Springer Optimization and Its Applications 110

Hoang Tuy

Convex Analysis and Global Optimization

Second Edition

Springer

Springer Optimization and Its Applications

VOLUME 110

Aims and Scope
Optimization has been expanding in all directions at an astonishing rate during the last few decades. New algorithmic and theoretical techniques have been developed, the diffusion into other disciplines has proceeded at a rapid pace, and our knowledge of all aspects of the field has grown even more profound. At the same time, one of the most striking trends in optimization is the constantly increasing emphasis on the interdisciplinary nature of the field. Optimization has been a basic tool in all areas of applied mathematics, engineering, medicine, economics, and other sciences.

The series *Springer Optimization and Its Applications* publishes undergraduate and graduate textbooks, monographs and state-of-the-art expository work that focus on algorithms for solving optimization problems and also study applications involving such problems. Some of the topics covered include nonlinear optimization (convex and nonconvex), network flow problems, stochastic optimization, optimal control, discrete optimization, multi-objective programming, description of software packages, approximation techniques and heuristic approaches.

More information about this series at http://www.springer.com/series/7393

Hoang Tuy

Convex Analysis and Global Optimization

Second Edition

Springer

Hoang Tuy
Vietnam Academy of Science
and Technology
Institute of Mathematics
Hanoi, Vietnam

ISSN 1931-6828 ISSN 1931-6836 (electronic)
Springer Optimization and Its Applications
ISBN 978-3-319-31482-2 ISBN 978-3-319-31484-6 (eBook)
DOI 10.1007/978-3-319-31484-6

Library of Congress Control Number: 2016934409

Mathematics Subject Classification (2010): 90-02, 49-02, 65K-10

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Preface to the First Edition

This book grew out of lecture notes from a course given at Graz University of Technology (autumn 1990), The Institute of Technology at Linköping University (autumn 1991), and the Ecole Polytechnique in Montréal (autumn 1993). Originally conceived of as a Ph.D. course, it attempts to present a coherent and mathematically rigorous theory of deterministic global optimization. At the same time, aiming at providing a concise account of the methods and algorithms, it focuses on the main ideas and concepts, leaving aside too technical details which might affect the clarity of presentation.

Global optimization is concerned with finding global solutions to nonconvex optimization problems. Although until recently convex analysis has been developed mainly under the impulse of convex and local optimization, it has become a basic tool for global optimization. The reason is that the general mathematical structure underlying virtually every nonconvex optimization problem can be described in terms of functions representable as differences of convex functions (dc functions) and sets which are differences of convex sets (dc sets). Due to this fact, many concepts and results from convex analysis play an essential role in the investigation of important classes of global optimization problems. Since, however, convexity in nonconvex optimization problems is present only partially or in the "other" way, new concepts have to be introduced and new questions have to be answered. Therefore, a new chapter on dc functions and dc sets is added to the traditional material of convex analysis. Part I of this book is an introduction to convex analysis interpreted in this broad sense as an indispensable tool for global optimization.

Part II presents a theory of deterministic global optimization which heavily relies on the dc structure of nonconvex optimization problems. The key subproblem in this approach is to transcend an incumbent, i.e., given a solution of an optimization problem (the best so far obtained), check its global optimality, and find a better solution, if there is one. As it turns out, this subproblem can always be reduced to solving a dc inclusion of the form $x \in D \setminus C$, where D, C are two convex sets. Chapters 4–6 are devoted to general methods for solving concave and dc programs through dc inclusions of this form. These methods include successive partitioning and cutting, outer approximation and polyhedral annexation, or combination of the concepts. The last two chapters discuss methods for exploiting special structures in global optimization. Two aspects of nonconvexity deserve particular attention: the rank of nonconvexity, i.e., roughly speaking the number of nonconvex variables, and the degree of nonconvexity, i.e., the extent to which a problem fails to be convex. Decomposition methods for handling low rank nonconvex problems are presented in Chap. 7, while nonconvex quadratic problems, i.e., problems involving only nonconvex functions which in a sense have lowest degree of nonconvexity, are discussed in the last Chap. 8.

I have made no attempt to cover all the developments to date because this would be merely impossible and unreasonable. With regret I have to omit many interesting results which do not fit in with the mainstream of the book. On the other hand, much new material is offered, even in the parts where the traditional approach is more or less standard.

I would like to express my sincere thanks to many colleagues and friends, especially R.E. Burkard (Graz Technical University), A. Migdalas and P. Värbrand (University of Linköping), and B. Jaumard and P. Hansen (University of Montréal), for many fruitful discussions we had during my enjoyable visits to their departments, and P.M. Pardalos for his encouragement in publishing this book. I am particularly grateful to P.T. Thach and F.A. Al-Khayyal for many useful remarks and suggestions and also to P.H. Dien for his efficient help during the preparation of the manuscript.

Hanoi, Vietnam
October 1997

Hoang Tuy

Preface to the Second Edition

Over the 17 years since the publication of the first edition of this book, tremendous progress has been achieved in global optimization. The present revised and enlarged edition is to give an up-to-date account of the field, in response to the needs of research, teaching, and applications in the years to come.

In addition to the correction of misprints and errors, much new material is offered. In particular, several recent important advances are included: a modern approach to minimax, fixed point and equilibrium, a robust approach to optimization under nonconvex constraints, and also a thorough study of quadratic programming with a single quadratic constraint.

Moreover, three new chapters are added: monotonic optimization (Chap. 11), polynomial optimization (Chap. 12), and optimization under equilibrium constraints (Chap. 13). These topics have received increased attention in recent years due to many important engineering and economic applications.

Hopefully, as main reference in deterministic global optimization, this book will replace both its first edition and the 1996 (last) edition of the book *Global Optimization: Deterministic Approaches* by Reiner Horst and Hoang Tuy.

Hanoi, Vietnam
September 2015

Hoang Tuy

Contents

Glossary of Notation

$\mathbb{R}$	The real line
$\mathbb{R}^n$	Real n-dimensional space
aff E	Affine hull of the set E
conv E	Convex hull of the set E
cl C	Closure of the set C
ri C	Relative interior of the set C
∂C	Relative boundary of a convex set C
rec(C)	Recession cone of a convex set C
$N_C(x^0)$	(Outward) normal cone of a convex set C at point $x^0 \in C$
E°	Polar of set E
$X^\natural$	Lower conjugate of set X
dom$f$	Effective domain of function $f(x)$
epi$f$	Epigraph of $f(x)$
$f^*(p)$	Conjugate of function $f(x)$
$f^\natural(p)$	Quasiconjugate of function $f(x)$
$\delta_C(x)$	Indicator function of C
$s_C(x)$	Support function of C
$d_C(x)$	Distance function of C
$\partial f(x^0)$	Subdifferential of f at x^0
$\partial_\varepsilon f(x^0)$	ε-subdifferential of f at x^0
$\partial^\natural f(x^0)$	Quasi-supdifferential of f at x^0
LMI	Linear matrix inequality
SDP	Semidefinite program
$DC(\Omega)$	Linear space formed by dc functions on Ω
BB	Branch and Bound
BRB	Branch Reduce and Bound
OA	Outer Approximation
PA	Polyhedral Annexation
CS	Cone Splitting
SIT	Successive Incumbent Transcending
BCP	Basic Concave Program

SBCP	Separable Basic Concave Program
GCP	General Concave Program
LRC	Linear Reverse Convex Program
CDC	Canonical DC Program
COP	Continuous Optimization Problem
MO	Canonical Monotonic Optimization Problem
MPEC	Mathematical Programming with Equilibrium Constraints
MPAEC	MPEC with an Affine Variational Inequality
VI	Variational Inequality
OVI	Optimization with Variational Inequality constraint
GBP	General Bilevel Programming
CBP	Convex Bilevel Programming
LBP	Linear Bilevel Programming
OE	Optimization over Efficient Set

Part I
Convex Analysis

Chapter 1
Convex Sets

1.1 Affine Sets

Let a, b be two points of $\mathbb{R}^n$. The set of all $x \in \mathbb{R}^n$ of the form

$$x = (1 - \lambda)a + \lambda b = a + \lambda(b - a), \quad \lambda \in \mathbb{R} \tag{1.1}$$

is called the *line through a and b*. A subset M of $\mathbb{R}^n$ is called an *affine set* (or *affine manifold*) if it contains every line through two points of it, i.e., if $(1-\lambda)a+\lambda b \in M$ for every $a \in M, b \in M$, and every $\lambda \in \mathbb{R}$. An affine set which contains the origin is a subspace.

Proposition 1.1 *A nonempty set M is an affine set if and only if $M = a + L$, where $a \in M$ and L is a subspace.*

Proof Let M be an affine set and $a \in M$. Then $M = a + L$, where the set $L = M - a$ contains 0 and is an affine set, hence is a subspace. Conversely, let $M = a + L$, with $a \in M$ and L a subspace. For any $x, y \in M, \lambda \in \mathbb{R}$ we have $(1 - \lambda)x + \lambda y = a + (1 - \lambda)(x - a) + \lambda(y - a)$. Since $x - a, y - a \in L$ and L is a subspace, it follows that $(1 - \lambda)(x - a) + \lambda(y - a) \in L$, hence $(1 - \lambda)x + \lambda y \in M$. Therefore, M is an affine set. □

The subspace L is said to be *parallel* to the affine set M. For a given nonempty affine set M the parallel subspace L is uniquely defined. Indeed, if $M = a' + L'$, then $L' = M - a' = L + (a - a')$, but $a' = a' + 0 \in M = a + L$, hence $a' - a \in L$, and so $L' = L + (a - a') = L$.

The dimension of the subspace L parallel to an affine set M is called the *dimension* of M. A point $a \in \mathbb{R}^n$ is an affine set of dimension 0 because the subspace parallel to $M = \{a\}$ is $L = \{0\}$. A line through two points a, b is an affine set of dimension 1, because the subspace parallel to it is the one-dimensional subspace

H. Tuy, *Convex Analysis and Global Optimization*, Springer Optimization and Its Applications 110, DOI 10.1007/978-3-319-31484-6_1

$\{x = \lambda(b - a)|\ \lambda \in \mathbb{R}\}$. An $(n-1)$-dimensional affine set is called a *hyperplane*, or a *plane* for short.

Proposition 1.2 *Any r-dimensional affine set is of the form*

$$M = \{x|\ Ax = b\},$$

where $b \in \mathbb{R}^m, A \in \mathbb{R}^{m\times n}$ *such that rank*$A = n - r$. *Conversely, any set of the above form is an r-dimensional affine set.*

Proof Let M be an r-dimensional affine set. For any $x^0 \in M$, the set $L = M - x^0$ is an r-dimensional subspace, so $L = \{x|\ Ax = 0\}$, where A is some $m \times n$ matrix of rank $n - r$. Hence $M = \{x|\ A(x - x^0) = 0\} = \{x|\ Ax = b\}$, with $b = Ax^0$. Conversely if M is a set of this form then, for any $x^0 \in M$, we have $Ax^0 = b$, hence $M = \{x|A(x - x^0) = 0\} = x^0 + L$, with $L = \{x|\ Ax = 0\}$. Since rank$A = n - r$, L is an r-dimensional subspace. □

Corollary 1.1 *Any hyperplane is a set of the form*

$$H = \{x|\ \langle a, x\rangle = \alpha\}, \tag{1.2}$$

where $a \in \mathbb{R}^n \setminus \{0\}, \alpha \in \mathbb{R}$. *Conversely, any set of the above form is a hyperplane.*

The vector a in (1.2) is called the *normal* to the hyperplane H. Of course a and α are defined up to a nonzero common multiple, for a given hyperplane H. The line $\{\lambda a|\ \lambda \in \mathbb{R}\}$ meets H at a point λa such that $\langle a, \lambda a\rangle = \alpha$, hence $\lambda = \frac{\alpha}{\|a\|^2}$. Thus, the distance from 0 to a hyperplane $\langle a, x\rangle = \alpha$ is equal to $|\lambda|\|a\| = \frac{|\alpha|}{\|a\|}$.

The intersection of a family of affine sets is again an affine set. Given a set $E \subset \mathbb{R}^n$, there exists at least an affine set containing E, namely $\mathbb{R}^n$. Therefore, the intersection of all affine sets containing E is the smallest affine set containing E. This set is called the *affine hull* of E and denoted by aff E.

Proposition 1.3 *The affine hull of a set E is the set consisting of all points of the form*

$$x = \lambda_1 x^1 + \ldots + \lambda_k x^k,$$

such that $x^i \in E,\ \lambda_1 + \ldots + \lambda_k = 1$, *and k a natural number.*

Proof Let M be the set of all points of the above form. If $x, y \in M$, for example, $x = \sum_{i=1}^k \lambda_i x^i,\ y = \sum_{j=1}^h \mu_j y^j$, with $x^i, y^j \in E,\ \sum_{i=1}^k \lambda_i = \sum_{j=1}^h \mu_j = 1$, then for any $\alpha \in \mathbb{R}$ we can write

$$(1-\alpha)x + \alpha y = \sum_{i=1}^k (1-\alpha)\lambda_i x^i + \sum_{j=1}^h \alpha\mu_j y^j$$

with $\sum_{i=1}^k (1-\alpha)\lambda_i + \sum_{j=1}^h \alpha\mu_j = (1-\alpha)+\alpha = 1$, hence $(1-\alpha)x + \alpha y \in M$. Thus M is affine, and hence $M \supset \text{aff}E$. On the other hand, it is easily seen that if $x = \sum_{i=1}^k \lambda_i x^i$ with $x^i \in E$, $\sum_{i=1}^k \lambda_i = 1$, then $x \in \text{aff}E$ (for $k = 2$ this follows from the definition of affine sets, for $k \geq 3$ this follows by induction). So $M \subset \text{aff}E$, and therefore, $M = \text{aff}E$. □

Proposition 1.4 *The affine hull of a set of k points $x^1, \ldots, x^k$ $(k > r)$ in $\mathbb{R}^n$ is of dimension r if and only if the $(n+1) \times k$ matrix*

$$\begin{bmatrix} x^1 & x^2 & \cdots & x^k \\ 1 & 1 & \cdots & 1 \end{bmatrix} \tag{1.3}$$

is of rank $r+1$.

Proof Let $M = \text{aff}\{x^1, \ldots, x^k\}$. Then $M = x^k + L$, where L is the smallest space containing $x^1 - x^k, \ldots, x^{k-1} - x^k$. So M is of dimension r if and only if L is of dimension r, i.e., if and only if the matrix $[x^1 - x^k, \ldots, x^{k-1} - x^k]$ is of rank r. This in turn is equivalent to requiring that the matrix (1.3) be of rank $r+1$. □

We say that k points $x^1, \ldots, x^k$ are *affinely independent* if $\text{aff}\{x^1, \ldots, x^k\}$ has dimension $k-1$, i.e., if the vectors $x^1 - x^k, \ \ldots, x^{k-1} - x^k$ are linearly independent, or equivalently, the matrix (1.3) is of rank k.

Corollary 1.2 *The affine hull S of a set of k affinely independent points $\{x^1, \ldots, x^k\}$ in $\mathbb{R}^n$ is a $(k-1)$-dimensional affine set. Every point $x \in S$ can be uniquely represented in the form*

$$x = \sum_{i=1}^k \lambda_i x^i, \quad \sum_{i=1}^k \lambda_i = 1. \tag{1.4}$$

Proof By Proposition 1.3 any point $x \in S$ has the form (1.4). If $x = \sum_{i=1}^k \mu_i x^i$ is another representation, then $\sum_{i=1}^{k-1} (\lambda_i - \mu_i)(x^i - x^k) = 0$, hence $\lambda_i = \mu_i$ $\forall i = 1, \ldots, k$. □

Corollary 1.3 *There is a unique hyperplane passing through n affinely independent points $x^1, x^2, \ldots, x^n$ in $\mathbb{R}^n$. Every point of this hyperplane can be uniquely represented in the form*

$$x = \sum_{i=1}^n \lambda_i x^i, \quad \sum_{i=1}^n \lambda_i = 1.$$

1.2 Convex Sets

Given two points $a, b \in \mathbb{R}^n$, the set of all points $x = (1-\lambda)a + \lambda b$ such that $0 \leq \lambda \leq 1$ is called the (closed) *line segment* between a and b and denoted by $[a, b]$.

A set $C \subset \mathbb{R}^n$ is called *convex* if it contains any line segment between two points of it; in other words, if $(1-\lambda)a + \lambda b \in C$ whenever $a, b \in C,\ 0 \le \lambda \le 1$.

Affine sets are obviously convex. Other particular examples of convex sets are halfspaces which are sets of the form

$$\{x|\ \langle a, x\rangle \le \alpha\}, \quad \{x|\ \langle a, x\rangle \ge \alpha\}$$

(*closed halfspaces*), or

$$\{x|\ \langle a, x\rangle < \alpha\}, \quad \{x|\ \langle a, x\rangle > \alpha\}$$

(*open halfspaces*), where $a \in \mathbb{R}^n, \alpha \in \mathbb{R}$.

A point x such that $x = \sum_{i=1}^k \lambda_i a^i$ with $a^i \in \mathbb{R}^n,\ \lambda_i \ge 0, \sum_{i=1}^k \lambda_i = 1$, is called a *convex combination* of $a^1, \ldots, a^k \in \mathbb{R}^n$.

Proposition 1.5 *A set $C \subset \mathbb{R}^n$ is convex if and only if it contains all the convex combinations of its elements.*

Proof It suffices to show that if C is a convex set then

$$a^i \in C, i = 1, \ldots, k,\ \lambda_i > 0,\ \sum_{i=1}^k \lambda_i = 1 \Rightarrow \sum_{i=1}^k \lambda_i a^i \in C.$$

Indeed, this is obvious for $k = 2$. Assuming this is true for $k = h - 1 \ge 2$ and letting $\alpha = \sum_{i=1}^{h-1} \lambda_i > 0$ we can write

$$\sum_{i=1}^h \lambda_i a^i = \sum_{i=1}^{h-1} \lambda_i a^i + \lambda_h a^h = \alpha \sum_{i=1}^{h-1} \frac{\lambda_i}{\alpha} a^i + \lambda_h a^h.$$

Since $\sum_{i=1}^{h-1}(\lambda_i/\alpha) = 1$, by the induction hypothesis $y = \sum_{i=1}^{h-1}(\lambda_i/\alpha)a^i \in C$. Next, since $\alpha + \lambda_h = 1$, we have $\alpha y + \lambda_h a^h \in C$. Thus the above property holds for every natural k. □

Proposition 1.6 *The intersection of any collection of convex sets is convex. If C, D are convex sets then $C + D := \{x + y|\ x \in C, y \in D\}$, $\alpha C = \{\alpha x|\ x \in C\}$ (and hence, also $C - D := C + (-1)D$) are convex sets.*

Proof If $\{C_\alpha\}$ is a collection of convex sets, and $a, b \in \cap_\alpha C_\alpha$, then for each α we have $a \in C_\alpha, b \in C_\alpha$, therefore $[a, b] \subset C_\alpha$ and hence, $[a, b] \subset \cap_\alpha C_\alpha$.

If C, D are convex sets, and $a = x + y, b = u + v$ with $x, u \in C, y, v \in D$, then $(1-\lambda)a + \lambda b = [(1-\lambda)x + \lambda u] + [(1-\lambda)y + \lambda v] \in C + D$, for any $\lambda \in [0, 1]$, hence $C + D$ is convex. The convexity of αC can be proved analogously. □

A subset M of $\mathbb{R}^n$ is called a *cone* if

$$x \in M,\ \lambda > 0 \Rightarrow \lambda x \in M.$$

The origin 0 itself may or may not belong to M. A set $a+M$, which is the translate of a cone M by $a \in \mathbb{R}^n$, is also called a cone with *apex* at a. A cone M which contains no line is said to be *pointed*; in this case, 0 is also called the *vertex* of M, and $a+M$ is a cone with vertex at a.

Proposition 1.7 *A subset M of $\mathbb{R}^n$ is a convex cone if and only if*

$$\lambda M \subset M \quad \forall \lambda > 0 \tag{1.5}$$

$$M + M \subset M. \tag{1.6}$$

Proof If M is a convex cone then (1.5) holds by definition of a cone; furthermore, from $x, y \in M$ we have $\frac{1}{2}(x+y) \in M$, hence by (1.5), $x+y \in M$. Conversely, if (1.5) and (1.6) hold, then M is a cone, and for any $x, y \in M, \lambda \in (0,1)$ we have $(1-\lambda)x \in M, \lambda y \in M$, hence $(1-\lambda)x + \lambda y \in M$, so M is convex. □

Corollary 1.4 *A subset M of $\mathbb{R}^n$ is a convex cone if and only if it contains all the positive linear combinations of its elements.*

Proof Properties (1.6) and (1.5) mean that M is closed under addition and positive scalar multiplication, i.e., M contains all elements of the form $\sum_{i=1}^k \lambda_i a^i$, with $a^i \in M, \lambda_i > 0$. □

Corollary 1.5 *Let E be a convex set. The set $\{\lambda x |\ x \in E, \lambda > 0\}$ is the smallest convex cone which contains E.*

Proof Since any cone containing E must contain this set it suffices to show that this set is a convex cone. But this follows from the preceding corollary and the fact that $\sum_i \alpha_i x^i = \lambda \sum_i (\alpha_i/\lambda) x^i$, with $\lambda = \sum_i \alpha_i$. □

The convex cone obtained by adjoining 0 to the cone mentioned in Corollary 1.5 is often referred to as the cone *generated* by E and is denoted by cone E.

1.3 Relative Interior and Closure

A *norm* in $\mathbb{R}^n$ is by definition a mapping $\|\cdot\|$ from $\mathbb{R}^n$ to $\mathbb{R}$ which satisfies

$$\begin{array}{l} \|x\| \geq 0 \quad \forall x \in \mathbb{R}^n; \quad \|x\| = 0 \ \text{ only if } x = 0 \\ \|\alpha x\| = |\alpha| \|x\| \quad \forall x \in \mathbb{R}^n,\ \alpha \in \mathbb{R} \qquad\qquad 1 \\ \|x+y\| \leq \|x\| + \|y\| \quad x, y \in \mathbb{R}^n. \end{array} \tag{1.7}$$

Familiar examples of norms in $\mathbb{R}^n$ are the l_p-norms

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}, \quad 1 \leq p < \infty,$$

and the l_∞-norm

$$\|x\|_\infty = \max_{1\le i\le n} |x_i|.$$

The l_2-norm, which can also be defined via the inner product by means of $\|x\| = \sqrt{\langle x, x\rangle}$, is also called the *Euclidean norm*. The l_∞-norm is also called the *Tchebycheff norm*.

Given a norm $\|\cdot\|$ in $\mathbb{R}^n$, the *distance* between two points $x, y \in \mathbb{R}^n$ is the nonnegative number $\|x-y\|$; the *ball* of center a and radius r is the set $\{x \in \mathbb{R}^n | \ \|x-a\| \le r\}$.

Topological concepts in $\mathbb{R}^n$ such as open set, closed set, interior, closure, neighborhood, convergent sequence, etc., are defined in $\mathbb{R}^n$ with respect to a given norm. However, the following proposition shows that the topologies in $\mathbb{R}^n$ defined by different norms are all equivalent.

Proposition 1.8 *For any two norms $\|\cdot\|, \|\cdot\|'$ in $\mathbb{R}^n$ there exist constants $c_2 \ge c_1 > 0$ such that*

$$c_1\|x\| \le \|x\|' \le c_2\|x\| \quad \forall x \in \mathbb{R}^n.$$

We omit the proof which can be found, e.g., in Ortega and Rheinboldt (1970).

Denote the closure and the interior of a set C by cl$C$ and int$C$, respectively. Recall that $a \in \text{cl}C$ if and only if every ball around a contains at least one point of C, or equivalently, a is the limit of a sequence of points of C; $a \in \text{int}C$ if and only if there is a ball around a entirely included in C. An immediate consequence of Proposition 1.8 is the following alternative characterization of interior points of convex sets in $\mathbb{R}^n$.

Corollary 1.6 *A point a of a convex set $C \subset \mathbb{R}^n$ is an interior point of C if for every $x \in \mathbb{R}^n$ there exists $\alpha > 0$ such that $a + \alpha(x - a) \in C$.*

Proof Without loss of generality we can assume that $a = 0$. Let e^i be the i-unit vector of $\mathbb{R}^n$. The condition implies that for each $i = 1, \dots, n$ there is $\alpha_i > 0$ such that $\alpha_i e^i \in C$ and $\beta_i > 0$ such that $-\beta_i e^i \in C$. Now let $\alpha = \min\{\alpha_i, \beta_i, i = 1, \dots, n\}$ and consider the set $B = \{x| \ \|x_i\|_1 \le \alpha\}$. For any $x \in B$, since $x = \sum_{i=1}^n x_i e^i$ with $\sum_{i=1}^n \frac{|x_i|}{\alpha} \le 1$, if we define $u^i = \alpha e^i$ for $x_i > 0$ and $u^i = -\alpha e^i$ for $x_i \le 0$ then $u^1, \dots, u^n \in C$ and we can write $x = \sum_{i=1}^n \frac{|x_i|}{\alpha} u^i + (1 - \sum_{i=1}^n \frac{|x_i|}{\alpha})0$. From the convexity of C it follows that $x \in C$. Thus, $B \subset C$, and since B is a ball in the l_1-norm, 0 is an interior point of C. □

A point a belongs to the *relative interior* of a convex set $C \subset \mathbb{R}^n$: $a \in \text{ri}C$, if it is an interior point of C relative to aff$C$. The set difference (cl$C$) \ (ri$C$) is called the *relative boundary* of C and denoted by ∂C.

Proposition 1.9 *Any nonempty convex set $C \subset \mathbb{R}^n$ has a nonempty relative interior.*

Proof Let $M = \text{aff}C$ with $\dim M = r$, so there exist $r+1$ affinely independent elements $x^1, \ldots, x^{r+1}$ of C. We show that $a = \frac{1}{r+1}\sum_{i=1}^{r+1} x^i$ is a relative interior of C. Indeed, any $x \in M$ is of the form $x = \sum_{i=1}^{r+1} \lambda_i x^i$, with $\sum_{i=1}^{r+1} \lambda_i = 1$, so $a + \alpha(x-a) = \sum_{i=1}^{r+1} \mu_i x^i$ with

$$\mu_i = (1-\alpha)\frac{1}{r+1} + \alpha\lambda_i, \quad \sum_{i=1}^{r+1} \mu_i = 1.$$

For $\alpha > 0$ sufficiently small, we have $\mu_i > 0, i = 1, \ldots, r+1$, hence $a + \alpha(x-a) \in C$. Therefore, by Corollary 1.6, $a \in \text{ri}C$. □

The *dimension* of a convex set is by definition the dimension of its affine hull. A convex set C in $\mathbb{R}^n$ is said to have *full dimension* if $\dim C = n$. It follows from the above Proposition that a convex set C in $\mathbb{R}^n$ has a nonempty interior if and only if it has full dimension.

Proposition 1.10 *The closure and the relative interior of a convex set are convex.*

Proof Let C be a convex set, $a, b \in \text{cl}C$, say $a = \lim_{\nu\to\infty} x^\nu, b = \lim_{\nu\to\infty} y^\nu$, where $x^\nu, y^\nu \in C$ for every ν. For every $\lambda \in [0,1]$, we have $(1-\lambda)x^\nu + \lambda y^\nu \in C$, hence, $(1-\lambda)a + \lambda b = \lim[(1-\lambda)x^\nu + \lambda y^\nu] \in \text{cl}C$. Thus, if $a, b \in \text{cl}C$ then $[a,b] \subset \text{cl}C$, proving the convexity of $\text{cl}C$. Now let $a, b \in \text{ri}C$, so that there is a ball B around 0 such that $(a+B) \cap \text{aff}C$ and $(b+B) \cap \text{aff}C$ are contained in C. For any $x = (1-\lambda)a + \lambda b$, with $\lambda \in (0,1)$ we have $(x+B) \cap \text{aff}C = (1-\lambda)(a+B) \cap \text{aff}C + \lambda(b+B) \cap \text{aff}C \subset C$, so $x \in \text{ri}C$. This proves the convexity of $\text{ri}C$. □

Let $C \subset \mathbb{R}^n$ be a convex set containing the origin. The function $p_C : \mathbb{R}^n \to \mathbb{R}$ such that

$$p_C(x) = \inf\{\lambda > 0 |\, x \in \lambda C\}$$

is called the *gauge* or *Minkowski functional* of C.

Proposition 1.11 *The gauge $p_C(x)$ of a convex set $C \subset \mathbb{R}^n$ containing 0 in its interior satisfies*

(i) $p_C(\alpha x) = \alpha p_C(x) \quad \forall x \in \mathbb{R}^n,\ \alpha \geq 0$ *(positive homogeneity),*
(ii) $p_C(x+y) \leq p_C(x) + p_C(y) \quad \forall x, y \in C$ *(subadditivity),*
(iii) $p_C(x)$ *is continuous,*
(iv) $\text{int}C = \{x |\, p_C(x) < 1\} \subset C \subset \text{cl}C = \{x |\, p_C(x) \leq 1\}$.

Proof

(i) Since $0 \in \text{int}C$ it is easy to see that $p_C(x)$ is finite everywhere. For every $\alpha \geq 0$ we have $p_C(\alpha x) = \inf\{\lambda > 0 |\, \alpha x \in \lambda C\} = \alpha \inf\{\mu > 0 |\, x \in \mu C\} = \alpha p_C(x)$.
(ii) If $x \in \lambda C, y \in \mu C$ $(\lambda, \mu > 0)$, i.e., $x = \lambda x',\ y = \mu y'$ with $x', y' \in C$, then

$$x + y = (\lambda + \mu)[\frac{\lambda}{\lambda + \mu}x' + \frac{\mu}{\lambda + \mu}y'] \in (\lambda + \mu)C.$$

Thus, $p_C(x + y) \le \lambda + \mu$ for all $\lambda, \mu > 0$ such that $x \in \lambda C,\ y \in \mu C$. Hence $p_C(x + y) \le p_C(x) + p_C(y)$.

(iii) Consider any $x^0 \in \mathbb{R}^n$ and let $x - x^0 \in B_\delta := \{u |\ \|u\| \le \delta\}$, where $\delta > 0$ is so small that the ball B_δ is contained in εC. Then, $x - x^0 \in \varepsilon C$, consequently $p_C(x - x^0) \le \varepsilon$, hence $p_C(x) \le p_C(x^0) + p_C(x - x^0) \le p_C(x^0) + \varepsilon$; on the other hand, $x^0 - x \in \varepsilon C$ implies similarly $p_C(x^0) \le p_C(x) + \varepsilon$. Thus, $|p_C(x) - p_C(x^0)| \le \varepsilon$ whenever $x - x^0 \in B_\delta$, proving the continuity of p_C at any $x^0 \in \mathbb{R}^n$.

(iv) This follows from the continuity of $p_C(x)$. □

Corollary 1.7 *If* $a \in \text{int}C, b \in \text{cl}C$, *then* $x = (1-\lambda)a + \lambda b \in \text{int}C$ *for all* $\lambda \in [0, 1)$. *If a convex set C has nonempty interior then* cl(int$C$) = cl$C$, int(cl$C$) = int$C$.

Proof By translating we can assume that $0 \in \text{int}C$, so $\text{int}C = \{x |\ p_C(x) < 1\}$. Then $p_C(x) \le (1 - \lambda)p_C(a) + \lambda p_C(b) < 1$ whenever $p_C(a) < 1, p_C(b) \le 1$, and $0 \le \lambda < 1$. The second assertion is a straightforward consequence of (iv). □

Let C be a convex set in $\mathbb{R}^n$. A vector $y \neq 0$ is called a *direction of recession* of C if

$$\{x + \lambda y |\ \lambda \ge 0\} \subset C \quad \forall x \in C. \tag{1.8}$$

Lemma 1.1 *The set of all directions of recession of C is a convex cone. If C is closed then (1.8) holds provided* $\{x + \lambda y |\ \lambda \ge 0\} \subset C$ *for one* $x \in C$.

Proof The first assertion can easily be verified. If C is closed, and $\{x + \lambda y |\ \lambda \ge 0\} \subset C$ for one $x \in C$ then for any $x' \in C, \mu \ge 0$ we have $x' + \mu y \in C$ because $x' + \mu y = \lim_{\lambda \to +\infty} z_\lambda$, where $z_\lambda = \left(1 - \frac{\mu}{\lambda+\mu}\right)x' + \frac{\mu}{\lambda+\mu}(x + \lambda y) \in C$. □

The convex cone formed by all directions of recession and the vector zero is called the *recession cone* of C and is denoted by rec$C$.

Proposition 1.12 *Let* $C \subset \mathbb{R}^n$ *be a convex set containing a point a in its interior.*

(i) *For every* $x \neq a$ *the halfline* $\Gamma(a, x) = \{a + \lambda(x - a) |\ \lambda > 0\}$ *either entirely lies in C or it cuts the boundary* ∂C *at a unique point* $\sigma(x)$, *such that every point in the line segment* $[a, \sigma(x)]$, *except* $\sigma(x)$, *is an interior point of C.*

(ii) *The mapping* $\sigma(x)$ *is defined and continuous on the set* $\mathbb{R}^n \setminus (a + M)$, *where* $M = \text{rec}(\text{int}C) = \text{rec}(\text{cl}C)$ *is a closed convex cone.*

Proof

(i) Assuming, without loss of generality, that $0 \in \text{int}C$, it suffices to prove the proposition for $a = 0$. Let $p_C(x)$ be the gauge of C. For every $x \neq a$, if $p_C(x) = 0$ then $x \in \mu C$ for all arbitrarily small $\mu > 0$, i.e., $\lambda x \in C$ for all $\lambda > 0$, hence the halfline $\Gamma(a, x)$ entirely lies in C. If $p_C(x) > 0$ then $p_C(\lambda x) = \lambda p_C(x) = 1$

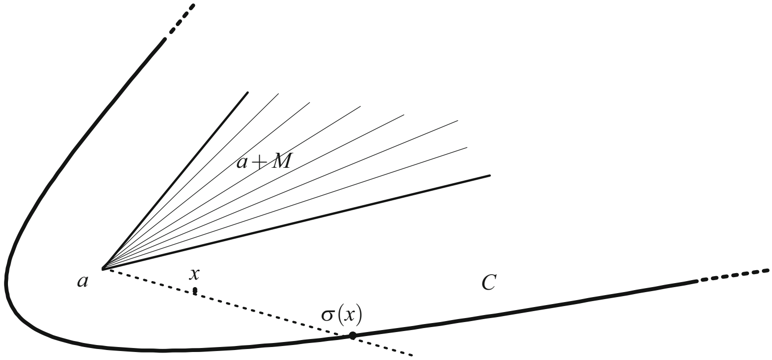


Fig. 1.1 The mapping $\sigma(x)$

just for $\lambda = 1/p_C(x)$. So $\Gamma(a,x)$ cuts ∂C at a unique point $\sigma(x) = x/p_C(x)$, such that for all $0 \le \lambda < 1/p_C(x)$ we have $p_C(\lambda x) < 1$ and hence $\lambda x \in \mathrm{int}C$.

(ii) The domain of definition of σ is $\mathbb{R}^n \setminus M$, with $M = \{x |\, p_C(x) = 0\}$. The set M is a closed convex cone (Fig. 1.1) because $p_C(x)$ is continuous and $p_C(x) = 0$, $p_C(y) = 0$ implies $0 \le p_C(\lambda x + \mu y) \le \lambda p_C(x) + \mu p_C(y) = 0$ for all $\lambda > 0$ and $\mu > 0$. Furthermore, the function $\sigma(x) = x/p_C(x)$ is continuous at any x where $p_C(x) > 0$. It remains to prove that M is the recession cone of both int$C$ and cl$C$. It is plain that $\mathrm{rec}(\mathrm{int}C) \subset M$. Conversely, if $y \in M$ and $x \in \mathrm{int}C$ then $x = \alpha a + (1-\alpha)c$, for some $c \in C$ and $0 < \alpha < 1$. Since $a + \lambda y \in C$ for all $\lambda > 0$, it follows that $x + \lambda y = \alpha a + (1-\alpha)c + \lambda y = \alpha(a + \frac{\lambda}{\alpha}y) + (1-\alpha)c \in C$ for all $\lambda > 0$, i.e., $y \in \mathrm{rec}(\mathrm{int}C)$. Thus, $M \subset \mathrm{rec}(\mathrm{int}C)$. Finally, any $x \in \partial C$ is the limit of some sequence $\{x^k\} \subset \mathrm{int}C$. If $y \in M$ then for every $\lambda > 0$: $x^k + \lambda y \in C,\ \forall k$, hence by letting $k \to \infty$, $x + \lambda y \in \mathrm{cl}C$. This means that $M \subset \mathrm{rec}(\mathrm{cl}C)$, and since the converse inclusion is obvious we must have $M = \mathrm{rec}(\mathrm{cl}C)$. □

Note that if C is neither closed nor open then rec$C$ may be a proper subset of $\mathrm{rec}(\mathrm{int}C) = \mathrm{rec}(\mathrm{cl}C)$. An example is the set $C = \{(x_1, x_2) |\, x_1 > 0,\ x_2 > 0\} \cup \{(0,0)\}$, for which $\mathrm{rec}C = C,\ \mathrm{rec}(\mathrm{int}C) = \mathrm{cl}C$.

Corollary 1.8 *A nonempty closed convex set $C \subset \mathbb{R}^n$ is bounded if and only if its recession cone is the singleton $\{0\}$.*

Proof By considering C in aff$C$, we can assume that C has an interior point (Proposition 1.6) and that $0 \in \mathrm{int}C$. If C is bounded then by Proposition 1.12, $\sigma(x)$ is defined on $\mathbb{R}^n \setminus \{0\}$, i.e., $\mathrm{rec(int)}C = \{0\}$. Conversely, if $\mathrm{rec(int)}C = \{0\}$, then $p_C(x) > 0\ \forall x \neq 0$. Letting $K = \min\{p_C(x) |\ \|x\| = 1\}$, we then have, for every $x \in C:\ p_C(x) = \|x\| p_C\left(\frac{x}{\|x\|}\right) \le 1$, hence, $\|x\| \le \dfrac{1}{p_C\left(\frac{x}{\|x\|}\right)} \le \dfrac{1}{K}$. □

1.4 Convex Hull

Given any set $E \subset \mathbb{R}^n$, there exists a convex set containing E, namely $\mathbb{R}^n$. The intersection of all convex sets containing E is called the *convex hull* of E and denoted by conv$E$. This is in fact the smallest convex set containing E.

Proposition 1.13 *The convex hull of a set $E \subset \mathbb{R}^n$ consists of all convex combinations of its elements.*

Proof Let C be the set of all convex combinations of E. Clearly $C \subset \text{conv}E$. To prove the converse inclusion, it suffices to show that C is convex (since obviously $C \supset E$). If $x = \sum_{i\in I} \lambda_i a^i,\ y = \sum_{j\in J} \mu_j b^j$ with $a^i, b^j \in E$ and $0 \le \alpha \le 1$ then

$$(1-\alpha)x + \alpha y = \sum_{i\in I}(1-\alpha)\lambda_i a^i + \sum_{j\in J}\alpha\mu_j b^j.$$

Since

$$\begin{aligned}&\sum_{i\in I}(1-\alpha)\lambda_i + \sum_{j\in J}\alpha\mu_j \\ &\quad = (1-\alpha)\sum_{i\in I}\lambda_i + \alpha\sum_{j\in J}\mu_j = (1-\alpha)+\alpha = 1,\end{aligned}$$

it follows that $(1-\alpha)x + \alpha y \in C$. Therefore, C is convex. □

The convex hull S of a set of k affinely independent points $a^1, \dots, a^k$ in $\mathbb{R}^n$, written $S = [a^1, \dots, a^k]$, is called a $(k-1)$-*simplex* spanned by $a^1, \dots, a^k$. By Corollary 1.2 this is a $(k-1)$-dimensional set every point of which is uniquely represented as a convex combination of $a^1, \dots, a^k$:

$$x = \sum_{i=1}^{k}\lambda_i a^i, \quad \text{with} \quad \sum_{i=1}^{k}\lambda_i = 1,\ \ \lambda_i \ge 0,\ i = 1, \dots, k.$$

The numbers $\lambda_i = \lambda_i(x), i = 1, \dots, \lambda_k$ are called the *barycentric coordinates* of x in the simplex S. For any two points x, x' in S, since

$$\|\lambda(x) - \lambda(x')\| \le \left\|\sum_{i=1}^{k}\lambda_i(x) - \sum_{i=1}^{k}\lambda_i(x')\right\| = \|x - x'\|$$

the 1-1 map $x \mapsto \lambda(x)$ is continuous.

Proposition 1.13 shows that any point $x \in \text{conv}E$ can be expressed as a convex combination of a finite number of points of E. The next proposition which gives an upper bound of the minimal number of points in such expressions is a fundamental dimensionality result in convexity theory.

Theorem 1.1 (Caratheodory's Theorem) *Let E be a set contained in an affine set of dimension k. Then any vector $x \in$ conv$E$ can be represented as a convex combination of $k+1$ or fewer elements of E.*

Proof By Proposition 1.13, for any $x \in \text{conv}E$ there exists a finite subset of E, say $a^1, \ldots, a^m \in E$, such that

$$\sum_{i=1}^{m} \lambda_i a^i = x, \ \sum_{i=1}^{m} \lambda_i = 1, \ \lambda_i \geq 0.$$

If $\lambda^* = (\lambda_1^*, \ldots, \lambda_m^*)$ is a basic solution (in the sense of linear programming theory) of the above linear system in $\lambda_1, \ldots, \lambda_m$, then at most $k+1$ components of λ^* are positive. For example, $\lambda_i^* = 0$ for all $i > h$ where $h \leq k+1$, i.e., x can be represented as a convex combination of $a^1, \ldots, a^h$, with $h \leq k+1$. □

Corollary 1.9 *The convex hull of a compact set is compact.*

Proof Let E be a compact set and $\{x^k\} \subset \text{conv}E$. By the above theorem, $x^k = \sum_{i=1}^{n+1} \lambda_{i,k} a^{i,k}$, with $a^{i,k} \in E, \ \lambda_{i,k} \geq 0, \ \sum_{i=1}^{n+1} \lambda_{i,k} = 1$. Since $E \times [0,1]$ is compact, there exists a subsequence $\{k_\nu\}$ such that

$$\lambda_{i,k_\nu} a^{i,k_\nu} \to \mu_i a^i, \ a^i \in E, \ \mu_i \in [0,1].$$

We then have $x^{k_\nu} \to \sum_{i=1}^{n+1} \mu_i a^i \in \text{conv}E$, proving the corollary. □

Remark 1.1 The convex hull of a closed set E may not be closed, as shown by the example of a set $E \subset \mathbb{R}^2$ consisting of a straightline L and a point $a \notin L$ (Fig. 1.2).

Theorem 1.2 (Shapley–Folkman's Theorem) *Let* $\{S_i, i = 1, \ldots, m\}$ *be any finite family of sets in* $\mathbb{R}^n$. *For every* $x \in \text{conv}\left(\sum_{i=1}^{m} S_i\right)$ *there exists a subfamily* $\{S_i, i \in I\}$, *with* $|I| \leq n$, *such that*

$$x \in \sum_{i \notin I} S_i + \sum_{i \in I} \text{conv}(S_i).$$

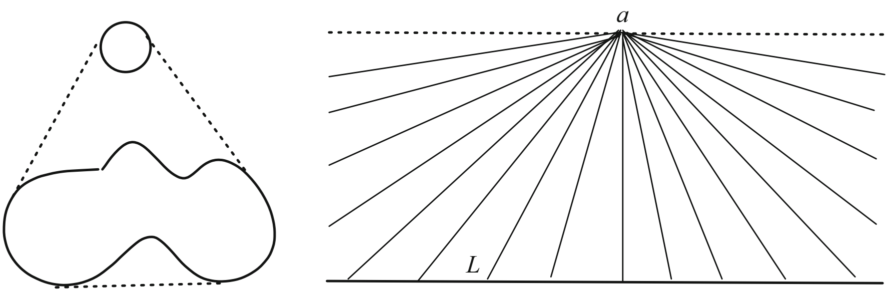


Fig. 1.2 Convex hull of a closed set

Roughly speaking, if m is very large and every S_i is sufficiently small then the sum $\sum_{i=1}^m S_i$ differs little from its convex hull.

Proof Observe that $\text{conv}(\sum_{i=1}^m S_i) = \sum_{i=1}^m \text{conv} S_i$. Indeed, define the linear mapping $\varphi : \mathbb{R}^{n\times m} \to \mathbb{R}^n$ such that $\varphi((x^1,\dots,x^m)) = \sum_{i=1}^m x^i$. Then $\sum_{i=1}^m S_i = \varphi(\prod_{i=1}^m S_i)$ and by linearity of φ :

$$\text{conv}\left(\varphi\left(\prod_{i=1}^m S_i\right)\right) = \varphi\left(\text{conv}\prod_{i=1}^m S_i\right) = \varphi\left(\prod_{i=1}^m \text{conv} S_i\right),$$

proving our assertion. So any $x \in \text{conv}(\sum_{i=1}^m S_i)$ is of the form $x = \sum_{i=1}^m x^i$, with $x^i \in \text{conv} S_i$. Let x^i be a convex combination of $x^{ij} \in S_i, j \in J_i$. Then there exist λ_{ij} satisfying

$$\sum_{i=1}^m \sum_{j\in J_i} \lambda_{ij} x^{ij} = x$$

$$\sum_{j\in J_i} \lambda_{ij} = 1 \quad i = 1,\dots,m.$$

$$\lambda_{ij} \geq 0 \quad i = 1,\dots,m,\ j \in J_i.$$

Let $\lambda^* = (\lambda^*_{ij})$ be a basic solution of this linear system in λ_{ij}, so that at most $n+m$ components of λ^* are positive. Denote by I the set of all i such that $\lambda^*_{ij} > 0$ for at least two $j \in J_i$, and let $|I| = k$. Then λ^* has at least $2k + (m-k)$ positive components, therefore $n+m \geq 2k+(m-k)$, and hence $k \leq n$. Since for every $i \notin I$ we have $\lambda^*_{ij_i} = 1$ for some $j_i \in J_i$ and $\lambda^*_{ij} = 0$ for all other $j \in J_i$, it follows that

$$x = \sum_{i\notin I} x^{ij_i} + \sum_{i\in I}\sum_{j\in J_i} \lambda^*_{ij} x^{ij} \in \sum_{i\notin I} S_i + \sum_{i\in I} \text{conv} S_i.$$

This is the desired expression. □

We close this section by a proposition generalizing Caratheodory's Theorem. A set $X \subset \mathbb{R}^n$ is called a *Caratheodory core* of a set $E \subset \mathbb{R}^n$ if there exist a set $B \subset \mathbb{R}^n$, an interval $\Delta \subset R$ along with a mapping $\alpha : X \to \Delta$ such that $tB \subset t'B$ for $t, t' \in \Delta, t < t'$, and

$$E = \bigcup_{x\in X} (x + \alpha(x)B).$$

A case of interest for the applications is when B is the unit ball and $\Delta = R_+$, so that $E = \bigcup_{x\in X} B(x, \alpha(x))$ where $B(x,r)$ denotes the ball of center x and radius r. Since a trivial Caratheodory core of any set E is $X = E$ with B being the unit ball in $\mathbb{R}^n$ and $\Delta = \{0\}$, Caratheodory's Theorem appears to be a special case of the following general proposition (Thach and Konno 1996):

Proposition 1.14 *If a set E in $\mathbb{R}^n$ has a Caratheodory core X contained in a k-dimensional affine set then any vector of* conv$E$ *is a convex combination of $k+1$ or fewer elements of E.*

Proof For any $y \in \text{conv}(E)$ there exists a finite set of vectors $\{y^1, \ldots, y^m\} \subset E$ such that

$$y = \sum_{i=1}^m \lambda_i y^i, \ \lambda_i \geq 0, \ \sum_{i=1}^m \lambda_i = 1. \tag{1.9}$$

Since $y^i \in E$, there is $x^i \in X$ such that $y^i \in x^i + \alpha_i B, \ \alpha_i := \alpha(x^i)$. Setting $x = \sum_{i=1}^m \lambda_i x^i$, $\theta = \sum_{i=1}^m \lambda_i \alpha_i$, we have

$$(x, \theta) = \sum_{i=1}^m \lambda_i (x^i, \alpha_i).$$

Consider the linear program

$$\max \sum_{i=1}^m t_i \alpha_i \ :$$

$$\sum_{i=1}^m t_i x^i = x, \quad \sum_{i=1}^m t_i = 1 \tag{1.10}$$

$$t_i \geq 0, \ i = 1, \ldots, m. \tag{1.11}$$

and let $t^* := (t_1^*, \ldots, t_m^*)$ be a basic optimal solution of it. Since X is contained in an affine set of dimension k the matrix of the system (1.10)

$$\begin{bmatrix} x^1 & x^2 & \cdots & x^m \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

is of rank at most $k+1$ (Proposition 1.4). Therefore, t^* must satisfy $m-(k+1)$ constraints (1.11) as equalities. That is, $t_i^* = 0$ for at least $m-(k+1)$ indices i. Without loss of generality we can assume that $t_i^* = 0$ for $i = k+2, \ldots, m$, so by setting $\theta^* = \sum_{i=1}^{k+1} t_i^* \alpha_i$, we have

$$(x, \theta^*) = \sum_{i=1}^{k+1} t_i^* (x^i, \alpha_i), \ t_i^* \geq 0, \ \sum_{i=1}^{k+1} t_i^* = 1.$$

Here $\theta \leq \theta^*$ (because t^* is optimal to the above program while $t_i = \lambda_i, i = 1, \ldots, m$ is a feasible solution). Hence $\theta B \subset \theta^* B$ and we can write

$$\begin{aligned} y = \sum_{i=1}^m \lambda_i y^i &\in \sum_{i=1}^m \lambda_i (x^i + \alpha_i B) \\ &= \sum_{i=1}^m \lambda_i x^i + \sum_{i=1}^m \lambda_i \alpha_i B \end{aligned}$$

$$
\begin{aligned}
&= x + \theta B \subset x + \theta^* B \\
&= \sum_{i=1}^{k+1} t_i^* x^i + \sum_{i=1}^{k+1} t_i^* \alpha_i B \\
&= \sum_{i=1}^{k+1} t_i^* (x^i + \alpha_i B).
\end{aligned}
$$

We have thus found $k+1$ vectors $u^i := x^i + \alpha_i B \in E,\ i = 1, \ldots, k+1$, such that $y = \sum_{i=1}^{k+1} t_i^* u^i$ with $t_i^* \geq 0,\ \sum_{i=1}^{k+1} t_i^* = 1$. □

Note that in the above proposition, the dimension k of aff$X$ can be much smaller than the dimension of aff$E$. For instance, if $X = [a, b]$ is a line segment in $\mathbb{R}^n$ and B is the unit ball while $\alpha(a) = \alpha(b) = 1,\ \alpha(x) = 0\ \forall x \notin \{a, b\}$ (so that E is the union of a line segment and two balls centered at the endpoints of this segment) then aff$E = \mathbb{R}^n$ but aff$X = \mathbb{R}$. Thus in this case, by Proposition 1.14, any point of conv$E$ can be represented as the convex combination of only two points of E (and not $n+1$ points as by applying Caratheodory's Theorem.

1.5 Separation Theorems

Separation is one of the most useful concepts of convexity theory. The first and fundamental fact is the following:

Lemma 1.2 *Let C be a nonempty convex subset of $\mathbb{R}^n$. If $0 \notin C$ then there exists a vector $t \in \mathbb{R}^n$ such that*

$$\sup_{x \in C} \langle t, x \rangle \leq 0, \quad \inf_{x \in C} \langle t, x \rangle < 0. \tag{1.12}$$

Proof The proposition is obvious for $n = 1$. Arguing by induction, assume it is true for $n = k-1$ and prove it for $n = k$. Define

$$
\begin{gathered}
C_0 = \{x \in C |\ x_n = 0\}, \\
C_+ = \{x \in C |\ x_n > 0\}, \quad C_- = \{x \in C |\ x_n < 0\}.
\end{gathered}
$$

If $x_n = 0$ for all $x \in C$ then C can be identified with a set in $\mathbb{R}^{n-1}$, so the Lemma is true by the induction assumption. If $C_- = \emptyset$ but $C_+ \neq \emptyset$, then the vector $t = (0, \ldots, 0, -1)$ satisfies (1.12). Analogously, if $C_+ = \emptyset$ and $C_- \neq \emptyset$ we can take $t = (0, \ldots, 0, 1)$. Therefore, we may assume that both C_- and C_+ are nonempty. For any $x \in C_+, x' \in C_-$ we have

$$\lambda \left(\frac{x}{x_n} - \frac{x'}{x'_n} \right) \in C_0 \tag{1.13}$$

for $\lambda > 0$ such that $\frac{\lambda}{x_n} - \frac{\lambda}{x'_n} = 1$. Thus, the set

$$C_* = \{(x_1, \ldots, x_{n-1}) | \ (x_1, \ldots, x_{n-1}, 0) \in C\}$$

is a nonempty convex set in $\mathbb{R}^{n-1}$ which does not contain the origin. By the induction hypothesis there exists a vector $t^* \in \mathbb{R}^{n-1}$ and a point $(\hat{x}_1, \ldots, \hat{x}_{n-1}) \in C_*$ satisfying

$$\sum_{i=1}^{n-1} t_i^* \hat{x}_i < 0, \quad \sup \sum_{i=1}^{n-1} t_i^* x_i | \quad (x_1, \ldots, x_{n-1}) \in C_* \ \leq 0. \tag{1.14}$$

From (1.13) and (1.14), for all $x \in C_+, x' \in C_-$ we deduce

$$\sum_{i=1}^{n-1} t_i^* \left(\frac{x_i}{x_n} - \frac{x'_i}{x'_n} \right) \leq 0,$$

i.e., $p(x) \leq q(x')$, where

$$p(x) = \sum_{i=1}^{n-1} t_i^* \frac{x_i}{x_n} \ (x \in C_+), \quad q(x') = \sum_{i=1}^{n-1} t_i^* \frac{x'_i}{x_n} \ (x' \in C_-).$$

Setting

$$p = \sup_{x \in C_+} p(x), \qquad q = \inf_{x \in C_-} q(x)$$

we then conclude $p \leq q$. Therefore, if we now take $t = (t_1^*, \ldots, t_{n-1}^*, t_n)$, with $-t_n \in [p, q]$ then $p(x) + t_n \leq 0 \ \forall x \in C_+$ and $t_n + q(x) \geq 0 \ \forall x \in C_-$. This implies

$$\sum_{i=1}^{n} t_i x^i \leq 0 \quad \forall x \in C,$$

completing the proof since $\langle t, \hat{x} \rangle < 0$ for $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_{n-1}, 0)$. □

We say that two nonempty subsets C, D of $\mathbb{R}^n$ are *separated* by the hyperplane $\langle t, x \rangle = \alpha$ ($t \in \mathbb{R}^n \setminus \{0\}$) if

$$\sup_{x \in C} \langle t, x \rangle \leq \alpha \leq \inf_{y \in D} \langle t, y \rangle. \tag{1.15}$$

Theorem 1.3 (First Separation Theorem) *Two disjoint nonempty convex sets C, D in $\mathbb{R}^n$ can be separated by a hyperplane.*

Proof The set $C - D$ is convex and satisfies $0 \notin C - D$, so by the previous Lemma, there exists $t \in \mathbb{R}^n \setminus \{0\}$ such that $\langle t, x - y \rangle \leq 0 \ \forall x \in C, y \in D$. Then (1.15) holds for $\alpha = \sup_{x \in C} \langle t, x \rangle$. □

Remark 1.2 Before drawing some consequences of the above result, it is useful to recall a simple property of linear functions:

A linear function $l(x) := \langle t, x\rangle$ with $t \in \mathbb{R}^n \setminus \{0\}$ never achieves its minimum (or maximum) over a set D at an interior point of D; if it is bounded above (or below) over an affine set then it is constant on this affine set.

It follows from this fact that if the set D in Theorem 1.3 is open then (1.15) implies

$$\langle t, x\rangle \le \alpha < \langle t, y\rangle \quad \forall x \in C,\ y \in D.$$

Furthermore:

Corollary 1.10 *If an affine set C does not meet an open convex set D then there exists a hyperplane containing C and not meeting D.*

Proof Let $\langle t, x\rangle = \alpha$ be the hyperplane satisfying (1.15). By the first part of the Remark above, we must have $\langle t, x\rangle = \text{const}\ \forall x \in C$, i.e., C is contained in the hyperplane $\langle t, x - x^0\rangle = 0$, where x^0 is any point of C. By the second part of the Remark, $\langle t, y\rangle < \alpha\ \forall y \in D$ (hence $D \cap C = \emptyset$), since if $\langle t, y^0\rangle = \alpha$ for some $y^0 \in D$ then this would mean that α is the maximum of the linear function $\langle t, x\rangle$ over the open set D. □

Lemma 1.3 *Let C be a nonempty convex set in $\mathbb{R}^n$. If C is closed and $0 \notin C$ then there exists a vector $t \in \mathbb{R}^n \setminus \{0\}$ and a number $\alpha < 0$ such that*

$$\langle t, x\rangle \le \alpha < 0 \quad \forall x \in C.$$

Proof Since C is closed and $0 \notin C$, there is a ball D around 0 which does not intersect C. By the above theorem, there is $t \in \mathbb{R}^n \setminus \{0\}$ satisfying (1.15). Then $\alpha < 0$ because $0 \in D$. □

We say that two nonempty subsets C, D of $\mathbb{R}^n$ are *strongly separated* by a hyperplane $\langle t, x\rangle = \alpha,\ (t \in \mathbb{R}^n \setminus \{0\}, \alpha \in \mathbb{R})$ if

$$\sup_{x\in C}\langle t, x\rangle < \alpha < \inf_{y\in D}\langle t, y\rangle. \tag{1.16}$$

Theorem 1.4 (Second Separation Theorem) *Two disjoint nonempty closed convex sets C, D in $\mathbb{R}^n$ such that either C or D is compact can be strongly separated by a hyperplane.*

Proof Assume, e.g., that C is compact and consider the convex set $C - D$. If $z^k = x^k - y^k \to z$, with $x^k \in C,\ y^k \in D$, then by compactness of C there exists a subsequence $x^{k_\nu} \to x \in C$ and by closedness of D, $y^{k_\nu} = x^{k_\nu} - z^{k_\nu} \to x - z \in D$; hence $z = x - y$, with $x \in C,\ y \in D$. So the set $C - D$ is closed. Since $0 \notin C - D$, there exists by the above Lemma a vector $t \in \mathbb{R}^n \setminus \{0\}$ such that $\langle t, x - y\rangle \le \eta < 0\ \forall x \in C$, $y \in D$. Then $\sup_{x\in C}\langle t, x\rangle - \frac{\eta}{2} \le \inf_{y\in D}\langle t, y\rangle + \frac{\eta}{2}$. Setting $\alpha = \inf_{x\in C}\langle t, x\rangle + \frac{\eta}{2}$ yields (1.16). □

Remark 1.3 If the set C is a cone then one can take $\alpha = 0$ in (1.15). Indeed, for any given $x \in C$ we then have $\lambda x \in C\ \forall\lambda > 0$, so $\langle t, \lambda x\rangle \le \alpha\ \forall\lambda > 0$, hence $\langle t, x\rangle \le 0\ \forall x \in C$. Similarly, one can take $\alpha = 0$ in (1.16).

1.6 Geometric Structure

There are several important concepts related to the geometric structure of convex sets: supporting hyperplane and normal cone, face and extreme point.

1.6.1 Supporting Hyperplane

A hyperplane $H = \{x|\ \langle t, x\rangle = \alpha\}$ is said to be a *supporting hyperplane* to a convex set $C \subset \mathbb{R}^n$ if at least one point x^0 of C lies in H : $\langle t, x^0\rangle = \alpha$, and all points of C lie in one halfspace determined by H, say: $\langle t, x\rangle \le \alpha\ \forall x \in C$ (Fig. 1.3). The halfspace $\langle t, x\rangle \le \alpha$ is called a *supporting halfspace* to C.

Theorem 1.5 *Through every boundary point x^0 of a convex set $C \subset \mathbb{R}^n$ there exists at least one supporting hyperplane to C.*

Proof Since $x^0 \notin \mathrm{ri}C$, there exists by Theorem 1.3 a hyperplane $\langle t, x\rangle = \alpha$ such that

$$\langle t, x\rangle \le \alpha \le \langle t, x^0\rangle. \quad \forall x \in \mathrm{ri}C,$$

hence

$$\langle t, x - x^0\rangle \le 0 \qquad \forall x \in C.$$

Thus, the hyperplane $H = \{x|\ \langle t, x - x^0\rangle = 0\}$ is a supporting hyperplane to C at x^0. □

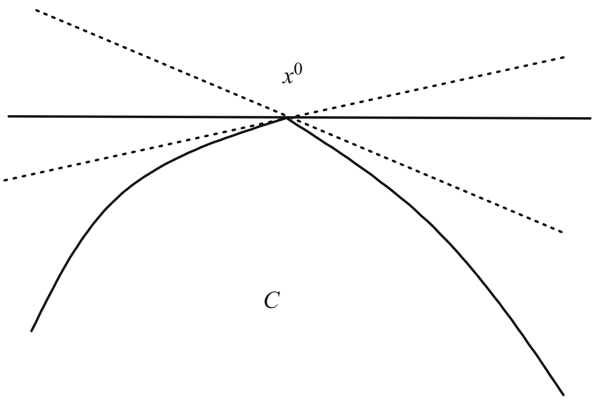


Fig. 1.3 Supporting hyperplane

The normal t to a supporting hyperplane to C at x^0 is characterized by the condition

$$\langle t, x - x^0 \rangle \le 0 \quad \forall x \in C. \tag{1.17}$$

A vector t satisfying this condition does not make an acute angle with any line segment in C having x^0 as endpoint. It is also called an *outward normal*, or simply a *normal*, to C at point x^0 ($-t$ is called an *inward normal*). The set of all vectors t normal to C at x^0 is called the *normal cone* to C at x^0 and is denoted by $N_C(x^0)$. It is readily verified that $N_C(x^0)$ is a convex cone which reduces to the singleton $\{0\}$ when x^0 is an interior point of C.

Proposition 1.15 *Let C be a closed convex set. For any $y^0 \notin C$ there exists $x^0 \in \partial C$ such that $y^0 - x^0 \in N_C(x^0)$.*

Proof Let $\{x^k\} \subset C$ be a sequence such that

$$\|y^0 - x^k\| \to \inf_{x \in C} \|y^0 - x\| \quad (k \to \infty).$$

The sequence $\{\|y^0 - x^k\|\}$ is bounded (because convergent), and hence so is the sequence $\{\|x^k\|\}$ because $\|x^k\| \le \|x^k - y^0\| + \|y^0\|$. Therefore, there is a subsequence $x^{k_\nu} \to x^0$. Since C is closed, we have $x^0 \in C$ and

$$\|y^0 - x^0\| = \min_{x \in C} \|y^0 - x\|.$$

It is easy to see that $y^0 - x^0 \in N_C(x^0)$. Indeed, for any $x \in C$ and $\lambda \in (0, 1)$ we have $x_\lambda := \lambda x + (1 - \lambda)x^0 \in C$, so $\|y^0 - x_\lambda\|^2 \ge \|y^0 - x^0\|^2$, i.e.,

$$\|(y^0 - x^0) - \lambda(x - x^0)\|^2 \ge \|y^0 - x^0\|^2.$$

Upon easy computation this yields

$$\lambda^2 \|x - x^0\|^2 - 2\lambda \langle y^0 - x^0, x - x^0 \rangle \ge 0$$

for all $\lambda \in (0, 1)$, hence $\langle y^0 - x^0, x - x^0 \rangle \le 0$, as desired. □

Note that the condition $y^0 - x^0 \in N_C(x^0)$ uniquely defines x^0. Indeed, if $y^0 - x^1 \in N_C(x^1)$ then from the inequalities

$$\langle y^0 - x^0, x^1 - x^0 \rangle \le 0, \quad \langle y^0 - x^1, x^0 - x^1 \rangle \le 0,$$

it follows that $\langle x^0 - x^1, x^1 - x^0 \rangle \ge 0$, hence $x^1 - x^0 = 0$. The point x^0 is called the *projection* of y^0 on the convex set C. As seen from the above, it is the point of C nearest to y^0.

Theorem 1.6 *A closed convex set C which is neither empty nor the whole space is equal to the intersection of all its supporting halfspaces.*

Proof Let E be this intersection. Obviously, $C \subset E$. Suppose $E \setminus C \neq \emptyset$ and let $y^0 \in E \setminus C$. By the above proposition, there exists a supporting halfspace to C which does not contain y^0, conflicting with $y^0 \in E$. Therefore, $C = E$. □

1.6.2 Face and Extreme Point

A convex subset F of a convex set C is called a *face* of C if any line segment in C with a relative interior point in F entirely lies in F, i.e., if

$$x, y \in C, (1-\lambda)x + \lambda y \in F, 0 < \lambda < 1 \Rightarrow [x, y] \subset F. \tag{1.18}$$

Of course, the empty set and C itself are faces of C. A *proper* face of C is a face which is neither the empty set nor C itself. A face of a convex cone is itself a convex cone.

Proposition 1.16 *The intersection of a convex set C with a supporting hyperplane H is a face of C.*

Proof The convexity of $F := H \cap C$ is obvious. If a segment $[a, b] \subset C$ has a relative interior point x in F and were not contained in F then H would strictly separate a and b, contrary to the definition of a supporting hyperplane. □

A zero-dimensional face of C is called an *extreme point* of C. In other words, an extreme point of C is a point $x \in C$ which is not a relative interior point of any line segment with two distinct endpoints in C.

If a convex set C has a halfline face, the direction of this halfline is called an *extreme direction* of C. The set of extreme directions of C is the same as the set of extreme directions of its recession cone.

It can easily be verified that a face of a face of C is also a face of C. In particular:

An extreme point (or direction) of a face of a convex set C is also an extreme point (or direction, resp.) of C.

Proposition 1.17 *Let E be an arbitrary set in $\mathbb{R}^n$. Every extreme point of $C =$ conv$E$ belongs to E.*

Proof Let x be an extreme point of C. By Proposition 1.13 $x = \sum_{i=1}^{k} \lambda_i x^i, x^i \in E, \lambda_i > 0, \sum_{i=1}^{k} \lambda_i = 1$. It is easy to see that $\lambda_i = 1$ (i.e., $x = x^i$) for some i. In fact, otherwise there is h with $0 < \lambda_h < 1$, so

$$x = \lambda_h x^h + (1-\lambda_h) y^h, \text{ with } y^h = \sum_{i \neq h} \frac{\lambda_i}{1-\lambda_h} x^i \in C$$

(because $\sum_{i \neq h} \frac{\lambda_i}{1-\lambda_h} = 1$), i.e., x is a relative interior point of the line segment $[x^h, y^h] \subset C$, conflicting with x being an extreme point. □

Given an arbitrary point a of a convex set C, there always exists at least one face of C containing a, namely C. The intersection of all faces of C containing a is obviously a face. We refer to this face as the *smallest face containing* a and denote it by F_a.

Proposition 1.18 *F_a is the union of a and all $x \in C$ for which there is a line segment $[x, y] \subset C$ with a as a relative interior point.*

Proof Denote the mentioned union by F. If $x \in F$, i.e., $x \in C$ and there is $y \in C$ such that $a = (1-\lambda)x + \lambda y$ with $0 < \lambda < 1$, then $[x, y] \subset F_a$ because F_a is a face. Thus, $F \subset F_a$. It suffices, therefore, to show that F is a face. By translating if necessary, we may assume that $a = 0$. It is easy to see that F is a convex subset of C. Indeed, if $x^1, x^2 \in F$, i.e., $x^1 = -\lambda_1 y^1,\ x^2 = -\lambda_2 y^2$, with $y^1, y^2 \in C$ and $\lambda_1, \lambda_2 > 0$ then for any $x = (1-\alpha)x^1 + \alpha x^2$ $(0 < \alpha < 1)$ we have $x = -(1-\alpha)\lambda_1 y^1 - \alpha\lambda_2 y^2$, with

$$\frac{-1}{(1-\alpha)\lambda_1 + \alpha\lambda_2}x = \frac{(1-\alpha)\lambda_1}{(1-\alpha)\lambda_1 + \alpha\lambda_2}y^1 + \frac{\alpha\lambda_2}{(1-\alpha)\lambda_1 + \alpha\lambda_2}y^2 \in C,$$

hence $x \in F$. Thus, it remains to prove (1.18). Let $x, y \in C,\ u = (1-\lambda)x + \lambda y \in F$, $0 < \lambda < 1$. Since $u \in F$, there is $v \in C$ such that $u = -\mu v$, so

$$x = \frac{u}{1-\lambda} - \frac{\lambda y}{1-\lambda} = \frac{-\mu v}{1-\lambda} - \frac{\lambda y}{1-\lambda}.$$

We have

$$-\frac{(1-\lambda)x}{\mu+\lambda} = \frac{\mu}{\mu+\lambda}v + \frac{\lambda}{\mu+\lambda}y \in C,$$

so $x \in F$, and similarly, $y \in F$. Therefore, F is a face, completing the proof. □

An equivalent formulation of the above proposition is the following:

Proposition 1.19 *Let C be a convex set and $a \in C$. A face F of C is the smallest face containing a if and only if $a \in \mathrm{ri}F$.*

Proof Indeed, for any subset F of C the condition in Proposition 1.18 is equivalent to saying that F is a face and $a \in \mathrm{ri}F$. □

Corollary 1.11 *If a face F_1 of a convex set C is a proper subset of a face F_2 then* $\dim F_1 < \dim F_2$.

Proof Let a be any relative interior point of F_1. If $\dim F_1 = \dim F_2$, then $a \in \mathrm{ri}F_2$, and so both F_1 and F_2 coincide with the smallest face containing a. □

Corollary 1.12 *A face F of a closed convex set C is a closed set.*

Proof We may assume that $0 \in \mathrm{ri}F$, so that $F = F_0$. Let M be the smallest space containing F. Clearly, every $x \in M$ is of the form $x = -\lambda y$ for some $y \in F$ and $\lambda > 0$. Therefore, by Proposition 1.18, $F = M \cap C$ and since M and C are closed, so must be F. □

1.7 Representation of Convex Sets

Given a nonempty convex set C, the set $(\mathrm{rec}C) \cap (-\mathrm{rec}C)$ is called the *lineality space* of C. It is actually the largest subspace contained in $\mathrm{rec}C$ and consists of the zero vector and all the nonzero vectors y such that, for every $x \in C$, the line through x in the direction of y is contained in C. The dimension of the lineality space of C is called the *lineality* of C. A convex set C of lineality 0, i.e., such that $(\mathrm{rec}C) \cap (-\mathrm{rec}C) = \{0\}$, contains no line (and so is sometimes said to be *line-free*).

Proposition 1.20 *A nonempty closed convex set C has an extreme point if and only if it contains no line.*

Proof If C has an extreme point a then a is also an extreme point of $a + \mathrm{rec}C$, hence 0 is an extreme point of $\mathrm{rec}C$, i.e., $(\mathrm{rec}C) \cap (-\mathrm{rec}C) = \{0\}$. Conversely, suppose C is line-free and let F be a face of smallest dimension of C. Then F must also be line-free, and if F contains two distinct points x, y then the intersection of C with the line through x, y must have an endpoint a. Of course F_a cannot contain $[x, y]$, hence by Corollary 1.11, $\dim F_a < \dim F$, a contradiction. Therefore, F consists of a single point which is an extreme point of C. □

If a nonempty convex set C has a nontrivial lineality space L then for any $x \in C$ we have $x = u + v$, with $u \in L,\ v \in L^{\perp}$, where $L^{\perp}$ is the orthogonal complement of L. Clearly $v = x - u \in C - L = C + L \subset C$, hence C can be decomposed into the following sum:

$$C = L + (C \cap L^{\perp}), \tag{1.19}$$

in which the convex set $C \cap L^{\perp}$ contains no line, hence has an extreme point, by the above proposition.

We are now in a position to state the fundamental representation theorem for convex sets. In view of the above decomposition (1.19) it suffices to consider closed convex sets C with lineality zero. Denote the set of extreme points of C by $V(C)$ and the set of extreme directions of C by $U(C)$.

Theorem 1.7 *Let $C \subset \mathbb{R}^n$ be a closed convex set containing no line. Then*

$$C = conv V(C) + cone U(C). \tag{1.20}$$

In other words any $x \in C$ can be represented in the form

$$x = \sum_{i \in I} \lambda_i v^i + \sum_{j \in J} \mu_j u^j,$$

with I, J finite index sets, $v^i \in V(C), u^i \in U(C)$, and $\lambda_i \geq 0, \mu_j \geq 0,\ \sum_i \lambda_i = 1$.

Proof The theorem is obvious if $\dim C = 0$. Arguing by induction, assume that the theorem is true for $\dim C = n - 1$ and consider the case when $\dim C = n$. Let x be an arbitrary point of C and Γ be any line containing x. Since C is line-free, the set $\Gamma \cap C$ is either a closed halfline, or a closed line segment. In the former case, let $\Gamma = \{a + \lambda u |\ \lambda \geq 0\}$. Of course $a \in \partial C$, and there is a supporting hyperplane to C at a by Theorem 1.5. The intersection F of this hyperplane with C is a face of C of dimension less than n. By the induction hypothesis,

$$a \in \text{cone}U(F) + \text{conv}V(F) \subset \text{cone}U(C) + \text{conv}V(C).$$

Since $u \in \text{cone}U(C)$ and $x = a + \overline{\lambda} u$ for some $\overline{\lambda} \geq 0$, it follows that

$$\begin{aligned} x = a + \overline{\lambda} u &\in (\text{cone}U(C) + \text{conv}V(C)) + \text{cone}U(C) \\ &\subset \text{cone}U(C) + \text{conv}V(C). \end{aligned}$$

In the case when Γ is a closed line segment $[a, b]$, both a, b belong to ∂C, and by the same argument as above, each of these points belongs to $\text{cone}U(C) + \text{conv}V(C)$. Since this set is convex and $x \in [a, b],\ x$ belongs to this set too. □

An immediate consequence of Theorem 1.7 is the following important *Minkowski's Theorem* which corresponds to the special case $U(C) = \{0\}$.

Corollary 1.13 *A nonempty compact convex set C is equal to the convex hull of its extreme points.* □

In view of Caratheodory's Theorem one can state more precisely: *Every point of a nonempty compact convex set C of dimension k is a convex combination of* $k + 1$ *or fewer extreme points of* C.

Corollary 1.14 *A nonempty compact convex set C has at least one extreme point and every supporting hyperplane H to C contains at least one extreme point of C.*

Proof The set $C \cap H$ is nonempty, compact, convex, hence has an extreme point. Since by Proposition 1.16 this set is a face of C any extreme point of it is also an extreme point of C. □

1.8 Polars of Convex Sets

By Theorem 1.6 a closed convex set which is neither empty nor the whole space is fully determined by the set of its supporting hyperplanes. Its turns out that a duality correspondence can be established between convex sets and their sets of supporting hyperplanes.

For any set E in $\mathbb{R}^n$ the set

$$E^\circ = \{y \in \mathbb{R}^n |\ \langle y, x\rangle \le 1,\ \forall x \in E\}$$

is called the *polar* of E. If E is a closed convex cone, so that $\lambda E \subset E$ for all $\lambda \ge 0$, then the condition $\langle y, x\rangle \le 1,\ \forall x \in E$, is equivalent to $\langle y, x\rangle \le 0,\ \forall x \in E$. Therefore, the polar of a cone M is the cone

$$M^\circ = \{y \in \mathbb{R}^n |\ \langle y, x\rangle \le 0,\ \forall x \in M\}.$$

The polar of a subspace is the same as its orthogonal complement.

Proposition 1.21

(i) *The polar E° of any set E is a closed convex set containing the origin; if $E \subset F$ then $F^\circ \subset E^\circ$;*
(ii) $E \subset E^{\circ\circ}$ *; if E is closed, convex, and contains the origin then* $E^{\circ\circ} = E$;
(iii) $0 \in \mathrm{int}E$ *if and only if E° is bounded.*

Proof

(i) Straightforward.
(ii) If $x \in E$ then $\langle x, y\rangle \le 1\ \forall y \in E^\circ$, hence $x \in E^{\circ\circ}$. Suppose E is closed, convex and contains 0, but there is a point $u \in E^{\circ\circ} \setminus E$. Then by the Separation Theorem 1.4 there exists a vector y such that $\langle y, x\rangle \le 1,\ \forall x \in E$, while $\langle y, u\rangle > 1$. The former inequality (for all $x \in E$) means that $y \in E^\circ$; the latter inequality then implies that $u \notin E^{\circ\circ}$, a contradiction.
(iii) First note that the polar of the ball B_r of radius r around 0 is the ball $B_{1/r}$. Indeed, the distance from 0 to a hyperplane $\langle y, x\rangle = 1$ is $\dfrac{1}{\|y\|}$. Therefore, if $y \in (B_r)^\circ$, i.e., $\langle y, x\rangle \le 1\ \forall x \in B_r$ then $\dfrac{1}{\|y\|} \ge r$, hence $\|y\| \le 1/r$, i.e., $y \in B_{1/r}$, and conversely. This proves that $(B_r)^\circ = B_{1/r}$. Now, $0 \in \mathrm{int}E$ if and only if $B_r \subset E$ for some $r > 0$, hence if and only if $E^\circ \subset (B_r)^\circ = B_{1/r}$, i.e., E° is bounded. □

Proposition 1.22 *If $M_1, M_2 \subset \mathbb{R}^n$ are closed convex cones then*

$$(M_1 + M_2)^\circ = M_1^\circ \cap M_2^\circ; \quad (M_1 \cap M_2)^\circ = M_1^\circ + M_2^\circ.$$

Proof If $y \in (M_1+M_2)^\circ$ then for any $x \in M_1$, writing $x = x+0 \in M_1+M_2$ we have $\langle x, y\rangle \le 0$, hence $y \in M_1^\circ$; analogously, $y \in M_2^\circ$. Therefore $(M_1+M_2)^\circ \subset M_1^\circ \cap M_2^\circ$. Conversely, if $y \in M_1^\circ \cap M_2^\circ$ then for all $x^1 \in M_1, x^2 \in M_2$ we have $\langle x^1, y\rangle \le 0$, $\langle x^2, y\rangle \le 0$, hence $\langle x^1 + x^2, y\rangle \le 0$, implying that $y \in (M_1+M_2)^\circ$. Therefore, the first equality holds. The second equality follows, because by (ii) of the previous proposition, $M_1^\circ + M_2^\circ = (M_1^\circ + M_2^\circ)^{\circ\circ} = [(M_1^\circ)^\circ \cap (M_2^\circ)^\circ]^\circ = (M_1 \cap M_2)^\circ$. □

Given a set $C \subset \mathbb{R}^n$, the function $x \in \mathbb{R}^n \mapsto s_C(x) = \sup_{y\in C}\langle x, y\rangle$ is called the *support function* of C.

Proposition 1.23 *Let C be a closed convex set containing 0.*

(i) *The gauge of C is the support function of C°.*
(ii) *The recession cone of C and the closure of the cone generated by C° are polar to each other.*
(iii) *The lineality space of C and the subspace generated by C° are orthogonally complementary to each other.*

Dually, also, with C and C° interchanged.

Proof

(i) Since $(C^\circ)^\circ = C$ by Proposition 1.21, we can write

$$p_C(x) = \inf\{\lambda > 0 |\ \frac{x}{\lambda} \in C\} = \inf\{\lambda > 0 |\ \langle \frac{x}{\lambda}, y\rangle \le 1\ \forall y \in C^\circ\}$$
$$= \inf\{\lambda > 0 |\ \langle x, y\rangle \le \lambda\ \forall y \in C^\circ\} = \sup_{y\in C^\circ} \langle x, y\rangle$$

(ii) Since $0 \in C$, the recession cone of C is the largest closed convex cone contained in C. Its polar must then be the smallest closed convex cone containing C°, i.e., the closure of the convex cone generated by C°.
(iii) Similarly, the lineality space of C is the largest space contained in C. Therefore, its orthogonal complement is the smallest subspace containing C°. □

Corollary 1.15 *For any closed convex set $C \subset \mathbb{R}^n$ containing 0 we have*

$$\dim C^\circ = n - \text{lineality}C$$
$$\text{lineality}C^\circ = n - \dim C.$$

Proof The latter relation follows from (iii) of the previous proposition, while the former is obtained from the latter by noting that $(C^\circ)^\circ = C$. □

The number $\dim C - \text{lineality}C$ which can be considered as a measure of the nonlinearity of a convex set C is called the *rank* of C. From (1.19) we have $\text{rank}C = \dim(C \cap L^\perp)$, where L is the lineality space of C. The previous corollary also shows that for any closed convex set C :

$$\text{rank}C = \text{rank}C^\circ.$$

1.9 Polyhedral Convex Sets

A convex set is said to be *polyhedral* if it is the intersection of a finite family of closed halfspaces. A polyhedral convex set is also called a *polyhedron*. In other words, a polyhedron is the solution set of a finite system of linear inequalities of the form

$$\langle a^i, x\rangle \le b_i \quad i = 1, \dots, m, \tag{1.21}$$

or in matrix form:

$$Ax \le b, \tag{1.22}$$

where A is the $m \times n$ matrix of rows a^i and $b \in \mathbb{R}^m$. Since a linear equality can be expressed as two linear inequalities, a polyhedron is also the solution set of a system of linear equalities and inequalities of the form

$$\langle a^i, x\rangle = b_i, \quad i = 1, \dots, m_1$$
$$\langle a^i, x\rangle \le b_i, \quad i = m_1 + 1, \dots, m.$$

The *rank* of a system of linear inequalities of the form (1.22) is by definition the rank of the matrix A. When this rank is equal to the number of inequalities, we say that the inequalities are *linearly independent*.

Proposition 1.24 *A polyhedron $D \neq \emptyset$ defined by the system (1.22) has dimension r if and only if the subsystem of (1.22) formed by the inequalities which are satisfied as equalities by all points of D has rank $n - r$.*

Proof Let $I_0 = \{i|\ \langle a^i, x\rangle = b_i\ \forall x \in D\}$ and $M = \{x|\ \langle a^i, x\rangle = b_i\ i \in I_0\}$. Since $\dim M = n - r$, and $D \subset M$, it suffices to show that $M \subset \mathrm{aff} D$. We can assume that $I_1 := \{1, \dots, m\} \setminus I_0 \neq \emptyset$ for otherwise $D = M$. For each $i \in I_1$ take $x^i \in D$ such that $\langle a^i, x^i\rangle < b_i$ and let $x^0 = \frac{1}{q}\sum_{i\in I_1} x^i$, where $q = |I_1|$. Then clearly

$$x^0 \in M, \quad \langle a^i, x^0\rangle < b_i\ \forall i \in I_1. \tag{1.23}$$

Consider now any $x \in M$. Using (1.23) it is easy to see that for $\lambda > 0$ sufficiently small, we have

$$x^0 + \lambda(x - x^0) \in M, \quad \langle a^i,\ x^0 + \lambda(x - x^0)\rangle \le b_i\ \forall i \in I_1,$$

hence $y := x^0 + \lambda(x - x^0) \in D$. Thus, both points x^0 and y belong to D. Hence the line through x^0 and y lies in $\mathrm{aff} D$, and in particular, $x \in \mathrm{aff} D$. Since x is an arbitrary point of M this means that $M \subset \mathrm{aff} D$, and hence $M = \mathrm{aff} D$. $\square$

Corollary 1.16 *A polyhedron (1.22) is full-dimensional if and only if there is x^0 satisfying $\langle a^i, x^0\rangle < b_i \ \forall i = 1,\dots,m$.*

Proof This corresponds to the case $r = 0$, i.e., $I_1 = \{1,\dots,m\}$. Note that if for every $i \in I_1$ there exists $x^i \in D$ satisfying $\langle a^i, x^i\rangle < b_i$ then any $x^0 \in \text{ri}D$ (or x^0 constructed as in the above proof) satisfies $\langle a^i, x^0\rangle < b_i \ \forall i \in I_1$. □

1.9.1 Facets of a Polyhedron

As previously, denote by D the polyhedron (1.22) and let

$$I_0 = \{i|\ \langle a^i, x\rangle = b_i \ \forall x \in D\}.$$

Theorem 1.8 *A nonempty subset F of D is a face of D if and only if*

$$F = \{x|\ \langle a^i, x\rangle = b_i,\ i \in I; \quad \langle a^i, x\rangle \le b_i,\ i \notin I\}, \tag{1.24}$$

for some index set I such that $I_0 \subset I \subset \{1,\dots,m\}$.

Proof

(i) Any set F of the form (1.24) is a face. Indeed, F is obviously convex. Suppose there is a line segment $[y,z] \subset D$ with a relative interior point x in F. We have $z = (1-\lambda)x + \lambda y$ for some $\lambda < 0$, so if $\langle a^i, y\rangle < b_i$ for some $i \in I$ then $\langle a^i, z\rangle = (1-\lambda)\langle a^i, x\rangle + \lambda\langle a^i, y\rangle > (1-\lambda)b_i + \lambda b_i = b_i$, a contradiction. Therefore, $\langle a^i, y\rangle = b_i \ \forall i \in I$, and analogously, $\langle a^i, z\rangle = b_i \ \forall i \in I$. Hence $[y,z] \subset F$, proving that F is a face.

(ii) Any face F has the form (1.24). Indeed, by Proposition 1.19 $F = F_{x^0}$, where x^0 is any relative interior point of F. Denote by F' the set in the right-hand side of (1.24) for $I = \{i|\ \langle a^i, x^0\rangle = b_i\}$. According to what has just been proved, F' is a face, so $F' \supset F$. To prove the converse inclusion, consider any $x \in F'$ and let $y = x^0 + \lambda(x^0 - x)$. Since

$$\langle a^i, x^0\rangle = \langle a^i, x\rangle = b_i \ \forall i \in I; \quad \langle a^i, x^0\rangle < b_i \ \forall i \notin I$$

one has $\langle a^i, y\rangle = b_i \ \forall i \in I$, $\langle a^i, y\rangle \le b_i \ \forall i \notin I$ provided $\lambda > 0$ is sufficiently small. Then $y \in F'$ and $[x,y] \subset F' \subset D$. Since $x^0 \in F$ and x^0 is a relative interior point of $[x,y]$, while F is a face, this implies that $[x,y] \subset F$, and in particular, $x \in F$. Thus, any $x \in F'$ belongs to F. Therefore $F = F'$, completing the proof. □

A proper face of maximal dimension of a polyhedron is called a *facet*. By the above proposition, if F is a facet of D then there is a $k \notin I_0$ such that

$$F = \{x \in D | \langle a^k, x\rangle = b_k\}. \tag{1.25}$$

However, not every face of this form is a facet. An inequality $\langle a^k, x\rangle \le b_k$ is said to be *redundant* if the removal of this inequality from (1.21) does not affect the polyhedron D, i.e., if the system (1.21) is equivalent to

$$\langle a^i, x\rangle \le b_i, \quad i \in \{1, \ldots, m\} \setminus \{k\}.$$

Proposition 1.25 *If the inequality $\langle a^k, x\rangle \le b_k$ with $k \notin I_0$ is not redundant then (1.25) is a facet.*

Proof The non-redundance of the inequality $\langle a^k, x\rangle \le b_k$ means that there is an y such that

$$\langle a^i, y\rangle = b_i,\ i \in I_0, \quad \langle a^i, y\rangle \le b_i,\ i \notin (I_0 \cup \{k\}), \quad \langle a^k, y\rangle > b_k.$$

Let x^0 be a relative interior point of D, i.e., such that

$$\langle a^i, x^0\rangle = b_i,\ i \in I_0, \quad \langle a^i, x^0\rangle < b_i,\ i \notin I_0.$$

Then the line segment $[x^0, y]$ meets the hyperplane $\langle a^k, x\rangle = b_k$ at a point z satisfying

$$\langle a^i, z\rangle = b_i,\ i \in (I_0 \cup \{k\}), \quad \langle a^i, z\rangle < b_i,\ i \notin (I_0 \cup \{k\}).$$

This shows that F defined by (1.25) is of dimension $\dim D - 1$, hence is a facet. □

Thus, if a face (1.25) is not a facet, then the inequality $\langle a^k, x\rangle \le b_k$ can be removed without changing the polyhedron. In other words, the set of facet-defining inequalities, together with the system $\langle a^i, x\rangle = b_i, i \in I_0$, completely determines the polyhedron.

1.9.2 Vertices and Edges of a Polyhedron

An extreme point (0-dimensional face) of a polyhedron is also called a *vertex*, and a 1-dimensional face is also called an *edge*. The following characterizations of vertices and edges are immediate consequences of Theorem 1.8:

Corollary 1.17

(i) *A point $x \in D$ is a vertex if and only if it satisfies as equalities n linearly independent inequalities from (1.21).*
(ii) *A line segment (or halfline, or line) $\Gamma \subset D$ is an edge of D if and only if it is the set of points of D satisfying as equalities $n-1$ linearly independent inequalities from (1.21).*

Proof A set F of the form (1.24) is a k-dimensional face if and only if the system $\langle a^i, x\rangle = b_i,\ i \in I$, is of rank $n - k$. □

A vertex of a polyhedron $D \subset \mathbb{R}^n$ [defined by (1.21)] is said to be *nondegenerate* if it satisfies as equalities exactly n inequalities (which must then be linearly independent) from (1.21). Two vertices x^1, x^0 are said to be *neighboring* (or *adjacent*) if the line segment between them is an edge.

Theorem 1.9 *Let x^0 be a nondegenerate vertex of a full-dimensional polyhedron D defined by a system (1.21). Then there are exactly n edges of D emanating from x^0. If I is the set of indices of inequalities that are satisfied by x^0 as equalities, then for each $k \in I$ there is an edge emanating from x^0 whose direction z is defined by the system*

$$\langle a^k, z\rangle = -1, \quad \langle a^i, z\rangle = 0, \; i \in I \setminus \{k\}. \tag{1.26}$$

Proof By the preceding corollary a vector z defines an edge emanating from x^0 if and only if for all $\lambda > 0$ small enough $x^0 + \lambda z$ belongs to D and satisfies exactly $n-1$ equalities from the system $\langle a^i, x\rangle = b_i, \; i \in I$. Since $\langle a^i, x^0\rangle < b_i \; (i \notin I)$, we have $\langle a^i, \; x^0 + \lambda z\rangle \leq b_i \; (i \notin I)$ for all sufficiently small $\lambda > 0$. Therefore, it suffices that for some $k \in I$ we have $\langle a^i, \; x^0 + \lambda z\rangle = b_i \; (i \in I \setminus \{k\}), \; \langle a^k, \; x^0 + \lambda z\rangle < b_k$, i.e., that z (or a positive multiple of z) satisfies (1.26). Since the latter system is of rank n it fully determines the direction of z. We thus have exactly n edges emanating from x^0. □

A bounded polyhedron is called a *polytope*.

Proposition 1.26 *An r-dimensional polytope has at least $r + 1$ vertices.*

Proof A polytope is a compact convex set, so by Corollary 1.13 it is the convex hull of its vertex set $\{x^1, x^2, \ldots, x^k\}$. The affine hull of the polytope is then the affine hull of this vertex set, and by Proposition 1.4, if it is of dimension r then the matrix

$$\begin{bmatrix} x^1 & x^2 & \cdots & x^k \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

is of rank $r + 1$, hence $k \geq r + 1$. □

A polytope which is the convex hull of $r + 1$ affinely independent points $x^1, x^2, \ldots, x^{r+1}$ is called an *r-simplex* and is denoted by $[x^1, x^2, \ldots, x^{r+1}]$. By Proposition 1.17 any vertex of such a polytope must then be one of the points $x^1, x^2, \ldots, x^{r+1}$, and since there must be at least $r + 1$ vertices, these are precisely $x^1, x^2, \ldots, x^{r+1}$.

If D is an r-dimensional polytope then any set of $k \leq r + 1$ vertices of D determines an r-simplex which is an r-dimensional face of D. Conversely, any r-dimensional face of D is of this form.

Proposition 1.27 *The recession cone of a polyhedron (1.22) is the cone $M := \{x|\, Ax \leq 0\}$.*

Proof By definition $y \in \text{rec}D$ if and only if for any $x \in D$ and $\lambda > 0$ we have $x + \lambda y \in D$, i.e., $A(x + \lambda y) = Ax + \lambda Ay \le b$, i.e., $Ay \le 0$. Hence the conclusion. □

Thus, the extreme directions of a polyhedron (1.22) are the same as the extreme directions of the cone $Ax \le 0$.

Corollary 1.18 *A nonempty polyhedron (1.22) has at least a vertex if and only if* $\text{rank}A = n$. *It is a polytope if and only if the cone* $Ax \le 0$ *is the singleton* $\{0\}$.

Proof By Proposition 1.20 D has a vertex if and only if it contains no line, i.e., the recession cone $Ax \le 0$ contains no line, which in turns is equivalent to saying that $\text{rank}A = n$. The second assertion follows from Corollary 1.8 and the preceding proposition. □

1.9.3 Polar of a Polyhedron

Let x^0 be a solution of the system (1.21), so that $Ax^0 \le b$. Then $A(x - x^0) \le b - Ax^0$, with $b - Ax^0 \ge 0$. Therefore, by shifting the origin to x^0 and dividing, the system (1.21) can be put into the form

$$\langle a^i, x\rangle \le 1,\ i = 1,\dots,p, \quad \langle a^i, x\rangle \le 0,\ i = p+1,\dots,m. \tag{1.27}$$

Proposition 1.28 *Let P be the polyhedron defined by the system (1.27). Then the polar of P is the polyhedron*

$$Q = \text{conv}\{0, a^1,\dots,a^p\} + \text{cone}\{a^{p+1},\dots,a^m\}. \tag{1.28}$$

and conversely, the polar of Q is the polyhedron P.

Proof If $y \in Q$, i.e.,

$$y = \sum_{i=1}^m \lambda_i a^i, \text{ with } \sum_{i=1}^p \lambda_i \le 1,\ \lambda_i \ge 0, i = 1,\dots,m$$

then

$$\langle y, x\rangle = \sum_{i=1}^m \lambda_i \langle a^i, x\rangle \le 1 \quad \forall x \in P,$$

hence $y \in P^\circ$ (polar of P). Therefore, $Q \subset P^\circ$. To prove the converse inclusion, observe that for any $y \in Q^\circ$ we must have $\langle y, x\rangle \le 1\ \forall x \in Q$. In particular, for $i \le p$, taking $x = a^i$ yields $\langle a^i, y\rangle \le 1,\ i = 1,\dots,p$; while for $i \ge p+1$, taking $x = \theta a^i$ with arbitrarily large $\theta > 0$ yields $\langle a^i, y\rangle \le 0,\ i = p+1,\dots,m$. Thus, $Q^\circ \subset P$ and hence, by Proposition 1.21, $P^\circ \subset (Q^\circ)^\circ = Q$. This proves that $Q = P^\circ$. The equality $Q^\circ = P^{\circ\circ} = P$ then follows. □

Note the presence of the vector 0 in (1.28). For instance, if $P = \{x \in \mathbb{R}^2 | \ x_1 \leq 1, x_2 \leq 1\}$ then its polar is $P^\circ = \text{conv}\{0, e^1, e^2\}$ and not $\text{conv}\{e^1, e^2\}$. However, if P is bounded then $0 \in \text{int}P^\circ$ (Proposition 1.21), and 0 can be omitted in (1.28). As an example, it is easily verified that the polar of the unit l_∞-ball $\{x | \ \max_i |x_i| \leq 1\} = \{x | \ -1 \leq x_i \leq 1, \ i = 1, \ldots, n\}$ is the set $\text{conv}\{e^i, -e^i, i = 1, \ldots, n\}$, i.e., the unit l_1-ball $\{x | \ \sum_{i=1}^n |x_i| \leq 1\}$. This fact is often expressed by saying that the l_∞-norm and the l_1-norm are *dual* to each other.

Corollary 1.19 *If two full-dimensional polyhedrons P, Q containing the origin are polar to each other then there is a 1-1 correspondence between the set of facets of P not containing the origin and the set of nonzero vertices of Q; likewise, there is a 1-1 correspondence between the set of facets of P containing the origin and the set of extreme directions of Q.*

Proof One can suppose that P, Q are defined as in Proposition 1.28. Then an inequality $\langle a^k, x\rangle \leq 1, k \in \{1, \ldots, p\}$, is redundant in (1.27) if and only if the vector a^k can be removed from (1.28) without changing Q. Therefore, the equality $\langle a^k, x\rangle = 1$ defines a facet of P not containing 0 if and only if a^k is a vertex of Q. Likewise, for the facets passing through 0 (Fig. 1.4). □

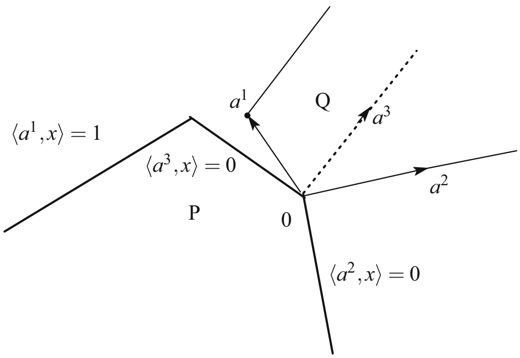


Fig. 1.4 Polyhedrons polar to each other

1.9.4 Representation of Polyhedrons

From Theorem 1.8 it follows that a polyhedron has finitely many faces, in particular finitely many extreme points and extreme directions. It turns out that this property completely characterizes polyhedrons in the class of closed convex sets.

Theorem 1.10 *For every polyhedron D there exist two finite sets $V = \{v^i, i \in I\}$ and $U = \{u^j, j \in J\}$ such that*

$$D = \text{conv}V + \text{cone}U. \tag{1.29}$$

In other words, D consists of points of the form

$$x = \sum_{i\in I} \lambda_i v^i + \sum_{j\in J} \mu_j u^j, \text{ with}$$

$$\sum_{i\in I} \lambda_i = 1, \quad \lambda_i \geq 0, \mu_j \geq 0, \ i \in I, j \in J.$$

Proof If L is the lineality of D then, according to (1.19):

$$D = L + C, \tag{1.30}$$

where $C := D \cap L^{\perp}$ is a polyhedron containing no line. By Theorem 1.7,

$$C = \text{conv}V + \text{cone}U_1, \tag{1.31}$$

where V is the set of vertices and U_1 the set of extreme directions of C. Now let U_0 be a basis of L, so that $L = \text{cone}\{U_0 \cup (-U_0)\}$. From (1.30) and (1.31) we conclude

$$D = \text{conv}V + \text{cone}U_1 + \text{cone}\{U_0 \cup (-U_0)\},$$

whence (1.29) by setting $U = U_0 \cup (-U_0) \cup U_1$. □

Lemma 1.4 *The convex hull of a finite set of halflines emanating from the origin is a polyhedral convex cone. In other words, given any finite set of vectors $U \subset \mathbb{R}^n$, there exists an $m \times n$ matrix A such that*

$$\text{cone}U = \{x|\ Ax \leq 0\} \tag{1.32}$$

Proof Let $U = \{u^j, j \in J\}$, $M = \text{cone}\{u^j, j \in J\}$. By Proposition 1.28 $M^{\circ} = \{x|\ \langle v^j, x\rangle \leq 0, j \in J\}$ and by Theorem 1.10 there exist $a^1, \ldots, a^m$ such that $M^{\circ} = \{x = \sum_{i=1}^m \lambda_i a^i|\ \lambda_i \geq 0, i = 1, \ldots, m\}$. Finally, again by Proposition 1.28, $(M^{\circ})^{\circ} = \{x|\ \langle a^i, x\rangle \leq 0, i = 1, \ldots, m\}$, which yields the desired representation by noting that $M^{\circ\circ} = M$. □

Theorem 1.11 *Given any two finite sets V and U in $\mathbb{R}^n$, there exist an $m \times n$ matrix A and a vector $b \in \mathbb{R}^m$ such that*

$$\text{conv}V + \text{cone}U = \{x|\ Ax \leq b\}.$$

Proof Let $V = \{v^i, i \in I\}, U = \{u^j, j \in J\}$. We can assume $V \neq \emptyset$, since the case $V = \emptyset$ has just been treated. In the space $\mathbb{R}^{n+1} = \mathbb{R}^n \times \mathbb{R}$, consider the set

$$K = \left\{ (x, t) = \sum_{i\in I} \lambda_i (v^i, 1) + \sum_{j\in J} \mu_j (u^j, 0)|\ \lambda_i \geq 0, \mu_j \geq 0 \right\}.$$

This set is contained in the halfspace $\{(x, t) \in \mathbb{R}^n \times \mathbb{R} | \ t \geq 0\}$ of $\mathbb{R}^{n+1}$ and if we set $D = \text{conv}V + \text{cone}U$ then: $x \in D \Leftrightarrow (x, 1) \in K$. By Lemma 1.4, K is a convex polyhedral cone, i.e., there exist (A, b) such that

$$K = \{(x, t) | \ Ax - tb \leq 0\}.$$

Thus,

$$D = \{x| \quad Ax \leq b\},$$

completing the proof. □

Corollary 1.20 *A closed convex set D which has only finitely many faces is a polyhedron.*

Proof Indeed, in view of (1.19), we can assume that D contains no line. Then D has a finite set V of extreme points and a finite set U of extreme directions, so by Theorem 1.7 it can be represented in the form $\text{conv}V + \text{cone}U$, and hence is a polyhedron by the above theorem. □

1.10 Systems of Convex Sets

Given a system of closed convex sets we would like to know whether they have a nonempty intersection.

Lemma 1.5 *If m closed convex proper sets $C_1, \dots, C_m$ of $\mathbb{R}^n$ have no common point in a compact convex set C then there exist closed halfspaces $H_i \supset C_i$ $(i = 1, \dots, m)$ with no common point in C.*

Proof The closed convex set C_m does not intersect the compact set $C' = C \cap (\cap_{i=1}^{m-1} C_i)$, so by Theorem 1.4 there exists a closed halfspace $H_m \supset C_m$ disjoint with C' (i.e., such that $C_1, \dots, C_{m-1}, H_m$ have no common point in C). Repeating this argument for the collection $C_1, \dots, C_{m-1}, H_m$ we find a closed halfspace $H_{m-1} \supset C_{m-1}$ such that $C_1, \dots, C_{m-2}, H_{m-1}, H_m$ have no common point in C, etc. Eventually we obtain the desired closed halfspaces $H_1, \dots, H_m$. □

Theorem 1.12 *If the closed convex sets $C_1, \dots, C_m$ in $\mathbb{R}^n$ have no common point in a convex set C but any $m - 1$ sets of the collection have a common point in C then there exists a point of C not belonging to any $C_i, i = 1, \dots, m$.*

Proof First, it can be assumed that the set C is compact. In fact, if it were not so, we could pick, for each $i = 1, \dots, m$, a point a^i common to C and the sets $C_j, j \neq i$. Then the set $C' = \text{conv}\{a^1, \dots, a^m\}$ is compact and the sets $C_1, \dots, C_m$ have no common point in C' (because $C' \subset C$); moreover, any $m - 1$ sets among these have a common point in C' (which is one of the points $a^1, \dots, a^m$). Consequently, without any harm we can replace C with C'.

Second, each $C_i, i = 1, \ldots, m$, must be a proper subset of $\mathbb{R}^n$ because if some $C_{i_0} = \mathbb{R}^n$ then a common point a^{i_0} of $C_i, i \neq i_0$, in C would be a common point of $C_1, \ldots, C_m$, in C. Therefore, by Lemma 1.5 there exist closed halfspaces $H_i \supset C_i, i = 1, \ldots, m$, such that $H_1, \ldots, H_m$ have no common point in C. Let $H_i = \{x | \langle h_i(x) \leq 0\}$, where $h_i(x)$ is an affine function. We contend that for every set $I \subset \{1, \ldots, m\}$ and every $q \notin I$ there exists $x^{qI} \in C$ satisfying

$$h_i(x^{qI}) \begin{cases} > 0 \text{ if } i = q & (1) \\ = 0 \text{ if } i \in I & (2) \\ \leq 0 \text{ if } i \neq q & (3) \end{cases}$$

In fact, for $I = \emptyset$ this is clear: since $H_i \supset C_i, i = 1, \ldots, m$, any $m - 1$ sets among the $H_i, i = 1, \ldots, m$, have a common point in C. Supposing this is true for $|I| = s$ consider the case $|I| = s+1$. For $p \in I, J = I \backslash \{p\}$ by the induction hypothesis there exist x^{pJ} and x^{qJ}. The convex set $\{x \in C | \langle h_i(x) = 0 \ (i \in J),\ h_i(x) \leq 0 \ i \notin \{p, q\}\}$ contains two points x^{pJ} and x^{qJ} satisfying $h_p(x^{pJ}) > 0$ and $h_p(x^{qJ}) \leq 0$, hence it must contain a point x^{qI} satisfying $h_p(x^{qI}) = 0$. Clearly x^{qI} satisfies (2) and (3), and hence, also (1) since the H_i have no common point in C.

For each q denote by x^q the point x^{qI} associated with $I = \{1, \ldots, m\} \setminus \{q\}$. It is easily seen that the point $x^* := \frac{1}{m} \sum_{q=1}^m x^q \in C$ satisfies $\forall i \quad h_i(x^*) = \frac{1}{m} \sum_{q=1}^m h_i(x^q) = \frac{1}{m} h_i(x^i) > 0$. This means x^* does not belong to any H_i, hence neither to any C_i. □

Corollary 1.21 *Let $C_1, C_2, \ldots, C_m$ be closed convex sets whose union is a convex set C. If any $m-1$ sets of the collection have a nonempty intersection, then the entire collection has a nonempty intersection.*

Proof If the entire collection had an empty intersection then by Theorem 1.12 there would exist a point of C not belonging to any C_i, a contradiction. □

Theorem 1.13 (Helly) *Let $C_1, \ldots, C_m$ be $m > n + 1$ closed convex subsets of $\mathbb{R}^n$. If any $n + 1$ sets of the collection have a nonempty intersection then the entire collection has a nonempty intersection.*

Proof Suppose $m = n + 2$, so that $m - 1 = n + 1$. For each i take an $a^i \in \cap_{j \neq i} C_j$. Since $m - 1 > n$ the vectors $a^i - a^1, i = 2, \ldots, m$ must be linearly dependent, i.e., there exist numbers $\lambda_i, i = 1, \ldots, m$, not all zero, such that $\sum_{i=2}^m \lambda_i (a^i - a^1) = 0$. Setting $\lambda_1 = -\sum_{i=2}^m \lambda_i$ yields $\sum_{i=1}^m \lambda_i a^i = 0$. With $J = \{i | \lambda_i \geq 0, \alpha = \sum_{i \in J} \lambda_i = -\sum_{i \notin J} \lambda_i > 0$ the latter equality can be written as

$$\frac{1}{\alpha} \sum_{i \in J} \lambda_i a^i = \frac{1}{\alpha} \sum_{i \notin J} (-\lambda_i) a^i.$$

For $i \in J$ we have $a^i \in \cap_{j \notin J} C_j$, so $x = \sum_{i \in J} \frac{\lambda_i}{\alpha} a^i \in \cap_{j \notin J} C_j$; for $i \notin J$ we have $a^i \in \cap_{j \in J} C_j$, so $x = \sum_{i \notin J} \frac{-\lambda_i}{\alpha} a^i \in \cap_{j \in J} C_i$. Therefore, $x \in \cap_{j=1}^m C_j$, proving the theorem for $m = n + 2$. The general case follows by induction on m. □

1.11 Exercises

1 Show that any affine set in $\mathbb{R}^n$ is the intersection of a finite collection of hyperplanes.

2 Let A be an $m \times n$ matrix. The *kernel* (*null-space*) of A is by definition the subspace $L = \{x \in \mathbb{R}^n | \; Ax = 0\}$. The *orthogonal complement* to L is the set $L^\perp = \{x \in \mathbb{R}^n | \; \langle x, y \rangle = 0 \; \forall y \in L\}$. Show that $L^\perp$ is a subspace and that $L^\perp = \{x \in \mathbb{R}^n | \; x = A^T u, u \in \mathbb{R}^m\}$.

3 Let M be a convex cone containing 0. Show that the affine hull of M is the set $M - M = \{x - y | \; x \in M, y \in M\}$ and that the largest subspace contained in M is the set $-M \cap M$.

4 Let D be a nonempty polyhedron defined by a system of linear inequalities $Ax \leq b$ where rank$A = r < n$. Show that the lineality of D is the subspace $E = \{x | \; Ax = 0\}$, of dimension $n - r$, and that $D = E + D_0$ where D_0 is a polyhedron having at least one extreme point.

5 Let $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping. If D is a (bounded) polyhedron in $\mathbb{R}^n$ then $\varphi(D)$ is a (bounded) polyhedron in $\mathbb{R}^m$. (Use Theorems 1.10 and 1.11.)

6 If C and D are polyhedrons in $\mathbb{R}^n$ then $C + D$ is also a polyhedron. ($C + D$ is the image of $E = \{(x, y) | \; x \in C, y \in D\}$ under the linear mapping $(x, y) \mapsto x + y$ from $\mathbb{R}^{2n}$ to $\mathbb{R}^n$.)

7 Let $\{C_i, i \in I\}$ be an arbitrary collection of nonempty convex sets in $\mathbb{R}^n$. Then a vector $x \in \mathbb{R}^n$ belongs to the convex hull of the union of the collection if and only if x is the convex combination of a finite number of vectors $y^i, i \in J$ such that $y^i \in C_i$ and $J \subset I$.

8 Let C_1, C_2 be two disjoint convex sets in $\mathbb{R}^n$, and $a \in \mathbb{R}^n$. Then at least one of the two sets $C_2 \cap \text{conv}(C_1 \cup \{a\})$, $C_1 \cap \text{conv}(C_2 \cup \{a\})$ is empty. (Argue by contradiction.)

9 Let $C_1, \ldots, C_m$ be open convex proper subsets of $\mathbb{R}^n$ with no common point in a convex proper subset D of $\mathbb{R}^n$. Then there exist open halfspaces $L_i \supset C_i, i = 1, \ldots, m$, and a closed halfspace $H \supset D$ such that $L_1, \ldots, L_m$ have no common point in H.

10 Let $a^1, \ldots, a^m$ be a set of m vectors in $\mathbb{R}^n$. Show that the convex hull of this set contains 0 in its interior if and only if either of the following conditions holds:

1. For any nonzero vector $x \in \mathbb{R}^n$ one has $\max\{\langle a^i, x\rangle \mid i = 1, \ldots, m\} > 0$.
2. Any polyhedron $\langle a^i, x\rangle \leq \alpha_i,\ i = 1, \ldots, m$ is bounded.

11 Let C be a closed convex set and let $y^0 \notin C$. If $x^0 \in C$ is the point of C nearest to y^0, i.e., such that $\|x^0 - y^0\| = \min_{x \in C} \|x - y^0\|$, then $\|x - x^0\| < \|x - y^0\|\ \forall x \in C$. (*Hint*: From the proof of Proposition 1.15 it can be seen that x^0 is characterized by the condition $x^0 - y^0 \in N_C(x^0)$, i.e., $\langle x^0 - y^0, x - x^0\rangle \geq 0\ \forall x \in C$.)

12 (Fejer) For any set E in $\mathbb{R}^n$, a point y^0 belongs to the closure of conv$E$ if and only if for every $y \in \mathbb{R}^n$ there is $x \in E$ such that $\|x - y^0\| \leq \|x - y\|$. (Use Exercise 10 to prove the "only if" part.)

13 The radius of a compact set S is defined to be the number $\mathrm{rad}(S) = \min_x \max_{y \in S} \|x - y\|$. Show that for any $x \in \mathrm{conv}S$ there exists $y \in S$ such that $\|x - y\| \leq \mathrm{rad}(S)$.

14 Show by a counter-example that Theorem 1.7 (Sect. 1.7) is not true even for polyhedrons if the condition that C be line-free is omitted.

15 Show by a counter-example in $\mathbb{R}^3$ that Corollary 1.19 is not true if P is not full dimensional.

Chapter 2
Convex Functions

2.1 Definition and Basic Properties

Given a function $f : S \to [-\infty, +\infty]$ on a nonempty set $S \subset \mathbb{R}^n$, the sets

$$\mathrm{dom} f = \{x \in S | f(x) < +\infty\}$$
$$\mathrm{epi} f = \{(x, \alpha) \in S \times \mathbb{R} | f(x) \le \alpha\}$$

are called the *effective domain* and the *epigraph* of $f(x)$, respectively. If $\mathrm{dom} f \neq \emptyset$ and $f(x) > -\infty$ for all $x \in S$, then we say that the function $f(x)$ is *proper*.

A function $f : S \to [-\infty, +\infty]$ is called *convex* if its epigraph is a convex set in $\mathbb{R}^n \times R$. This is equivalent to saying that S is a convex set in $\mathbb{R}^n$ and for any $x^1, x^2 \in S$ and $\lambda \in [0, 1]$, we have

$$f((1-\lambda)x^1 + \lambda x^2) \le (1-\lambda)f(x^1) + \lambda f(x^2) \tag{2.1}$$

whenever the right-hand side is defined. In other words (2.1) must always hold unless $f(x^1) = -f(x^2) = \pm\infty$. By induction it can be proved that if $f(x)$ is convex then for any finite set $x^1, \ldots, x^k \in S$ and any nonnegative numbers $\lambda_1, \ldots, \lambda_k$ summing up to 1, we have

$$f\left(\sum_{i=1}^k \lambda_i x^i\right) \le \sum_{i=1}^k \lambda_i f(x^i)$$

whenever the right-hand side is defined. A function $f(x)$ is said to be *concave* on S if $-f(x)$ is convex; *affine* on S if $f(x)$ is finite and both convex and concave. An affine function on $\mathbb{R}^n$ has the form $f(x) = \langle a, x\rangle + \alpha$, with $a \in \mathbb{R}^n, \alpha \in \mathbb{R}$, because its epigraph is a halfspace in $\mathbb{R}^n \times \mathbb{R}$ containing no vertical line.

H. Tuy, *Convex Analysis and Global Optimization*, Springer Optimization and Its Applications 110, DOI 10.1007/978-3-319-31484-6_2

For a given nonempty convex set $C \subset \mathbb{R}^n$ we can define the convex functions:

- the *indicator function* of C : $\delta_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{if } x \notin C \end{cases}$
- the *support function* (see Sect. 1.8) $s_C(x) = \sup_{y \in C} \langle y, x \rangle$
- the *distance function* $d_C(x) = \inf_{y \in C} \|x - y\|$.

The convexity of $\delta_C(x)$ is obvious; that of the two other functions can be verified directly or derived from Propositions 2.5 and 2.9 below.

Proposition 2.1 *If $f(x)$ is an improper convex function on $\mathbb{R}^n$ then $f(x) = -\infty$ at every relative interior point x of its effective domain.*

Proof By the definition of an improper convex function, $f(x^0) = -\infty$ for at least some $x^0 \in \text{dom} f$ (unless $\text{dom} f = \emptyset$). If $x \in \text{ri}(\text{dom} f)$ then there is a point $x' \in \text{dom} f$ such that x is a relative interior point of the line segment $[x^0, x']$. Since $f(x') < +\infty$, it follows from $x = \lambda x^0 + (1 - \lambda)x'$ with $\lambda \in (0, 1)$, that $f(x) \leq \lambda f(x^0) + (1 - \lambda) f(x') = -\infty$. □

From the definition it is straightforward that a function $f(x)$ on $\mathbb{R}^n$ is convex if and only if its restriction to every straight line in $\mathbb{R}^n$ is convex. Therefore, convex functions on $\mathbb{R}^n$ can be characterized via properties of convex functions on the real line.

Theorem 2.1 *A real-valued function $f(x)$ on an open interval $(a, b) \subset \mathbb{R}$ is convex if and only if it is continuous and possesses at every $x \in (a, b)$ finite left and right derivatives*

$$f'_-(x) = \lim_{t \uparrow 0} \frac{f(x + t) - f(x)}{t}, \quad f'_+(x) = \lim_{t \downarrow 0} \frac{f(x + t) - f(x)}{t}$$

such that $f'_+(x)$ is nondecreasing and

$$f'_-(x) \leq f'_+(x), \qquad f'_+(x^1) \leq f'_-(x^2) \textit{ for } x^1 < x^2. \tag{2.2}$$

Proof

(i) Let $f(x)$ be convex. If $0 < s < t$ and $x + t < b$ then the point $(x + s, f(x + s))$ is below the segment joining $(x, f(x))$ and $(x + t, f(x + t))$, so

$$\frac{f(x + s) - f(x)}{s} \leq \frac{f(x + t) - f(x)}{t}. \tag{2.3}$$

This shows that the function $t \mapsto [f(x + t) - f(x)]/t$ is nonincreasing as $t \downarrow 0$. Hence it has a limit $f'_+(x)$ (finite or $= -\infty$). Analogously, $f'_-(x)$ exists (finite or $= +\infty$). Furthermore, setting $y = x + s$, $t = s + r$, we also have

$$\frac{f(x + s) - f(x)}{s} \leq \frac{f(y + r) - f(y)}{r}, \tag{2.4}$$

which implies $f'_+(x) \leq f'_+(y)$ for $x < y$, i.e., $f'_+(x)$ is nondecreasing. Finally, writing (2.4) as

$$\frac{f(y-s)-f(y)}{-s} \leq \frac{f(y+r)-f(y)}{r}.$$

and letting $-s \uparrow 0, r \downarrow 0$ yields $f'_-(y) \leq f'_+(y)$, proving the left part of (2.2) and at the same time the finiteness of these derivatives. The continuity of $f(x)$ at every $x \in (a,b)$ then follows from the existence of finite $f'_-(x)$ and $f'_+(x)$. Furthermore, setting $x = x^1, y + r = x^2$ in (2.4) and letting $s, r \to 0$ yields the right part of (2.2).

(ii) Now suppose that $f(x)$ has all the properties mentioned in the Proposition and let $a < c < d < b$. Consider the function

$$g(x) = f(x) - f(c) - (x-c)\frac{f(d)-f(c)}{d-c}.$$

Since for any $x = (1-\lambda)c + \lambda d$, we have $g(x) = f(x) - f(c) - \lambda[f(d) - f(c)] = f(x) - [(1-\lambda)f(c) + \lambda f(d)]$, to prove the convexity of $f(x)$ it suffices to show that $g(x) \leq 0$ for any $x \in [c,d]$. Suppose the contrary, that the maximum of $g(x)$ over the segment $[c,d]$ is positive (this maximum exists because $f(x)$ is continuous). Let $e \in [c,d]$ be the point where this maximum is attained. Note that $g(c) = g(d) = 0$, (hence $c < e < d$) and from its expression, $g(x)$ has the same properties as $f(x)$, namely: $g'_-(x), g'_+(x)$ exist at every $x \in (c,d)$, $g'_-(x) \leq g'_+(x)$, $g'_+(x)$ is nondecreasing and $g'_+(x^1) \leq g'_-(x^2)$ for $x^1 \leq x^2$. Since $g(e) \geq g(x)\ \forall x \in [c,d]$, we must have $g'_-(e) \geq 0 \geq g'_+(e)$, consequently $g'_-(e) = g'_+(e) = 0$, and hence, since $g'_+(x)$ is nondecreasing, $g'_+(x) \geq 0\ \forall x \in [e,d]$. If $g'_-(y) \leq 0$ for some $y \in (e,d]$ then $g'_+(x) \leq g'_-(y) \leq 0$ hence $g'(x) = 0$ for all $x \in [e,y)$, from which it follows that $g(y) = g(e) > 0$. Since $g(d) = 0$, there must exist $y \in (e,d)$ with $g'_-(y) > 0$. Let $x^1 \in [y,d)$ be the point where $g(x)$ attains its maximum over the segment $[y,d]$. Then $g'_+(x^1) \leq 0$, contradicting $g'_+(y) \geq g'_-(y) > 0$. Therefore $g(x) \leq 0$ for all $x \in [c,d]$, as was to be proved. □

Corollary 2.1 *A differentiable real-valued function $f(x)$ on an open interval is convex if and only if its derivative f' is a nondecreasing function. A twice differentiable real-valued function $f(x)$ on an open interval is convex if and only if its second derivative f'' is nonnegative throughout this interval.* □

Proposition 2.2 *A twice differentiable real-valued function $f(x)$ on an open convex set C in $\mathbb{R}^n$ is convex if and only if for every $x \in C$ its Hessian matrix*

$$Q_x = (q_{ij}(x)), \quad q_{ij}(x) = \frac{\partial^2 f}{\partial x_i \partial x_j}(x_1, \ldots, x_n)$$

is positive semidefinite, i.e.,

$$\langle u, Q_x u\rangle \geq 0 \ \forall u \in \mathbb{R}^n.$$

Proof The function f is convex on C if and only if for each $a \in C$ and $u \in \mathbb{R}^n$ the function $\varphi_{a,u}(t) = f(a + tu)$ is convex on the open real interval $\{t|\ a + tu \in C\}$. The proposition then follows from the preceding corollary since an easy computation yields $\varphi''(t) = \langle u, Q_x u\rangle$ with $x = a + tu$. □

In particular, a *quadratic function*

$$f(x) = \frac{1}{2}\langle x, Qx\rangle + \langle x, a\rangle + \alpha,$$

where Q is a symmetric $n \times n$ matrix, is convex on $\mathbb{R}^n$ if and only if Q is positive semidefinite. It is concave on $\mathbb{R}^n$ if and only if its matrix Q is negative semidefinite.

Proposition 2.3 *A proper convex function f on $\mathbb{R}^n$ is continuous at every interior point of its effective domain.*

Proof Let $x^0 \in \text{int}(\text{dom} f)$. Without loss of generality one can assume $x^0 = 0$. By Theorem 2.1, for each $i = 1, \ldots, n$ the restriction of f to the open interval $\{t|\ x^0 + te^i \in \text{int}(\text{dom} f)\}$ is continuous relative to this interval. Hence for any given $\varepsilon > 0$ and for each $i = 1, \ldots, n$, we can select $\delta_i > 0$ so small that $|f(x) - f(x^0)| \leq \varepsilon$ for all $x \in [-\delta_i e^i, +\delta_i e^i]$. Let $\delta = \min\{\delta_i|\ i = 1, \ldots, n\}$ and $B = \{x|\ \|x\|_1 \leq \delta\}$. Denote $u^i = \delta e^i, u^{i+n} = -\delta e^i,\ i = 1, \ldots, n$. Then, as seen in the proof of Corollary 1.6, any $x \in B$ is of the form $x = \sum_{i=1}^{2n} \lambda_i u^i$, with $\sum_{i=1}^{2n} \lambda_i = 1, 0 \leq \lambda_i \leq 1$, hence $f(x) \leq \sum_{i=1}^{2n} \lambda_i f(u^i)$, and consequently, $f(x) - f(x^0) \leq \sum_{i=1}^{2n} \lambda_i [f(x^i) - f(x^0)]$. Therefore,

$$|f(x) - f(x^0)| \leq \sum_{i=1}^{2n} \lambda_i |f(x) - f(x^0)| \leq \varepsilon$$

for all $x \in B$, proving the continuity of $f(x)$ at x^0. □

Proposition 2.4 *Let f be a real-valued function on a convex set $C \subset \mathbb{R}^n$. If for every $x \in C$ there exists a convex open neighborhood U_x of x such that f is convex on $U_x \cap C$ then f is convex on C.*

Proof It suffices to show that for every $a \in C, u \in \mathbb{R}^n$, the function $\varphi(t) = f(a+tu)$ is convex on the interval $\Delta := \{t|\ a + tu \in C\}$. But from the hypothesis, this function is convex in a neighborhood of every $t \in \Delta$, hence is continuous and has left and right derivatives $\varphi'_-(t) \leq \varphi'_+(t)$ which are non decreasing in a neighborhood of every $t \in \Delta$. These derivatives thus exist and satisfy the conditions described in Theorem 2.1 on the whole interval Δ. Hence, $\varphi(t)$ is convex. □

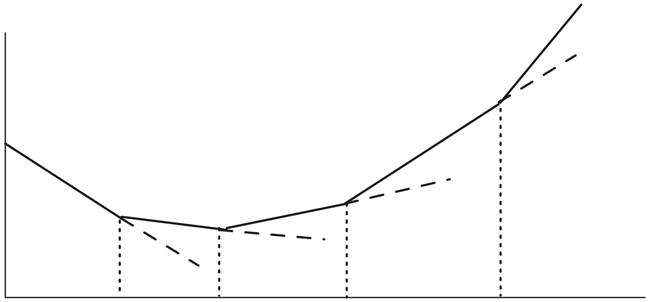


Fig. 2.1 Convex piecewise affine function

For example, let $f(x) : \mathbb{R} \to \mathbb{R}$ be a piecewise convex function on the real line, i.e., a function such that the real line can be partitioned into a finite number of intervals $\Delta_i, i = 1, \ldots, N$, in each of which $f(x)$ is convex. Then $f(x)$ is convex if and only if it is convex in the neighborhood of each breakpoint (endpoint of some interval Δ_i). In particular, a piecewise affine function on the real line is convex if and only if at each breakpoint the left slope is at most equal to the right slope (in other words, the sequence of slopes is nondecreasing) (see Fig. 2.1).

2.2 Operations That Preserve Convexity

A function with a complicated expression may be built up from a number of simpler ingredient functions via certain standard operations. The convexity of such a function can often be established indirectly, by proving that the ingredient functions are known convex functions, whereas the operations involved in the composition of the ingredient functions preserve convexity. It is therefore useful to be familiar with some of the most important operations which preserve convexity.

Proposition 2.5 *A positive combination of finitely many proper convex functions on $\mathbb{R}^n$ is convex. The upper envelope (pointwise supremum) of an arbitrary family of convex functions is convex.*

Proof If $f(x)$ is convex and $\alpha \geq 0$, then $\alpha f(x)$ is obviously convex. If f_1 and f_2 are proper convex functions on $\mathbb{R}^n$, then it is also evident that $f_1 + f_2$ is convex. This proves the first part of the proposition. The second part follows from the facts that if $f(x) = \sup\{f_i(x) |\ i \in I\}$, then $\text{epi} f = \cap_{i \in I} \text{epi} f_i$, and the intersection of a family of convex sets is a convex set. □

Proposition 2.6 *Let Ω be a convex set in $\mathbb{R}^n$, G a convex set in $\mathbb{R}^m$, $\varphi(x, y)$ a real-valued convex function on $\Omega \times G$. Then the function*

$$f(x) = \inf_{y \in G} \varphi(x, y)$$

is convex on Ω.

Proof Let $x^1, x^2 \in \Omega$ and $x = \lambda x^1 + (1-\lambda)x^2$ with $\lambda \in [0,1]$. For each $i = 1, 2$ select a sequence $\{y^{i,k}\} \subset G$ such that

$$\varphi(x^i, y^{i,k}) \to \inf_{y \in G} \varphi(x^i, y).$$

By convexity of φ,

$$f(x) \le \varphi(x, \lambda y^{1,k} + (1-\lambda)y^{2,k}) \le \lambda\varphi(x^1, y^{1,k}) + (1-\lambda)\varphi(x^2, y^{2,k}),$$

hence, letting $k \to \infty$ yields

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

□

Proposition 2.7 *If $g_i(x), i = 1, \ldots, m$, are concave positive functions on a convex set $C \subset \mathbb{R}^n$ then their geometric mean*

$$f(x) = \left[\prod_{i=1}^m g_i(x)\right]^{1/m} \tag{2.5}$$

is a concave function (so $-f(x)$ is a convex function) on C.

Proof Let $T = \{t \in \mathbb{R}^m_+ |\ \prod_{i=1}^m t_i \ge 1\}$. We show that for any fixed $x \in C$:

$$\left[\prod_{i=1}^m g_i(x)\right]^{1/m} = \frac{1}{m}\min_{t \in T}\left\{\sum_{i=1}^m t_i g_i(x)\right\}. \tag{2.6}$$

Indeed, observe that the right-hand side of (2.6) is equal to

$$\frac{1}{m}\min\left\{\sum_{i=1}^m t_i g_i(x) |\ \prod_{i=1}^m t_i = 1,\ t_i > 0,\ i = 1, \ldots, m\right\}$$

since if $\prod_{i=1}^m t_i > 1$ then by replacing t_i with $t'_i \le t_i$ such that $\prod_{i=1}^m t'_i = 1$, we can only decrease the value of $\sum_{i=1}^m t_i g_i(x)$. Therefore, it can be assumed that $\prod_{i=1}^m t_i = 1$ and hence

$$\prod_{i=1}^m t_i g_i(x) = \prod_{i=1}^m g_i(x). \tag{2.7}$$

Since $t_i g_i(x) > 0, i = 1, \ldots, m$, and for fixed x the product of these positive numbers is constant [$= \prod_{i=1}^m g_i(x)$ by (2.7)] their sum is minimal when these numbers are equal (by theorem on arithmetic and geometric mean). That is, taking account of (2.7), the minimum of $\sum_{i=1}^m t_i g_i(x)$ is achieved when $t_i g_i(x) = [\prod_{i=1}^m g_i(x)]^{1/m}$ $\forall i = 1,\dots,m$, hence (2.6). Since for fixed $t \in T$ the function $x \mapsto \varphi_t(x) := \frac{1}{m}\sum_{i=1}^m t_i g_i(x)$ is concave by Proposition 2.5, their lower envelope $\inf_{t\in T}\varphi_t(x) = [\prod_{i=1}^m g_i(x)]^{1/m}$ is concave by the same proposition. □

An interesting concave function of the class (2.5) is the geometric mean:

$$f(x) = \begin{cases} (x_1 x_2 \cdots x_n)^{1/n} & \text{if } x_1 \geq 0, \dots, x_n \geq 0 \\ -\infty & \text{otherwise} \end{cases}$$

which corresponds to the case when $g_i(x) = x_i$.

Proposition 2.8 *Let $g(x) : \mathbb{R}^n \to (-\infty, +\infty)$ be a convex function and let $\varphi(t) : \mathbb{R} \to (-\infty, +\infty)$ be a nondecreasing convex function. Then $f(x) = \varphi(g(x))$ is convex on $\mathbb{R}^n$.*

Proof The proof is straightforward. For any $x^1, x^2 \in \mathbb{R}^n$ and $\lambda \in [0, 1]$ we have

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

hence

$$\varphi(g((1-\lambda)x^1 + \lambda x^2)) \leq (1-\lambda)\varphi(g(x^1)) + \lambda\varphi(g(x^2)).$$

□

For example, by this proposition the function $f(x) = \sum_{i=1}^m c_i e^{g_i(x)}$ is convex if $c_i > 0$ and each $g_i(x)$ is convex proper.

Given the epigraph E of a convex function $f(x)$, one can restore $f(x)$ by the formula

$$f(x) = \inf\{t \mid (x, t) \in E\}. \tag{2.8}$$

Conversely, given a convex set $E \subset \mathbb{R}^{n+1}$ the function $f(x)$ defined by (2.8) is a convex function on $\mathbb{R}^n$ by Proposition 2.6. Therefore, if $f_1, \dots, f_m$ are m given convex functions, and $E \subset \mathbb{R}^{n+1}$ is a convex set resulting from some operation on their epigraphs $E_1, \dots, E_m$, then one can use (2.8) to define a corresponding new convex function $f(x)$.

Proposition 2.9 *Let $f_1, \dots, f_m$ be proper convex functions on $\mathbb{R}^n$. Then*

$$f(x) = \inf\left\{\sum_{i=1}^m f_i(x^i) \mid x^i \in \mathbb{R}^n,\ \sum_{i=1}^m x^i = x\right\}$$

is a convex function on $\mathbb{R}^n$.

Proof Indeed, $f(x)$ is defined by (2.8), where $E = E_1 + \cdots + E_m$ and $E_i = \text{epi} f_i$, $i = 1, \ldots, m$. $\square$

The above constructed function $f(x)$ is called the *infimal convolution* of the functions $f_1, \ldots, f_m$. For example, the convexity of the distance function $d_C(x) = \inf\{\|x - y\| \mid y \in C\}$ associated with a convex set C follows from the above proposition because $d_C(x) = \inf_y\{\|x-y\|+\delta_C(y)\} = \inf\{\|x^1\|+\delta_C(x^2) | \, x^1+x^2 = x\}$.

Let $g(x)$ now be a nonconvex function, so that its epigraph is a nonconvex set. The relation (2.8) where $E = \text{conv}(\text{epi} g)$ defines a function $f(x)$ called the *convex envelope* or *convex hull* of $g(x)$ and denoted by conv$g$. Since E is the smallest convex set containing the epigraph of g it is easily seen that conv$g$ is the largest convex function majorized by g.

When C is a subset of $\mathbb{R}^n$, the convex envelope of the function

$$g|C = \begin{cases} g(x) & \text{if } x \in C \\ +\infty & \text{if } x \notin C \end{cases}$$

is called the *convex envelope of g over C*.

Proposition 2.10 *Let $g : \mathbb{R}^n \to \mathbb{R}$. The convex envelope of g over a set $C \subset \mathbb{R}^n$ such that dim(aff C) $= k$ is given by*

$$f(x) = \inf\left\{\sum_{i=1}^{k+1} \lambda_i g(x^i) | \, x^i \in C, \lambda_i \geq 0, \sum_{i=1}^{k+1} \lambda_i = 1, \sum_{i=1}^{k+1} \lambda_i x^i = x\right\}.$$

Proof Let $X = C \times \{0\} \subset \mathbb{R}^n \times \mathbb{R},\ B = \{(0,t) | \, 0 \leq t \leq 1\} \subset \mathbb{R}^n \times \mathbb{R}$ and define $E = \{(x, g(x)) | \, x \in C\} = \bigcup_{(x,0)\in X}((x,0) + g(x)B)$. Then X is a Caratheodory core of E (cf Sect. 1.4) and since $\dim(\text{aff} X) = k$, by Proposition 1.14, we have

$$\text{conv}E = \left\{(x,t) = \sum_{i=1}^{k+1} \lambda_i(x^i, g(x^i)) | \, x^i \in C,\ \lambda_i \geq 0,\ \sum_{i=1}^{k+1} \lambda_i = 1\right\}.$$

But clearly for every $(x,t) \in \text{epi} g$ there is $\theta \leq t$ such that $(x,\theta) \in E$, so for every $(x,t) \in \text{conv}(\text{epi} g)$ there is $\theta \leq t$ such that $(x,\theta) \in \text{conv}E$. Therefore, $(\text{conv} g)(x) = \inf\{t | \, (x,t) \in \text{conv}(\text{epi} g)\} = \inf\{t | \, (x,t) \in \text{conv}E\} = \inf\{\sum_{i=1}^{k+1} \lambda_i g(x^i) | \, x^i \in C,$ $\lambda_i \geq 0, \sum_{i=1}^{k+1} \lambda_i x^i = x, \sum_{i=1}^{k+1} \lambda_i = 1\}$. $\square$

Corollary 2.2 *The convex envelope of a concave function $g : \mathbb{R}^n \to \mathbb{R}$ over a polytope D in $\mathbb{R}^n$ with vertex set V is the function*

$$f(x) = \min\left\{\sum_{i=1}^{n+1} \lambda_i g(v^i) | \, v^i \in V, \lambda_i \geq 0, \sum_{i=1}^{n+1} \lambda_i = 1, \sum_{i=1}^{n+1} \lambda_i v^i = x\right\}.$$

Proof By the above proposition, $(\text{conv}g)(x) \leq f(x)$ but any $x \in D$ is of the form $x = \sum_{i=1}^{n+1} \lambda_i v^i$, with $v^i \in V, \lambda_i \geq 0, \sum_{i=1}^{n+1} \lambda_i = 1$, hence $f(x) \leq \sum_{i=1}^{n+1} \lambda_i g(v^i)$ (by definition of $f(x)$), while $\sum_{i=1}^{n+1} \lambda_i g(v^i) \leq g(x)$ by the concavity of g. Therefore, $f(x) \leq g(x)\ \forall x \in D$, and since $f(x)$ is convex, we must have $f(x) \leq (\text{conv}g)(x)$, and hence $f(x) = (\text{conv}g)(x)$. □

2.3 Lower Semi-Continuity

Given a function $f : \mathbb{R}^n \to [-\infty, +\infty]$ the sets

$$\{x | f(x) \leq \alpha\}, \quad \{x | f(x) \geq \alpha\}$$

where $\alpha \in [-\infty, +\infty]$ are called *lower* and *upper level sets*, respectively, of f.

Proposition 2.11 *The lower (upper, resp.) level sets of a convex (concave, resp.) function $f(x)$ are convex.*

Proof This property is equivalent to

$$f((1 - \lambda)x^1 + \lambda x^2) \leq \max\{f(x^1), f(x^2)\} \quad \forall \lambda \in (0, 1) \tag{2.9}$$

for all $x^1, x^2 \in \mathbb{R}^n$, which is an immediate consequence of the definition of convex functions. □

Note that the converse of this proposition is not true. For example, a real-valued function on the real line which is nondecreasing has all its lower level sets convex, but may not be convex. A function $f(x)$ whose every nonempty lower level set is convex (or, equivalently, which satisfies (2.9) for all $x^1, x^2 \in \mathbb{R}^n$), is said to be *quasiconvex*. If every nonempty upper level set is convex, $f(x)$ is said to be *quasiconcave*.

Proposition 2.12 *For any proper convex function f :*

(i) *The maximum of f over any line segment is attained at one endpoint.*
(ii) *If $f(x)$ is finite and bounded above on a halfline, then its maximum over the halfline is attained at the origin of the halfline.*
(iii) *If $f(x)$ is finite and bounded above on an affine set then it is constant on this set.*

Proof

(i) Immediate from Proposition 2.11.
(ii) If $f(b) > f(a)$ then for any $x = b + \lambda(b - a)$ with $\lambda \geq 0$, we have $b = \frac{1}{1+\lambda}x + \frac{\lambda}{1+\lambda}a$, hence $(1+\lambda)f(b) \leq f(x) + \lambda f(a)$, (whenever $f(x) < +\infty$),

i.e., $f(x) \geq \lambda[f(b) - f(a)] + f(b)$, which implies $f(x) \to +\infty$ as $\lambda \to +\infty$. Therefore, if $f(x)$ is finite and bounded above on a halfline of origin a, one must have $f(b) \leq f(a)$ for every b on this halfline.

(iii) Let M be an affine set on which $f(x)$ is finite. If $f(b) > f(a)$ for $a, b \in M$, then by (ii), $f(x)$ is unbounded above on the halfline in M from a through b. Therefore, if $f(x)$ is bounded above on M, it must be constant on M. □

A function f from a set $S \subset \mathbb{R}^n$ to $[-\infty, +\infty]$ is said to be *lower semi-continuous* (l.s.c.) at a point $x \in S$ if

$$\liminf_{\substack{y \in S \\ y \to x}} f(y) \geq f(x).$$

It is said to be *upper semi-continuous* (u.s.c.) at $x \in S$ if

$$\limsup_{\substack{y \in S \\ y \to x}} f(y) \leq f(x).$$

A function which is both lower and upper semi-continuous at x is continuous at x in the ordinary sense.

Proposition 2.13 *Let S be a closed set in $\mathbb{R}^n$. For an arbitrary function $f : S \to [-\infty, +\infty]$ the following conditions are equivalent:*

(i) *The epigraph of f is a closed set in $\mathbb{R}^{n+1}$;*
(ii) *For every $\alpha \in \mathbb{R}$ the set $\{x \in S |\, f(x) \leq \alpha\}$ is closed;*
(iii) *f is lower semi-continuous throughout S.*

Proof (i)$\Rightarrow$ (ii). Let $x^\nu \in S, x^\nu \to x, f(x^\nu) \leq \alpha$. Since $(x^\nu, \alpha) \in \mathrm{epi} f$, it follows from the closedness of epi$f$ that $(x, \alpha) \in \mathrm{epi} f$, i.e., $f(x) \leq \alpha$, proving (ii).

(ii)$\Rightarrow$(iii). Let $x^\nu \in S, x^\nu \to x$. If $\lim_{\nu \to \infty} f(x^\nu) < f(x)$ then there is $\alpha < f(x)$ such that $f(x^\nu) \leq \alpha$ for all sufficiently large ν. From (ii) it would then follow that $f(x) \leq \alpha$, a contradiction. Therefore $\lim_{\nu \to \infty} f(x^\nu) \geq f(x)$, proving (iii).

(iii)$\Rightarrow$ (i). Let $(x^\nu, t^\nu) \in \mathrm{epi} f$, (i.e., $f(x^\nu) \leq t^\nu$) and $(x^\nu, t^\nu) \to (x, t)$. Then from (iii), we have $\liminf f(x^\nu) \geq f(x)$, hence $t \geq f(x)$, i.e., $(x, t) \in \mathrm{epi} f$. □

Proposition 2.14 *Let f be a l.s.c. proper convex function. Then all the nonempty lower level sets $\{x |\, f(x) \leq \alpha\}$, $\alpha \in \mathbb{R}$, have the same recession cone and the same lineality space. The recession cone is made up of 0 and the directions of halflines over which f is bounded above, while the lineality space is the space parallel to the affine set on which f is constant.*

Proof By Proposition 2.13 every lower level set $C_\alpha := \{x |\, f(x) \leq \alpha\}$ is closed. Let $\Gamma = \{\lambda u |\, \lambda \geq 0\}$. If f is bounded above on a halfline $\Gamma_a = a + \Gamma$, then $f(a) \in \mathbb{R}$ (because f is proper) and by Proposition 2.12 $f(x) \leq f(a)\ \forall x \in \Gamma_a$. For any nonempty $C_\alpha, \alpha \in \mathbb{R}$, consider a point $b \in C_\alpha$ and let $\beta = \max\{f(a), \alpha\}$ so that

$\beta \in \mathbb{R}$ and $\Gamma_a \subset C_\beta$. Since C_β is closed and $b \in C_\beta$ it follows from Lemma 1.1 that $\Gamma_b \subset C_\beta$, i.e., $f(x)$ is finite and bounded above on Γ_b. Then by Proposition 2.12, $f(x) \le f(b) \le \alpha\ \forall x \in \Gamma_b$, hence $\Gamma_b \subset C_\alpha$. Thus, if f is bounded above on a halfline Γ_a then Γ is a direction of recession for every nonempty $C_\alpha, \alpha \in R$. The converse is obvious. Therefore, the recession cone of C_α is the same for all α and is made up of 0 and all directions of halflines over which f is bounded above. The rest of the proposition is straightforward. □

The recession cone and the lineality space common to each lower level set of f are also called the *recession cone* and the *constancy space* of f, respectively.

Corollary 2.3 *If the lower level set $\{x|\, f(x) \le \alpha\}$ of a l.s.c. proper convex function f is nonempty and bounded for one α then it is bounded for every α.*

Proof Any lower level set of f is a closed convex set. Therefore, it is bounded if and only if its recession cone is the singleton $\{0\}$ (Corollary 1.8). □

Corollary 2.4 *If a l.s.c. proper convex function f is bounded above on a halfline then it is bounded above on every parallel halfline emanating from a point of dom$f$. If it is constant on a line then it is constant on every parallel line passing through a point of dom$f$.*

Proof Immediate. □

Proposition 2.15 *Let f be any proper convex function on $\mathbb{R}^n$. For any $y \in \mathbb{R}^n$, there exists $t \in \mathbb{R}$ such that (y, t) belongs to the lineality space of* epi$f$ *if and only if*

$$f(x + \lambda y) = f(x) + \lambda t \quad \forall x \in \mathrm{dom} f, \forall \lambda \in \mathbb{R}. \tag{2.10}$$

When f is l.s.c., this condition is satisfied provided for some $x \in$ dom$f$ *the function $\lambda \mapsto f(x + \lambda y)$ is affine.*

Proof (y, t) belongs to the lineality space of epi$f$ if and only if for any $x \in \mathrm{dom} f$: $(x, f(x)) + \lambda(y, t) \in \mathrm{epi} f\ \forall \lambda \in \mathbb{R}$, i.e., if and only if

$$f(x + \lambda y) - \lambda t \le f(x) \quad \forall \lambda \in \mathbb{R}.$$

By Proposition 2.12 applied to the proper convex function $\varphi(\lambda) = f(x+\lambda t)-\lambda t$, this is equivalent to saying that $\varphi(\lambda) =$ constant, i.e., $f(x + \lambda y) - \lambda t = f(x)\ \forall \lambda \in \mathbb{R}$. This proves the first part of the proposition. If f is l.s.c., i.e., epi$f$ is closed, then (y, t) belongs to the lineality space of epi$f$ provided for some $x \in \mathrm{dom} f$ the line $\{(x, f(x)) + \lambda(y, t)|\ \lambda \in \mathbb{R}\}$ is contained in epi$f$ (Lemma 1.1). □

The projection of the lineality space of epi$f$ on $\mathbb{R}^n$, i.e., the set of vectors y for which there exists t such that (y, t) belongs to the lineality space of epi$f$, is called the *lineality space* of f. The directions of these vectors y are called *directions in which f is affine*. The dimension of the lineality space of f is called the *lineality* of f.

By definition, the *dimension of a convex function* f is the dimension of its domain. The number $\dim f - \text{lineality} f$ which is a measure of the nonlinearity of f is then called the *rank* of f :

$$\text{rank} f = \dim f - \text{lineality} f. \tag{2.11}$$

Corollary 2.5 *If a proper convex function f of full dimension on $\mathbb{R}^n$ has rank k then there exists a $k \times n$ matrix B with $\text{rank} B = k$ such that for any $b \in B(\text{dom} f)$ the restriction of f to the affine set $Bx = b$ is affine.*

Proof Let L be the lineality space of f. Since $\dim L = n - k$, there exists a $k \times n$ matrix B of rank k such that $L = \{u \in \mathbb{R}^n | \, Bu = 0\}$. If $b = Bx^0$ for some $x^0 \in \text{dom} f$ then for any x such that $Bx = b$, we have $B(x - x^0) = 0$, hence, denoting by $u^1, \ldots, u^h$ a basis of L, $x = x^0 + \sum_{i=1}^h \lambda_i u^i$. By Proposition 2.15 there exist $t_i \in R$ such that $f(x^0 + \sum_{i=1}^h \lambda_i u^i) = f(x^0) + \sum_{i=1}^h \lambda_i t_i$ for all $\lambda \in \mathbb{R}^h$. Therefore, $f(x)$ is affine on the affine set $Bx = b$. □

Given any proper function f on $\mathbb{R}^n$, the function whose epigraph is the closure of $\text{epi} f$ is the largest l.s.c. minorant of f. It is called the *l.s.c. hull* of f or the *closure* of f, and is denoted by $\text{cl} f$. Thus, for a proper function f,

$$\text{epi}(\text{cl} f) = \text{cl}(\text{epi} f). \tag{2.12}$$

A proper convex function f is said to be *closed* if $\text{cl} f = f$ (so for proper convex functions, closed is synonym of l.s.c.). An improper convex function is said to be closed only if $f \equiv +\infty$ or $f \equiv -\infty$ (so if $f(x) = -\infty$ for some x then its closure is the constant function $-\infty$).

Proposition 2.16 *The closure of a proper convex function f is a proper convex function which agrees with f except perhaps at the relative boundary points of $\text{dom} f$.*

Proof Since $\text{epi} f$ is convex, $\text{cl}(\text{epi} f)$, i.e., $\text{epi}(\text{cl} f)$, is also convex (Proposition 1.10). Hence by Proposition 2.13, $\text{cl} f$ is a closed convex function. Now the condition (2.12) is equivalent to

$$\text{cl} f(x) = \liminf_{y \to x} f(y) \quad \forall x \in \mathbb{R}^n. \tag{2.13}$$

If $x \in \text{ri}(\text{dom} f)$ then by Proposition 2.3 $f(x) = \lim_{y \to x} f(y)$, hence, by (2.13), $f(x) = \text{cl} f(x)$. Furthermore, if $x \notin \text{cl}(\text{dom} f)$ then $f(y) = +\infty$ for all y in a neighborhood of x and the same formula (2.13) shows that $\text{cl} f(x) = +\infty$. Thus the second half of the proposition is true. It remains to prove that $\text{cl} f$ is proper. For every $x \in \text{ri}(\text{dom} f)$, since f is proper, $-\infty < \text{cl} f(x) = f(x) < +\infty$. On the other hand, if $\text{cl} f(x) = -\infty$ at some relative boundary point x of $\text{dom} f = \text{dom}(\text{cl} f)$ then for an arbitrary $y \in \text{ri}(\text{dom} f)$, we have $\text{cl} f\left(\frac{x+y}{2}\right) \le \frac{1}{2}\text{cl} f(x) + \frac{1}{2}\text{cl} f(y) = -\infty$. Noting that $\frac{x+y}{2} \in \text{ri}(\text{dom} f)$ this contradicts what has just been proved and thereby completes the proof. □

2.4 Lipschitz Continuity

By Proposition 2.3 a convex function is continuous relative to the relative interior of its domain. In this section we present further continuity properties of convex functions.

Proposition 2.17 *Let f be a convex function on $\mathbb{R}^n$ and D a polyhedron contained in $\mathrm{dom} f$. Then f is u.s.c. relative to D, so that if f is l.s.c. f is actually continuous relative to D.*

Proof Consider any $x \in D$. By translating if necessary, we can assume that $x = 0$. Let $e^1, \ldots, e^n$ be the unit vectors of $\mathbb{R}^n$ and C the convex hull of the set $\{e^1, \ldots, e^n, -e^1, \ldots, -e^n\}$. This set C is partitioned by the coordinate hyperplanes into simplices $S_i, i = 1, \ldots, 2^n$, with a common vertex at 0. Since D is a polyhedron, each $D_i = S_i \cap D$ is a polytope and obviously $D \cap C = \cup_{i=1}^{2^n} D_i$. Now let $\{x^k\} \subset D \cap C$ be any sequence such that $x^k \to 0, f(x^k) \to \gamma$. Then at least one D_i, say D_1, contains an infinite subsequence of $\{x^k\}$. For convenience we also denote this subsequence by $\{x^k\}$. If V_1 is the set of vertices of D_1 other than 0, then each x^k is a convex combination of 0 and elements of V_1 : $x^k = (1 - \sum_{v \in V_1} \lambda_v^k)0 + \sum_{v \in V_1} \lambda_v^k v$, with $\lambda_v^k \geq 0, \sum_{v \in V_1} \lambda_v^k \leq 1$. By convexity of f we can write

$$f(x^k) \leq (1 - \sum_{v \in V_1} \lambda_v^k) f(0) + \sum_{v \in V_1} \lambda_v^k f(v).$$

As $k \to +\infty$, since $x^k \to 0$, it follows that $\lambda_v^k \to 0 \ \forall v$, hence $\gamma \leq f(0)$, i.e.,

$$\limsup f(x^k) \leq f(0).$$

This proves the upper semi-continuity of f at 0 relative to D. □

Theorem 2.2 *For a proper convex function f on $\mathbb{R}^n$ the following assertions are equivalent:*

(i) *f is continuous at some point;*
(ii) *f is bounded above on some open set;*
(iii) $\mathrm{int}(\mathrm{epi} f) \neq \emptyset$;
(iv) $\mathrm{int}(\mathrm{dom} f) \neq \emptyset$ *and f is Lipschitzian on every bounded set contained in* $\mathrm{int}(\mathrm{dom} f)$;
(v) $\mathrm{int}(\mathrm{dom} f) \neq \emptyset$ *and f is continuous there.*

Proof (i) $\Rightarrow$ (ii) If f is continuous at a point x^0, then there exists an open neighborhood U of x^0 such that $f(x) < f(x^0) + 1$ for all $x \in U$.

(ii) $\Rightarrow$ (iii) If $f(x) \leq c$ for all x in an open set U, then $U \times [c, +\infty) \subset \mathrm{epi} f$, hence $\mathrm{int}(\mathrm{epi} f) \neq \emptyset$.

(iii) $\Rightarrow$ (iv) If $\mathrm{int}(\mathrm{epi} f) \neq \emptyset$, then there exists an open set U and an open interval $I \subset R$ such that $U \times I \subset \mathrm{epi} f$, hence $U \subset \mathrm{dom} f$, i.e., $\mathrm{int}(\mathrm{dom} f) \neq \emptyset$. Consider any compact set $C \subset \text{int}(\text{dom}f)$ and let B be the Euclidean unit ball. For every $r > 0$ the set $C + rB$ is compact, and the family of closed sets $\{(C + rB) \setminus \text{int}(\text{dom}f), r > 0\}$ has an empty intersection. In view of the compactness of $C + rB$ some finite subfamily of this family must have an empty intersection, hence for some $r > 0$, we must have $(C + rB) \setminus \text{int}(\text{dom}f) = \emptyset$, i.e., $C + rB \subset \text{int}(\text{dom}f)$. By Proposition 2.3 the function f is continuous on $\text{int}(\text{dom}f)$. Denote by μ_1 and μ_2 the maximum and the minimum of f over $C + rB$. Let x, x' be two distinct points in C and let $z = x + \frac{r(x-x')}{\|x-x'\|}$. Then $z \in C + rB \subset \text{int}(\text{dom}f)$. But

$$x = (1-\alpha)x' + \alpha z, \quad \alpha = \frac{\|x - x'\|}{r + \|x - x'\|}$$

and $z, x' \in \text{dom}f$, hence

$$f(x) \leq (1-\alpha)f(x') + \alpha f(z) = f(x') + \alpha(f(z) - f(x'))$$

and consequently

$$\begin{aligned} f(x) - f(x') &\leq \alpha(f(z) - f(x') \leq \alpha(\mu_1 - \mu_2) \\ &\leq \gamma\|x - x'\|, \quad \gamma = \frac{\mu_1 - \mu_2}{r}. \end{aligned}$$

By symmetry, we also have $f(x') - f(x) \leq \gamma\|x - x'\|$. Hence, for all x, x' such that $x \in C, x' \in C$:

$$|f(x) - f(x')| \leq \gamma\|x - x'\|,$$

proving the Lipschitz property of f over C.

(iv)$\Rightarrow$ (v) and (v)$\Rightarrow$ (i): obvious. □

2.5 Convex Inequalities

A *convex inequality* in x is an inequality of the form $f(x) \leq 0$ or $f(x) < 0$ where f is a convex function. Note that an inequality like $f(x) \geq 0$ or $f(x) > 0$, with $f(x)$ convex, is not convex but *reverse convex*, because it becomes convex only when reversed. A system of inequalities is said to be *consistent* if it has a solution, i.e., if there exists at least one value x satisfying all the inequalities; it is *inconsistent* otherwise. Many mathematical problems reduce to investigating the consistency (or inconsistency) of a system of inequalities.

Proposition 2.18 *Let $f_0, f_1, \ldots, f_m$ be convex functions, finite on some nonempty convex set $D \subset \mathbb{R}^n$. If the system*

$$x \in D, \; f_i(x) < 0 \;\; i = 0, 1, \ldots, m \tag{2.14}$$

is inconsistent, then there exist multipliers $\lambda_i \geq 0$, $i = 0, 1, \ldots, m$, *such that* $\sum_{i=0}^{m} \lambda_i > 0$ *and*

$$\lambda_0 f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \geq 0 \quad \forall x \in D. \tag{2.15}$$

If in addition

$$\exists x^0 \in D \quad f_i(x^0) < 0 \;\; i = 1, \ldots, m \tag{2.16}$$

then $\lambda_0 > 0$, *so that one can take* $\lambda_0 = 1$.

Proof Consider the set C of all vectors $y = (y_0, y_1, \ldots, y_m) \in \mathbb{R}^{m+1}$ for each of which there exists an $x \in D$ satisfying

$$f_i(x) < y_i, \quad i = 0, 1, \ldots, m. \tag{2.17}$$

As can readily be verified, C is a nonempty convex set and the inconsistency of the system (2.14) means that $0 \notin C$. By Lemma 1.2 there is a hyperplane in $\mathbb{R}^{m+1}$ properly separating 0 from C, i.e., a vector $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m) \neq 0$ such that

$$\sum_{i=0}^{m} \lambda_i y_i \geq 0 \quad \forall y = (y_0, y_1, \ldots, y_m) \in C. \tag{2.18}$$

If $x \in D$ then for every $\varepsilon > 0$, we have $f_i(x) < f_i(x) + \varepsilon$ for $i = 0, 1, \ldots, m$, so $(f_0(x) + \varepsilon, \ldots, f_m(x) + \varepsilon) \in C$ and hence,

$$\sum_{i=0}^{m} \lambda_i (f_i(x) + \varepsilon) \geq 0.$$

Since $\varepsilon > 0$ can be arbitrarily small, this implies (2.15). Furthermore, $\lambda_i \geq 0$, $i = 0, 1, \ldots, m$ because if $\lambda_j < 0$ for some j, then by fixing, for an arbitrary $x \in D$, all $y_i > f_i(x), i \neq j$, while letting $y_j \to +\infty$, we would have $\sum_{i=0}^{m} \lambda_i y_i \to -\infty$, contrary to (2.18). Finally, under (2.16), if $\lambda_0 = 0$ then $\sum_{i=1}^{m} \lambda_i > 0$, hence by (2.16) $\sum_{i=1}^{m} \lambda_i f_i(x^0) < 0$, contradicting the inequality $\sum_{i=1}^{m} \lambda_i f_i(x^0) \geq 0$ from (2.15). Therefore, $\lambda_0 > 0$ as was to be proved. □

Corollary 2.6 *Let D be a convex set in $\mathbb{R}^n$, g, f two convex functions finite on D. If $g(x^0) < 0$ for some $x^0 \in D$, while $f(x) \geq 0$ for all $x \in D$ satisfying $g(x) \geq 0$, then there exists a real number $\lambda \geq 0$ such that $f(x) + \lambda g(x) \geq 0 \; \forall x \in D$.*

Proof Apply the above Proposition for $f_0 = f, f_1 = g$. □

A more general result about inconsistent systems of convex inequalities is the following:

Theorem 2.3 (Generalized Farkas–Minkowski Theorem) *Let f_i, $i \in I_1 \subset \{1, \ldots, m\}$, be affine functions on $\mathbb{R}^n$, and let f_0 and f_i, $i \in I_2 := \{1, \ldots, m\} \setminus I_1$, be convex functions finite on some convex set $D \subset \mathbb{R}^n$. If there exists x^0 satisfying*

$$x^0 \in \mathrm{ri}D,\ f_i(x^0) \le 0\ (i \in I_1),\ f_i(x^0) < 0\ (i \in I_2). \tag{2.19}$$

while the system

$$x \in D,\ f_i(x) \le 0\ (i = 1, \ldots, m),\ f_0(x) < 0, \tag{2.20}$$

is inconsistent, then there exist numbers $\lambda_i \ge 0, i = 1, \ldots, m$, such that

$$f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \ge 0 \quad \forall x \in D. \tag{2.21}$$

Proof By replacing I_1 with $\{i|\ f_i(x^0) = 0\}$ we can assume $f_i(x^0) = 0$ for every $i \in I_1$. Arguing by induction on m, observe first that the theorem is true for $m = 1$. Indeed, in this case, if $I_1 = \emptyset$ the theorem follows from Proposition 2.18, so we can assume that $I_1 = \{1\}$, i.e., $f_1(x)$ is affine. In view of the fact $f_1(x^0) = 0$ for $x^0 \in \mathrm{ri}D$, if $f_1(x) \ge 0\ \forall x \in D$, then $f_1(x) = 0\ \forall x \in D$ and (2.21) holds with $\lambda = 0$. On the other hand, if there exists $x \in D$ satisfying $f_1(x) < 0$ then, since the system $x \in D, f_1(x) \le 0, f_0(x) < 0$ is inconsistent, again the theorem follows from Proposition 2.18. Thus, in any event the theorem is true when $m = 1$. Assuming now that the theorem is true for $m = k - 1 \ge 1$, consider the case $m = k$. The hypotheses of the theorem imply that the system

$$x \in D,\ f_i(x) \le 0\ (i = 1, \ldots, k-1),\ \max\{f_k(x), f_0(x)\} < 0 \tag{2.22}$$

is inconsistent, and since the function $\max\{f_k(x), f_0(x)\}$ is convex, there exists, by the induction hypothesis, $t_i \ge 0, i = 1, \ldots, k-1$, such that

$$\max\{f_k(x), f_0(x)\} + \sum_{i=1}^{k-1} t_i f_i(x) \ge 0 \quad \forall x \in D. \tag{2.23}$$

We show that this implies the inconsistency of the system

$$x \in D,\ \sum_{i=1}^{k-1} t_i f_i(x) \le 0,\ f_k(x) \le 0,\ f_0(x) < 0. \tag{2.24}$$

Indeed, from (2.23) it follows that no solution of (2.24) exists with $f_k(x) < 0$. Furthermore, if there exists $x \in D$ satisfying (2.24) with $f_k(x) = 0$ then, setting $x' = \alpha x^0 + (1 - \alpha)x$ with $\alpha \in (0, 1)$, we would have $x' \in D$, $\sum_{i=1}^{k-1} t_i f_i(x') \le 0, f_k(x') < 0$ and $f_0(x') \le \alpha f_0(x^0) + (1 - \alpha) f_0(x) < 0$ for sufficiently small $\alpha > 0$.

Since this contradicts (2.23), the system (2.24) is inconsistent and, again by the induction hypothesis, there exist $\theta \geq 0$ and $\lambda_k \geq 0$ such that

$$f_0(x) + \theta \sum_{i=1}^{k-1} t_i f_i(x) + \lambda_k f_k(x) \geq 0 \quad \forall x \in D. \tag{2.25}$$

This is the desired conclusion with $\lambda_i = \theta t_i \geq 0$ for $i = 1, \ldots, k-1$. $\square$

Corollary 2.7 (Farkas' Lemma) *Let A be an $m \times n$ matrix, and let $p \in \mathbb{R}^n$. If $\langle p, x\rangle \geq 0$ for all $x \in \mathbb{R}^n$ satisfying $Ax \geq 0$ then $p = A^T\lambda$ for some $\lambda = (\lambda_1, \ldots, \lambda_m) \geq 0$.*

Proof Apply the above theorem for $D = \mathbb{R}^n$, $f_0(x) = \langle p, x\rangle$, and $f_i(x) = -\langle a^i, x\rangle$, $i = 1, \ldots, m$, where a^i is the i-th row of A. Then there exist nonnegative $\lambda_1, \ldots, \lambda_m$ such that $\langle p, x\rangle - \sum_{i=1}^m \lambda_i \langle a^i, x\rangle \geq 0$ for all $x \in \mathbb{R}^n$, hence $\langle p, x\rangle - \sum_{i=1}^m \lambda_i \langle a^i, x\rangle = 0$ for all $x \in \mathbb{R}^n$, i.e., $p = \sum_{i=1}^m \lambda_i a^i$. $\square$

Theorem 2.4 *Let $f_1, \ldots, f_m$ be convex functions finite on some convex set D in $\mathbb{R}^n$, and let A be a $k \times n$ matrix, $b \in \mathrm{ri}A(D)$. If the system*

$$x \in D,\ Ax = b,\ f_i(x) < 0,\ i = 1, \ldots, m \tag{2.26}$$

is inconsistent, then there exist a vector $t \in \mathbb{R}^m$ and nonnegative numbers $\lambda_1, \ldots, \lambda_m$ summing up to 1 such that

$$\langle t, Ax - b\rangle + \sum_{i=1}^m \lambda_i f_i(x) \geq 0 \quad \forall x \in D. \tag{2.27}$$

Proof Define $E = \{x \in D \mid Ax = b\}$. Since the system

$$x \in E,\ f_i(x) < 0,\ i = 1, \ldots, m$$

is inconsistent, there exist, by Proposition 2.18, nonnegative numbers $\lambda_1, \ldots, \lambda_m$, not all zero, such that

$$\sum_{i=1}^m \lambda_i f_i(x) \geq 0 \quad \forall x \in E. \tag{2.28}$$

By dividing $\sum_{i=1}^m \lambda_i$, we may assume $\sum_{i=1}^m \lambda_i = 1$. Obviously, the convex function $f(x) := \sum_{i=1}^m \lambda_i f_i(x)$ is finite on D. Consider the set C of all $(y, y_0) \in \mathbb{R}^k \times \mathbb{R}$ for which there exists an $x \in D$ satisfying

$$Ax - b = y,\ f(x) < y_0.$$

Since by (2.28) $0 \notin C$, and since C is convex there exists, again by Lemma 1.2, a vector $(t, t_0) \in \mathbb{R}^k \times R$ such that

$$\inf_{(y,y_0)\in C}[\langle t, y\rangle + t_0 y_0] \geq 0, \qquad \sup_{(y,y_0)\in C}[\langle t, y\rangle + t_0 y_0] > 0. \tag{2.29}$$

If $t_0 < 0$ then by fixing $x \in D, y = Ax - b$ and letting $y_0 \to +\infty$ we would have $\langle t, y\rangle + t_0 y_0 \to -\infty$, contradicting (2.29). Consequently $t_0 \geq 0$. We now contend that $t_0 > 0$. Suppose the contrary, that $t_0 = 0$, so that $\langle t, y\rangle \geq 0\ \forall (y, y_0) \in C$, and hence

$$\langle t, y\rangle \geq 0 \quad \forall y \in A(D) - b.$$

Since by hypothesis $b \in \mathrm{ri}A(D)$, i.e., $0 \in \mathrm{ri}(A(D) - b)$, this implies $\langle t, y\rangle = 0\ \forall y \in A(D)$, hence $\langle t, y\rangle + t_0 y_0 = 0$ for all $(y, y_0) \in C$, contradicting (2.29). Therefore, $t_0 > 0$ and we can take $t_0 = 1$. Then the left inequality (2.29), where $y = Ax - b$, $y_0 = f(x) + \varepsilon$ for $x \in D$, yields the desired relation (2.27) by making $\varepsilon \downarrow 0$. □

2.6 Approximation by Affine Functions

A general method of nonlinear analysis is linearization, i.e., the approximation of convex functions by affine functions. The basis for this approximation is provided by the next result on the structure of closed convex functions which is merely the analytical equivalent form of the corresponding theorem on the structure of closed convex sets (Theorem 1.5).

Theorem 2.5 *A proper closed convex function f on $\mathbb{R}^n$ is the upper envelope (pointwise supremum) of the family of all affine functions h on $\mathbb{R}^n$ minorizing f.*

Proof We first show that for any $(x^0, t^0) \notin \mathrm{epi}f$ there exists $(a, \alpha) \in \mathbb{R}^n \times \mathbb{R}$ such that

$$\langle a, x\rangle - t < \alpha < \langle a, x^0\rangle - t^0 \quad \forall (x, t) \in \mathrm{epi}f. \tag{2.30}$$

Indeed, since $\mathrm{epi}f$ is a closed convex set there exists by Theorem 1.3 a hyperplane strongly separating (x^0, t^0) from $\mathrm{epi}f$, i.e., an affine function $\langle a, x\rangle + \gamma t$ such that

$$\langle a, x\rangle + \gamma t < \alpha < \langle a, x^0\rangle + \gamma t^0 \quad \forall (x, t) \in \mathrm{epi}f.$$

It is easily seen that $\gamma \leq 0$ because if $\gamma > 0$ then by taking a point $\overline{x} \in \mathrm{dom}f$ and an arbitrary $t \geq f(\overline{x})$, we would have $(\overline{x}, t) \in \mathrm{epi}f$, hence $\langle a, \overline{x}\rangle + \gamma t < \alpha$ for all $t \geq f(\overline{x})$, which would lead to a contradiction as $t \to +\infty$. Furthermore, if $x^0 \in \mathrm{dom}f$ then $\gamma = 0$ would imply $\langle a, x^0\rangle < \langle a, x^0\rangle$, which is absurd. Hence, in this case, $\gamma < 0$, and by dividing a and α by $-\gamma$, we can assume $\gamma = -1$, so that (2.30) holds. On the other hand, if $x^0 \notin \mathrm{dom} f$ and $\gamma = 0$, we can consider an $x^1 \in \mathrm{dom} f$ and $t^1 < f(x^1)$, so that $(x^1, t^1) \notin \mathrm{epi} f$ and by what has just been proved, there exists $(b, \beta) \in \mathbb{R}^n \times \mathbb{R}$ satisfying

$$\langle b, x\rangle - t < \beta < \langle b, x^1\rangle - t^1 \quad \forall (x,t) \in \mathrm{epi} f.$$

For any $\theta > 0$ we then have for all $(x,t) \in \mathrm{epi} f$:

$$\langle b + \theta a, x\rangle - t = (\langle b, x\rangle - t) + \theta \langle a, x\rangle < \beta + \theta\alpha,$$

while $\langle b + \theta a, x^0\rangle - t^0 = (\langle b, x^0\rangle - t^0) + \theta\langle a, x^0\rangle > \beta + \theta\alpha$ for sufficiently large θ because $\alpha < \langle a, x^0\rangle$. Thus, for $\theta > 0$ large enough, setting $a' = b + \theta a$, $\alpha' = \beta + \theta\alpha$, we have

$$\langle a', x\rangle - t < \alpha' < \langle a', x^0\rangle - t^0 \quad \forall (x,t) \in \mathrm{epi} f,$$

i.e., (a', α') satisfies (2.30). Note that (2.30) implies $\langle a, x\rangle - \alpha \le f(x)\ \forall x$, i.e., the affine function $h(x) = \langle a, x\rangle - \alpha$ minorizes $f(x)$. Now let $\mathbf{Q}$ be the family of all affine functions h minorizing f. We contend that

$$f(x) = \sup\{h(x) |\ h \in \mathbf{Q}\}. \tag{2.31}$$

Suppose the contrary, that $f(x^0) > \mu = \sup\{h(x) |\ h \in \mathbf{Q}\}$ for some x^0. Then $(x^0, \mu) \notin \mathrm{epi} f$ and by the above there exists $(a, \alpha) \in \mathbb{R}^n \times R$ satisfying (2.30) for $t^0 = \mu$. Hence, $h(x) = \langle a, x\rangle - \alpha \in \mathbf{Q}$ and $\alpha < \langle a, x^0\rangle - \mu$, i.e., $h(x^0) = \langle a, x^0\rangle - \alpha > \mu$, a contradiction. Thus (2.31) holds, as was to be proved. □

Corollary 2.8 *For any function $f : \mathbb{R}^n \to [-\infty, +\infty]$ the closure of the convex hull of f is equal to the upper envelope of all affine functions minorizing f.*

Proof An affine function h minorizes f if and only if it minorizes $\mathrm{cl}(\mathrm{conv} f)$, hence the conclusion. □

Proposition 2.19 *Any proper convex function f has an affine minorant. If $x^0 \in$* $\mathrm{int}(\mathrm{dom} f)$ *then an affine minorant h exists which is exact at x^0, i.e., such that $h(x^0) = f(x^0)$.*

Proof Indeed, the collection $\mathbf{Q}$ in Theorem 2.5 for $\mathrm{cl}(f)$ is nonempty. If $x^0 \in \mathrm{int}(\mathrm{dom} f)$ then $(x^0, f(x^0))$ is a boundary point of the convex set $\mathrm{epi} f$. Hence by Theorem 1.5 there exists a supporting hyperplane to $\mathrm{epi} f$ at this point, i.e., there exists $(a, \alpha) \in \mathbb{R}^n \times \mathbb{R}$ such that either $a \neq 0$ or $\alpha \neq 0$ and

$$\langle a, x\rangle - \alpha t \le \langle a, x^0\rangle - \alpha f(x^0) \quad \forall (x,t) \in \mathrm{epi} f.$$

As in the proof of Theorem 2.5, it is readily seen that $\alpha \le 0$. Furthermore, if $\alpha = 0$ then the above relation implies that $\langle a, x\rangle \le \langle a, x^0\rangle$ for all x in some open neighborhood U of x^0 contained in $\mathrm{dom} f$, and hence that $a = 0$, a contradiction.

Therefore, $\alpha < 0$, so we can take $\alpha = -1$. Then $\langle a, x\rangle - t \leq \langle a, x^0\rangle - f(x^0)$ for all $(x,t) \in \text{epi}f$ and the affine function

$$h(x) = \langle a, x - x^0\rangle + f(x^0)$$

satisfies $h(x) \leq f(x)\ \forall x$ and $h(x^0) = f(x^0)$. □

2.7 Subdifferential

Given a proper function f on $\mathbb{R}^n$, a vector $p \in \mathbb{R}^n$ is called a *subgradient* of f at a point x^0 if

$$\langle p, x - x^0\rangle + f(x^0) \leq f(x) \quad \forall x. \tag{2.32}$$

The set of all subgradients of f at x^0 is called the *subdifferential* of f at x^0 and is denoted by $\partial f(x^0)$ (Fig. 2.2). The function f is said to be *subdifferentiable* at x^0 if $\partial f(x^0) \neq \emptyset$.

Theorem 2.6 *Let f be a proper convex function on $\mathbb{R}^n$. For any bounded set $C \subset \text{int}(\text{dom}f)$ the set $\bigcup_{x\in C} \partial f(x)$ is nonempty and bounded. In particular, $\partial f(x^0)$ is nonempty and bounded at every $x^0 \in \text{int}(\text{dom}f)$.*

Proof By Proposition 2.19 if $x^0 \in \text{int}(\text{dom}f)$ then f has an affine minorant $h(x)$ such that $h(x^0) = f(x^0)$, i.e., $h(x) = \langle p, x - x^0\rangle + f(x^0)$ for some $p \in \partial f(x^0)$. Thus, $\partial f(x^0) \neq \emptyset$ for every $x^0 \in \text{int}(\text{dom}f)$. Consider now any bounded set $C \subset \text{int}(\text{dom}f)$. As we saw in the proof of Theorem 2.2, there is $r > 0$ such that $C + rB \subset \text{int}(\text{dom}f)$, where B denotes the Euclidean unit ball. By definition, for any

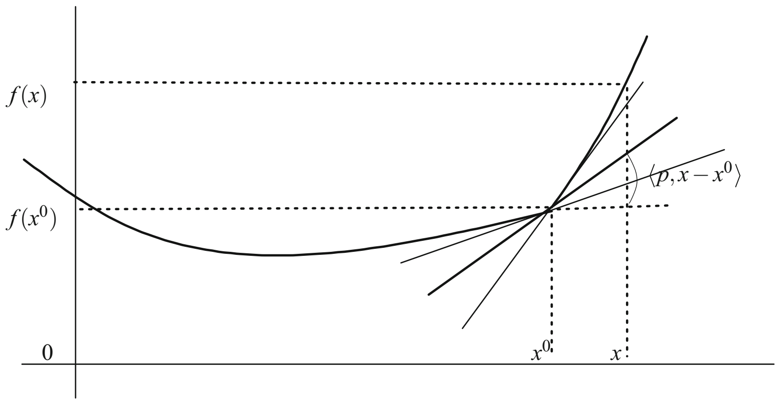


Fig. 2.2 The set $\partial f(x^0)$

$x \in C$ and $p \in \partial f(x)$, we have $\langle p, y-x\rangle + f(x) \le f(y)\ \forall y$, but by Theorem 2.2, there exists $\gamma > 0$ such that $|f(x)-f(y)| \le \gamma\|y-x\|$ for all $y \in C + rB$. Hence $|\langle p, y-x\rangle| \le \gamma\|y-x\|$ for all $y \in C+rB$, i.e., $|\langle p, u\rangle| \le \gamma\|u\|$ for all $u \in B$. By taking $u = p/\|p\|$ this implies $\|p\| \le \gamma$, so the set $\bigcup_{x\in C}\partial f(x)$ is bounded. □

Corollary 2.9 *Let f be a proper convex function on $\mathbb{R}^n$. For any bounded convex subset C of* $\mathrm{int}(\mathrm{dom} f)$ *there exists a positive constant γ such that*

$$f(x) = \sup\{h(x) |\ h \in \mathbf{Q}_0\} \quad \forall x \in C, \tag{2.33}$$

where every $h \in \mathbf{Q}_0$ has the form $h(x) = \langle a, x\rangle - \alpha$ with $\|a\| \le \gamma$.

Proof It suffices to take as $\mathbf{Q}_0$ the family of all affine functions $h(x) = \langle a, x-y\rangle + f(y)$, with $y \in C,\ a \in \partial f(y)$. □

Corollary 2.10 *Let $f : D \to \mathbb{R}$ be a convex function defined and continuous on a convex set D with nonempty interior. If the set $\cup\{\partial f(x) |\ x \in \mathrm{int}D\}$ is bounded, then f can be extended to a finite convex function on $\mathbb{R}^n$.*

Proof For each point $y \in \mathrm{int}D$ take a vector $p_y \in \partial f(y)$ and consider the affine function $h_y(x) = f(y) + \langle p_y, x-y\rangle$. The function $\tilde{f}(x) = \sup\{h_y(x) |\ y \in \mathrm{int}D\}$ is convex on $\mathbb{R}^n$ as the upper envelope of a family of affine functions. If a is any fixed point of D then $h_y(x) = f(y) + \langle p_y, a-y\rangle + \langle p_y, x-a\rangle \le f(a) + \langle p_y, x-a\rangle \le f(a) + \|p_y\|.\|x-a\|$. Since $\|p_y\|$ is bounded on $\mathrm{int}D$ the latter inequality shows that $-\infty < \tilde{f}(x) < +\infty\ \forall x \in \mathbb{R}^n$. Thus, $\tilde{f}(x)$ is a convex finite function on $\mathbb{R}^n$. Finally, since obviously $\tilde{f}(x) = f(x)$ for every $x \in \mathrm{int}D$ it follows from the continuity of both $f(x)$ and $\tilde{f}(x)$ on D that $\tilde{f}(x) = f(x)\ \forall x \in D$. □

Example 2.1 (*Positively Homogeneous Convex Function*) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a positively homogeneous convex function, i.e., a convex function $f : \mathbb{R}^n \to \mathbb{R}$ such that $f(\lambda x) = \lambda f(x)\ \forall \lambda > 0$. Then

$$\partial f(x^0) = \{p \in \mathbb{R}^n \mid \langle p, x^0\rangle = f(x^0),\ \langle p, x\rangle \le f(x)\ \forall x\} \tag{2.34}$$

Proof if $p \in \partial f(x^0)$ then $\langle p, x-x^0\rangle + f(x^0) \le f(x)\ \forall x$. Setting $x = 2x^0$ yields $\langle p, x^0\rangle + f(x^0) \le 2f(x^0)$, i.e., $\langle p, x^0\rangle \le f(x^0)$, then setting $x = 0$ yields $-\langle p, x^0\rangle \le -f(x^0)$, hence $\langle p, x^0\rangle = f(x^0)$. (Note that this condition is trivial and can be omitted if $x^0 = 0$). Furthermore, $\langle p, x-x^0\rangle = \langle p, x\rangle - \langle p, x^0\rangle = \langle p, x\rangle - f(x^0)$, hence $\langle p, x\rangle \le f(x)\ \forall x$. Conversely, if p belongs to the set on the right-hand side of (2.34) then obviously $\langle p, x-x^0\rangle \le f(x) - f(x^0)$, so $p \in \partial f(x^0)$. □

If, in addition, $f(-x) = f(x) \ge 0\ \forall x$ then the condition $\langle p, x\rangle \le f(x)\ \forall x$ is equivalent to $|\langle p, x\rangle| \le f(x)\ \forall x$. In particular:

1. If $f(x) = \|x\|$ (Euclidean norm) then

$$\partial f(x^0) = \begin{cases} \{p \mid \ \|p\| \le 1\} \text{ (unit ball)} & \text{if } x^0 = 0 \\ \{x^0/\|x^0\|\} & \text{if } x^0 \ne 0. \end{cases} \tag{2.35}$$

2. If $f(x) = \max\{|x_i| \mid i = 1, \ldots, n\}$ (Tchebycheff norm) then

$$\partial f(x^0) = \begin{cases} \text{conv}\{\pm e_1, \ldots, \pm e_n\} & \text{if } x^0 = 0 \\ \text{conv}\{(\text{sign} x_i^0)x_i^0 \mid i \in I_{x^0}\} & \text{if } x^0 \neq 0, \end{cases} \tag{2.36}$$

where $I_x = \{i \mid |x_i| = f(x)\}$.

3. If Q is a symmetric positive semidefinite matrix and $f(x) = \sqrt{\langle x, Qx\rangle}$ (elliptic norm) then

$$\partial f(x^0) = \begin{cases} \{p \mid \langle p, x\rangle \leq \sqrt{\langle x, Qx\rangle}\ \forall x\} & \text{if } x^0 \in \text{Ker} Q \\ \{(Qx^0)/\sqrt{\langle x^0, Qx^0\rangle}\} & \text{if } x^0 \notin \text{Ker} Q. \end{cases} \tag{2.37}$$

Example 2.2 (*Distance Function*) Let C be a closed convex set in $\mathbb{R}^n$, and $f(x) = \min\{\|y - x\| \mid y \in C\}$. Denote by $\pi_C(x)$ the projection of x on C, so that $\|\pi_C(x) - x\| = \min\{\|y - x\| \mid y \in C\}$ and $\langle x - \pi_C(x), y - \pi_C(x)\rangle \leq 0\ \forall y \in C$ (see Proposition 1.15). Then

$$\partial f(x^0) = \begin{cases} N_C(x^0) \cap B(0, 1) & \text{if } x^0 \in C \\ \left\{\frac{x^0 - \pi_C(x^0)}{\|x^0 - \pi_C(x^0)\|}\right\} & \text{if } x^0 \notin C, \end{cases} \tag{2.38}$$

where $N_C(x^0)$ denotes the outward normal cone of C at x^0 and $B(0, 1)$ the Euclidean unit ball.

Proof Let $x^0 \in C$, so that $f(x^0) = 0$. Then $p \in \partial f(x^0)$ implies $\langle p, x - x^0\rangle \leq f(x)\ \forall x$, hence, in particular, $\langle p, x - x^0\rangle \leq 0\ \forall x \in C$, i.e., $p \in N_C(x^0)$; furthermore, $\langle p, x - x^0\rangle \leq f(x) \leq \|x - x^0\|\ \forall x$, hence $\|p\| \leq 1$, i.e., $p \in B(0, 1)$. Conversely, if $p \in N_C(x^0) \cap B(0, 1)$ then $\langle p, x - \pi_C(x)\rangle \leq \|x - \pi_C(x)\| \leq f(x)$, and $\langle p, \pi_C(x) - x^0\rangle \leq 0$, consequently $\langle p, x - x^0\rangle = \langle p, x - \pi_C(x)\rangle + \langle p, \pi_C(x) - x^0\rangle \leq f(x) = f(x) - f(x^0)$ for all x, and so $p \in \partial f(x^0)$.

Turning to the case $x^0 \notin C$, observe that $p \in \partial f(x^0)$ implies $\langle p, x - x^0\rangle + f(x^0) \leq f(x)\ \forall x$, hence, setting $x = \pi_C(x^0)$ yields $\langle p, \pi_C(x^0) - x^0\rangle + \|\pi_C(x^0) - x^0\| \leq 0$, i.e., $\langle p, x^0 - \pi_C(x^0)\rangle \geq \|x^0 - \pi_C(x^0)\|$. On the other hand, setting $x = 2x^0 - \pi_C(x^0)$ yields $\langle p, x^0 - \pi_C(x^0)\rangle + \|\pi_C(x^0) - x^0\| \leq 2\|\pi_C(x^0) - x^0\|$, i.e., $\langle p, x^0 - \pi_C(x^0)\rangle \leq \|\pi_C(x^0) - x^0\|$. Thus, $\langle p, x^0 - \pi_C(x^0)\rangle = \|\pi_C(x^0) - x^0\|$ and consequently $p = \frac{x^0 - \pi_C(x^0)}{\|x^0 - \pi_C(x^0)\|}$. Conversely, the last equality implies $\langle p, x^0 - \pi_C(x^0)\rangle = \|x^0 - \pi_C(x^0)\| = f(x^0)$, $\langle p, x - \pi_C(x)\rangle \leq \|x - \pi_C(x)\| = f(x)$, hence $\langle p, x - x^0\rangle + f(x^0) = \langle p, x - x^0\rangle + \langle p, x^0 - \pi_C(x^0)\rangle = \langle p, x - \pi_C(x^0)\rangle = \langle p, x - \pi_C(x)\rangle + \langle p, \pi_C(x) - \pi_C(x^0)\rangle, \leq \|x - \pi_C(x)\| = f(x)$ for all x (note that $\langle p, \pi_C(x) - \pi_C(x^0)\rangle \leq 0$ because $p \in N_C(\pi_C(x^0))$). Therefore, $\langle p, x - x^0\rangle + f(x^0) \leq f(x)$ for all x, proving that $p \in \partial f(x^0)$. □

Observe from the above examples that there is a unique subgradient (which is just the gradient) at every point where f is differentiable. This is actually a general fact which we are now going to establish.

Let $f : \mathbb{R}^n \to [-\infty, +\infty]$ be any function and let x^0 be a point where f is finite. If for some $u \neq 0$ the limit (finite or infinite)

$$\lim_{\lambda \downarrow 0} \frac{f(x^0 + \lambda u) - f(x^0)}{\lambda}$$

exists, then it is called the *directional derivative* of f at x^0 in the direction u, and is denoted by $f'(x^0; u)$.

Proposition 2.20 *Let f be a proper convex function and $x^0 \in \mathrm{dom} f$. Then:*

(i) *$f'(x^0; u)$ exists for every direction u and satisfies*

$$f'(x^0; u) = \inf_{\lambda > 0} \frac{f(x^0 + \lambda u) - f(x^0)}{\lambda}; \tag{2.39}$$

(ii) *The function $u \mapsto f'(x^0; u)$ is convex and homogeneous and $p \in \partial f(x^0)$ if and only if*

$$\langle p, u \rangle \leq f'(x^0, u) \quad \forall u. \tag{2.40}$$

(iii) *If f is continuous at x^0 then $f'(x^0; u)$ is finite and continuous at every $u \in \mathbb{R}^n$, the subdifferential $\partial f(x^0)$ is compact and*

$$f'(x^0; u) = \max\{\langle p, u \rangle \mid p \in \partial f(x^0)\}. \tag{2.41}$$

Proof

(i) For any given $u \neq 0$ the function $\varphi(\lambda) = f(x^0 + \lambda u)$ is proper convex on the real line, and $0 \in \mathrm{dom}\varphi$. Therefore, its right derivative $\varphi'_+(0) = f'(x^0; u)$ exists (but may equal $+\infty$ if 0 is an endpoint of $\mathrm{dom}\varphi$). The relation (2.39) follows from the fact that $[\varphi(\lambda) - \varphi(0)]/\lambda$ is nonincreasing as $\lambda \downarrow 0$.

(ii) The homogeneity of $f'(x^0; u)$ is obvious. The convexity then follows from the relations

$$\begin{aligned} f'(x^0; u + v) &= \inf_{\lambda > 0} \frac{f(x^0 + \frac{\lambda}{2}(u + v)) - f(x^0)}{\frac{\lambda}{2}} \\ &\leq \inf_{\lambda > 0} \frac{f(x^0 + \lambda u) - f(x^0) + f(x^0 + \lambda v) - f(x^0)}{\lambda} \\ &= f'(x^0; u) + f'(x^0; v). \end{aligned}$$

Setting $x = x^0 + \lambda u$ we can turn the subgradient inequality (2.32) into the condition

$$\langle p, u \rangle \leq [f(x^0 + \lambda u) - f(x^0)]/\lambda \quad \forall u, \ \forall \lambda > 0,$$

which is equivalent to $\langle p, u \rangle \leq \inf_{\lambda > 0}[f(x^0 + \lambda u) - f(x^0)]/\lambda, \ \forall u$, i.e., by (i), $\langle p, u \rangle \leq f'(x^0; u) \ \forall u$.

(iii) If f is continuous at x^0 then there is a neighborhood U of 0 such that $f(x^0+u)$ is bounded above on U. Since by (i) $f'(x^0;u) \le f(x^0+u) - f(x^0)$, it follows that $f'(x^0;u)$ is also bounded above on U, and hence is finite and continuous on $\mathbb{R}^n$ (Theorem 2.2). The Condition (2.40) then implies that $\partial f(x^0)$ is closed and hence compact because it is bounded by Theorem 2.6. In view of the homogeneity of $f'(x^0;u)$, an affine minorant of it which is exact at some point must be of the form $\langle p, u\rangle$, with $\langle p,u\rangle \le f'(x^0;u)\ \forall u$, i.e., by (ii), $p \in \partial f(x^0)$. By Corollary 2.9, we then have $f'(x^0;u) = \max\{\langle p,u\rangle \mid p \in \partial f(x^0)\}$. □

According to the usual definition, a function f is *differentiable* at a point x^0 if there exists a vector $\nabla f(x^0)$ (the *gradient* of f at x^0) such that

$$f(x^0+u) = f(x^0) + \langle \nabla f(x^0), u\rangle + o(\|u\|).$$

This is equivalent to

$$\lim_{\lambda \downarrow 0} \frac{f(x^0+\lambda u) - f(x^0)}{\lambda} = \langle \nabla f(x^0), u\rangle, \quad \forall u \neq 0,$$

so the directional derivative $f'(x^0;u)$ exists, and is a linear function of u.

Proposition 2.21 *Let f be a proper convex function and $x^0 \in \mathrm{dom} f$. If f is differentiable at x^0 then $\nabla f(x^0)$ is its unique subgradient at x^0.*

Proof If f is differentiable at x^0 then $f'(x^0;u) = \langle \nabla f(x^0), u\rangle$, so by (ii) of Proposition 2.20, a vector p is a subgradient f at x^0 if and only if $\langle p,u\rangle \le \langle \nabla f(x^0), u\rangle\ \forall u$, i.e., if and only if $p = \nabla f(x^0)$. □

One can prove conversely that if f has a unique subgradient at x^0 then f is differentiable at x^0 (see, e.g., Rockafellar 1970).

2.8 Subdifferential Calculus

A convex function f may result from some operations on convex functions $f_i, i \in I$. (cf Sect. 2.2). It is important to know how the subdifferential of f can be computed in terms of the subdifferentials of the f_i' s.

Proposition 2.22 *Let f_i, $i = 1,\dots,m$, be proper convex functions on $\mathbb{R}^n$. Then for every $x \in \mathbb{R}^n$:*

$$\partial\left(\sum_{i=1}^m f_i(x)\right) \supset \sum_{i=1}^m \partial f_i(x).$$

If there exists a point $a \in \cap_{i=1}^m \mathrm{dom} f_i$, *where every function* f_i, *except perhaps one, is continuous, then the above inclusion is in fact an equality for every* $x \in \mathbb{R}^n$.

Proof It suffices to prove the proposition for $m = 2$ because the general case will follow by induction. Furthermore, the first part is straightforward, so we only need to prove the second part. If $p \in \partial(f_1 + f_2)(x^0)$, then the system

$$x - y = 0,\ f_1(x) + f_2(y) - f_1(x^0) - f_2(x^0) - \langle p, x - x^0\rangle < 0$$

is inconsistent. Define $D = \mathrm{dom} f_1 \times \mathrm{dom} f_2$ and $A(x, y) := x - y$. By hypothesis, f_1 is continuous at $a \in \mathrm{dom} f_1 \cap \mathrm{dom} f_2$, so there is a ball U around 0 such that $a + U \subset \mathrm{dom} f_1$, hence $U = (a + U) - a \subset \mathrm{dom} f_1 - \mathrm{dom} f_2 = A(D)$, i.e., $0 \in \mathrm{int} A(D)$. Therefore, by Theorem 2.4 there exists $t \in \mathbb{R}^n$ such that

$$\langle t, x - y\rangle + [f_1(x) + f_2(y) - f_1(x^0) - f_2(x^0) - \langle p, x - x^0\rangle] \geq 0$$

for all $x \in \mathbb{R}^n$ and all $y \in \mathbb{R}^n$. Setting $y = x^0$ yields $\langle p - t, x - x^0\rangle \leq f_1(x) - f_1(x^0)$ $\forall x \in \mathbb{R}^n$, i.e., $p - t \in \partial f_1(x^0)$. Then setting $x = x^0$ yields $\langle t, y - x^0\rangle \leq f_2(y) - f_2(x^0)$ $\forall y \in \mathbb{R}^n$, i.e., $t \in \partial f_2(x^0)$. Thus, $p = (p - t) + t \in \partial f_1(x^0) + \partial f_2(x^0)$, as was to be proved. □

Proposition 2.23 *Let* $A : \mathbb{R}^n \to \mathbb{R}^m$ *be a linear mapping and* g *be a proper convex function on* $\mathbb{R}^m$. *Then for every* $x \in \mathbb{R}^n$:

$$A^T \partial g(Ax) \subset \partial(g \circ A)(x).$$

If g *is continuous at some point in* $Im(A)$ (*the range of* A) *then the above inclusion is in fact an equality for every* $x \in \mathbb{R}^n$.

Proof The first part is straightforward. To prove the second part, consider any $p \in \partial(g \circ A)(x^0)$. Then the system

$$Ax - y = 0,\ g(y) - g(Ax^0) - \langle p, x - x^0\rangle < 0,\ x \in \mathbb{R}^n,\ y \in \mathbb{R}^m$$

is inconsistent. Define $D = \mathbb{R}^n \times \mathrm{dom} g$, $B(x, y) = Ax - y$. Since there is a point $b \in \mathrm{Im} A \cap \mathrm{int}(\mathrm{dom} g)$, we have $b \in \mathrm{int} B(D)$, so by Theorem 2.4 there exists $t \in \mathbb{R}^m$ such that

$$\langle t, Ax - y\rangle + g(y) - g(Ax^0) - \langle p, x - x^0\rangle \geq 0$$

for all $x \in \mathbb{R}^n$ and all $y \in \mathbb{R}^m$. Setting $y = 0$ then yields $\langle A^T t - p, x\rangle - g(Ax^0) + \langle p, x^0\rangle \geq 0$ $\forall x \in \mathbb{R}^n$, hence $p = A^T t$, while setting $x = x^0$ yields $\langle t, y - Ax^0\rangle \leq g(y) - g(Ax^0)$, i.e., $t \in \partial g(Ax^0)$. Therefore, $p \in A^T \partial g(Ax^0)$. □

Proposition 2.24 *Let* $g(x) = (g_1(x), \ldots, g_m(x))$, *where each* g_i *is a convex functions from* $\mathbb{R}^n$ *to* $\mathbb{R}$, *let* $\varphi : \mathbb{R}^m \to \mathbb{R}$ *be a convex function, component-wise* *increasing, i.e., such that* $\varphi(t) \geq \varphi(t')$ *whenever* $t_i \geq t'_i,\ i = 1, \ldots, m$. *Then the function* $f = \varphi \circ g$ *is convex and*

$$\partial f(x) = \left\{ \sum_{i=1}^{m} s_i p^i \mid p^i \in \partial g_i(x), (s_1, \ldots, s_m) \in \partial\varphi(g(x)) \right\}. \tag{2.42}$$

Proof The convexity of $f(x)$ follows from an obvious extension of Proposition 2.8 which corresponds to the case $m = 1$. To prove (2.42), let $p = \sum_{i=1}^m s_i p^i$ with $p^i \in \partial g_i(x^0), s \in \partial\varphi(g(x^0))$. First observe that $\langle s, y - g(x^0)\rangle \leq \varphi(y) - \varphi(g(x^0))\ \forall y \in \mathbb{R}^m$ implies, for all $y = g(x^0) + u$ with $u \leq 0:\ \langle s, u\rangle \leq \varphi(g(x^0)+u) - \varphi(g(x^0)) \leq 0\ \forall u \leq 0$, hence $s \geq 0$. Now $\langle p, x - x^0\rangle = \sum_{i=1}^m s_i \langle p^i, x - x^0\rangle \leq \sum_{i=1}^m s_i[g_i(x) - g_i(x^0)] = \langle s, g(x) - g(x^0)\rangle \leq \varphi(g(x)) - \varphi(g(x^0)) = f(x) - f(x^0)$ for all $x \in \mathbb{R}^n$. Therefore $p \in \partial f(x^0)$, i.e., the right-hand side of (2.42) is contained in the left-hand side. To prove the converse, let $p \in \partial f(x^0)$, so that the system

$$x \in \mathbb{R}^n,\ y \in \mathbb{R}^m,\ g_i(x) < y_i \quad i = 1, \ldots, m \tag{2.43}$$

$$\varphi(y) - \varphi(g(x^0)) - \langle p, x - x^0\rangle < 0 \tag{2.44}$$

is inconsistent, while the system (2.43) has a solution. By Proposition 2.18 there exists $s \in \mathbb{R}^m_+$ such that

$$\varphi(y) - \varphi(g(x^0)) - \langle p, x - x^0\rangle + \langle s, g(x) - y\rangle \geq 0$$

for all $x \in \mathbb{R}^n, y \in \mathbb{R}^m$. Setting $x = x^0$ yields $\varphi(y) - \varphi(g(x^0)) \geq \langle s, y - g(x^0)\rangle$ for all $y \in \mathbb{R}^m$, which means that $s \in \partial\varphi(g(x^0))$. On the other hand, setting $y = g(x^0)$ yields $\langle p, x - x^0\rangle \leq \sum_{i=1}^m s_i[g_i(x) - g_i(x^0)]$ for all $x \in \mathbb{R}^n$, which means that $p \in \partial(\sum_{i=1}^m s_i g_i(x^0))$, hence by Proposition 2.22, $p = \sum_{i=1}^m s_i p^i$ with $p^i \in \partial g_i(x^0)$. □

Note that when $\varphi(y)$ is differentiable at $g(x)$ the above formula (2.42) is similar to the classical chain rule, namely:

$$\partial(\varphi \circ g)(x) = \sum_{i=1}^{m} \frac{\partial\varphi}{\partial y_i}(g(x))\partial g_i(x).$$

Proposition 2.25 *Let* $f(x) = \max\{g_1(x), \ldots, g_m(x)\}$, *where each* g_i *is a convex function from* $\mathbb{R}^n$ *to* $\mathbb{R}$. *Then*

$$\partial f(x) = \mathrm{conv}\{\cup \partial g_i(x) |\ i \in I(x)\}, \tag{2.45}$$

where $I(x) = \{i \mid f(x) = g_i(x)\}$.

Proof If $p \in \partial f(x^0)$ then the system

$$g_i(x) - f(x^0) - \langle p, x - x^0\rangle < 0 \quad i = 1, \ldots, m$$

is inconsistent. By Proposition 2.18, there exist $\lambda_i \geq 0$ such that $\sum_{i=1}^m \lambda_i = 1$ and $\sum_{i=1}^m \lambda_i[g_i(x) - f(x^0) - \langle p, x - x^0\rangle] \geq 0$. Setting $x = x^0$, we have

$$\sum_{i \notin I(x^0)} \lambda_i[g_i(x^0) - f(x^0)] \geq 0,$$

with $g_i(x^0) - f(x^0) < 0$ for every $i \notin I(x^0)$. This implies that $\lambda_i = 0$ for all $i \notin I(x^0)$. Hence

$$\sum_{i \in I(x^0)} \lambda_i[g_i(x) - g_i(x^0) - \langle p, x - x^0\rangle] \geq 0$$

for all $x \in \mathbb{R}^n$ and so $p \in \partial(\sum_{i \in I(x^0)} \lambda_i g_i(x^0))$. By Proposition 2.22, $p = \sum_{i \in I(x^0)} p^i$, with $p^i \in \partial g_i(x^0)$. Thus $\partial f(x^0) \subset \text{conv}\{\cup \partial g_i(x^0) | \; i \in I(x^0)\}$. The converse inclusion can be verified in a straightforward manner. □

2.9 Approximate Subdifferential

A proper convex function f on $\mathbb{R}^n$ may have an empty subdifferential at certain points. In practice, however, one often needs only a concept of approximate subdifferential .

Given a positive number $\varepsilon > 0$, a vector $p \in \mathbb{R}^n$ is called an *ε-subgradient* of f at point x^0 if

$$\langle p, x - x^0\rangle + f(x^0) \leq f(x) + \varepsilon \quad \forall x. \tag{2.46}$$

The set of all ε-subgradients of f at x^0 is called the *ε-subdifferential* of f at x^0, and is denoted by $\partial_\varepsilon f(x^0)$ (Fig. 2.3).

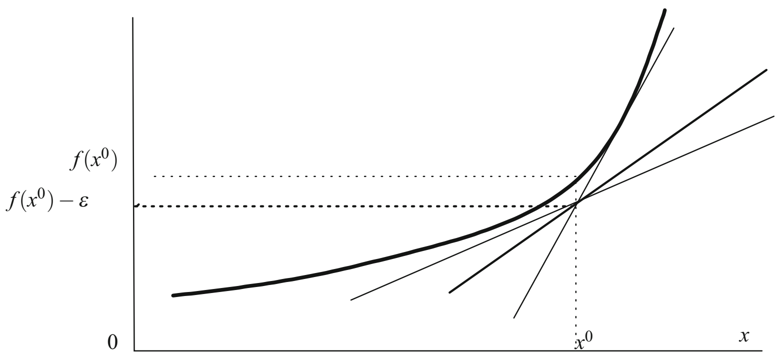


Fig. 2.3 The set $\partial_\varepsilon f(x^0)$

Proposition 2.26 *For any proper closed convex function f on $\mathbb{R}^n$, and any given $\varepsilon > 0$ the ε-subdifferential of f at any point $x^0 \in \mathrm{dom} f$ is nonempty. If C is a bounded subset of int(dom f) then the set $\cup_{x\in C}\partial_\varepsilon f(x)$ is bounded.*

Proof Since the point $(x^0, f(x^0) - \varepsilon) \notin \mathrm{epi} f$, there exists $p \in \mathbb{R}^n$ such that $\langle p, x - x^0\rangle < f(x) - (f(x^0) - \varepsilon) \quad \forall x$ (see (2.30) in the proof of Theorem 2.5). Thus, $\partial_\varepsilon f(x^0) \neq \emptyset$. The proof of the second part of the proposition is analogous to that of Theorem 2.6. □

Note that $\partial_\varepsilon f(x^0)$ is unbounded when x^0 is a boundary point of $\mathrm{dom} f$.

Proposition 2.27 *Let $x^0, x^1 \in \mathrm{dom} f$. If $p \in \partial_\varepsilon f(x^0)$ $(\varepsilon \geq 0)$ then $p \in \partial_\eta f(x^1)$ for $\eta = f(x^1) - f(x^0) - \langle p, x^1 - x^0\rangle + \varepsilon \geq 0$.*

Proof If $\langle p, x - x^0\rangle \leq f(x) - f(x^0) + \varepsilon$ for all x then $\langle p, x - x^1\rangle = \langle p, x - x^0\rangle + \langle p, x^0 - x^1\rangle \leq f(x) - f(x^0) + \varepsilon - \langle p, x^1 - x^0\rangle = f(x) - f(x^1) + \eta$ for all x. □

A function $f(x)$ is said to be *strongly convex* on a convex set C if there exists $r > 0$ such that

$$\begin{aligned} &f((1-\lambda)x^1 + \lambda x^2) \\ &\quad\leq (1-\lambda)f(x^1) + \lambda f(x^2) - (1-\lambda)\lambda r\|x^1 - x^2\|^2 \end{aligned} \tag{2.47}$$

for all $x^1, x^2 \in C$, and all $\lambda \in [0, 1]$. The number $r > 0$ is then called the *modulus of strong convexity* of $f(x)$. Using the identity

$$(1-\lambda)\lambda\|x^1 - x^2\|^2 = (1-\lambda)\|x^1\|^2 + \lambda\|x^2\|^2 - \|(1-\lambda)x^1 + \lambda x^2\|^2$$

for all $x^1, x^2 \in \mathbb{R}^n$, and all $\lambda \in [0, 1]$, it is easily verified that a convex function $f(x)$ is strongly convex with modulus of strong convexity r if and only if the function $f(x) - r\|x\|^2$ is convex.

Proposition 2.28 *If $f(x)$ is a strongly convex function on $\mathbb{R}^n$ with modulus of strong convexity r then for any $x^0 \in \mathbb{R}^n$ and $\varepsilon \geq 0$:*

$$\partial f(x^0) + B(0, 2\sqrt{r\varepsilon}) \subset \partial_\varepsilon f(x^0), \tag{2.48}$$

where $B(0, \alpha)$ denotes the ball of radius α around 0.

Proof Let $p \in \partial f(x^0)$. Since $F(x) := f(x) - r\|x\|^2$ is convex and $p - 2rx^0 \in \partial F(x^0)$ we can write

$$f(x) - f(x^0) - r(\|x\|^2 - \|x^0\|^2) \geq \langle p - 2rx^0, x - x^0\rangle$$

for all x, hence

$$f(x) - f(x^0) \geq \langle p, x - x^0\rangle + r\|x - x^0\|^2.$$

Let us determine a vector u such that

$$r\|x-x^0\|^2 \geq \langle u, x-x^0\rangle - \varepsilon \ \ \forall x \in \mathbb{R}^n. \tag{2.49}$$

The convex quadratic function $r\|x-x^0\|^2 - \langle u, x-x^0\rangle$ achieves its minimum at the point $\bar{x}$ such that $2r(\bar{x}-x^0) - u = 0$, i.e., $\bar{x}-x^0 = \frac{u}{2r}$. This minimum is equal to $r[\frac{u}{2r}]^2 - \frac{\|u\|^2}{2r} = \frac{-\|u\|^2}{4r}$. Thus, by choosing u such that $\|u\|^2 \leq 4r\varepsilon$, i.e., $u \in B(0, 2\sqrt{r\varepsilon})$ we will have (2.49), hence $p+u \in \partial_\varepsilon f(x^0)$. $\square$

Corollary 2.11 *Let $f(x) = \frac{1}{2}\langle x, Qx\rangle + \langle x, a\rangle$, where $a \in \mathbb{R}^n$,and Q is an $n \times n$ symmetric positive definite matrix. Let $r > 0$ be the smallest eigenvalue of Q. Then*

$$Qx + a + u \in \partial_\varepsilon f(x)$$

for any $u \in \mathbb{R}^n$ such that $\|u\| \leq 2\sqrt{r\varepsilon}$.

Proof Clearly $f(x) - r\|x\|^2$ is convex (as its smallest eigenvalue is nonnegative), so $f(x)$ is a strongly convex function to which the above proposition applies. $\square$

2.10 Conjugate Functions

Given an arbitrary function $f : \mathbb{R}^n \to [-\infty, +\infty]$, we consider the set of all affine functions h minorizing f. It is natural to restrict ourselves to proper functions, because an improper function either has no affine minorant (if $f(x) = -\infty$ for some x) or is minorized by every affine function (if $f(x)$ is identical to $+\infty$).

Observe that if $\langle p, x\rangle - \alpha \leq f(x)$ for all x then $\alpha \geq \langle p, x\rangle - f(x) \ \ \forall x$. The function

$$f^*(p) = \sup_{x\in\mathbb{R}^n}\{\langle p, x\rangle - f(x)\}. \tag{2.50}$$

which is clearly closed and convex is called the *conjugate* of f.

For instance, the conjugate of the function $f(x) = \delta_C(x)$ (indicator function of a set C, see Sect. 2.1) is the function $f^*(p) = \sup_{x\in C}\langle p, x\rangle = s_C(p)$ (support function of C). The conjugate of an affine function $f(x) = \langle c, x\rangle - \alpha$ is the function

$$f^*(p) = \sup_x\{\langle p, x\rangle - \langle c, x\rangle + \alpha\} = \begin{cases} +\infty, & p \neq c \\ \alpha, & p = c. \end{cases}$$

Two less trivial examples are the following:

Example 2.3 The conjugate of the proper convex function $f(x) = e^x, \ x \in \mathbb{R}$, is by definition $f^*(p) = \sup_x\{px - e^x\}$. Obviously, $f^*(p) = 0$ for $p = 0$ and $f^*(p) = +\infty$ for $p < 0$. For $p > 0$, the function $px - e^x$ achieves a maximum at $x = \xi$ satisfying $p = e^\xi$, so $f^*(p) = p\log p - p$. Thus,

$$f^*(p) = \begin{cases} 0, & p = 0 \\ +\infty, & p < 0 \\ p\log p - p, & p > 0. \end{cases}$$

The conjugate of $f^*(p)$ is in turn the function $f^{**}(x) = \sup_p\{px - f^*(p)\} = \sup\{px - p\log p + p \mid p > 0\} = e^x$.

Example 2.4 The conjugate of the function $f(x) = \frac{1}{\alpha}\sum_{i=1}^n |x_i|^\alpha,\ 1 < \alpha < +\infty$, is

$$f^*(p) = \frac{1}{\beta}\sum_{i=1}^n |p_i|^\beta,\ 1 < \beta < +\infty,$$

where $1/\alpha + 1/\beta = 1$. Indeed,

$$f^*(p) = \sup_x \left\{\sum_{i=1}^n p_i x_i - \frac{1}{\alpha}\sum_{i=1}^n |x_i|^\alpha\right\}.$$

By differentiation, we find that the supremum on the right-hand side is achieved at $x = \xi$ satisfying $p_i = |\xi_i|^{\alpha-1}\mathrm{sign}\xi_i$, hence

$$f^*(p) = \sum_{i=1}^n |\xi_i|^\alpha \left(1 - \frac{1}{\alpha}\right) = \frac{1}{\beta}\sum_{i=1}^n |p_i|^\beta.$$

Proposition 2.29 *Let $f : \mathbb{R}^n \to [-\infty, +\infty]$ be an arbitrary proper function. Then:*

(i) $f(x) + f^*(p) \ge \langle p, x\rangle \quad \forall x \in \mathbb{R}^n,\ \forall p \in \mathbb{R}^n$;
(ii) $f^{**}(x) \le f(x)\ \forall x$, *and* $f^{**} = f$ *if and only if* f *is convex and closed;*
(iii) $f^{**}(x) = \sup\{h(x) \mid h$ affine, $h \le f\}$, *i.e.,* $f^{**}(x)$ *is the largest closed convex function minorizing* $f(x)$: $f^{**} = \mathrm{cl(conv)}f$.

Proof (i) is obvious, and from (i) $f^{**}(x) = \sup_p\{\langle p, x\rangle - f^*(p)\} \le f(x)$. Let Q be the set of all affine functions majorized by f. For every $h \in Q$, say $h(x) = \langle p, x\rangle - \alpha$, we have $\langle p, x\rangle - \alpha \le f(x)\ \forall x$, hence $\alpha \ge \sup_x\{\langle p, x\rangle - f(x)\} = f^*(p)$, and consequently $h(x) \le \langle p, x\rangle - f^*(p)\ \forall x$. Thus

$$\sup\{h \mid h \in Q\} \le \sup_p\{\langle p, x\rangle - f^*(p)\} = f^{**}, \tag{2.51}$$

If f is convex and closed then, since it is proper, by Theorem 2.5 it is just equal to the function on the left-hand side of (2.51) and since $f \ge f^{**}$, it follows that $f = f^{**}$. Conversely, if $f = f^{**}$ then f is the conjugate of f^*, hence is convex and closed. Turning to (iii) observe that if $Q = \emptyset$ then $f^*(p) = \sup_x\{\langle p, x\rangle - f(x)\} = +\infty$ for all p and consequently, $f^{**} \equiv -\infty$. But in this case, $\sup\{h|\ h \in Q\} = -\infty$, too, hence $f^{**} = \sup\{h|\ h \in Q\}$. On the other hand, if there is $h \in Q$ then from (2.51) $h \leq f^{**}$; conversely, if $h \leq f^{**}$ then $h \leq f$ (because $f^{**} \leq f$). In this case, since $f^{**}(x) \leq f(x) < +\infty$ at least for some x, and $f^{**}(x) > -\infty\ \forall x$ (because there is an affine function minorizing f^{**}), it follows that f^{**} is proper. By Theorem 2.5, then $f^{**} = \sup\{h|\ h \text{ affine } h \leq f^{**}\} = \sup\{h|\ h \in Q\}$. This proves the equality in (iii), and so, by Corollary 2.8, $f^{**} = \text{cl}(\text{conv})f$. □

2.11 Extremal Values of Convex Functions

The smallest and the largest values of a convex function on a given convex set are often of particular interest.

Let $f : \mathbb{R}^n \to [-\infty, +\infty]$ be an arbitrary function, and C an arbitrary set in $\mathbb{R}^n$. A point $x^0 \in C \cap \text{dom}f$ is called a *global minimizer* of $f(x)$ on C if $-\infty < f(x^0) \leq f(x)$ for all $x \in C$. It is called a *local minimizer* of $f(x)$ on C if there exists a neighborhood $U(x^0)$ of x^0 such that $-\infty < f(x^0) \leq f(x)$ for all $x \in C \cap U(x^0)$. The concepts of *global maximizer* and *local maximizer* are defined analogously. For an arbitrary function f on a set C we denote the set of all global minimizers (maximizers) of f on C by $\text{argmin}_{x \in C} f(x)$ ($\text{argmax}_{x \in C} f(x)$, resp.). Since $\min_{x \in C} f(x) = -\max_{x \in C}(-f(x))$ the theory of the minimum (maximum) of a convex function is the same as the theory of the maximum (minimum, resp.) of a concave function.

2.11.1 Minimum

Proposition 2.30 *Let C be a nonempty convex set in $\mathbb{R}^n$, and $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Any local minimizer of f on C is also global. The set $\underset{x \in C}{\text{argmin}} f(x)$ is a convex subset of C.*

Proof Let $x^0 \in C$ be a local minimizer of f and $U(x^0)$ be a neighborhood such that $f(x^0) \leq f(x)\ \forall x \in C \cap U(x^0)$. For any $x \in C$ we have $x_\lambda := (1-\lambda)x^0 + \lambda x \in C \cap U(x^0)$ for sufficiently small $\lambda > 0$. Then $f(x^0) \leq f(x_\lambda) \leq (1-\lambda)f(x^0) + \lambda f(x)$, hence $f(x^0) \leq f(x)$, proving the first part of the proposition. If $\alpha = \min f(C)$ then $\text{argmin}_{x \in C} f(x)$ coincides with the set $C \cap \{x|\ f(x) \leq \alpha\}$ which is a convex set by the convexity of $f(x)$ (Proposition 2.11). □

Remark 2.1 A real-valued function f on a convex set C is said to be *strictly convex* on C if

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

for any two distinct points $x^1, x^2 \in C$ and $0 < \lambda < 1$. For such a function f the set $\text{argmin}_{x \in C} f(x)$, if nonempty, is a singleton, i.e., a strictly convex function $f(x)$ on C has at most one minimizer over C. In fact, if there were two distinct minimizers x^1, x^2 then by strict convexity $f(\frac{x^1+x^2}{2}) < f(x^1)$, which is impossible.

Proposition 2.31 *Let C be a convex set in $\mathbb{R}^n$, and $f : \mathbb{R}^n \to [-\infty, +\infty]$ a convex function which is finite on C. For a point $x^0 \in C$ to be a minimizer of f on C it is necessary and sufficient that*

$$0 \in \partial f(x^0) + N_C(x^0), \tag{2.52}$$

*where $N_C(x^0)$ denotes the (outward) normal cone of C at x^0. (*cf Sect. 1.6*)*

Proof If (2.52) holds there is $p \in \partial f(x^0) \cap -N_C(x^0)$. For every $x \in C$, since $p \in \partial f(x^0)$, we have $\langle p, x - x^0\rangle \leq f(x) - f(x^0)$, i.e., $f(x^0) \leq f(x) - \langle p, x - x^0\rangle$; on the other hand, since $p \in -N_C(x^0)$, we have $\langle p, x - x^0\rangle \geq 0$, hence $f(x^0) \leq f(x)$, i.e., x^0 is a minimizer. Conversely, if $x^0 \in \operatorname{argmin}_{x\in C} f(x)$ then the system

$$(x, y) \in C \times \mathbb{R}^n,\ x - y = 0,\ f(y) - f(x^0) < 0$$

is inconsistent. Define $D := C \times \mathbb{R}^n$ and $A(x, y) := x - y$, so that $A(D) = C - \mathbb{R}^n$. For any ball U around 0, $x^0 + U \subset \mathbb{R}^n$, hence $U = x^0 - (x^0 + U) \subset A(D)$, and so $0 \in \operatorname{int} A(D)$. Therefore, by Theorem 2.4, there exists a vector $p \in \mathbb{R}^n$ such that

$$\langle p, x - y\rangle + f(y) - f(x^0) \geq 0 \quad \forall (x, y) \in C \times \mathbb{R}^n.$$

Letting $y = x^0$ yields $\langle p, x - x^0\rangle \geq 0\ \forall x \in C$, i.e., $p \in -N_C(x^0)$, then letting $x = x^0$ yields $f(y) - f(x^0) \geq \langle p, y - x^0\rangle\ \forall y \in \mathbb{R}^n$, i.e., $p \in \partial f(x^0)$. Thus, $p \in -N_C(x^0) \cap \partial f(x^0)$, completing the proof. □

Corollary 2.12 *Under the assumptions of the above proposition, an interior point x^0 of C is a minimizer if and only if $0 \in \partial f(x^0)$.*

Proof Indeed, $N_C(x^0) = \{0\}$ if $x^0 \in \operatorname{int} C$. □

Proposition 2.32 *Let C be a nonempty compact set in $\mathbb{R}^n$, $f : C \to \mathbb{R}$ an arbitrary continuous function, f^c the convex envelope of f over C. Then any global minimizer of $f(x)$ on C is also a global minimizer of $f^c(x)$ on* $\operatorname{conv} C$.

Proof Let $x^0 \in C$ be a global minimizer of $f(x)$ on C. Since f^c is a minorant of f, we have $f^c(x^0) \leq f(x^0)$. If $f^c(x^0) < f(x^0)$ then the convex function $h(x) = \max\{f(x^0), f^c(x)\}$ would be a convex minorant of f larger than f^c, which is impossible. Thus, $f^c(x^0) = f(x^0)$ and $f^c(x) = h(x)\ \forall x \in \operatorname{conv} C$. Hence, $f^c(x^0) = f(x^0) \leq f^c(x)\ \forall x \in \operatorname{conv} C$, i.e., x^0 is also a global minimizer of $f^c(x)$ on $\operatorname{conv} C$. □

2.11.2 Maximum

In contrast with the minimum, a local maximum of a convex function may not be global. Generally speaking, local information is not sufficient to identify a global maximizer of a convex function.

Proposition 2.33 *Let C be a convex set in $\mathbb{R}^n$, and $f : C \to \mathbb{R}$ be a convex function. If $f(x)$ attains its maximum on C at a point $x^0 \in \mathrm{ri}C$ then $f(x)$ is constant on C. The set $\operatorname*{argmax}_{x\in C} f(x)$ is a union of faces of C.*

Proof Suppose that f attains its maximum on C at a point $x^0 \in \mathrm{ri}C$ and let x be an arbitrary point of C. Since $x^0 \in \mathrm{ri}C$ there is $y \in C$ such that $x^0 = \lambda x + (1-\lambda)y$ for some $\lambda \in (0,1)$. Then $f(x^0) \le \lambda f(x) + (1-\lambda)f(y)$, hence $\lambda f(x) \ge f(x^0) - (1-\lambda) f(y) \ge f(x^0) - (1-\lambda)f(x^0) = \lambda f(x^0)$. Thus $f(x) \ge f(x^0)$, hence $f(x) = f(x^0)$, proving the first part of the proposition. The second part follows, because for any maximizer x^0 there is a face F of C such that $x^0 \in \mathrm{ri}F$: then by the previous argument, any point of this face is a global minimizer. □

Proposition 2.34 *Let C be a closed convex set, and $f : C \to \mathbb{R}$ be a convex function. If C contains no line and $f(x)$ is bounded above on every halfline of C then*

$$\sup\{f(x)|\ x \in C\} = \sup\{f(x)|\ x \in V(C)\},$$

where $V(C)$ is the set of extreme points of C. If the maximum of $f(x)$ is attained at all on C, it is attained on $V(C)$.

Proof By Theorem 1.7, $C = \mathrm{conv}V(C) + K$, where K is the convex cone generated by the extreme directions of C. Any point of C which is actually not an extreme point belongs to a halfline emanating from some $v \in V(C)$ in the direction of a ray of K. Since $f(x)$ is finite and bounded above on this halfline, its maximum on the halfline is attained at v (Proposition 2.12, (ii)). Therefore, the supremum of $f(x)$ on C is reduced to the supremum on $\mathrm{conv}V(C)$. The conclusion then follows from the fact that any $x \in \mathrm{conv}V(C)$ is of the form $x = \sum_{i\in I} \lambda_i v^i$, with $|I| < +\infty, v^i \in V(C), \lambda_i \ge 0, \sum_{i\in I} \lambda_i = 1$, hence $f(x) \le \sum_{i\in I} \lambda_i f(v^i) \le \max_{i\in I} f(v^i)$. □

Corollary 2.13 *A real-valued convex function $f(x)$ on a polyhedron C containing no line is either unbounded above on some unbounded edge or attains its maximum at an extreme point of C.*

Corollary 2.14 *A real-valued convex function $f(x)$ on a compact convex set C attains its maximum at an extreme point of C.*

The latter result is in fact true for a wider class of functions, namely for *quasiconvex* functions. As was defined in Sect. 2.3, these are functions $f : \mathbb{R}^n \to [-\infty, +\infty]$ such that for any real number α, the level set $L_\alpha := \{x \in \mathbb{R}^n | f(x) \le \alpha\}$ is convex, or equivalently, such that

$$f((1-\lambda)x^1 + \lambda x^2) \le \max\{f(x^1), f(x^2)\} \tag{2.53}$$

for any $x^1, x^2 \in C$ and any $\lambda \in [0,1]$.

To see that Corollary 2.14 extends to quasiconvex functions, just note that a compact convex set C is the convex hull of its extreme points (Corollary 1.13), so any $x \in C$ can be represented as $x = \sum_{i \in I} \lambda_i v^i$, where v^i are extreme points, $\lambda_i \geq 0$ and $\sum_{i \in I} \lambda_i = 1$. If $f(x)$ is a quasiconvex function finite on C, and $\alpha = \max_{i \in I} f(v^i)$, then $v^i \in C \cap L_\alpha$, $\forall i \in I$, hence $x \in C \cap L_\alpha$, because of the convexity of the set $C \cap L_\alpha$. Therefore, $f(x) \leq \alpha = \max_{i \in I} f(v^i)$, i.e., the maximum of f on C is attained at some extreme point.

A function $f(x)$ is said to be *quasiconcave* if $-f(x)$ is quasiconvex. A convex (concave) function is of course quasiconvex (quasiconcave), but the converse may not be true, as can be demonstrated by a monotone nonconvex function of one variable. Also it is easily seen that an upper envelope of a family of quasiconvex functions is quasiconvex, but the sum of two quasiconvex functions may not be quasiconvex.

2.12 Minimax and Saddle Points

2.12.1 Minimax

Given a function $f(x, y) : C \times D \to \mathbb{R}$ we can compute $\inf_{x \in C} f(x, y)$ and $\sup_{y \in D} f(x, y)$. It is easy to see that there always holds

$$\sup_{y \in D} \inf_{x \in C} f(x, y) \leq \inf_{x \in C} \sup_{y \in D} f(x, y).$$

Indeed, $\inf_{x \in C} f(x, y) \leq f(z, y)\ \forall z \in C, y \in D$, so $\sup_{y \in D} \inf_{x \in C} f(x, y) \leq \sup_{y \in D} f(z, y)\ \forall z \in C$, hence $\sup_{y \in D} \inf_{x \in C} f(x, y) \leq \inf_{z \in C} \sup_{y \in D} f(z, y)$.

We would like to know when the reverse inequality is also true, i.e., when there holds the *minimax equality*

$$\gamma := \sup_{y \in D} \inf_{x \in C} f(x, y) = \inf_{x \in C} \sup_{y \in D} f(x, y) := \eta.$$

Investigations on this question date back to von Neumann (1928). A classical result of his states that if C, D are compact and $f(x, y)$ is convex in x, concave in y and continuous in each variable then the minimax equality holds. Since minimax theorems have found important applications, there has been a great deal of work on the extension of von Neumann's theorem. Almost all these extensions are based either on the separation theorem of convex sets or a fixed point argument. The most important result in this direction is due to Sion (1958). Later a more general minimax theorem was established by Tuy (1974) without any appeal to separation or fixed point argument. We present here a simplified version of the latter result which is also a refinement of Sion's theorem.

Theorem 2.7 *Let C, D be two closed convex sets in $\mathbb{R}^n, \mathbb{R}^m$, respectively, and let $f(x, y) : C \times D \to \mathbb{R}$ be a function quasiconvex, lower semi-continuous in x and quasiconcave, upper semi-continuous in y. Assume that*

(*) *There exists a finite set $N \subset D$ such that* $\sup_{y \in N} f(x, y) \to +\infty$ *as* $x \in C$, $\|x\| \to +\infty$.

Then there holds the minimax equality

$$\inf_{x\in C}\sup_{y\in D} f(x,y) = \sup_{y\in D}\inf_{x\in C} f(x,y). \tag{2.54}$$

Proof Since the inequality $\inf_{x\in C}\sup_{y\in D} f(x,y) \geq \sup_{y\in D}\inf_{x\in C} f(x,y)$ is trivial, it suffices to show the reverse inequality:

$$\inf_{x\in C}\sup_{y\in D} f(x,y) \leq \sup_{y\in D}\inf_{x\in C} f(x,y). \tag{2.55}$$

Let $\eta := \sup_{y\in D}\inf_{x\in C} f(x,y)$. If $\eta = +\infty$ then (2.55) is obvious, so we can assume $\eta < +\infty$. For an arbitrary $\alpha > \eta$ define

$$C_\alpha(y) = \{x \in C |\, f(x,y) \leq \alpha\}.$$

Since $\sup_{y\in D}\inf_{x\in C} < \alpha$, we have $C_\alpha(y) \neq \emptyset\ \forall y \in D$. If we can show that

$$\bigcap_{y\in D} C_\alpha(y) \neq \emptyset, \tag{2.56}$$

i.e., there is $x \in C$ satisfying $f(x,y) \leq \alpha$ for all $y \in D$, then $\inf_{x\in C}\sup_{y\in D} \leq \alpha$ and since this is true for every $\alpha > \eta$ it will follow that $\inf_{x\in C}\sup_{y\in D} \leq \eta$, proving (2.55). Thus, all is reduced to establishing (2.56). This will be done in three stages. To simplify the notation, from now on we shall omit the subscript α and write simply $C(a), C(b)$, etc…

I. Let us first show that for every pair $a, b \in D$ the two sets $C(a)$ and $C(b)$ intersect. Assume the contrary, that

$$C(a) \cap C(b) = \emptyset. \tag{2.57}$$

Consider an arbitrary $\lambda \in [0,1]$ and let $y_\lambda = (1-\lambda)a + \lambda b$. If $x \in C(y_\lambda)$, i.e., $f(x,y_\lambda) \leq \alpha$ then $\min\{f(x,a), f(x,b)\} \leq f(x,y_\lambda) \leq \alpha$ by quasiconcavity of $f(x,.)$, hence, either $f(x,a) \leq \alpha$ or $f(x,b) \leq \alpha$. Therefore,

$$C(y_\lambda) \subset C(a) \cup C(b). \tag{2.58}$$

Since $C(y_\lambda)$ is convex it follows from (2.58) that $C(y_\lambda)$ cannot meet both sets $C(a)$ and $C(b)$ which are disjoint by assumption (2.57). Consequently, for every $\lambda \in [0,1]$, one and only one of the following alternatives occurs:

$$\text{(a) } C(y_\lambda) \subset C(a); \qquad \text{(b) } C(y_\lambda) \subset C(b).$$

Denote by $M_a(M_b$, resp.) the set of all $\lambda \in [0,1]$ satisfying (a) (satisfying (b), resp.). Clearly $0 \in M_a, 1 \in M_b, M_a \cup M_b = [0,1]$ and, analogously to (2.58):

$$C(y_\lambda) \subset C(y_{\lambda_1}) \cup C(y_{\lambda_2}) \quad \forall \lambda \in [\lambda_1, \lambda_2]. \tag{2.59}$$

Therefore, $\lambda \in M_a$ implies $[0,\lambda] \subset M_a$, and $\lambda \in M_b$ implies $[\lambda, 1] \subset M_b$. Let $s = \sup M_a = \inf M_b$ and assume, for instance, that $s \in M_a$ (the argument is similar if $s \in M_b$). We show that (2.57) leads to a contradiction.

Since $\alpha > \eta \geq \inf_{x\in C} f(x, y_s)$, we have $f(\overline{x}, y_s) < \alpha$ for some $\overline{x} \in C$. By upper semi-continuity of $f(\overline{x}, .)$ there is $\varepsilon > 0$ such that $f(\overline{x}, y_{s+\varepsilon}) < \alpha$ and so $\overline{x} \in C(y_{s+\varepsilon})$. But $\overline{x} \in C(y_s) \subset C(a)$, hence $C(y_{s+\varepsilon}) \subset C(a)$, i.e., $s + \varepsilon \in M_a$, contradicting the definition of s. Thus (2.57) cannot occur and we must have $C(a) \cap C(b) \neq \emptyset$ for all $a, b \in C, \alpha < \gamma$.

II. We now show that any finite collection $C(y^1), \ldots, C(y^k)$ with $y^1, \ldots, y^k \in D$ has a nonempty intersection. By (I) this is true for $k = 2$. Assuming this is true for $k = h-1$ let us consider the case $k = h$. Set $C' = C(y^h), C'(y) = C' \cap C(y)$. From part I we know that $C'(y) \neq \emptyset$ for every $y \in D$. This means that for all $\alpha > \eta : \ \forall y \in D \quad \exists x \in C' \quad f(x,y) \leq \alpha$, so that $\sup_{y\in D} \inf_{x\in C'} f(x,y) \leq \alpha$. Since $\eta = \sup_{y\in D} \inf_{x\in C} f(x,y) \leq \sup_{y\in D} \inf_{x\in C'} f(x,y)$, it follows that $\eta = \sup_{y\in D} \inf_{x\in C'} f(x,y)$. So all the hypotheses of the theorem still hold when C is replaced by C'. It then follows from the induction assumption that the sets $C'(y^1), \ldots, C'(y^{h-1})$ have a nonempty intersection. Thus the family $\{C_\alpha(y), y \in D\}$ has the finite intersection property.

III. Finally, for every $y \in D$ let $C^*(y) = \{x \in C | \, f(x,y) \leq \alpha, \sup_{z\in N} f(x,z) \leq \alpha\}$. Then $C^*(y) \subset C^N := \{x \in C | \ \sup_{z\in N} f(x,z) \leq \alpha\}$ and the set C^N is compact because if it were not so there would exist a sequence $x^k \in C^N$ such that $\|x^k| \to +\infty$, contradicting assumption (*). On the other hand, for any finite set $E \subset D$ clearly $\cap_{y\in E} C^*(y) = \cap_{y\in E\cup N} C(y)$, so by part II $\cap_{y\in E} C^*(y) \neq \emptyset$, i.e., the family $\{C^*(y), y \in D\}$ has the finite intersection property. Since every $C^*(y)$ is a subset of the compact set C^N it follows that $\cap_{y\in D} C_\alpha(y) = \cap_{y\in D} C^*(y) \neq \emptyset$, i.e. (2.56) must hold. □

Remark 2.2 In view of the symmetry in the roles of x, y Theorem 2.7 still holds if instead of condition (*) one assumes that

(!) *There exists a finite set* $M \subset C$ *such that* $\inf_{x\in M} f(x,y) \to -\infty$ *as* $y \in D, \|y\| \to +\infty$.

The proof is analogous, using $D_\alpha(x) := \{y \in D | \, f(x,y) \geq \alpha\}$ with $\alpha > \gamma := \inf_{x\in C} \sup_{y\in D} f(x,y)$ (instead of $C_\alpha(y)$ with $\alpha < \eta := \sup_{y\in D} \inf_{x\in C} f(x,y)$) and proving that $\cap_{x\in C} D_\alpha(x) \neq \emptyset$.

2.12.2 Saddle Point of a Function

A pair $(\bar{x}, \bar{y}) \in C \times D$ is called a *saddle point* of the function $f(x, y) : C \times D \to \mathbb{R}$ if

$$f(\bar{x}, y) \le f(\bar{x}, \bar{y}) \le f(x, \bar{y}) \quad \forall x \in C,\ \forall y \in D. \tag{2.60}$$

This means

$$\min_{x \in C}(x, \bar{y}) = f(\bar{x}, \bar{y}) = \max_{y \in D} f(\bar{x}, y),$$

so in a neighborhood of the point $(\bar{x}, \bar{y}, f(\bar{x}, \bar{y}))$ the set epi$f := \{(x, y, t) |\ (x, y) \in C \times D,\ t \in \mathbb{R},\ f(x, y) \le t\}$ reminds the image of a saddle (or a mountain pass).

Proposition 2.35 *A point $(\bar{x}, \bar{y}) \in C \times D$ is a saddle point of $f(x, y) : C \times D \to \mathbb{R}$ if and only if*

$$\max_{y \in D} \inf_{x \in C} f(x, y) = \min_{x \in C} \sup_{y \in D} f(x, y). \tag{2.61}$$

Proof We have

$$\begin{aligned}
(2.60) &\Leftrightarrow \sup\nolimits_{y \in D} f(\bar{x}, y) \le f(\bar{x}, \bar{y}) \le \inf\nolimits_{y \in C} f(x, \bar{y}) \\
&\Leftrightarrow \inf\nolimits_{x \in C} \sup\nolimits_{y \in D} f(x, y) \le \sup\nolimits_{y \in D} f(\bar{x}, y) \le f(\bar{x}, \bar{y}) \\
&\quad \le \inf\nolimits_{x \in C} f(x, \bar{y}) \le \sup\nolimits_{y \in D} \inf\nolimits_{x \in C} f(x, y).
\end{aligned}$$

Since always $\inf_{x \in C} \sup_{y \in D} f(x, y) \ge \sup_{y \in D} \inf_{x \in C} f(x, y)$ we must have equality everywhere in the above sequence of inequalities. Therefore, (2.60) must be equivalent to (2.61). □

As a consequence of Theorem 2.7 and Remark 2.2 we can now state

Proposition 2.36 *Assume that $C \subset \mathbb{R}^n$, $D \subset \mathbb{R}^m$ are nonempty closed convex sets and the function $f(x, y) : C \times D \to \mathbb{R}$ is quasiconvex l.s.c in x and quasiconcave u.s.c. in y. If both the following conditions hold:*

(i) *There exists a finite set $N \subset D$ such that $\sup_{y \in N} f(x, y) \to +\infty$ as $x \in C,\ \|x\| \to +\infty$;*
(ii) *There exists a finite set $M \subset C$ such that $\inf_{x \in M} f(x, y) \to +\infty$ as $y \in D,\ \|y\| \to +\infty$;*

then there exists a saddle point $(\bar{x}, \bar{y})$ for the function $f(x, y)$.

Proof If (i) holds, we have $\min_{x \in C} \sup_{y \in D} f(x, y) = \sup_{y \in D} \inf_{x \in C} f(x, y)$ by Theorem 2.7. If (ii) holds, we have $\inf_{x \in C} \sup_{y \in D} f(x, y) = \max_{y \in D} \inf_{x \in C} f(x, y)$ by Remark 2.2. Hence the condition (2.61). □

2.13 Convex Optimization

We assume that the reader is familiar with convex optimization problems, i.e., optimization problems of the form

$$\min\{f(x)|\ g_i(x) \le 0\ (i = 1, \ldots, m),\ h_j(x) = 0\ (j = 1, \ldots, p),\ x \in \Omega\}, \tag{2.62}$$

where $\Omega \subset \mathbb{R}^n$ is a closed convex set, $f, g_1, \ldots, g_m$ are convex functions finite on an open domain containing Ω, $h_1, \ldots, h_m$ are affine functions.

In this section we study a generalization of problem (2.62), where the convex inequality constraints are understood in a generalized sense.

2.13.1 Generalized Inequalities

a. Ordering Induced by a Cone

A cone $K \subset \mathbb{R}^n$ induces a partial ordering $\preceq_K$ on $\mathbb{R}^n$ such that

$$x \preceq_K y \quad \Leftrightarrow \quad y - x \in K.$$

We also write $x \succeq_K y$ to mean $y \preceq_K x$. In the case $K = \mathbb{R}^n_+$ this is the usual ordering $x \le y \ \Leftrightarrow x_i \le y_i,\ i = 1, \ldots, m.$

The following properties of the ordering $\preceq_K$ are straightforward:

(i) transitivity: $x \preceq_K y,\ y \preceq_K z \Rightarrow x \preceq_K z$;
(ii) reflexivity: $x \preceq_K x\ \forall x \in \mathbb{R}^n$;
(iii) preservation under addition: $x \preceq_K y,\ x' \preceq_K y' \ \Rightarrow x + x' \preceq_K y + y'$;
(iv) preservation under nonnegative scaling: $x \preceq_K y,\ \alpha > 0 \ \Rightarrow \alpha x \preceq_K \alpha y$.

The ordering induced by a cone K is of particular interest when K is *closed, solid* (i.e., has nonempty interior), and *pointed* (i.e., contains no line: $x \in K \Rightarrow -x \notin K$). In that case a relation $x \preceq_K y$ is called a *generalized inequality*. We also write $x \prec_K y$ to mean that $y - x \in \mathrm{int}K$ and call such a relation a *strict generalized inequality*.

Generalized inequalities enjoy the following important properties:

(i) $x \preceq_K y,\ y \preceq_K x \ \Rightarrow x = y$;
(ii) $x \not\prec_K x$;
(iii) $x^k \preceq_K y^k,\ x^k \to x,\ y^k \to y \Rightarrow x \preceq_K y$;
(iv) If $x \prec_K y$ then for u, v small enough $x + u \prec_K y + v$;
(v) $x \prec_K y,\ u \preceq_K \Rightarrow x + u \prec_K y + v$;
(vi) The set $\{z|\ x \preceq_K z \preceq_K y\}$ is bounded.

For instance, (ii) is true because $x \prec_K x$ implies $0 = x - x \in \mathrm{int}K$, which is impossible as K is pointed; (vi) holds because the set $E = \{z|\ x \preceq_K z \preceq_K y\}$

is closed, so if there exists $\{z^k\} \subset E, \|z^k\| \to +\infty$ then, up to a subsequence, $z^k/\|z^k\| \to u$ with $\|u\| = 1$ and since $\frac{z^k}{\|z^k\|} \in \frac{x}{\|z^k\|} + K, \frac{z^k}{\|z^k\|} \in \frac{x}{\|z^k\|} - K$, letting $k \to +\infty$ yields $u \in K, u \in -K$, hence by (i) $u = 0$ conflicting with $\|u\| = 1$.

b. Dual Cone

The *dual cone* of a cone K is by definition the cone

$$K^* = \{y \in \mathbb{R}^n | \ \langle x, y\rangle \geq 0 \ \forall x \in K\}.$$

Note that $K^* = -K^o$, where K^o is the polar of K (Sect. 1.8). As can easily be seen:

(i) K^* is a closed convex cone;
(ii) K^* is also the dual cone of the closure of K;
(iii) $(K^*)^* = \mathrm{cl}K$;
(iv) If $x \in \mathrm{int}K$ then $\langle x, y\rangle > 0 \ \forall y \in K^* \setminus \{0\}$.

A cone K is said to be *self-dual* if $K^* = K$. Clearly the orthant $\mathbb{R}^n_+$ is a self-dual cone.

Lemma 2.1 *If K is a closed convex cone then*

$$y \notin K \ \Leftrightarrow \ \exists \lambda \in K^* \ \langle \lambda, y\rangle < 0.$$

Proof By (iii) above $K = (K^*)^*$ so $y \notin K$ if and only if $y \notin (K^*)^*$, hence if and only if there exists $\lambda \in K^*$ satisfying $\langle \lambda, y\rangle < 0$. □

c. *K*-Convex Functions

Given a convex cone $K \subset \mathbb{R}^m$ inducing an ordering $\preceq_K$ on $\mathbb{R}^m$ a map $g : \mathbb{R}^n \to \mathbb{R}^m$ is called K-convex if for every $x^1, x^2 \in \mathbb{R}^n$ and $0 \leq \alpha \leq 1$ we have the generalized inequality

$$g(\alpha x^1 + (1-\alpha)x^2) \preceq_K \alpha g(x^1) + (1-\alpha)g(x^2). \tag{2.63}$$

For instance, the map $g : \mathbb{R}^n \to \mathbb{R}^m$ is $\mathbb{R}^m_+$-convex (or component-wise convex) if each function $g_i(x), i = 1 \ldots, m$, is convex.

Lemma 2.2 *If $g : \mathbb{R}^n \to \mathbb{R}^m$ is a K-convex map then for every $\lambda \in K^*$ the function $\langle \lambda, g(x)\rangle = \sum_{i=1}^m \lambda_i g_i(x)$ is convex in the usual sense and the set $\{x \in \mathbb{R}^n | \ g(x) \preceq_K 0\}$ is convex.*

Proof For every $x^1, x^2 \in \mathbb{R}^n$, we have by (2.63)

$$\alpha g(x^1) + (1-\alpha)g(x^2) - g(\alpha x^1 + (1-\alpha)x^2) \in K.$$

Since $\lambda \in K^*$ it follows that

$$\langle \lambda, \alpha g(x^1) + (1-\alpha)g(x^2) - g(\alpha x^1) + (1-\alpha)g(x^2)\rangle \geq 0,$$

hence $\alpha\langle \lambda, g(x^1)\rangle + (1-\alpha)\langle \lambda, g(x^1)\rangle \geq \langle \lambda, g(\alpha x^1) + (1-\alpha)g(x^2)\rangle$, proving that the function $\langle \lambda, g(x)\rangle$ is convex.

Further, by (2.2), if $g(x^1) \preceq_K 0,\ g(x^2) \preceq_K 0$ then

$$g(\alpha x^1) + (1-\alpha)x^2) \preceq_K \alpha g(x^1) + (1-\alpha)g(x^2) \preceq_K 0,$$

proving that the set $\{x|\ g(x) \preceq_K 0\}$ is convex. □

2.13.2 Generalized Convex Optimization

A *generalized convex optimization problem* is a problem of the form

$$\min\{f(x)|\ g_i(x) \preceq_{K_i} 0\ (i = 1, \ldots, m),\ h(x) = 0,\ x \in \Omega\}, \tag{2.64}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function, K_i is a closed, solid, pointed convex cone in $\mathbb{R}^{s_i}$, $g_i : \mathbb{R}^n \to \mathbb{R}^{s_i}$, $i = 1, \ldots, m$, is K_i-convex, finite on the whole $\mathbb{R}^n$, $h : \mathbb{R}^n \to \mathbb{R}^p$ is an affine map, and $\Omega \subset \mathbb{R}^n$ is a closed convex set. By Lemma 2.2 the constraint set of this problem is convex, so this is also a problem of minimizing a convex function over a convex set.

Let $\lambda_i \in K_i^*$ be the Lagrange multiplier associated with the generalized inequality $g_i(x) \preceq_{K_i} 0$ and $\mu \in \mathbb{R}^p$ the Lagrange multiplier associated with the equality $h(x) = 0$. So the Lagrangian of the problem is the function

$$L(x, \lambda, \mu) := f(x) + \sum_{i=1}^{m} \langle \lambda_i, g_i(x)\rangle + \langle \mu, h(x)\rangle,$$

where $\lambda_i \in K_i^*, i = 1, \ldots, m$, and $\mu \in \mathbb{R}^p$. The dual Lagrange function is

$$\varphi(\lambda, \mu) = \inf_{x \in \Omega} L(x, \lambda, \mu) = \inf_{x \in \Omega}\{f(x) + \sum_{I=1}^{m} \langle \lambda_i, g_i(x)\rangle + \langle \mu, h(x)\rangle\}.$$

Since $\varphi(\lambda, \mu)$ is the lower envelope of a family of affine functions in (λ, μ), it is a concave function. Setting $K^* = K_1^* \times \cdots \times K_m^*,\ \lambda = (\lambda_1, \ldots, \lambda_m)$ we can show that

$$\sup_{\lambda \in K^*, \mu \in \mathbb{R}^p} L(x, \lambda, \mu)$$
$$= \begin{cases} f(x) & \text{if } g_i(x) \preceq_{K_i} 0\ (i = 1, \ldots, m), h(x) = 0; \\ +\infty & \text{otherwise} \end{cases}$$

In fact, if $g_i(x) \preceq_{K_i} 0$ $(i = 1, \ldots, m), h(x) = 0$ then the supremum is attained for $\lambda = 0, \mu = 0$ and equals $f(x)$. On the other hand, if $h(x) \neq 0$ there is $\mu \in \mathbb{R}^p$ satisfying $\langle \mu, h(x) \rangle > 0$ and for $\lambda = 0$ the supremum is equal to $\sup_{\theta>0}\{f(x) + \langle \theta\mu, h(x) \rangle\} = +\infty$. If there is an $i = 1, \ldots, m$ with $g_i(x) \not\preceq_{K_i} 0$ then by Lemma 2.1 there is $\lambda_i \in K_i^*$ satisfying $\langle \lambda_i, g_i(x) \rangle > 0$, hence for $\mu = 0, \lambda_j = 0$ $\forall j \neq i$, the supremum equals $\sup_{\theta>0}\{f(x) + \theta\langle \lambda_i, g_i(x) \rangle\} = +\infty$.

Thus, the problem (2.64) can be written as

$$\inf_{x\in\Omega} \sup_{\lambda\in K^*,\mu\in\mathbb{R}^p} L(x, \lambda, \mu).$$

The dual problem is

$$\sup_{\lambda\in K^*,\mu\in\mathbb{R}^p} \inf_{x\in\Omega} L(x, \lambda, \mu),$$

that is,

$$\sup_{\lambda\in K^*,\mu\in\mathbb{R}^p} \varphi(\lambda, \mu). \tag{2.65}$$

Theorem 2.8

(i) (weak duality) *The optimal value in the dual problem never exceeds the optimal value in the primal problem (2.64).*

(ii) (strong duality) *The optimal values in the two problems are equal if the Slater condition holds, i.e., if*

$$\exists \bar{x} \in \mathrm{int}\Omega \quad h(\bar{x}) = 0,\ g_i(\bar{x}) \prec_{K_i} 0,\ i = 1, \ldots, m.$$

Proof (i) is straightforward, we need only prove (ii). By Lemma 2.2 for every $\lambda_i \in K_i^*$ the function $\langle \lambda_i, g_i(x) \rangle$ is convex (in the usual sense), finite on $\mathbb{R}^n$ and hence continuous. So $L(x, \lambda, \mu)$ is a convex continuous function in $x \in \Omega$ for every fixed $(\lambda, \mu) \in D := K_1^* \times \cdots \times K_m^* \times \mathbb{R}^p$ and affine in $(\lambda, \mu) \in D$ for every fixed $x \in h(\Omega)$.

Since $h : \mathbb{R}^n \to \mathbb{R}^p$ is an affine map, without loss of generality it can be assumed that $h(\Omega) = \mathbb{R}^p$. Since $h(\bar{x}) = 0$ and $\bar{x} \in \mathrm{int}\Omega$, we have $0 \in \mathrm{int}\Omega$, so for every $j = 1, \ldots, p$ there is an $a^j \in \Omega$ such that $h(a^j)$ has its j-th component equal to 1, and all other components equal to 0. Then for sufficiently small $\varepsilon > 0$, we have $x^i := \bar{x} + \varepsilon(a^j - \bar{x}) \in \Omega,\ \hat{x}^j := \bar{x} - \varepsilon(a^j - \bar{x}) \in \Omega$, and so

$$g_i(x^j) < 0\ (i = 1, \ldots, m),\ h_j(x^j) > 0,\ h_j(x^j) = 0\ \forall i \neq j$$
$$g_i(\hat{x}^j) < 0\ (i = 1, \ldots, m),\ h_j(\hat{x}^j) < 0,\ h_i(\hat{x}^j) = 0\ \forall i \neq j.$$

Let $M = \{x^j, \hat{x}^j, j = 1, \ldots, p\}$. For every $(\lambda, \mu) \in D \setminus \{0\}$ we can write

$$\rho(\lambda, \mu) := \min_{x\in M}[\sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)] < 0$$

hence, $\theta := \max\{\rho(\lambda,\mu)|\ \lambda \in \mathbb{R}^m_+,\ \mu \in \mathbb{R}^p, \|\lambda\| + \|\mu\| = 1\} < 0$. Consequently,

$$\min_{x\in M} L(x,(\lambda,\mu)) \le \max_{x\in M} f(x) + \min_{x\in M}\{\sum_{i=1}^m \lambda_i g_i(x) + \sum_{j=1}^p \mu_j h_j(x)\}$$
$$\le \max_{x\in M} f(x) + (\|\lambda\| + \|\mu\|)\theta \to -\infty$$

as $(\lambda,\mu) \in D, \|\lambda\| + \|\mu\| \to +\infty$. So the function $L(x,(\lambda,\mu))$ satisfies condition (!) in Remark 2.2 with $C = \Omega, y = (\lambda,\mu)$. By virtue of this Remark,

$$\inf_{x\in\Omega} \sup_{(\lambda,\mu)\in D} L(x,(\lambda,\mu)) = \max_{(\lambda,\mu)\in D} \inf_{x\in\Omega} L(x,(\lambda,\mu)),$$

completing the proof of (ii). □

The difference between the optimal values in the primal and dual problems is called the *duality gap*. Theorem 2.8 says that for problem (2.64) the duality gap is never negative and is zero if Slater condition is satisfied.

2.13.3 Conic Programming

Let $K \subset \mathbb{R}^m$ be a closed, solid, and pointed convex cone, let $c \in \mathbb{R}^n$ and let A be an $m \times n$ matrix, and $b \in \mathbb{R}^m$. Then the following optimization problem:

$$\min\{\langle c,x\rangle |\ Ax \preceq_K b\} \tag{2.66}$$

is called a *conic programming problem*.

Since $A(\alpha x^1 + (1-\alpha)x^2) - [\alpha Ax^1 + (1-\alpha)Ax^2] = 0 \in K$, i.e.,

$$A(\alpha x^1 + (1-\alpha)x^2) \succeq_K \alpha Ax^1 + (1-\alpha)Ax^2,$$

for all $x^1, x^2 \in \mathbb{R}^n, 0 \le \alpha \le 1$, the map $x \mapsto b - Ax$ is K-convex. So a conic program is nothing but a special case of the above considered problem (2.64) when $m = 1, K_1 = K, f(x) = \langle c,x\rangle, g_1(x) = b - Ax, h \equiv 0, \Omega = \mathbb{R}^n$.

The Lagrangian of problem (2.66) is

$$L(x,y) = \langle c,x\rangle + \langle y, b - Ax\rangle = \langle c - A^T y, x\rangle + \langle b,y\rangle \quad (y \succeq_K 0).$$

But, as can easily be seen,

$$\inf_{x\in\mathbb{R}^n} L(x,y) = \begin{cases} \langle b,y\rangle \text{ if } A^T y = x \\ -\infty \quad \text{otherwise} \end{cases}$$

so the dual of the conic programming problem (2.66) is the problem

$$\max\{\langle b, y\rangle \mid A^T y = c,\ y \succeq_K 0\}. \tag{2.67}$$

Clearly linear programming is a special case of conic programming when $K = \mathbb{R}^n_+$. However, while strong duality always holds for linear programming (except only when both the primal and the dual problems are infeasible), it is not so for conic programming. By Theorem 2.8 a sufficient condition for strong duality in conic programming is

$$\exists \overline{x} \quad A\overline{x} \succ_K b.$$

2.14 Semidefinite Programming

2.14.1 The SDP Cone

Given a symmetric $n \times n$ matrix $A = [a_{ij}]$ (i.e., a matrix A such that $A^T = A$) the *trace* of A, written $\mathrm{Tr}(A)$, is the sum of its diagonal elements:

$$\mathrm{Tr}(A) := \sum_{i=1}^{n} a_{ii}.$$

From the definition it is readily seen that

$$\mathrm{Tr}(\alpha A + \beta B) = \alpha \mathrm{Tr}(A) + \beta \mathrm{Tr}(B), \quad \mathrm{Tr}(AB) = \mathrm{Tr}(BA)$$
$$\mathrm{Tr}(A) = \lambda_1 + \cdots + \lambda_n$$

where $\lambda_i, i = 1, \ldots, n$, are the eigenvalues of A, the latter equality being derived from the development of the characteristic polynomial $\det(\lambda I_n - A)$.

Consider now the space $\mathbf{S}^n$ of all $n \times n$ symmetric matrices. Using the concept of trace we can introduce an *interior product*[1] in $\mathbf{S}^n$ defined as follows:

$$\langle A, B\rangle = \mathrm{Tr}(AB) = \sum_{i,j} a_{ij} b_{ij} = \mathrm{vec}(A)^T \mathrm{vec}(B),$$

where $\mathrm{vec}(A)$ denotes the $n \times n$ column vector whose elements are elements of the matrix A in the order $a_{11}, \cdots, a_{1n}, a_{21}, \cdots, a_{2n}, \cdots, a_{n1}, a_{nn}$. The norm of a matrix A associated with this inner product is the *Frobenius norm* given by

[1] Sometimes also written as $A \bullet B$ and called *dot product*.

$$\|A\|_F = (\langle A, A\rangle)^{1/2} = (\mathrm{Tr}(A^T A)^{1/2} = \left(\sum_{i,j} a_{ij}^2\right)^{1/2}.$$

The set of semidefinite positive matrices $X \in \mathbf{S}^n$ forms a convex cone $\mathbf{S}^n_+$ called the *SDP cone* (semidefinite positive cone). For any two matrices $A, B \in \mathbf{S}^n$ we write $A \preceq B$ to mean that $B - A$ is semidefinite positive. In particular, $A \succeq 0$ means A is a semidefinite positive matrix.

Lemma 2.3 *The cone $\mathbf{S}^n_+$ is closed, convex, solid, pointed, and self-dual, i.e.,*

$$\mathbf{S}^n_+ = (\mathbf{S}^n_+)^*.$$

Proof We prove only the self-dual property. By definition

$$(\mathbf{S}^n_+)^* = \{Y \in \mathbf{S}^n_+ \mid \mathrm{Tr}(YX) \geq 0 \ \forall X \in \mathbf{S}^n_+\}.$$

Let $Y \in \mathbf{S}^n_+$. For every $X \in \mathbf{S}^n_+$, we have

$$\mathrm{Tr}(YX) = \mathrm{Tr}(YX^{1/2}X^{1/2}) = \mathrm{Tr}(X^{1/2}YX^{1/2}) \geq 0,$$

where the last inequality holds because the matrix $X^{1/2}YX^{1/2}$ is semidefinite positive. Therefore, $\mathbf{S}^n_+ \subset (\mathbf{S}^n_+)^*$. Conversely, let $Y \in (\mathbf{S}^n_+)^*$. For any $u \in \mathbb{R}^n$, we have

$$\mathrm{Tr}(Yuu^T) = \sum_{i=1}^n y_{1i}u_iu_1 + \cdots + \sum_{i=1}^n y_{ni}u_iu_n = \sum_{i,j=1}^n y_{ij}u_iu_j = u^TYu.$$

The matrix uu^T is obviously semidefinite positive, while the trace of Yuu^T is nonnegative by assumption, so the product u^TYu is nonnegative. Since u is arbitrary, this means $Y \in \mathbf{S}^n_+$. Hence, $(\mathbf{S}^n_+)^* \subset \mathbf{S}^n_+$. □

A map $F : \mathbb{R}^n \to \mathbf{S}^m$ is said to be *convex*, or more precisely, *convex with respect to matrix inequalities* if it is $\mathbf{S}^m_+$-convex, i.e., such that for every $X, Y \in \mathbf{S}^n$ and $0 \leq t \leq 1$

$$F(tX + (1-t)Y) \preceq tF(X) + (1-t)F(Y).$$

For instance, the map $X \mapsto X^2$ is convex because for every $u \in \mathbb{R}^m$ the function $u^TX^2u = \|Xu\|^2$ is convex quadratic with respect to the components of X and so

$$u^T(\lambda X + (1-\lambda)X)^2u \leq \lambda u^TX^2u + (1-\lambda)u^TX^2u,$$

which implies $(\lambda X + (1-\lambda)X)^2 \preceq \lambda X^2 + (1-\lambda)X^2$. Analogously, the function $X \mapsto XX^T$ is convex.

2.14.2 Linear Matrix Inequality

A linear matrix inequality (LMI) is a generalized inequality

$$A_0 + x_1 A_1 + \cdots + x_n A_n \preceq 0$$

where $x \in \mathbb{R}^n$ is the variable and $A_i \in \mathbf{S}^p, i = 0, 1, \ldots, n$, are given $p \times p$ symmetric matrices. The inequality sign $\preceq$ is understood with respect to the cone $\mathbf{S}^p_+$: the notation $P \preceq 0$ means the matrix P is semidefinite negative. Obviously, $A(x) := A_0 + \sum_{k=1}^n x_k A_k \in \mathbf{S}^p_+$ and each element of this matrix is an affine function of x :

$$A(x) = [a_{ij}(x)],\ a_{ij}(x) = a^0_{ij} + \sum_{k=1}^n a^k_{ij} x_k.$$

Therefore an LMI can also be defined as an inequality of the form

$$A(x) \preceq 0,$$

where $A(x)$ is a square symmetric matrix whose every element is an affine function of x.

By definition

$$A(x) \preceq 0 \Leftrightarrow \langle y, A(x)y \rangle \leq 0 \ \forall y \in \mathbb{R}^p,$$

so setting $C := \{x \in \mathbb{R}^n |\, A(x) \preceq 0\}$, we have

$$C = \cap_{y \in \mathbb{R}^p} \{x \in \mathbb{R}^n |\ \langle y, A(x)y \rangle \leq 0\}.$$

Since for every fixed y the set $\{x \in \mathbb{R}^n |\ \langle y, A(x)y \rangle \leq 0\}$ is a halfspace, we see that the solution set C of an LMI is a closed convex set. In other words, an LMI is nothing but a specific convex inequality which is equivalent to an infinite system of linear inequalities.

Obviously, the inequality $A(x) \succeq 0$ is also an LMI, determining a convex constraint for x. Furthermore, a finite system of LMI's of the form

$$A^{(1)}(x) \preceq 0, \ldots, A^{(m)}(x) \preceq 0$$

can be equivalently written as the single LMI

$$\mathrm{Diag}(A^{(1)}(x), \cdots, A^{(m)}(x)) \preceq 0.$$

2.14.3 SDP Programming

A *semidefinite program* (*SDP*) is a problem of minimizing a linear function under an LMI constraint, i.e., a problem of the form

$$(SDP) \qquad \min\{\langle c, x\rangle \mid A_0 + x_1A_1 + \ldots + x_nA_n \preceq 0\}.$$

Clearly this is a special case of generalized convex optimization. Specifically, (*SDP*) can be rewritten in the form (2.64), with $m = 1, f(x) = \langle c, x\rangle, g_1(x) = A_0 + x_1A_1 + \ldots + x_nA_n, K_1 = \mathbf{S}^p_+$.

To form the dual problem to (*SDP*) we associate with the LMI constraint a dual variable $Y \in (\mathbf{S}^p_+)^* = \mathbf{S}^p_+$ (see Lemma 2.3), so the Lagrangian is

$$L(x, Y) = c^Tx + \mathrm{Tr}(Y(A_0 + x_1A_1 + \cdots + x_nA_n)).$$

The dual function is

$$\varphi(Y) = \inf\{L(x, Y) \mid x \in \mathbb{R}^n\}.$$

Since $L(x, Y)$ is affine in x it is unbounded below, except if it is identically zero, i.e., if $c_i + \mathrm{Tr}(YA_i) = 0, i = 1, \ldots, n$, in which case $L(x, Y) = \mathrm{Tr}(A_0Y)$. Therefore,

$$\varphi(Y) = \begin{cases} \mathrm{Tr}(A_0Y \text{ if } \mathrm{Tr}(A_iY) + c_i = 0, i = 1, \ldots, n \\ -\infty \qquad\qquad\qquad \text{otherwise} \end{cases}$$

Consequently, the dual of (*SDP*) is

$$(SDD) \qquad \max\{\mathrm{Tr}(A_0Y) \mid \mathrm{Tr}(A_iY) + c_i = 0, i = 1, \ldots, n, Y = Y^T \succeq 0\}.$$

Writing this problem in the form

$$\max\{\langle A_0, Y\rangle \mid \langle A_i, Y\rangle = -c_i, i = 1, \ldots, , Y \succeq 0\} \tag{2.68}$$

we see that (*SDP*) reminds a linear program in standard form.

By Theorem 2.8 strong duality holds for (*SDP*) if Slater condition is satisfied

$$\exists \overline{x} \quad A_0 + \overline{x}_1A_1 + \cdots + \overline{x}_nA_n. \tag{2.69}$$

2.15 Exercises

1 Let $f(x)$ be a convex function and $C = \mathrm{dom} f \subset \mathbb{R}^n$. Show that for all $x^1, x^2 \in C$ and $\lambda \notin [0, 1]$ such that $\lambda x^1 + (1 - \lambda)x^2 \in C$:

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

2 A real-valued function $f(t), -\infty < t < +\infty$, is strictly convex (cf Remark 2.1) if it has a strictly monotonically increasing derivative $f'(t)$. Apply this result to $f(t) = e^t$ and show that for any positive numbers $\alpha_1, \dots \alpha_k$;

$$\left(\prod_{i=1}^{k} \alpha_i\right)^{1/k} \leq \frac{1}{k}\sum_{i=1}^{k} \alpha_i$$

with equality holding only when $\alpha_1 = \dots = \alpha_k$.

3 A function $f(x)$ is strongly convex (cf Sect. 2.9) on a convex set C if and only if there exists $r > 0$ (modulus of strong convexity) such that $f(x) - r\|x\|^2$ is convex. Show that:

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

for all $x^1, x^2 \in C$ and $\lambda \notin [0, 1]$ such that $\lambda x^1 + (1-\lambda)x^2 \in C$.

4 Show that if $f(x)$ is strongly convex on $\mathbb{R}^n$ (with modulus of strong convexity r) then for any $x^0 \in \mathbb{R}^n$ and $p \in \partial f(x^0)$:

$$f(x) - f(x^0) \geq \langle p, x - x^0\rangle + r\|x - x^0\|^2 \ \forall x \in \mathbb{R}^n.$$

5 Notations being the same as in Exercise 4, show that for any $x^1, x^2 \in \mathbb{R}^n$ and $p^1 \in \partial f(x^1), p^2 \in \partial f(x^2)$:

$$\langle p^1 - p^2, x^1 - x^2\rangle \geq r\|x^1 - x^2\|^2.$$

6 If $f(x)$ is strongly convex on a convex set C then for any $x^0 \in C$ the level set $\{x \in C | \, f(x) \leq f(x^0)\}$ is bounded.

7 Let C be a nonempty convex set in $\mathbb{R}^n$, $f : C \to \mathbb{R}$ a convex function, Lipschitzian with constant L on C. The function

$$F(x) = \inf\{f(y) + L\|x - y\| \,|\, y \in C\}$$

is convex, Lipschitzian with constant L on the whole space and satisfies $F(x) = f(x) \ \forall x \in C$.

8 Show that if $\partial_\varepsilon f(x^0)$ is a singleton for some $x^0 \in \mathrm{dom} f$ and $\varepsilon > 0$, then $f(x)$ is an affine function.

9 For a proper convex function f : $p \in \partial f(x^0)$ if and only if $(p, -1)$ is an outward normal to the set epi$f$ at point $(x^0, f(x^0))$.

10 Let M be a nonempty set in $\mathbb{R}^n$, $h : M \to \mathbb{R}$ an arbitrary function, E an $n \times n$ matrix. The function

$$\varphi(x) = \max\{\langle x, Ey\rangle - h(y) |\ y \in M\}$$

is convex and for every $x^0 \in \mathrm{dom}\varphi$, if $y^0 \in \mathrm{argmax}_{y \in M}\{\langle x^0, Ey\rangle - h(y)\}$ then $Ey^0 \in \partial\varphi(x^0)$.

11 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function, X a closed convex subset of $\mathbb{R}^n, A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{m\times p}, c \in \mathbb{R}^m$. Show that the function $\varphi(y) = \min\{f(x)|\ Ax + By \leq c, x \in X\}$ is convex and for every $y^0 \in \mathrm{dom}\varphi$, if λ is a Kuhn–Tucker vector for the problem $\min\{f(x)|\, Ax + By^0 \leq c, x \in X\}$ then the vector $B^T\lambda$ is a subgradient of φ at y^0.

Chapter 3
Fixed Point and Equilibrium

3.1 Approximate Minimum and Variational Principle

It is well known that a real-valued function $f : \mathbb{R}^n \to \mathbb{R}$ which is l.s.c. on a nonempty closed set $X \subset \mathbb{R}^n$ attains a minimum over X if X is compact. This is true, more generally, if f is *coercive* on X, i.e., if

$$f(x) \to +\infty \text{ as } x \in X, \|x\| \to +\infty. \tag{3.1}$$

In fact, for an arbitrary $a \in X$ the set $C = \{x \in X|\, f(x) \leq f(a)\}$ is closed and if it were unbounded there would exist a sequence $\{x^k\} \subset X$ such that $f(x^k) \leq f(a), \|x^k\| \to +\infty$ while by (3.1) this would imply $f(x^k) \to +\infty$, a contradiction. So C is compact and the minimum of $f(x)$ over C, i.e., the minimum of $f(x)$ on X, exists.

If condition (3.1) fails, $f(x)$ may not have a minimum over X. In that case one must be content with the following concept of approximate minimum: Given $\varepsilon > 0$ a point x_ε is called an ε-*approximate minimizer* of $f(x)$ on X if

$$\inf_{x \in X} f(x) \leq f(x_\varepsilon) \leq \inf_{x \in X} f(x) + \varepsilon. \tag{3.2}$$

An ε-approximate minimizer on X always exists if $f(x)$ is bounded below on X. For differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$ a minimizer $\bar{x}$ of $f(x)$ on $\mathbb{R}^n$ satisfies $\nabla f(\bar{x}) = 0$. Is there some similar condition for an ε-approximate minimizer? The next propositions give an answer to this question.

Theorem 3.1 (Ekeland Variational Principle (Ekeland 1974)) *Let $f : X \to \mathbb{R} \cup \{+\infty\}$ be a function on a closed set $X \subset \mathbb{R}^n$ which is finite at at least one point, l.s.c. and bounded below on X. Let $\varepsilon > 0$ be given. Then for every ε-approximate minimizer x_ε of f and every $\lambda > 0$ there exists x^* satisfying*

H. Tuy, *Convex Analysis and Global Optimization*, Springer Optimization and Its Applications 110, DOI 10.1007/978-3-319-31484-6_3

$$f(x^*) \leq f(x_\varepsilon); \qquad \|x^* - x_\varepsilon\| \leq \lambda; \tag{3.3}$$

$$f(x^*) \leq f(x) + \tfrac{\varepsilon}{\lambda}\|x - x^*\| \ \forall x \in X. \tag{3.4}$$

Proof We follow the elegant proof in Hiriart-Urruty (1983). Clearly the function $g(x) = f(x) + \frac{\varepsilon}{\lambda}\|x - x_\varepsilon\|$ is l.s.c. and coercive on X, so it has a minimizer x^* on X, i.e., a point $x^* \in X$ such that

$$f(x^*) + \frac{\varepsilon}{\lambda}\|x^* - x_\varepsilon\| \leq f(x) + \frac{\varepsilon}{\lambda}\|x - x_\varepsilon\| \quad \forall x \in X. \tag{3.5}$$

Letting $x = x_\varepsilon$ yields

$$f(x^*) + \frac{\varepsilon}{\lambda}\|x^* - x_\varepsilon\| \leq f(x_\varepsilon), \tag{3.6}$$

hence $f(x^*) \leq f(x_\varepsilon)$. Further, $\inf_{x\in X} f(x) + \frac{\varepsilon}{\lambda}\|x^* - x_\varepsilon\| \leq f(x^*) + \frac{\varepsilon}{\lambda}\|x^* - x_\varepsilon\| \leq f(x_\varepsilon) \leq \inf_{x\in X} f(x) + \varepsilon$ by the definition of x_ε, hence $\|x^* - x_\varepsilon\| \leq \lambda$. Finally, from (3.5) it follows that

$$\begin{aligned} f(x^*) &\leq f(x) + \frac{\varepsilon}{\lambda}\{\|x - x_\varepsilon\| - \|x^* - x_\varepsilon\|\} \\ &\leq f(x) + \frac{\varepsilon}{\lambda}\|x - x^*\| \quad \forall x \in X. \end{aligned}$$

□

Note that by (3.3) x^* is also an ε-approximate minimizer which is not too distant from x_ε; moreover by (3.4) x^* is an *exact* minimizer of the function $f(x) + \frac{\varepsilon}{\lambda}\|x - x^*\|$ which differs from $f(x)$ just by a 'perturbation' quantity $\frac{\varepsilon}{\lambda}\|x - x^*\|$. The larger λ the smaller this perturbation quantity and the closer x^* to an exact minimizer, but x^* may be farther from x_ε. In practice, the choice of λ depends on the circumstances. Usually the value $\lambda = \sqrt{\varepsilon}$ is suitable.

Corollary 3.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function which is bounded below on $\mathbb{R}^n$. For every ε-approximate minimizer x_ε of f there exists an ε-approximate minimizer x^* such that*

$$\|x^* - x_\varepsilon\| \leq \sqrt{\varepsilon}, \qquad \|\nabla f(x^*)\| \leq \sqrt{\varepsilon},$$

Proof From (3.4) for $x = x^* + tu, u \in \mathbb{R}^n, t < 0$ and $\lambda = \sqrt{\varepsilon}$ we have $f(x^*) \leq f(x^* + tu) + \sqrt{\varepsilon}\|tu\|$. Therefore, $\frac{f(x^*+tu)-f(x^*)}{t} \leq \sqrt{\varepsilon}\|u\|$, and letting $t \to +\infty$ yields $\langle \nabla f(x^*), u\rangle \leq \sqrt{\varepsilon}\|u\| \ \forall u \in \mathbb{R}^n$, hence $\|\nabla f(x^*)\| \leq \sqrt{\varepsilon}$. □

Let X, Y be two sets in $\mathbb{R}^n, \mathbb{R}^m$, respectively. A *set-valued map* f from X to Y is a map which associates with each point $x \in X$ a set $f(x) \subset Y$ called the *value* of f at x. Theorem 3.1 implies the following fixed point theorem due to Caristi (1976):

Theorem 3.2 *Let P be a set-valued map from a closed set $X \subset \mathbb{R}^n$ into itself and let $f : X \to [0, +\infty]$ be a l.s.c. function finite at at least one point. If*

$$(\forall x \in X)(\exists y \in P(x)) \qquad \|x - y\| \le f(x) - f(y) \tag{3.7}$$

then P has a fixed point, i.e., a point $x^* \in P(x^*)$.

Proof Let x^* be the point that exists according to Theorem 3.1 for $\varepsilon = 1 < \lambda$. We have $f(x^*) \le f(x) + \frac{\|x - x^*\|}{\lambda}$ $\forall x \in X$, so $f(x^*) < f(x) + \|x - x^*\|$ $\forall x \in X \setminus \{x^*\}$. On the other hand, by hypothesis there exists $y \in P(x^*)$ satisfying $\|x^* - y\| \le f(x^*) - f(y)$, i.e., $f(x^*) \ge f(y) + \|y - x^*\|$. This implies $y = x^*$, so $x^* \in P(x^*)$. $\square$

It can be shown that Theorem 3.2 in turn implies Theorem 3.1 (Exercise 1), so these two theorems are in fact equivalent.

Heuristically, a set-valued map P with values $P(x) \neq \emptyset$ can be conceived of as a dynamic system whose potential function value at position x is $f(x)$. The assumption (3.7) means that when the system moves from position x to position y its potential diminishes at least by a quantity equal to the distance from x to y. The fixed point x^* corresponds to an equilibrium—a position from where no move to a position with lower potential is possible.

3.2 Contraction Principle

For any set $A \subset \mathbb{R}^n$ and $\delta > 0$ define $d(x, A) := \inf_{y \in A} \|x - y\|, A_\delta := \{x \in \mathbb{R}^n |\ d(x, A) \le \delta\}$. If A, B are two subsets of $\mathbb{R}^n$ the number

$$d(A, B) := \inf\{\delta > 0 |\ A \subset B_\delta, B \subset A_\delta\}$$

is called the *Hausdorff distance* between A and B. From this definition it readily follows that: if $d(A, B) \le \delta$ then $d(x, B) \le \delta$ $\forall x \in A$ (because $A \subset B_\delta$) and $d(y, A) \le \delta$ $\forall y \in B$ (because $B \subset A_\delta$).

A set-valued map P from a set $X \subset \mathbb{R}^n$ to $\mathbb{R}^n$ is said to be a *contraction* if there exists $\theta,\ 0 < \theta < 1$, such that

$$(\forall x, x' \in X) \qquad d(P(x), P(x')) \le \theta \|x - x'\|. \tag{3.8}$$

Lemma 3.1 *If a set-valued map P from a closed set $X \subset \mathbb{R}^n$ to X is a contraction then the function $f(x) = d(x, P(x))$ is l.s.c. on X.*

Proof We have to show that if $d(x^k, P(x^k)) \le \alpha,\ \|x^k - x_0| \to 0$ then $d(x^0, P(x^0)) \le \alpha$. By definition of $d(x^k, P(x^k))$ we can select $y^k \in P(x^k)$ such that $\|x^k - y^k\| \le d(x^k, P(x^k)) + 1/k$. Then $d(y^k, P(x^0)) \le d(P(x^k), P(x^0)) \le \theta \|x^k - x^0\|$ and therefore

$$d(x^0, P(x^0)) \le \|x^0 - x^k\| + \|x^k - y^k\| + d(y^k, P(x^0))$$

$$\leq \|x^0 - x^k\| + d(x^k, P(x^k)) + 1/k + \theta\|x^k - x^0\|$$
$$\leq \|x^0 - x^k\| + \alpha + 1/k + \theta\|x^k - x^0\| \to \alpha$$

as $k \to +\infty$. Consequently, $d(x^0, P(x^0)) \leq \alpha$. □

The next theorem extends the well-known Banach contraction principle to set-valued maps.

Theorem 3.3 (Nadler 1969) *Let P be a set-valued map from a closed set $X \subset \mathbb{R}^n$ into itself, such that $P(x)$ is a nonempty closed set for every $x \in X$. If there exists a number $\theta, 0 < \theta < 1$, satisfying (3.8) then for every $a \in X$ and every $\alpha \in (\theta, 1)$ there exists a point $x^* \in P(x^*)$ such that $\|x^* - a\| \leq \frac{\|a-P(a)\|}{\alpha-\theta}$.*

Proof By Lemma 3.1 the function $f(x) = d(x, P(x))$ is l.s.c., so the set $E := \{x \in X|\ (\alpha - \theta)\|x - a\| \leq f(a) - f(x)\}$ is closed and nonempty (as $a \in E$). We show that

$$(\forall x \in E)(\exists y \in P(x) \cap E) \qquad (\alpha - \theta)\|x - y\| \leq f(x) - f(y). \tag{3.9}$$

Let $x \in X$. Since $\alpha < 1$ by definition of $d(x, P(x))$ there exists $y \in P(x)$ satisfying

$$\alpha\|x - y\| \leq d(x, P(x)). \tag{3.10}$$

On the other hand, since $y \in P(x)$ it follows from (3.8) that

$$d(y, P(y)) \leq d(P(x), P(y)) \leq \theta\|x - y\|. \tag{3.11}$$

Adding (3.10) and (3.11) yields

$$\alpha\|x - y\| + d(y, P(y)) \leq d(x, P(x)) + \theta\|x - y\|,$$

hence $(\alpha - \theta)\|x - y\| \leq f(x) - f(y)$, proving (3.9).

Finally, since $x \in E$ means $(\alpha - \theta)\|x - a\| \leq f(a) - f(x)$, we have

$$(\alpha - \theta)\|y - a\| \leq (\alpha - \theta)(\|x - y\| + \|x - a\|)$$
$$\leq f(x) - f(y) + f(a) - f(x) = f(a) - f(y).$$

So $y \in E$ and all the conditions in Caristi Theorem 3.2 are satisfied for the set E, the map $P(x) \cap E$ and the function $f(x) = d(x, P(x))$. Therefore, there exists $x^* \in P(x^*) \cap E$,, i.e., $x^* \in P(x^*)$ and $\|x^* - a\| \leq \frac{f(a)-f(x^*)}{\alpha-\theta} = \frac{f(a)}{\alpha-\theta}$ by noting that $f(x^*) = 0$ as $x^* \in P(x^*)$. □

3.3 Fixed Point and Equilibrium

Solving an equation $f(x) = 0$ amounts to finding a fixed point of the mapping $F(x) = x - f(x)$. This explains the important role of fixed point theorems in analysis and optimization. In the preceding section we have already encountered two fixed point theorems: Caristi theorem which is equivalent to Ekeland variational principle and Nadler contraction theorem which is an extension of Banach fixed point principle. In this section we will study topological fixed points and their connection with the theory of equilibrium.

The foundation of topological fixed point theorems lies in a famous proposition due to the three Polish mathematicians B. Knaster, C. Kuratowski, and S. Mazurkiewicz (Knaster et al. 1929) and for this reason called the KKM Lemma.

3.3.1 The KKM Lemma

Let $S := [a^0, a^1, \ldots, a^n]$ be an n-simplex spanned by the vertices $a^0, a^1, \ldots, a^n$. For every set $I \subset \{0, 1, \ldots, n\}$ the set $\mathrm{conv}\{a^i |\ i \in I\}$ is a *face* of S, more precisely the face spanned by the vertices $a^i, i \in I$, or the face opposite the vertices $a^j, j \notin I$. A face opposite a vertex a^i is also called a *facet*. For example, $[a^1, \ldots, a^n]$ is the facet opposite a^0. A *simplicial subdivision* of S is a decomposition of it into a finite number of n-simplices such that any two simplices either do not intersect or intersect along a common face.

A *simplicial subdivision* (briefly, a subdivision) of S can be constructed inductively as follows.

For a 1-simplex $[a^0, a^1]$ the subdivision is formed by two simplices $[a^0, \frac{1}{2}(a^0 + a^1)]$ and $[\frac{1}{2}(a^0 + a^1), a^1]$. Suppose subdivisions have been defined for k-simplices with $k < n$. The subdivision of the n-simplex $[a^0, a^1, \ldots, a^n]$ consists of all n-simplices spanned each by a set of the form $\sigma \cup \{c\}$, where σ is the vertex set of an $(n-1)$-simplex in the subdivision of a facet of S and $c = \frac{1}{n+1}(a^0 + a^1 + \ldots + a^n)$. It can easily be checked that the required condition of a subdivision is satisfied, namely: any two simplices in the subdivision either do not intersect or intersect along a common face.

The just constructed subdivision is referred to as subdivision of degree 1. By applying the subdivision operation to each simplex in a subdivision of degree 1 a more refined subdivision is obtained which is referred to as subdivision of degree 2, and so on. By continuing the process, subdivisions of larger and larger degrees are obtained. It is not hard to see that the maximal diameter of simplices in a subdivision of degree k tends to zero as $k \to +\infty$.

Consider now an n-simplex $S = [a^0, a^1, \ldots, a^n]$ and a subdivision of S consisting of m simplices

$$S_1,\ S_2,\ \ldots,\ S_m. \tag{3.12}$$

To avoid confusion we will refer to the simplices in (3.12) as subsimplices in the subdivision. Let σ_k be the vertex set of S_k. The set $V = \cup_{k=1}^m \sigma_k$ is actually the set of subdivision points of the subdivision. A map $\ell : V \to \{0, 1, 2, \ldots, n\}$ is called a *labeling* of the subdivision and a subsimplex S_k is said to be *complete* if $\ell(\sigma_k) = \{0, 1, 2, \ldots, n\}$.

Lemma 3.2 *Let $\ell : V \to \{0, 1, 2, \ldots, n\}$ be a labeling such that if $z \in V$ belongs to a facet $[a^{i_0}, a^{i_1}, \ldots, a^{i_k}]$ of S then $\ell(z) \in \{i_0, i_1, \ldots, i_k\}$. Then there exists an odd number of complete subsimplices.*

Proof The proposition is trivial if S is 0-dimensional, i.e., if S is a single point. Arguing by induction, suppose the proposition is true for any $(n-1)$-dimensional S and check it for an n-dimensional $S = [a^0, a^1, \ldots, a^n]$. Consider any subsimplex S_k. A facet of S_k will be called a fair facet if its vertex set U satisfies $\ell(U) = \{1, \ldots, n\}$. Let ρ_k denote the number of fair facets of the subsimplex S_k. It can easily be seen that if S_k is complete then $\rho_k = 1$, otherwise $\rho_k = 2$. Therefore the number of complete subsimplices will be odd if so is the number $\rho = \sum_{k=1}^m \rho_k$. Let ρ' be the number of fair facets of subsimplices that lie entirely in the facet $[a^1, \ldots, a^n]$ of S. Clearly each of these fair facets is in fact a complete subsimplex in the subdivision of the $(n-1)$-simplex $[a^1, \ldots, a^n]$ (with respect to the labeling of $[a^1, \ldots, a^n]$ induced by ℓ). By the induction hypothesis ρ' is an odd number. The number of remaining fair facets of subsimplices (3.12) is then $\rho - \rho'$. But each of these fair facets is shared by two subsimplices in (3.12), so it appears in two subsimplices in (3.12) and is counted twice in the sum ρ. Therefore, $\rho - \rho'$ is even, and hence, ρ is odd.

□

The above proposition, due to Sperner (1928), is often referred to as Sperner Lemma. We are now in a position to state and prove the KKM Lemma.

Proposition 3.1 (KKM Lemma) *Let $\{a^0, a^1, \ldots, a^n\}$ be any finite subset of $\mathbb{R}^m$. If $L_0, L_1, \ldots, L_n$ are closed subsets of $\mathbb{R}^m$ such that for every set $I \subset \{0, 1, \ldots, n\}$*

$$\text{conv}\{a^i, i \in I\} \subset \cup_{i \in I} L_i, \tag{3.13}$$

then $\cap_{i=0}^n L_i \neq \emptyset$.

Proof First assume, as in the classical formulation of this proposition, that $a^0, a^1, \ldots, a^n$ are affinely independent. The set $S = [a^0, a^1, \ldots, a^n]$ is then an n-simplex and for each $I \subset \{0, 1, \ldots, n\}$ the set $\text{conv}\{a^i, i \in I\}$ is a face of S.

Consider a simplicial subdivision of degree k of S. Let x be a subdivision point, i.e., a vertex of some subsimplex in the subdivision, and $\text{conv}\{a^i |\ i \in I\}$ be the smallest face of S containing x. By assumption $x \in \cup_{i \in I} L_i$, so there exists $i \in I$ such that $x \in L_i$. We then set $\ell(x) = i$. Clearly this labeling ℓ of the subdivision points satisfies the condition of Lemma 3.2. Therefore, there exists a subsimplex $[x^{0,k}, x^{1,k}, \ldots, x^{n,k}]$ with $\ell(x^{i,k}) = i, i = 0, 1, \ldots, n$. By compactness of S there exists

$x^* \in S$ such that, up to a subsequence, $x^{0,k} \to x^*$. As $k \to +\infty$, the diameter of the simplex $[x^{0,k}, x^{1,k}, \dots, x^{n,k}]$ tends to zero, hence $\|x^{0,k} - x^{i,k}\| \to 0$, i.e.,

$$x^{i,k} \to x^*, \quad i = 0, 1, \dots, n.$$

Since $x^{i,k} \in L_i \ \forall k$ and the set L_i is closed, we conclude $x^* \in L_i, \ i = 0, 1, \dots, n$.

Turning to the general case when $a^0, a^1, \dots, a^n$ may not be affinely independent let $K := \text{conv}\{a^0, a^1, \dots, a^n\}$ and $S := [e^0, e^1, \dots, e^n]$ where e^i is the i-th unit vector of $\mathbb{R}^n$, i.e., the vector such that $e^i_i = 1, e^i_j = 0 \ \forall j \neq i$. Noting that for every $x \in S$ we have $x = \sum_{i=0}^n x_i$ with $x_i \geq 0, \sum_{i=0}^n x_i = 1$, we can define the continuous map $\phi : S \to K$ by $\phi(x) = \sum_{i=0}^n x_i a^i$. For each i let $M_i := \{x \in S | \ \phi(x) \in L_i \cap K\}$. By continuity of ϕ each M_i is a closed set and from the assumption (3.13) it can easily be deduced that for every $I \subset \{0, 1, \dots, n\}, \ \text{conv}\{e^i, i \in I\} \subset \cup_{i \in I} M_i$. Since $e^0, e^1, \dots, e^n$ are affinely independent, by the above there exists $\bar{x} \in M_i \ \forall i = 0, 1, \dots, n$, hence $\phi(\bar{x}) \in L_i \ \forall i = 0, 1, \dots, n$. □

Remark 3.1 To get insight into condition (3.13) consider the case when $a^i = e^i, i = 0, 1, \dots, n$. Imagine that a sum of value 1 is to be distributed among $n + 1$ persons $i = 0, 1, \dots, n$. A sharing program in which person i receives a share equal to x_i can be represented by a point $x = (x_0, x_1, \dots, x_n)$ of the simplex $S = [e^0, e^1, \dots, e^n]$. It may happen that for each person not every sharing program is acceptable. Let L_i denote the set of sharing programs x acceptable for person i (each L_i may naturally be assumed to be a closed set). Then condition (3.13), i.e., $\text{conv}\{e^i, i \in I\} \subset \cup_{i \in I} L_i$ simply states that for any set $I \subset \{0, 1, \dots, n + 1\}$ every sharing program in which only people in the group I have positive shares is acceptable for at least someone of them. The KKM Lemma says that under these natural assumptions there exists a sharing program acceptable for everybody. Thus a very simple common sense underlies this proposition whose proof, as we saw, is quite sophisticated.

3.3.2 General Equilibrium Theorem

Given a nonempty subset $C \subset \mathbb{R}^n$, a nonempty subset $D \subset \mathbb{R}^m$ and a real-valued function $F(x, y) : C \times D \to \mathbb{R}$, the *general equilibrium problem* is to find a point $\bar{x}$ such that

$$\bar{x} \in C, \ F(\bar{x}, y) \geq 0 \quad \forall y \in D.$$

Such a point $\bar{x}$ is referred to as an *equilibrium point* (shortly, equilibrium) for the system (C, D, F).

In the particular case when $C = D$ is a nonempty closed convex set this problem was first introduced by Blum and Oettli (1994) and since then has been extensively studied.

The KKM Lemma can be viewed as a proposition asserting the existence of equilibrium in a basic particular case. In fact given a finite set $\Omega \subset \mathbb{R}^m$, and for each $y \in \Omega$ a closed set $L(y) \subset \mathbb{R}^m$, the KKM Lemma gives conditions for the existence of an $\overline{x} \in \cap_{y\in\Omega} L(y)$. If we define $F(x,y) = \chi_y(x) - 1$ where $\chi_y(x)$ is the characteristic function of the closed set $L(y)$ (so that $F(x,y) = 0$ for $x \in L(y)$ and $F(x,y) = -1$ otherwise), then $\overline{x} \in \cap_{y\in\Omega} L(y)$ means $\overline{x} \in \Omega,\ F(\overline{x},y) \geq 0\ \forall y \in \Omega$. So $\overline{x}$ is an equilibrium point for the system (Ω,Ω,F).

Based on the KKM Lemma we now prove a general existence condition for equilibrium in a system (C,D,F). Recall that given a set $X \subset \mathbb{R}^n$ and a set $Y \subset \mathbb{R}^m$, a *set-valued map* f from X to Y is a map which associates with each point $x \in X$ a set $f(x) \subset Y$, called the value of f at x. A set-valued map f from X to Y is said to be *closed* if its graph

$$\{(x,y)|\ x \in X, y \in f(x)\}$$

is closed in $\mathbb{R}^n \times \mathbb{R}^m$, i.e., if $x^\nu \in X, y^\nu \in f(x^\nu)$ and $x^\nu \to x, y^\nu \to y$ always imply $y \in f(x)$.

Theorem 3.4 (General Equilibrium Theorem) *Let C be a nonempty compact subset of $\mathbb{R}^n$, D a nonempty closed convex subset of $\mathbb{R}^m$, and $F(x,y) : C \times D \to \mathbb{R}$ a function which is upper semi-continuous in x and quasiconvex in y. If there exists a closed set-valued map φ from D to C with nonempty convex compact values such that $\inf_{y\in D} \inf_{x\in\varphi(y)} F(x,y) \geq 0$ then the system (C,D,F) has an equilibrium point, i.e., there exists an $\overline{x}$ such that*

$$\overline{x} \in C,\quad F(\overline{x},y) \geq 0\ \forall y \in D. \tag{3.14}$$

Proof For every $y \in D$ define $C(y) := \{x \in C|\ F(x,y) \geq 0\}$. By assumption $F(x,y) \geq 0\ \forall x \in \varphi(y), \forall y \in D$, so $\varphi(y) \subset C(y)\ \forall y \in D$. Since $\varphi(y) \neq \emptyset$, this implies $C(y) \neq \emptyset$; furthermore, $C(y)$ is compact because C is compact and $F(.,y)$ is u.s.c. So to prove (3.14), i.e., that $\cap_{y\in D} C(y) \neq \emptyset$, it will suffice to show that the family $\{C(y), y \in D\}$ has the finite intersection property, i.e., for any finite set $\{a^1,\dots,a^k\} \subset D$, we have

$$\cap_{i=1}^k C(a^i) \neq \emptyset. \tag{3.15}$$

Let $\Omega := \{a^i, i = 1,\dots,k\}$ and for each $i = 1,\dots,k$ define $L(a^i) := \{y \in \mathrm{conv}\Omega|\ \varphi(y) \cap C(a^i) \neq \emptyset\}$ (note that $\mathrm{conv}\Omega \subset D$ because D is convex). It is easily seen that $L(a^i)$ is a closed subset of D. Indeed, let $\{y^\nu\} \subset L(a^i)$ be a sequence such that $y^\nu \to y$. Then $y^\nu \in \mathrm{conv}\Omega$, so $y \in \mathrm{conv}\Omega$ because the set $\mathrm{conv}\Omega$ is closed. Further, for each ν there is $x^\nu \in \varphi(y^\nu) \cap C(a^i)$. Since $C(a^i)$ is compact we have, up to a subsequence, $x^\nu \to x$ for some $x \in C(a^i)$ and since the map φ is closed, $x \in \varphi(y)$. Consequently, $x \in \varphi(y) \cap C(a^i)$, hence $\varphi(y) \cap C(a^i) \neq \emptyset$, i.e., $y \in L(a^i)$.

For any nonempty set $I \subset \{1,\dots,k\}$, let $D_I := \{y \in \mathrm{conv}\Omega |\ C(y) \subset \cup_{i\in I} C(a^i)\} = \{y \in \mathrm{conv}\Omega |\ F(x,y) < 0\ \forall x \notin \cup_{i\in I} C(a^i)\}$. In view of the quasiconvexity of $F(x,.)$ the set D_I is convex and since $a^i \in D_I\ \forall i \in I$ it follows that

$$\mathrm{conv}\{a^i, i \in I\} \subset D_I \subset \{y \in \mathrm{conv}\Omega |\ \varphi(y) \subset \cup_{i\in I} C(a^i)\}. \tag{3.16}$$

Therefore,

$$\mathrm{conv}\{a^i, i \in I\} \subset \cup_{i\in I}\{y \in \mathrm{conv}\Omega |\ \varphi(y) \cap C(a^i) \neq \emptyset\} \subset \cup_{i\in I} L(a^i). \tag{3.17}$$

Since this holds for every set $I \subset \{1,\dots,k\}$, by KKM Lemma, we have $\cap_{i=1}^k L(a^i) \neq \emptyset$. So there exists $\bar{y} \in \mathrm{conv}\Omega$ such that $\varphi(\bar{y}) \cap C(a^i) \neq \emptyset \quad \forall i = 1,\dots,k$.

Now for each $i = 1,\dots,k$ take an $x^i \in \varphi(\bar{y}) \cap C(a^i)$. It is not hard to see that for every set $I \subset \{1,\dots,k\}$, we have

$$M := \mathrm{conv}\{x^i, i \in I\} \subset \cup_{i\in I} C(a^i). \tag{3.18}$$

In fact, assume the contrary, that $S := \{x \in M |\ x \notin \cup_{i\in I} C(a^i)\} \neq \emptyset$. Since $\bar{y} \in \mathrm{conv}\Omega$, by (3.16) $\varphi(\bar{y}) \subset \cup_{i=1}^k C(a^i)$, and since $\varphi(\bar{y})$ is convex it follows that

$$M \subset \varphi(\bar{y}) \subset \cup_{i=1}^k C(a^i). \tag{3.19}$$

Therefore, $S = \{x \in M |\ x \in \cup_{i=1}^k C(a^i) \setminus \cup_{i\in I} C(a^i)\} = \{x \in M |\ x \in \cup_{i\notin I} C(a^i)\} = M \cap (\cup_{i\notin I} C(a^i))$. Since each set $C(a^i), i = 1,\dots,k$, is compact, so is the set $\cup_{i\notin I} C(a^i)$, hence S is a closed subset of M. On the other hand, $M \setminus S = \{x \in M |\ x \in \cup_{i\in I} C(a^i)\} = M \cap (\cup_{i\in I} C(a^i))$, so $M \setminus S$, too, is a closed subset of M. But $M \setminus S \neq \emptyset$ because $x^i \in C(a^i)\ \forall i \in I$. So M is the union of two disjoint nonempty closed sets S and $M \setminus S$, conflicting with the convexity of M. Therefore $S = \emptyset$, proving (3.18).

Since (3.18) holds for every set $I \subset \{1,\dots,k\}$, again by KKM Lemma (this time with $\Omega = \{x^1,\dots,x^k\}, L(x^i) = C(a^i)$), we have $\cap_{i=1}^k C(a^i) \neq \emptyset$, i.e., (3.15) and hence, $\cap_{y\in D} C(a) \neq \emptyset$, as was to be proved. □

Remark 3.2 (i) Theorem 3.4 still holds if instead of the compactness of C we only assume the existence of a $\bar{y} \in D$ such that the set $\{x \in C |\ F(x,\bar{y}) \geq 0\}$ is compact (or, equivalently, such that $F(x,\bar{y}) \to -\infty$ as $x \in C, \|x\| \to +\infty$). Indeed, for any finite set $N \subset D$ we can write $\cap_{y\in N\cup\{\bar{y}\}} C(y) = \cap_{y\in D} C'(y)$, where $C'(y) = C(y) \cap C(\bar{y})$, so the family $\{C'(y), y \in D\}$ has the finite intersection property; since $C'(y)$ is compact it follows that $\cap_{y\in D} C'(y) \neq \emptyset$, hence $\cap_{y\in D} C(y) \neq \emptyset$ and the above proof goes through.

(ii) By replacing $F(x,y)$ with $F(x,y) - \alpha$, where α is any real number Theorem 3.4 can also be reformulated as follows:

Let $C \subset \mathbb{R}^n$ be a nonempty compact set, $D \subset \mathbb{R}^m$ a nonempty closed convex set, and $F(x, y) : C \times D \to \mathbb{R}$ a function upper semi-continuous in x and quasiconvex in y; let φ be a closed set-valued map from D to C with nonempty convex compact values. Then for any real number α we have the implication

$$\inf_{y \in D} \inf_{x \in \varphi(y)} F(x, y) \geq \alpha \Rightarrow \exists \overline{x} \in C \ \inf_{y \in D} F(\overline{x}, y) \geq \alpha. \tag{3.20}$$

Or equivalently,

$$\max_{x \in C} \inf_{y \in D} F(x, y) \geq \inf_{y \in D} \inf_{x \in \varphi(y)} F(x, y). \tag{3.21}$$

Remark 3.3 For a heuristic interpretation of Theorem 3.4, consider a company (a person, a community) with an utility function $f(x, y)$ that depends upon a variable $x \in C$ under its control and a variable $y \in D$ outside its control. For each given $y \in D$, the best for the company is certainly to select an $x^y \in C$ maximizing $F(x^y, y)$ over C, i.e., such that $f(x^y, y) - f(x, y) \geq 0 \ \forall x \in C$. Theorem 3.4 tells under which conditions there exists an $\overline{x} \in C$ such that $f(\overline{x}, y) - f(x, y) \geq 0 \ \forall x \in C \ \forall y \in D$, i.e., an $\overline{x} = x^y$ for any whatever $y \in D$.

As it turns out, such kind of situation underlies the equilibrium mechanism in most cases of interest. For instance, a weaker variant of Theorem 3.4 with both C, D compact and $F(x, y)$ continuous on $C \times D$ was shown many years ago to be a generalization of the Walras Excess Demand Theorem (Tuy 1976). Below we show how various important existence propositions pertaining to equilibrium can be obtained as immediate consequences of Theorem 3.4.

3.3.3 Equivalent Minimax Theorem

First an alternative equivalent form of Theorem 3.4 is the following minimax theorem:

Theorem 3.5 *Let C be a nonempty compact subset of $\mathbb{R}^n$, D a nonempty closed convex subset of $\mathbb{R}^m$, and $F(x, y) : C \times D \to \mathbb{R}$ a function which is u.s.c. in x and quasiconvex in y. If there exists a closed set-valued map φ from D to C with nonempty compact convex values such that for every $\alpha < \gamma := \inf_{y \in D} \sup_{x \in C} F(x, y)$ we have $\varphi(y) \subset C_\alpha(y) := \{x \in C | \ F(x, y) \geq \alpha\}$ then the following minimax equality holds:*

$$\max_{x \in C} \inf_{y \in D} F(x, y) = \inf_{y \in D} \sup_{x \in C} F(x, y). \tag{3.22}$$

Proof Clearly, for every $\alpha < \gamma$ the assumption $\varphi(y) \subset C_\alpha(y) \ \forall y \in D$ means that $\inf_{y \in D} \inf_{x \in \varphi(y)} F(x, y) \geq \alpha$ while by Theorem 3.4 the latter inequality implies $\max_{x \in C} \inf_{y \in D} F(x, y) \geq \alpha$. Therefore

$$\max_{x \in C} \inf_{y \in D} F(x, y) \geq \gamma := \inf_{y \in D} \sup_{x \in C} F(x, y),$$

whence (3.22) because the converse inequality is trivial. Thus Theorem 3.4 implies Theorem 3.5. The converse can be proved in the same manner. □

Corollary 3.2 *Let C be a nonempty compact convex set in $\mathbb{R}^n$, D a nonempty closed convex set in $\mathbb{R}^m$, and $F(x, y) : C \times D \to \mathbb{R}$ a u.s.c. function which is quasiconcave in x and quasiconvex in y. Then there holds the minimax equality (3.22).*

Proof For every $\alpha < \gamma$ define $\varphi(y) := \{x \in C |\ F(x, y) \geq \alpha\}$. Then $\varphi(y)$ is a nonempty closed convex set. If $(x^\nu, y^\nu) \in C \times D$, $(x^\nu, y^\nu) \to (\overline{x}, \overline{y})$, and $x^\nu \in \varphi(y^\nu)$, i.e., $F(x^\nu, y^\nu) \geq \alpha$ then by upper semi-continuity of $F(x, y)$, we must have $F(\overline{x}, \overline{y}) \geq \alpha$, i.e., $\overline{x} \in \varphi(\overline{y})$. So the set-valued map φ is closed and $\varphi(y) = C_\alpha(y)\ \forall y \in D$. The conclusion follows by Theorem 3.5. □

Note that Corollary 3.2 differs from Sion's minimax theorem (Theorem 2.7) only by the continuity properties imposed on $F(x, y)$. Actually, with a little bit more skill Sion's theorem can also be derived directly from the General Equilibrium Theorem (Exercise 6).

3.3.4 Ky Fan Inequality

When $D = C$ and $\varphi(y)$ is single-valued Theorem 3.4 yields

Theorem 3.6 (Ky Fan 1972) *Let C be a nonempty compact convex set in $\mathbb{R}^n$, and $F(x, y) : C \times C \to \mathbb{R}$ a function which is u.s.c in x and quasiconvex in y. If $\varphi(y)$ is a continuous map from C into C there exists $\overline{x} \in C$ such that*

$$\inf_{y \in C} F(\overline{x}, y) \geq \inf_{y \in C} F(\varphi(y), y). \tag{3.23}$$

In the special case $\varphi(y) = y\ \forall y \in C$ this inequality becomes

$$\inf_{y \in C} F(\overline{x}, y) \geq \inf_{y \in C} F(y, y).$$

If, in addition, $F(y, y) = 0\ \forall y \in C$ then

$$\inf_{y \in C} F(\overline{x}, y) \geq 0.$$

Proof The inequality (3.21) reduces to (3.23) when $\varphi(y)$ is a singleton. □

In the special case $F(x, y) = \|x - y\|$ we have from (3.23) $\min_{y \in C} \|x - y\| \geq \min_{y \in C} \|\varphi(y) - y\|$, hence $0 = \min_{y \in C} \|\varphi(y) - y\|$, which means there exists $\overline{y} \in C$ such that $\overline{y} = \varphi(\overline{y})$. Thus, as a consequence of Ky Fan inequality we get *Brouwer fixed point theorem*: a continuous map φ from a compact convex set C into itself has a fixed point. Actually, the following more general fixed point theorem is also a direct consequence of Theorem 3.4:

3.3.5 Kakutani Fixed Point

Theorem 3.7 (Kakutani 1941) *Let C be a compact convex subset of $\mathbb{R}^n$ and φ a closed set-valued map from C to C with nonempty convex values $\varphi(x) \subset C\ \forall x \in C$. Then φ has a fixed point, i.e., a point $\overline{x} \in \varphi(\overline{x})$.*

Proof The function $F(x, y) : C \times C \to \mathbb{R}$ defined by $F(x, y) = \|x - y\|$ is continuous on $C \times C$ and convex in y for every fixed x. By Theorem 3.4 (Remark 3.2, (3.21)) there exists $\overline{x} \in C$ such that

$$\inf_{y \in C} \|\overline{x} - y\| \geq \inf_{y \in C} \inf_{x \in \varphi(y)} \|x - y\|.$$

Since $\inf_{y \in C} \|\overline{x} - y\| = 0$, for every natural k we have $\inf_{y \in C} \inf_{x \in \varphi(y)} \|x - y\| \leq 1/k$, hence there exist $y^k \in C, x^k \in \varphi(y^k) \subset C$ such that $\|y^k - x^k\| \leq 1/k$. By compactness of C there exist $\overline{x}, \overline{y}$ such that, up to a subsequence, $x^k \to \overline{x} \in C,\ y^k \to \overline{x} \in C$. Since the map φ is closed it follows that $\overline{x} \in \varphi(\overline{x})$. □

Historically, Brouwer Theorem was established in 1912. Kakutani Theorem was derived from Brouwer Theorem almost three decades later, via a quite involved machinery. The above proof of Kakutani Theorem is perhaps the shortest one.

3.3.6 Equilibrium in Non-cooperative Games

Consider an n-person game in which the player i has a strategy set $X_i \subset \mathbb{R}^{m_i}$. When the player i chooses a strategy $x_i \in X_i$, the situation of the game is described by the vector $x = (x_1, \ldots, x_n) \in \prod_{i=1}^n X_i$. In that situation the player i obtains a payoff $f_i(x)$.

Assume that each player does not know which strategy is taken by the other players. A vector $\overline{x} \in X := \prod_{i=1}^n X_i$ is called a *Nash equilibrium* if for every $i = 1, \ldots, n$:

$$f_i(\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_n) = \max_{x_i \in X_i} f_i(\overline{x}_1, \ldots, \overline{x}_{i-1}, x_i, \overline{x}_{i+1}, \ldots, \overline{x}_n).$$

In other words, $\overline{x} \in X$ is a Nash equilibrium if

$$f_i(\overline{x}) = \max_{x_i \in X_i} f_i(\overline{x}|_i x_i),$$

where for any $z \in X_i$, $\overline{x}|_i z$ denotes the vector $\overline{x} \in \prod_{i=1}^n X_i$ in which $\overline{x}_i$ has been replaced by z.

Theorem 3.8 (Nash Equilibrium Theorem (Nash 1950)) *Assume that for each $i = 1, \dots, n$ the set X_i is convex compact and the function $f_i(x)$ is continuous on X and concave in x_i. Then there exists a Nash equilibrium.*

Proof The set $X := \prod_{i=1}^n X_i$ is convex, compact as the Cartesian product of n convex compact sets. Consider the function

$$F(x, y) := \sum_{i=1}^{n} [f_i(x) - f_i(x|_i y_i)], \quad x \in X, \ y \in X. \tag{3.24}$$

This is a function continuous in x and convex in y. The latter is due to the assumption that for each $i = 1, \dots, n$ the function $f_i(x)$ is concave in x_i, which means that the function $z \mapsto f_i(x|_i z)$ is concave. Since $y|_i y_i = y$ we also have $F(y, y) = 0 \ \forall y \in X$. Therefore, by Ky Fan inequality Theorem 3.6, there exists $\overline{x} \in X$ satisfying

$$F(\overline{x}, y) = \sum_{i=1}^{n} [f_i(\overline{x}) - f_i(\overline{x}|_i y_i)] \geq 0 \quad \forall y \in X.$$

Fix an arbitrary i and take $y \in X$ such that $y_j = \overline{x}_j$ for $j \neq i$. Writing the above inequality as

$$f_i(\overline{x}) - f_i(\overline{x}|_i y_i) + \sum_{j \neq i} [f_j(\overline{x}) - f_j(\overline{x}|_j \overline{x}_j)] \geq 0 \quad \forall y \in X,$$

and noting that $\overline{x}|_j \overline{x}_j = \overline{x}$ we deduce the inequality $f_i(\overline{x}) \geq f_i(\overline{x}|y_i) \ \forall y_i \in X_i$. Since this holds for every $i = 1, \dots, n$, $\overline{x}$ is a Nash equilibrium. □

3.3.7 Variational Inequalities

As we saw in Chap. 2, Proposition 2.31, a minimizer $\overline{x}$ of a convex function $f : \mathbb{R}^n \to \mathbb{R}$ over a convex set $C \subset \mathbb{R}^n$ must satisfy the condition

$$0 \in \partial f(\overline{x}) + N_C(\overline{x}), \tag{3.25}$$

where $\partial f(\overline{x})$ is the subdifferential of f at $\overline{x}$ and $N_C(\overline{x}) = \{y \in \mathbb{R}^n | \ \langle y, x - \overline{x} \rangle \leq 0 \ \forall x \in C\}$ is the outward normal cone to C at $\overline{x}$. The above condition can be equivalently written as

$$\exists \overline{y} \in \partial f(\overline{x}) \quad \langle \overline{y}, x - \overline{x} \rangle \geq 0 \ \forall x \in C.$$

A relation like (3.25) is called a *variational inequality*. Replacing the map $x \mapsto \partial f(x)$ with a given arbitrary set-valued map Ω from a nonempty closed convex set

C to $\mathbb{R}^n$, we consider the *variational inequality* for (C, Ω):

$$(VI(C,\Omega)) \qquad \text{Find } \overline{x} \in C \text{ satisfying } 0 \in \Omega(\overline{x}) + N_C(\overline{x}).$$

Alternatively, following Robinson (1979) we can consider the *generalized equation*

$$0 \in \Omega(x) + N_C(x). \tag{3.26}$$

Although only a condensed form of writing $VI(C, \Omega)$, the generalized equation (3.26) is often a more efficient tool for studying many questions of nonlinear analysis and optimization.

Based on Theorem 3.4 we can now derive a solvability condition for the variational inequality $VI(C, \Omega)$ or the generalized equation (3.26). For that we need a property of set-valued maps.

A set-valued map $f(x)$ from a set $X \subset \mathbb{R}^n$ to a set $Y \subset \mathbb{R}^m$ is said to be *upper semi-continuous* if for every $x \in X$ and every open set $U \supset f(x)$ there exists a neighborhood W of x such that $f(z) \subset U \ \forall z \in W \cap X$.

Proposition 3.2 (i) *An upper semi-continuous set-valued map f from a set X to a set Y which has closed values is a closed map. Conversely, a closed set-valued map f from X to a compact set Y is upper semi-continuous.*
(ii) *If $f(x)$ is an upper semi-continuous map from a compact set X to Y with nonempty compact values then $f(X)$ is compact.*

Proof (i) Let f be an upper semi-continuous map with closed values. If $(x^\nu, y^\nu) \in X \times Y$ is a sequence such that $(x^\nu, y^\nu) \to (x, y) \in X \times Y$, then for any neighborhood U of $f(x)$ we will have $y^\nu \in U$ for all large enough ν, so $y \in f(x)$. Conversely, let f be a closed set-valued map from X to a compact set Y. If f is not upper semi-continuous. There would exist $x^0 \in X$ and an open set $U \supset f(x^0)$ such that for every k there exists $x^k, y^k \in f(x^k)$ satisfying $\|x^k - x^0\| \leq 1/k, y^k \notin U$. Since Y is compact, by taking a subsequence if necessary, $y^k \to y \in Y$ and since $f(x^0) \subset U$, clearly $y \notin f(x^0)$, conflicting with the map being closed.
(ii) Let $W_t, t \in T$, be an open covering of $f(X)$. For each fixed $x \in X$ since $f(x)$ is compact there exists a finite covering $W_i, i \in I(x) \subset T$. In view of the upper semi-continuity of the map f there exists an open ball $V(y)$ around x such that $f(z) \subset \cup_{i \in I(x)} W_i \ \forall z \in V(x)$. Then by compactness of Y there exists a finite set $E \subset Y$ such that Y is entirely covered by $\cup_{x \in E} V(x)$. So the finite family $W_t, t \in \cup\{W_i, i \in I(x), x \in E\}$ covers $f(X)$, proving the compactness of this set. □

Theorem 3.9 *Let C be a compact convex set in $\mathbb{R}^n$, Ω an upper semi-continuous set-valued map from C to $\mathbb{R}^n$ with nonempty compact convex values. Then the variational inequality $VI(C, \Omega)$ [the generalized equation (3.26)] has a solution, i.e., there exists $\overline{x} \in C$ and $\overline{y} \in \Omega(\overline{x})$ such that*

$$\langle \overline{y}, x - \overline{x} \rangle \geq 0 \quad \forall x \in C.$$

Proof By Proposition 3.2 Ω is a closed map and its range $D = \Omega(C)$ is compact. Define $F(x, y) = \langle x, y\rangle - \min_{z\in C}\langle z, y\rangle$. This function is continuous in y and convex in x. By Theorems3.4, (3.21), there exists $\overline{y} \in D$ such that

$$\inf_{x\in C} F(x, \overline{y}) \geq \inf_{x\in C} \inf_{y\in\Omega(x)} F(x, y).$$

But $\inf_{x\in C} F(x, \overline{y}) = \min_{x\in C}\langle x, \overline{y}\rangle - \min_{z\in C}\langle z, \overline{y}\rangle = 0$, hence

$$\inf_{x\in C} \inf_{y\in\Omega(x)} F(x, y) \leq 0.$$

Since C is compact, for every $k = 1, 2, \ldots,$ there exists $x^k \in C$ satisfying

$$\inf_{y\in\Omega(x^k)} F(x^k, y) \leq 1/k,$$

hence there exists $x^k \in C,\ y^k \in \Omega(x^k)$ such that $F(x^k, y^k) \leq 1/k$, i.e.,

$$\langle x^k, y^k\rangle - \min_{z\in C}\langle z, y^k\rangle \leq 1/k.$$

Let $(\overline{x}, \overline{y})$ be a cluster point of (x^k, y^k). Then $\overline{x} \in C,\ \overline{y} \in \Omega(\overline{x})$ because the map Ω is closed, and $\langle \overline{x}, \overline{y}\rangle - \min_{z\in C}\langle z, \overline{y}\rangle \leq 0$, i.e., $\langle \overline{y}, x - \overline{x}\rangle \geq 0\ \forall x \in C$ as to be proved. □

3.4 Exercises

1 Derive Ekeland variational principle from Theorem 3.2. (Hint: Let x_ε be an ε-approximate minimizer of $f(x)$ on X and $\lambda > 0$. Define $P(x) = \{y \in X|\ \frac{\varepsilon}{\lambda}\|y - x_\varepsilon)\| \leq f(x_\varepsilon) - f(y)\}$ and show that if Ekeland's theorem is false there exists for every $x \in X$ an $y \in P(x)$ satisfying $\frac{\varepsilon}{\lambda}\|x - y\| < f(x) - f(y)$. Then apply Caristi theorem to derive a contradiction.)

2 Let S be an n-simplex spanned by $a^0, a^1, \ldots, a^n$. For each $i = 0, 1, \ldots, n$ let F_i be the facet of S opposite the vertex a^i (i.e., the $(n-1)$-simplex spanned by n points $a^j, j \neq i$), and L_i a given closed subset of S. Show that the KKM Lemma is equivalent to saying that if the following condition is satisfied:

$$S \subset \cup_{i=0}^n L_i, \quad F_i \subset L_i,\ \forall i,$$

then $\cap_{i=0}^n L_i \neq \emptyset$.

3 Let $S, F_i, L_i, i = 0, 1, \ldots, n$ be as in the previous exercise. Show that the KKM Lemma is equivalent to saying that if the following condition is satisfied:

$$S \subset \cup_{i=0}^n L_i, \quad a^i \in L_i,\ F_i \cap L_i = \emptyset,$$

then $\cap_{i=0}^n L_i \neq \emptyset$.

4 Derive the KKM Lemma from Brouwer fixed point theorem. (Hint: Let $S, L_0, L_1, \ldots, L_n$ be as in the previous exercise, $\rho(x, L_i) = \min\{\|x - y\|| \; y \in L_i\}$. For $k > 0$ set $H_i = \{x \in S| \; \rho(x, L_i) \geq 1/k$ and define a continuous map $f : S \to S$ by $f_i(x) = \frac{x_{i-1}\rho(x,H_i)}{\sum_{i=1}^n x_i \rho(x,H_i)}$ with the convention $x_{-1} = x_n$. Show that if $\overline{x} = f(\overline{x})$ then $f_i(\overline{x}) > 0 \; \forall i$, hence $\rho(\overline{x}, H_i) > 0$, i.e., $\rho(\overline{x}, L_i) \leq 1/k$ for every $i = 0, 1, \ldots, n$. Since k can be arbitrary, deduce that $\cap_{i=0}^n L_i \neq \emptyset$.)

5 Let $C_1, \ldots, C_n$ be n closed convex sets in $\mathbb{R}^k$ such that their union is a convex set C. Show that if the intersection of $n-1$ arbitrary sets among them is nonempty then the intersection of the n sets is nonempty. (Hint: Let $a^i \in L_i := \cap_{j \neq i} C_j$ and apply KKM to the set $\{a^1, \ldots, a^n\}$ and the collection of closed sets $\{L_1, \ldots, L_n\}$.)

6 Let C be a compact convex subset of $\mathbb{R}^n$, D a closed convex subset of $\mathbb{R}^m$, $f(x, y) : C \times D \to \mathbb{R}$ a function which is quasi convex, l.s.c. in x and quasiconcave u.s.c. in y. Show that for every $\alpha > \eta := \sup_{y \in D} \inf_{x \in C} f(x, y)$ the set-valued map φ from D to C such that $\varphi(y) := \{x \in C| \; f(x, y) \leq \alpha\} \; \forall y \in D$ is closed and using this fact deduce Sion's Theorem from Theorem 3.4. (Hint: show that the set-valued map φ is u.s.c., hence closed because C is compact.)

Chapter 4
DC Functions and DC Sets

4.1 DC Functions

Convexity is a nice property of functions but, unfortunately, it is not preserved under even such simple algebraic operations as scalar multiplication or lower envelope. In this chapter we introduce the *dc structure* (also called the *complementary convex structure*) which is the common underlying mathematical structure of virtually all nonconvex optimization problems.

Let Ω be a convex set in $\mathbb{R}^n$. We say that a function is *dc on* Ω if it can be expressed as the difference of two convex functions on Ω, i.e., if $f(x) = f_1(x) - f_2(x)$, where f_1, f_2 are convex functions on Ω.

Of course, convex and concave functions are particular examples of dc functions. A quadratic function $f(x) = \langle x, Qx\rangle$, where Q is a symmetric matrix, is a dc function which may be neither convex nor concave. Indeed, setting $x = Uy$ where $U = [u^1, \ldots, u^n]$ is the matrix of normalized eigenvectors of Q, we have $U^TQU = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, hence $f(x) = f(Uy) = \langle Uy, QUy\rangle = \langle y, U^TQUy\rangle$, so that $f(x) = f_1(x) - f_2(x)$ with

$$f_1(x) = \sum_{\lambda_i \geq 0} \lambda_i y_i^2, \quad f_2(x) = -\sum_{\lambda_i < 0} \lambda_i y_i^2, \quad y = U^{-1}x.$$

Actually most functions encountered in practice are dc, as will be shortly demonstrated.

Functions which can be represented as differences of convex functions were considered some 40 years ago by Alexandrov (1949, 1950) and Landis (1951), and some time later by Hartman (1959) who proved a number of important properties to be discussed below. For our purpose, the first fundamental property of these functions is their stability relative to operations frequently used in optimization theory.

H. Tuy, *Convex Analysis and Global Optimization*, Springer Optimization and Its Applications 110, DOI 10.1007/978-3-319-31484-6_4

Proposition 4.1 *If $f_i(x)$, $i = 1, \ldots, m$, are dc functions on Ω then the following functions are also dc on Ω:*

(i) $\sum_{i=1}^{m} \alpha_i f_i(x)$, *for any real numbers* α_i;
(ii) $g(x) = \max\{f_1(x), \ldots, f_m(x)\}$;
(iii) $h(x) = \min\{f_1(x), \ldots, f_m(x)\}$.

Proof We only prove (ii) because (i) is trivial and the proof of (iii) is similar. Let $f_i(x) = p_i(x) - q_i(x)$ with p_i, q_i convex on Ω. From the obvious equality $f_i = p_i + \sum_{j \neq i} q_j - \sum_j q_j$ it follows that $\max\{f_1, \ldots, f_m\} = \max_i [p_i + \sum_{j \neq i} q_j] - \sum_j q_j$ which is a difference of two convex functions since the upper envelope and the sum of finitely many convex functions are convex. □

Thus, the set of dc functions on Ω forms the smallest linear space containing all convex functions on Ω. This linear space, denoted by $DC(\Omega)$, is stable under the operations of upper and lower envelope of finitely many functions.

An inequality of the form $f(x) \leq 0$, where the function $f(x)$ is convex, is called a convex inequality (because the set of all x satisfying this inequality is a convex set). If $f(x)$ is concave, then the inequality is called *complementary convex* or *reverse convex* because its solution set is the complement of a convex set. Thus, a reverse convex inequality is also an inequality of the form $f(x) \geq 0$, where $f(x)$ is convex. If $f(x)$ is a dc function then the inequality $f(x) \leq 0$ is called a *dc inequality*. Of course the inequality $f(x) \geq 0$ is then also a dc inequality because $-f(x)$ is still dc.

An important consequence of Proposition 4.1 is that any finite system of dc inequalities, whether conjunctive or disjunctive, is equivalent to a single dc inequality. Indeed, a conjunctive system

$$g_i(x) \leq 0, \quad i = 1, \ldots, m$$

can be written as

$$g(x) := \max\{g_1(x), \ldots, g_m(x)\} \leq 0 \tag{4.1}$$

while a disjunctive system

$$g_i(x) \leq 0 \quad \text{for at least one } i = 1, \ldots, m$$

is equivalent to

$$g(x) := \min\{g_1(x), \ldots, g_m(x)\} \leq 0. \tag{4.2}$$

Furthermore, if $g(x) = p(x) - q(x)$ with $p(x), q(x)$ convex, then, introducing an additional variable t, the dc inequality (4.1) [or (4.2)] can be split into two inequalities:

$$p(x) - t \leq 0, \quad t - q(x) \leq 0, \tag{4.3}$$

where the first is a convex inequality, and the second is a *reverse convex inequality*. Thus, any finite system of dc inequalities can be rewritten as a system of one convex inequality and one reverse convex inequality. As will be seen later, this property allows a compact description of a wide class of nonconvex optimization problems and provides the basis for a unified theory of global optimization.

4.2 Universality of DC Functions

The next property shows that virtually all the most frequently encountered functions in practice are dc. As usual, $C^k(\mathbb{R}^n)$ (or simply C^k) denotes the class of functions on $\mathbb{R}^n$ continuously differentiable up to order k.

Proposition 4.2 *Every function $f \in C^2(\mathbb{R}^n)$ is dc on any compact convex set $\Omega \subset \mathbb{R}^n$.*

Proof We show that, given any compact convex set $\Omega \subset \mathbb{R}^n$, the function $g(x) = f(x) + \rho\|x\|^2$ becomes convex on Ω when ρ is sufficiently large (then $f(x) = g(x) - \rho\|x\|^2$ yields a dc representation of f). Indeed, since $\langle u, \nabla^2 g(x)u\rangle = \langle u, \nabla^2 f(x)u\rangle + \rho\|u\|^2$, if ρ is so large that

$$-\min\{\langle u, \nabla^2 f(x)u\rangle \mid x \in \Omega, \|u\| = 1\} \le \rho$$

then $\langle u, \nabla^2 g(x)u\rangle \ge 0,\ \forall u$, hence $g(x)$ is convex by Proposition 2.2. □

Corollary 4.1 *For any continuous function $f(x)$ on a compact convex set Ω and for any $\varepsilon > 0$ there exists a dc function $g(x)$, such that*

$$\max_{x\in\Omega} |f(x) - g(x)| \le \varepsilon.$$

Proof By Weierstrass theorem there exists a polynomial $g(x)$ satisfying the required condition, and obviously $g(x) \in C^2$. □

Thus $DC(\Omega)$ is dense in $C(\Omega)$, the Banach space of continuous functions on Ω, equipped with the sup norm. Though dc functions occur frequently in practice, they often appear in a hidden, not directly recognizable, form. To help to identify dc functions in various situations we prove some further properties of these functions.

A function $f : D \to \mathbb{R}$ defined on a convex set $D \subset \mathbb{R}^n$ is said to be *locally dc* if for every $x \in D$ there exist a convex open neighborhood U of x and a pair of convex functions g, h on U such that $f|U = g|U - h|U$. Note that since U can be assumed bounded, the sets $\cup\{\partial g(y) \mid y \in U\}$ and $\cup\{\partial h(y) \mid y \in U\}$ are bounded by Theorem 2.6, so by Corollary 2.10, the functions g, h can be assumed convex finite on all of $\mathbb{R}^n$.

Proposition 4.3 *A locally dc function on a convex, open, or closed, set D is dc on D.*

Proof We shall restrict ourselves to the case when D is a compact convex set (a proof for the general case can be found in Hartman (1959); see also Ellaia (1984)). From the hypothesis and the compactness of D one can find a finite set $\{x_1, \ldots, x_k\} \subset D$ together with open convex neighborhoods $U_1, \ldots, U_k$ of these points covering D, and convex functions $h_i : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, \ldots, k)$ such that $(f + h_i)|U_i$ is convex. Let $h = \sum_{i=1}^k h_i$ and consider the function $g = f + h$. For each i, $(f + h_i)|U_i$ is convex, hence $g|U_i = (f + h_i)|U_i + (\sum_{j \neq i} h_j)|U_i$ is convex. By Proposition 2.4, g is convex on D, i.e., $f = g - h$ with g, h convex. $\square$

A mapping $F = (f_1, f_2, \ldots, f_m) : \Omega \to \mathbb{R}^m$ defined on a convex set $\Omega \subset \mathbb{R}^n$ is said to be dc on Ω: $F \in DC(\Omega)$, if $f_i \in DC(\Omega)$ for every $i = 1, \ldots, m$.

Proposition 4.4 *Let $\Omega_1 \subset \mathbb{R}^n, \Omega_2 \subset \mathbb{R}^m$ be convex sets such that Ω_1 is open or closed, Ω_2 is open. If $F_1 : \Omega_1 \to \Omega_2, F_2 : \Omega_2 \to \mathbb{R}^k$ are dc mappings then $F_2 \circ F_1 : \Omega_1 \to \mathbb{R}^k$ is also a dc mapping.*

Proof It suffices to show that if $F = (f_1, \ldots, f_m) : \Omega_1 \to \Omega_2$ is dc and $g : \Omega_2 \to \mathbb{R}$ is convex then $g(f_1, \ldots, f_m) \in DC(\Omega_1)$. Let $x \in \Omega_1$ and $y = F(x) \in \Omega_2$. The convex function $g(y)$ can be represented in a neighborhood U_2 of y as pointwise supremum of a family of affine functions: $g(y) = \sup_t \ell_t$ where $\ell_t = a_{0t} + a_{1t}y_1 + \cdots + a_{mt}y_m$ $(y_1, \ldots, y_m$ are coordinates of y in $\mathbb{R}^m)$ and $M = \sup_{i,t} |a_{it}| < +\infty$ (cf. Corollary 2.9). Let $f_i(x) = f_i^+(x) - f_i^-(x)$ in a neighborhood U_1 of x such that $F(U_1) \subset U_2$. Then

$$
\begin{aligned}
&\ell_t(f_1, \ldots, f_m) \\
&\quad = a_{0t} + \sum_{i=1}^m a_{it} f_i^+ - \sum_{i=1}^m a_{it} f_i^- \\
&\quad = \left[a_{0t} + \sum_{i=1}^m (M + a_{it}) f_i^+ + \sum_{i=1}^m (M - a_{it}) f_i^- \right] \\
&\qquad -M \sum_{i=1}^m (f_i^+ + f_i^-) = p_t - q
\end{aligned}
$$

with p_t and q convex and q independent of t. Then $g(f_1, \ldots, f_m) = \sup_t \ell_t(f_1, \ldots, f_m) = \sup_t (p_t - q) = \sup_t p_t - q = p - q$, i.e., $g(f_1, \ldots, f_m)$ is locally dc on Ω_1. Hence, by Proposition 4.3, $g \circ f \in DC(\Omega_1)$. $\square$

Corollary 4.2 *Let Ω_1, Ω_2 be as in Proposition 4.4. If $F_1 : \Omega_1 \to \Omega_2$ is dc on $\Omega_1 \subset \mathbb{R}^n$ and $F_2 : \Omega_2 \to \mathbb{R}^k$ is C^2-smooth, then $F_2 \circ F_1$ is dc on Ω_1. In particular,* *the product of two dc functions is dc; if* $f(x)$ *is dc on an open (or closed) convex set* Ω *and* $f(x) \neq 0$ *for all* $x \in \Omega$ *then* $\frac{1}{f(x)}$ *and* $|f(x)|^{1/m}$ *are dc on* Ω.

Proof Indeed, C^2-smooth functions are dc. □

The above properties explain why an overwhelming majority of functions of practical interest are dc.

Remark 4.1 Following McCormick (1972) a function $f(x)$ is called a *factorable function* if it results from a finite sequence of compositions of transformed sums and products of simple functions of one variable. In other words, a factorable function is a function which can be obtained as the last in a sequence of functions $f_1, f_2, \ldots,$ built up as follows:

$$f_i(x) = x_i \quad (i = 1, \ldots, n) \tag{4.4}$$

and for $k > n, f_k$ is one of the forms:

$$f_k(x) = f_l(x) + f_j(x) \quad \text{for some } l, j < k \tag{4.5}$$

$$f_k(x) = f_l(x) \times f_j(x) \quad \text{for some } l, j < k \tag{4.6}$$

$$f_k(x) = F(f_j(x)) \quad \text{for some } j < k, \tag{4.7}$$

where $F : \mathbb{R} \to \mathbb{R}$ is a simple dc function of one variable, such as $F(t) = t^p, F(t) = e^t, F(t) = \log|t|, F(t) = \sin t$, etc.

From the above corollary it also easily follows that a factorable function is a dc function.

Remark 4.2 When a function f is dc and $f = g - h$ is a dc representation of f, then for any convex function u, $f = (g + u) - (h + u)$ gives another dc representation of f. By taking u to be any strictly convex function, we see that a dc function is always expressible as a difference of two *strictly* convex functions.

4.3 DC Representation of Basic Composite Functions

In many questions, it is important to know how to write effectively a dc function as a difference of convex functions. This problem of (effective) *dc representation* which is far from being simple will be discussed in this and the next sections for some particular classes of functions.

To begin with, observe that many functions expressing cost, utility, attraction, repulsion, etc., in applied fields (e.g., location theory, molecular conformation, and quadratic assignment problems) are compositions of convex or concave functions with convex or concave monotone functions. These functions are dc by Proposition 4.4. Let us show how to find their dc representations (Tuy 1996). First recall from Proposition 2.8 that: if $h : M \to \mathbb{R}_+$ is a convex (concave, resp.) function on a convex subset M of $\mathbb{R}^m$ and if $q : \mathbb{R}_+ \to \mathbb{R}$ is a convex (concave, resp.) nondecreasing function, then $q(h(x))$ is a convex (concave, resp.) function on M.

Proposition 4.5 *Let $h : M \to \mathbb{R}_+$ be a convex function on a convex compact subset M of $\mathbb{R}^m$. If $q : \mathbb{R}_+ \to \mathbb{R}$ is a convex nonincreasing function such that $q'_+(0) > -\infty$, then $q(h(x))$ is a dc function on M:*

$$q(h(x)) = g(x) - Kh(x),$$

where $g(x)$ is a convex function and K is a positive constant satisfying $K \geq |q'_+(0)|$ ($q'_+(t)$ denotes the right derivative of $q(t)$ at point t).

Proof We have $q'_+(0) \leq q'_+(t) \leq 0 \quad \forall t \geq 0$, therefore $\tilde{q}(t) = q(t) + Kt$ satisfies $\tilde{q}'_+(t) = q'_+(t) + K \geq q'_+(0) + K \geq 0 \quad \forall t \geq 0$. So $\tilde{q}$ is a convex nondecreasing function and by Proposition 2.8 that has just been recalled above, $\tilde{q}(h(x))$ is a convex function on M. Since $\tilde{q}(h(x)) = q(h(x)) + Kh(x)$, the result follows. $\square$

Analogously: *Let $h : M \to \mathbb{R}_+$ be a convex function as in Proposition 4.5. If $q : \mathbb{R}_+ \to \mathbb{R}$ is a concave nondecreasing function such that $q'_+(0) < +\infty$, then $q(h(x))$ is a dc function on M:*

$$q(h(x)) = Kh(x) - g(x),$$

where $g(x)$ is a convex function and K is a positive constant satisfying $K \geq |q'_+(0)|$.

Using Proposition 4.5, one can see, for example, that the function $we^{-\theta\|x-a\|}$ (with $w > 0,\ \theta > 0$) is dc, since it is equal to $q(\|x-a\|)$ and $q(t) = we^{-\theta t}$ is a convex decreasing function with $q'_+(0) = -\theta w > -\infty$. A generalization of Proposition 4.5 is the following:

Proposition 4.6 *Let $h(x) = u(x) - v(x)$ where $u, v : M \to \mathbb{R}_+$ are convex functions on a compact convex set $M \subset \mathbb{R}^m$ such that $h(x) \geq 0\ \forall x \in M$. If $q : \mathbb{R}_+ \to \mathbb{R}$ is a convex nonincreasing function such that $q'_+(0) > -\infty$ then $q(h(x))$ is a dc function on M:*

$$q(h(x)) = g(x) - K[u(x) + v(x)],$$

where $g(x) = q(h(x)) + K[u(x) + v(x)]$ is a convex function and K is a constant satisfying $K \geq |q'_+(0)|$.

Proof By convexity of $q(t)$, for any $\theta \in \mathbb{R}_+$ we have $q(t) \geq q(\theta) + q'_+(\theta)(t-\theta)$, with equality holding for $\theta = t$. Therefore,

$$q(t) = \sup_{\theta \in R_+} \{q(\theta) + (t-\theta)q'_+(\theta)\} = \sup_{\theta \in R_+} \{q(\theta) - \theta q'_+(\theta) + tq'_+(\theta)\},$$

and consequently,

$$\begin{aligned}
&q(u(x) - v(x)) \\
&\quad = \sup_{\theta \in R_+} \{q(\theta) - \theta q'_+(\theta) + (K + q'_+(\theta))u(x) + (K - q'_+(\theta))v(x)\} \\
&\qquad -K[u(x) + v(x)] = g(x) - K[u(x) + v(x)].
\end{aligned}$$

We contend that $g(x) = \sup_{\theta \in R_+} \{q(\theta) - \theta q'_+(\theta) + (K + q'_+(\theta))u(x) + (K - q'_+(\theta))v(x)\}$ is convex. Indeed, since $q(t)$ is convex, $q'_+(\theta) \geq q'_+(0)$ and hence, $K + q'_+(\theta) \geq K + q'_+(0) \geq 0$ for all $\theta \geq 0$; furthermore, since $q(t)$ is nonincreasing, $q'_+(\theta) \leq 0$ and hence $K - q'_+(\theta) \geq K > 0$ for all $\theta \geq 0$. It follows that for each fixed $\theta \in \mathbb{R}^+$ the function $x \mapsto q(\theta) - \theta q'_+(\theta) + (K + q'_+(\theta))u(x) + (K - q'_+(\theta))v(x)$ is convex and $g(x)$, as the pointwise supremum of a family of convex functions, is itself convex. □

Analogously: *Let $h(x) = u(x) - v(x)$ where $u, v : M \to \mathbb{R}_+$ are convex functions on a compact convex set $M \subset \mathbb{R}^m$ such that $h(x) \geq 0$ $\forall x \in M$. If $q : \mathbb{R}_+ \to \mathbb{R}$ is a concave nondecreasing function such that $q'_+(0) < +\infty$, then $q(h(x))$ is a dc function on M:*

$$q(h(x)) = K[u(x) + v(x)] - g(x),$$

where $g(x) = K[u(x) + v(x)] - q(h(x))$ is a convex function and K is a constant satisfying $K \geq |q'_+(0)|$.

Proposition 4.7 *Let $h(x) = u(x) - v(x)$ where $u, v : M \to \mathbb{R}_+$ are convex functions on a compact convex set $M \subset \mathbb{R}^m$ such that $0 \leq h(x) \leq a$ $\forall x \in M$. If $q : [0, a] \to \mathbb{R}$ is a convex nondecreasing function such that $q'_-(a) < +\infty$ then $q(h(x))$ is a dc function on M:*

$$q(h(x)) = g(x) - K[a + v(x) - u(x)],$$

where $g(x) = q(h(x)) + K[a + v(x) - u(x)]$ is a convex function and K is a constant satisfying $K \geq q'_-(a)$.

Proof Define $p : [0, a] \to \mathbb{R}$ by $p(t) = q(a - t)$. Clearly, $p(t)$ is convex nonincreasing function and $q(t) = p(a - t)$ so that $q(h(x)) = p(a - h(x))$. The conclusion follows from Proposition 4.6. □

Thus, under mild conditions, a convex (or concave) monotone function of a dc function over a compact convex set is a dc function, whose dc representation can easily be obtained. In the next section we shall see that a rather wide class of functions of a real variable can be represented in the form $f(t) = p(t) - q(t)$ where $p(t), q(t)$ are convex nondecreasing functions. The above results then provide a dc representation of composite functions of the form $f(h(x))$, where $h(x)$ is a dc function on $\mathbb{R}^n$.

4.4 DC Representation of Separable Functions

Obviously, a separable function $f(x) = \sum_{i=1}^{n} f_i(x_i)$ is dc if each univariate function $f_i(t), t \in R$ is dc, so the question of dc representation for separable functions reduces to that of univariate functions. A classical result in this area is the following:

Proposition 4.8 *A function $f : [0,a] \to \mathbb{R}$ is representable in the form $f = p - q$, where p and q are convex and $p'_+(0), p'_-(a), q'_+(0), q'_-(a)$ are all finite, if and only if*

$$f(t) = f(0) + \int_0^t r(\theta)d\theta, \tag{4.8}$$

for some function of bounded variation $r : [0,a] \to \mathbb{R}$.

Proof See, e.g., Roberts and Varberg (1973). □

As is well known from real analysis, a function $r : [0,a] \to \mathbb{R}$ of bounded variation can always be decomposed into a difference of two monotone nondecreasing functions: $r(t) = \pi(t) - \rho(t)$, where $\pi(t) = V_0^t(r)$ is the total variation of r in the interval $[0,t]$. Hence the dc representation of a function of the class (4.8) is as follows:

$$f(t) = \left(f(0) + \int_0^t \pi(\theta)d\theta\right) - \int_0^t \rho(\theta)d\theta. \tag{4.9}$$

This classical representation, however, is not always convenient. Therefore, it is useful to know alternative representations for some classes of frequently encountered univariate functions.

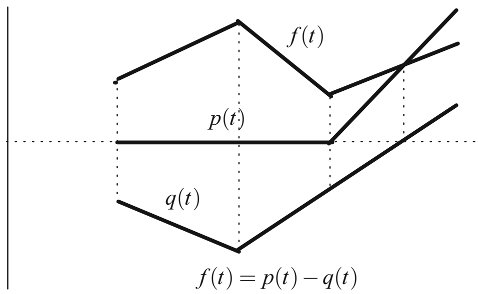
One such class includes continuous piecewise affine functions $f(t) : [0,a] \to \mathbb{R}$ (such as those encountered, for instance, in the Telepak problem (Zadeh 1974)). Let $0 < c_1 < c_2 < \ldots < c_k < a$ be the sequence of breakpoints. Since in the neighborhood of each $t \in [0,a]$, $f(t)$ is either affine (if t is not a breakpoint), or concave, or convex, it follows that $f(t)$ is locally dc, and hence dc on $[0,a]$. Denote $U_i = [c_{i-1}, c_{i+1}]$, with $c_0 = 0, c_{k+1} = a$, and let $I = \{i \mid f(t) \text{ is concave at } c_i\}$. For each $i \in I$ take the convex function $\xi_i(t) : [0,a] \to R$ that agrees with $-f(t)$ at every $t \in U_i$ and is affine on each interval $(0, c_i)$ and (c_i, a).

Proposition 4.9 *A dc representation of a piecewise affine function $f(t)$ is $f(t) = p(t) - q(t)$, where $p(t)$ is convex piecewise affine and*

$$q(t) = \sum_{i \in I} \xi_i(t).$$

Proof Clearly $q(t)$ is convex. To show that $f(t) + q(t)$ is convex, observe that for every $i \in I$:

Fig. 4.1 DC decomposition of piecewise affine functions



$$(f+q)|U_i = (f+\xi_i)|U_i + \left(\sum_{j\in I\setminus\{i\}} \xi_j\right)|U_i$$
$$= \left(\sum_{j\in I\setminus\{i\}} \xi_j\right)|U_i.$$

Since every ξ_j with $j \in I \setminus \{i\}$ is affine on U_i, it follows that $f(t) + q(t)$ is affine on every $U_i, i \in I$. On the other hand, for $i \notin I$:

$$(f+q)|U_i = \left(f + \sum_{j\in I} \xi_j\right)|U_i,$$

and since on such U_i, $f(t)$ is convex while every ξ_j with $j \in I$ is affine,it follows that $f(t)+q(t)$ is also convex on every $U_i, i \notin I$. Thus, every $t \in [0, a]$ has a neighborhood in which $f(t) + q(t)$ is convex. Hence, $f(t) + q(t)$ is convex on $[0, a]$, as was to be proved (Fig. 4.1). □

Proposition 4.9 can be extended to continuous functions $f(t) : [0, a] \to \mathbb{R}$ that are piecewise convex–concave in the following sense: there exists a sequence $e_0 := 0 < e_1 < e_2 < \ldots < e_k < a := e_{k+1}$ such that for each interval $V_i = [e_i, e_{i+1}], i = 0, \ldots, k$, the function $f_i(t) := f(t)|V_i$ is convex or concave, with finite $f'_+(e_i), f'_-(e_{i+1})$, and moreover, $f_i(t)$ is convex whenever $f_{i-1}(t)$ is concave and conversely. Denote $I = \{i \mid f_i(t) \text{ is concave}\}$. For each $i \in I$ let $\psi_i(t) : [0, a] \to \mathbb{R}$ be the continuous convex function which agrees with $-f(t)$ for $t \in V_i$ and is affine outside V_i. Then, by an analogous argument to the proof of Proposition 4.9, it is easily proved that $f(t) = p(t) - q(t)$, where

$$q(t) = \sum_{i\in I} \psi_i(t)$$

is convex on each interval $U_i = [e_{i-1}, e_{i+1}], i = 1, \ldots, k$, and hence convex on the whole segment $[0, a]$.

In particular this representation applies to *S-shaped functions*, i.e., continuous functions $f(t) : [0, a] \to \mathbb{R}$, which are convex (or concave, resp.) in a certain interval $[0, c], 0 < c < a$, and concave (convex, resp.) in the interval $[c, a]$. For example, if $f(t)$ is convex in $[0, c]$ and concave in $[c, a]$ then one can take $q(t)$ to be a convex function which agrees with $-f(t)$ for $t \geq c$, and is affine for $t < c$.

4.5 DC Representation of Polynomials

Since a polynomial has continuous derivatives of any order, it follows from Proposition 4.2 that:

Proposition 4.10 *Any polynomial in $x \in \mathbb{R}^n$ is a dc function on $\mathbb{R}^n$.* □

A nontrivial problem, however, is how to represent effectively a polynomial as a difference of two convex polynomials. For quadratic functions (polynomials of degree 2), the answer is provided by the following well-known result:

Proposition 4.11 *Let $f(x) = \langle x, Qx\rangle$ be an indefinite quadratic form, where Q is a symmetric indefinite matrix with spectral radius $\rho(Q)$. Then for any $\lambda \geq \rho(Q)$ the function $g(x) = f(x) + \lambda\|x\|^2$ is convex.*

Proof Since the matrix $Q + \lambda I$ is obviously positive semidefinite, the function $g(x) = \langle x, (Q + \lambda I)x\rangle$ is convex. □

It follows that any indefinite quadratic function $f(x)$ can be represented as $f(x) = g(x) - \lambda\|x\|^2$, where $g(x) = f(x) + \lambda\|x\|^2$ is convex. If we know the matrix U of normalized eigenvectors of Q, so that $U^T QU = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, then setting $x = Uy, F(y) = f(Uy)$, we can convert $f(x)$ into a *separable* function:

$$F(y) = \sum_{i=1}^{n} \lambda_i y_i^2$$

which is obviously dc. A useful consequence of the above result is the following (see, e.g., Thach 1992):

Corollary 4.3 *Any system of quadratic inequalities*

$$f_i(x) := \langle x, Q^i x\rangle + \langle b^i, x\rangle + d_i \leq 0, \quad i = 1, \ldots, m,$$

where Q^i are symmetric matrices, $b^i \in \mathbb{R}^n$, $d_i \in \mathbb{R}$, is equivalent to a single dc inequality

$$g(x) - \lambda\|x\|^2 \leq 0,$$

where $g(x)$ is a convex function and $\lambda \geq \max_{i=1,\ldots,m} \rho(Q^i)$.

Proof Since the ith inequality can be rewritten as $g_i(x) - \lambda\|x\|^2 \leq 0$, with $g_i(x)$ convex, the system is equivalent to

$$\max_{i=1,\dots,m} g_i(x) \leq \lambda\|x\|^2$$

which is the desired result for $g(x) = \max_{i=1,\dots,m} g_i(x)$. $\square$

Thus, the question of dc representation of quadratic polynomials is solved rather satisfactorily. For polynomials of degrees higher than 2, the results are much less simple. It is well known that for any polynomial $P(x)$ and any compact convex set $\Omega \subset \mathbb{R}^n$ the function $P(x) + \rho\|x\|^2$ is convex on Ω if $\rho \geq -\min\{\langle u, \nabla^2 P(x)u\rangle \mid x \in \Omega, \|u\| = 1\}$. However, the estimation of such a lower bound for ρ is often computationally expensive. In practice, we are interested in the dc representation of a polynomial $P(x)$ over a set $x \geq a$, so by translating if necessary, we may consider $P(x)$ only in the orthant $x \geq 0$.

Proposition 4.12 *For any two convex functions $f_1, f_2 : \mathbb{R}^n_+ \to \mathbb{R}_+$ a dc representation of their product is the following:*

$$f_1(x)f_2(x) = \frac{1}{2}[f_1(x) + f_2(x)]^2 - \frac{1}{2}[f_1^2(x) + f_2^2(x)].$$

Proof Clearly $f_1(x) + f_2(x)$ is convex, nonnegative-valued, and since the univariate function $q(t) = t^2$ is convex increasing on $\mathbb{R}_+$, by Proposition 2.8 each term in the above difference is convex. $\square$

Using this Proposition and the fact that for every natural m the univariate function $t \to t^m$ is convex on $\mathbb{R}_+$, one can easily obtain a dc representation for any monomial of the form $x_1^{p_1}x_2^{p_2}\dots x_n^{p_n}$, and hence a dc representation for $P(x)$ on $\mathbb{R}^n_+$.

4.6 DC Sets

In this and the next sections, by *dc set* we mean a set $M \subset \mathbb{R}^n$ for which there exist convex functions $g, h : \mathbb{R}^n \to \mathbb{R}$ such that $M = \{x \mid g(x) \leq 0, h(x) \geq 0\}$ (so $M = D \setminus C$, where $D = \{x \mid g(x) \leq 0\}$ and $C = \{x \mid h(x) < 0\}$). Since $M = \{x \mid \max[g(x), -h(x)] \leq 0\}$, it follows that a dc set can also be defined by a dc inequality. Conversely, if a set S is defined by a dc inequality: $S = \{x \mid g(x) - h(x) \leq 0\}$, with $g(x)$ and $h(x)$ convex, then clearly $S = \{x \mid (x, t) \in M \text{ for some } t\}$, where $M = \{(x, t) \in \mathbb{R}^n \times \mathbb{R} \mid g(x) - t \leq 0, t - h(x) \leq 0\}$; therefore, S can be obtained as the projection on $\mathbb{R}^n$ of the dc set $M \subset \mathbb{R}^{n+1}$.

Surprisingly, dc sets are not so different from arbitrary closed sets. This can be seen from the observation by Asplund (1973) that for any closed set $S \subset \mathbb{R}^n$, if $d(x, S) = \inf\{\|x - y\| \mid y \in S\}$ denotes the distance from x to S then the function $x \mapsto \|x\|^2 - d^2(x, S) = \sup\{2xy - \|y\|^2 \mid y \in S\}$ is convex. Thus, $S = \{x \mid d^2(x, S) \leq 0\}$ where $d^2(x, S)$ is a dc function.

For the applications, however, a drawback of the function $d^2(x,S)$ is that it is often too difficult to compute; furthermore, $d^2(x,S) = 0$ for any $x \in S$, so the representation $S = \{x|\ d^2(x,S) = 0\}$ does not make any difference between interior and boundary points of S. The following results by Thach (1993a) attempt to remove this drawback in a way to allow computational developments (see Part II, Sect. 7.5).

Let $h : \mathbb{R}^n \to \mathbb{R}$ be a *strictly convex* function (cf. Sect. 2.11), i.e., a function such that $h((1-\lambda)x^1 + \lambda x^2) < (1-\lambda)h(x^1) + \lambda h(x^2)$ whenever $x^1 \neq x^2, 0 < \lambda < 1$ (for instance, one can take $h(x) = \|x\|^2$). Given a nonempty closed set S in $\mathbb{R}^n$, define

$$d^2(x) = \inf_{y \in S, p \in \partial h(x)} [h(y) - h(x) - \langle p, y - x \rangle]. \tag{4.10}$$

Lemma 4.1 *We have*

(i) $d(x) = 0 \quad \forall x \in S;$
(ii) $d(x) > 0 \quad \forall x \notin S;$
(iii) *If* $x^k \to x$ *and* $d(x^k) \to 0$ *as* $k \to \infty$ *then* $x \in S$.

Proof (i) is obvious because $h(y) - h(x) - \langle p, y - x \rangle \geq 0 \quad \forall y, \forall p \in \partial h(x)$.

(ii) follows from (iii). Indeed, if $x \notin S$ and $d(x) = 0$ then by taking $x^k = x \quad \forall k$, we see that $d(x^k) \to 0$, hence by (iii), $x \in S$, a contradiction. Thus, we need only prove (iii). Let $x^k \to x$ and $d(x^k) \to 0$, so that there exist y^k and $p^k \in \partial h(x^k)$ satisfying

$$h(y^k) - h(x^k) - \langle p^k, y^k - x^k \rangle \to 0.$$

By a boundedness property of subdifferentials of convex functions (Theorem 2.6), we may assume, by taking a subsequence if necessary, that $p^k \to p \in \partial h(x)$. Then

$$h(y^k) - h(x) - \langle p, y^k - x \rangle \to 0 \ (k \to \infty). \tag{4.11}$$

Clearly the function $y \mapsto \tilde{h}(y) = h(y) - h(x) - \langle p, y - x \rangle$ satisfies $\tilde{h}(x) = 0 \leq \tilde{h}(y) \quad \forall y$. It follows that $\tilde{h}(y) > 0 \quad \forall y \neq x$ for, by the strict convexity of $\tilde{h}$, if $\tilde{h}(y) = 0$ for some $y \neq x$ then $\tilde{h}(\frac{x+y}{2}) < 0$, a contradiction. Thus the level set $C_0 = \{y \mid \ \tilde{h}(y) \leq 0\} = \{x\}$ is bounded, and consequently, so is the level set $C_1 = \{y \mid \ \tilde{h}(y) \leq 1\}$. But in view of (4.11) we may assume $\tilde{h}(y^k) \leq 1 \ \forall k$. Therefore, the sequence $\{y^k\}$ is bounded. For any cluster point $\overline{y}$ of this sequence we have $\tilde{h}(\overline{y}) = 0$, hence $\overline{y} = x$ and since S is closed and $y^k \in S$, it follows that $x = \overline{y} \in S$. $\square$

Now let θ be any positive number and $r : \mathbb{R}^n \to \mathbb{R}_+$ any function such that:

$$r(x) = 0 \quad \forall x \in S; \tag{4.12}$$

$$0 < r(y) \leq \min\{\theta, d(y)\} \quad \forall y \notin S. \tag{4.13}$$

Define

$$g_S(x) = \sup_{y \notin S, p \in \partial h(y)} \{h(y) + \langle p, x - y \rangle + r^2(y)\}. \tag{4.14}$$

Proposition 4.13 *Let* $h : \mathbb{R}^n \to \mathbb{R}$ *be any given strictly convex function. For every closed set* $S \subset \mathbb{R}^n$ *the function* $g_S(x)$ *defined by (4.14) is closed, convex, finite everywhere and satisfies*

$$S = \{x \in \mathbb{R}^n \mid \; g_S(x) - h(x) \leq 0\}. \tag{4.15}$$

Proof We will assume that S is neither empty nor the whole space (otherwise the proposition is trivial). The function $g_S(x)$ is closed, convex as the pointwise supremum of a family of affine functions $x \mapsto h(y) + \langle p, x - y \rangle + r^2(y)$. It is finite everywhere because for every $x \in \mathbb{R}^n$:

$$g_S(x) \leq \sup_{y \notin S, p \in \partial h(y)} \{h(y) + \langle p, x - y \rangle + \theta^2\} \leq h(x) + \theta^2 < +\infty.$$

If $y \notin S$ then, since $r(y) > 0$, it follows that

$$g_S(y) \geq h(y) + \langle p, y - y \rangle + r^2(y) > h(y).$$

On the other hand, if $x \in S$ then for all $y \notin S, \; p \in \partial h(y)$, we have

$$h(y) + \langle p, x - y \rangle + r^2(y) \leq h(y) + \langle p, x - y \rangle + h(x) - h(y) - \langle p, x - y \rangle \leq h(x),$$

hence, $g_S(x) \leq h(x) \; \forall x \in S$. □

Remark 4.3 If we take $h(x) = \|x\|^2$ then $\partial h(y) = \{2y\}$, hence $d(y) = \inf\{\|x - y\| \mid x \in S\}$ is the distance from y to S. Following Eaves and Zangwill (1971), a function $r : \mathbb{R}^n \to \mathbb{R}_+$ satisfying (4.12) and (4.13) can be called a *separator* for the set S. Thus, given any closed set $S \subset \mathbb{R}^n$ and a separator $r(y)$ for S we can describe S as the solution set of the dc inequality $g_S(x) - \|x\|^2 \leq 0$, where

$$g_S(x) = \sup_{y \notin S} \{r^2(y) + 2xy - \|y\|^2\}.$$

Later, in Sect. 7.5, we will see how this representation can be used for solving continuous optimization problems.

Corollary 4.4 (Thach 1993a) *Any closed set in* $\mathbb{R}^n$ *is the projection on* $\mathbb{R}^n$ *of a dc set in* $\mathbb{R}^{n+1}$ (Fig. 4.2).

Proof From (4.15) $S = \{x \mid \; \exists (x, t) \in \mathbb{R}^{n+1}, \; g_S(x) \leq t, t \leq h(x)\}$, so S is the projection of $D \setminus C \subset \mathbb{R}^{n+1}$ on $\mathbb{R}^n$, for $D = \{(x, t) \mid g_S(x) \leq t\}, C = \{(x, t) \mid h(x) < t\}$. □

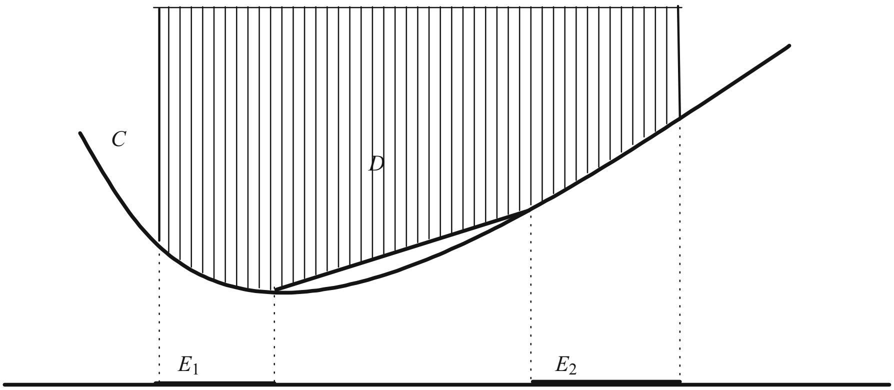


Fig. 4.2 $S := E_1 \cup E_2 = \mathrm{pr}(D \setminus \mathrm{int}C)$ with $D =$ shaded area

Corollary 4.5 *For any lower semi-continuous function $f(x)$ there exists a convex function $g(x)$ such that the inequality $f(x) \le 0$ is equivalent to the dc inequality*

$$g(x) - \|x\|^2 \le 0.$$

Proof Indeed, $S = \{x | f(x) \le 0\}$ is a closed set. □

A fundamental property of convex sets which is the cornerstone of the whole convex duality theory is the possibility of defining a nonempty closed convex proper subset of $\mathbb{R}^n$ as the intersection of a family of halfspaces (Theorem 1.6).

An analogous property holds, with, however, reverse convex sets replacing halfspaces, for an arbitrary nonempty closed proper subset of $\mathbb{R}^n$.

A set M is called *reverse convex* or *complementary convex* if it is defined by a reverse convex inequality, i.e., $M = \{x \mid \ g(x) \ge 0\}$, where $g(x)$ is a convex function.

Proposition 4.14 *Let $h : \mathbb{R}^n \to \mathbb{R}$ be any given strictly convex function. Let S be a nonempty closed proper subset of $\mathbb{R}^n$.*

(i) *For every $y \in \mathbb{R}^n \setminus S$ there exists an affine function $l(x)$ on $\mathbb{R}^n$ such that*

$$l(y) - h(y) > 0, \qquad l(x) - h(x) \le 0 \quad \forall x \in S. \tag{4.16}$$

(ii) *There exists a family of affine functions $l_i(x), i = 1, 2, \ldots,$ such that*

$$S = \bigcap_{i=1}^{\infty} \{x \mid \ l_i(x) - h(x) \le 0\}. \tag{4.17}$$

Proof (i) Choose a function $r : \mathbb{R}^n \to \mathbb{R}_+$ satisfying (4.12), (4.13) and such that

$$\text{whenever } y^k \to y \text{ and } r(y^k) \to 0 \text{ as } k \to \infty, \text{ then } y \in S.$$

By Lemma 4.1 the function $r(y) = d(y)$ satisfies this condition. Let $p \in \partial h(y), l(x) = h(y) + \langle p, x - y\rangle + r^2(y)$. Clearly, $l(y) = h(y) + r^2(y) > h(y)$, while from (4.14): $l(x) \le g_S(x) \le h(x) \quad \forall x \in S$. We will say that the reverse convex set $D(y) = \{x \mid \ l(x) - h(x) \le 0\}$ separates S from y.

(ii) Consider a grid of points $\{y^i \mid \ i = 1, 2, \dots\} \subset \mathbb{R}^n \setminus S$ which is dense in $\mathbb{R}^n \setminus S$ (e.g., the grid of all points with rational coordinates in $\mathbb{R}^n \setminus S$). For each y^i let $p^i \in \partial h(y^i), \ \ l_i(x) = h(y^i) + \langle p^i, x - y^i\rangle + r^2(y^i)$ so that the set $D_i = \{x \mid \ l_i(x) - h(x) \le 0\}$ separates S from y^i. We contend that

$$S = \bigcap_{i=1}^{\infty} D_i.$$

Indeed, by D denote the set on the right-hand side. It is plain that $S \subset D$, so we need only to prove the converse containment. Suppose $y \in D$ but $y \notin S$. Because the grid $\{y^i \mid \ i = 1, 2, \dots\}$ is dense in $\mathbb{R}^n \setminus S$, there is a sequence in this grid (which for convenience we also denote by $y^i, i = 1, 2, \dots$) such that $y^i \to y$ as $i \to \infty$. We have $y \in D_i \quad \forall i$, i.e.,

$$h(y^i) + \langle p^i, y - y^i\rangle + r^2(y^i) \le h(y) \quad \forall i.$$

Since $y^i \to y, \ h(y^i) \to h(y)$, it follows that $r^2(y^i) \to 0$, which by Lemma 4.1 conflicts with $y \notin S$. Therefore, $D \subset S$, and hence $S = D$. □

Thus, an arbitrary closed set which is neither empty nor the whole space can be defined by means of a strictly convex function and a family of affine functions. Later (in Part II) we will see that using this property certain approximation procedures commonly used in convex optimization (such as linearization and outer approximation of convex sets by polyhedrons) can be extended to continuous nonconvex optimization.

4.7 Convex Minorants of a DC Function

A nonconvex inequality $f(x) \le 0, x \in D$, where D is a convex set in $\mathbb{R}^n$, can often be handled by replacing it with a convex inequality $c(x) \le 0$, where $c(x)$ is a convex minorant of $f(x)$ on D. The latter inequality is then called a convex relaxation of the former. Of course, the tightest relaxation is obtained when $c(x) = \mathrm{conv} f(x)$, the convex envelope, i.e., the largest convex minorant, of $f(x)$.

If $f(x)$ is a dc function for which a dc representation $f(x) = g(x) - h(x)$ is available, then a convex minorant of $f(x)$ on a polytope $D \subset \mathbb{R}^n$ with vertex set V is provided by $g(x) + \mathrm{conv}_D(-h)(x)$, i.e., according to Corollary 2.2,

$$g(x) - \sup\left\{\sum_{i=1}^{n+1} \lambda_i h(v^i) |\ v^i \in V, \lambda_i \ge 0, \sum_{i=1}^{n+1} \lambda_i = 1, \sum_{i=1}^{n+1} \lambda_i v^i = x\right\}.$$

Most often, however, $f(x)$ is given as a factorable function (see Sect. 3.3) whose explicit dc representation is not readily available and generally very hard to compute. In such cases, using the factorable structure the computation of a convex minorant (or concave majorant) of $f(x)$ can be reduced to solving the following two problems (McCormick 1982):

PROBLEM 4.1 Given a function $F(t)$ of one variable, and a function $t(x)$, defined on a convex set $S \subset \mathbb{R}^n$, find a convex minorant (concave majorant) of $F(t(x))$ on the set $\{x \in S|\ a \leq t(x) \leq b\}$.

Assume that there are available a convex minorant $p(x)$ and a concave majorant $q(x)$ of $t(x)$. Denote the convex envelope and the concave envelope of $F(t)$ on the segment $a \leq t \leq b$ by $\varphi(t)$ and $\Phi(t)$, respectively. Let $\alpha := \inf\{F(t)|a \leq t \leq b\} = F(t_{\min})$, $\beta := \sup\{F(t)|a \leq t \leq b\} = F(t_{\max})$. Also for any three real numbers u, v, w denote by $\mathrm{mid}[u, v, w]$ the middle number, i.e., the number which is between the two others.

Proposition 4.15 *The functions*

$$\varphi(\mathrm{mid}[p(x), q(x), t_{\min}]),\ \ \Phi(\mathrm{mid}[p(x), q(x), t_{\max}])$$

provide a convex minorant and a concave majorant of $F(t(x))$ *on* $\{x \in S|\ a \leq t(x) \leq b\}$.

Proof It suffices to prove for the convex minorant, because the proof for the concave majorant is analogous. Clearly the convex envelope $\varphi(t)$ of $F(t)$ can be broken up into the sum of three terms: the convex decreasing part, the convex increasing part, and the minimum value. The convex decreasing part is

$$\varphi^D(t) = \varphi(\min(t, t_{\min})) - \alpha,$$

and the convex increasing part is

$$\varphi^I(t) = \varphi(\max(t, t_{\min})) - \alpha.$$

Then $\varphi(t) = \varphi^D(t) + \varphi^I(t) + \alpha$. Observe that since $\varphi^D(t)$ is decreasing and $t(x) \leq q(x)$ we have $\varphi^D(t(x)) \geq \varphi(\min(q(x), t_{\min})) - \alpha$. Similarly, $\varphi^I(t) \geq \varphi(\max(p(x), t_{\min})) - \alpha$. Hence,

$$\begin{aligned} F(t(x)) &\geq \varphi(t(x)) \\ &= \varphi^D(t(x)) + \varphi^I(t(x)) + \alpha \\ &\geq \varphi(\min(q(x), t_{\min})) + \varphi(\max(p(x), t_{\min})) - \alpha \qquad (4.18) \\ &= \varphi(\mathrm{mid}[p(x), q(x), t_{\min}]). \end{aligned}$$

The convexity of this function follows from the expression (4.18) because $\varphi(\min\ (q(x), t_{\min}))$ is convex as a convex decreasing function of the concave function $\min(q(x), t_{\min})$ and $\varphi(\max(p(x), t_{\min}))$ is convex as a convex increasing function of the convex function $\max(p(x), t_{\min})$. □

PROBLEM 4.2 Given two functions $u(x), v(y)$ on $\mathbb{R}^n, \mathbb{R}^m$ resp., find a convex minorant (concave majorant) of the product $f(x) = u(x)v(y)$ on the set $\{(x,y)|\ p \leq u(x) \leq q, r \leq v(y) \leq s\}$.

Lemma 4.2 *The system of inequalities*

$$p \leq u \leq q \quad r \leq v \leq s, \tag{4.19}$$

where $p < q, r < s$ is equivalent to

$$\min\{su + pv - ps,\ ru + qv - qr\} \geq uv \tag{4.20}$$

$$uv \geq \max\{ru + pv - pr,\ su + qv - qs\}. \tag{4.21}$$

Proof The inequalities (4.19) imply

$$(u-p)(v-r) \geq 0, \quad (q-u)(s-v) \geq 0 \tag{4.22}$$

$$(u-p)(v-s) \leq 0, \quad (u-q)(v-r) \leq 0. \tag{4.23}$$

Conversely, from (4.22)–(4.23) it follows that $(u-p)[(v-r)-(v-s)] = (u-p)(s-r) \geq 0$, hence $p \leq u$ because $r < s$; similarly, $u \leq q, r \leq v \leq s$, i.e., (4.19). Thus, (4.19) is equivalent to (4.22)–(4.23). Expanding the latter yields precisely (4.20)–(4.21). □

Proposition 4.16 *Let $p \geq 0, r \geq 0$. If $u(x), v(y)$ are convex (concave, resp.) and finite on $\mathbb{R}^n, \mathbb{R}^m$ resp. then a convex minorant (concave majorant, resp.) of the product $u(x)v(y)$ on the set $\{(x,y)|\ p \leq u(x) \leq q,\ r \leq v(y) \leq s\}$ is given by the function*

$$\max\{ru(x) + pv(y) - pr,\ su(x) + qv(y) - qs\} \tag{4.24}$$

$$(\min\{su(x) + pv(y) - ps,\ ru(x) + qv(y) - qr\}, \text{resp.}). \tag{4.25}$$

Proof If $u(x), v(y)$ are convex, the function $\max\{ru(x) + pv(y) - pr,\ su(x) + qv(y) - qs\}$ is convex and by the above Lemma it is a minorant of $u(x)v(y)$ for $p \leq u(x) \leq q, r \leq v(y) \leq s$. Similarly, if $u(x), v(y)$ are concave, the function $\min\{su(x) + pv(y) - ps,\ ru(x) + qv(y) - qr\}$ is a concave majorant of $u(x)v(y)$. □

Proposition 4.17 *Let $p \geq 0, r \geq 0$ and suppose there exist convex functions $\varphi(x), \psi(y)$ and concave functions $\Phi(x), \Psi(y)$ such that $p \leq \varphi(x) \leq u(x) \leq \Phi(x) \leq q$ and $r \leq \psi(y) \leq v(y) \leq \Psi(y) \leq s$ for all (x,y) in a convex set*

$D \subset \mathbb{R}^n \times \mathbb{R}^m$. Then a convex minorant and a concave majorant of the product $u(x)v(y)$ on the set D are provided by the functions

$$\max\{r\varphi(x) + p\psi(y) - pr,\ s\varphi(x) + q\psi(y) - qs\},$$

$$\min\{s\Phi(x) + p\Psi(y) - ps,\ r\Phi(x) + q\Psi(y) - qr\}.$$

Proof Immediate from the above. □

Corollary 4.6 *Let $p \geq 0, r \geq 0$. The convex and concave envelopes of the function $f(x,y) = xy$ on the rectangle $\{(x,y) \in \mathbb{R}^2 | \, p \leq x \leq q, r \leq y \leq s\}$ are the functions*

$$\max\{rx + py - pr,\ sx + qy - qs\}, \tag{4.26}$$

$$\min\{sx + py - ps,\ rx + qy - qr\}. \tag{4.27}$$

A concave majorant of the function $\frac{x}{y}$ on the rectangle $\{(x,y) \in \mathbb{R}^2 | \, p \leq x \leq q, r \leq y \leq s\}$ is the function

$$\min\left\{sx + \frac{p}{y} - \frac{p}{r},\ \frac{x}{s} + \frac{q}{y} - \frac{q}{s}\right\}.$$

Proof This follows from Proposition 4.16 when $u(x) = x, v(y) = y$ and when $u(x), v(y) = \frac{1}{y}$ (note that the function $\frac{1}{y}$ is concave). □

4.8 The Minimum of a DC Function

As we saw in Sect. 2.11 any local minimum of a convex function on a convex set C is also a global minimum. This important property is no longer true for general dc functions but it partially extends to quadratic functions, i.e., to functions of the form $f(x) = \frac{1}{2}\langle x, Qx\rangle + \langle a, x\rangle + \alpha$ where Q is a symmetric $n \times n$ matrix.

Proposition 4.18 *Any local minimum (or maximum, resp.) of a quadratic function $f(x)$ on affine set $E \subset \mathbb{R}^n$ is a global minimum (maximum, resp.).*

Proof If x^0 is a local minimizer of $f(x)$ on E, there exists a ball B centered at x^0 such that $f(x^0) \leq f(x)\ \forall x \in B \cap E$. For any $z \in E \setminus \{x^0\}$ let L be the line through x^0 and z, i.e., $L = \{x \in \mathbb{R}^n | \, x = x^0 + t(z - x^0), t \in (-\infty, +\infty)\}$. Then the restriction of f on L is the function $\varphi(t) = f(x^0 + t(z - x^0)), t \in \mathbb{R}$. Since $f(x)$ is a quadratic function on $\mathbb{R}^n$, $\varphi(t)$ is a quadratic function of the real variable t. This function attains a local minimum at $t = 0$, which must then be a global minimum as the minimum of a quadratic function of a real variable. So $\varphi(0) \leq \varphi(t)\ \forall t \in R$, in particular $f(z) = \varphi(1) \geq \varphi(0) = f(x^0)$, i.e., $f(z) \geq f(x^0)\ \forall z \in E$. □

Proposition 4.19 *A quadratic function which is bounded below on $\mathbb{R}^n$ must be convex. The minimum (or maximum) of an indefinite quadratic function on a compact convex set C is always attained at a boundary point (which may not be an extreme point).*

Proof If a quadratic function $f(x)$ is not convex there exist two distinct points $a, b \in \mathbb{R}^n$ and a number $t \in (0,1)$ such that $f((1-t)a+tb) > (1-t)f(a)+tf(b)$. So the function $\varphi(t) := f((1-t)a+tb), t \in R$, which is the restriction of f on the line through a, b, is not convex. Then $\varphi(t)$ must be concave because a quadratic function on R is either convex or concave. Hence, $\varphi(t)$ is unbounded below on R, conflicting with $f(x)$ being bounded below on $\mathbb{R}^n$. This proves the first part of the proposition. Turning to the second part, without loss of generality we can assume that the affine hull of C is the whole space $\mathbb{R}^n$. Since $f(x)$ is continuous and C is compact convex, $f(x)$ attains a minimum on C. If this minimum were an interior point of C, it would be a local minimum of $f(x)$ on $\mathbb{R}^n$, and hence, by Proposition 4.18, a global minimum. So $f(x)$ would be bounded below on $\mathbb{R}^n$, and would be convex, contrary to its indefiniteness. Analogously for the maximum. $\square$

By Proposition 2.31, a point x^0 achieves the minimum of a proper convex function $f(x)$ on $\mathbb{R}^n$ if and only if $0 \in \partial f(x^0)$. For general dc functions, a criterion for the global minimum is the following (Hiriart-Urruty 1989a,b; see also Tuy and Oettli 1994):

Proposition 4.20 *Let* $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *be an arbitrary proper function and* $h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *a convex proper l.s.c. function. A point* $x^0 \in \mathrm{dom} g \cap \mathrm{dom} h$ *is a global minimizer of* $g(x) - h(x)$ *on* $\mathbb{R}^n$ *if and only if*

$$\partial_\varepsilon h(x^0) \subset \partial_\varepsilon g(x^0) \quad \forall \varepsilon > 0. \tag{4.28}$$

Proof Without loss of generality we may assume that $g(x^0) = h(x^0) = 0$. Then x^0 is a global minimizer of $g(x) - h(x)$ on $\mathbb{R}^n$ if and only if $h(x) \le g(x)\ \forall x \in \mathbb{R}^n$. It is easy to see that the latter condition is equivalent to (4.28). Indeed, observe that an affine function $\varphi(x) = \langle a, x - x^0\rangle + \varphi(x^0)$ is a minorant of a proper function $f(x)$ if and only if $a \in \partial_\varepsilon f(x^0)$ with $\varepsilon = f(x^0) - \varphi(x^0)$. Now, since $h(x)$ is a proper convex l.s.c. function on $\mathbb{R}^n$, it is known that $h(x) = \sup\{\varphi(x) |\ \varphi \in \mathbf{Q}\}$, where $\mathbf{Q}$ denotes the collection of all affine minorants of $h(x)$ (Theorem 2.5). Hence, the condition $h(x) \le g(x)\ \forall x \in \mathbb{R}^n$ is equivalent to saying that every $\varphi \in \mathbf{Q}$ is a minorant of g, i.e., by the above observation, every $a \in \partial_\varepsilon h(x^0)$ belongs to $\partial_\varepsilon g(x^0)$, for every $\varepsilon > 0$. $\square$

Remark 4.4 Geometrically, the condition $h(x) \le g(x)\ \forall x \in \mathbb{R}^n$ which expresses the global minimality of x^0 means

$$G \subset H, \tag{4.29}$$

where G, H are the epigraphs of the functions g, h, respectively. Since h is proper convex and l.s.c., H is a nonempty closed convex set in $\mathbb{R}^n \times \mathbb{R}$ and it is obvious that (4.29) holds if and only if every closed halfspace in $\mathbb{R}^n \times \mathbb{R}$ that contains H also contains G (see Strekalovski 1987, 1990, 1993). The latter condition, if restricted to nonvertical halfspaces, is nothing but a geometric expression of (4.28).

The next result exhibits an important duality relationship in dc minimization problems which has found many applications (Toland 1978).

Proposition 4.21 *Let g, h be as in Proposition 4.20. Then*

$$\inf_{x \in \mathrm{dom} g} \{g(x) - h(x)\} = \inf_{u \in \mathrm{dom} h^*} \{h^*(u) - g^*(u)\}. \tag{4.30}$$

If u^0 is a minimizer of $h^*(u) - g^*(u)$ then any $x^0 \in \partial g^*(u^0)$ is a minimizer of $g(x) - h(x)$.

Proof Suppose $g(x) - h(x) \geq \gamma$, i.e., $g(x) \geq h(x) + \gamma$ for all $x \in \mathrm{dom} g$. Then $g^*(u) = \sup_{x \in \mathrm{dom} g}\{\langle u, x\rangle - g(x)\} \leq \sup_{x \in \mathrm{dom} g}\{\langle u, x\rangle - h(x)\} - \gamma \leq h^*(u) - \gamma$, i.e., $h^*(u) - g^*(u) \geq \gamma$. Conversely, the latter inequality implies (in an analogous manner) $g^{**}(x) - h^{**}(x) \geq \gamma$, hence $g(x) - h(x) \geq \gamma$. This proves (4.30). Suppose now that u^0 is a minimizer of $h^*(u) - g^*(u)$ and let $x^0 \in \partial g^*(u^0)$, hence $u^0 \in \partial g(x^0)$. Since $h^*(u^0) + h(x^0) \geq \langle u^0, x^0\rangle$, we have

$$\begin{aligned} g(x^0) - h(x^0) &\leq g(x^0) - \langle u^0, x^0\rangle + h^*(u^0) \\ &\leq g(x) - \langle u^0, x\rangle + h^*(u^0) \\ &\leq g(x) - [\langle u^0, x\rangle - h^*(u^0)] \leq g(x) - h(x) \end{aligned}$$

for all $x \in \mathrm{dom} g$. □

4.9 Exercises

1 Let C be a compact set in $\mathbb{R}^n$. For every $x \in \mathbb{R}^n$ denote by $d(x) = \inf\{\|x - y\| \mid y \in \partial C\}$. Show that $d(x)$ is a dc function such that if C is a convex set then $d(x)$ is concave on C but convex on any convex subset of $\mathbb{R}^n \setminus C$.

Find a dc representation of the functions:

2 $P(x, y) = \sum_{j=1}^n [a_j x_j + b_j y_j + c_j x_j^2 + d_j y_j^2 + p_j x_j y_j + q_j x_j y_j^2 + r_j x_j^2 y_j]$ for $(x, y) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+$.

3

$$f(t) = \begin{cases} \alpha - wt & 0 \leq t \leq \delta \\ (\alpha - w\delta)(1 - \frac{t-\delta}{\eta-\delta}) & \delta \leq t \leq \eta \\ 0 & t \geq \eta \end{cases}$$

for $t \geq 0$ (α, δ, η are positive numbers, and $\delta < \eta$).

4

$$f(x) = \sum_{i=1}^n \frac{r_i(x_i)}{\beta_0 + \sum_{j=1}^k \beta_j x_j}, \quad x \geq 0,$$

where each $r_i(t)$ is a strictly increasing S-shaped function (convex for $t \in [0, \alpha_i]$, concave for $t \geq \alpha_i$, with $\alpha_i > 0$), and $\beta_i, i = 0, 1, \ldots, n$ are positive numbers.

5 $f(x) = \left(x_1 + \frac{1}{x_2}\right)^{1/2} + \frac{4}{(x_1 - x_2)^2}, \quad x \in \mathbb{R}^2_+.$

6 A function $f : C \to \mathbb{R}$ (where C is an open convex subset of $\mathbb{R}^n$) is said to be *weakly convex* if there exists a constant $r > 0$ such that for all $x^1, x^2 \in C$ and all $\lambda \in [0, 1]$:

$$f((1-\lambda)x^1 + \lambda x^2) \leq (1-\lambda)f(x^1) + \lambda f(x^2) + (1-\lambda)\lambda r\|x^1 - x^2\|^2.$$

Show that f is weakly convex if and only if it is a dc function of the form $f(x) = g(x) - r\|x\|^2$, where $r > 0$.

7 Show that a function $f : \mathbb{R}^n \to \mathbb{R}$ is weakly convex if and only if there exists a constant $r > 0$ such that for every $x^0 \in \mathbb{R}^n$ one can find a vector $y \in \mathbb{R}^n$ satisfying $f(x) - f(x^0) \geq \langle y, x - x^0\rangle - r\|x - x^0\|^2 \ \forall x \in \mathbb{R}^n$.

8 Find a convex minorant for the function $f(x) = e^{-(x_1+x_2-x_3)^2} + x_1x_2x_3$ on $-1 \leq x_i \leq +1,\ i = 1, 2, 3$.

9 Same question for the function $f(x) = \log(x_1 + x_2^2)[\min\{x_1, x_2\}]$ on $1 \leq x_i \leq 3, i = 1, 2$.

10 (1) If $\overline{x}$ is a local minimizer of the function $f(x) = g(x) - h(x)$ over $\mathbb{R}^n$ then $0 \in \partial g(\overline{x}) \underline{*} \partial h(\overline{x})$ where $A \underline{*} B = \{x |\ x + B \subset A\}$.

(2) If $\overline{x}$ is a global minimizer of the dc function $f = g - h$ over $\mathbb{R}^n$ then any point $x^* \in \partial h(\overline{x})$ achieves a minimum of the function $h^* - g^*$ over domg^*.

Part II
Global Optimization

Chapter 5
Motivation and Overview

5.1 Global vs. Local Optimum

We will be concerned with optimization problems of the general form

$$(P) \qquad \min\{f(x) \mid \quad g_i(x) \leq 0 \ (i = 1, \ldots, m), \ x \in X\},$$

where X is a closed convex set in $\mathbb{R}^n$, $f : \Omega \to \mathbb{R}$ and $g_i : \Omega \to \mathbb{R}, \ i = 1, \ldots, m$, are functions defined on some set Ω in $\mathbb{R}^n$ containing X. Denoting the *feasible set* by D:

$$D = \{x \in X | \ g_i(x) \leq 0, \ i = 1, \ldots, m\}$$

we can also write the problem as

$$\min\{f(x) | \ x \in D\}.$$

A point $x^* \in D$ such that (Fig. 5.1)

$$f(x^*) \leq f(x), \ \forall x \in D,$$

is called a *global optimal solution* (global minimizer) of the problem. A point $x' \in D$ such that there exists a neighborhood W of x' satisfying

$$f(x') \leq f(x), \ \forall x \in D \cap W$$

is called a *local optimal solution* (local minimizer). When the problem is *convex*, i.e., when all the functions $f(x)$ and $g_i(x), i = 1, \ldots, m$, are convex, any local minimizer is global (Proposition 2.30), but this is no longer true in the general case. A problem is said to be *multiextremal* when it has many local minimizers

H. Tuy, *Convex Analysis and Global Optimization*, Springer Optimization and Its Applications 110, DOI 10.1007/978-3-319-31484-6_5

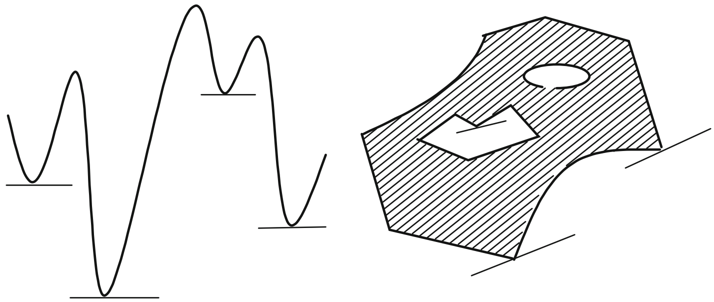


Fig. 5.1 Local and global minimizers

with different objective function values (so that a local minimizer may fail to be global; Fig. 5.1). Throughout the sequel, unless otherwise stated it will always be understood that the minimum required in problem (P) is a global one, and we will be mostly interested in the case when the problem is multiextremal. Often we will assume X compact and $f(x), g_i(x), i = 1, \ldots, m$, continuous on X, (so that if D is nonempty then the existence of a global minimum is guaranteed); but occasionally the case where $f(x)$ is discontinuous at certain boundary points of D will also be discussed, especially when this turns out to be important for the applications.

By its very nature, global optimization requires quite different methods from those used in conventional nonlinear optimization, which so far has been mainly concerned with finding local optima. Strictly speaking, local methods of nonlinear programming which rely on such concepts as derivatives, gradients, subdifferentials and the like, may even fail to locate a local minimizer. Usually these methods compute a Karush–Kuhn–Tucker (KKT) point, but in many instances most KKT points are not even local minimizers. For example, the problem:

$$\text{minimize } -\sum_{i=1}^{n}(c_i x_i + x_i^2) \text{ subject to } -1 \le x_i \le 1 \ (i = 1, \ldots, n),$$

where the c_i are sufficiently small positive numbers, has up to 3^n KKT points, of which only $2^n + 1$ points are local minimizers, the others are saddle points.

Of course the above example reduces to n univariate problems $\min\{-c_i x_i - x_i^2 | -1 \le x_i \le 1\}, i = 1, \ldots, n$, and since the strictly quasiconcave function $-c_i x_i - x_i^2$ attains its unique minimum over the segment $[-1, 1]$ at $x_i = 1$, it is clear that $(1, 1, \ldots, 1)$ is the unique global minimizer of the problem. This situation is typical: although certain multiextremal problems may be easy (among them instances of problems which have been shown to be NP-hard, see Chap. 7), their optimal solution is most often found on the basis of exploiting favorable global properties of the functions involved rather than by conventional local methods. In any case, easy multiextremal problems are exceptions. Most nonconvex global optimization problems are difficult, and their main difficulty is often due to multiextremality.

From the computational complexity viewpoint, even the problem which only differs from the above mentioned example in that the objective function is a general quadratic concave function is NP-hard (Sahni 1974). It should be noted, however, that, as shown in Murty and Kabadi (1987), to verify that a given point is an unconstrained local minimizer of a nonconvex function is already NP-hard. Therefore, even if we were to restrict our goal to obtaining a local minimizer only, this would not significantly lessen the complexity of the problem. On the other hand, as optimization models become widely used in engineering, economics, and other sciences, an increasing number of global optimization problems arise from applications that cannot be solved successfully by standard methods of linear and nonlinear programming and require more suitable methods. Often, it may be of considerable interest to know at least whether there exists any better solution to a practical problem than a given solution that has been obtained, for example, by some local method. As will be shown in Sect. 5.5, this is precisely the central issue of global optimization.

5.2 Multiextremality in the Real World

The following examples illustrate how multiextremal nonconvex problems arise in different economic, engineering, and management activities.

Example 5.1 (Production–Transportation Planning) Consider r factories producing a certain good to satisfy the demands d_j, $j = 1, \ldots, m$, of m destination points. The cost of producing t units of goods at factory i is a concave function $g_i(t)$ (economy of scale), the transportation cost is linear, with unit cost c_{ij} from factory i to destination point j. In addition, if the destination point j receives $t \neq d_j$ units, then a shortage penalty $h_j(t)$ must be paid which is a decreasing nonnegative convex function of t in the interval $[0, d_j)$. Under these conditions, the problem that arises is the following one where a dc function has to be minimized over a polyhedron:

$$
\begin{aligned}
\text{minimize} \quad & \sum_{i=1}^{r}\sum_{j=1}^{m} c_{ij}x_{ij} + \sum_{i=1}^{r} g_i(y_i) + \sum_{j=1}^{m} h_j(z_j) \\
\text{s.t.} \quad & \sum_{j=1}^{m} x_{ij} = y_i \quad (i = 1, \ldots, r) \\
& \sum_{i=1}^{r} x_{ij} = z_j \quad (j = 1, \ldots, m) \\
& x_{ij}, y_i, z_j \geq 0 \quad \forall i, j.
\end{aligned}
$$

In a general manner nonconvexities in economic processes arise from economies of scale (concave cost functions) or increasing returns (convex utility functions).

Example 5.2 (Location–Allocation) A number of p facilities providing the same service have to be constructed to serve n users located at points $a^j, j = 1, \ldots, n$ on the plane. Each user will be served by one of the facilities, and the transportation cost is a linear function of the distance. The problem is to determine the locations $x^i \in \mathbb{R}^2, i = 1, \ldots, p$, of the facilities so as to minimize the weighted sum of all transportation costs from each user to the facility that serves this user. Introduced by Cooper (1963), this problem was shown to be NP-hard by Megiddo and Supowwit (1984). Originally it was formulated as consisting simultaneously of a nontrivial combinatorial part (allocation, i.e., partitioning the set of users into groups to be served by each facility) and a nonlinear part (location, i.e., finding the optimal location of the facility serving each group). The corresponding mathematical formulation of the problem is (Brimberg and Love 1994):

$$\begin{array}{ll} \min_{x,u} & \sum_{i=1}^{p} \sum_{j=1}^{N} w_{ij} \|a^j - x^i\| \\ \text{s.t.} & \sum_{i=1}^{p} w_{ij} = r_j,\ j = 1, \ldots, N \\ & w_{ij} \geq 0,\ i = 1, \ldots, p,\ j = 1, \ldots, N \\ & x^i \in S,\ i = 1, \ldots, p, \end{array}$$

where w_{ij} is an unknown weight representing the flow from point a^j to facility i, S is a rectangular domain in $\mathbb{R}^2$ where the facilities must be located.

With this traditional formulation the problem is exceedingly complicated as it involves a large number of variables and a highly nonconvex objective function. However, observing that to minimize the total cost, each user must be served by the closest facility, and that the distance from a user j to the closest facility is obviously

$$d_j(x^1, \ldots, x^p) = \min_{i=1,\ldots,p} \|a^j - x^i\| \tag{5.1}$$

the problem can be formulated more simply as

$$\text{minimize} \ \sum_{j=1}^{n} w_j d_j(x^1, \ldots, x^p) \ \text{ subject to } \ x^i \in \mathbb{R}^2, i = 1, \ldots, p,$$

where $w_j > 0$ is the weight assigned to user j. Since each $d_j(.)$ is a dc function (pointwise minimum of a set of convex functions $\|a^j - x^i\|$, see Sect. 3.1), the problem amounts to minimizing a dc function over $(\mathbb{R}^2)^p$.

Example 5.3 (Engineering Design) Random fluctuations inherent in any fabrication process (e.g., integrated circuit manufacturing) may result in a very low production yield. To help the designer to minimize the influence of these random fluctuations, a method consists in maximizing the yield by centering the nominal value of design parameters in the region of acceptability. In more precise terms, this *design centering problem* (Groch et al. 1985; Vidigal and Director 1982) can be formulated as follows. Suppose the quality of a manufactured item can be characterized by an n-dimensional parameter and an item is accepted when the parameter is contained in a given region $M \subset \mathbb{R}^n$. If x is the nominal value of the parameter, and y its actual value, then usually y will deviate from x and for every fixed x the probability that the deviation $\|x - y\|$ does not exceed a given level r monotonically increases with r. Under these conditions, for a given nominal value of x the production yield can be measured by the maximal value of $r = r(x)$ such that

$$B(x, r) := \{y : \|x - y\| \leq r\} \subset M$$

and to maximize the production yield, one has to solve the problem

$$\text{maximize } \; r(x) \; \text{ subject to } \; x \in \mathbb{R}^n. \tag{5.2}$$

Often, the norm $\|.\|$ is ellipsoidal, i.e., there exists a positive definite matrix Q such that $\|x - y\|^2 = (x - y)^T Q(x - y)$. Denoting $D = \mathbb{R}^n \setminus M$, we then have

$$\begin{aligned} r^2(x) &= \inf\{(x - y)^T Q(x - y) : y \notin D\} \\ &= \inf\{x^T Qx + y^T Qy - 2x^T Qy : y \notin D\}, \end{aligned}$$

so $r^2(x) = x^T Qx - h(x)$ with $h(x) = \sup\{2x^T Qy - y^T Qy : y \notin D\}$. Since for each $y \notin D$ the function $x \mapsto 2x^T Qy - y^T Qy$ is affine, $h(x)$ is a convex function and therefore, $r^2(x)$ is a dc function. Using nonlinear programming methods, we can only compute KKT points which may even fail to be local minima, so the design centering problem is another instance of multiextremal optimization problem (see, e.g., Thach 1988). Problems similar to (5.2) appear in engineering design, coordinate measurement technique, etc.

Example 5.4 (Pooling and Blending) In oil refineries, a two-stage process is used to form final products $j = 1, \ldots, J$ from given oil components $i = 1, \ldots, I$ (see Fig. 5.2). In the first stage, intermediate products are obtained by combining the components supplied from different sources and stored in special tanks or pools. In the second stage, these intermediate products are combined to form final products of prescribed quantities and qualities. A certain quantity of an initial component can go directly to the second stage. If x_{il} denotes the amount of component i allocated

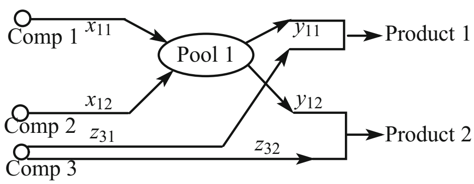


Fig. 5.2 The pooling problem

to pool l, y_{lj} the amount going from pool l to product j, z_{ij} the amount of component i going directly to product j, p_{lk} the level of quality k (relative sulfur content level) in pool l, and C_{ik} the level of quality k in component i, then these variables must satisfy (see, e.g., Ben-Tal et al. 1994; Floudas and Aggarwal 1990; Foulds et al. 1992) (Fig. 5.2):

$$\left.\begin{array}{l}\sum_l x_{il} + \sum_j z_{ij} \le A_i \\ \sum_i x_{il} - \sum_j y_{lj} = 0\end{array}\right\} \quad \text{(component balance)}$$

$$\begin{array}{l}\sum_i x_{il} \le S_l \qquad \text{(pool balance)} \\ \sum_l (p_{lk} - P_{jk}) y_{lj} + \sum_i (C_{ij} - P_{jk}) z_{ij} \le 0 \quad \text{(pool quality)} \\ \sum_l y_{lj} + \sum_i z_{ij} \le D_j \quad \text{(product demands constraints)} \\ x_{il}, y_{lj}, z_{ij}, p_{lk} \ge 0,\end{array}$$

where A_i, S_l, P_{jk}, D_j are the upper bounds for component availabilities, pool sizes, product qualities, and product demands, respectively, and C_{ik} the level of quality k in component i. Given the unit prices c_i, d_j of component i and product j, the problem is to determine the variables $x_{il}, y_{lj}, z_{ij}, p_{lk}$ satisfying the above constraints with maximum profit

$$-\sum_i \sum_l c_i x_{il} + \sum_l \sum_j d_j y_{lj} + \sum_i \sum_j (d_j - c_i) z_{ij}.$$

Thus, a host of optimization problems of practical interest in economics and engineering involve nonconvex functions and/or nonconvex sets in their description. Nonconvex global optimization problems arise in many disciplines including control theory (robust control), computer science (VLSI chip design and databases), mechanics (structural optimization), physics (nuclear design and microcluster phenomena in thermodynamics), chemistry (phase and chemical reaction equilibrium, molecular conformation), ecology (design and cost allocation for waste treatment systems), ... Not to mention many mathematical problems of importance which can be looked at and studied as global optimization problems. For instance, finding a fixed point of a map $F : \mathbb{R}^n \to \mathbb{R}^n$ on a set $D \subset \mathbb{R}^n$, which is a recurrent problem in applied mathematics, reduces to finding a global minimizer of the function $\|x - F(x)\|$ over D; likewise, solving a system of equations $g_i(x) = 0$ $(i = 1, \dots, m)$ amounts to finding a global minimizer of the function $f(x) = \max_i |g_i(x)|$.

5.3 Basic Classes of Nonconvex Problems

Optimization problems without any particular structure are hopelessly difficult. Therefore, our interest is focussed first of all on problems whose underlying mathematical structure is more or less amenable to analysis and computational study. This point of view leads us to consider two basic classes of nonconvex global optimization problems: dc optimization and monotonic optimization.

5.3.1 DC Optimization

A dc optimization (dc programming) problem is a problem of the form

$$\left|\begin{array}{l}\text{minimize}\, f(x)\\ \text{subject to} \quad x \in D,\ h_i(x) \le 0\ (i = 1, \ldots, m),\end{array}\right. \tag{5.3}$$

where D is a convex set in $\mathbb{R}^n$, and each of the functions $f(x), h_1(x), \ldots, h_m(x)$ is a dc function on $\mathbb{R}^n$. Most optimization problems of practical interest are dc programs (see the examples in Sect. 4.2).

The class of dc optimization problems includes several important subclasses: concave minimization, reverse convex programming, non convex quadratic optimization among others.

a. Concave Minimization (Convex Maximization)

$$\text{minimize}\, f(x) \quad \text{subject to} \quad x \in D, \tag{5.4}$$

where $f(x)$ is a concave function (i.e., $-f(x)$ is a convex function) and D is a closed convex set given by an explicit system of convex inequalities: $g_i(x) \le 0,\ i = 1, \ldots, m$.

This is a global optimization problem with the simplest structure. Since (5.4) can be written as $\min\{t|\ x \in D, f(x) \le t\}$, it is equivalent to minimizing t under the constraint $(x,t) \in G \setminus H$ with $G = D \times \mathbb{R}$ and $H = \{(x,t)|\, f(x) > t\}$ (H is a convex set because the function $f(x)$ is concave).

Of particular interest is the linearly constrained concave minimization problem (i.e., problem (5.4) when the constraint set D is a polyhedron given by a system of linear inequalities) which plays a central role in global optimization and has been extensively studied in the last four decades.

Since the most important property of $f(x)$ that should be exploited in problem (5.4) is its quasiconcavity (expressed in the fact that every upper level set $\{x|\ f(x) \ge \alpha\}$ is convex), rather than its concavity, it is convenient to slightly extend problem (5.4) by assuming $f(x)$ to be only quasiconcave. The corresponding problem, called *quasiconcave minimization* (or *quasiconvex maximization*), does not differ much from concave minimization. Also, by considering quasiconcave rather than concave minimization, more insight can be obtained into the relationship between this problem and other nonconvex optimization problems.

b. Reverse Convex Programming

$$\text{minimize}\, f(x) \quad \text{subject to} \quad x \in D,\ h(x) \geq 0, \tag{5.5}$$

where $f(x), h(x)$ are convex finite functions on $\mathbb{R}^n$ and D is a closed convex set in $\mathbb{R}^n$. Setting

$$C = \{x|\ h(x) \leq 0\},$$

the problem can be rewritten as

$$\text{minimize}\, f(x) \quad \text{subject to} \quad x \in D \setminus \text{int}C.$$

This differs from a conventional convex program only by the presence of the *reverse convex constraint* $h(x) \geq 0$. Writing the problem as $\min\{t|\ f(x) \leq t,\ x \in D,\ h(x) \geq 0\}$, we see that it amounts to minimizing t over the set $G \setminus \text{int}H$, with $G = \{(x,t)|\, f(x) \leq t,\ x \in D\},\ H = C \times \mathbb{R}$.

Just as concave minimization can be extended to quasiconcave minimization, so can problem (5.5) be extended to the case when $f(x)$ is quasiconvex. As indicated above, concave minimization can be converted to reverse convex programming by introducing an additional variable. For many years, it was commonly believed (but never established) that conversely, a reverse convex program should be equivalent to a concave minimization problem via some simple transformation. It was not until 1991 that a duality theory for nonconvex optimization was developed (Thach 1991), within which framework the two problems: quasiconcave minimization over convex sets and quasiconvex minimization over complements of convex sets are dual to each other, so that solving one can be reduced to solving the other and vice versa.

c. Quadratic Optimization

An important subclass of dc programs is constituted by *quadratic nonconvex programs* which, in their most general formulation, consist in minimizing a quadratic function under nonconvex quadratic constraints, i.e., a problem (5.3) in which

$$f(x) = \frac{1}{2}\langle x, Q^0 x\rangle + \langle c^0, x\rangle$$

$$h_i(x) = \frac{1}{2}\langle x, Q^i x\rangle + \langle c^i, x\rangle \quad i = 1, \dots, m,$$

where $Q^i,\ i = 0, 1, \dots, m$, are symmetric $n \times n$ matrices, and at least one of these matrices is indefinite or negative semidefinite.

5.3.2 *Monotonic Optimization*

This is the class of problems of the form

$$\left|\begin{array}{l}\text{maximize } f(x)\\ \text{subject to } g(x) \le 0 \le h(x),\ x \in \mathbb{R}^n_+,\end{array}\right. \tag{5.6}$$

where $f(x), g(x), h(x)$ are *increasing* functions on $\mathbb{R}^n_+$. A function $f : \mathbb{R}^n_+ \to \mathbb{R}$ is said to be increasing if $f(x) \le f(x')$ whenever $x, x' \in \mathbb{R}^n_+, x \le x'$.

Monotonic optimization is a new important chapter of global optimization which has been developed in the last 15 years. It is easily seen that if $\mathscr{F}$ denotes the family of increasing functions on $\mathbb{R}^n_+$ then

$$f_1(x), f_2(x) \in \mathscr{F} \Rightarrow \min\{f_1(x), f_2(x)\} \in \mathscr{F};\ \max\{f_1(x), f_2(x)\} \in \mathscr{F}.$$

These properties are similar to those of the family of dc functions, so a certain parallel exists between Monotonic Optimization and DC Optimization.

5.4 General Nonconvex Structure

In the deterministic, much more than in the stochastic and other approaches to global optimization, it is essential to understand the *mathematical structure* of the problem being investigated. Only a deep understanding of this structure can provide insight into the most relevant properties of the problem and suggest efficient methods for handling it.

Despite the great variety of nonconvex optimization problems, most of them share a common mathematical structure, which is the *complementary convex* structure as described in (5.5). Consider, for example, the general dc programming problem (5.3). As we saw in Sect. 4.1, the system of dc constraints $h_i(x) \le 0, i = 1, \ldots, m$ is equivalent to the single dc constraint:

$$h(x) = \max_{i=1,\ldots,m} h_i(x) \le 0. \tag{5.7}$$

If $h(x) = p(x) - q(x)$ then the constraint (5.7) in turn splits into two constraints: one convex constraint $p(x) \le t$ and one reverse convex constraint $q(x) \ge t$. Thus, assuming $f(x)$ linear and setting $z = (x, t)$, the dc program (5.3) can be reformulated as the reverse convex program:

$$\min\{l(z) |\ z \in G \setminus \text{int}H\}, \tag{5.8}$$

where $l(z) = f(x)$ is linear, $G = \{z = (x, t) | x \in D, p(x) \le t\}$, and $H = \{z = (x, t) |\ q(x) \ge t\}$. This reverse convex program (5.8) is also referred to as a *canonical dc programming problem.*

Consider next the more general problem of continuous optimization

$$\text{minimize } f(x) \quad \text{subject to} \quad x \in S, \tag{5.9}$$

where S is a compact set in $\mathbb{R}^n$ and $f(x)$ is a continuous function on $\mathbb{R}^n$. As previously, by writing the problem as $\min\{t|\ f(x) \leq t, x \in S\}$ and changing the notation, we can assume that the objective function $f(x)$ in (5.9) is a linear function. Then, by Proposition 4.13, if $r : \mathbb{R}^n \to \mathbb{R}_+$ is any *separator* for S (for instance, $r(x) = \inf\{\|x - y\| \mid y \in S\}$) the set S can be described as

$$S = \{x|\ g(x) - \|x\|^2 \leq 0\},$$

where $g(x) = \sup_{y \notin S}\{r^2(y) + 2xy - \|y\|^2\}$ is a convex function on $\mathbb{R}^n$. Therefore, problem (5.9) is again equivalent to a problem of the form (5.8).

To sum up, a complementary convex structure underlies every continuous global optimization problem. Of course, this structure is not always apparent and even when it has been made explicit still a lot of work remains to be done to bring it into a form amenable to efficient computational analysis. Nevertheless, the attractive feature of the complementary convex structure is that, on the one hand, it involves convexity, and therefore brings to bear analytical tools from convex analysis, like subdifferential, supporting hyperplane, polar set, etc. On the other hand, since convexity is involved in the "wrong" (reverse) direction, these tools must be used in some specific way and combined with combinatorial tools like cutting planes, relaxation, restriction, branch and bound, etc. In this sense, global optimization is a sort of bridge between nonlinear programming and combinatorial (discrete) optimization.

Remark 5.1 The connection between global optimization and combinatorial analysis is also apparent from the fact that any 0-1 integer programming problem can be converted into a global optimization problem. Specifically, a discrete constraint of the form

$$x_i \in \{0, 1\} \quad (i = 1, \ldots, n)$$

can be rewritten as

$$x \in G \setminus H,$$

with G being the hypercube $[0, 1]^n$ and H the interior of the ball circumscribing the hypercube, or alternatively as

$$0 \leq x_i \leq 1, \quad \sum_{i=1}^{n} x_i(x_i - 1) \geq 0.$$

5.5 Global Optimality Criteria

The essence of an iterative method of optimization is that we have a trial solution x^0 to the problem and look for a better one. If the method used is some standard procedure of nonlinear programming and x^0 is a KKT point, there is no clue to proceed further. We must then stop the procedure even though there is no guarantee that x^0 is a global optimal solution. It is this ability to become trapped at a stationary point which causes the failure of classical methods of nonlinear programming and motivates the need to develop global search procedures. Therefore, any global search procedure must be able to address the fundamental question of how to *transcend* a given feasible solution (the current incumbent, which may be a stationary point). In other words, the core of global optimization is the following issue:

Given a solution x^0 (the incumbent), check whether it is a global optimal solution, and if it is not, find a better feasible solution.

As in classical optimization theory, one way to deal with this issue is by devising criteria for recognizing a global optimum. For unconstrained dc optimization problems, we already know a global optimality criterion by Hiriart-Urruty (Proposition 4.20), namely: Given a proper convex l.s.c. function $h(x)$ on $\mathbb{R}^n$ and an arbitrary proper function $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, a point x^0 such that $g(x^0) < +\infty$ is an unconstrained global minimizer of $g(x) - h(x)$ if and only if

$$\partial_\varepsilon h(x^0) \subset \partial_\varepsilon g(x^0) \quad \forall \varepsilon > 0. \tag{5.10}$$

We now present two global optimality criteria for the general problem

$$\min\{f(x) |\, x \in S\}, \tag{5.11}$$

where the set $S \subset \mathbb{R}^n$ is assumed nonempty, compact, and robust, i.e., $S = \mathrm{cl}(\mathrm{int}S)$, and $f(x)$ is a continuous function on S. Since $f(x)$ is then bounded from below on S, without loss of generality we can also assume $f(x) > 0\ \forall x \in S$.

Proposition 5.1 *A point $x^0 \in S$ is a global minimizer of $f(x)$ over S if and only if $\alpha := f(x^0)$ satisfies either of the following conditions:*

(i) *The sequence $r(\alpha, k), k = 1, 2, \dots,$ is bounded (Falk 1973a), where*

$$r(\alpha, k) = \int_S \left[\frac{\alpha}{f(x)}\right]^k dx.$$

(ii) *The set $E := \{x \in S |\, f(x) < \alpha\}$ has null Lebesgue measure (Zheng 1985).*

Proof Suppose x^0 is not a global minimizer, so that $f(x) < \alpha$ for some $x \in S$. By continuity we will have $f(x') < \alpha$ for all x' sufficiently close to x. On the other hand, since S is robust, in any neighborhood of x there is an interior point of S. Therefore, by taking a point $x' \in \mathrm{int}S$ sufficiently close to x we will have $f(x') < \alpha$. Let W be a neighborhood of x' contained in S such that $f(y) < \alpha$ for all $y \in W$. Then $W \subset E$, hence $\mathrm{mes}E \geq \mathrm{mes}W > 0$. Conversely, if $\mathrm{mes}E > 0$, then $E \neq \emptyset$, hence x^0 is not a global minimizer. This proves that (ii) is a necessary and sufficient condition for the optimality of x^0. We now show that (i) is equivalent to (ii). If $\mathrm{mes}E > 0$, then since $E = \cup_{i=1}^{+\infty}\{x \in S |\, \frac{\alpha}{f(x)} \geq 1 + \frac{1}{i}\}$ there is i_0 such that $E_0 = \{x \in S |\, \frac{\alpha}{f(x)} \geq 1 + \frac{1}{i_0}\}$ has mes$E_0 = \mu > 0$. This implies

$$\int_S \left[\frac{\alpha}{f(x)}\right]^k dx \geq \mu(1 + 1/i_0)^k \rightarrow +\infty \text{ as } k \rightarrow +\infty.$$

On the other hand, if mes$E = 0$ then

$$\int_S \left[\frac{\alpha}{f(x)}\right]^k dx = \int_{S\setminus E} \left[\frac{\alpha}{f(x)}\right]^k dx \leq \text{mes}(S \setminus E) \quad \forall k,$$

proving that (i) $\Leftrightarrow$ (ii). □

Using Falk's criterion it is even possible, when the global optimizer is unique, to derive a global criterion which in fact yields a closed formula for this unique global optimizer (Pincus 1968).

Proposition 5.2 *Assume, in addition to the previously specified conditions, that $f(x)$ has a unique global minimizer x^0 on S. Then for every $i = 1, \ldots, n$:*

$$\frac{\int_S x_i e^{-kf(x)} dx}{\int_S e^{-kf(x)} dx} \rightarrow x_i^0 \quad (k \rightarrow +\infty). \tag{5.12}$$

Proof We show that for every $i = 1, \ldots, n$:

$$\varphi_i(k) := \frac{\int_S |x_i - x_i^0| e^{-kf(x)} dx}{\int_S e^{-kf(x)} dx} \rightarrow 0 \quad (k \rightarrow +\infty). \tag{5.13}$$

For any fixed $\varepsilon > 0$ let $W_\varepsilon = \{x \in S \mid \ \|x - x^0\| < \varepsilon\}$ and write

$$\varphi_i(k) = \frac{\int_{W_\varepsilon} |x_i - x_i^0| e^{-kf(x)} dx}{\int_S e^{-kf(x)} dx} + \frac{\int_{S\setminus W_\varepsilon} |x_i - x_i^0| e^{-kf(x)} dx}{\int_S e^{-kf(x)} dx}.$$

Since x^0 is the unique minimizer of $f(x)$ over S, clearly $\min\{f(x) | \ x \in S\setminus W_\varepsilon\} > f(x^0)$ and there is $\delta > 0$ such that $\max\{f(x^0) - f(x) \mid \ x \in S \setminus W_\varepsilon\} < -\delta$. Then

$$\frac{\int_{W_\varepsilon} |x_i - x_i^0| e^{-kf(x)} dx}{\int_S e^{-kf(x)} dx} \leq \varepsilon \frac{\int_{W_\varepsilon} e^{-kf(x)} dx}{\int_S e^{-kf(x)} dx} \leq \varepsilon. \tag{5.14}$$

On the other hand, setting $M = \max_{x \in S} \|x - x^0\|$, we have

$$\frac{\int_{S\setminus W_\varepsilon} |x_i - x_i^0| e^{-kf(x)} dx}{\int_S e^{-kf(x)} dx} \leq M \int_{S\setminus W_\varepsilon} \frac{e^{k[f(x^0) - f(x)]} dx}{\int_S e^{k[f(x^0)-f(x)]} dx} \tag{5.15}$$

$$\leq \frac{M \text{mes}(S \setminus W_\varepsilon) e^{-k\delta}}{\int_S e^{k[f(x^0)-f(x)]} dx}. \tag{5.16}$$

The latter ratio tends to 0 as $k \to +\infty$ because $\int_S e^{k[f(x^0)-f(x)]}dx$ is bounded by Proposition 5.1 (Falk's criterion for x^0 to be the global minimizer of $e^{f(x)}$). Since $\varepsilon > 0$ is arbitrary, this together with (5.14) implies (5.13). □

Remark 5.2 Since $f(x) > 0 \ \forall x \in S$, a minimizer of $f(x)$ over S is also a maximizer of $\frac{1}{f(x)}$ and conversely. Therefore, from the above propositions we can state, under the same assumptions as previously (S compact and robust and $f(x)$ continuous positive on S):

(i) A point $x^0 \in S$ with $f(x^0) = \alpha$ is a global *maximizer* of $f(x)$ over S if and only if the sequence

$$s(\alpha, k) := \int_S \left[\frac{f(x)}{\alpha}\right]^k dx, \ k = 1, 2, \ldots,$$

is bounded.

(ii) If $f(x)$ has a unique global *maximizer* x^0 over S then

$$x_i^0 = \lim_{k \to +\infty} \frac{\int_S x_i e^{kf(x)} dx}{\int_S e^{kf(x)} dx}, \quad i = 1, \ldots, n.$$

5.6 DC Inclusion Associated with a Feasible Solution

In view of the difficulties with computing multiple integrals the above global optimality criteria are of very limited applicability. On the other hand, using the complementary convex structure common to most optimization problems, it is possible to formulate global optimality criteria in such a way that in most cases transcending an incumbent solution x^0 reduces to checking a *dc inclusion* of the form

$$D \subset C, \tag{5.17}$$

where D, C are closed convex subsets of $\mathbb{R}^n$ depending in general on x^0. Furthermore, this inclusion is such that from any $x \in D \setminus C$ it is easy to derive a better feasible solution than x^0.

Specifically, for the concave minimization problem (5.4), $C = \{x | \, f(x) \geq f(x^0)\}$ and D is the constraint set. Clearly these sets are closed convex, and $x^0 \in D$ is optimal if and only if $D \subset C$, while any $x \in D \setminus C$ yields a better feasible solution than x^0.

Consider now the canonical dc program (5.8):

$$\min\{l(x)|\ x \in G \setminus \mathrm{int}H\}, \tag{5.18}$$

where G and H are closed convex subsets of $\mathbb{R}^n$. As shown in Sect. 4.3 this is the general form to which one can reduce virtually any global optimization problem to be considered in this book. Assume that the feasible set $D = G \setminus \mathrm{int}H$ is *robust*, i.e., satisfies $D = \mathrm{cl}(\mathrm{int}D)$ and that the problem is genuinely nonconvex, which means that

$$\min\{l(x)|\ x \in G\} < \min\{l(x)|\ x \in G \setminus \mathrm{int}H\}. \tag{5.19}$$

Proposition 5.3 *Under the stated assumptions a feasible solution x^0 to (5.18) is global optimal if and only if*

$$\{x \in G|\ l(x) \leq l(x^0)\} \subset H. \tag{5.20}$$

Proof Suppose there exists $x \in G \cap \{x|\ l(x) \leq l(x^0)\} \setminus H$. From (5.19) we have $l(z) < l(x)$ for at least some $z \in G \cap \mathrm{int}H$. Then the line segment $[z, x]$ meets the boundary ∂H at some point $x' \in G \setminus \mathrm{int}H$, i.e., at some feasible point x'. Since $x \notin H$, we must have $x' = \lambda x + (1-\lambda)z$ with $0 < \lambda < 1$, hence $l(x') = \lambda l(x) + (1-\lambda) l(z) < l(x) \leq l(x^0)$. Thus, if (5.20) does not hold then we can find a feasible solution x' better than x^0. Conversely suppose there exists a feasible solution x better than x^0. Let W be a ball around x such that $l(x') < l(x^0)$ for all $x' \in W$. Since D is robust one can find a point $x' \in W \cap \mathrm{int}D$. Then clearly $x' \in G \setminus H$ and $l(x') < l(x^0)$, contrary to (5.20). Therefore, if (5.20) holds, x^0 must be global optimal. □

Thus, for problem (5.18) transcending x^0 amounts to solving an inclusion (5.17) where $D = \{x \in G|\ l(x) \leq l(x^0)\}$ and $C = H$. If $x \in G \setminus H$ then the point $x' \in \partial H \cap [z, x]$ yields a better feasible solution than x^0, where z is any point of $D \cap \mathrm{int}H$ with $l(z) < l(x)$ [this point exists by virtue of (5.19)].

We shall refer to the problem of checking an inclusion $D \subset C$ or, equivalently, solving the inclusion $x \in D \setminus C$, as the *DC feasibility problem*. Note that for the concave minimization problem the set $C = \{x|\ f(x) \geq f(x^0)\}$ depends on x^0, whereas for the canonical dc program the set $D = \{x \in G|\ l(x) \leq l(x^0)\}$ depends on x^0.

5.7 Approximate Solution

In practice, the computation of an exact global optimal solution may be too expensive while most often we only need the global optimum within some prescribed tolerance $\varepsilon \geq 0$. So problem

$$(P) \qquad\qquad \min\{f(x)|\ x \in D\}$$

can be considered solved if a feasible solution $\overline{x}$ has been found such that

$$f(x) \geq f(\overline{x}) - \varepsilon, \quad \forall x \in D. \tag{5.21}$$

Such a feasible solution is called a global *ε-optimal solution*. Setting $\Delta f = \max f(D) - \min f(D)$, the relative error made by accepting $f(\overline{x})$ as the global optimum is

$$\frac{f(\overline{x}) - \min f(D)}{\Delta f} \leq \frac{\varepsilon}{\Delta f} \leq \frac{\varepsilon}{M},$$

where M is a lower bound for Δf.

With tolerance ε the problem of transcending a feasible solution x^0 then consists in the following:

(∗) *Find a feasible solution z such that $f(z) < f(x^0)$ or else prove that x^0 is a global ε-minimizer.*

A typical feature of many global optimization problems which results from their complementary convex structure is that a subset X (typically, a finite subset of the boundary) of the feasible domain D is known to contain at least one global minimizer and is such that starting from any point $z \in D$ we can find, by local search, a point $x \in X$ with $f(x) \leq f(z)$. Under these circumstances, if a procedure is available for solving (∗) then (P) can be solved according to a *two phase scheme* as follows:

Let $z^0 \in D$ be an initial feasible solution. Set $k = 0$.

Phase 1 (Local phase). Starting from z^k search for a point $x^k \in X$ such that $f(x^k) \leq f(z^k)$. Go to Phase 2.

Phase 2 (Global phase). Solve (∗) for $\overline{x} = x^k$, by the available procedure. If the output of this procedure is the global ε-optimality of x^k, then terminate. Otherwise, the procedure returns a feasible point z such that $f(z) < f(x^k)$. Then set $z^k \leftarrow z$, increase k by 1 and go back to Phase 1.

Phase 1 can be carried out by using any local optimization or nondeterministic method (heuristic or stochastic). Only Phase 2 is what distinguishes deterministic approaches from the other ones. In the next chapters we shall discuss various methods for solving inclusions of the form (∗), i.e., practically for carrying out Phase 2, for various classes of problems. Clearly, the completion of one cycle of the above described iterative scheme achieves a move from one feasible point $x^k \in X$ to a feasible point $x^{k+1} \in X$ such that $f(x^{k+1}) < f(x^k)$. If X is finite then the process is guaranteed to terminate after finitely many cycles by a global ε-optimal solution. In the general case, there is the possibility of the process being infinite. Fortunately, however, for most problems it turns out that the iterative process can be organized so that, whenever infinite, it will converge to a global optimum; in other words, after a sufficient number of iterations it will yield a solution as close to the global optimum as desired.

5.8 When is Global Optimization Motivated?

Since global optimization methods are generally expensive, they should be used only when they are really motivated. Before applying a sophisticated global optimization algorithm to a given problem, it is therefore necessary to perform elementary operations (such as fixing certain variables to eliminate them, tightening bounds, and changing the variables) for improving or simplifying the formulation wherever possible. In some simple cases, an obvious transformation may reduce a seemingly nonconvex problem to a convex or even linear program. For instance, a problem of the form

$$\min\{-\sqrt{1+(l(x))^2}|\ x \in D\},$$

where D is a polyhedron and $l(x)$ a linear function nonnegative on D is merely a disguise of the linear program $\max\{l(x)|\ x \in D\}$. More generally, problems such as $\min\{\Phi(l(x)|\ x \in D\}$ where D is a polyhedron, $l(x)$ a linear function, while $\Phi(u)$ is a nondecreasing function of u on the set $l(D)$ should be recognized as linear programs and solved as such.[1]

There are, however, circumstances when more or less sophisticated transformations may be required to remove the apparent nonconvexities and reformulate the problem in a form amenable to known efficient solution methods. Therefore, *preprocessing* is often a necessary preliminary phase before solving any global optimization problem.

Example 5.5 Consider the following problem encountered in the theoretical evaluation of certain fault-tolerant shortest path distributed algorithms (Bui 1993):

$$(Q) \qquad \min \sum_{i=1}^{n} \frac{(\sigma - 2x_i)y_i + x_i}{(\delta - y_i)x_i} : \tag{5.22}$$

$$\sum_{i=1}^{n} x_i \geq a, \quad \sum_{i=1}^{n} y_i \geq b \tag{5.23}$$

$$\eta \leq x_i \leq \gamma, \quad \eta \leq y_i \leq \delta, \tag{5.24}$$

where $0 < n\eta \leq b$, $0 < \gamma < 1 < \sigma$ and $0 < \delta < 1$. Assume that $n\gamma \geq a, n\delta \geq b$ (otherwise the problem is infeasible). The objective function looks rather complicated, and one might be tempted to study the problem as a difficult nonconvex optimization problem. Close scrutiny reveals, however, that certain variables can be fixed and eliminated. Specifically, by rewriting the objective function as

$$\sum_{i=1}^{n} \left[\frac{\sigma}{x_i} \left(\frac{\delta}{\delta - y_i} - 1 \right) + \frac{1 - 2\delta}{\delta - y_i} + 2 \right]. \tag{5.25}$$

[1] Unfortunately, such linear programs disguised in "nonlinear" global optimization problems have been and still are sometimes used for "testing" global optimization algorithms.

it is clear that a feasible solution (x, y) of (Q) can be optimal only if $x_i = \gamma,\ i = 1, \ldots, n$. Hence, we can set $x_i = \gamma,\ i = 1, \ldots, n$. Problem (Q) then becomes

$$(Q*) \qquad \min \sum_{i=1}^{n} \left[\frac{\sigma}{\gamma} \left(\frac{\delta}{\delta - y_i} - 1 \right) + \frac{1 - 2\delta}{\delta - y_i} + 2 \right] : \tag{5.26}$$

$$\sum_{i=1}^{n} y_i \geq b, \quad \eta \leq y_i \leq \delta. \tag{5.27}$$

The objective function of $(Q*)$ can further be rewritten as

$$\sum_{i=1}^{n} \left[\frac{\mu}{\delta - y_i} - \sigma/\gamma + 2 \right]$$

with $\mu = (\sigma\delta)/\gamma + 1 - 2\delta) = (1 - \delta) + \delta(\sigma/\gamma - 1) > 0$. Since each term $\frac{1}{\delta - y_i}$ is a convex function of y_i, it follows that $(Q*)$ is a convex program. Furthermore, due to the special structure of this program, an optimal solution of it can be computed in closed form by standard convex mathematical programming methods. Note, however, that if the inequalities (5.23) are replaced by equalities then the problem becomes a truly nonconvex optimization problem.

Example 5.6 (Geometric Programming) An important class of optimization problems called *geometric programming* problems, which have met with many engineering applications (see, e.g., Duffin et al. 1967), have the following general formulation:

$$\min\{g_0(x) |\ g_i(x) \leq 1\ (i = 1, \ldots, m), x = (x_1, \ldots, x_n) > 0\},$$

where each $g_i(x)$ is a *posynomial*, i.e., a function of the form

$$g_i(x) = \sum_{j=1}^{T_i} c_{ij} \prod_{k=1}^{n} (x_k)^{a_{ijk}} \quad i = 0, 1, \ldots, m$$

with $c_{ij} > 0\ \forall i = 1, \ldots, m, j = 1, \ldots, T_i$. Setting

$$x_k = \exp(t_k),\ g_i(x) = \sum_{j=1}^{T_i} c_{ij} \exp\left(\sum_k a_{ijk} t_k \right) := h_i(t),$$

one can rewrite this problem into the form

$$\min\{\log h_0(t) |\ \log h_i(t) \leq 0 \quad i = 1, \ldots, m\}.$$

It is now not hard to see that each function $h_i(t)$ is convex on $\mathbb{R}^n$. Indeed, if the vectors $a^{ij} = (a_{ij1}, \ldots, a_{ijn}) \in \mathbb{R}^n$ are defined, then

$$h_i(t) = \sum_{j=1}^{T_i} c_{ij} e^{\langle a^{ij}, t\rangle}$$

and, since $c_{ij} > 0$, to prove the convexity of $h_i(t)$ it suffices to show that for any vector $a \in \mathbb{R}^n$, the function $e^{\langle a,t\rangle}$ is convex. But the function $\varphi(y) = e^y$ is convex increasing on the real line $-\infty < y < +\infty$, while $l(t) = \langle a, t\rangle$ is linear in t, therefore, by Proposition 2.8, the composite function $\varphi(l(t)) = e^{\langle a,t\rangle}$ is convex.

Thus a geometric program is actually a convex program of a special form. Very efficient specialized methods exist for solving this problem (Duffin et al. 1967). Note, however, that if certain c_{ij} are negative, then the problem becomes a difficult dc optimization problem (see, e.g., McCormick 1982).

Example 5.7 (Generalized Linear Programming) A *generalized linear program* is a problem of the form:

$$\text{(GLP)} \qquad \left| \begin{array}{l} \min cx : \\ \sum_{j=1}^n y^j x_j = b \\ y^j \in Y_j \quad j = 1, \ldots, n \\ x \geq 0, \end{array} \right. \tag{5.28}$$

where $x \in \mathbb{R}^n, y^j \in \mathbb{R}^m (j = 1, \ldots, n),\ b \in \mathbb{R}^m$, and $Y_j \subset \mathbb{R}^m$ $(j = 1, \ldots, n)$ are polyhedrons. The name comes from the fact that this problem reduces to an ordinary linear program when the vectors $y^1, \ldots, y^n$ are constant, i.e., when each Y_j is a singleton. Since y^j are now variables, the constraints are bilinear. In spite of that, the analogy of the problem with ordinary linear programs suggests an efficient algorithm for solving it which can be considered as an extension of the classical simplex method (Dantzig 1963).

A special case of (GLP) is the following problem which arises, e.g., from certain applications in agriculture:

$$\text{(SP)} \qquad \min \sum_{i=1}^n c_i x_i : \tag{5.29}$$

$$x \in D \subset \mathbb{R}^n \tag{5.30}$$

$$x_i = u_i v_i \quad i = 1, \ldots, p \leq n, \tag{5.31}$$

$$u \in [r, s] \subset \mathbb{R}^p_{++}, \quad v \in S \subset \mathbb{R}^p_+, \tag{5.32}$$

where D, S are given polytopes in $\mathbb{R}^n, \mathbb{R}^p$, respectively. Setting $u_i = x_i, v_i = 1$ for all $i > p$ and substituting (5.31) into (5.29) and (5.30), we see that the problem is to minimize a bilinear function in u, v under linear and bilinear constraints. However, this natural transformation would lead to a complicated nonlinear problem, whereas a much simpler solution method is by reducing it to a (GLP) as follows. Let

$$D = \left\{ v \in \mathbb{R}^p_+ | \sum_{j=1}^{p} \alpha_{ij} v_j = e_i \ (i = 1, \ldots, m) \right\}, \tag{5.33}$$

$$S = \left\{ x \in \mathbb{R}^n_+ | \sum_{j=1}^{n} \beta_{kj} x_j = d_k \ (k = 1, \ldots, l) \right\}. \tag{5.34}$$

By rewriting the constraints (5.30)–(5.32) as

$$\begin{aligned} &\sum_{j=1}^{p} \alpha_{ij} \frac{x_j}{u_j} = e_i \quad i = 1, \ldots, m \\ &\sum_{j=1}^{n} \beta_{kj} x_j = d_k \quad k = 1, \ldots, l \\ &r_j \le u_j \le s_j \quad j = 1, \ldots, p \\ &x_j \ge 0 \quad j = 1, \ldots, n, \end{aligned}$$

we obtain a (GLP) where $b = \begin{bmatrix} e \\ d \end{bmatrix} \in \mathbb{R}^{m+l}$,

$$y^j \in \mathbb{R}^{m+l}, \quad y^j_h = \begin{cases} \alpha_{hj}/u_j & h = 1, \ldots, m, \ j = 1, \ldots, p \\ 0 & h = 1, \ldots, m, \ j = p+1, \ldots, n \\ \beta_{h-m,j} & h = m+1, \ldots, m+l, \ j = 1, \ldots, n \end{cases}$$

and Y_j is a polyhedron in $\mathbb{R}^{m+l}$ defined by

$$\alpha_{hj}/r_j \le y^j_h \le \alpha_{hj}/s_j \quad h \in I^+_j \tag{5.35}$$

$$\alpha_{hj}/s_j \le y^j_h \le \alpha_{hj}/r_j \quad h \in I^-_j \tag{5.36}$$

$$I^+_j = \{i | \alpha_{ij} > 0\}, \quad I^-_j = \{i | \alpha_{ij} < 0\}. \tag{5.37}$$

An interesting feature worth noticing of this problem is that, despite the bilinear constraints, the set Ω of all x satisfying (5.31), (5.32) is a polytope, so that (SP) is actually a *linear program* (Thieu 1992). To see this, for every fixed $u \in [r, s]$ denote by $\tilde{u} : \mathbb{R}^p \to \mathbb{R}^p$ the affine map that carries $v \in \mathbb{R}^p$ to $\tilde{u}(v) \in \mathbb{R}^p$ with

$\tilde{u}_i(v) = u_i v_i,\ i = 1, \dots, p$. Then the set $\tilde{u}(S)$ is a polytope, as it is the image of a polytope under an affine map. Now let $u^j, j \in J$, be the vertices of the rectangle $[r, s]$ and G be the convex hull of $\cup_{j\in J}\tilde{u}^j(S)$, $\overline{G} = \{x \in \mathbb{R}^n |\ (x_1, \dots, x_p) \in G\}$. If $x \in \overline{G}$, then $x_i = \sum_{j\in J} \lambda_j \tilde{u}^j_i(v^j), i = 1, \dots, p$ with $v^j \in S,\ \lambda_j \geq 0, \sum_{j\in J} \lambda_j = 1$ (Exercise 7, Chap. 1), and so

$$x_i = \sum_{j\in J} \lambda_j u^j_i v^j_i, \quad i = 1, \dots, p. \tag{5.38}$$

Since $u^j_i > 0, v^j_i \geq 0$, one can have $\sum_{j\in J} \lambda_j v^j_i = 0$ only if $x_i = 0$, so one can write $x_i = u_i v_i (i = 1, \dots, p)$, with $v = \sum_{j\in J} \lambda_j v^j \in S$, and

$$u_i = \begin{cases} \dfrac{x_i}{v_i} & \text{if } v_i > 0 \\ r_i & \text{otherwise.} \end{cases}$$

Clearly, $r_i \leq u_i \leq s_i,\ i = 1, \dots, p$ because $r_i v_i \leq x_i \leq s_i v_i$ by (5.38). Thus $x \in G$ implies that $x \in \Omega$. Conversely, if $x \in \Omega$, i.e., $x_i = u_i v_i,\ i = 1, \dots, p$ for some $u \in [r, s], v \in S$, then $u = \sum_{j\in J} \lambda_j u^j$ with $\lambda_j \geq 0, \sum_{j\in J} \lambda_j = 1$, hence $x_i = \sum_{j\in J} \lambda_j (u^j_i v_i),\ i = 1, \dots, p$, which means that $x \in G$. Therefore $\Omega = G$, and so the feasible set of (SP) is a polytope. Though this result implies that (*SP*) can be solved as a linear program, the determination of the constraints in this linear program [i.e., the linear inequalities describing the feasible set] may be cumbersome computationally. It should also be noted that the problem will become highly nonconvex if the rectangle $[r, s]$ in the condition $u \in [r, s]$ is replaced by an arbitrary polytope.

Example 5.8 (Semidefinite Programming) In recent years *linear matrix inequality* (LMI) techniques have emerged as powerful numerical tools for the analysis and design of control systems. Recall from Sect. 2.12.2 that an LMI is any matrix inequality of the form

$$Q(x) := Q_0 + \sum_{i=1}^{n} x_i Q_i \preceq 0, \tag{5.39}$$

where $Q_i, i = 0, 1, \dots, m$ are real symmetric $m \times m$ matrices, and for any two real symmetric $m \times m$ matrices A, B the notation $A \preceq B$ ($A \prec B$, resp.) means that the matrix $A - B$ is negative semidefinite (definite, resp.), i.e., the largest eigenvalue of $A - B$, $\lambda_{\max}(A - B)$, is nonpositive (negative, resp.). A *semidefinite program* is a linear program under LMI constraints, i.e., a problem of the form

$$\min\{cx | \quad Q(x) \preceq 0\}, \tag{5.40}$$

where $Q(x)$ is as in (5.39). Obviously, $Q(x)$ is negative semidefinite if and only if

$$\sup\{u^T Q(x)u|\ u \in \mathbb{R}^m\} \leq 0,$$

Since for fixed u the function $\varphi_u(x) := u^T Q(x)u$ is affine in x, it follows that the set $G = \{x \in \mathbb{R}^n|\ \varphi_u(x) \leq 0\ \forall u \in \mathbb{R}^m\}$ (the feasible set of (5.40)) is convex. Therefore, the problem (5.40) is a convex program which can be solved by specialized interior point methods (see, e.g., Boyd and Vandenberghe 2004). Note that the problem $\min\{cx|\ Q(x) \succeq 0\}$ is also a semidefinite program , since $Q(x) \succeq 0$ if and only if $-Q(x) \preceq 0$.

Many problems from control theory reduce to nonconvex global optimization problems under LMI constraints. For instance, the "robust performance problem" (Apkarian and Tuan 1999) can be formulated as that of minimizing γ subject to

$$\sum_{k=1}^{N} z_i A_{0i} + \sum_{i=1}^{m}(u_i A_i + v_i A_{i+m}) + \gamma B \prec 0 \tag{5.41}$$

$$u_i = \gamma x_i,\ x_i v_i = \gamma \quad i = 1, \ldots, m \tag{5.42}$$

$$x \in [a, b],\ z \in [-d, d]^N, \tag{5.43}$$

where A_{0i}, A_i, B are real symmetric matrices (the constraint (5.41) is convex, while (5.42) are bilinear).

5.9 Exercises

1 Let C, D be two compact convex sets in $\mathbb{R}^2$ such that $D \subset \text{int}C$. Show that the distance $d(x) = \inf\{\|x - y\| : y \notin C\}$ from a point $x \in D$ to the boundary of C is a concave function. Suppose C is a lake, and D an island in the lake. Show that finding the shortest bridge connecting D with the mainland (i.e., $\mathbb{R}^2 \setminus \text{intC}$) is a concave minimization problem. If C, D are merely compact (not necessarily convex), show that this is a dc optimization problem.

2 (Weber problem with attraction and repulsion)

1. A single facility is to be constructed to serve n users located at points $a^j \in D$ on the plane (where D is a rectangle in $\mathbb{R}^2$). If the facility is located at $x \in D$, then the attraction of the facility to user j is $q_j(h_j(x))$, where $h_j(x) = \|x - a^j\|$ is the distance from x to a^j and $q_j : R \to \mathbb{R}$ is a convex decreasing function (the farther x is away from a^j the less attractive it looks to user j). Show that the problem of finding the location of the facility with maximal total attraction $\sum_{j=1}^{n} q_j(h_j(x))$ is a dc optimization problem.
2. In practice, some of the points a^j may be repulsion rather than attraction points for the facility. For example, there may be in the area D a garbage dump, or

sewage plant, or nuclear plant, and one may wish the facility to be located as far away from these points as possible. If J_1 is the set of attraction points, J_2 the set of repulsion points, then in typical situations one may have

$$q_j[h_j(x)] = \begin{cases} \alpha_j - w_j\|x - a^j\| & j \in J_1, \\ w_j e^{-\theta_j\|x-a^j\|} & j \in J_2, \end{cases}$$

where $\alpha_j > 0$ is the maximal attraction of point $j \in J_1$, $\theta_j > 0$ is the rate of decay of the repulsion of point $j \in J_2$ and $w_j > 0$ for all j. Show that in this case the problem of maximizing the function

$$\sum_{j\in J_1} q_j[h_j(x)] - \sum_{j\in J_2} q_j[h_j(x)]$$

is still a dc optimization problem. (When $J_2 = \emptyset$ the problem is known as Weber's problem).

3 In the previous problem, show that if there is an index $i_* \in J_1$ such that $w_{i_*} \geq \sum_{j=1, j\neq i_*}^{n} w_j$ then a^{i_*} is the optimal location of the facility ("majority theorem").

4 (Competitive location) In Exercise 2, suppose that $J_2 = \emptyset$ (no repulsion point), and the attraction of the new facility to user j is defined as follows. There exist already in the area several facilities and the new facility is meant to compete with these existing facilities. If $\delta_j > 0$ ($\eta_j > \delta_j$, resp.) is the distance from user j to the nearest (farthest, resp.) existing facility, then the attraction of the new facility to user j is a function $q_j(t)$ of the distance t from the new facility to user j such that

$$q_j(t) = \begin{cases} \alpha_j - w_j t & 0 \leq t \leq \delta_j \\ (\alpha_j - w_j\delta_j)\left(1 - \frac{t-\delta_j}{\eta_j - \delta_j}\right) & \delta_j \leq t \leq \eta_j \\ 0 & t \geq \eta_j \end{cases}$$

(for user j the new facility is attractive as long as it is closer than any existing facility, but this attraction quickly decreases when the new facility is farther than some of the existing facilities and becomes zero when it is farther than the farthest existing facility). Denoting by $h_j(x)$ the distance from the unknown location x of the new facility to user j, show that to maximize the total attraction one has again to solve a dc optimization problem.

5 In Example 5.3 on engineering design prove that $r(x) = \max\{r|\ B(x,r) \subset M\} = \inf\{\|x - y\| : y \in \partial M\}$ (distance from x to ∂M). Derive that $r(x)$ is a dc function (cf Exercise 1, Chap. 4), and hence that finding the maximal ball contained in a given compact set $M \subset \mathbb{R}^n$ reduces to maximizing a dc function under a dc constraint.

6 Consider a system of disjunctive convex inequalities $x \in \bigcup_{i=1}^{n} C_i$, where $C_i = \{x : g_i(x) \leq 0\}$. Show that this system is equivalent to a single dc inequality $p(x) - q(x) \leq 0$ and find the two convex functions $p(x), q(x)$.

7 In Example 5.4 introduce the new variables q_{il} such that $x_{il} = q_{il} \sum_j y_{lj}$ and transform the pooling problem into a problem in q_{il}, y_{lj}, z_{ij} (eliminate the variables x_{il}).

8 (Optimal design of a water distribution network) Assuming that a water distribution network has a single source, a single demand pattern, and a new pumping facility at the source node, the problem is to determine the pump pressure H, the flow rate q_i and the head losses J_i along the arcs $i = 1, \ldots, s$, so as to minimize the cost (see, e.g., Fujiwara and Khang 1990):

$$f(q, H, J) := b_1 \sum_i L_i (KL_i/J_i)^{\beta/\alpha} (q_i/c)^{\beta\lambda/\alpha} + b_2 |a_1|^{e_1} H^{e_2} + b_3 |a_1| H$$

under the constraints

$$\begin{array}{lll}
\sum_{i \in \text{in}(k)} q_i - \sum_{i \in \text{out}(k)} q_i = a_k & k = 1, \ldots, n & \text{(flow balance)} \\
\sum_{i \in \text{loop } p} \pm J_i = 0 & p = 1, \ldots, m & \text{(head loss balance)} \\
\sum_{i \in r(k)} \pm J_i \leq H + h_1 - h_k^{\min} & p = 1, \ldots, m & \text{(hydraulic requirement)}
\end{array}$$

$$\left.\begin{array}{l} q_i^{\max} \geq q_i \geq q_i^{\min} \geq 0 \quad i = 1, \ldots, s \\ H^{\max} \geq H \geq H^{\min} \geq 0 \end{array}\right\} \text{(bounds)}$$

$$\left.\begin{array}{l} KL_i (q_i/c)^{\lambda} / d_{\max}^{\alpha} \leq J_i \\ J_i \leq KL_i (q_i/c)^{\lambda} / d_{\min}^{\alpha} \end{array}\right\} i = 1, \ldots, s \quad \text{(Hazen–Williams equations)}$$

(L_i is the length of arc i, K and c are the Hazen–Williams coefficients, $-a_1$ is the supply water rate at the source and $b_1, b_2, b_3, e_1, e_2, \beta$ are constants such that $0 < e_1, e_2 < 1$ and $1 < \beta < 2.5, \lambda = 1.85$ and $\alpha = 4.87$). By the change of variables $p_i = \log q_i, \pi_i = \log J_i$, transform the first term of the sum defining $f(q, H, J)$ into a convex function of p, π and reduce the problem to a dc optimization problem.

9 Solve the problem in Example 5.5.

10 Solve the problem

$$\begin{array}{l}
\text{minimize} \quad 3x_1 + 4x_2 \quad \text{subject to} \\
y_{11} x_1 + 2x_2 + x_3 = 5; \quad y_{12} x_1 + x_2 + x_4 = 3 \\
x_1, x_2 \geq 0; \ y_{11} - y_{12} \leq 1; \ 0 \leq y_{11} \leq 2; \ 0 \leq y_{12} \leq 2.
\end{array}$$

Chapter 6
General Methods

Methods for solving global optimization problems combine in different ways some basic techniques: partitioning, bounding, cutting, and approximation by relaxation or restriction. In this chapter we will discuss two general methods: branch and bound (BB) and outer approximation (OA).

A basic operation involved in a BB procedure is partitioning, more precisely, successive partitioning.

6.1 Successive Partitioning

Partitioning is a technique consisting in generating a process of subdivision of an initial set (which is either a simplex or a cone, or a rectangle) into several subsets of the same type (simplices, cones, rectangles, respectively) and reiterating this operation for each of the partition sets as many times as needed. There are three basic types of partitioning : simplicial, conical, and rectangular.

6.1.1 Simplicial Subdivision

Consider a $(p-1)$-simplex $S = [u^1, \ldots, u^p] \subset \mathbb{R}^n$ and an arbitrary point $v \in S$, so that

$$v = \sum_{i=1}^{p} \lambda_i u^i, \quad \sum_{i=1}^{p} \lambda_i = 1, \quad \lambda_i \geq 0 \ (i = 1, \ldots, p).$$

Let $I = \{i| \ \lambda_i > 0\}$. It is a simple matter to check that for every $i \in I$, the points $u^1, \ldots, u^{i-1}, v, u^{i+1}, \ldots, u^p$ are vertices of a simplex $S_i \subset S$ and that:

H. Tuy, *Convex Analysis and Global Optimization*, Springer Optimization and Its Applications 110, DOI 10.1007/978-3-319-31484-6_6

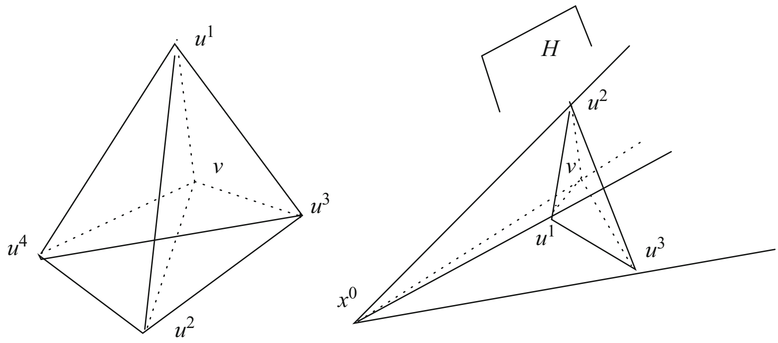


Fig. 6.1 Simplicial and conical subdivision

$$(\text{ri}S_i) \cap (\text{ri}S_j) = \emptyset \quad \forall j \neq i; \quad \bigcup_{i \in I} S_i = S.$$

We say that the simplices $S_i, i \in I$ form a *radial partition(subdivision)* of the simplex S via v (Fig. 6.1). Each S_i will be referred to as a partition member or a "child" of S. Clearly, the partition is *proper*, i.e., consists of at least two members, if and only if v does not coincide with any u^i. An important special case is when v is a point of a *longest* edge of the simplex S, i.e., $v \in [u^k, u^h]$, where

$$\|u^k - u^h\| = \max\{\|u^i - u^j\| | \ i < j; \ i, j = 1, \ldots, p\}$$

($\|\cdot\|$ denotes any given norm in $\mathbb{R}^n$, for example, $\|x\| = \max\{|x_1|, \ldots, |x_n|\}$). If $v = \alpha u^k + (1 - \alpha)u^h$ with $0 < \alpha \leq 1/2$ then the partition is called a *bisection* of ratio α (S is divided into two subsimplices such that the ratio of the volume of the smaller subsimplex to that of S equals α). When $\alpha = 1/2$, the bisection is called *exact*.

Consider now an infinite sequence of nested simplices $S_1 \supset S_2 \supset \cdots \supset S_k \supset \cdots$ such that S_{k+1} is a child of S_k in a partition via some $v^k \in S_k$. Such a sequence is called a *filter* (of simplices) and the simplex $S_\infty = \bigcap_{k=1}^\infty S_k$ the *limit* of the filter. Let $\delta(S)$ denote the diameter of S, i.e., the length of its longest edge. The filter is said to be *exhaustive* if $\cap_{k=1}^\infty S_k$ is a singleton, i.e., if $\lim_{k \to \infty} \delta(S_k) = 0$ (the filter shrinks to a single point). An important question about a filter concerns the conditions under which it is exhaustive.

Lemma 6.1 (Basic Exhaustiveness Lemma) *Let* $S_k = [u^{k1}, \ldots, u^{kn}], k = 1, 2, \ldots,$ *be a filter of simplices such that* S_{k+1} *is a child of* S_k *in a subdivision via* $v^k \in S_k$. *Assume that:*

(i) *For infinitely many* k *the subdivision of* S_k *is a bisection;*
(ii) *There exists a constant* $\rho \in (0, 1)$ *such that for every* k*:*

$$\max\{\|v^k - u^{ki}\| : \ i = 1, \ldots, n\} \leq \rho\delta(S_k).$$

Then the filter is exhaustive.

Proof Denote $\delta(S_k) = \delta_k$. Since $\delta_{k+1} \leq \delta_k$, it follows that δ_k tends to some limit δ as $k \to \infty$. Arguing by contradiction, suppose $\delta > 0$, so there exists t such that $\rho\delta_k < \delta$ for all $k \geq t$. By assumption (ii) we can write

$$\max\{\|v^k - u^{ki}\| : i = 1, \ldots, n\} \leq \rho\delta_k < \delta \leq \delta_k, \quad \forall k \geq t. \tag{6.1}$$

Let us color every vertex of S_t "black" and for any $k > t$ let us color "white" every vertex of S_k which is not black. Clearly, every white vertex of S_k must be a point v^h for some $h \in \{t, \ldots, k-1\}$, hence, by (6.1), $\max\{\|v^h - u^{hi}\| : i = 1, \ldots, n\} \leq \rho\delta_h < \delta \leq \delta_k$. So any edge of S_k incident to a white vertex must have a length less than δ_k and cannot be a longest edge of S_k. In other words, for $k \geq t$ a longest edge of S_k must have two black endpoints. Now denote $K = \{k \geq t | \ S_k \text{ is bisected}\}$. Then for $k \in K$, v^k belongs to a longest edge of S_k, hence, according to what has just been proved, to an edge of S_k joining two black vertices. Since v^k becomes a white vertex of S_{k+1}, it follows that for $k \in K$, S_{k+1} has one white vertex more than S_k. On the other hand, in the whole subdivision process the number of white vertices never decreases. That means, after one bisection the number of white vertices increases at least by 1. Therefore, after a certain number of bisections, we come up with a simplex S_k with only white vertices. Then, according to (6.1), every edge of S_k has length less than δ, conflicting with $\delta \leq \delta_k = \delta(S_k)$. Therefore, $\delta = 0$, as was to be proved. □

It should be emphasized that condition (ii) is essential for the exhaustiveness of the filter (see Exercise 1 of this chapter). This condition implies in particular that the bisections mentioned in (i) are bisections of ratio no less than $1 - \rho > 0$. A filter of simplices satisfying (i) but not necessarily (ii) enjoys the following weaker property of "semi-exhaustiveness" which, however, may sometimes be more useful.

Theorem 6.1 (Basic Simplicial Subdivision Theorem) *Let $S_k = [u^{k1}, \ldots, u^{kn}]$, $k = 1, 2, \ldots$, be a filter of simplices such that S_{k+1} is a child of S_k in a subdivision via a point $v^k \in S_k$. Then*

(i) *Each sequence $\{u^{ki}, k = 1, 2, \ldots\}$ has an accumulation point u^i such that $S_\infty := \cap_{k=1}^{\infty} S^k$ is the convex hull of $u^1, \ldots, u^n$;*
(ii) *The sequence $\{v^k, k = 1, 2, \ldots\}$ has an accumulation point $v \in \{u^1, \ldots, u^n\}$;*
(iii) *If for infinitely many k the subdivision of S_k is a bisection of ratio no less than $\alpha > 0$ then an accumulation point of the sequence $\{v^k, k = 1, 2, \ldots\}$ is a vertex of $S_\infty = [u^1, \ldots, u^n]$.*

Proof (i) Since S_1 is compact and $\{u^{ki}, k = 1, 2, \ldots\} \subset S_1$ there exists a subsequence $\{k_s\} \subset \{1, 2, \ldots\}$ such that for every $i = 1, \ldots, n$, $u^{k_s i} \to u^i$ as $s \to +\infty$. If $x \in S_\infty = [u^1, \ldots, u^n]$ then there exist $\alpha_i \geq 0$ such that $\sum_{i=1}^n \alpha_i = 1$ and $x = \sum_{i=1}^n \alpha_i u^i = \lim_{s\to+\infty} x_s$, for $x_s = \sum_{i=1}^n \alpha_i u^{k_s i} \in S^{k_s}$, hence $x \in S^{k_s}$. Since this holds for every s and $\{S^k\}$ is a nested sequence, it follows that $x \in \cap_{k=1}^{\infty} S^k$. So $S_\infty \subset \cap_{k=1}^{\infty} S^k$. Conversely for any s if $x \in S^{k_s}$, then with $\alpha_{si} \geq 0$ satisfying $\sum_{i=1}^{n} \alpha_{si} = 1$ we have $x = \sum_{i=1}^{n} \alpha_{si} u^{k_s i} \to \sum_{i=1}^{n} \alpha_i u^i$ as $s \to +\infty$, hence $x \in [u^1, \ldots, u^n] = S_\infty$. Therefore, $S_\infty = \cap_{k=1}^{\infty} S^k = [u^1, \ldots, u^n]$.

(ii) S^k is a child of S^{k-1} in a subdivision via v^{k-1}, so a vertex of S^k, say u^{ki_k}, must be v^{k-1}. Since $i_k \in \{1, \ldots, n\}$ there exists a subsequence $\{k_s\} \subset \{1, 2, \ldots\}$ such that i_{k_s} is the same for all s, for instance, $i_{k_s} = 1\ \forall s$. Therefore $v^{k_s - 1} = u^{k_s 1}\ \forall s$. Since u^1 is an accumulation point of the sequence $\{u^{k1}\}$ we can assume that $u^{k_s 1} \to u^1$ as $s \to +\infty$. Hence, $v^{k_s - 1} \to u^1$ as $s \to +\infty$.

(iii) This is obviously true if $\delta(S_\infty) = 0$, i.e., if S_∞ is a singleton. Consider the case $\delta(S_\infty) > 0$ which by Lemma 6.1 can occur only if condition (ii) of this Lemma does not hold, i.e., if for every $s = 1, 2, \ldots$ there is a k_s such that

$$\max\{\|v^{k_s} - u^{k_s i}\| \mid i = 1, \ldots, n\} \geq \rho_s \delta(S_{k_s}),$$

where $\rho_s \to 1 (s \to +\infty)$. Since S_1 is compact, by taking a subsequence if necessary we may assume that $v^{k_s} \to v, u^{k_s i} \to u^i (i = 1, \ldots, n)$. Then letting $s \to \infty$ in the previous inequality yields $\max_{i=1,\ldots,n} \|v - u^i\| = \delta(S_\infty)$. Since for each i the function $x \mapsto \|x - u^i\|$ is strictly convex, so is the function $x \mapsto \max_{i=1,\ldots,n} \|x - u^i\|$. Therefore, a maximizer of the latter function over the polytope $\mathrm{conv}\{u^1, \ldots, u^n\}$ must be a vertex of it: $v \in \mathrm{vert} S_\infty$. □

Incidentally, the proof has shown that one can have $\delta(S_k) > 0$ only if infinitely many subdivisions are not bisections of ratio $\geq \alpha$. Therefore:

Corollary 6.1 *A filter of simplices $S_k, k = 1, 2, \ldots$ such that every S_{k+1} is a child of S_k via a bisection of ratio no less than $\alpha > 0$ is exhaustive.* □

A finite collection $\mathscr{N}$ of simplices covering a set Ω is called a (simplicial) *net* for Ω. A net for Ω is said to be a *refinement* of another net if it is obtained by subdividing one or more members of the latter and replacing them with their partitions. A (simplicial) subdivision process for Ω is a sequence of (simplicial) nets $\mathscr{N}_1, \mathscr{N}_2, \ldots$, each of which, except the first one, is a refinement of its predecessor.

Lemma 6.2 *An infinite subdivision process $\mathscr{N}_1, \mathscr{N}_2, \ldots$, generates at least one filter.*

Proof Since $\mathscr{N}_1$ has finitely many members, at least one of these, say S_{k_1}, has infinitely many descendants. Since S_{k_1} has finitely many children, one of these, say S_{k_2}, has infinitely many descendants. Continuing this way we shall generate an infinite nested sequence which is a filter. □

A subdivision process is said to be *exhaustive* if every filter $\{S_k, k = 1, 2, \ldots,\}$ generated by the process is exhaustive. From Corollary 6.1 it follows that a subdivision process consisting only of bisections of ratio at least α is exhaustive.

6.1.2 Conical Subdivision

Turning to conical subdivision, consider a fixed cone $M_0 \subset \mathbb{R}^n$ vertexed at x^0 (throughout this chapter, by *cone* in $\mathbb{R}^n$ we always mean a polyhedral cone having exactly n edges). To simplify the notation assume $x^0 = 0$. Let H be a fixed hyperplane meeting each edge of M_0 at a point distinct from 0. For any cone $M \subset M_0$ the set $H \cap M$ is an $(n-1)$-simplex $S = [u^1, \dots, u^n]$ not containing 0. We shall refer to this simplex S as the *base* of the cone M. Then any subdivision of the simplex S via a point $v \in S$ induces in an obvious way a *subdivision* (splitting) of the cone M into subcones, via the ray through v (Fig. 6.1). A subdivision of M is called *bisection* (of ratio α) if it is induced by a bisection (of ratio α) of S. A filter of cones, i.e., an infinite sequence of nested cones each of which is a child of its predecessor by a subdivision, is said to be *exhaustive* if it is induced by an exhaustive filter of simplices, i.e., if their intersection is a ray.

Theorem 6.2 (Basic Conical Subdivision Theorem) *Let C be a compact convex set in $\mathbb{R}^n$, containing 0 in its interior, $\{M_k,\ k = 1, 2, \dots\}$ a filter of cones vertexed at 0 and having each exactly n edges. For each k let $z^{k1}, \dots, z^{kn}$ be the intersection points of the n edges of M_k with ∂C (the boundary of C). If for each k M_{k+1} is a child of M_k in a subdivision via the ray through a point $q^k \in [z^{k1}, \dots, z^{kn}]$ then one accumulation point of the sequence $\{q^k, k = 1, 2, \dots,\}$ belongs to ∂C.*

Proof Let $S_k = [u^{k1}, \dots, u^{kn}]$ be the base of M_k. Then S_{k+1} is a child of S_k in a subdivision via a point $v^k \in S_k$, with v^k being the point where S_k meets the halfline from 0 through q^k. For every $i = 1, \dots, n$ let u^i be an accumulation point of the sequence $\{u^{ki}, k = 1, 2, \dots\}$. By Theorem 6.1, (ii), there exists u^i such that $v^{k_s} \to u^i$ for some subsequence $\{k_s\} \subset \{1, 2, \dots\}$. Therefore, $q^{k_s} \to z^i$ where $z^i = \lim_{s\to\infty} z^{k_s i} \in \partial C$ (Fig. 6.2). □

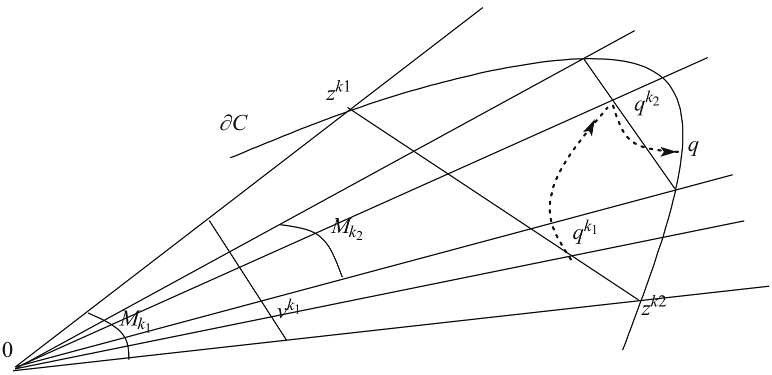


Fig. 6.2 Basic conical subdivision theorem

6.1.3 Rectangular Subdivision

A rectangle in $\mathbb{R}^n$ is a set of the form $M = [r, s] = \prod_{i=1}^{n}[r_i, s_i] = \{x \in \mathbb{R}^n | \ r_i \leq x_i \leq s_i \ (i = 1, \ldots, n)\}$. There are several ways to subdivide a rectangle $M = [r, s]$ into subrectangles. However, in global optimization procedures, we will mainly consider partitions through a hyperplane parallel to one facet of the given rectangle. Specifically, given a rectangle M, a partition of M is defined by giving a point $v \in M$ together with an index $j \in \{1, \ldots, n\}$: the partition sets are then the two subrectangles M_- and M_+ determined by the hyperplane $x_j = v_j$, namely:

$$M_- = \{x | \ r_j \leq x_j \leq v_j, \ r_i \leq x_i \leq s_i(i \neq j)\};$$

$$M_+ = \{x | \ v_j \leq x_j \leq s_j, \ r_i \leq x_i \leq s_i(i \neq j)\}.$$

We will refer to this partition as a *partition via* (v, j) (Fig. 6.3). A nice property of this partition method is expressed in the following:

Lemma 6.3 *Let* $\{M_k := [r^k, s^k], k \in K\}$ *be a filter of rectangles such that* M_k *is partitioned via* (v^k, j_k). *Then there exists a subsequence* $K_1 \subset K$ *such that* $j_k = j_0 \ \forall k \in K_1$ *and, as* $k \to +\infty, k \in K_1$,

$$r^k_{j_0} \to r_{j_0}, \ s^k_{j_0} \to s_{j_0}, \ v^k_{j_0} \to \overline{v}_{j_0} \in \{r_{j_0}, s_{j_0}\}. \tag{6.2}$$

Proof Without loss of generality we can assume $K = \{1, 2, \ldots\}$. Since $j_k \in \{1, \ldots, n\}$ there exists a j_0 such that $j_k = j_0$ for all k forming an infinite subsequence $K_1 \subset K$. Because of boundedness we may assume $v^k_{j_0} \to \overline{v}_{j_0}$ $(k \to +\infty, k \in K_1)$. Furthermore, since $r^k_{j_0} \leq r^{k+1}_{j_0} \leq r^h_{j_0} \leq s^h_{j_0} \leq s^{k+1}_{j_0} \leq s^k_{j_0}$ for all $h > k$, we have $r^k_{j_0} \to r_{j_0}, r^{k+1}_{j_0} \to r_{j_0}$, and $s^k_{j_0} \to s_{j_0}, s^{k+1}_{j_0} \to s_{j_0}$ $(k \to +\infty)$. But $v^k_{j_0} \in \{r^{k+1}_{j_0}, s^{k+1}_{j_0}\}$, hence, by passing to the limit, $\overline{v}_{j_0} \in \{r_{j_0}, s_{j_0}\}$. $\square$

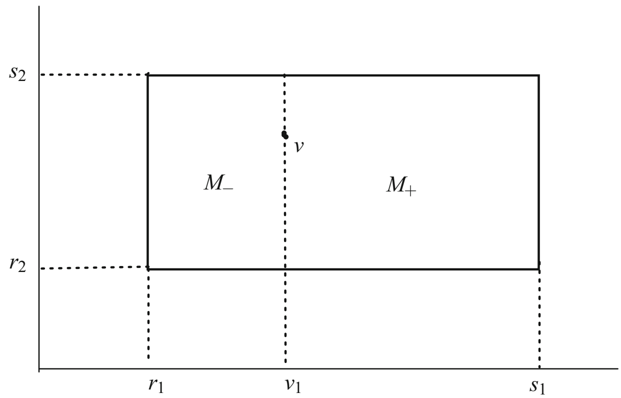


Fig. 6.3 Rectangular subdivision

Theorem 6.3 (Basic Rectangular Subdivision Theorem) *Let* $\{M_k = [r^k, s^k], k \in K\}$ *be a filter of rectangles as in Lemma 6.3, where*

$$j_k \in \text{argmax}\{\eta_j^k | j = 1, \ldots, n\}, \quad \eta_j^k = \min\{v_j^k - r_j^k, s_j^k - v_j^k\}. \tag{6.3}$$

Then at least one accumulation point of the sequence of subdivision points $\{v^k\}$ *coincides with a corner of the limit rectangle* $M_\infty = \cap_{k=1}^{+\infty} M_k$.

Proof Let K_1 be the subsequence mentioned in Lemma 6.3. One can assume that $r^k \to r,\ s^k \to s, v^k \to \overline{v}$ as $k \to +\infty, k \in K_1$. From (6.2) and (6.3) it follows that for all $j = 1, \ldots, n$:

$$\eta_j^k \leq \eta_{j_0}^k \to 0 \ (k \to +\infty, k \in K_1),$$

hence $\overline{v}_j \in \{r_j, s_j\}\ \forall j$, which means that $\overline{v}$ is a corner of $M_\infty = [r, s]$. □

The subdivision of a rectangle $M = [r, s]$ via (v, j) is called a *bisection of ratio* α if the index j corresponds to a longest side of M and v is a point of this side such that $\eta_j = \min(v_j - r_j, s_j - v_j) = \alpha(s_j - r_j)$.

Corollary 6.2 *Let* $\{M_k := [r^k, s^k]\}$ *be any filter of rectangles such that* M_{k+1} *is a child of* M_k *in a bisection of ratio no less than a positive constant* $\alpha > 0$. *Then the filter is exhaustive, i.e., the diameter* $\delta(M_k)$ *of* M_k *tends to zero as* $k \to \infty$.

Proof Indeed, by the previous theorem, one accumulation point $\overline{v}$ of the sequence of subdivision points v^k must be a corner of the limit rectangle. This can occur only if $\delta(M_k) \to 0$. □

Theorem 6.4 (Adaptive Rectangular Subdivision Theorem) *For each* $k = 1, 2, \ldots$, *let* u^k, v^k *be two points in the rectangle* $M_k = [r^k, s^k]$ *and let* $w^k = \frac{1}{2}(u^k + v^k)$. *Define* j_k *by*

$$|u_{j_k}^k - v_{j_k}^k| = \max_{i=1,\ldots,n} |u_i^k - v_i^k|.$$

If for every $k = 1, 2, \ldots,\ M_{k+1}$ *is a child of* M_k *in the subdivision via* (w^k, j_k) *then there exists an infinite subsequence* $K \subset \{1, 2, \ldots\}$ *such that* $\|u^k - v^k\| \to 0$ *as* $k \in K, k \to +\infty$.

Proof By Theorem 6.3 there exists an infinite subsequence $K \subset \{1, 2, \ldots\}$ such that $j_k = j_0\ \forall k \in K,\ w_{j_0}^k \to \overline{w}_{j_0} \in \{r_{j_0}, s_{j_0}\}$ as $k \in K, k \to +\infty$. Since $w_{j_0}^k = \frac{1}{2}(u_{j_0}^k + v_{j_0}^k)$ it follows that $|u_{j_0}^k - v_{j_0}^k| \to 0$ as $k \in K, k \to +\infty$, i.e., $\max_{i=1,\ldots,n} |u_i^k - v_i^k| \to 0$ as $k \in K, k \to_\infty$. Hence, $\|v^k - u^k\| \to 0$ as $k \in K, k \to \infty$. □

The term "adaptive" comes from the adaptive role of this subdivision rule in the context of BB algorithms for solving problems with "nice" constraints, see Sect. 6.2 below. Its aim is to drive the difference $\|u^k - v^k\|$ to zero as $k \to +\infty$. We refer to it as an *adaptive subdivision via* (u^k, v^k).

6.2 Branch and Bound

Consider a problem

$$(P) \qquad \min\{f(x)|\ x \in D\},$$

where $f(x)$ is a real-valued continuous function on $\mathbb{R}^n$ and D is a closed bounded subset of $\mathbb{R}^n$.

In general it is not possible to compute an exact global optimal solution of this problem by a finite procedure. Therefore, in many cases one must be content with only a feasible solution with objective function value sufficiently near to the global optimal value. Given a tolerance $\varepsilon > 0$ a feasible solution $\overline{x}$ such that

$$f(\overline{x}) \leq \min\{f(x)|\ x \in D\} - \varepsilon$$

is called a *global ε-optimal solution* of the problem.

A BB method for solving problem (P) can be outlined as follows:

- Start with a set (simplex or rectangle) $M_0 \supset D$ and perform a (simplicial or rectangular, resp.) subdivision of M_0, obtaining the partition sets $M_i, i \in I$. If a feasible solution is available let $\overline{x}$ be the best feasible solution available and $\gamma = f(\overline{x})$. Otherwise, let $\gamma = +\infty$.
- For each partition set M_i compute a lower bound $\beta(M_i)$ for $f(M_i \cap D)$, i.e., a number $\beta(M_i) = +\infty$ if $M \cap D = \emptyset$, $\beta(M_i) \leq f(M \cap D)$ otherwise. Update the current best feasible solution $\overline{x}$ and the current best feasible value γ.
- Among all the current partition sets select a partition set $M_* \in \mathrm{argmin}_{i\in I}\beta(M_i)$. If $\gamma - \beta(M_*) \leq \varepsilon$ stop: $\overline{x} := CBS$ is a global ε-optimal solution. Otherwise, further subdivide M_*, add the new partition sets to the old ones to form a new refined partitioning of M_0.
- For each new partition set M compute a lower bound $\beta(M)$. Undate $\overline{x}, \gamma$ and repeat the process.

More precisely we can state the following

6.2.1 *Prototype BB Algorithm*

Select a (simplicial or rectangular) subdivision rule and a tolerance $\varepsilon \geq 0$.

Initialization. Start with an initial set (simplex or rectangle) M_0 containing the feasible set D. If any feasible solution is readily available, let $\overline{x}^0$ be the best among such solutions and $\gamma_0 = f(\overline{x}^0)$. Otherwise, set $\gamma_0 = +\infty$. Set $\mathscr{P}_0 = \{M_0\}, \mathscr{S}_0 = \{M_0\}, k = 0$.

Step 1. (Bounding) For each set $M \in \mathscr{P}_k$ compute a lower bound $\beta(M)$ for $f(M \cap D)$, i.e., a number such that

$$\beta(M) = +\infty \text{ if } M \cap D = \emptyset, \quad \beta(M) \leq \inf f(M \cap D) \text{ otherwise.} \tag{6.4}$$

If some feasible solutions have been known thus far, let $\overline{x}^k$ be the best among all of them and set $\gamma_k = f(\overline{x}^k)$. Otherwise, set $\gamma_k = +\infty$.

Step 2. (Pruning) Delete every $M \in \mathscr{S}_k$ for which $\beta(M) \geq \gamma_k - \varepsilon$ and let $\mathscr{R}_k$ be the collection of remaining sets of $\mathscr{S}_k$.

Step 3. (Termination Criterion) If $\mathscr{R}_k = \emptyset$, terminate: $\overline{x}^k$ is a global ε-optimal solution. Otherwise, select $M_k \in \operatorname{argmin}\{\beta(M)|\ M \in \mathscr{R}_k$.

Step 4. (Branching) Subdivide M_k according to the chosen subdivision rule. Let $\mathscr{P}_k$ be the partition of M_k. Reset $\mathscr{S}_k \leftarrow (\mathscr{R}_k \setminus \{M_k\}) \cup \mathscr{P}_k$. If any new feasible solution is available, update the current best feasible solution $\overline{x}^k$ and $\gamma_k = f(\overline{x}^k)$. Set $k \leftarrow k+1$ and return to Step 1.

We say that *bounding is consistent with branching* (briefly, consistent) if

$$\beta(M) - \min\{f(x)|\ x \in M \cap D\} \to 0 \ \text{ as } \operatorname{diam} M \to 0. \tag{6.5}$$

Proposition 6.1 *If the subdivision is exhaustive and bounding is consistent, then the BB algorithm is convergent, i.e., it can be infinite only when* $\varepsilon = 0$ *and then* $\beta(M_k) \to v(\mathrm{P}) := \min\{f(x)|\ x \in D\}$ *as* $k \to +\infty$.

Proof First observe that if the algorithm stops at Step 3 then $\gamma_k < \beta(M) + \varepsilon$ for every $M \in \mathscr{R}_k$, so $\gamma_k < +\infty$, and $\gamma_k = f(\overline{x}^k)$. Since $\overline{x}^k$ is feasible and satisfies $f(\overline{x}^k) < \beta(M)+\varepsilon \leq f(M \cap D)+\varepsilon$ for every $M \in \mathscr{S}_k$, [see (6.4)], $\overline{x}^k$ is indeed a global ε-optimal solution. Suppose now that Step 3 never occurs, so that the algorithm is infinite. Since $\beta(M_k) < \gamma_k + \varepsilon$, it follows from condition (6.4) that $M_k \cap D \neq \emptyset$. So there exists $x^k \in M_k \cap D$. Let $z^k \in M_k \cap D$ be a minimizer of $f(x)$ over $M_k \cap D$ (the points x^k and z^k exist, but may not be known effectively). We now contend that $\operatorname{diam}(M_k) \to 0$ as $k \to +\infty$. In fact, for every partition set M let $\sigma(M)$ be the generation index of M defined by induction as follows: $\sigma(M_0) = 1$; whenever M' is a child of M then $\sigma(M') = \sigma(M) + 1$. It is easily seen that $\sigma(M_k) \to +\infty$ as $k \to +\infty$. But $\sigma(M_k) = m$ means that $M_k = \cap_{i=1}^m S_i$ where $S_1 = M_0, S_m = M_k,\ S_i \in \mathscr{S}_i,\ S_{i+1}$ is a child of S_i. Hence, since the subdivision process is exhaustive, $\operatorname{diam}(M_k) \to 0$ as $k \to +\infty$. We then have $z^k - x^k \to 0$, so z^k and x^k tend to a common limit $x^* \in D$. But by (6.5), $f(z^k) - \beta(M_k) \to 0$, i.e., $\beta(M_k) \to f(x^*)$, hence, in view of (6.4), $f(x^*) \leq v(\mathrm{P})$. Since $x^* \in D$ the latter implies $f(x^*) = v(\mathrm{P})$, i.e., $\beta(M_k) \to v(\mathrm{P})$). Therefore, for k large enough $\gamma_k = f(\overline{x}^k) \leq f(x^k) < \beta(M_k)+\varepsilon$, a contradiction. □

As it is apparent from the above proof, the condition $\beta(M) = +\infty$ if $M \cap D = \emptyset$ in (6.4) is **essential** in a BB algorithm: if it fails to hold, $M_k \cap D$ may be empty for some k, in which case $x^* \in \cap_{h \geq k}(M_h \cap D)$ is infeasible, hence $\beta(M_k) \not\to v(\mathrm{P})$, despite (6.5). In the general case, however, it may be very difficult to decide whether a set $M_k \cap D$ is empty or not. It may also happen that computing a feasible solution is as hard as solving the problem itself, and then no feasible solution is available effectively at each iteration. Under these conditions, when the algorithm stops (case $\varepsilon > 0$) it only provides an approximate value of v(P), namely a value $\beta(M_k) < v(\mathrm{P}) + \varepsilon$; when the algorithm is infinite (case $\varepsilon = 0$) it only provides a sequence of infeasible solutions converging to a global optimal solution.

6.2.2 Advantage of a Nice Feasible Set

All the above difficulties of the BB algorithm disappear when the problem has a *nice* feasible set D. By this we mean that for any simplex or rectangle M a point $x \in M \cap D$ can be computed at cheap cost whenever it exists. More importantly, in most cases when the feasible set is nice a rectangular adaptive subdivision process can be used which ensures a faster convergence than an exhaustive subdivision process does.

Specifically suppose for each rectangle M_k two points x^k, y^k in M_k can be determined such that

$$x^k \in M_k \cap D, \quad y^k \in M_k, \tag{6.6}$$

$$f(y^k) - \beta(M_k) = o(\|x^k - y^k\|). \tag{6.7}$$

Example 6.1 Let $\varphi_{M_k}(x)$ be an underestimator of $f(x)$ tight at $y^k \in M_k$ (i.e., a continuous function $\varphi_{M_k}(x)$ satisfying $\varphi_{M_k}(x) \leq f(x)\ \forall x \in M_k \cap D,\ \varphi_{M_k}(y^k) = f(y^k)$); $\beta(M_k) = \min\{\varphi_{M_k}(x) |\, x \in M_k \cap D\}$, while $x^k \in \operatorname{argmin}\{\varphi_{M_k}(x) |\, x \in M_k \cap D\}$. (As is easily checked, $f(y^k) - \beta(M_k) = \varphi_{M_k}(y^k) - \varphi_{M_k}(x^k) \to 0$ as $x^k - y^k \to 0$.)

Example 6.2 Let $f(x) = f_1(x) - f_2(x),\ x^k \in \operatorname{argmin}\{f_1(x) |\, x \in M_k \cap D\},\ y^k \in \operatorname{argmax}\{f_2(x) |\, x \in M_k\},\ \beta(M_k) = f_1(x^k) - f_2(y^k)$. (If $f_1(x)$ is continuous then $f(y^k) - \beta(M_k) = f_1(y^k) - f_2(y^k) - [f_1(x^k) - f_2(y^k)] = f_1(y^k) - f_1(x^k) \to 0$ as $x^k - y^k \to 0$)

Under these conditions, the following *adaptive bisection* rule can be used for subdividing M_k: Bisect M_k via (w^k, j_k) where $w^k = \frac{1}{2}(x^k + y^k)$ and j_k satisfies

$$\|y^k_{j_k} - x^k_{j_k}\| = \max_{j=1,\dots,n} \|y^k_j - x^k_j\|.$$

Proposition 6.2 *A rectangular BB algorithm using an adaptive bisection rule is convergent, i.e., it produces a global ε-optimal solution in finitely many steps.*

Proof By Theorem 6.4 there exists a subsequence $\{k_\nu\}$ such that

$$y^{k_\nu} - x^{k_\nu} \to 0\ (\nu \to +\infty) \tag{6.8}$$

$$\lim_{\nu \to +\infty} y^{k_\nu} = \lim_{\nu \to +\infty} x^{k_\nu} = x^* \in D. \tag{6.9}$$

From (6.6) and (6.7) we deduce $\beta(M_{k_\nu}) \to f(x^*)$ as $\nu \to +\infty$. But, since $x^* \in D$, we have $\beta(M_{k_\nu}) \leq f(x^*)$. Therefore, $\beta(M_{k_\nu}) \to v(\mathrm{P})\ (\nu \to +\infty)$, proving the convergence of the algorithm. □

We shall refer to a BB algorithm using an adaptive (exhaustive, resp.) subdivision rule as an *adaptive* (*exhaustive* , resp.) BB algorithm.

To sum up, global optimization problems with all the nonconvexity concentrated in the objective function are generally easier to handle by BB algorithm than those with nonconvex constraints. This suggests that in dealing with nonconvex global optimization problems one should try to use valid transformations to shift, in one way or another, all the nonconvexity in the constraints to the objective function. Classical Lagrange multipliers or penalty methods, as well as decomposition methods are various ways to exploit this idea. Most of these methods, however, work under restrictive conditions. In the next chapter we will present a more practical way to handle hard constraints in dc optimization problems.

6.3 Outer Approximation

One of the earliest methods of nonconvex optimization consists in approximating a given problem by a sequence of easier relaxed problems constructed in such a way that the sequence of solutions of these relaxed problems converges to a solution of the given problem. This approach, first introduced in convex programming in the late 1950s (Cheney and Goldstein 1959; Kelley 1960), was later extended under the name of *outer approximation*, to concave minimization under convex constraints (Hoffman 1981; Thieu et al. 1983; Tuy 1983; Tuy and Horst 1988) and to more general nonconvex optimization problems (Mayne and Polak 1984; Tuy 1983, 1987a). Referring to the fact that cutting planes are used to eliminate unfit solutions of relaxed problems, this approach is sometimes also called *a cutting plane method*. It should be noted, however, that in an outer approximation procedure cuts are always conjunctive, i.e., the set that results from the cuts is always the intersection of all the cuts performed.

In this section we will present a flexible framework of outer approximation (OA) applicable to a wide variety of global optimization problems (Tuy 1995b), including DC optimization and monotonic optimization.

Consider the general problem of searching for an element of an unknown set $\Omega \subset \mathbb{R}^n$ (for instance, Ω is the set of optimal solutions of a given problem). Suppose there exist a closed set $G \supset \Omega$ and a family $\mathscr{P}$ of particular sets $P \supset G$, such that for each $P \in \mathscr{P}$ a point $x(P) \in P$ (called *a distinguished* point associated with P) can be defined satisfying the following conditions:

A1. $x(P)$ always exists and can be computed if $\Omega \neq \emptyset$, and whenever a sequence of distinguished points $x^1 = x(P_1),\ x^2 = x(P_2), \ldots$, converges to a point $\overline{x} \in G$ then $\overline{x} \in \Omega$ (in particular, $x(P) \in G$ implies that $x(P) \in \Omega$).

A2. Given any distinguished point $z = x(P), P \in \mathscr{P}$, we can recognize when $z \in G$ and if $z \notin G$, we can construct an affine function $l(x)$ (called a "cut") such that $P' = P \cap \{x|\ l(x) \leq 0\} \in \mathscr{P}$ and $l(x)$ strictly separates z from G, i.e., satisfies

$$l(z) > 0, \qquad l(x) \leq 0 \ \ \forall x \in G.$$

In traditional applications Ω is the set of optimal solutions of an optimization problem, G the feasible set, $\mathscr{P}$ a family of polyhedrons enclosing G, and $x(P)$ an optimal solution of the relaxed problem obtained by replacing the feasible set with P. However, it should be emphasized that in the most general case $\Omega, G, \mathscr{P}, x(P)$ can be defined in quite different ways, provided the above requirements A1, A2 are fulfilled. For instance, in several interesting applications, the set G may be defined only implicitly or may even be unknown, while $x(P)$ need not be a solution of any relaxed problem.

Under assumptions A1, A2 a natural method for finding a point of Ω is the following.

6.3.1 Prototype OA Procedure

Initialization. Start with a set $P_1 \in \mathscr{P}$. Set $k = 1$.

Step 1. Find the distinguished point $x^k = x(P_k)$ (by (A1)). If $x(P_k)$ does not exist then terminate: $\Omega = \emptyset$. If $x(P_k) \in G$ then terminate: $x(P_k) \in \Omega$. Otherwise, continue.

Step 2. Using (A2) construct an affine function $l_k(x)$ such that $P_{k+1} = P_k \cap \{x|\ l_k(x) \leq 0\} \in \mathscr{P}$ and $l_k(x)$ strictly separates x^k from G , i.e., satisfies

$$l_k(x^k) > 0, \quad l_k(x) \leq 0\ \forall x \in G. \tag{6.10}$$

Set $k \leftarrow k + 1$ and return to Step 1.

The above scheme generates a nested sequence of sets $P_k \in \mathscr{P}$:

$$P_1 \supset P_2 \supset \ldots \supset P_k \supset \ldots \supset G$$

approximating G more and more closely from outside, hence the name given to the method (Fig. 6.4). The approximation is refined at each iteration but only around the current distinguished point (considered the most promising at this stage).

An OA procedure as described above may be infinite (for instance, if $x(P)$ exists and $x(P) \notin G$ for every $P \in \mathscr{P}$). An infinite OA procedure is said to be *convergent* if it generates a sequence of distinguished points x^k, every accumulation point $\overline{x}$ of which belongs to G [hence solves the problem by (A1)]. In practice, the procedure is stopped when x^k has become sufficiently close to G, which necessarily occurs for sufficiently large k if the procedure is convergent.

The study of the convergence of an OA procedure relies on properties of the cuts $l_k(x) \leq 0$ satisfying condition (6.10). Let

$$l_k(x) = \langle p^k, x\rangle + \alpha_k, \quad \bar{l}_k(x) = \frac{l_k(x)}{\|p^k\|} = \langle \overline{p}^k, x\rangle + \overline{\alpha}_k.$$

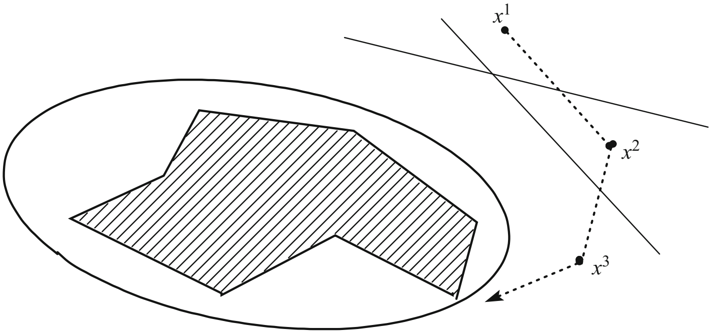


Fig. 6.4 Outer approximation

Lemma 6.4 (Outer Approximation Lemma) *If the sequence* $\{x^k\}$ *is bounded then*

$$\bar{l}_k(x^k) \to 0 \quad (k \to +\infty).$$

Proof Suppose the contrary, that $\bar{l}_{k_q}(x^{k_q}) \geq \eta > 0$ for some infinite subsequence $\{k_q\}$. Since $\{x^k\}$ is bounded we can assume $x^{k_q} \to \bar{x}$ $(q \to +\infty)$. Clearly, for every k,

$$\bar{l}_k(x^k) = \bar{l}_k(\bar{x}) + \langle \bar{p}^k, x^k - \bar{x}\rangle.$$

But for every fixed k, $\bar{l}_k(x^{k_q}) \leq 0$ $\forall k_q > k$ and letting $q \to +\infty$ yields $\bar{l}_k(\bar{x}) \leq 0$. Therefore, $0 < \eta \leq \bar{l}_{k_q}(x^{k_q}) \leq \langle \bar{p}^{k_q}, x^{k_q} - \bar{x}\rangle \to 0$, a contradiction. □

Theorem 6.5 (Basic Outer Approximation Theorem) *Let* G *be an arbitrary closed subset of* $\mathbb{R}^n$ *with nonempty interior, let* $\{x^k\}$ *be an infinite sequence in* $\mathbb{R}^n$*, and for each* k *let* $l_k(x) = \langle p^k, x\rangle + \alpha_k$ *be an affine function satisfying (6.10). If the sequence* $\{x^k\}$ *is bounded and for every* k *there exist* $w^k \in G$*, and* $y^k \in [w^k, x^k] \setminus \text{int}G$*, such that* $l_k(y^k) \geq 0$*,* $\{w^k\}$ *is bounded and, furthermore, every accumulation point of* $\{w^k\}$ *belongs to the interior of* G*, then*

$$x^k - y^k \to 0 \quad (k \to +\infty). \tag{6.11}$$

In particular, if a sequence $\{y^k\}$ *as above can be identified whose every accumulation point belongs to* G *then the OA procedure converges.*

Proof By (6.10) $-\langle \bar{p}^k, x^k\rangle < \bar{\alpha}_k \leq -\langle \bar{p}^k, w^k\rangle$, so the sequence $\bar{\alpha}_k$ is bounded. We can then select an infinite subsequence $\{k_q\}$ such that $\bar{\alpha}_{k_q} \to \alpha$, $x^{k_q} \to \bar{x}$, $y^{k_q} \to y$, $\bar{p}^{k_q} \to p$, $w^{k_q} \to w$ ($\|p\| = 1$ and $w \in \text{int}G$). So $\bar{l}_{k_q}(x) \to l(x) := px + \alpha$ $\forall x$. Since $l(\bar{x}) = \lim \bar{l}_{k_q}(\bar{x}) = \lim[\bar{l}_{k_q}(x^{k_q}) + \langle \bar{p}^{k_q}, \bar{x} - x^{k_q}\rangle] = \lim \bar{l}_{k_q}(x^{k_q})$, by Lemma 6.4 we have $l(\bar{x}) = 0$. On the other hand, from (6.10), $l(x) \leq 0$ $\forall x \in G$, and since

$w \in \text{int}G$ and $l(x) \not\equiv 0$, we must have $l(w) < 0$. From the hypothesis $l_k(y^k) \geq 0$ it follows that $\langle \overline{p}^k, y^k \rangle + \overline{\alpha}_k \geq 0$, hence $l(y) \geq 0$. But $y = \theta w + (1-\theta)\overline{x}$ for some $\theta \in [0,1]$, hence $l(y) = \theta l(w) + (1-\theta) l(\overline{x}) = \theta l(w)$, and since $l(y) \geq 0,\ l(w) < 0$, this implies that $\theta = 0$. Therefore, $y = \overline{x}$. □

An important special case is the following:

Corollary 6.3 *Let* $G = \{x |\ g(x) \leq 0\}$, *where* $g : \mathbb{R}^n \to \mathbb{R}$ *is a convex function and* $\{x^k\} \subset \mathbb{R}^n \setminus G$ *a bounded sequence. If the affine functions* $l_k(x)$ *satisfying (6.10) are such that* $l_k(x) = \langle p^k, x - y^k \rangle + \alpha_k$,

$$p^k \in \partial g(y^k),\ y^k \in [w^k, x^k] \setminus \text{int}G,\ 0 \leq \alpha_k \leq g(y^k), \tag{6.12}$$

where $\{w^k\}$ *is a bounded sequence every accumulation point of which belongs to the interior of G, and* $g(y^k) - \alpha_k \to 0$ $(k \to +\infty)$, *then every accumulation point* $\overline{x}$ *of* $\{x^k\}$ *belongs to G.*

Proof Since $\{x^k\}$ and $\{w^k\}$ are bounded, $\{y^k\}$ is bounded, too. Therefore, $\{p^k\}$ is bounded (Theorem 2.6), and by Lemma 6.4, $l_k(x^k) \to 0$. Furthermore, by Theorem 6.5, $x^k - y^k \to 0$, hence $\alpha_k = l_k(x^k) - \langle p^k, x^k - y^k \rangle \to 0$. But by hypothesis, $g(y^k) - \alpha_k \to 0$, consequently, $g(y^k) \to 0$. If $x^{k_q} \to \overline{x}$ then by Theorem 6.5 $y^{k_q} \to y = \overline{x}$, so $g(\overline{x}) = \lim g(y^k) = 0$, i.e., $\overline{x} \in G$. □

In a standard outer approximation procedure w^k is fixed and $\alpha_k = g(y^k)\ \forall k$. But as will be illustrated in subsequent sections, by allowing more flexibility in the choice of w^k and α_k the above theorems can be used to establish convergence under much weaker assumptions than usual.

6.4 Exercises

1 Construct a sequence of nested simplices M_k such that: for every k of an infinite sequence $\Delta \subset \{1, 2, \ldots\}$, M_{k+1} is a child of M_k by an exact bisection and nevertheless the intersection of all $M_k, k = 1, 2, \ldots$, is not a singleton. Thus, in Lemma 6.1 (which is fundamental to The Basic Subdivision Theorem) the condition (i) alone is not sufficient for the filter M_k to be exhaustive.

2 Prove directly (i.e., without using Lemma 6.1) that if a sequence of nested simplices $M_k, k = 1, 2, \ldots$, is constructed such that M_{k+1} is a child of M_k in a bisection of ration $\alpha > 0$ then the intersection of all M_k is a singleton. (Hint: observe that in a triangle $[a, b, c]$ where $[a, b]$ is the longest side, with length of $\delta > 0$, and v is a point of this side such that $\min\{\|v - a\|, \|v - b\|\} \geq \alpha\delta$ with $0 < \alpha \leq 1/2$ then there exists a constant $\theta \in (0, 1)$ such that $\|v - c\| \leq \theta\delta$; in fact, $\theta = \frac{\|w-s\|}{\delta}$ where w is a point of $[a, b]$ such that $\|w - a\| = \alpha\delta$ and s is the third vertex of an equilateral triangle whose two other vertices are a, b.)

3 Show that Corollary 6.2 can be strengthened as follows: Let $\{M_k := [r^k, s^k]\} \subset \mathbb{R}$ be any filter of rectangles in $\mathbb{R}^n$ such that for infinitely many k, M_{k+1} is a child of M_k in a bisection of ratio no less than a positive constant $\alpha > 0$. Then the filter is exhaustive, i.e., $\mathrm{diam} M_k \to 0$ as $k \to +\infty$. (Hint: use Lemma 6.3; This result is in contrast with what happens for filters of simplices, see Exercise 1).

4 Consider a sequence of nonempty polyhedrons $D_1 \supset \ldots \supset D_k \supset \ldots$ such that $D_{k+1} = D_k \cap \{x|\ l_k(x) \geq 0\}$, where $l_k(x)$ is an affine function. Show that if there exist a number $\varepsilon > 0$ along with for each k a point $x^k \in D_k \setminus D_{k+1}$ such that the distance from x^k to the hyperplane $H_k = \{x|\ l_k(x) = 0\}$ is at least ε then the sequence $D_1, \ldots, D_k, \ldots$ must be finite.

5 Suppose in Step 1 of the Prototype BB Algorithm, we take

$$\beta(M) \leq \inf f(M \cap P),$$

where P is a pregiven set containing D. Does the algorithm work with this modification?

Chapter 7
DC Optimization Problems

7.1 Concave Minimization Under Linear Constraints

One of the most frequently encountered problems of global optimization is *Concave Minimization* (or *Concave Programming*) which consists in minimizing a concave function $f : \mathbb{R}^n \to \mathbb{R}$ over a nonempty closed convex set $D \subset \mathbb{R}^n$. In this and the next chapters we shall focus on the *Basic Concave Programming (BCP) Problem* which is a particular variant of the concave programming problem when all the constraints are linear, i.e., when D is a polyhedron. Assuming $D = \{x| \ Ax \leq b, x \geq 0\}$, where A is an $m \times n$ matrix, and b an m-vector, the problem can be stated as

$$(BCP) \qquad \text{minimize } f(x) \tag{7.1}$$

$$\text{subject to } Ax \leq b \tag{7.2}$$

$$x \geq 0. \tag{7.3}$$

This is the earliest problem of nonconvex global optimization studied by deterministic methods (Tuy 1964). Numerous decision models in operations research, mathematical economics, engineering design, etc., lead to formulations such as (BCP). Important special (BCP) models are plant location problems and also concave min-cost network flow problems which have many applications. Other models which originally are not concave can be transformed into equivalent (BCP)s. On the other hand, techniques for solving (BCP) play a central role in global optimization, as will be demonstrated in subsequent chapters.

Recall that the concavity of $f(x)$ means that for any x', x'' in $\mathbb{R}^n$ we have

$$f((1-\lambda)x' + \lambda x'') \geq (1-\lambda)f(x') + \lambda f(x'') \quad \forall \lambda \in [0, 1].$$

H. Tuy, *Convex Analysis and Global Optimization*, Springer Optimization and Its Applications 110, DOI 10.1007/978-3-319-31484-6_7

Since $f(x)$ is finite throughout $\mathbb{R}^n$, it is continuous (Proposition 2.3) . Later we shall discuss the case when $f(x)$ is defined only on D and may even be discontinuous at certain boundary points of D.

Of course, all the difficulty of (BCP) is caused by the nonconvexity of $f(x)$. However, for the purpose of minimization, the concavity of $f(x)$ implies several useful properties listed below:

(i) For any real γ the level set $\{x| f(x) \geq \gamma\}$ is convex (Proposition 2.11).
(ii) The minimum of $f(x)$ over any line segment is attained at one of the endpoints of the segment (Proposition 2.12).
(iii) If $f(x)$ is bounded below on a halfline, then its minimum over the halfline is attained at the origin of the halfline (Proposition 2.12).
(iv) A concave function is constant on any affine set M where it is bounded below (Proposition 2.12).
(v) If $f(x)$ is unbounded below on some halfline then it is unbounded below on any parallel halfline (Corollary 2.4).
(vi) If the level set $\{x| f(x) \geq \gamma\}$ is nonempty and bounded for one γ, it will be bounded for all γ (Corollary 2.3).

The first two properties, which are actually equivalent, characterize quasiconcavity. When D is bounded, these turn out to be the most fundamental properties for the construction of solution methods for (BCP), so many results in this chapter still remain valid for quasiconcave minimization over polytopes.

Let γ be a real number such that the set $C(\gamma) := \{x| f(x) \geq \gamma\}$ is nonempty. By quasiconcavity and continuity of $f(x)$ this set is closed and convex. To find a feasible point $x \in D$ such that $f(x) < \gamma$ if there is one (in other words, to *transcend* γ) we must solve the dc inclusion

$$x \in D \setminus C(\gamma). \tag{7.4}$$

This inclusion depends upon the parameter γ. Since $C(\gamma) \subset C(\gamma')$ whenever $\gamma \geq \gamma'$, the optimal value in (BCP) is the value

$$\inf\{\gamma| \quad D \setminus C(\gamma) \neq \emptyset\} \tag{7.5}$$

A first property to be noted of this *parametric dc inclusion* is that the range of γ can be restricted to a finite set of real numbers. This follows from Property (iii) listed above, or rather, from the following extension of it:

Proposition 7.1 *Either the global minimum of $f(x)$ over D is attained at one vertex of D, or $f(x)$ is unbounded below over some extreme direction of D.*

Proof Let z be a point of D, chosen so that $z \in \operatorname{argmin}_{x \in D} f(x)$ if $f(x)$ is bounded on D and $f(z) < f(v)$ for any vertex v of D otherwise. From convex analysis (Theorem 2.6) we know that the convex function $-f(x)$ has a subgradient at z, i.e., a vector $-p$ such that $\langle p, x - z\rangle \geq f(x) - f(z)\ \forall x \in \mathbb{R}^n$. Consider the linear program

$$\text{minimize } \langle p, x - z\rangle \quad \text{subject to} \quad x \in D. \tag{7.6}$$

Since the objective function vanishes at $x = z$, by solving this linear program we obtain either an extreme direction u of D such that $\langle p, u\rangle < 0$ or a basic optimal solution v^0 such that $\langle p, v^0 - z\rangle \le 0$. In the first case, for any x we have $f(x+\lambda u) - f(z) \le \langle p, x + \lambda u - z\rangle = \langle p, x - x^*\rangle + \lambda\langle p, u\rangle \to -\infty$ as $\lambda \to +\infty$, i.e., $f(x)$ is unbounded below over any halfline of direction u. In the second case, v^0 is a vertex of D such that $f(v^0) - f(z) \le \langle p, v^0 - z\rangle \le 0$, hence $f(v^0) \le f(z)$. By the choice of z this implies that the second case cannot occur if $f(x)$ is unbounded below over D. Therefore, $f(x)$ has a minimum over D and z is a global minimizer. Since then $f(v^0) = f(z)$, the vertex v^0 is also a global minimizer. □

By this proposition the search for an optimal solution can be restricted to the finite set of vertices and extreme directions of D, thus reducing (BCP) to a combinatorial optimization problem. Furthermore, if a vertex v^0 is such that $f(v^0) \le f(v)$ for all neighboring vertices v then by virtue of the same property v^0 achieves the minimum of $f(x)$ over the polytope spanned by v^0 and all its neighbors, and hence is a local minimizer of $f(x)$ on D. Therefore, given any point $z \in D$ one can compute a vertex $\overline{x}$ which is a local minimizer such that $f(\overline{x}) \le f(z)$ by the following simple procedure:

Take a subgradient $-p$ of $-f(x)$ at z. Solve (7.6) . If an extreme direction u is found with $\langle p, u\rangle < 0$ then (BCP) is unbounded. Otherwise, a vertex v^0 is obtained with $f(v^0) \le f(z)$. Starting from v^0, pivot from a vertex to a better adjacent one, and repeat until a vertex v^k is reached which is no worse than any of its neighbors. Then $\overline{x} = v^k$ is a local minimizer.

Thus, according to the two phase scheme described in Chap. 5 (Sect. 5.7), to solve (BCP) within a tolerance $\varepsilon \ge 0$ we can proceed as follows:

Let $z \in D$ be an initial feasible solution.

Phase 1 (Local phase). Starting from z check whether the optimal value is unbounded ($= -\infty$), and if it is not, search for a vertex $\overline{x}$ of D such that $f(\overline{x}) \le f(z)$. Go to Phase 2.

Phase 2 (Global phase). Solve the inclusion (7.4) for $\gamma = f(\overline{x}) - \varepsilon$. If the inclusion has no solution, i.e., $D \setminus C(\gamma) = \emptyset$, then terminate: $\overline{x}$ is a global ε-optimal solution of (BCP). Otherwise, a feasible point z' is obtained such that $f(z') < f(\overline{x}) - \varepsilon$. Then set $z \leftarrow z'$, and go back to Phase 1.

Since a polyhedron has finitely many vertices, this scheme terminates after finitely many cycles and for sufficiently small $\varepsilon > 0$ a global ε-optimal solution is actually an exact global optimal solution. The key point, of course, is how to solve the dc inclusion (7.4).

7.1.1 Concavity Cuts

Most methods of global optimization combine in different ways some basic techniques: cutting, partitioning, approximating (by relaxation or restriction), and dualizing. In this chapter we will discuss successive partitioning methods which combine cutting and partitioning techniques.

A combinatorial tool which has proven to be useful in the study of many nonconvex optimization problems is the concept of *concavity cut* first introduced in Tuy (1964). This is a device based on the concavity of the objective function and used to exclude certain portions of the feasible set in order to avoid wasteful search on non-promising regions.

Without loss of generality, throughout the sequel we will assume that the feasible polyhedron D in (BCP) is full-dimensional. Consider the inclusion (7.4) for a given value of γ (for instance, $\gamma = f(\overline{x}) - \varepsilon$, where $\overline{x}$ is the incumbent feasible solution at a given stage). Let x^0 be a vertex of D such that $f(x^0) > \gamma$. Suppose that the convex set $C(\gamma)$ is bounded and that a cone M vertexed at x^0 is available that contains D and has exactly n edges. Since $x^0 \in \operatorname{int}C(\gamma)$, these edges meet the boundary of $C(\gamma)$ at n uniquely determined points $z^1, \ldots, z^n$ which are affinely independent (because so are the directions of the n edges of M). Let $\pi(x - x^0) = 1$ be the hyperplane passing through these n points. From

$$\pi(z^i - x^0) = 1 \quad (i = 1, \ldots, n), \tag{7.7}$$

we deduce

$$\pi = eQ^{-1}, \tag{7.8}$$

where $Q = (z^1 - x^0, \ldots, z^n - x^0)$ is the nonsingular matrix with columns $z^i - x^0$, $i = 1, \ldots, n$, and $e = (1, \ldots, 1)$ is a row vector of n ones.

Proposition 7.2 *Any point $x \in D$ such that $f(x) < \gamma$ must lie in the halfspace*

$$eQ^{-1}(x - x^0) > 1. \tag{7.9}$$

Proof Let $S = [x^0, z^1, \ldots, z^n]$ denote the simplex spanned by $x^0, z^1, \ldots, z^n$. Clearly, $S = M \cap \{x |\ eQ^{-1}(x - x^0) \le 1\}$. Since $x^0, z^1, \ldots, z^n \in C(\gamma)$, it follows that $S \subset C(\gamma)$, in other words, $f(x) \ge \gamma$ for all $x \in S$. Therefore, if $x \in D$ satisfies $f(x) < \gamma$ then $x \in M$ but $x \notin S$, and hence, x must satisfy (7.9). □

Corollary 7.1 *The linear inequality*

$$eQ^{-1}(x - x^0) \ge 1 \tag{7.10}$$

excludes x^0 without excluding any point $x \in D$ such that $f(x) < \gamma$.

We call the inequality (7.10) (or the hyperplane H of equation $eQ^{-1}(x - x^0) = 1$) a *γ-valid cut for* (f, D), constructed at x^0 (Fig. 7.1).

Corollary 7.2 *If $\gamma = f(\overline{x}) - \varepsilon$ for some $\overline{x} \in D$ and it so happens that*

$$\max\{eQ^{-1}(x - x^0) |\ x \in D\} \le 1 \tag{7.11}$$

then $f(x) \ge f(\overline{x}) - \varepsilon$ for all $x \in D$, i.e., $\overline{x}$ is a global ε-optimal solution of (BCP).

Thus, (7.11) provides a sufficient condition for global ε-optimality which can easily be checked by solving the linear program $\max\{eQ^{-1}(x - x^0) |\ x \in D\}$.

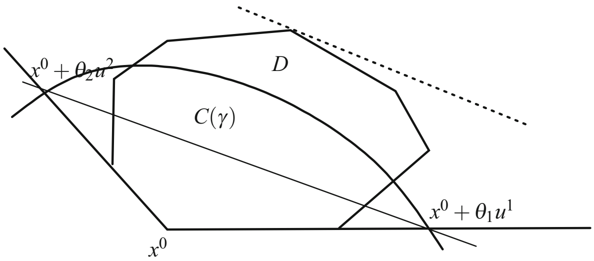


Fig. 7.1 γ-valid concavity cut

Remark 7.1 Denote by u^i, $i = 1, \ldots, n$, the directions of the edges of M and by $U = (u^1, \ldots, u^n)$ the nonsingular matrix with columns $u^1, \ldots, u^n$, so that

$$M = \{x |\, x = x^0 + Ut,\ t = (t_1 \ldots, t_n)^T \geq 0\};$$

i.e., $x \in M$ if and only if $t = U^{-1}(x - x^0) \geq 0$. Let $z^i = x^0 + \theta_i u^i$, where $\theta_i > 0$ is defined from the condition $f(x^0 + \theta_i u^i) = \gamma$. Then $Q = U\mathrm{diag}(\theta_1, \ldots, \theta_n)$, hence $Q^{-1} = \mathrm{diag}(\frac{1}{\theta_1}, \cdots, \frac{1}{\theta_n})U^{-1}$ and

$$eQ^{-1}(x - x^0) = (\frac{1}{\theta_1}, \cdots, \frac{1}{\theta_n})(t_1, \ldots, t_n)^T = \sum_{i=1}^{n} \frac{t_i}{\theta_i}.$$

Thus, if $z^i = x^0 + \theta_i u^i$, $i = 1, \ldots, n$ then the cut (7.10) can also be written in the form

$$\sum_{i=1}^{n} \frac{t_i}{\theta_i} \geq 1, \quad t = U^{-1}(x - x^0), \tag{7.12}$$

which is often more convenient than (7.10) .

Remark 7.2 Given a real number $\gamma < f(x^0)$ and any point $x \neq x^0$, if

$$\theta = \sup\{\lambda |\, f(x^0 + \lambda(x - x^0)) \geq \gamma\}$$

then the point $x^0 + \theta(x - x^0)$ is called the γ-*extension* of x with respect to x^0 (or simply, the γ-*extension* of x if x^0 is clear from the context). The boundedness of $C(\gamma)$ ensures that $\theta < +\infty$ (thus, each z^i defined above is the γ-extension of $v^i = x^0 + u^i$ with respect to x^0). Without assuming the boundedness of $C(\gamma)$, there may exist an index set $I \subset \{1, \ldots, n\}$ such that each ith edge, $i \in I$, of the cone M is entirely contained in $C(\gamma)$. In this case, for each $i \in I$ the γ-extension of $v^i = x^0 + u^i$ is infinite, but $x^0 + \theta_i u^i \in C(\gamma)$ for all $\theta_i > 0$, so the inequality (7.12) still defines a γ-valid cut for whatever large θ_i $(i \in I)$. Making $\theta_i = +\infty$ $(i \in I)$ in (7.12), we then obtain the cut

$$\sum_{i \notin I} \frac{t_i}{\theta_i} \geq 1, \quad t = U^{-1}(x - x^0). \tag{7.13}$$

We will refer to this cut (which exists, regardless of whether $C(\gamma)$ is bounded or not) as the γ*-valid concavity cut* for (f, D) at x^0 (or simply concavity cut, when f, D, γ, x^0 are clear from the context).

7.1.2 Canonical Cone

The construction of the cut (7.13) supposes the availability of a cone $M \supset D$ vertexed at x^0 and having exactly n edges. When $x^0 = 0$ (which implies $b \geq 0$), M can be taken to be the nonnegative orthant $x \geq 0$. In the general case, to construct M we transform the problem to the space of nonbasic variables relative to x^0, in the following way. Recall that $D = \{x \in \mathbb{R}^n |\ Ax \leq b, x \geq 0\}$ where A is an $m \times n$ matrix and b an m-vector. Introducing the slack variables $s = b - Ax$, denote $y = (x, s) \in \mathbb{R}^{n+m}$. Let $y_B = (y_i, i \in B)$ be the vector of basic variables and $y_N = (y_i, i \in N)$ the vector of nonbasic variables relative to the basic feasible solution $y^0 = (x^0, s^0), s^0 = b - Ax^0(|B| = m, |N| = n$ because $\dim D = n$ as assumed). The associated simplex tableau then yields

$$y_B = y_B^0 - W y_N, \quad y_B \geq 0, \quad y_N \geq 0 \tag{7.14}$$

where $W = [w_{ij}]$ is an $m \times n$ matrix and the basic solution y^0 corresponds to $y_N = 0$. By setting $y_N = t$ the constraints (7.14) become $Wt \leq y_B^0, t \geq 0$, so the polyhedron D is contained in the orthant $t \geq 0$. Note that, since $x_i = y_i$ for $i = 1, \ldots, n$, from (7.14) we derive

$$x = x^0 + Ut,$$

where U is an $n \times n$ matrix of elements $u_{ij}, i = 1, \ldots, n, j \in N$, such that

$$u_{ij} = \begin{cases} -w_{ij} & i \in B \\ 1 & i = j \\ 0 & \text{otherwise.} \end{cases}$$

Thus, the orthant $t \geq 0$ corresponds, in the original space, to the cone

$$M = \{x |\ x = x^0 + Ut, t \geq 0\} \tag{7.15}$$

which contains D, is vertexed at x^0 and has exactly n edges of directions represented by the n columns $u^j, j = 1, \dots, n$, of U. For convenience, in the sequel we will refer to the cone (7.15) as the *canonical cone* associated with x^0.

Concavity cuts of the above kind have been developed for general concave minimization problems. When the objective function $f(x)$ has some additional structure, as in concave quadratic minimization and bilinear programming problems, it is possible to significantly strengthen the cuts (Konno 1976a,b; see Chap. 10, Sect. 10.4). Also note that concavity cuts, originally designed for concave programming, have been used in other contexts (Balas 1971, 1972; Glover 1972, 1973a,b). Given the inclusion (7.4), where $\gamma = f(\overline{x}) - \varepsilon$, one can check whether the incumbent $\overline{x}$ is global ε-optimal by using the sufficient optimality condition in Corollary 7.2. This suggests a cutting method for solving the inclusion which consists in iteratively reducing the feasible set by concavity cuts, until a global ε-optimal solution is identified (see e.g. Bulatov 1977, 1987). Unfortunately, it has been observed that, when applied repeatedly the concavity cuts often become shallower and shallower, and may even leave untouched a significant portion of the feasible set. To make the cuts deeper, the idea is to combine cutting with splitting the feasible set into smaller and smaller parts. As illustrated in Fig. 7.2, by splitting the initial cone M_0 into sufficiently small subcones and applying concavity cuts to each of the subcones separately, it is possible to exclude a portion of the feasible set as close to $D \cap C(\gamma)$ as desired, and obtain an accurate approximation to the set $D \setminus C(\gamma)$.

7.1.3 Procedure DC

With the preceding background, let us now consider the fundamental *DC feasibility problem* that was formulated in Sect. 7.1 as the key to the solution of (*BCP*):

($*$) *Given two closed convex sets C, D in $\mathbb{R}^n$, either find a point*

$$x \in D \setminus C, \tag{7.16}$$

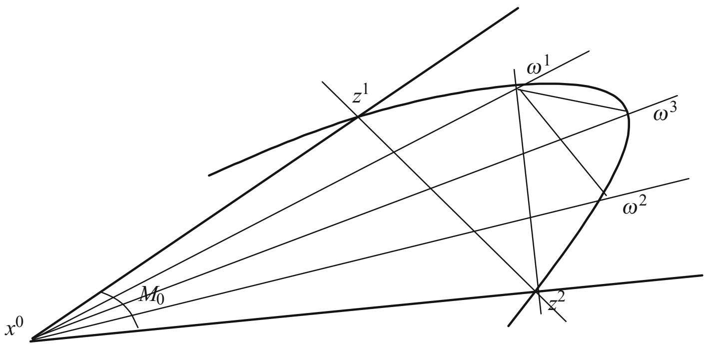


Fig. 7.2 Cut and split process

or else prove that

$$D \subset C. \tag{7.17}$$

As shown in Sect. 7.1, if $\overline{x}$ is the best feasible solution of (BCP) known thus far (the incumbent), then solving $(*)$ for $D = \{x|\ Ax \leq b,\ x \geq 0\}$ and $C = \{x|\ f(x) \geq f(\overline{x})\}$ amounts to transcending $\overline{x}$. According to the general two phase scheme for solving (BCP) within a given tolerance $\varepsilon > 0$, it suffices, however, to find a point $x \in D \setminus C$ (i.e., a better feasible solution than $\overline{x}$) or else to establish that $D \subset C_\varepsilon$ for $C_\varepsilon = \{x|\ f(x) \geq f(\overline{x}) - \varepsilon\}$ (i.e., that $\overline{x}$ is a global ε-optimal solution). Therefore, we will discuss a slight modification of Problem $(*)$, by relaxing the condition (7.17) to

$$D \subset C_\varepsilon, \tag{7.18}$$

where C_ε is a closed convex set such that

$$C_0 = C, \quad C \subset \mathrm{int}C_\varepsilon \text{ if } \varepsilon > 0. \tag{7.19}$$

In other words, we will consider the problem

(#) *For a given closed convex set C_ε satisfying (7.19) find a point $x \in D \setminus C$ or else prove that $D \subset C_\varepsilon$.*

We will assume that C_ε is bounded, since by restricting ourselves to the case when (BCP) has a finite optimal solution it is always possible to take $C_\varepsilon = \{x|\ f(x) \geq f(\overline{x}) - \varepsilon,\ \|x\| \leq \rho\}$ with sufficiently large $\rho > 0$.

Let x^0 be a vertex of D such that $f(x^0) \leq f(z)$ for all neighboring vertices z and $f(x^0) > f(\overline{x}) - \varepsilon$ (i.e., $x^0 \in \mathrm{int}C_\varepsilon$) and let $M_0 = \{x|\ x = x^0 + U_0t, t \geq 0\}$ be the canonical cone associated with x^0 as defined above (so the columns $u^{01}, \ldots, u^{0n}$ of U_0 are the directions of the n edges of M_0). By $S_0 = [v^{01}, \ldots, v^{0n}]$ denote the simplex spanned by $v^{0i} = x^0 + u^{0i},\ i = 1, \ldots, n$.

For a given subcone M of M_0 with base $S = [v^1, \ldots, v^n] \subset Z_0$, i.e., a subcone generated by n halflines from x^0 through $v^1, \ldots, v^n$, let U be the nonsingular matrix of columns $u^i = v^i - x^0,\ i = 1, \ldots, n$. If $\theta_i = \max\{\lambda|\ x^0 + \lambda u^i \in C_\varepsilon\}$ (so that $x^0 + \theta_i u^i$ is the γ-extension of $x^0 + u^i$ for $\gamma = f(\overline{x}) - \varepsilon$), then, as we saw from (7.12), the inequality

$$\sum_{i=1}^{n} \frac{t_i}{\theta_i} \geq 1, \quad t = U^{-1}(x - x^0)$$

defines a γ-valid cut for $(f, D \cap M)$ at x^0. Consider the linear program

$$\max\{\sum_{i=1}^{n} \frac{t_i}{\theta_i}| \quad t = U^{-1}(x - x^0) \geq 0,\ Ax \leq b,\ x \geq 0\}.$$

i.e.,

$$LP(D, M) \qquad \max\{\sum_{i=1}^{n} \frac{t_i}{\theta_i}| \quad \tilde{A}Ut \leq \tilde{b},\ t \geq 0\} \tag{7.20}$$

with

$$\tilde{A} = \begin{bmatrix} A \\ -I \end{bmatrix}, \; \tilde{b} = \begin{bmatrix} b - Ax^0 \\ x^0 \end{bmatrix}. \tag{7.21}$$

Denote the optimal value and a basic optimal solution of this linear program in t by $\mu(M)$ and $\bar{t}(M)$, respectively, and let $\omega(M) = x^0 + U\bar{t}(M)$.

Proposition 7.3 *If $\mu(M) \leq 1$ then $D \cap M \subset C_\varepsilon$, i.e., $f(x) \geq f(\bar{x}) - \varepsilon$ for all points $x \in D \cap M$. If $\mu(M) > 1$ and $f(\omega(M)) \geq f(\bar{x}) - \varepsilon$ then $\omega(M)$ does not lie on any edge of M.*

Proof Clearly $\mu(M)$ and $\omega(M)$ are the optimal value and a basic optimal solution of the linear program

$$\text{maximize } eQ^{-1}(x - x^0) \quad \text{subject to} \quad x \in D \cap M,$$

where $Q = U\text{diag}(\theta^1, \ldots, \theta^n)$ (i.e., Q is the matrix of columns $\theta_1 u^1, \ldots, \theta_n u^n$). The first assertion then follows from Proposition 7.2. To prove the second assertion, suppose that $\mu(M) > 1$ and $f(\omega(M)) \geq f(\bar{x}) - \varepsilon$ while $\omega(M)$ belongs to some jth edge. Since the hyperplane $\sum_{i=1}^{n} \frac{t_i}{\theta_i} = 1$ cuts this edge at the point $x^0 + \theta_j u^j$ we must have $\omega(M) = x^0 + \mu(M)(\theta_j u^j) = x^0 + \theta u^j$ for $\theta = \mu(M)\theta_j > \theta_j$ which contradicts the definition of θ_j because $f(\omega(M)) \geq f(\bar{x}) - \varepsilon$. □

From this proposition it follows that if M_0 can be partitioned into a number of subcones such that $\mu(M) \leq 1$ for every subcone, then $D \subset C_\varepsilon$, i.e., $\bar{x}$ is a global ε-optimal solution. Furthermore, if $\mu(M) > 1$ for some subcone M while $f(\omega(M)) \geq f(\bar{x}) - \varepsilon$, then the subdivision of M via the ray through $\omega(M)$ will be proper. We can now state the following procedure for the master problem (#), assuming $D = \{x |\, Ax \leq b, x \geq 0\}$, $C = \{x |\, f(x) \geq f(\bar{x})\}$, and $\bar{x}$ is a vertex of D.

7.1.3.1 Procedure DC

Let x^0 be a vertex of D which is a local minimizer, such that $x^0 \in \text{int}C_\varepsilon$, let $M_0 = \{x |\, x = x^0 + U_0 t, t \geq 0\}$ be a canonical cone associated with x^0, $S_0 = [x^0 + u^{01}, \ldots, x^0 + u^{0n}]$, where $u^{01}, \ldots, u^{0n}$ are the columns of U_0. Set $\mathscr{P} = \{M_0\}$, $\mathscr{N} = \{M_0\}$.

Step 1. For each cone $M \in \mathscr{P}$, of base $S = [x^0 + u^1, \ldots, x^0 + u^n] \subset S_0$ let U be the matrix of columns $u^1, \ldots, u^n$. Compute θ_i such that $x^0 + \theta_i u^i \in \partial C_\varepsilon (i = 1, \ldots, n)$ and solve the linear program [see (7.20)]:

$$LP(D, M) \qquad \max\{\sum_{i=1}^{n} \frac{t_i}{\theta_i} |\quad \tilde{A}Ut \leq \tilde{b}, t \geq 0\}$$

to obtain the optimal value $\mu(M)$ and a basic optimal solution $\bar{t}(M)$. Let $\omega(M) = x^0 + U\bar{t}(M)$.

Step 2. If for some $M \in \mathscr{P}, \omega(M) \notin C$, i.e., $f(\omega(M)) < f(\bar{x})$, then terminate: $\omega(M)$ yields a point $z \in D \setminus C$.

Step 3. Delete every cone $M \in \mathscr{N}$ for which $\mu(M) \leq 1$ and let $\mathscr{R}$ be the collection of all remaining cones.

Step 4. If $\mathscr{R} = \emptyset$ then terminate: $D \subset C_\varepsilon(\bar{x}$ is a global ε-optimal solution of (BCP)).

Step 5. Let $M_* \in \text{argmax}\{\mu(M)|\ M \in \mathscr{R}\}$, with base S_*. Select a point $v^* \in S_*$ which does not coincide with any vertex of S_*. Split S_* via v^*, obtaining a partition $\mathscr{P}_*$ of M_*.

Step 6. Set $\mathscr{N} \leftarrow (\mathscr{R} \setminus \{M_*\}) \cup \mathscr{P}_*$, $\mathscr{P} \leftarrow \mathscr{P}_*$ and return to Step 1.

Incorporating the above procedure into the two phase scheme for solving (BCP) (see Sect. 5.7), we obtain the following:

CS (Cone Splitting) Algorithm for (BCP)

Let z^0 be an initial feasible solution.

Phase 1. Compute a vertex $\bar{x}$ which is a local minimizer such that $f(\bar{x}) \leq f(z^0)$.

Phase 2. Call Procedure DC with $C{=}\{x|\, f(x) \geq f(\bar{x})\}$, $C_\varepsilon{=}\{x|\, f(x) \geq f(\bar{x}){-}\varepsilon\}$.

(a) If the procedure returns a point $z \in D$ such that $f(z) < f(\bar{x})$ then set $z^0 \leftarrow z$ and go back to Phase 1. (Optionally: with $\gamma = f(z)$ and x^0 being the vertex of D where Procedure DC has been started, construct a γ-valid cut $\pi(x-x^0) \geq 1$ for (f, D) at x^0 and reset $D \leftarrow D \cap \{x|\ \pi(x-x^0) \geq 1\}$).

(b) If the procedure establishes that $D \subset C_\varepsilon$, then $\bar{x}$ is a global ε-optimal solution.

With regard to convergence and efficiency, a key point in Procedure DC is the subdivision strategy, i.e., the rule of how the cone M_* in Step 5 must be subdivided.

Let M_k denote the cone M_* at iteration k. Noting that $\omega(M_k)$ does not lie on any edge of M_k, a natural strategy that comes to mind is to split the cone M_k upon the halfline from x^0 through $\omega(M_k)$, i.e., to subdivide its base S_k via the intersection point v^k of S_k with this halfline. A subdivision of this kind is referred to as an ω-*subdivision* and the strategy which prescribes ω-subdivision in every iteration is called the ω-strategy. This natural strategy was used in earliest works in concave minimization (Tuy 1964; Zwart 1974) and although it seemed to work quite well in practice (Hamami and Jacobsen 1988), its convergence was never rigorously established, until the works of Jaumard and Meyer (2001), Locatelli (1999), and most recently Kuno and Ishihama (2015). Now we can sate the following:

Theorem 7.1 *The CS Algorithm with ω-subdivisions terminates after finitely many steps, yielding a global ε-optimal solution of (BCP).*

Proof Thanks to Theorem 6.2, a new proof for this result can be given which is much shorter than all previously published ones in the above cited works. In fact, consider iteration k of Procedure DC with ω-subdivisions. The cone M_k (i.e., M_* in step 5) is subdivided via the halfline from x^0 through $\omega^k := \omega(M_k)$. Let $z^{k1}, \ldots, z^{kn}$

be the points where the edges of M_k meet the boundary ∂C_ε of C_ε and let q^k be the point where the halfline from x^0 through ω^k meets the simplex $[z^{k1}, \ldots, z^{kn}]$. By Theorem 6.2 if the procedure is infinite there is an infinite subsequence $\{k_s\}$ such that $q^{k_s} \to q \in \partial C_\varepsilon$ as $s \to \infty$. By continuity of $f(x)$ it follows that $f(q^{k_s}) \to f(q) = f(\overline{x}) - \varepsilon$, so $f(\omega^{k_s}) < f(\overline{x})$ for sufficiently large k_s. Since this is the stopping criterion in Step 2 of the procedure, the algorithm terminates after finitely many iterations by the stopping criterion in Step 4. □

Remark 7.3 It is not hard to show that $\omega^k = \mu(M_k)q^k$. From the above proof there is an infinite subsequence $\{k_s\}$ such that $\|\omega^{k_s} - q^{k_s}\| \to 0$, i.e., $(\mu(M_{k_s}) - 1)\|q^{k_s}\| \to 0$, hence

$$\mu(M_{k_s}) \to 1 \ (s \to \infty). \tag{7.22}$$

Later we will see that this condition is sufficient to ensure convergence of the CS Algorithm viewed as a Branch and Bound Algorithm.

Remark 7.4 The CS Algorithm for solving (BCP) consists of a finite number of cycles each of which involves a Procedure DC for transcending the current incumbent. Any new cycle is actually a **restart** of the algorithm, whence the name CS restart given sometimes to the algorithm. Due to restart, the set $\mathcal{N}_k$ (the collection of cones to be stored) can be kept within reasonable size to avoid storage problems. Furthermore, the vertex x^0 used to initialize Procedure DC in each cycle needs not be fixed throughout the whole algorithm, but may vary from cycle to cycle. When started from x^0 Procedure DC concentrates the search over the part of the boundary of D contained in the canonical cone associated with x^0, so a change of x^0 means a change of search directions. This **multistart** effect may sometimes drastically improve the convergence speed. On the other hand, further flexibility can be given to the algorithm by modifying Step 2 of Procedure DC as follows:

Step 2. (Modified) If for some $M \in \mathcal{N}, \omega(M) \notin C$, i.e., $f(\omega(M)) < f(\overline{x})$ then, optionally: (a) either set $z^0 \leftarrow \omega(M)$ and return to Phase 1, or (b) compute a vertex $\hat{x}$ of D such that $f(\hat{x}) \leq f(\omega(M))$, reset $\overline{x} \leftarrow \hat{x}$. and go to Step 3.

With this modification, as soon as in Step 2 of Procedure DC a better feasible solution is found which is better than the incumbent, the algorithm either returns to Phase 1 (with $z^0 \leftarrow \omega(M)$) or goes to Step 3 (with an updated incumbent) to continue the current Procedure DC. In general, the former option ("return to Phase 1") should be chosen when the set $\mathcal{N}$ has reached a critical size. The efficiency of the multirestart strategy is illustrated in the following example:

Example 7.1 Consider the problem

$$\begin{aligned} \text{minimize } & x_1 - 10x_2 + 10x_3 + x_8 \\ & \qquad -x_1^2 - x_2^2 - x_3^2 - x_4^2 - 7x_5^2 - 4x_6^2 - x_7^2 - 2x_8^2 \\ & \qquad\qquad +2x_1x_2 + 6x_1x_5 + 6x_2x_5 + 2x_3x_4 \end{aligned}$$

$$\text{subject to} \left| \begin{array}{l} x_1 + 2x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 \leq 8 \\ 2x_1 + x_2 + x_3 \leq 9 \\ x_3 + x_4 + x_5 \leq 5 \\ 0.5x_5 + 0.5x_6 + x_7 + 2x_8 \leq 3 \\ 2x_2 - x_3 - 0.5x_4 \leq 5 \\ x_1 \leq 6,\ x_i \geq 0 \quad i = 1, 2, \ldots, 8 \end{array} \right.$$

A local minimizer of the problem is the vertex $x^0 = (0, 0, 0, 0, 5, 1, 0, 0)$ of the feasible polyhedron, with objective function value -179. Suppose that Procedure DC is applied to transcend this local minimizer. Started from x^0 the procedure runs through 37 iterations, generating 100 cones, without detecting any better feasible solution than x^0 nor producing evidence that x^0 is already a global optimal solution. By restarting the procedure from the new vertex $x^0 \leftarrow (0, 2.5, 0, 0, 0, 0, 0, 0)$ we find after eight iterations that no better feasible solution than x^0 exists. Therefore, x^0 is actually a global minimizer.

7.1.4 ω-Bisection

As said above, for some time the convergence status of conical algorithms using ω-subdivisions was unsettled. It was then natural to look for other subdivisions which could secure convergence, even at some cost for efficiency. The simplest subdivision that came to mind was bisection: at each iteration the cone M_* is split along the halfline from x^0 through the midpoint of a longest edge of the simplex S_*. In fact, a first conical algorithm using bisections was developed (Thoai and Tuy 1980) whose convergence was proved rigorously. Subsequently, however, it was observed that the convergence achieved with bisections is too slow. This motivated the *Normal Subdivision* (Tuy 1991a) which is a kind of hybrid strategy, using ω-subdivisions in most iterations and bisections occasionally, just enough to prevent jamming and eventually ensure convergence.

At the time when the pure ω-subdivision was efficient, yet uncertain to converge, Normal Subdivision was the best hybrid subdivision ensuring both convergence and efficiency (Horst and Tuy 1996). However, once the convergence of pure ω-subdivision has been settled, bisections are no longer needed to prevent jamming in conical algorithms. On the other hand, an advantage of bisections over other kinds of subdivision is that each bisection creates only two new partition sets, so that using bisections may help to keep the growth of the collection $\mathscr{N}$ of accumulated partition sets within more manageable limits. Moreover, recent investigations by Kuno and Ishihama (2015) show that bisections can be built in such a way to take account as well of current problem structure just like ω-subdivisions.

The following concept of *ω-bisection* is a slight modification of that introduced by Kuno–Ishihama. The modification allows us to dispense with assuming strict concavity of $f(x)$ and also to simplify the convergence proof.

Let M_k be the cone to be subdivided at iteration k of Procedure DC, $S_k = [u^{k1}, \ldots, u^{kn}] \subset \mathbb{R}^n$ its base, v^k the point where S_k meets the ray through $\omega^k = \omega(M_k)$. Let $v^k = \sum_{i=1}^n \lambda_i^k u^{ki}$ where $\sum_{i=1}^n \lambda_i^k = 1$, $\lambda_i^k \geq 0$ $(i = 1, \ldots, n)$. Take $\alpha > 0$, and for every k define

$$\theta_i^k = \frac{\lambda_i^k + \alpha}{1 + n\alpha}, i = 1, \ldots, n, \quad \tilde{v}^k = \sum_{i=1}^n \theta_i^k u^{ki}.$$

Clearly $\theta_i^k > 0$ $\forall k, i = 1, \ldots, n$, $\sum_{i=1}^n \theta_i^k = 1$, so $\tilde{v}^k \in S_k$. Let $[u^{kr_k}, u^{ks_k}]$ be a longest edge of the simplex S_k and

$$w^k = \frac{\theta_{r_k}^k u^{kr_k} + \theta_{s_k}^k u^{ks_k}}{\theta_{r_k}^k + \theta_{s_k}^k}. \tag{7.23}$$

The *ω-bisection* of M_k is then defined to be the subdivision of M_k via the halfine from x^0 through w^k (or, in other words, the subdivision of M_k induced by the subdivision of S_k via w^k).

Since $\theta_i^k > 0$ $\forall i$ the set $\{\tilde{v}^k, u^{ki}, i \notin \{r_k, s_k\}\}$ is affinely independent and, as is easily seen, w^k is just the point where the segment $[u^{kr_k}, u^{ks_k}]$ meets the affine manifold through $\tilde{v}^k, u^{ki}, i \notin \{r_k, s_k\}$.

Theorem 7.2 *The CS Algorithm with ω-bisections terminates after finitely many steps, yielding a global ε-optimal solution of (BCP).*

Proof Since $\frac{\theta_{r_k}^k}{\theta_{r_k}^k + \theta_{s_k}^k} \geq \frac{\lambda_{r_k}^k + \alpha}{1 + n\alpha} \geq \frac{\alpha}{1 + n\alpha}$ the ω-bisection of S_k is a bisection of S_k of ratio $\geq \frac{\alpha}{1+n\alpha} > 0$, so by Corollary 6.1 the set $\cap_{k=1}^{\infty} S_k$ is a singleton $\{v\}$. Hence, if $z^{k1}, \ldots, z^{kn}$ are the points where the halflines from x^0 through $u^{k1}, \ldots, u^{kn}$ meet ∂C_ε while $q^k, \hat{q}^k$ are the points where the halfline from x^0 through v^k meets the simplex $[z^{k1}, \ldots, z^{kn}]$ and ∂C_ε, respectively, then $\|q^k - \hat{q}^k\| \to 0$ as $k \to \infty$. Therefore there is $q \in \partial C_\varepsilon$ such that $q^k, \hat{q}^k \to q$ as $k \to \infty$. Since $\omega^k := \omega(M_k) \in [q^k, \hat{q}^k]$ it follows that $\omega^k \to q \in \partial C_\varepsilon$. By reasoning just as in the last part of the proof of Theorem 6.1 we conclude that the CS Algorithm with ω-bisections terminates after finitely many iterations, yielding a global ε-optimal solution. □

From the computational experiments reported in Kuno and Ishihama (2015) the CS Algorithm with ω-bisection seems to compete favorably with the CS Algorithm using ω-subdivisions.

7.1.5 Branch and Bound

Let M be a cone vertexed at a point $x^0 \in D$ and having n edges of directions $u^1, \ldots, u^n$. A lower bound of $f(M \cap D)$ to be used in a conical BB algorithm for solving (BCP) can be computed as follows.

Let γ be the current incumbent value. Assuming that $\gamma - \varepsilon < f(x^0)$ (otherwise x^0 is already a global ε-optimal solution of (BCP) compute

$$\theta_i = \sup\{\lambda | f(x^0 + \lambda u^i) \geq \gamma - \varepsilon\},\ i = 1, \ldots, n, \tag{7.24}$$

and by $\mu = \mu(M)$ denote the optimal value of the linear program $LP(D, M)$ given by (7.20). If $y^i = x^0 + \mu\theta_i u^i, i = 1, \ldots, n$ then clearly $D \cap M \subset [x^0, y^1, \ldots, y^n]$, so by concavity of $f(x)$ a lower bound of $f(D \cap M)$ can be taken to be the value

$$\beta(M) = \begin{cases} \gamma - \varepsilon & \text{if } \mu(M) \leq 1 \\ \min\{f(y^1), \ldots, f(y^n)\} & \text{otherwise.} \end{cases} \tag{7.25}$$

(see Fig. 7.3). Using this lower bounding $M \mapsto \beta(M)$ together with the conical ω-subdivision rule we can then formulate the following BB (branch and bound) version of the CS Algorithm discussed in Sect. 5.4.

Conical BB Algorithm for (BCP)

Initialization Compute a feasible solution z^0.

Phase 1 Compute a vertex $\overline{x}$ of D which is a local minimizer such that $f(\overline{x}) \leq f(z^0)$.

Phase 2 Take a vertex x^0 of D (local minimizer) such that $f(x^0) > f(\overline{x}) - \varepsilon$. Let $M_0 = \{x |\ x = x^0 + U_0 t, t \geq 0\}$ be the canonical cone associated with x^0, $S_0 = [x^0 + u^{01}, \ldots, x^0 + u^{0n}]$, where $u^{0i}, i = 1, \ldots, n$ are the columns of U_0. Set $\mathscr{P} = \{M_0\}, \mathscr{N} = \{M_0\}, \gamma = f(\overline{x})$.

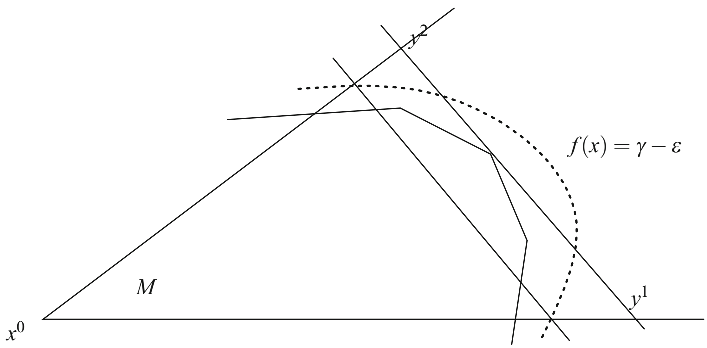


Fig. 7.3 Computation of $\beta(M)$

Step 1. (*Bounding*) For each cone $M \in \mathscr{P}$, of base $S = [z^1, \ldots, z^n] \subset S_0$ let U be the matrix of columns $u^i = z^i - x^0, i = 1, \ldots, n$. Compute $\theta_i, i = 1, \ldots, n$ according to (7.24) and solve the linear program [see (7.20)]:

$$LP(D, M) \qquad \max \left\{ \sum_{i=1}^{n} \frac{t_i}{\theta_i} | \quad \tilde{A}Ut \leq \tilde{b}, t \geq 0 \right\} \tag{7.26}$$

to obtain the optimal value $\mu(M)$ and a basic optimal solution $\bar{t}(M)$. Let $\omega(M) = x^0 + U\bar{t}(M)$. Compute $\beta(M)$ as in (7.25).

Step 2. (*Options*) If for some $M \in \mathscr{P}$, $f(\omega(M)) < \gamma$, then do either of the options:

(a) with $\gamma \leftarrow f(\omega(M))$ construct a $(\gamma - \varepsilon)$-valid cut for (f, D) at x^0 : $\langle \pi, x - x^0 \rangle \geq 1$, reset $D \leftarrow D \cap \{x | \langle \pi, x - x^0 \rangle \geq 1\}$, $z^0 \leftarrow \omega(M)$, and go back to Phase 1;

(b) compute a vertex $\hat{x}$ of D such that $f(\hat{x}) \leq f(\omega(M))$, reset $\bar{x} \leftarrow \hat{x}$, $\gamma \leftarrow f(\hat{x})$ and go to Step 3.

Step 3. (*Pruning*) Delete every cone $M \in \mathscr{N}$ for which $\beta(M) \geq \gamma - \varepsilon$ and let $\mathscr{R}$ be the collection of all remaining cones.

Step 4. (*Termination Criterion*) If $\mathscr{R} = \emptyset$, then terminate: $\bar{x}$ is a global ε-optimal solution of (BCP).

Step 5. (*Branching*) Let $M_* \in \operatorname{argmin}\{\beta(M) | M \in \mathscr{R}\}$, with base S_*. Perform an ω-subdivision of M_* and let $\mathscr{P}_*$ be the partition of M_*.

Step 6. (*New Net*) Set $\mathscr{N} \leftarrow (\mathscr{R} \setminus \{M_*\}) \cup \mathscr{P}_*$, $\mathscr{P} \leftarrow \mathscr{P}_*$ and return to Step 1.

Theorem 7.3 *The above algorithm can be infinite only if $\varepsilon = 0$ and then any accumulation point of the sequence $\{\bar{x}^k\}$ is a global optimal solution of (BCP).*

Proof Since D has finitely many vertices, if the algorithm is infinite, a Phase 2 must be infinite and generate a filter $\{M_k, k \in K\}$. Let γ_k be the current value of γ, and let $\mu^k = \mu(M_k), \omega^k = \omega(M_k), \bar{\gamma} = \lim \gamma_k$. By formula 7.22 (which extends easily to the case when $\gamma_k \searrow \bar{\gamma}$), there exists $K_1 \subset K$ such that $\mu_k \to 1, \omega^k \to \bar{\omega}$, with $\bar{\omega}$ satisfying $f(\bar{\omega}) = \bar{\gamma} - \varepsilon$ as $k \to \infty, k \in K_1$. But in view of Step 2, $f(\omega^k) \geq \gamma_k$, hence $f(\bar{\omega}) \geq \bar{\gamma}$, which is possible only if $\varepsilon = 0$. Moreover, without loss of generality we can assume $\beta(M_k) = f(x^0 + \mu_k \theta_{1k} u^{1k})$, and so, in that case, $\gamma_k - \beta(M_k) \leq f(\theta_{1k} u^{1k}) - f(\mu_k \theta_{1k} u^{1k}) \to 0$ as $k \to \infty, k \in K_1$. Therefore, by Proposition 6.1 the algorithm converges. □

Remark 7.5 Option (a) in Step 2 allows the algorithm to be *restarted* when storage problems arise in connection with the growth of the current set of cones. On the other hand, with option (b) the algorithm becomes a unified procedure. This flexibility sometimes enhances the practicability of the algorithm. Note, however, that the BB algorithm requires more function evaluations than the CS Algorithm (for each cone M the CS algorithm only requires the value of $f(x)$ at $\omega(M)$, while the BB Algorithm requires the values of $f(x)$ at n points $y^1, \ldots, y^n$). This should be taken into account when functions evaluations are costly, as it occurs in problems where the objective function $f(x)$ is implicitly defined via some auxiliary problem (for instance, $f(x) = \min\{\langle Qx, y\rangle |\ y \in G\}$, where $Q \in \mathbb{R}^{p\times n}$ and G is a polytope in $\mathbb{R}^p$).

7.1.6 Rectangular Algorithms

Aside from simplicial and conical subdivisions, rectangular subdivisions are also used in branch and bound algorithms for (BCP). Especially, rectangular subdivisions are often preferred when the objective function is such that a lower bound of its values over a rectangle can be obtained in a straightforward manner. The latter occurs, in particular, when the function is *separable*, i.e., has the form $f(x) = \sum_{i=1}^n f_i(x_i)$.

Consider now the Separable Basic Concave Program

$$(SBCP) \qquad \min\{f(x) := \sum_{j=1}^n f_j(x_j) |\quad x \in D\}, \tag{7.27}$$

where D is a polytope contained in the rectangle $[c, d] = \{x : c \le x \le d\}$ and each function $f_j(t)$ is concave and continuous on the interval $[c_j, d_j]$. Separable programming problems of this form appear quite frequently in production, transportation, and planning models. Also note that any quadratic function can be made separable by an affine transformation (see Chap. 4, Sect. 4.5).

Proposition 7.4 *If* $f(x) = \sum_{j=1}^n f_j(x_j)$ *is a separable concave function and* $M = [r; s]$ *is any rectangle with* $c \le r, s \le d$*, then there is an affine minorant of* $f(x)$ *on* M *which agrees with* $f(x)$ *at the vertices of* M*. This function is given by*

$$\varphi_M(x) = \sum_{j=1}^n \varphi_{M,j}(x_j),$$

where, for every j*,* $\varphi_{M,j}(.)$ *is the affine function of one variable that agrees with* $f_j(.)$ *at the endpoints of the interval* $[r_j, s_j]$*, i.e.,*

$$\varphi_{M,j}(t) = f_j(r_j) + \frac{f_j(s_j) - f_j(r_j)}{s_j - r_j}(t - r_j). \tag{7.28}$$

Proof That $\varphi_M(x)$ is affine and minorizes $f(x)$ over M is obvious. On the other hand, if x is a vertex of M then, for every j, we have either $x_j = r_j$ or $x_j = s_j$, hence $\varphi_{M,j}(x_j) = f_j(x_j),\ \forall j$, and therefore, $f(x) = \varphi_M(x)$. □

On the basis of this proposition a lower bound for $f(x)$ over a rectangle $M = [r, s]$ can be computed by solving the linear program

$$LP(D, M) \qquad \min\{\varphi_M(x) |\quad x \in D \cap M\}. \tag{7.29}$$

If $\omega(M)$ and $\beta(M)$ denote a basic optimal solution and the optimal value of this linear program, then $\beta(M) \le \min f(D \cap M)$, with equality holding when $f(\omega(M)) = \varphi_M(\omega(M))$. It is also easily seen that if a rectangle $M = [r; s]$ is entirely outside D then $D \cap M$ is empty and M can be discarded from consideration ($\beta(M) = +\infty$). On the other hand, if M is entirely contained in D then $D \cap M = M$, hence the minimum of $f(x)$ over $D \cap M$ is achieved at a vertex of M and since $\varphi_M(x)$ agrees with $f(x)$ at the vertices of M, it follows that $\beta(M) = \min f(D \cap M)$. Therefore, in a branch and bound rectangular algorithm for (*SBCP*) only those rectangles intersecting the boundary of D will need to be investigated, so that the search will essentially be concentrated on this boundary.

Proposition 7.5 *Let $\{M_k, k \in K\}$ be a filter of rectangles as in Lemma 6.3. If*

$$j_k \in \operatorname{argmax}\{f_j(\omega_j^k) - \varphi_{M_k,j}(\omega_j^k) | j = 1, \ldots, n\} \tag{7.30}$$

then there exists a subsequence $K_1 \subset K$ such that

$$f(\omega^k) - \varphi_{M_k}(\omega^k) \to 0 \ (k \to \infty, k \in K_1). \tag{7.31}$$

Proof By Lemma 6.3, we may assume, without loss of generality, that $j_k = 1 \ \forall k \in K_1$ and $\omega_1^k - r_1^k \to 0$ as $k \to \infty, k \in K_1$ (the case $\omega_1^k - s_1^k \to 0$ is similar). Since $\varphi_{M_k,1}(r_1^k) = f_1(r_1^k)$, it follows that $f_1(\omega_1^k) - \varphi_{M_k,1}(\omega_1^k) \to 0$. The choice of $j_k = 1$ by (7.30) then implies that $f_j(\omega_j^k) - \varphi_{M_k,j}(\omega_j^k) \to 0$ $(j = 1, \ldots, n)$, hence (7.31). □

The above results provide the basis for the following algorithm, essentially due to Falk and Soland (1969):

BB Algorithm for (SBCP)

Initialization Let $\overline{x}^0$ be the best feasible solution available, $\mathscr{P}_1 = \mathscr{N}_1 = \{M_0 := [c, d]\}$. Set $k = 1$.

Step 1. (*Bounding*) For each rectangle $M \in \mathscr{P}_k$ solve the linear program $LP(D, M)$ given by (7.29) to obtain its optimal value $\beta(M)$ and a basic optimal solution $\omega(M)$.

Step 2. (*Incumbent*) Update the incumbent by setting $\overline{x}^k$ equal to the best among $\overline{x}^{k-1}$ and all $\omega(M), M \in \mathscr{P}_k$.

Step 3. (*Pruning*) Delete every $M \in \mathscr{N}_k$ such that $\beta(M) \ge f(\overline{x}^k) - \varepsilon$. Let $\mathscr{R}_k$ be the collection of remaining members of $\mathscr{N}_k$.

Step 4. (*Termination Criterion*) If $\mathscr{R}_k = \emptyset$, then terminate: $\overline{x}^k$ is a global ε-optimal solution of (SBCP).

Step 5. (*Branching*) Choose $M_k \in \operatorname{argmin}\{\beta(M) | M \in \mathscr{R}_k\}$. Let $\omega^k = \omega(M_k)$. Choose j_k according to (7.30) and subdivide M_k via (ω^k, j_k). Let $\mathscr{P}_{k+1}$ be the partition of M_k.

Step 6. (*New Partitioning*) Set $\mathscr{N}_{k+1} = (\mathscr{R}_k \setminus \{M_k\}) \cup \mathscr{P}_{k+1}$. Set $k \leftarrow k + 1$ and go back to Step 1.

Theorem 7.4 *The above algorithm can be infinite only if $\varepsilon = 0$ and in this case every accumulation point of the sequence $\{\overline{x}^k\}$ is a global optimal solution of (SBCP).*

Proof Proposition 7.5 ensures that every filter $\{M_k, k \in K\}$ contains an infinite subsequence $\{M_k, k \in K_1\}$ such that

$$f(\omega^k) - \beta(M_k) \to 0 \quad (k \to \infty, k \in K_1).$$

The conclusion then follows from Proposition 5.1. □

Remark 7.6 The subdivision via (ω^k, j_k) as used in the above algorithm is also referred to as ω-*subdivision*. Note that if M_k is chosen as in Step 5, and ω^k coincides with a corner of M_k then, as we saw above [see comments following (7.29)], $f(\omega^k) = \beta(M_k)$, hence ω^k solves (*SBCP*). Intuitively this suggests that convergence would be faster if the subdivision tends to bring ω^k more rapidly to a corner of M_k, i.e., if ω-subdivision is used instead of bisection as in certain similar rectangular algorithms (see, e.g., Kalantari and Rosen 1987). In fact, although bisection also guarantees convergence (in view of its exhaustiveness, by Corollary 6.2), for (*SBCP*) it seems to be generally less efficient than ω-subdivision. Also note that just as with conical or simplicial subdivisions, rectangular procedures using ω-subdivision do not need any special device to ensure convergence.

7.2 Concave Minimization Under Convex Constraints

Consider the following *General Concave Minimization Problem*:

$$(GCP) \qquad \min\{f(x)| \quad x \in D\}, \tag{7.32}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a concave function, and $D = \{x| \ g(x) \leq 0\}$ is a compact convex set defined by the convex function $g : \mathbb{R}^n \to \mathbb{R}$. Since $f(x)$ is continuous and D is compact an optimal solution of (*GCP*) exists.

Using the outer approximation (OA) approach described in Chap. 5 we will reduce (*GCP*) to solving a sequence of (*BCP*)s. First it is not hard to check that the conditions (A1) and (A2) required in the Prototype OA Algorithm are fulfilled. Let $G = D$, let Ω be the set of optimal solutions to (*GCP*), $\mathscr{P}$ the collection of polyhedrons that contain D. Since $G = D$ is a compact convex set, it is easily seen that Condition A2 is fulfilled by taking $l(x)$ to be an affine function strictly separating z from G. The latter function $l(x)$ can be chosen in a variety of different ways. For instance, according to Corollary 6.3, if $g(x)$ is differentiable then to separate $x^k = x(P_k)$ from $G = D$ one can take

$$l_k(x) = \langle \nabla g(x^k), x - x^k \rangle + g(x^k) \tag{7.33}$$

[so $y^k = x^k,\ \alpha_k = g(x^k)$ in (6.12)]. If an interior point x^0 of D is readily available, the following more efficient cut (Veinott 1967) should be used instead:

$$l_k(x) = \langle p^k, x - y^k\rangle + g(y^k), \quad y^k \in [x^0, x^k] \setminus \mathrm{int}G, p^k \in \partial g(y^k). \tag{7.34}$$

Thus, to apply the outer approximation method to (GCP) it only remains to define a distinguished point $x^k := x(P_k)$ for every approximating polytope P_k so as to fulfill Condition A1.

7.2.1 Simple Outer Approximation

In the simple outer approximation method for solving (GCP), $x^k = x(P_k)$ is chosen to be a vertex of P_k minimizing $f(x)$ over P_k. In other words, denoting the vertex set of P_k by V_k one defines

$$x^k \in \mathrm{argmin}\{f(x)|\ x \in V_k\}. \tag{7.35}$$

Since $f(x^k) \leq \min f(D)$ it is obvious that $\overline{x} = \lim_{\nu\to\infty} x^{k_\nu} \in D$ implies that $\overline{x}$ solves (GCP), so Condition A1 is immediate.

A practical implementation of the simple OA method thus requires an efficient procedure for computing the vertex set V_k of P_k. At the beginning, P_1 can be taken to be a simple polytope (most often a simplex) with a readily available vertex set V_1. Since P_k is obtained from P_{k-1} by adding a single linear constraint, all we need is a subroutine for solving the following subproblem of *on-line vertex enumeration*:

Given the vertex set V_k of a polytope P_k and a linear inequality $l_k(x) \leq 0$, compute the vertex set V_{k+1} of the polytope

$$P_{k+1} = P_k \cap \{x|\ l_k(x) \leq 0\}.$$

Denote

$$V_k^- = \{v \in V_k|\ l_k(x) < 0\},\ V_k^+ = \{v \in V_k|\ l_k(x) > 0\}.$$

Proposition 7.6 *$V_{k+1} = V_k^- \cup W$, where $w \in W$ if and only if it is the intersection point of the hyperplane $l_k(x) = 0$ with an edge of P_k joining an element of V_k^+ with an element of V_k^-.*

Proof The inclusion $V_k^- \subset V_{k+1}$ is immediate from the fact that a point of a polyhedron of dimension n is a vertex if and only if it satisfies as equalities n linearly independent constraints (Corollary 1.17). To prove that $W \subset V_{k+1}$ let w be a point of an edge $[v, u]$ joining $v \in V_k^+$ with $u \in V_k^-$. Then w must satisfy as equalities $n-1$ linearly independent constraints of P_k (Corollary 1.17). Since all these constraints are binding for v while $l_k(v) > 0$ (because $v \in V_k^+$), the inequality $l_k(x) \leq 0$

cannot be a linear combination of these constraints. Hence w satisfies n linearly independent constraints defining P_{k+1}, i.e., $w \in V_{k+1}$. Conversely if $w \in V_{k+1}$ then either $l_k(w) < 0$ (i.e., $w \in V_k^-$), or else $l_k(w) = 0$. In the latter case, $n-1$ linearly independent constraints of P_k must be binding for w, so w is the intersection of the edge determined by these $n-1$ constraints with the hyperplane $l_k(y) = 0$. Obviously, one endpoint of this edge must be some $v \in V_k$ such that $l_k(v) > 0$. $\square$

Assume now that the polytopes P_k and P_{k+1} are nondegenerate and for each vertex $v \in V_k$ the list $N(v)$ of all its neighbors (adjacent vertices) together with the index set $I(v)$ of all binding constraints for v is available. Also assume that $|V_k^+| \leq |V_k^-|$ (otherwise, interchange the roles of these sets). Based on Proposition 7.6, $V_{k+1} = V_k^- \cup W$, where W (set of new vertices) can be computed as follows (Chen et al. 1991):

On-Line Vertex Enumeration Procedure

Step 1. Set $W \leftarrow \emptyset$ and for all $v \in V_k^+$:

- for all $u \in N(v) \cap V_k^-$:
  - compute $\lambda \in [0,1]$ such that $l_k(\lambda v + (1-\lambda)u) = 0$;
  - set $w = \lambda v + (1-\lambda)u$; $N(w) = \{u\}, N(u) \leftarrow (N(u) \setminus \{v\}) \cup \{w\}$; $I(w) = (I(v) \cap I(u)) \cup \{\nu\}$ where ν denotes the index of the new constraint.
  - set $W \leftarrow W \cup \{w\}$.

Step 2. For all $v \in W, u \in W$:

- if $I(u) \cap I(v) = n-1$ then set $N(v) \leftarrow N(v) \cup \{u\}$, $N(u) \leftarrow N(u) \cup \{v\}$.

In the context of outer approximation, if P_k is nondegenerate then P_{k+1} can be made nondegenerate by slightly perturbing the hyperplane $l_k(x) = 0$ to prevent it from passing through any vertex of P_k. In this way, by choosing P_1 to be nondegenerate, one can, at least in principle, arrange to secure nondegeneracy during the whole process. Earlier variants of on-line vertex enumeration procedures have been developed by Falk and Hoffman (1976), Thieu et al. (1983), Horst et al. (1988), and Thieu (1989).

Remark 7.7 If an interior point x^0 of D is available, the point $y^k \in [x^0, x^k] \cap \partial D$ is feasible and $\overline{x}^k \in \operatorname{argmin}\{f(y^1), \ldots, f(y^k)\}$ is the incumbent at iteration k. Therefore, a global ε-optimal solution is obtained when $f(\overline{x}^k) - f(x^k) \leq \varepsilon$, which must occur after finitely many iterations. The corresponding algorithm was first proposed by Hoffman (1981). A major difficulty with simple OA methods for global optimization is the rapid growth of the set V_k which creates serious storage problems and often makes the computation of V_{k+1} a formidable task.

Remark 7.8 When D is a polyhedron, $g(x) = \max_{i=1,\ldots,m}(\langle a^i, x\rangle - b_i)$, so if $x^k \notin D$, then $\langle a^{i_k}, x^k\rangle - b_{i_k} > 0$ for $i_k \in \operatorname{argmax}\{\langle a^i, x^k\rangle - b_i |\ i = 1, \ldots, m\}$, and $\nabla g(x^k) = a^{i_k}$. Hence the cut (7.33) is $l_k(x) = \langle a^{i_k}, x\rangle - b_{i_k} \leq 0$, and coincides thus with the constraint the most violated by x^k. Therefore, for (BCP) a simple OA method produces an exact optimal solution after finitely many iterations. Despite this and several other positive features, the uncontrolled growth of V_k is a handicap which prevents simple OA methods from competing favorably with branch and bound methods, except on problems of small size. In a subsequent section a dual version of OA methods called "polyhedral annexation" will be presented which offers restart and multistart possibilities, and hence, can perform significantly better than simple OA procedures in many circumstances. Let us also mention a method for solving (*BCP*) via "collapsing polytopes" proposed by Falk and Hoffman (1986) which can be interpreted as an OA method with the attractive feature that at each iteration only $n + 1$ new vertices are to be computed for P_k. Unfortunately, strong requirements on nondegeneracy limit the applicability of this ingenious algorithm.

7.2.2 Combined Algorithm

By replacing $f(x)$, if necessary, with $\min\{\rho, f(x)\}$ where $\rho > 0$ is sufficiently large, we may assume that the upper level sets of the concave function $f(x)$ are bounded. According to (7.35), the distinguished point x^k of P_k in the above OA scheme is a global optimal solution of the relaxed problem

$$(SP_k) \qquad\qquad \min\{f(x) |\ x \in P_k\}.$$

The determination of x^k in (7.35) involves the computation of the vertex set V_k of P_k. However, as said in Remark 7.7, a major difficulty of this approach is that, for many problems the set V_k may grow very rapidly and attain a prohibitively large size. To avoid computing V_k an alternative approach is to compute x^k by using a BB Algorithm, e.g., the BB conical procedure, for solving (SP_k). A positive feature of this approach that can be exploited to speed up the convergence is that the relaxed problem (SP_k) differs from (SP_{k-1}) only by one single additional linear constraint. Due to this fact, it does not pay to solve each of these problems to optimality. Instead, it may suffice to execute only a restricted number of iterations (even one single iteration) of the conical procedure for (SP_k), and take $x^k = \omega(M_k)$, where M_k is the current candidate for branching. At the next iteration $k + 1$, the current conical partition in the conical procedure for (SP_k) may be used to start the conical procedure for (SP_{k+1}). This idea of combining OA with BB concepts leads to the following hybrid procedure (Fig. 7.4).

Combined OA/BB Conical Algorithm for (GCP)

Initialization Take an n-simplex $[s^1, \ldots, s^{n+1}]$ inscribed in D and an interior point x^0 of this simplex. For each $i = 1, \ldots, n + 1$ compute the intersection $\bar{s}^i$ of the halfline from x^0 through s^i with the boundary of D, Let $\bar{x}^1 \in \operatorname{argmin}\{f(\bar{s}^1), \ldots, f(\bar{s}^{n+1})\}$, $\gamma_1 = f(\bar{x}^1)$, Let M_i be the cone with vertex at x^0 and n edges passing through $\bar{s}^j, j \neq i$, $\mathcal{N}_1 = \{M_i, i = 1, \ldots, n + 1\}$, $\mathcal{P}_1 = \mathcal{N}_1$. (For any subcone of a cone $M \in \mathcal{P}_1$ its base is taken to be the simplex spanned by the n intersection points of the edges of M with the boundary of D) Let P_1 be an initial polytope enclosing D. Set $k = 1$.

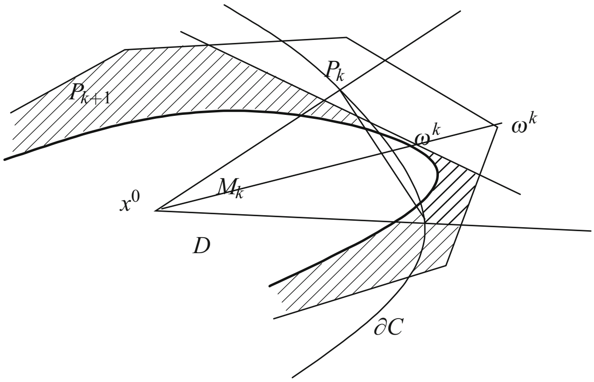


Fig. 7.4 Combined algorithm

Step 1. (*Bounding*) For each cone $M \in \mathscr{P}_k$ of base $[u^1, \ldots, u^n]$ compute the point $x^0 + \theta_i u^i$ where the ith edge of M meets the surface $f(x) = \gamma_k$. Let U be the matrix of columns $u^1, \ldots, u^n$. Solve the linear program

$$LP(P_k, M) \qquad \max \left\{ \sum_{i=1}^{n} \frac{t_i}{\theta_i} \middle| \quad x^0 + Ut \in P_k, \ t \geq 0 \right\}$$

to obtain the optimal value $\mu(M)$ and a basic optimal solution $\bar{t}(M)$. Let $\omega(M) = x^0 + U\bar{t}(M)$. If $\mu(M) > 1$, compute $z^i = x^0 + \mu(M)\theta_i u^i, \ i = 1, \ldots, n$ and

$$\beta(M) = \min\{f(z^1), \ldots, f(z^n)\}. \tag{7.36}$$

Step 2. (*Pruning*) Delete every cone $M \in \mathscr{N}_k$ such that $\mu(M) \leq 1$ or $\beta(M) \geq \gamma_k - \varepsilon$. Let $\mathscr{R}_k$ be the collection of remaining members of $\mathscr{N}_k$.

Step 3. (*Termination criterion 1*) If $\mathscr{R}_k = \emptyset$ terminate: $\overline{x}^k$ is a global ε-optimal solution.

Step 4. (*Termination criterion 2*) Select $M_k \in \operatorname{argmin}\{\beta(M) | \ M \in \mathscr{R}_k\}$. If $\omega^k := \omega(M_k) \in D$ terminate: ω^k is a global optimal solution.

Step 5. (*Branching*) Subdivide M_k via the halfline from x^0 through ω^k, obtaining the partition $\mathscr{P}_{k+1}$ of M_k.

Step 6. (*Outer approximation*) Compute $p^k \in \partial g(\overline{\omega}^k)$, where $\overline{\omega}^k$ is the intersection of the halfline from x^0 through ω^k with the boundary of D and define

$$P_{k+1} = P_k \cap \{x | \ \langle p^k, x - \overline{\omega}^k \rangle + g(\overline{\omega}^k) \leq 0\}.$$

Step 7. (*Updating*) Let $\overline{x}^{k+1}$ be the best of the feasible points $\overline{x}^k, \overline{\omega}^k$. Let $\gamma_{k+1} = f(\overline{x}^{k+1})$, $\mathscr{N}_{k+1} = (\mathscr{R}_k \setminus \{M_k\}) \cup \mathscr{P}_{k+1}$. Set $k \leftarrow k + 1$ and go back to Step 1.

Theorem 7.5 *The Combined Algorithm can be infinite only if $\varepsilon = 0$, and in that case every accumulation point of the sequence $\{\overline{x}^k\}$ is a global optimal solution of (GCP).*

Proof First observe that for any k we have $\beta(M_k) \leq \min f(D)$ while $f(\omega^k) \geq \min\{f(z^{k1}), \ldots, f(z^{kn})\} = \beta(M_k)$. Therefore, when Step 4 occurs ω^k is effectively a global optimal solution. We now show that for $\varepsilon > 0$ if Step 5 never occurs the current best value γ_k decreases by at least a quantity ε after finitely many iterations. In fact, suppose that $\gamma_k = \gamma_{k_0}$ for all $k \geq k_0$. Let $C = \{x \in \mathbb{R}^n | f(x) \geq \gamma_{k_0}\}$ and by z^{ki} denote the intersection point of the ith edge of M_k with ∂C. Let $\mu_k = \mu(M_k)$ and by q^k denote the intersection point of the halfline from x^0 through ω^k with the simplex $[z^{k1}, \ldots, z^{kn}]$. By Theorem 6.2 there exists a subsequence $\{k_s\} \subset \{1, 2 \ldots, \}$ and a point $q \in \partial C$ such that $q^{k_s} \to q$ as $s \to +\infty$. Moreover, $\mu_{k_s} \to 1$ as $s \to +\infty$ [see (7.22) in Remark 7.3]. Since $\omega^{k_s} - x^0 = \mu_{k_s}(q^{k_s} - x^0)$ it follows that $\omega^{k_s} \to q \in \partial C$, hence $f(\omega^{k_s}) \to f(q) = \gamma_{k_0}$ as $s \to +\infty$. But $f(\omega^k) \geq \beta(M_k)\ \forall k$, hence $\beta(M_{k_s}) \leq f(\omega^{k_s}) \to \gamma_{k_0}$ as $s \to +\infty$. Therefore, for s sufficiently large we will have $\beta(M_{k_s} \geq \gamma_{k_0} - \varepsilon = \gamma_{k_s} - \varepsilon$, and then the algorithm stops by the criterion in Step 3 yielding γ_{k_0} as the global ε-optimal value and $\overline{x}^{k_s}$ as a global ε-optimal solution. Thus if γ_{k_0} is not yet the global ε-optimal value, we must have $\gamma_k \leq \gamma_{k_0} - \varepsilon$ for some $k > k_0$. Since $\min f(D) < +\infty$, this proves the finiteness of the algorithm. □

Remark 7.9 If for solving (SP_k) the CS Algorithm is used instead of the BB algorithm, then we have a Combined OA/CS Conical Algorithm for (GCP). In Step 1 of the latter algorithm, the values θ_i are computed from the condition $f(x^0 + \theta_i u^i) = \gamma_k - \varepsilon$, and there is no need to compute $\beta(M)$ while in Step 2 a cone M is deleted if $\mu(M) \leq 1$. The OA/CS Algorithm involves less function evaluations than the OA/BB Algorithm.

Combined algorithms of the above type were introduced for general global optimization problems in Tuy and Horst (1988) under the name of restart BB Algorithms. A related technique of combining BB with OA developed in Horst et al. (1991) leads essentially to the same algorithm in the case of (GCP).

7.3 Reverse Convex Programming

A reverse convex inequality is an inequality of the form

$$h(x) \geq 0, \tag{7.37}$$

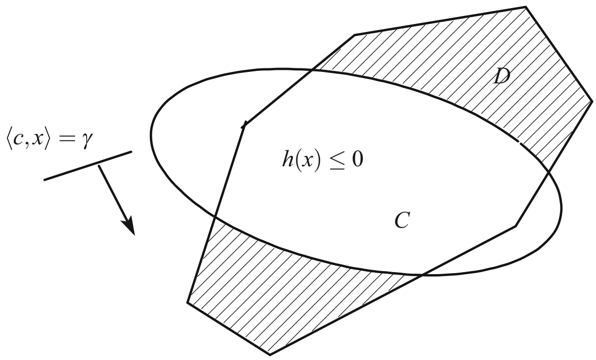
where $h : \mathbb{R}^n \to \mathbb{R}$ is a convex function. When such an inequality is added to the constraints of a linear program $\min\{cx | Ax \leq b, x \leq 0\}$, the latter becomes a complicated nonconvex global optimization problem

$$(LRC) \qquad \text{minimize } cx \tag{7.38}$$

$$\text{subject to } Ax \leq b, \quad x \geq 0 \tag{7.39}$$

$$h(x) \geq 0. \tag{7.40}$$

Fig. 7.5 Reverse convex constraint



The difficulty is brought on by the additional reverse convex constraint which destroys the convexity and sometimes the connectedness of the feasible set (Fig. 7.5). Even when $h(x)$ is as simple as a convex quadratic function, testing the feasibility of the constraints (7.39)–(7.40) is already NP-hard. This is no surprise since (*BCP*), which is NP-hard, can be written as a special (*LRC*):

$$\min\{t| \quad Ax \le b,\ x \ge 0,\ t - f(x) \ge 0\}.$$

Reverse convex constraints occur in many problems of mechanics, engineering design, economics, control theory, and other fields. For instance, in engineering design a control variable x may be subject to constraints of the type $\|x - d\| \ge \delta$ (, i.e., $h(x) := \|x - d\| - \delta \ge 0$), or $\langle p^i, x\rangle \ge \alpha_i$ for at least one $i \in I$ (i.e., $h(x) = \max\limits_{i\in I}(\langle p^i, x\rangle - \alpha_i) \ge 0$). In combinatorial optimization a discrete constraint such as $x_i \in \{0, 1\}$ for all $i \in I$ amounts to requiring that $0 \le x_i \le 1, i \in I$, and $h(x) := \sum_{i\in I} x_i(x_i - 1) \ge 0$. Also, in bilevel linear programming problems, where the variable is $x = (y, z) \in \mathbb{R}^p \times \mathbb{R}^q$, a constraint of the form $z \in \mathrm{argmax}\{dz|Py+Qz \le s\}$ can be expressed as $h(x) := \varphi(y) - dz \ge 0$, where $\varphi(y) := \max\{dz|Qz \le Py - s\}$ is a convex function of y.[1]

7.3.1 General Properties

Setting $D = \{x|Ax \le b, x \ge 0\}$, $C = \{x|h(x) \le 0\}$, we can write the problem as

$$(LRC): \qquad \min\{cx| \quad x \in D \setminus \mathrm{int}C\}. \tag{7.41}$$

[1]In the above formulation the function $h(x)$ is assumed to be finite throughout $\mathbb{R}^n$; if this condition may not be satisfied (as in the last example) certain results below must be applied with caution.

Here D is a polyhedron and C is a closed convex set. For the sake of simplicity we will assume D nonempty and C bounded. Furthermore, it is natural to assume that

$$\min\{cx|x \in D\} < \min\{cx|x \in D \setminus \mathrm{int}C\}. \tag{7.42}$$

which simply means that the reverse convex constraint $h(x) \geq 0$ is essential, so that (LRC) does not reduce to the trivial underlying linear program. Thus, if w is a basic optimal solution of the linear program then

$$w \in D \cap \mathrm{int}C \quad \text{and} \quad cw < cx, \ \forall x \in D \setminus \mathrm{int}C. \tag{7.43}$$

Proposition 7.7 *For any $z \in D \setminus C$ there exists a better feasible solution y lying on the intersection of the boundary ∂C of C with an edge E of the polyhedron D.*

Proof We show how to compute y satisfying the desired conditions. If $p \in \partial h(z)$ then the polyhedron $\{x \in D|\ \langle p, x-z\rangle + h(z) = 0\}$ is contained in the feasible set, and a basic optimal solution z^1 of the linear program $\min\{cx|\ x \in D,\ \langle p, x-z\rangle + h(z) = 0\}$ will be a vertex of the polyhedron $P = \{x \in D|\ cx \leq cz^1\}$. But in view of (7.43) z^1 is not optimal to the linear program $\min\{cx|\ x \in P\}$. Therefore, starting from z^1 and using simplex pivots for solving the latter program we can move from a vertex of D to a better adjacent vertex until we reach a vertex u with an adjacent vertex v such that $h(u) \geq 0$ but $h(v) < 0$. The intersection of ∂C with the line segment $[u, v]$ will be the desired y. □

Corollary 7.3 (Edge Property) *If (LRC) is solvable then at least one optimal solution lies on the intersection of the boundary of C with an edge of D.*

If $\mathscr{E}$ denotes the collection of all edges of D and for every $E \in \mathscr{E}$, $\sigma(E)$ is the set of extreme points of the line segment $E \cap C$, then it follows from Corollary 7.3 (Hillestad and Jacobsen 1980b) that an optimal solution must be sought only among the finite set

$$X = \bigcup\{\sigma(E)|\ E \in \mathscr{E}\}. \tag{7.44}$$

We next examine how to recognize an optimal solution. Problem (LRC) is said to be *regular (stable)* if for every feasible point x there exists, in every neighborhood of x, a feasible point x' such that $h(x') > 0$. Since by slightly moving x' towards an interior point of D, if necessary, one can make $x' \in \mathrm{int}D$, this amounts to saying that the feasible set $S = D \setminus \mathrm{int}C$ satisfies $S = \mathrm{cl}(\mathrm{int}D)$, i.e., it is *robust*. For example, the problem depicted in Fig. 7.5 is regular, but the one in Fig. 7.6 is not (the point $x = opt$ which is an isolated feasible point does not satisfy the required condition). By specializing Proposition 5.3 to (LRC) we obtain

Theorem 7.6 (Optimality Criterion) *In order that a feasible point $\overline{x}$ be a global optimal solution it is necessary that*

$$\{x \in D|\ cx \leq c\overline{x}\} \subset C. \tag{7.45}$$

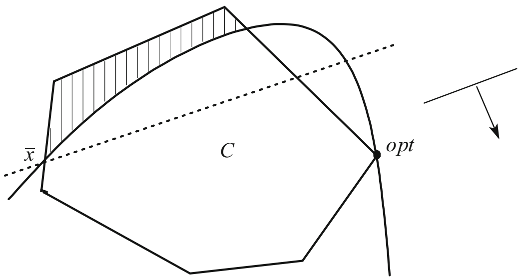


Fig. 7.6 Nonregular problem

This condition is also sufficient if the problem is regular. □

For any real number γ let $D(\gamma) = \{x \in D|\ cx \leq \gamma\}$. Since $D(\gamma) \subset D(\gamma')$ whenever $\gamma \leq \gamma'$, it follows from the above theorem that under the regularity assumption , the optimal value in (LRC) is

$$\inf\{\gamma|\quad D(\gamma) \setminus C \neq \emptyset\}. \tag{7.46}$$

Also note that by letting $\gamma = c\bar{x}$ condition (7.45) is equivalent to

$$\max\{h(x)|\ x \in D(\gamma)\} \leq 0, \tag{7.47}$$

where the maximization problem on the left-hand side is a (BCP) since $h(x)$ is convex. In the general case when regularity is not assumed, condition (7.45) [i.e. (7.47)] may hold without $\bar{x}$ being a global optimal solution, as illustrated in Fig. 7.6.

Fortunately, a slight perturbation of (LRC) can always make it regular, as shown in the following:

Proposition 7.8 *For $\varepsilon > 0$ sufficiently small, the problem*

$$\text{minimize} \quad cx \quad \text{subject to} \quad x \in D,\ h(x) + \varepsilon(\|x\|^2 + 1) \geq 0 \tag{7.48}$$

is regular and its optimal value tends to the optimal value of (LRC) as $\varepsilon \to 0$.

Proof. Since the vertex set V of D is finite, there exists $\varepsilon_0 > 0$ so small that $\varepsilon \in (0, \varepsilon_0)$ implies that $h_\varepsilon(x) := h(x) + \varepsilon(\|x\|^2 + 1) \neq 0,\ \forall x \in V$. Indeed, if $V_1 = \{x \in V|h(x) < 0\}$ and ε_0 satisfies $h(x) + \varepsilon_0(\|x\|^2 + 1) < 0,\ \forall x \in V_1$, then whenever $0 < \varepsilon < \varepsilon_0$ we have $h(x) + \varepsilon(\|x\|^2 + 1) < 0,\ \forall x \in V_1$, while $h(x) + \varepsilon(\|x\|^2 + 1) \geq 0 + \varepsilon > 0,\ \forall x \in V \setminus V_1$. Now consider problem (7.48) for $0 < \varepsilon < \varepsilon_0$ and let $x \in D$ be such that $h_\varepsilon(x) \geq 0$. If $x \notin V$ then x is the midpoint of some line segment $\Delta \subset D$, and since the function $h_\varepsilon(x)$ is strictly convex, any neighborhood of x must contain a point x' of Δ such that $h_\varepsilon(x') > 0$. On the other hand, if $x \in V$, then $h_\varepsilon(x) > 0$, and any point x' of D sufficiently near to x will satisfy $h_\varepsilon(x') > 0$. Thus, given any feasible solution x of problem (7.48) there exists a feasible solution x' arbitrarily close to x such that $h_\varepsilon(x') > 0$. This means that problem (7.48) is regular. If x_ε is a global optimal solution of (7.48) then obviously $cx_\varepsilon \leq \gamma := \min\{cx|x \in D, h(x) \geq 0\}$ and since $h_\varepsilon(x_\varepsilon) \geq 0$ and $h_\varepsilon(x_\varepsilon) - h(x_\varepsilon) \to 0$ as $\varepsilon \to 0$, if $\overline{x}$ is any cluster point of $\{x_\varepsilon\}$ then $h(\overline{x}) \geq 0$, i.e., $\overline{x}$ is feasible to (LRC), hence $c\overline{x} = \gamma$. □

Note that the function $h_\varepsilon(x)$ in the above perturbed problem (7.48) is *strictly convex*. This additional feature may sometimes be useful.

An alternative approach for handling nonregular problems is to replace the original problem by the following:

$$(LRC_\varepsilon) \qquad \min\{cx| \quad x \in D,\ h(x) \geq -\varepsilon\}. \tag{7.49}$$

A feasible solution x_ε to the latter problem such that

$$cx_\varepsilon \leq \min\{cx|\, x \in D, h(x) \geq 0\} \tag{7.50}$$

is called an *ε-approximate optimal solution* to (LRC). The next result shows that for $\varepsilon > 0$ sufficiently small an ε-approximate optimal solution is as close to an optimal solution as desired.

Proposition 7.9 *If for each $\varepsilon > 0$, x_ε is an ε- approximate optimal solution of (LRC), then any accumulation point $\overline{x}$ of the sequence $\{x_\varepsilon\}$ solves (LRC).*

Proof Clearly $h(\overline{x}) \geq 0$, hence $\overline{x}$ is feasible to (LRC). For any feasible solution x of (LRC), the condition (7.50) implies that $c\overline{x} \leq \min\{cx|\, x \in D, h(x) \geq 0\}$, hence $\overline{x}$ is an optimal solution of (LRC). □

Observe that if $C_\varepsilon = \{x|\ h(x) \leq -\varepsilon\}$ then a feasible solution x_ε to the problem (7.49) such that

$$\{x \in D|\ cx \leq cx_\varepsilon\} \subset C_\varepsilon \tag{7.51}$$

is an ε-approximate optimal solution of (LRC). This together with Proposition 7.7 suggests that, within a tolerance $\varepsilon > 0$ (so small that still $x^0 \in \mathrm{int}C_\varepsilon$), transcending an incumbent value γ in (LRC) reduces to solving the following problem, analogous to problem (#) for (BCP) in Sect. 7.1, Subsect. 7.1.3:

($\natural$) *Given two closed convex sets $D(\gamma) = \{x \in D|\ cx \leq \gamma\}$ and $C_\eta = \{x|\ h(x) \leq -\eta\}$ $(0 < \eta < \varepsilon)$, find a point $z \in D(\gamma) \setminus C_\varepsilon$ or else prove that $D(\gamma) \subset C_\eta$.*

Note that here $C_\varepsilon \subset C_\eta$. By Proposition 7.7, from a point $z \in D(\gamma) \setminus C_\varepsilon$ one can easily derive a feasible point y to (LRC_ε) with $cy < \gamma$; on the other hand, if $D(\gamma) \subset C_\eta$, then γ is an ε-approximate optimal value because $0 < \eta < \varepsilon$.

7.3.2 Solution Methods

Thus the search for an ε-approximate optimal solution of (*LRC*) reduces to solving the *parametric dc feasibility problem*

$$\inf\{\gamma| \quad D(\gamma) \setminus C_\varepsilon \neq \emptyset\}. \tag{7.52}$$

analogous to (7.5) for (*BCP*). This leads to a method for solving (*LRC*) which is essentially the same as that used for solving (*BCP*).

In the following X_ε denotes the set analogous to the set X in (7.44) but defined for problem (LRC_ε) instead of (*LRC*).

CS (Cone Splitting) Restart Algorithm for (LRC)

Initialization Solve the linear program $\min\{cx|\ x \in D\}$ to obtain a basic optimal solution w. If $h(w) \geq -\varepsilon$, terminate: w is an ε-approximate optimal solution. Otherwise:

(a) If a vertex z^0 of D is available such that $h(z^0) > -\varepsilon$, go to Phase 1.
(b) Otherwise, set $\gamma = \infty, D(\gamma) = D$, and go to Phase 2.

Phase 1. Starting from z^0 search for a feasible point $\overline{x} \in X_\varepsilon$ [see (7.44)]. Let $\gamma = c\overline{x}, D(\gamma) = \{x \in D|\ cx \leq \gamma\}$. Go to Phase 2.

Phase 2. Call Procedure DC to solve problem (♮) (i.e., apply Procedure DC in Sect. 6.1 with $D(\gamma)$, C_ε and C_η in place of D, C, C_ε resp.).

(1) If a point $z \in D(\gamma) \setminus C_\varepsilon$ is obtained, set $z^0 \leftarrow z$ and go to Phase 1.
(2) If $D(\gamma) \subset C_\eta$ then terminate: $\overline{x}$ is an ε-approximate optimal solution (when $\gamma < +\infty$), or the problem is infeasible (when $\gamma = +\infty$).

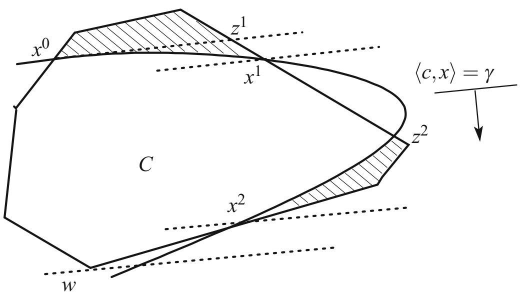


Fig. 7.7 The FW/BW algorithm

Remark 7.10 In Phase 1, the point $\overline{x} \in X$ can be computed by performing a number of simplex pivots as indicated in the proof of Proposition 7.7. In Phase 2, the initial cone of Procedure DC is the canonical cone associated with the vertex $x^0 = w$ of D (w is also a vertex of the polyhedron $D(\gamma)$). Also note that $\varepsilon > 0$, whereas the algorithm does not assume regularity of the problem. However, if the problem is regular, then the algorithm with $\eta = \varepsilon = 0$ will find an exact global optimal solution after finitely many steps, though may require infinitely many steps to identify it as such.

Since the current feasible solution moves in Phase I forward from the region $h(x) > 0$ to the frontier $h(x) = 0$, and in Phase 2 backward from this frontier to the region $h(x) > 0$, the algorithm is also called the **forward–backward (FW/BW) algorithm** (Fig. 7.7).

Example 7.2

$$\begin{array}{lll} \text{Minimize } -2x_1 + x_2 & & \\ \text{subject to } x_1 + x_2 & \leq & 10 \\ \quad -x_1 + 2x_2 & \leq & 8 \\ \quad x_1 - x_2 & \leq & 4 \\ \quad -2x_1 - 3x_2 & \leq & -6 \\ \quad x_1, x_2 & \geq & 0 \\ \quad x_1^2 - x_1x_2 + x_2^2 - 6x_1 & \geq & 0\ . \end{array}$$

The optimal solution $\bar{x} = (4.3670068; 5.6329932)$ is obtained after one cycle (Fig. 7.8).

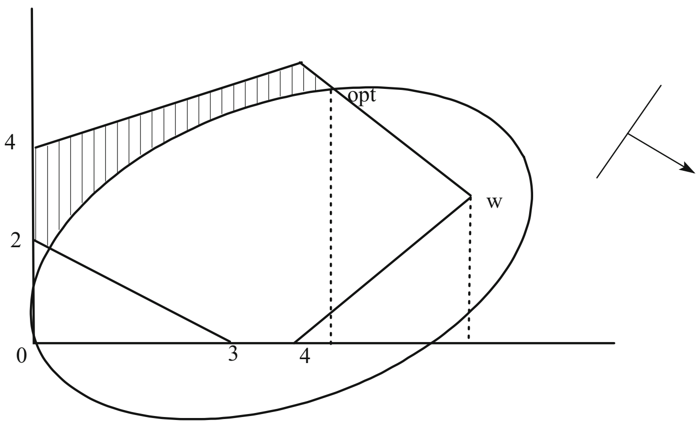


Fig. 7.8 An example

Just as the CS Algorithm for (BCP) admits of a BB version, the CS Algorithm for (LRC) can be transformed into a BB Algorithm by using lower bounds to define the selection operation in Phase 2 .

Let M_0 be the canonical cone associated with the vertex $x^0 = w$ of D, and let $z^{i0}, i = 1, \ldots, n$, be the points where the ith edge of M_0 meets the boundary ∂C of the closed convex set $C = \{x|\ h(x) \leq 0\}$. Consider a subcone M of M_0, of base

$S = [z^1, \ldots, z^n] \subset S_0$. Let $y^i = x^0 + \theta_i u^i$ be the point where the ith edge of M meets ∂C and let U be the nonsingular $n \times n$ matrix of columns $u^1, \ldots, u^n$. Let $\tilde{A}$ and $\tilde{b}$ be defined as in (7.20).

Proposition 7.10 *A lower bound for* $\min\{cx|\ x \in D \cap M, h(x) \geq 0\}$ *is given by the optimal value* $\beta(M)$ *of the linear program*

$$\min\left\{\langle c, x^0 + Ut\rangle | \quad \tilde{A}Ut \leq \tilde{b},\ \sum_{i=1}^{n} \frac{t_i}{\theta_i} \geq 1,\ t \geq 0\right\}. \tag{7.53}$$

Furthermore, if a basic optimal solution $\bar{t}(M)$ *of this linear program is such that* $\omega(M) = x^0 + U\bar{t}(M)$ *lies on some edge of* M *then* $\omega(M) \in S$*, and hence* $\beta(M) = \min\{cx|\ x \in M \cap S\}$*, where* $S = D \setminus \mathrm{int}C$ *is the feasible set of (LRC).*

Proof Clearly $x \in M \setminus \mathrm{int}C$ if and only if $x = x^0 + Ut,\ t \geq 0,\ \sum_{i=1}^n \frac{t_i}{\theta_i} \geq 1$. Since $M \cap S \subset (M \cap D) \setminus \mathrm{int}C$, a lower bound of $\min\{cx|\ x \in M \cap S\}$ is given by the optimal value $\beta(M)$ of the linear program

$$\min\{\langle c, x^0 + Ut\rangle | \quad x^0 + Ut \in D,\ t \geq 0,\ \sum_{i=1}^{n} \frac{t_i}{\theta_i} \geq 1\} \tag{7.54}$$

which is nothing but (7.53). To prove the second part of the proposition, observe that if $\omega(M)$ lies on the ith edge it must satisfy $\omega(M) = x^0 + \lambda(y^i - x^0)$ for some $\lambda \geq 1$ and since $x^0 \in D, \omega(M) \in D$ it follows by convexity of D that $y^i \in D$, hence $y^i \in S$. In view of assumption (7.42) this implies that $cx^0 < cy^i$, i.e., $c(y^i - x^0) > 0$, and hence $\langle c, \omega(M) - y^i\rangle = \langle c, (\lambda - 1)y^i\rangle > 0$ if $\lambda > 1$. Since $c\omega(M) \leq cy^i$ from the definition of $\omega(M)$ we must have $\lambda = 1$, i.e., $\omega(M) \in D \setminus \mathrm{int}C$. □

As a consequence of this proposition, when $\omega(M)$ is not feasible to (*LRC*) then it does not lie on any edge of M, so the subdivision of M upon the halfline from x^0 through $\omega(M)$ is proper. Also, if all y^i , $i = 1, \ldots, n$, belong to D, then $\omega(M)$ is decidedly feasible (and is one of the y^i), hence $\beta(M)$ is the exact minimum of cx over the feasible portion in M and M will be fathomed in the branch and bound process. Thus, in a conical procedure using the above lower bounding, only those cones will actually remain for consideration which contain boundary points of the feasible set. In other words, the search will concentrate on regions which are potentially of interest.

In the next algorithm, let $\varepsilon \geq 0$ be the tolerance.

Conical BB Algorithm for (LRC)

Initialization Solve the linear program $\min\{cx|\ x \in D\}$ to obtain a basic optimal solution w. If $h(w) \geq -\varepsilon$, terminate: w is an ε-approximate optimal solution. Otherwise:

(a) If a vertex z^0 of D is available such that $h(z^0) > -\varepsilon$, go to Phase 1.
(b) Otherwise, set $\gamma_1 = \infty, D(\gamma_1) = D$, and go to Phase 2.

Phase 1. Starting from z^0 search for a point $\overline{x}^1 \in X_\varepsilon$ [see (7.44)]. Let $\gamma_1 = c\overline{x}^1, D(\gamma_1) = \{x \in D|\ cx \leq \gamma_1\}$. Go to Phase 2.

Phase 2 Let M_0 be the canonical cone associated with the vertex $x^0 = w$ of D, let $\mathscr{N}_1 = \{M_0\}, \mathscr{P}_1 = \{M_0\}$. Set $k = 1$.

Step 1. (*Bounding*) For each cone $M \in \mathscr{P}_k$ whose base is $[x^0 + u^1, \ldots, x^0 + u^n]$ and whose edges meet ∂C at points $y^i = x^0 + \theta_i u^i,\ i = 1, \ldots, n$, solve the linear program (7.53), where U denotes the matrix of columns $u^i = y^i - x^0,\ i = 1, \ldots, n$. Let $\beta(M), \overline{t}(M)$ be the optimal value and a basic optimal solution of this linear program, and let $\omega(M) = x^0 + U\overline{t}(M)$.

Step 2. (*Options*) If some $\omega(M), M \in \mathscr{P}_k$, are feasible to ($LRC_\varepsilon$) (7.49) and satisfy $\langle c, \omega(M)\rangle < \gamma_k$, then reset γ_k equal to the smallest among these values and $\overline{x}^k \leftarrow \arg\gamma_k$ (i.e., the x such that $\langle c, x\rangle = \gamma_k$). Go to Step 3 or (optionally) return to Phase 1.

Step 3. (*Pruning*) Delete every cone $M \in \mathscr{N}_k$ such that $\beta(M) \geq \gamma_k$ and let $\mathscr{R}_k$ be the collection of remaining cones.

Step 4. (*Termination Criterion*) If $\mathscr{R}_k = \emptyset$, then terminate:

(a) If $\gamma_k < +\infty$, then x^k is an ε-approximate optimal solution.
(b) Otherwise, (LRC) is infeasible.

Step 5. (*Branching*) Select $M_k \in \operatorname{argmin}\{\beta(M)|\ M \in \mathscr{R}_k\}$. Subdivide M_k along the halfline from x^0 through $\omega^k = \omega(M_k)$. Let $\mathscr{P}_{k+1}$ be the partition of M_k.

Step 6. (*New Net*) Set $\mathscr{N}_k \leftarrow (\mathscr{R}_k \setminus \{M_k\}) \cup \mathscr{P}_{k+1}$, $k \leftarrow k + 1$ and return to Step 1.

Theorem 7.7 *The above algorithm can be infinite only if $\varepsilon = 0$, and then any accumulation point of each of the sequences $\{\omega^k\}, \{\overline{x}^k\}$ is a global optimal solution of (LRC).*

Proof. Since the set X_ε [see (7.44)] is finite, it suffices to show that Phase 2 is finite, unless $\varepsilon = 0$. If Phase 2 is infinite it generates at least a filter $\{M_k, k \in K\}$. Denote by $y^{ki},\ i = 1, \ldots, n$ the intersections of the edges of M_k with ∂C, by q^k the intersection of the halfline from x^0 through $\omega^k := \omega(M_k)$ with the hyperplane through $y^{k1}, \ldots, y^{kn}$. By Theorem 6.2, there exists at least one accumulation point of $\{q^k\}$, say $q = \lim q^k$ $(k \to \infty, k \in K_1 \subset K)$, such that $q \in \partial C$. We may assume that $\omega^k \to \overline{\omega}$ $(k \to \infty, k \in K_1)$. But, as a result of Steps 2 and 3, $\omega^k \in \operatorname{int} C_\varepsilon$, hence $\overline{\omega} \in \partial C_\varepsilon$. On the other hand, $q^k \in [x^0, \omega^k]$, hence $q \in [x^0, \overline{\omega}]$. This is possible only if $\varepsilon = 0$ and $\overline{\omega} = q \in \partial C$, so that $\overline{\omega}$ is feasible. Since, furthermore, $c\omega^k \leq \min\{cx|\ x \in D \setminus \operatorname{int} C\}$, $\forall k$, it follows that $\overline{\omega}$, and hence any accumulation point of the sequence $\{x^k\}$ is a global optimal solution. □

Remark 7.11 It should be emphasized that the above algorithm does not require regularity of the problem, even when $\varepsilon = 0$. However, in the latter case, since every ω^k is infeasible, the procedure may approach a global optimal solution through a sequence of infeasible points only. This situation is illustrated in Fig. 7.9, where

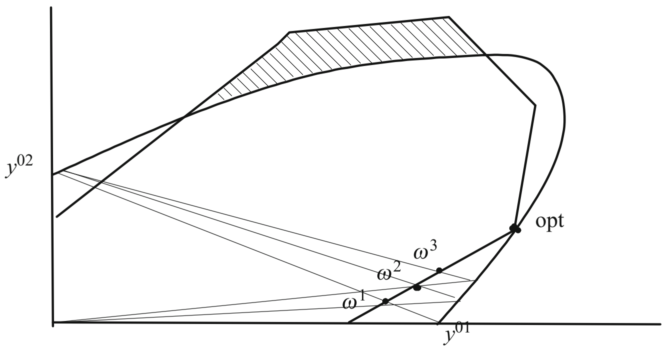


Fig. 7.9 Convergence through infeasible points

an isolated global optimal solution of a nonregular problem is approached by a sequence of infeasible points $\omega^1, \omega^2, \ldots$ A conical algorithm for (LRC) using bisections throughout was originally proposed by Muu (1985).

7.4 Canonical DC Programming

In its general formulation (cf. Sect. 5.3) the *Canonical DC Programming* problem is to find

$$(CDC) \quad \min\{f(x) |\; g(x) \leq 0 \leq h(x)\}, \tag{7.55}$$

where $f, g, h : \mathbb{R}^n \to R$ are convex functions. Clearly (LRC) is a special case of (CDC) when the objective function $f(x)$ is linear.

Proposition 7.11 *Any dc optimization problem can be reduced to the canonical form (CDC) at the expense of introducing at most two additional variables.*

Proof An arbitrary dc optimization problem

$$\min\{f_1(x) - f_2(x) |\; g_{i,1}(x) - g_{i,2}(x) \leq 0,\; i = 1, \ldots, m\} \tag{7.56}$$

can be rewritten as

$$\min\{f_1(x) - z |\; z - f_2(x) \leq 0,\; g(x) \leq 0,\; i = 1, \ldots, m\},$$

where the objective function is convex (in x, z) and $g(x) = \max_{i=1,\ldots,m}[g_{i,1}(x) - g_{i,2}(x)]$ is a dc function (cf. Proposition 4.1). Next the dc inequality $g(x) = p(x) - q(x) \leq 0$ with $p(x), q(x)$ convex can be split into two inequalities:

$$p(x) - t \leq 0, \quad t - q(x) \leq 0, \tag{7.57}$$

where the first is a convex inequality, and the second is a reverse convex inequality. By changing the notation if necessary, one obtains the desired canonical form. □

Whereas the application of OA technique to (GCP) is almost straightforward, it is much less obvious for (CDC) and requires special care to ensure convergence. By hypothesis the sets

$$D := \{x|\ g(x) \le 0\}, \quad C := \{x|\ h(x) \le 0\}$$

are closed and convex. For simplicity we will assume D bounded, although with minor modifications the results can be extended to the unbounded case.

If an optimal solution w of the convex program $\min\{f(x)|\ g(x) \le 0\}$ satisfies $h(w) \ge 0$, then the problem is solved (w is a global optimal solution). Therefore, without loss of generality we may assume that there exists a point w satisfying

$$w \in \mathrm{int}D \cap \mathrm{int}C, \quad f(x) > f(w)\ \forall x \in D \setminus \mathrm{int}C. \tag{7.58}$$

The next important property is an immediate consequence of this assumption.

Proposition 7.12 (Boundary Property) *Every global optimal solution lies on $D \cap \partial C$.*

Proof Let z^0 be any feasible solution. If $z^0 \notin \partial C$, then the line segment $[w, z^0]$ meets ∂C at a point $z^1 = (1-\lambda)w + \lambda z^0$ such that $0 < \lambda < 1$. By convexity we have from (7.58), $f(z^1) \le (1-\lambda)f(w) + \lambda f(z^0) < f(z^0)$, so z^1 is a better feasible solution than z^0. □

Just like (LRC), Problem (CDC) is said to be *regular* if the feasible set $S = D \setminus \mathrm{int}C$ is robust or, which amounts to the same,

$$D \setminus \mathrm{int}C = \mathrm{cl}(D \setminus C). \tag{7.59}$$

Theorem 7.6 for (LRC) extends to (CDC):

Theorem 7.8 (Optimality Criterion) *In order that a feasible solution $\overline{x}$ to (CDC) be global optimal it is necessary that*

$$\{x \in D|\, f(x) \le f(\overline{x})\} \subset C. \tag{7.60}$$

This condition is also sufficient if the problem is regular.

Proof Analogous to the proof of Proposition 5.3, with obvious modifications ($f(x)$ instead of $l(x)$). □

To exploit the above optimality criterion, it is convenient to introduce the next concept. Given $\varepsilon > 0$ (the tolerance), a vector $\overline{x}$ is said to be an *ε-approximate optimal solution* to (*CDC*) if

$$\overline{x} \in D, \quad h(\overline{x}) \geq -\varepsilon, \tag{7.61}$$

$$f(\overline{x}) \leq \min\{f(x) |\, x \in D,\ h(x) \geq 0\}. \tag{7.62}$$

Clearly, as $\varepsilon_k \downarrow 0$, any accumulation point of a sequence $\{\overline{x}^k\}$ of ε_k-approximate optimal solutions to (*CDC*) yields an exact global optimal solution. Therefore, in practice one should be satisfied with an ε-approximate optimal solution for ε sufficiently small. Denote $C_\varepsilon = \{x|\ h(x) \leq -\varepsilon\}, D(\gamma) = \{x \in D|\, f(x) \leq \gamma\}$. In view of (7.58) it is natural to require that

$$w \in \mathrm{int}D \cap \mathrm{int}C_\varepsilon, \qquad f(x) > f(w) \quad \forall x \in D \setminus \mathrm{int}C_\varepsilon. \tag{7.63}$$

Define $\overline{\gamma} = \inf\{f(x)|\, x \in D, h(x) > -\varepsilon\}$ and let

$$G = D(\overline{\gamma}), \qquad \Omega = D(\overline{\gamma}) \setminus \mathrm{int}C_\varepsilon.$$

By (7.63) and Proposition 7.12, it is easily seen that Ω coincides with the set of ε-approximate optimal solutions of (*CDC*), so the problem amounts to searching for a point $\overline{x} \in \Omega$. Denote by $\mathscr{P}$ the family of polytopes P in $\mathbb{R}^n$ for which there exists $\gamma \in [\overline{\gamma}, +\infty]$ satisfying $G \subset D(\gamma) \subset P$. For every $P \in \mathscr{P}$ define

$$x(P) \in \mathrm{argmax}\{h(x)|\, x \in V\}, \tag{7.64}$$

where V is the vertex set of P. Let us verify that Conditions A1 and A2 stated in Sect. 6.3 are satisfied. Condition A1 is obvious because $x(P)$ exists and can be computed provided $\Omega \neq \emptyset$; moreover, since $P \supset G \supset \Omega$, we must have $h(x(P)) \geq -\varepsilon$, so any accumulation point $\overline{x}$ of a sequence $x(P_k), k = 1, 2, \ldots$, satisfies $h(\overline{x}) \geq -\varepsilon$, and hence $\overline{x} \in \Omega$ whenever $\overline{x} \in G$.

To verify A2, consider any $z = x(P)$ associated to a polytope P such that $D(\gamma) \subset P$ for some $\gamma \in [\overline{\gamma}, +\infty]$. As we just saw, $h(z) \geq -\varepsilon$. If $h(z) < 0$ then $\max\{h(x)|\, x \in P\} < 0$, hence $D(\gamma) \subset \{x|\ h(x) < 0\}$, which implies that γ is an $\varepsilon-$approximate optimal value (if $\gamma < +\infty$) or the problem is infeasible (if $\gamma = +\infty$). If $h(z) \geq 0$ then $\max\{g(z), h(z) + \varepsilon\} > 0$ and since $\max\{g(w), h(w) + \varepsilon\} < 0$, we can compute a point y such that

$$y \in [w, z], \quad \max\{g(y), h(y) + \varepsilon\} = 0.$$

Two cases are possible:

(a) $g(y) = 0$: since $g(w) < 0$ this event may occur only if $g(z) > 0$ and so $z = x(P)$ can be separated from G by a cut $l(x) := \langle p, x - y\rangle \leq 0$ with $p \in \partial g(y)$.
(b) $h(y) = -\varepsilon$: then $g(y) \leq 0$, so y is an ε-approximate solution in the sense of (7.61). Furthermore, since $y = (1-\lambda)w + \lambda z, 0 < \lambda < 1)$, it follows that $f(y) \leq (1-\lambda)f(w) + \lambda f(z) < f(z)$, so $z = x(P)$ can be separated from G by a cut $l(x) := \langle p, x - y\rangle \leq 0$ with $p \in \partial f(y)$.

In either case, if we set $\gamma' = \min\{\gamma, f(y)\}$ then $P' = P \cap \{x|\ l(x) \leq 0\} \supset D(\gamma')$, i.e., $P' \in \mathscr{P}$.

Thus, both Conditions A1 and A2 are satisfied, and hence the outer approximation scheme described in Sect. 6.3 can be applied to solve (CDC).

7.4.1 Simple Outer Approximation

We are led to the following OA Procedure (Fig. 7.10) for finding an ε-approximate optimal solution of (CDC):

OA Algorithm for (CDC)

Step 0. Let $\overline{x}^1$ be the best feasible solution available, $\gamma_1 = f(\overline{x}^1)$ (if no feasible solution is known, set $\overline{x}^1 = \emptyset, \gamma_1 = +\infty$). Take a polytope $P_0 \supset D$ and let $P_1 = \{x \in P_0|\ f(x) \leq \gamma_1\}$. Determine the vertex set V_1 of P_1. Set $k = 1$.

Step 1. Compute $x^k \in \operatorname{argmax}\{h(x)|\ x \in V_k\}$. If $h(x^k) < 0$, then terminate:

(a) If $\gamma_k < +\infty, \overline{x}^k$ is an ε-approximate optimal solution;
(b) If $\gamma_k = +\infty$, the problem is infeasible.

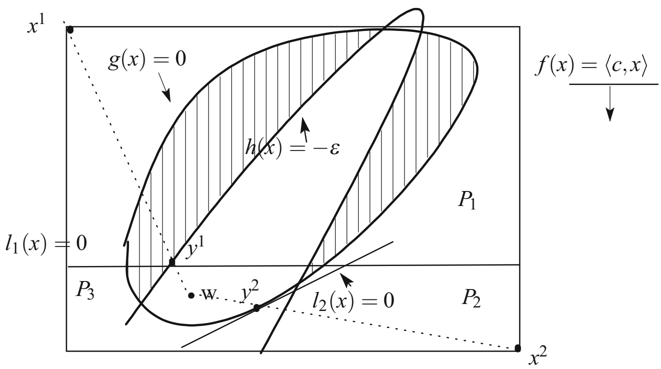


Fig. 7.10 Outer approximation algorithm for (CDC)

Step 2. Compute $y^k \in [w, x^k]$ such that $\max\{g(y^k), h(y^k) + \varepsilon\} = 0$. If $g(y^k) = 0$ then set $\gamma_{k+1} = \gamma_k, \overline{x}^{k+1} = \overline{x}^k$ and let

$$l_k(x) := \langle p^k, x - y^k \rangle, \quad p^k \in \partial g(y^k). \tag{7.65}$$

Step 3. If $h(y^k) = -\varepsilon$, then set $\gamma_{k+1} = \min\{\gamma_k, f(y^k)\}, \overline{x}^{k+1} = y^k$ if $f(y^k) \leq \gamma_k$, $\overline{x}^{k+1} = \overline{x}^k$ otherwise. Let

$$l_k(x) := \langle p^k, x - y^k \rangle, \quad p^k \in \partial f(y^k). \tag{7.66}$$

Step 4. Compute the vertex set V_{k+1} of $P_{k+1} = P_k \cap \{x |\ l_k(x) \leq 0\}$, set $k \leftarrow k+1$ and go back to Step 1.

Theorem 7.9 *The OA Algorithm for* (*CDC*) *can be infinite only if $\varepsilon = 0$ and in this case, if the problem is regular any accumulation point of the sequence $\{\overline{x}^k\}$ will solve (CDC).*

Proof Suppose that the algorithm is infinite and let $\overline{x} = \lim_{q\to+\infty} x^{k_q}$. It can easily be checked that all conditions of Theorem 6.5 are fulfilled ($w \in \mathrm{int}G$ by (7.58); $y^k \in [w, x^k] \setminus \mathrm{int}G$, $l_k(y^k) = 0$). By this theorem, $\overline{x} = \lim y^{k_q}$, therefore $h(\overline{x}) = \lim h(y^{k_q}) \leq -\varepsilon$. On the other hand, $h(x^k) \geq 0$, hence $h(\overline{x}) \geq 0$. This is a contradiction unless $\varepsilon = 0$, Since $g(y^k) \leq 0\ \forall k$ it follows that $\overline{x} \in D$. Thus $\overline{x}$ is a feasible solution. Now let $\tilde{\gamma} = \lim \gamma_k$. Since for every k, $h(x^k) = \max\{h(x) |\ x \in P_k\}$ and $D(\gamma_k) \subset P_k$ it follows that $0 = h(\overline{x}) = \max\{h(x) |\ x \in D(\tilde{\gamma})\}$. Hence $D(\tilde{\gamma}) \subset C$. By the optimality criterion (7.60) this implies, under the regularity assumption, that $\overline{x}$ is a global optimal solution and $\tilde{\gamma} = \overline{\gamma}$, the optimal value. Furthermore, since $\gamma_k = f(\overline{x}^k)$, any accumulation point of the sequence $\{\overline{x}^k\}$ is also a global optimal solution. □

Remark 7.12 No regularity condition is required when $\varepsilon > 0$. It can easily be checked that the convergence proof is still valid if in Step 3 of certain iterations the cut is taken with $p^k \in \partial h(y^k)$ instead of $p^k \in \partial f(y^k)$. This flexibility should be used to monitor the speed of convergence.

Remark 7.13 The above algorithm presupposes the availability of a point $w \in \mathrm{int}D \cap \mathrm{int}C$, i.e., satisfying $\max\{g(w), h(w)\} < 0$. Such a point, if not readily available, can be computed by methods of convex programming. It is also possible to modify the algorithm so as to dispense with the computation of w (Tuy 1995b). An earlier version of the above algorithm with $\varepsilon = 0$ was developed in (Tuy 1987a). Several modified versions of this algorithm have also been proposed (e.g., Thoai 1988) which, however, were later shown to be nonconvergent. In fact, the convergence of cutting methods for (*CDC*) is a more delicate question than it seems.

7.4.2 Combined Algorithm

By translating if necessary, assume that (7.58) holds with $w = 0$, and the feasible set $D \setminus \text{int}C$ lies entirely in the orthant $M_0 = \mathbb{R}^n_+$. It is easily verified that the above OA algorithm still works if, in the case when $\max\{h(x)|\ x \in P_k\} > 0$, $x(P_k)$ is taken to be any point $x^k \in P_k$ satisfying $h(x^k) > -\varepsilon$ rather than a maximizer of $h(x)$ over P_k as in (7.64). As shown in Sect. 6.3 such a point x^k can be computed by applying the Procedure DC to the inclusion

$$(SP_k) x \in P_k \setminus C_\varepsilon.$$

For $\varepsilon > 0$, after finitely many steps this procedure either returns a point x^k such that $h(x^k) > -\varepsilon$, or establishes that $\max\{h(x)|\ x \in P_k\} < 0$ (the latter implies that the current incumbent is an ε-approximate optimal solution if $\gamma_k < +\infty$, or that the problem is infeasible if $\gamma_k = +\infty$). Furthermore, since P_{k+1} differs from P_k by only one additional linear constraint, the last conical partition in the DC procedure for (SP_k) can be used to start the procedure for (SP_{k+1}). We can thus formulate the following combined algorithm (recall that $w = 0$):

Combined OA/CS Conical Algorithm for (CDC)

Initialization Let $\gamma_1 = c\overline{x}^1$, where $\overline{x}^1$ is the best feasible solution available (if no feasible solution is known, set $\overline{x}^1 = \emptyset$, $\gamma_1 = +\infty$). Take a polytope $P_0 \supset D$ and let $P_1 = \{x \in P_0|\ cx \le \gamma_1\}$. Set $\mathscr{N}_1 = \mathscr{P}_1 = \{\mathbb{R}^n_+\}$ (any subcone of $\mathbb{R}^n_+$ will be assumed to have its base in the simplex spanned by the units vectors $\{e^i, i = 1, \dots, n\}$). Set $k = 1$.

Step 1. (*Evaluation*) For each cone $M \in \mathscr{P}_k$ of base $[u^1, \dots, u^n]$ compute the point $\theta_i u^i$ where the ith edge of M meets ∂C. Let U be the matrix of columns $u^1, \dots, u^n$. Solve the linear program

$$\max\left\{\sum_{i=1}^n \frac{t_i}{\theta_i} |\quad Ut \in P_k,\ t \ge 0\right\}$$

to obtain the optimal value $\mu(M)$ and a basic optimal solution $\bar{t}(M)$. Let $\omega(M) = U\bar{t}(M)$.

Step 2. (*Screening*) Delete every cone $M \in \mathscr{N}_k$ such that $\mu(M) < 1$, and let $\mathscr{R}_k$ be the set of remaining cones.

Step 3. (*Termination Criterion 1*) If $\mathscr{R}_k = \emptyset$, then terminate: $\overline{x}^k$ is an ε-approximate optimal solution of (*CDC*) (if $\gamma_k < +\infty$), or the problem is infeasible (if $\gamma_k = \infty$).

Step 4. (*Termination Criterion 2*) Select $M_k \in \text{argmax}\{\mu(M)|\ M \in \mathscr{R}_k\}$. If $\omega^k = \omega(M_k) \in D(\gamma^k) := \{x \in D|\ f(x) \le \gamma^k\}$, then terminate: ω^k is a global optimal solution.

Step 5. (*Branching*) Subdivide M_k along the ray through $\omega^k = \omega(M_k)$. Let $\mathscr{P}_{k+1}$ be the partition of M_k.

Step 6. (*Outer Approximation*) If $h(\omega(M)) > -\varepsilon$ for some $M \in \mathscr{R}_k$ then let $x^k = \omega(M), y^k \in [0, x^k]$ such that $\max\{g(y^k), h(y^k) + \varepsilon\} = 0$.

(a) If $g(y^k) = 0$ let $l_k(x) \leq 0$ be the cut (7.65), $\overline{x}^{k+1} = \overline{x}^k, \gamma_{k+1} = \gamma_k$.
(b) Otherwise, let $l_k(x) \leq 0$ be the cut (7.66), $\overline{x}^{k+1} = y^k, \gamma_{k+1} = cy^k$.

Step 7. (*Updating*) Let $\mathscr{N}_{k+1} = (\mathscr{R}_k \setminus \{M_k\}) \cup \mathscr{P}_{k+1},\ P_{k+1} = \{x \in P_k |\ l_k(x) \leq 0\}$. Set $k \leftarrow k + 1$ and go back to Step 1.

Note that in Step 2, a cone M is deleted if $\mu(M) < 1$ (strict inequality), so $\mathscr{R}_k = \emptyset$ implies that $\max\{h(x)|\ x \in P_k\} < 0$. It is easy to prove that, analogously to the Combined Algorithm for (*GCP*) (Theorem 7.5), the above Combined Algorithm for (*CDC*) can be infinite only if $\varepsilon = 0$ and in this case, if the problem is regular, every accumulation point of the sequence $\{\overline{x}^k\}$ is a global optimal solution of (*CDC*).

7.5 Robust Approach

Consider the general dc optimization problem

$$(P) \qquad \min\{f(x)|\ x \in [a,b],\ g_i(x) \leq 0,\ i = 1, \ldots, m\},$$

where $f, g_1, \ldots, g_m$ are dc functions on $\mathbb{R}^n$, $a, b \in \mathbb{R}^n$ and $[a,b] = \{x \in \mathbb{R}^n |\ a \leq x \leq b\}$. If $f(x) = f_1(x) - f_2(x)$ with f_1, f_2 being convex functions, we can rewrite (*P*) as $\min\{f_1(x) - t|\ x \in [a,b],\ g_i(x) \leq 0, i = 1, \ldots, m,\ t - f_2(x) \leq 0\}$, so by changing the notation it can always be assumed that the objective function $f(x)$ in (*P*) is convex.

In the most general case, a common approach for solving problem (*P*) is to use a suitable OA or BB procedure for generating a sequence of infeasible solutions x^k such that, as $k \to +\infty$, x^k tends to a feasible solution x^*, while $f(x^k)$ tends to the optimal value of the problem. Under these conditions x^* is a global optimal solution, and *theoretically* the problem can be considered solved.

Practically, however, the algorithm has to be stopped at some x^k which is expected to be close enough to the optimal solution. On the one hand, x^k must be approximately feasible, on the other hand, the value $f(x^k)$ must approach the optimal value $v(P)$. In precise terms the former condition means that, given a tolerance $\varepsilon > 0$, one must have $x^k \in [a,b], g_i(x^k) \leq \varepsilon, i = 1, \ldots, m$; the latter condition means that, given a tolerance $\eta > 0$, one must have $f(x^k) \leq v(P) + \eta$. A point x^k satisfying both these conditions is referred to as an (ε, η)-*approximate optimal solution*, or for $\eta = \varepsilon$ an ε-*approximate optimal solution*.

Everything in this approximation scheme seems to be in order, so computing an (ε, η)-approximate optimal solution has become a common practice for solving nonconvex global optimization problems. Close scrutiny reveals, however, pitfalls behind this (ε, η)-approximation concept.

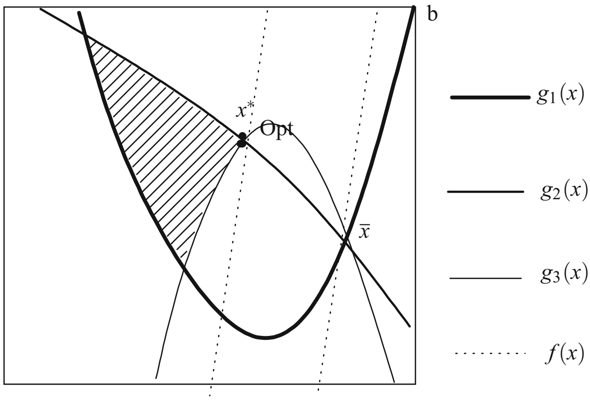


Fig. 7.11 Inadequate ε-approximate optimal solution

Consider, for example, the quadratic program depicted in Fig. 7.11, where the objective function $f(x)$ is linear and the nonconvex constraint set is defined by the inequalities $a \le x \le b,\ g_i(x) \le 0\ (i = 1, 2, 3)$. The optimal solution is x^*, while the point $\overline{x}$ is infeasible, but almost feasible:

$$g_1(\overline{x}) = 0,\ g_2(\overline{x}) = 0,\ g_3(\overline{x}) = \alpha$$

for some small $\alpha > 0$. If $\varepsilon > \alpha$ then $\overline{x}$ is feasible with tolerance ε and will be accepted as an ε-approximate optimal solution, though it is quite far off the optimal solution x^*. But if $\varepsilon < \alpha$ then $\overline{x}$ is no longer feasible with tolerance ε and the ε-approximate solution will come close to x^*.

Thus, even for regular problems (problems with no isolate feasible solution), the ε-approximation approach may give an incorrect optimal solution if ε is not sufficiently small. The trouble is that in practice we often do not know what exactly means "sufficiently small," i.e., we do not know how small the tolerance should be to guarantee a correct approximate optimal solution.

Things are much worse when the problem happens to have a global optimal solution which is an isolated feasible solution. In that case a slight perturbation of the data may cause a drastic change of the global optimal solution; consequently, a small change of the tolerances ε, η may cause a drastic change of the (ε, η)-approximate optimal solution. Due to this instability an isolated global optimal solution is very difficult to compute and very difficult to implement when computable. To avoid such annoying situation a common practice is to restrict consideration to problems with a *robust* feasible set, i.e., a feasible set D satisfying

$$D = \mathrm{cl}(\mathrm{int} D), \tag{7.67}$$

where cl and int denote the closure and the interior, respectively. Unfortunately, this condition (7.67) which implies the absence of isolated feasible solutions is generally very hard to check for a given nonconvex problem. Practically we often have to deal with feasible sets which are not known à priori to contain isolated points or not.

Therefore, from a practical point of view only nonisolated feasible solutions of (P) should be considered and instead of trying to find a global optimal solution of (P) it would be more reasonable to look for the best nonisolated feasible solution. Embodying this idea, the robust approach is devised for computing the best nonisolated feasible solution without preliminary knowledge about condition (7.67) holding or not. First let us introduce some basic concepts.

Let D^* denote the set of nonisolated points (i.e., accumulation points) of $D := \{x \in [a,b] | \ g(x) \leq 0\}$. A solution $x^* \in D^*$ is called an *essential optimal solution* of (P) if

$$f(x^*) \leq f(x) \quad \forall x \in D^*.$$

In other words, an essential optimal solution is an optimal solution of the problem

$$\min\{f(x) | \ x \in D^*\}.$$

For given $\varepsilon > 0, \eta > 0$, a point $x \in [a,b]$ satisfying $g(x) \leq -\varepsilon$ is said to be *ε-essential feasible*, and a nonisolated feasible solution x^* is called *essential* (ε, η)-*optimal* if

$$f(x^*) \leq f(x) + \eta \quad \forall x \in D_\varepsilon,$$

where $D_\varepsilon := \{x \in [a,b] | \ g(x) \leq -\varepsilon\}$ is the set of all ε-essential feasible solutions. Clearly for $\varepsilon = \eta = 0$ an essential (ε, η)-optimal solution is a nonisolated feasible point which is optimal.

To find the best nonisolated feasible solution the robust approach proceeds by successive improvement of the current incumbent. Specifically, let $w \in \mathbb{R}^n$ be any point such that $f(w) - \varepsilon > f(x) \ \forall x \in D$; in each stage of the process, one has to solve the following subproblem of incumbent transcending:

(SP) *Given an $\overline{x} \in \mathbb{R}^n$, find a nonisolated feasible solution $\hat{x}$ of (P) such that $f(\hat{x}) \leq f(\overline{x}) - \varepsilon$, or else establish that none such $\hat{x}$ exists.*

Clearly, if $\overline{x} = w$ then an answer to (SP) would give a nonisolated feasible solution or else identify essential infeasibility of the problem (P). If $\overline{x}$ is the best nonisolated feasible solution currently available then an answer to (SP) would give a new nonisolated feasible solution $\hat{x}$ with $f(\hat{x}) \leq f(\overline{x}) - \varepsilon$, or else identify $\overline{x}$ as an essential ε-optimal solution.

Since the essential optimal value is bounded above in view of the compactness of D, by successively solving a finite sequence of subproblems (SP) we will finally come up with an essential ε-optimal solution, or an evidence that the problem is essentially infeasible. The key reduces to investigating the incumbent transcending subproblem (SP).

Setting $g(x) = \max_{i=1,\ldots,m} g_i(x)$, consider the following pair of problems:

$$\min\{f(x) | \ g(x) \leq \varepsilon, \ x \in [a,b]\}, \qquad (P_\varepsilon)$$

$$\min\{g(x) | \ f(x) \leq \gamma, \ x \in [a,b]\}, \qquad (Q_\gamma)$$

where the objective and constraint functions are interchanged.

Due to the fact that $f(x)$ is convex a key property of problem (Q_γ) is that its feasible set is a convex set, so (Q_γ) has no isolated feasible solution and computing a feasible solution of it can be done at cheap cost. Therefore, as shown in Sect. 7.5, an adaptive BB Algorithm can be devised that, for any given tolerance $\eta > 0$, can compute an η-optimal solution of (Q_γ) in finitely many steps (Proposition 5.2). Furthermore, such an η-optimal solution is stable under small changes of η.

For our purpose of solving problem (P) this property of (Q_γ) together with the following relationship between (P_ε) and (Q_γ) turned out to be crucial. Let $\min(P_\varepsilon)$, $\min(Q_\gamma)$ denote the optimal values of problem (P_ε), (Q_γ), respectively.

Proposition 7.13 *For every $\varepsilon > 0$, if $\min(Q_\gamma) > \varepsilon$ then $\min(P_\varepsilon) > \gamma$.*

Proof If $\min(Q_\gamma) > \varepsilon$, then any $x \in [a,b]$ such that $g(x) \leq \varepsilon$ must satisfy $f(x) > \gamma$, hence, by compactness of the feasible set of (P_ε), $\min\{f(x)|\ g(x) \leq \varepsilon,\ x \in [a,b]\} > \gamma$. $\square$

This almost trivial proposition simply expresses the interchangeability between objective and constraint in many practical situations (Tuy 1987a, 2003). Exploiting this interchangeability, the basic idea of robust approach is to replace the original problem (P), possibly very difficult, by a sequence of easier, stable, problems (Q_γ), where the parameter γ can be iteratively adjusted until a stable (robust) solution to (P) is obtained.

7.5.1 Successive Incumbent Transcending

As shown above, a key step towards finding an essential (ε,η)-optimal solution of problem (P) is to deal with the following question which is the subproblem (SP) when $\gamma = f(\overline{x}) - \varepsilon$:

$(SP\gamma)$ *Given a real number γ, and $\varepsilon > 0$, check whether problem* (P) *has a nonisolated feasible solution x satisfying $f(x) \leq \gamma$, or else establish that no ε-essential feasible solution x exists such that $f(x) \leq \gamma$.*

Based on the convexity of $f(x)$, consider an adaptive BB Algorithm for solving problem (Q_γ). As described in the proof of Proposition 6.2, such an algorithm generates at each iteration k two points $x^k \in M_k,\ y^k \in M_k$ satisfying

$$f(x^k) \leq \gamma, \quad g(y^k) - \beta(M_k) \to 0 \quad \text{as } k \to +\infty,$$

where $\beta(M_k) \leq v(Q_\gamma)$ (note that the objective function is $g(x)$, the constraints are $f(x) \leq \gamma,\ x \in [a,b]$).

Proposition 7.14 *Let $\varepsilon > 0$ be given. Either $g(x^k) < 0$ for some k or $\beta(M_k) > -\varepsilon$ for some k. In the former case, x^k is a nonisolated feasible solution of (P) satisfying $f(x^k) \leq \gamma$. In the latter case, no ε-essential feasible solution x of (P) exists such that $f(x) \leq \gamma$ (so, if $\gamma = f(\overline{x}) - \eta$ for a given $\eta > 0$ and a nonisolated feasible solution $\overline{x}$ then $\overline{x}$ is an essential (ε,η)-optimal solution).*

Proof If $g(x^k) < 0$, then x^k is an essential feasible solution of (P) satisfying $f(x^k) \leq \gamma$, because by continuity $g(x) < 0$ for all x in a neighborhood of x^k. If $\beta(M_k) > -\varepsilon$, then $v(Q_\gamma) > -\varepsilon$, hence, by Proposition 7.13, $v(P_\varepsilon) > \gamma$.

It remains to show that either $g(x^k) < 0$ for some k, or $\beta(M_k) > -\varepsilon$ for some k. Suppose the contrary that

$$g(x^k) \geq 0, \ \beta(M_k) \leq -\varepsilon \quad \forall k. \tag{7.68}$$

As we saw in Sect. 6.1 (Theorem 6.4), the sequence x^k, y^k generated by an adaptive BB Algorithm satisfies, for some infinite subsequence $\{k_\nu\} \subset \{1, 2, \ldots\}$,

$$g(y^{k_\nu}) - \beta(M_{k_\nu}) \to 0 \ (\nu \to +\infty), \tag{7.69}$$

$$\lim_{\nu\to+\infty} x^{k_\nu} = \lim_{\nu\to+\infty} y^{k_\nu} = x^* \quad \text{with } f(x^*) \leq \gamma, \ x^* \in [a, b]. \tag{7.70}$$

Since by (7.68) $g(x^k) \geq 0 \ \forall k$, it follows that $g(x^*) \geq 0$. On the other hand, by (7.69) and (7.68) we have $g(x^*) = \lim_{\nu\to+\infty} \beta(M_{k_\nu}) \leq -\varepsilon$. This contradiction shows that the event (7.68) is impossible, completing the proof. □

Thus, with the stopping criterium " $g(x^k) < 0$ or $\beta(M_k) > -\varepsilon$ " an adaptive BB procedure for solving (Q_γ) will help to solve the subproblem $(SP\gamma)$. Note that $\beta(M_k) = \min\{\beta(M) | \ M \in \mathscr{R}_k\}$ ($\mathscr{R}_k$ being the collection of partition sets remaining for exploration at iteration k). So if the deletion (pruning) criterion in each iteration of this procedure is " $\beta(M) > -\varepsilon$ " then the stopping criterion " $\beta(M_k) > -\varepsilon$ " is nothing but the usual criterion " $\mathscr{R}_k = \emptyset$ ".

For brevity an adaptive BB Algorithm for solving (Q_γ) with the deletion criterion "$\beta(M) > -\varepsilon$ " and stopping criterion " $g(x^k) < 0$ " will be referred to as *Procedure* $(SP\gamma)$. Using this procedure, problem (P) can be solved according to the following scheme:

Successive Incumbent Transcending (SIT) Scheme
Start with $\gamma = \gamma_0$, where $\gamma_0 = f(a) - \eta$.

Call Procedure $(SP\gamma)$. If a nonisolated feasible solution $\bar{x}$ of (P) is obtained with $f(\bar{x}) \leq \gamma$, reset $\gamma \leftarrow f(\bar{x}) - \eta$ and repeat. Otherwise, stop: if $\gamma = f(\bar{x}) - \eta$ for a nonisolated feasible solution $\bar{x}$ then $\bar{x}$ is an essential (ε, η)-optimal solution; if $\gamma = \gamma_0$ then problem (P) has no ε-essential feasible solution.

Since $f(D)$ is compact and $\eta > 0$ this scheme is necessarily finite.

7.5.2 The SIT Algorithm for DC Optimization

For the sake of convenience let us rewrite problem (P) in the form:

$$(P) \qquad \min\{f(x) | \ g_1(x) - g_2(x) \leq 0, \ x \in [a, b]\},$$

where f, g_1, g_2 are convex functions.

As shown above (Proposition 7.14), Procedure ($SP\gamma$) for problem (P) is an adaptive BB algorithm for solving problem (Q_γ), with deletion criterion "$\beta(M) > -\varepsilon$" and stopping criterion "$g(x^k) < 0$".

Incorporating Procedure ($SP\gamma$) into the SIT Scheme we can state the following SIT Algorithm for (P):

SIT Algorithm for (P)

Select tolerances $\varepsilon > 0, \eta > 0$. Let $\gamma_0 = f(a) - \eta$.

Step 0. Set $\mathscr{P}_1 = \{M_1\}, M_1 = [a, b], \mathscr{R}_1 = \emptyset, \gamma = \gamma_0$. Set $k = 1$.

Step 1. For each box (hyperrectangle) $M \in \mathscr{P}_k$:

- Reduce M, i.e., find a box $[p, q] \subset M$ as small as possible satisfying $\min\{g(x)|\, f(x) \le \gamma,\ x \in [p, q]\} = \min\{g(x)|\, f(x) \le \gamma,\ x \in M\}$ and set $M \leftarrow [p, q]$.
- Compute a lower bound $\beta(M)$ for $g(x)$ over the feasible solutions in $[p, q]$.
  $\beta(M)$ must be such that: $\beta(M) < +\infty \Rightarrow \{x \in M|\, f(x) \le \gamma\} \ne \emptyset$, see (5.4).

Delete every M such that $\beta(M) > -\varepsilon$.

Step 2. Let $\mathscr{P}'_k$ be the collection of boxes that results from $\mathscr{P}_k$ after completion of Step 1. Let $\mathscr{R}'_k = \mathscr{R}_k \cup \mathscr{P}'_k$.

Step 3. If $\mathscr{R}'_k = \emptyset$, then *terminate*: if $\gamma = \gamma_0$ the problem (P) is ε-essential infeasible; if $\gamma = f(\overline{x}) - \eta$ for some nonisolated feasible solution $\overline{x}$, then $\overline{x}$ is an essential (ε, η)-optimal solution of (P).

[Step 4.] If $\mathscr{R}'_k \ne \emptyset$, let $M_k \in \operatorname{argmin}\{\beta(M)|\ M \in \mathscr{R}'_k\},\ \beta_k = \beta(M_k)$. Determine $x^k \in M_k$ and $y^k \in M_k$ such that

$$f(x^k) \le \gamma, \quad g(y^k) - \beta(M_k) = o(\|x^k - y^k\|). \tag{7.71}$$

If $g(x^k) < -\varepsilon$, go to Step 5. If $g(x^k) \ge -\varepsilon$, go to Step 6.

Step 5. x^k is a nonisolated feasible solution satisfying $f(x^k) \le \gamma$.
If $\gamma = \gamma_0$ define $\overline{x} = x^k,\ \gamma = f(\overline{x}) - \eta$ and go to Step 6. If $\gamma = f(\overline{x}) - \eta$ and $f(x^k) < f(\overline{x})$ reset $\overline{x} \leftarrow x^k$ and go to Step 6.

Step 6. Divide M_k into two subboxes by the adaptive bisection, i.e., bisect M_k via (v^k, j_k), where $j_k \in \operatorname{argmax}\{|y^k_j - x^k_j| : j = 1, \dots, n\}, v^k = \frac{1}{2}(x^k + y^k)$.
Let $\mathscr{P}_{k+1}$ be the collection of these two subboxes of M_k, $\mathscr{R}_{k+1} = \mathscr{R}'_k \setminus \{M_k\}$.
Increment k, and return to Step 1.

Theorem 7.10 *The SIT Algorithm terminates after finitely many steps, yielding an essential (ε, η)-optimal solution or an evidence that the problem has no ε-essential feasible solution.*

Proof Follows from the above discussion. □

To help practical implementation, it may be useful to discuss in more detail two operations important for the efficiency of the algorithm: the box reduction in Step 1 and the computation of x^k and y^k in Step 4.

Box Reduction
Let $M = [p, q]$ be a box. We wish to determine a box red$M = [p', q'] \subset M$ such that

$$\{x \in [p', q'] | f(x) \leq \gamma,\ g_1(x) - g_2(x) \leq \varepsilon\} = \{x \in M | f(x) \leq \gamma,\ g_1(x) - g_2(x) \leq \varepsilon\}. \tag{7.72}$$

Select a convex function $\overline{g}(x)$ underestimating $g(x) = g_1(x) - g_2(x)$ on M and tight at a point $y \in M$, i.e., such that $\overline{g}(y) = g(y)$. For example, $\overline{g}(x) = g_1(x) - g_2(y)$, where y is a corner of the hyperrectangle $[p, q]$ maximizing $g_2(x)$ over $[p, q]$). Then p', q' are determined by solving, for $i = 1, \ldots, n$, the convex programs:

$$p'_i = \min\{x_i |\ x \in [p, q],\ f(x) \leq \gamma,\ \overline{g}(x) \leq \varepsilon\},$$
$$q'_i = \max\{x_i |\ x \in [p, q],\ f(x) \leq \gamma,\ \overline{g}(x) \leq \varepsilon\}.$$

Bounding and Condition (7.71)
With $\overline{g}(x)$ being a convex underestimator of $g(x) = g_1(x) - g_2(x)$ over $M = [p, q]$, tight at a point $y \in$ red$M$ define $\beta(M) = \min\{\overline{g}(x) |\ f(x) \leq \gamma,\ x \in \text{red}M\}$. So for each k there is a point $y^k \in M_k$ such that $\overline{g}(y^k) = g(y^k)$. Also define $x^k \in \text{argmin}\{\overline{g}(x) |\ f(x) \leq \gamma,\ x \in \text{red}M_k\}$, so that $\overline{g}(x^k) = \beta(M_k)$. If $x^k = y^k$ then $\beta(M_k) = \min\{g(x) |\ x \in M_k\}$ so, as can easily be checked, condition (7.71) is satisfied.

Alternative Method: Define

$$x^M \in \text{argmin}\{g_1(x) | f(x) \leq \gamma,\ x \in \text{red}M\},$$
$$y^M \in \text{argmax}\{g_2(x) | f(x) \leq \gamma,\ x \in \text{red}M\},$$
$$\beta(M) = g_1(x^M) - g_2(y^M).$$

The points $x^k = x^{M_k}, y^k = y^{M_k}$ are such that $x^k = y^k$ implies that $\min\{g_1(x) - g_2(x) |\ x \in M_k\} = 0$, hence it is easy to verify that condition (7.71) is satisfied.

Special Case: When f, g are polynomials or signomials, one can define at each iteration k a convex program (e.g., a *SDP*)

$$(L_k) \qquad \max\{\ell_k(x, u) | f(x) \leq \gamma,\ (x, u) \in C_k\},$$

where $u \in \mathbb{R}^{m_k}, C_k$ is a compact set in $M_k \times \mathbb{R}^{m_k}$, and $\ell_k(x, u)$ is a linear function such that the problem $\max\{g(x) | \tilde{f}(x) \leq \gamma,\ x \in M_k\}$ is equivalent to

$$(NL_k) \qquad \max\{\ell_k(x,u) | f(x) \le \gamma,\ (x,u) \in C_k,\ r_i(x,u) \ge 0\ (i \in I_k)\},$$

where $r_i(x,u) \ge 0,\ i \in I_k$, are nonconvex constraints (Lasserre 2001; Sherali and Adams 1999). Furthermore, (L_k) can be selected so that there exists $y^k \in M_k$ satisfying

$$(y^k,u) \in C_k \Rightarrow r_i(y^k,u) \ge 0 \ \ (i \in I_k) \tag{7.73}$$

(see, e.g., Tuy (1998, Proposition 8.12)). In that case, setting

$$\beta(M_k) = \max\{\ell_k(x,u) | f(x) \le \gamma,\ (x,u) \in C_k\}$$

and taking an optimal solution (x^k,u^k) of (L_k) one has a couple $x^k,\ y^k$ satisfying (7.73) and

$$x^k \in M_k,\ f(x^k) \le \gamma,\ (x^k,u^k) \in C_k,\ \ell_k(x^k,u^k) = \beta(M_k). \tag{7.74}$$

If $x^k = y^k$ then from (7.73) and $(x^k,u^k) \in C_k$ it follows that $r_i(x^k,u) \ge 0$ $(i \in I_k)$, so (x^k,u^k) solves (NL_k) and hence x^k solves the subproblem $\min\{g(x) | \tilde{f}(x) \le \gamma,\ x \in M_k\}$, i.e., $g(x^k) = \ell_k(x^k,u^k) = \beta(M_k)$. It can then easily be shown that $g(y^k) - \beta(M_k) \to 0$ as $x^k - y^k \to 0$, so that x^k, y^k satisfy condition (7.71).

7.5.3 Equality Constraints

So far we assumed that all nonconvex constraints are of inequality type: $g_i(x) \le 0,\ i = 1,\dots,m$. If there are equality constraints such as

$$h_j(x) = 0,\ j = 1,\dots,s$$

with all functions $h_1,\dots,h_s$ nonlinear, the SIT algorithm for whatever $\varepsilon > 0$ will conclude that there is no ε-essential feasible solution. This does not mean, however, that the problem has no nonisolated feasible solution. In that case, since an exact solution to a nonlinear system of equations cannot be expected to be computable in finitely many steps, one should be content with replacing every equality constraint $h_j(x) = 0$ by an approximate one

$$|h_j(x)| \le \delta,$$

where $\delta > 0$ is the tolerance. The above approach can then be applied to the resulting set of inequality constraints.

7.5.4 Illustrative Examples

To illustrate the practicality of the robust approach, we give two numerical examples.

Example 7.3 Minimize x_1 subject to

$$\begin{aligned}(x_1 - 5)^2 + 2(x_2 - 5)^2 + (x_3 - 5)^2 &\geq 18 \\ (x_1 + 7 - 2x_2)^2 + 4(2x_1 + x_2 - 11)^2 + 5(x_3 - 5)^2 &\geq 100\end{aligned}$$

This problem has one isolated feasible point $(1, 3, 5)$ which is also the optimal solution. With $\varepsilon = \eta = 0.01$ the essential (ε, η)-optimal solution (3.747692, 7.171420, 2.362317) with objective function value 3.747692. is found by the SIT Algorithm after 15736 iterations.

Example 7.4 (n=4) Minimize $(3 + x_1x_3)(x_1x_2x_3x_4 + 2x_1x_3 + 2)^{2/3}$ subject to

$$\begin{aligned}
&-3(2x_1x_2 + 3x_1x_2x_4)(2x_1x_3 + 4x_1x_4 - x_2) \\
&\quad -(x_1x_3 + 3x_1x_2x_4)(4x_3x_4 + 4x_1x_3x_4 + x_1x_3 - 4x_1x_2x_4)^{1/3} \\
&\qquad +3(x_4 + 3x_1x_3x_4)(3x_1x_2x_3 + 3x_1x_4 + 2x_3x_4 - 3x_1x_2x_4)^{1/4} \leq -309.219315 \\
&-2(3x_3 + 3x_1x_2x_3)(x_1x_2x_3 + 4x_2x_4 - x_3x_4)^2 \\
&\quad +(3x_1x_2x_3)(3x_3 + 2x_1x_2x_3 + 3x_4)^4 - (x_2x_3x_4 + x_1x_3x_4)(4x_1 - 1)^{3/4} \\
&\qquad -3(3x_3x_4 + 2x_1x_3x_4)(x_1x_2x_3x_4 + x_3x_4 - 4x_1x_2x_3 - 2x_1)^4 \leq -78243.910551 \\
&-3(4x_1x_3x_4)(2x_4 + 2x_1x_2 - x_2 - x_3)^2 \\
&\quad +2(x_1x_2x_4 + 3x_1x_3x_4)(x_1x_2 + 2x_2x_3 + 4x_2 - x_2x_3x_4 - x_1x_3)^4 \leq 9618 \\
&0 \leq x_i \leq 5 \;\; i = 1, 2, 3, 4.
\end{aligned}$$

This problem would be difficult to solve by Sherali-Adams' RLT method or Lassere's method which both would require introducing a very large number of additional variables.

For $\varepsilon = \eta = 0.01$ the essential (ε, η)-optimal solution

$$x^{\text{essopt}} = (4.994594,\ 0.020149,\ 0.045424,\ 4.928073)$$

with objective function value 5.906278 is found by the SIT Algorithm at iteration 667, and confirmed as such at iteration 866.

7.5.5 Concluding Remark

Global optimization problems with nonconvex constraint sets are difficult in two major aspects: (1) a feasible solution may be very hard to determine; (2) the optimal solution may be an isolated feasible solution, in which case a correct solution cannot be computed by a finite procedure and implemented successfully.

To cope with these difficulties the robust approach consists basically in finding a nonisolated feasible solution and improving it step by step. The resulting algorithm works its way to the best nonisolated optimal solution through a number of cycles of incumbent transcending. A major advantage of it, apart from stability, is that when prematurely stopped it may still provide a good nonisolated feasible solution, in contrast to other current methods which are almost useless in that case.

7.6 DC Optimization in Design Calculations

An important class of nonconvex optimization problems encountered in the applications has the following general formulation:

(P) Given a compact set $S \subset \mathbb{R}^n$ with int$S \neq \emptyset$ and a gauge $p : \mathbb{R}^n \to \mathbb{R}_+$ find the largest value r for which there is an $x \in S$ satisfying

$$B(x;r) := \{y \in \mathbb{R}^n | \, p(y-x) \leq r\} \subset S. \tag{7.75}$$

Here by *gauge* is meant a nonnegative positively homogeneous and convex function. As can be immediately seen, a gauge $p : \mathbb{R}^n \to \mathbb{R}_+$ is actually the gauge of the convex set $\{x | \, p(x) \leq 1\}$ in the sense earlier defined in Sect. 1.3.

Problems of the above kind arise in many design calculations. A typical example is the design centering problem mentioned in Example 5.3, Sect. 5.2. Specifically, suppose in a fabrication process the quality of a manufactured item is characterized by an n-dimensional parameter. An item is acceptable if this parameter lies in some region of acceptability S. If $x \in \mathbb{R}^n$ is the designed value of the parameter, then, due to unavoidable fluctuations, the actual value y of it will usually deviate from x, so that $p(y-x) > 0$. For each fixed value of x the value of y must lie in $B(x;r)$ in order to be acceptable, so the production yield can be measured by the maximal value of $r = r(x)$ such that $B(x;r)$ is still contained in S. Under these conditions, to maximize the yield the designer has to solve the problem (7.75).

This is an inherently difficult problem involving two levels of optimization: for every fixed value $x \in S$ compute

$$r(x) := \max\{r | \, B(x;r) \subset S\}, \tag{7.76}$$

then determine the value $\bar{x}$ that achieves

$$\bar{r} := \max\{r(x) | \, x \in S\}. \tag{7.77}$$

Except in special cases both these two subproblems are nonconvex. Even in the relatively simple case when S is convex nonpolyhedral and $p(x) = \|x\|$ (the Euclidean norm), the subproblem (7.76), for fixed x, is known to be NP-hard (Shiau 1984). In view of these difficulties, for some time most of the papers published on this subject either concentrated on some particular aspects of the problem or were confined to the search for a local optimum, under rather restrictive conditions.

If no further structure is given on the data, the general problem of design centering is essentially intractable by deterministic methods. Fortunately, in many practically important cases (see, e.g., Vidigal and Director (1982)) it can be assumed that

1. $p(x) = x^T Ax$,
   where A is a symmetric positive definite matrix;
2. $S = C \cap D_1 \cap \cdots \cap D_m$,
   where C is a closed convex set in $\mathbb{R}^n$ with $\mathrm{int}C \neq \emptyset$ and $D_i = \mathbb{R}^n \setminus \mathrm{int}C_i$ $(i = 1, \ldots, m)$ and C_i is a closed convex subset of $\mathbb{R}^n$.

Below we present a solution method proposed by Thach (1988) for problem (P) under these assumptions.

7.6.1 DC Reformulation

For every $x \in \mathbb{R}^n$ define

$$g(x) = \sup_{y \notin S}(2x^T Ay - y^T Ay).$$

Proposition 7.15 (Thach 1988) *The function $g(x)$ is convex, finite on $\mathbb{R}^n$ and the general design centering problem is equivalent to the unconstrained dc maximization problem*

$$\max\{x^T Ax - g(x) |\ x \in \mathbb{R}^n\}. \tag{7.78}$$

Proof For every $x \in \mathbb{R}^n$ we can write

$$\begin{aligned} x^T Ax - g(x) &= \inf_{y \notin \mathrm{int}S}(x^T Ax - 2x^T Ay + y^T Ay + y^T Ay) \\ &= \inf_{y \notin \mathrm{int}S}(y - x)^T A(y - x) \qquad (7.79) \\ &= \inf_{y \notin \mathrm{int}S} p^2(y - x). \end{aligned}$$

But for $x \in S$ it is easily seen from (7.76) that $r(x) = \inf_{y \notin \mathrm{int}S} p(y - x)\ \forall x \in S$. On the other hand, for $x \notin S$, by making $y = x$ in (7.79) we obtain $x^T Ax - g(x) = 0$ since $p^2(y - x) \geq 0$. Hence,

$$x^T Ax - g(x) = \begin{cases} r^2(x) & x \in S \\ 0 & x \notin S \end{cases} \tag{7.80}$$

This shows that problem (7.77) is equivalent to (7.78). The finiteness of $g(x)$ follows from (7.80) and that of $r(x)$. Finally, since for every fixed y the function $x \mapsto 2x^T Ay - y^T Ay$ is affine, $g(x)$ is convex as the pointwise supremum of a family of affine functions. □

Remark 7.14 (i) In view of (7.80) problem (7.78) is the same as the constrained optimization problem

$$\max\{x^T Ax - g(x) |\ x \in S\}.$$

(ii) For $x \in S$ we can also write $r(x) = \min_{y \in \partial S} p(y-x)$, where $\partial S := S \setminus \mathrm{int} S$ is the boundary of S. Therefore, from (7.79),

$$g(x) = \max_{y \in \partial S}\{2x^T Ay - y^T Ay) \qquad \forall x \in S.$$

7.6.2 Solution Method

A peculiar feature of the dc optimization problem (7.78) is that the convex function $g(x)$ is not given explicitly but is defined as the optimal value in the optimization problem

$$\max\{2x^T Ay - y^T Ay |\ y \notin S\}. \tag{7.81}$$

However, from the relation $\mathbb{R}^n \setminus \mathrm{int} S = (\cup_{i=1}^m C_i) \cup (\mathbb{R}^n \setminus \mathrm{int} C)$ it is easily seen that

$$g(x) = \max\{g_0(x), g_1(x), \ldots, g_m(x)\}$$

with

$$g_0(x) = \max\{2x^T Ay - y^T Ay |\ y \in \mathbb{R}^n \setminus \mathrm{int} C\}; \tag{7.82}$$

$$g_i(x) = \max\{2x^T Ay - y^T Ay |\ y \in C_i\},\ i = 1, \ldots, m. \tag{7.83}$$

Thus, the subproblem (7.81) splits into $m+1$ simpler ones. Every subproblem (7.83) is a convex program (maximizing a concave function $y \mapsto 2x^T Ay - y^T Ay$ over a closed convex set C_i), so it can be solved by standard algorithms. The subproblem (7.82), though harder (maximizing a concave function over a closed reverse convex set), can be solved by reverse convex programming methods developed in Section 7.3. In that way the value of $g(x)$ at every point x can be computed.

Also the next proposition helps to find a subgradient of $g(x)$ at every x.

Proposition 7.16 *If $\overline{x} \notin S$, then $2A\overline{x} \in \partial g(\overline{x})$. If $\overline{x} \in S$, then $2A\overline{y} \in \partial g(\overline{x})$, where $\overline{y}$ is an optimal solution of (7.81) for $x = \overline{x}$.*

Proof By $\overline{y}$ denote an optimal solution of (7.81) for $x = \overline{x}$. Then $g(\overline{x}) = 2\overline{x}^T A\overline{y} - \overline{y}^T A\overline{y}$, so

$$g(x) - g(\overline{x}) \geq (2x^T A\overline{y} - \overline{y}^T A\overline{y}) - (2\overline{x}^T A\overline{y} - \overline{y}^T A\overline{y}) = (x - \overline{x})^T 2A\overline{y} \quad \forall x \in \mathbb{R}^n.$$

This implies that $2A\overline{y} \in \partial g(\overline{x})$ for every $\overline{x} \in \mathbb{R}^n$. In particular, when $\overline{x} \notin S$ we have $g(\overline{x}) = \overline{x}^T A\overline{x}$, so $y = \overline{x}$ is an optimal solution of (7.81), hence $2A\overline{x} \in \partial g(\overline{x})$. □

We are now in a position to derive a solution method for the dc optimization problem (7.78) equivalent to the general design centering problem (P).

Let $f(x) := x^T Ax - g(x)$. Starting with an n-simplex Z_0 containing the compact set S we shall construct a sequence of functions $f_k(.)$, $k = 0, 1, \ldots$, such that

(i) $f_k(.)$ overestimates $f(.)$, i.e., $f_k(x) \geq f(x)$ *all* $x \in S$;
(ii) Each problem $\max\{f_k(x)|\ x \in Z_0\}$ can be solved by available efficient algorithms;
(iii) $\max\{f_k(x)|\, x \in Z_0\} \to \max\{f(x)|\, x \in S\}$ as $k \to +\infty$.

Specifically, we can state the following:

Algorithm
Select an n-simplex $Z_0 \supset S$, a point $x^0 \in S$ and $y^0 = x^0$. Set $k = 1$.

Step 1. Let $f_k(x) = x^T Ax - \max\{\langle 2Ay^i, x - x^i\rangle + g(x^i)|\ i = 0, 1, \cdots, k-1\}$. Solve the kth approximating problem

$$(P_k) \qquad \max\{f_k(x)|\, x \in Z_0\},$$

obtaining x^k. If $x^k \in S$ go to Step 2, otherwise go to Step 3.

Step 2. Solve the kth subproblem

$$(S_k) \qquad \max\{2(x^k)^T Ay - y^T Ay|\ y \notin \mathrm{int} S\}$$

obtaining an optimal solution y^k and the optimal value $g(x^k)$.
If $f_k(x^k) = (x^k)^T Ax^k - g(x^k)$, stop: x^k solves (P). Otherwise, set $k \leftarrow k+1$ and go back to Step 1.

Step 3. Let $y^k = x^k$. Set $k \leftarrow k + 1$ and go back to Step 1.

Proposition 7.17 *If the above algorithm stops at Step 2 the current x^k solves the design centering (P). If it generates an infinite sequence x^k then any cluster point of this sequence solves (P).*

Proof First observe that $f_{k-1}(x) \geq f_k(x) \geq f(x)$ $\forall x$ and that

$$f_k(x^i) = f(x^i),\ i = 0, 1, \ldots, k-1.$$

In fact, since $2Ay^i \in \partial g(x^i)$ by Proposition 7.16, it follows that $\langle 2Ay^i, x - x^i\rangle + g(x^i) \le g(x), i = 0, 1, \dots, k-1$, hence $f_k(x) \ge f_{k+1}(x) \ge f(x)\ \forall x \in \mathbb{R}^n$. For $x = x^i$ we have $g(x^i) \ge \langle 2Ay^i, x^i - x^j\rangle + g(x^j),\ i, j = 0, 1, \dots, k-1$. Hence, $f_k(x^i) \le (x^i)^T Ax^i - g(x^i) = f(x^i)$, which implies that $f_k(x^i) = f(x^i)$.

Now suppose that at some iteration $k, f_k(x^k) = (x^k)^T Ax^k - g(x^k)$, i.e., $f_k(x^k) = f(x^k)$. Then, since x^k solves (P_k) we have, for any $x \in Z_0 :\ f(x) \le f_k(x) \le f_k(x^k) = f(x^k)$, so x^k solves (P), justifying the stopping criterium in Step 2. If the algorithm is infinite, let $\overline{x} = \lim_{\nu\to+\infty} x^{k_\nu}$. As we just saw, $f_{k+1}(x^{k+1}) = \max\{f_{k+1}(x) |\ x \in Z_0\} \le \max\{f_k(x) |\ x \in Z_0\} = f_k(x^k)$, so $\{f_k(x^k)\}$ is an increasing sequence bounded above by $\max\{f(x) |\ x \in S\}$. Therefore, there exists $\gamma = \lim_{k\to+\infty} f_j(x^k)$. But for any $i < k :$

$$(x^{k_\nu})^T Ax^{k_\nu} - \langle 2Ay^i, x^{k_\nu} - x^i\rangle - g(x^i) \ge f_{k_\nu}(x^{k_\nu}) \ge f_{k_\nu}(x) \ge f(x).$$

By letting $\nu \to +\infty$ this yields

$$\overline{x}^T A\overline{x} - \langle 2Ay^i, \overline{x} - x^i\rangle - g(x^i) \ge \gamma \ge f(\overline{x}) = \overline{x}A\overline{x} - g(\overline{x}). \tag{7.84}$$

Since $y^k = x^k$, i.e., y^k solves (S_k) we have, by Remark 7.14, $y^k \in \partial S \subset S$, so the sequence $\{y^{k_\nu}\}$ is bounded. By passing to subsequences if necessary, we may assume $y^{k_\nu} \to \overline{y}$. Obviously, $2A\overline{y} \in \partial g(\overline{x})$. Setting $i = k_\nu$ in (7.84) and letting $\nu \to +\infty$ yields

$$\overline{x}^T A\overline{x} - \langle 2A\overline{y}, \overline{x} - \overline{x}\rangle - g(\overline{x}) \ge \gamma \ge \overline{x}^T A\overline{x} - g(\overline{x}),$$

hence $\gamma = \overline{x}^T A\overline{x} - g(\overline{x}) = f(\overline{x})$. Since $f_{k_\nu}(x^{k_\nu}) = \max\{f_{k_\nu}(x) |\ x \in Z_0\} \ge \max\{f(x) |\ x \in Z_0\}$, it follows that

$$f(\overline{x}) = \max\{f(x) |\ x \in Z_0\} = \max\{f(x) |\ x \in \mathbb{R}^n\}$$

(see Remark 7.14, (ii)). □

Remark 7.15 The dc maximization problem (7.78) can be rewritten as

$$\max\{x^T Ax - t |\ x \in S,\ g(x) \le t\}. \tag{7.85}$$

It is not hard to see that the above algorithm can also be interpreted as an outer approximation method for solving this convex maximization (concave minimization) under convex constraints (see Sect. 7.2). In fact, the kth approximating problem (P_k) is equivalent to the problem

$$\max\{x^T Ax - t |\ x \in Z_0,\ \langle 2Ay^i, x - x^i\rangle + g(x^i) \le t, i = 0, \dots, k-1\}. \tag{7.86}$$

Noting that $2Ay^i \in \partial g(x^i)$ (Proposition 7.16) we have $\langle 2Ay^i, x - x^i\rangle + g(x^i) \le g(x)\ \forall x \in \mathbb{R}^n$, so the linear constraints in (7.86) simply define the polytope outer approximating the convex set $\{x |\ g(x) \le t\}$ at iteration k of an outer approximation procedure for solving (7.85).

7.7 DC Optimization in Location and Distance Geometry

7.7.1 *Generalized Weber's Problem*

The classical Weber's problem is to locate a single facility in the plane that minimizes the sum of weighted Euclidean distances to the locations of N known users. This problem can be generalized to the case of p facilities as follows.

Suppose p facilities (providing the same service) are designed to serve N users located at points $a^1, \ldots, a^n$ in the plane. We wish to determine the locations $x^1, \ldots, x^p \in \mathbb{R}^2$ of the facilities so as to minimize the total cost (travel time, transport cost for customers, etc.), or to maximize the total attraction (utility, number of customers, etc.). The cost or attraction is a given function of the distances form the users to the facilities. Traditionally, the standard mathematical formulation of this problem is (Brimberg and Love 1994):

$$
\begin{aligned}
\min\nolimits_{x,u} \quad & \sum_{i=1}^{p}\sum_{j=1}^{N} w_{ij} d(x^i, a^j) \\
\text{s.t.} \quad & \sum_{i=1}^{p} w_{ij} = r_j,\ j = 1, \ldots, N \\
& w_{ij} \geq 0,\ i = 1, \ldots, p,\ j = 1, \ldots, N \\
& x^i \in S,\ \ i = 1, \ldots, p,
\end{aligned}
$$

where w_{ij} is unknown weight representing the flow from facility i to the fixed point (warehouse) $a^j, j = 1, \ldots, N$; $d(x, a)$ is the distance function which estimates the travel distance between any two points $x, a \in \mathbb{R}^2$, and S is a rectangular domain in $\mathbb{R}^2$ where the facilities must be located.

A serious disadvantage of the above formulation is that it involves a huge number of variables (pN variables w_{ij} and $2p$ variables x_1^k, x_2^k) and a highly nonconvex objective function. Because of that the problem is very difficult to handle, even for small values of p and N. In fact, while nonlinear programming methods can be trapped at a KKT point which may even not be a local optimum, the difficulties considerably increase when the model is further extended to deal with complicated but more realistic objective functions (e.g., when the cost function is measured by a monotone rather than linear function of the distances, or when the facilities are attractive only for a number of users and repelling for others).

Using the dc approach it is possible to significantly reduce the number of variables, and express the objective function in a much simpler form. Noticing that in an optimal solution each user must be served by the closest facility and setting

$$h_j(x^1, \ldots, x^p) = \min_{i=1,\ldots,p} \|x^i - a^j\|, \tag{7.87}$$

we can rewrite the problem as

$$\min \sum_{j=1}^{N} r_j h_j(x^1, \ldots, x^p) \qquad \text{s.t.} \qquad x^i \in S \subset \mathbb{R}^2,\ i = 1, \ldots, p. \tag{7.88}$$

Since clearly

$$h_j(x) = \sum_{i=1}^{p} \|x^i - a^j\| - \max_k \sum_{i \neq k} \|x^i - a^j\|,$$

where the functions $\sum_{i=1}^{p} \|x^i - a^j\|, \max_k \sum_{i \neq k} \|x^i - a^j\|$ are convex, it follows that (7.88) is a dc optimization problem. One obvious advantage of this formulation is that it uses only $2p$ variables $x_1^i, x_2^i, i = 1, \ldots, p$ instead of $pN + 2p$ variables as in the standard formulation. This allows large-scale problems which usually are out of reach of standard nonlinear methods to be solved successfully by dc optimization methods. For instance, in Tuy et al. (1995a,b,c) and Al-Khayyal et al. (1997, 2002), good computational results have been reported for problems with $N = 100{,}000, p = 1$ and $N = 10{,}000, p = 2, \quad N = 1000, p = 3$. The traditional approach would have difficulty even in solving problems with $N = 500, p = 2$, which would require more than 1000 variables to be introduced.

Facility with Attraction and Repulsion

In more realistic models, the attraction of facility i to user j at distance t away is measured by a convex decreasing function $q_j(t)$. Furthermore, for some users the attraction effect is positive, for others it is negative (which amounts to a repulsion). Let J_+ be the set of attraction points (attractors), J_- the set of repulsion points (repellers). The problem is to find the locations of facilities so as to maximize the total effect, i.e.,

$$\max \left\{ \sum_{j \in J_+} q_j(h_j(X)) - \sum_{j \in J_-} q_j(h_j(X)) |\ X = (x^1, \ldots, x^p) \in S^p \subset (\mathbb{R}^2)^p \right\},$$

where $h_j(X)$ is given by (7.87). Using Proposition 4.6 it is easy to find an explicit dc representation of the objective function. Hence, the problem is a dc optimization just as (7.88). For detail the interested reader is referred to Chen et al. (1992, 1992b), Tuy et al. (1992), and also Maranas and Floudas (1994) and Al-Khayyal et al. (1997, 2002), where results in solving problems with up to 100,000 attractors and repellers have been reported. A local approach to this problem under more restrictive assumptions ($J = \emptyset$) can be found in Idrissi et al. (1988).

Maximin Location

The maximin location problem is to determine the locations of p facilities so as to maximize the minimum distance form a user to the closest facility. Examples are obnoxious facilities, such as nuclear plants, garbage dumps, sewage plants, etc. With $h_j(x^1, \ldots, x^p)$ defined by (7.87) the mathematical formulation of the problem is

$$\max\{ \min_{j=1,\ldots,N} h_j(x^1, \ldots, x^p) |\ x^i \in S,\ i = 1, \ldots, p\}. \tag{7.89}$$

An analogous problem, when locating emergency facilities, such as fire stations, hospitals, patrol car centers for a security guard company, etc., consists in minimizing the maximum distance from a user to the nearest facility. With the same $h_j(x^1, \ldots, x^p)$ as above the latter problem can be formulated as

$$\min\{\max_{j=1,\ldots,N} h_j(x^1, \ldots, x^p)|\ x^i \in S,\ i = 1, \ldots, p\}. \tag{7.90}$$

and is often referred to as the *p-center problem*.

Since $h_j(x^1, \ldots, x^p)$ is a dc function, it follows that $\min_{j=1,\ldots,N} h_j(x^1, \ldots, x^p)$ and $\max_{j=1,\ldots,N} h_j(x^1, \ldots, x^p)$ are also dc functions with easily computable explicit dc representations. Therefore, (7.89) and (7.90) are dc optimization problems which can be efficiently solved by currently available methods (see Al-Khayyal et al. 1997 for the case $p = 1$)

7.7.2 Distance Geometry Problem

A problem which has applications in molecular conformation and other questions such as surveying and satellite ranging, data visualization and pattern recognition, etc., is the *multidimensional scaling problem* or *distance geometry problem*. It consists in finding p objects $x^1, \ldots, x^p$ in $\mathbb{R}^n$ such that the quantity

$$V_p(x^1, \ldots, x^p) := \sum_{i<j} w_{ij}(\delta_{ij}^2 - \|x^i - x^j\|^2)^2 \tag{7.91}$$

is smallest, where $\Delta = (\delta_{ij})$, $W = (w_{ij})$ are symmetric matrices of order p such that

$$\delta_{ij} = \delta_{ji} \geq 0, \quad w_{ij} = w_{ji} \geq 0, \quad (i < j); \qquad \delta_{ii} = w_{ii} = 0\ (i = 1, \ldots, p).$$

By writing the problem as

$$\begin{array}{ll} \min & (\sum_{i<j} w_{ij}\|x^i - x^j\|^2 - 2\sum_{j<j} w_{ij}\delta_{ij}\|x^i - x^j\|) \\ \text{s.t.} & x^i \in \mathbb{R}^n, i = 1, \ldots, p \end{array} \tag{7.92}$$

or, alternatively, as

$$\min \sum_{i,j} w_{ij}t_{ij}^2 \left| \begin{array}{l} -t_{ij} \leq \delta_{ij}^2 - \|x^i - x^j\|^2 \leq t_{ij}\ (\forall i < j) \\ x^i \in \mathbb{R}^n,\ i = 1, \ldots, p \end{array} \right.$$

we again obtain a dc optimization problem of a particular type which can be solved by methods of nonconvex quadratic programming (see Chap. 10). An efficient local dc optimization called DCA has been developed by Tao (2003) for solving the multidimensional scaling problem on the basis of the dc formulation (7.92) (see also An (2003) where a combined DCA-smoothing technique is applied).

7.7.3 Continuous p-Center Problem

The continuous p-center problem differs from the p-center problem only in that the set of users is a rectangle $S \subset \mathbb{R}^2$ rather than a finite set of points. It consists of finding the locations $x^1, \ldots, x^p$ of p facilities in S so as to minimize the maximal distance from a point $x \in S$ to the nearest facility, i.e.,

$$\min \max_{y \in S} h_y(x^1, \ldots, x^p) \left| \begin{array}{l} h_y(x^1, \ldots, x^p) = \min_{i=1,\ldots,p} \|x^i - y\| \\ x^i \in S, i = 1, \ldots, p. \end{array} \right. \tag{7.93}$$

For each j define the Voronoi (Dirichlet) cell

$$D_j := \{x | \ \|x - x^j\| \leq \|x - x^i\| \ \forall i = 1, \ldots, p\}.$$

Computational results of (Suzuki and Okabe 1995) for $p = 30$ and $p = 49$ suggest the conjecture that for large values of p most Voronoi tend to be *hexagons*, only a few are *pentagons*.

7.7.4 Distribution of Points on a Sphere

In closing this section we should mention a challenging problem which has in recent time attracted the attention of mathematicians, biologists, chemists, and physicists, working on such fields as viral morphology, crystallography, molecular structure, and electrostatics (Saff and Kuijlaars 1997). Like many famous problems of mathematics, its formulation is rather simple: find the configuration of p points on a sphere in $\mathbb{R}^3$ so that

$$\min V_p(x^1, \ldots, x^p) := \sum_{i<j} \frac{1}{\|x^i - x^j\|} \quad \text{s.t.} \quad \|x^i\| = 1, \ = 1, \ldots, p. \tag{7.94}$$

Physically $V_p(x^1, \ldots, x^p)$ represents the energy of p charged particles that repel each other according to Coulomb's law. An optimal solution $X = (x^1, \ldots, x^p)$ of the problem (7.94) is called a p-tuple of *elliptic Fekete points*. Computational experiments have shown that for $32 \leq p \leq 200$ optimal configuration arrange points according to *hexagonal–pentagon pattern* (i.e., all but exactly 12 Voronoi cells are hexagonal, the exceptional cells being pentagons).

A related problem is to find a configuration of p points on the sphere that maximizes the product

$$W_p(x^1, \ldots, x^p) := \prod_{i<j} \|x^i - x^j\|. \tag{7.95}$$

The difficulty of problem (7.94) is that there are many local minima that are not global; in addition the local minimizers have objective function values very close to the global minimum, which make it very hard to determine the exact value of the global minimum.

Observe that $\frac{1}{\|x^i - x^j\|} = q(\|x^i - x^j\|)$ with $q = 1/t$ convex decreasing function for $t \geq \delta$. Hence, by Proposition 4.5 problem (7.94) can be rewritten as

$$\min\{\sum_{i<j}(g(x^1, x^2) - K\|x^i - x^j\|)|\ \|x^i\| = 1,\ i = 1, \ldots, p\},$$

where $g(t)$ is convex, K some positive constant. This is a dc minimization problem over a sphere which should be practically solvable for small values of p. Analogously, the problem

$$\max\{\prod_{i<j} \|x^i - x^j\| \mid \|x^i\| = 1,\ i = 1, \ldots, p\},$$

or equivalently,

$$\max\{\sum_{i<j} \log(\|x^i - x^j\|) \mid \|x^i| = 1,\ i = 1, \ldots, p\}$$

can be written as a dc optimization problem over a sphere:

$$\max\{\sum_{i<j}(L\|x^i - x^j\| - h_{ij}(x^i, x^j)) \mid \|x^i\| = 1,\ i = 1, \ldots, p\}$$

since $\log(\|x^i - x^j\|) = q(\|x^i - x^j\|)$ with $q(t) = \log t$ (concave increasing for $t \geq \eta > 0$).

7.8 Clustering via dc Optimization

The classification of objects into groups or clusters is important in many fields of operations research and applied sciences. Existing approaches to this problem include statistical, machine learning, mathematical programming and in particular, linear and integer programming. Among them approaches based on mathematical programming techniques are most effective (Bradley et al. 1997; Mangasarian 1997).

The problem of cluster analysis can be formulated as follows. For a given set of points $a^i, i = 1, \ldots, m$ in $\mathbb{R}^n$ find cluster points $x^l, l = 1, \ldots, p$ in $\mathbb{R}^n$ such that the sum of the minima over $l \in \{1, \ldots, p\}$ of the distance between each point a^i and the cluster centers $x^l, l = 1, \ldots, p$ is minimized. Using one-norm for measuring distances, the problem is

$$\begin{array}{ll} \text{minimize} \ \sum_{i=1}^m \min_{l=1,\ldots,p} \langle e, y^{il} \rangle & \\ \text{subject to} \ -y^{il} \le a^i - x^l \le y^{il} \quad i = 1, \ldots, m; \ l = 1, \ldots, p \end{array} \tag{7.96}$$

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

Since for each $i = 1, \ldots, m$ the function $\min_{l=1,\ldots,p} \langle e, y^{il} \rangle$ is concave piecewise linear, (7.96) is a piecewise linear concave minimization problem studied in Section 7.2. The main difficulty of this approach is that in practice m is too large for the problem to be solved efficiently. Therefore, exploiting the special form of the objective function (Bradley, Mangasarian and Street 1997) transformed the problem into an equivalent bilinear program, by rewriting it as

$$\begin{array}{ll} \min & \sum_{i=1}^m \sum_{l=1}^p \langle e, y^{il} t_{il} \rangle \\ \text{s.t.} & -y^{il} \le a^i - x^l \le y^{il} \quad i = 1, \ldots, m; \ l = 1, \ldots, p \\ & \sum_{l=1}^p t_{il} = 1, \ t_{il} \ge 0 \quad i = 1, \ldots, m; \ l = 1, \ldots, p \end{array} \tag{7.97}$$

This bilinear program is then solved by an algorithm which alternates between solving a linear program in the variable t and a linear program in the variables (x, y). Although this method terminates after finitely many iterations at a stationary point satisfying the necessary optimality conditions, this stationary point is not guaranteed to be a global optimal solution.

Furthermore, when 2-norm is used instead of 1-norm, the problem can no longer formulated as a concave minimization. Instead, it becomes a considerably harder problem of Weber multisource location problem. In Mangasarian (1997) the problem (7.97) is solved by a p-mean algorithm which again does not guarantee a global optimal solution.

Below we present a global optimization approach via dc optimization by Tuy et al. (2001). Another dc optimization approach based on the DCA algorithm by P.D Tao was developed by An et al. (2007).

7.8.1 Clustering as a dc Optimization

Since at optimality each point a^i must be assigned the cluster center closest to it, the clustering problem can be formulated as

$$\min \sum_{i=1}^m \min_{l=1,\ldots,p} \|x^l - a^i\| \quad \text{s.t.} \quad x^l \in S \subset \mathbb{R}^n, \ l = 1, \ldots, p. \tag{7.98}$$

where S is a compact convex set in $\mathbb{R}^n$. Without loss of generality we can assume that S is the unit simplex of $\mathbb{R}^n_+$. By a usual transformation,

$$\min_{l=1,\ldots,p} \|x^i - a^i\| = \sum_{l=1}^{p} \|x^l - a^i\| - \max_r \sum_{l\neq r} \|x^l - a^i\|$$

where $\sum_{l=1}^p \|x^l - a^i\|$ and $\max_r \sum_{l\neq r} \|x^l - a^i\|$ are convex functions. Therefore, (7.98) is actually the dc optimization problem

$$\begin{array}{ll} \min \sum_{l=1}^p \|x^l - a^i\| - \max_r \sum_{l\neq r} \|x^l - a^i\| \\ \text{s.t.} \qquad x^l \in S,\ l = 1,\ldots,p. \end{array} \tag{7.99}$$

A common method for solving this problem is by branch and bound. There are $p \times n$ variables $x^l_j, j = 1,\ldots,n, l = 1,\ldots,p$. For the sake of clarity any vector in $X = \prod_{l=1}^p X_l, X_l \in \mathbb{R}^n$, will be denoted by a boldface low case letter, e.g., $\mathbf{x} = (x^1,\ldots,x^p)$ where $x^l \in X_l$. We also agree that $\mathbf{x}_l = x^l$. The problem takes on the form

$$\min\{f(\mathbf{x}) := g(\mathbf{x}) - h(\mathbf{x}) |\ \mathbf{x} \in \mathbf{S}\}, \tag{7.100}$$

where

$$g(\mathbf{x}) = \sum_{i=1}^{m}\sum_{l=1}^{p} g_{il}\mathbf{x}), \quad g_{il}\mathbf{x}) = \|\mathbf{x}_l - a^i\|,$$

$$h(\mathbf{x}) = \sum_{i=1}^{m} h_i(\mathbf{x}), \quad h_i(\mathbf{x}) = \max_r \sum_{l\neq r} \|\mathbf{x}_l - a^i\|,$$

and $\mathbf{S}$ is a simplex in $X = \prod_{l=1}^p X_l,\ X_l = \mathbb{R}^n$.

a. Bounding

Let M be a subsimplex of $\mathbf{S}$, with vertex set $V(M)$ (note that $|V(M)| = pn + 1$). To compute a lower bound $\beta(M)$ for the value of $f(\mathbf{x})$ over M we replace the problem by a linear relaxation constructed as follows.

For each $\mathbf{v} \in V(M)$ let $p_{il}(\mathbf{v}) \in \partial g_{il}(\mathbf{v})$, i.e.,

$$p_{il}(\mathbf{v}) = \begin{cases} \frac{\mathbf{v}_l - a^i}{\|\mathbf{v}_l - a^i\|} & \text{if } \mathbf{v}_l \neq a^i, \\ 0 & \text{otherwise} \end{cases}$$

Then an affine minorant of $g_{il}(\mathbf{v})$ is $\langle p_{il}(\mathbf{v}), \mathbf{x}_l - \mathbf{v}_l\rangle + \|\mathbf{v}_l - a^i\|$, so a minorant of $g(\mathbf{x})$ is

$$\varphi(\mathbf{x}) = \max_{\mathbf{v}\in V(M)} \sum_{i=1}^{m}\sum_{l=1}^{p}[\langle p_{il}(\mathbf{v}), \mathbf{x}_l - \mathbf{v}_l\rangle + \|\mathbf{v}_l - a^i\|.$$

Also let $\psi_i(\mathbf{x})$ be a minorant of $h_i(\mathbf{x})$ provided by the affine function which matches with $h_i(\mathbf{x})$ at every $\mathbf{v} \in V(M)$, i.e.,

$$\psi_i(\mathbf{x}) = \sum_{\mathbf{v}\in V(M)} \lambda_v \max_r \sum_{l\neq r} \|\mathbf{v}^i - a^i\| \text{ for every}$$

$$\mathbf{x} = \sum_{\mathbf{v}\in V(M)} \lambda_{\mathbf{v}}\mathbf{v},\ \sum_{\mathbf{v}\in V(M)} \lambda_{\mathbf{v}} = 1,\ \lambda_{\mathbf{v}} \geq 0.$$

Proposition 7.18 *A lower bound of $f(\mathbf{x})$ over M is given by the value*

$$\beta(M) = \left|\begin{array}{l} \min_{t,\lambda}\{t - \sum_{i=1}^{m}\sum_{\mathbf{v}\in V(M)} \lambda_{\mathbf{v}} \max_r \sum_{l\neq r} \|\mathbf{v}_l - a^i\|\} \ s.t. \\ \sum_{i=1}^{m}\sum_{l=1}^{p}[\langle p_{il}(\mathbf{v}), \mathbf{x}_l - \mathbf{v}_l\rangle + \|\mathbf{v}_l - a^i\|] \leq t\ \forall \mathbf{v} \in V(M), \\ \mathbf{x}_l = \sum_{\mathbf{v}\in V(M)} \lambda_{\mathbf{v}}\mathbf{v}_l,\ \ \sum_{\mathbf{v}\in V(M)} \lambda_{\mathbf{v}} = 1,\ \lambda_{\mathbf{v}} \geq 0 \end{array}\right. \tag{7.101}$$

Proof A relaxation of the subproblem associated with M is

$$\min\{\varphi(\mathbf{x}) - \sum_{i=1}^{m} \psi_i(\mathbf{x})|\ \mathbf{x} \in M\} \tag{7.102}$$

which can also be written as

$$\min\{t - \sum_{i=1}^{m} \psi_i(\mathbf{x})|\ \varphi(\mathbf{x}) \leq t,\ \mathbf{x} \in M\}$$

and in this form it coincides with (7.101). □

We shall refer to the subproblem (7.101) with optimal value $\beta(M)$ as the *bounding subproblem* associated with M. Note that if $(\bar{t}, \overline{\lambda})$ is an optimal solution of the linear problem (7.101) then an optimal solution of the bounding subproblem (7.102) is $\mathbf{x}(M) = \sum_{\mathbf{v}\in V(M)} \overline{\lambda}_{\mathbf{v}}\mathbf{v}$.

Proposition 7.19 *We have $\varphi(\mathbf{x}) = g(\mathbf{x})$ and $\psi_i(\mathbf{x}) = h_i(\mathbf{x})$ $\forall i = 1, \ldots, m$ at every vertex of M. In particular, if $\mathbf{x}(M)$ is a vertex of M, i.e., if all but one $\overline{\lambda}_{\mathbf{v}}$, $\mathbf{v} \in V(M)$, are zeros, then $\overline{\mathbf{x}}(M)$ is a minimizer of $f(\mathbf{x})$ over M.*

Proof Obvious from the definition of $\varphi(\mathbf{x})$ and $\psi_i(\mathbf{x})$. □

b. Branching

Branching is performed using ω-subdivision. Specifically, let $\overline{\mathbf{x}}(M) = \sum_{\mathbf{v}\in V(M)} \overline{\lambda}_{\mathbf{v}}\mathbf{v}$ be an optimal solution of the relaxed problem (7.102) and let $J(M) = \{\mathbf{v} \in V(M)||\overline{\lambda}_{\mathbf{v}} > 0\}$. Then subdivide M via the point $\overline{\mathbf{x}}(M)$.

Theorem 7.11 *The BB Algorithm with the above bounding and the ω-bisection for branching is convergent. To be precise, if M_k is the partition set at iteration k, and $\overline{\mathbf{x}}^k$ is the current best solution, then a cluster point $\overline{\mathbf{x}}$ of the sequence $\{\overline{\mathbf{x}}^k\}$ yields an optimal solution of the problem.*

Proof The algorithm works exactly the same way as that for solving the equivalent concave minimization under convex constraints

$$\min\{t - h(\mathbf{x})|\ g(\mathbf{x}) - t \leq 0,\ \mathbf{x} \in \mathbf{S}\}.$$

The convergence of the algorithm then follows from Theorem 7.4 on the convergence of the BB Algorithm for (SBCP). □

Computational testing of the above algorithm has been carried out on the Wisconsin Diagnostic Breast Cancer Database created by Wolberg et al. (1994, 1995,a,b). This database consists of 569 vectors in $\mathbb{R}^{30}$, which should be arranged into two clusters (malignant/benign). The original number of feature parameters ($n = 30$) is too large for a direct application of global optimization methods, but it is argued that one can restrict consideration to four most informative parameters. So the clustering problem is solved for $m = 569, n = 4, p = 2$. The effectiveness of the method is confirmed by the performed computational experiments: in each set of these the two clusters obtained turned out to be correct with 96 % accuracy. For details, see Tuy et al. (2001).

7.9 Exercises

1 Consider the problem $\min\{\sum_{j=1}^{n} f_j(x_j) : x \in D\}$, where $D \subset \mathbb{R}^n_+$ is a polytope and

$$f_j(t) = \begin{cases} 0 & \text{if } t = 0, \\ d_j + c_j t & \text{if } t > 0. \end{cases}$$

(d_j is a fixed cost). Show that there exists an $\varepsilon > 0$ such that this problem is equivalent to a (BCP) with $f(x) = \sum_{j=1}^{n} \tilde{f}_j(x_j)$, where $\tilde{f}_j : \mathbb{R} \to \mathbb{R}$ is a concave function which is linear in the interval $(-\infty, \varepsilon)$ and agrees with $f_j(t)$ at points $t = 0$ and $t \in (\varepsilon, +\infty)$.

2 Let $\varepsilon \geq 0$ be a given number. A feasible solution $\mathbf{x}$ of (BCP) is said to be an ε-approximate (global optimal) solution if for any feasible x there exists at least one feasible point y such that $\|y-x\| \leq \varepsilon$ and $f(y) \geq f(\overline{x})$ (for $\epsilon = 0$, this reduces to the usual concept of global optimal solution). Consider the following cutting method for *(BCP)*:

Take a vertex x^0 of D. Set $D_0 = D, k = 0$.

(1) Let $\overline{x}^k \in \operatorname{argmin}\{f(x^i), i = 0, \ldots, k\}$, $\gamma_k = f(\overline{x}^k)$. Compute a γ_k-valid cut for (f, D_k) at $x^k : \pi^k(x - x^k) \geq 1$ and solve the linear program

$$\text{maximize} \quad \pi^k(x - x^k) \quad \text{subject to} \quad x \in D_k$$

obtaining its optimal value μ_k and a basic optimal solution x^{k+1}.

(2) If $\mu_k \leq 1 + \varepsilon\|p^k\|$, then terminate: accept $\overline{x}^k$ as an ε-approximate solution.

Otherwise, let $D_{k+1} = D_k \cap \{x : \pi^k(x - x^k) \geq 1 + \varepsilon\|\pi^k\|\}$ (geometrically, this amounts to add a slice of thickness ε to each cut), set $k \leftarrow k + 1$ and go to 1).

Show that this cutting method is finite if $\varepsilon > 0$. Does this method always give an ε-approximate solution of *(BCP)*?

3 Show that if the problem (*BCP*) has a unique global optimal solution then Procedure DC, in which the initial vertex x^0 is precisely the global optimal solution, cannot be infinite.

4 Solve the problem

$$\text{maximize} \quad (x_1 - \frac{1}{2})^2 + (x_2 - \frac{2}{3})^2 \quad \text{subject to}$$

$$-x_1 + x_2 \leq 1; \quad x_1 + x_2 \leq \frac{5}{2}; \quad \frac{3}{2}x_1 - x_2 \leq \frac{5}{4}; \quad \frac{1}{2}x_1 - x_2 \leq \frac{1}{4};$$

$$x_1 \geq 0; \quad 0 \leq x_2 \leq \frac{3}{2}.$$

5 Solve the problem

$$\text{minimize} \quad -x^2 - y^2 - (z-1)^2 \quad \text{subject to}$$

$$x + y - z \leq 0; \quad -x + y - z \leq 0; \quad -6x + y + z \leq 1.9$$

$$12x + 5y + 12z \leq 22.8; \quad 12x + 12y + 7z \leq 17.1; \quad y \geq 0.$$

6 Prove that (*LRC*) is equivalent to the parametric convex maximization problem: Find the smallest value θ such that

$$\max\{h(x) |\, Ax \leq b,\ x \geq 0,\ cx \leq \theta\} \geq 0.$$

7 Solve the (*LRC*):

$$\begin{aligned}
&\text{maximize} \quad -x_2 \quad \text{subject to} \\
&-x_1 + 2x_2 \leq 2; \quad 0 \leq x_1; \quad 0 \leq x_2 \leq 2 \\
&4(x_1 - 2)^2 + 4(x_2 - 2)^2 \geq 9; \quad -(x_1 + 3)^2 - x_2^2 \geq -\frac{81}{4}
\end{aligned}$$

8 Consider the *bilevel linear programming* problem:

$$\begin{aligned}
&\text{minimize} \quad 3x_1 + 2x_2 + y_1 + y_2 \quad \text{subject to} \\
&x_1 + x_2 + y_1 + y_2 \leq 4, \quad x_1 \geq 0, x_2 \geq 0 \\
&(y_1, y_2) \in \operatorname{argmin}\{4y_1 + y_2 : 3x_1 + 5x_2 + 6y_1 + 2y_2 \geq 15, y_1 \geq 0, y_2 \geq 0\}.
\end{aligned}$$

Restate this problem as a (LRC) and solve it by the FW/BW or BB Algorithm.

9 Show that the problem of minimizing a convex function $f(x)$ under the reverse convex constraint $h(x) \geq 0$ is always regular. Describe an algorithm for solving it.

10 Consider the *linear complementarity problem*: Given an $n \times n$ real matrix Q and a vector $p \in \mathbb{R}^n$ find a vector $x \in \mathbb{R}^n$ such that $Qx + p \geq 0, x \geq 0, \langle x, Qx + p \rangle = 0$. Show that this problem can be converted into a (*BCP*).

Chapter 8
Special Methods

Many problems have a special structure which should be exploited appropriately for their efficient solution. This chapter discusses special methods adapted to special problems of dc optimization and extensions.

8.1 Polyhedral Annexation

A drawback of OA algorithms is that, due to their stiff structure, these algorithms do not allow restart, nor multistart. Since a new linear constraint is added at each iteration while restart is not possible, the vertex set of the current outer approximating polytope may quickly reach a prohibitive size. Furthermore, OA methods require for their convergence certain conditions on the functions involved which may fail to hold in many problems of interest. It turns out that most of these difficulties can be overcome or mitigated by applying the OA strategy not to the original problem but, instead, to an equivalent problem in the dual space. This approach is based on the duality correspondence between convex sets and their polars, and is often referred to as *Polyhedral Annexation* (PA) or *Inner Approximation* (IA) (Tuy 1990).

*8.1.1 Procedure DC**

In Chap. 7 (Sect. 7.1), we saw how the core of many global optimization problems is a *DC feasibility problem* of the following form:

(DC) Given two closed convex sets C and D, find a point $x \in D \setminus C$ or else prove that $D \subset C$.

H. Tuy, *Convex Analysis and Global Optimization*, Springer Optimization and Its Applications 110, DOI 10.1007/978-3-319-31484-6_8

As in Chap. 6 (Sect. 6.3), assume that

$$0 \in D \cap (\mathrm{int} C). \tag{8.1}$$

By Proposition 1.21, condition (8.1), along with the convexity and closedness of D and C imply that $D = D^{\circ\circ}, C = C^{\circ\circ}$, so this polarity correspondence gives rise to a dual problem which will be referred to as the *dual DC feasibility problem:*

($C^{\circ}D^{\circ}$) Find a point $v \in C^{\circ} \setminus D^{\circ}$ or else prove that $C^{\circ} \subset D^{\circ}$.

Theorem 8.1 *The two problems (DC) and ($C^{\circ}D^{\circ}$) are equivalent in the following sense:*

(i) *$D \subset C$ if and only if $C^{\circ} \subset D^{\circ}$;*
(ii) *For each $v \in C^{\circ} \setminus D^{\circ}$ there exists a point $z \in D \setminus C$; conversely for each $z \in D \setminus C$ there exists a point $v \in C^{\circ} \setminus D^{\circ}$.*

Proof Since both D and C are closed convex sets, (i) follows immediately from Proposition 1.21. If $v \in C^{\circ} \setminus D^{\circ}$, then $\langle v, x\rangle \leq 1,\ \forall x \in C$ while $\langle v, z\rangle > 1$ for at least one $z \in D$, hence $z \in D \setminus C$. Conversely, if $z \in D \setminus C$, then $z \in D^{\circ\circ} \setminus C^{\circ\circ}$, i.e., $\langle u, z\rangle \leq 1\ \forall u \in D^{\circ}$ while $\langle v, z\rangle > 1$ for at least one $v \in C^{\circ}$, hence $v \in C^{\circ} \setminus D^{\circ}$. □

Thus, the DC feasibility problem can be solved by solving the dual DC feasibility problem and vice versa. As it often happens in optimization theory, the dual problem may turn out to be easier than the original one. In any case, the joint study of a dual pair of problems may give more insight into both and may suggest more efficient methods for handling each of them. Note that since $0 \in \mathrm{int} C$, the set C° is compact by Proposition 1.21. Furthermore, if D is compact then $0 \in \mathrm{int} D^{\circ}$, i.e., $0 \in C^{\circ} \cap \mathrm{int} D^{\circ}$, so there is a full symmetry in the above duality correspondence.

Setting $G = C^{\circ}, \Omega = C^{\circ} \setminus D^{\circ}$, we see that the dual DC feasibility problem is to find an element of $\Omega \subset G$, where G is a closed convex set. Let us examine how the OA method can be applied to solve this problem. Assume that $D = \{x |\ Ax \leq b, x \geq 0\}$ is a polyhedron and $x^0 = 0$ is a vertex of D satisfying (8.1). Let M be the canonical cone associated with $x^0 = 0$ (see Sect. 7.1). For every $i = 1, \ldots, n$ the i-th edge of M either meets ∂C at some point z^i or is entirely contained in C. In the latter case, let z^i be any point on this edge, arbitrarily far away from 0. Define the convex function

$$v \in \mathbb{R}^n \mapsto \mu(v) = \max\{\langle v, x\rangle |\ x \in D\}. \tag{8.2}$$

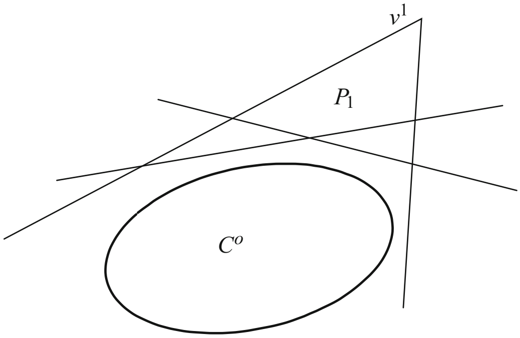
Lemma 8.1 *Let $S_1 = [0, z^1, \ldots, z^n]$ and let P be any polyhedron such that*

$$C^{\circ} \subset P \subset P_1 := S_1^{\circ}. \tag{8.3}$$

If V is the vertex set of P then

$$[\mu(v) \leq 1 \quad \forall v \in V] \quad \Rightarrow \quad C^{\circ} \subset D^{\circ}. \tag{8.4}$$

Fig. 8.1 Initial polyhedron P_1



Proof By Proposition 1.28, $P_1 = \{y|\ \langle z^i, y\rangle \le 1,\ i = 1,\dots,n\}$, so by Proposition 1.27 the recession cone of P_1 is the cone $\{y|\ \langle z^i, y\rangle \le 0,\ i = 1,\dots,n\} = M^\circ$ (note that $M = \text{cone}\{z^1,\dots,z^n\}$). Since $D \subset M$, any $u \in M^\circ$ satisfies $\mu(u) \le 0$. From (8.3) for any halfline $\{v + \lambda u|\ \lambda \ge 0\} \subset P$, we have $u \in \text{rec}P_1 = M^\circ$, hence $\mu(v + \lambda u) \le \max\{\langle v, x\rangle|\ x \in D\} + \lambda \max\{\langle u, x\rangle|\ x \in D\} \le \mu(v)$. So the convex function $\mu(v)$ is bounded on every halfline in P. Furthermore, P_1 has a unique vertex v^1 determined by

$$v^1 = eU^{-1}, \quad U = [z^1,\dots,z^n] \tag{8.5}$$

(e is the row vector of n ones, see (7.8)), so that $P_1 = M^\circ + v^1$ and by Proposition 1.20 P_1 contains no line (Fig. 8.1). By Proposition 2.34, then, $\max\{\mu(v)|\ v \in P\} = \max\{\mu(v)|\ v \in V\}$, and so $\mu(v) \le 1\ \forall v \in V$ implies that $\max\{\mu(v)|\ v \in P\} \le 1$, hence $P \subset D^\circ$, i.e., (8.4) because $C^\circ \subset P$. $\square$

By $\mathscr{P}$ now denote the family of polyhedrons P satisfying (8.3) and for each polyhedron $P_k \in \mathscr{P}$ with vertex set V_k define the distinguished point v^k of P_k to be the point

$$v^k \in \text{argmax}\{\mu(v)|\ v \in V_k\}. \tag{8.6}$$

By Lemma 8.1, if $\mu(v^k) \le 1$, then $C^\circ \subset D^\circ$ and the dual DC feasibility problem is solved. Otherwise, let $\mu(v^k) = \langle v^k, x^k\rangle,\ x^k \in D$. There are two possibilities:

(a) $x^k \notin C$: then $x^k \in D \setminus C$ and the problem is also solved.
(b) $x^k \in C$: then $v^k \notin C^\circ$ (because $\langle v^k, x^k\rangle > 1$) and the point $\hat{x}^k$ where the ray through x^k meets ∂C satisfies $\langle v^k, \hat{x}^k\rangle \ge \mu(v^k) > 1$, so the inequality

$$\langle \hat{x}^k, y\rangle \le 1 \tag{8.7}$$

holds for all $y \in C^\circ$ but not for $y = v^k$. Therefore, the cut (8.7) determines a polytope $P_{k+1} = \{y \in P_k|\ \langle \hat{x}^k, y\rangle \le 1\} \in \mathscr{P}$ such that $v^k \notin P_{k+1}$.

We can thus formulate the following OA procedure for solving the dual DC feasibility problem.

Procedure DC* (*D* Polyhedron, *C* Closed Convex Set)
Initialization. Select a vertex x^0 of D such that $x^0 \in \mathrm{int}C$ and let M be the associated canonical cone. Set $D \leftarrow D - x^0,\ C \leftarrow C - x^0,\ M \leftarrow M - x^0$. For each $i = 1, \ldots, n$ let z^i be the point where the i-edge of M meets ∂C (or any point of this edge if it does not meet ∂C). Let $P_1 = \{y |\ \langle z^i, y\rangle \le 1,\ i = 1, \ldots, n\},\ U = [z^1, \ldots, z^n]$, $v^1 = eU^{-1}$, $V_1 = \{v^1\}, \overline{V}_1 = V_1$. Set $k = 1$.

Step 1. For each $v \in \overline{V}_k$ solve

$$LP(v, D) \qquad\qquad \max\{\langle v, x\rangle |\ x \in D\},$$

to obtain its optimal value $\mu(v)$ and a basic optimal solution $x(v)$.

Step 2. Let $v^k \in \mathrm{argmax}\{\mu(v)|\ v \in V_k\}$. If $\mu(v^k) \le 1$, then terminate: $C^\circ \subset D^\circ$, hence $D \subset C$.

Step 3. If $\mu(v^k) > 1$ and $x^k := x(v^k) \notin C$, then terminate: $x^k \in D \setminus C$.

Step 4. If $\mu(v^k) > 1$ and $x^k \in C$, compute $\theta_k \ge 1$ such that $\hat{x}^k = \theta_k x^k \in \partial C$ and define

$$P_{k+1} = P_k \cap \{y |\ \theta_k \langle x^k, y\rangle \le 1\}. \tag{8.8}$$

Step 5. Compute the vertex set V_{k+1} of P_{k+1} and let $\overline{V}_{k+1} = V_{k+1} \setminus V_k$. Set $k \leftarrow k + 1$ and go back to Step 1.

Theorem 8.2 *Procedure DC* terminates after finitely many steps yielding a point of $D \setminus C$ or proving that $D \subset C$.*

Proof Each iteration generates an x^k which is a vertex of D (basic optimal solution of $LP(v^k, D)$). But $\hat{x}^k$ is distinct from all $\hat{x}^{k-1}, \ldots, \hat{x}^1$. Indeed, v^k satisfies all the constraints $\langle \hat{x}^i, y\rangle - 1 \le 0,\ i = 1, \ldots, k-1$, because $v^k \in P_k$, but violates the constraint $\langle \hat{x}^k, y\rangle - 1 \le 0$ because $\langle v^k, \hat{x}^k\rangle \ge \mu(v^k) > 1$. Therefore, x^k is distinct from all $x^i, i = 1, \ldots, k-1$. This implies that the number of iterations cannot exceed the number of vertices of D. □

Remark 8.1 All the linear programs $LP(v, D)$ have the same constraints. This should make the solution of these programs an easy task. If 0 is not a vertex but an interior point of D (so that $0 \in \mathrm{int}D \cap \mathrm{int}C$), then one can take the initial polyhedron to be $P_1 = \{y |\ \langle z^i, y\rangle \le 1,\ i = 1, \ldots, n+1\}$, where $z^1, \ldots, z^{n+1}$ are $n + 1$ points on ∂C determining an n-simplex containing 0 in its interior. In this case V_1 consists of the $n + 1$ vertices of P_1.

Remark 8.2 If we denote by S_k the polyhedron such that $S_k^\circ = P_k$ (i.e., $S_k = P_k^\circ$), then assuming $x^0 = 0$ and using Propositions 1.21 and 1.28 it can be verified that

$$S_1 \subset S_2 \subset \ldots \subset S_k \subset \ldots \subset C, \quad S_{k+1} = \mathrm{conv}(S_k \cup \{\hat{x}^k\}).$$

Indeed, from $P_{k+1} = \{y|\ \langle z^i, y\rangle \leq 1,\ i = 1,\dots,n;\ \langle \hat{x}^j, y\rangle \leq 1,\ j = 1,\dots,k\}$ it follows that for all $k = 1, 2, \dots$:

$$S_{k+1} = \text{conv}\{0, z^1, \dots, z^n,\ \hat{x}^1, \dots, \hat{x}^k\},$$

and hence, $S_{k+1} = \text{conv}(S_k \cup \{\hat{x}^k\})$.

The above procedure thus amounts to constructing a sequence of expanding polyhedrons contained in C, such that S_{h+1} is obtained from S_h by "annexing" $\hat{x}^h$ to S_h, until a polyhedron S_k is obtained which entirely covers D (in which case $D \subset C$) or a point $x^k \in D$ is found outside C (then $x^k \in D \setminus C$). This idea of *polyhedral annexation* (which gives the name to the method) was in fact suggested in Tuy (1964). On the other hand, since $S_1, S_2, \dots$ form a sequence of polyhedrons approximating the convex set C from the interior, the procedure can also be regarded as an *inner approximation method*.

Remark 8.3 With minor modifications Procedure DC* extends to the case where D may not be bounded. For instance, in this case, if the point x^k or $\hat{x}^k$ in Step 3 is at infinity in a direction u^k then the inequality in (8.8) should be replaced by

$$\langle u^k, y\rangle \leq 0.$$

When C has a nontrivial lineality space L, then $r := \dim C^\circ = n - \dim L < n$, so the dual DC feasibility problem is actually a problem in dimension $r < n$ and should be easier to solve than the original problem. We shall examine in Chap. 8 how the PA Algorithm can be streamlined to exploit this structure.

8.1.2 Polyhedral Annexation Method for (BCP)

To obtain an algorithm for (*BCP*) it only remains to incorporate Procedure DC* into the usual two phase scheme. Furthermore, to enhance efficiency, each time a better feasible solution than the incumbent is found, we can introduce a concavity cut to reduce the feasible polyhedron to be considered in the next cycle. This leads to the following:

PA (Polyhedral Annexation) Algorithm for (*BCP*)

Phase 1 Compute a vertex $\overline{x}$ of D.
Phase 2 Call Procedure DC* with $C = \{x|f(x) \geq f(\overline{x})\}$ (take the initial vertex x^0 to be such that $f(x^0) > f(\overline{x})$).

(a) If the procedure returns a vertex $z \in D \setminus C$, then set $\overline{x} \leftarrow z$ and go back to Phase 1. (Optionally: with $\gamma = f(z)$ construct a γ-valid cut $\pi(x - x^0) \geq 1$ for (f, D) at x^0 and reset $D \leftarrow D \cap \{x|\ \pi(x - x^0) \geq 1\}$).
(b) If the procedure establishes that $D \subset C$, then $\overline{x}$ is an optimal solution.

Theorem 8.3 *The PA Algorithm for (BCP) terminates after finitely many steps at a global minimizer.*

Proof Straightforward from the finiteness of Procedure DC*. □

Remark 8.4 Each new cycle is essentially a **restart**. Just as with the CS Restart Algorithm in Chap. 6, restart is a device to avoid a excessive growth of V_k. Moreover, since a new vertex x^0 is used to initialize Phase 2 and the region that remains to be explored is reduced by a cut, restart can also increase the chance of transcending the incumbent. In practical implementations restart with a new cut and a new x^0 is recommended when Phase 2 seems to drag on and $|V_k|$ approaches a critical limit.

Remark 8.5 In the above statement of the PA Algorithm we implicitly assumed that D is bounded. To extend the procedure to the case when D may be unbounded, it suffices to make some minor modifications indicated in Remark 8.3. It should be emphasized that in contrast to OA methods, the PA Algorithm does not require the computation of subgradients. Nor does it require the function $f(x)$ to be finite on an open set containing D (though it would be more efficient in this case). Another advantage of the PA Algorithm which will be demonstrated in Chap. 9 is that it can be used as a technique for reducing the dimension of the problem when the lineality of C is positive.

Example 8.1

$$\begin{aligned} \text{Minimize} \quad & f(x) := -(x_1 - 4.2)^2 - (x_2 - 1.9)^2 \\ \text{subject to} \quad & -x_1 + x_2 \leq 3 \\ & x_1 + x_2 \leq 11 \\ & 2x_1 - x_2 \leq 16 \\ & -x_1 - x_2 \leq -1 \\ & x_2 \leq 5 \\ & x_1 \geq 0, \; x_2 \geq 0. \end{aligned}$$

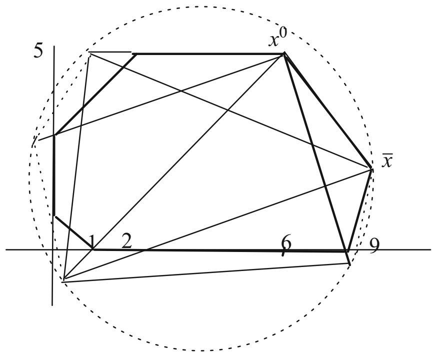
The algorithm is initialized from $z = (1.0, 0.5)$.

Phase 1 finds a vertex local minimizer $\overline{x} = (9.0, 2.0)$. In Phase 2 which is started from the vertex $x^0 = (6.0, 5.0)$, Procedure DC* with $C = \{x : f(x) \geq f(\overline{x})\}$ terminates after two iterations with the conclusion $D \subset C$. Hence the optimal solution is $\overline{x} = (9.0, 2.0)$ (Fig. 8.2).

8.1.3 Polyhedral Annexation Method for (LRC)

Since Procedure DC* solves the DC feasibility problem, it can also be incorporated into a two phase scheme for solving linear programs with an additional reverse convex constraint:

Fig. 8.2 Solving an example



$$(LRC) \qquad\qquad \min\{cx|\ x \in D \setminus \mathrm{int}C\}$$

where $D = \{x|\ Ax \leq b, x \geq 0\}$ is a polytope and $C = \{x|\ h(x) \leq 0\}$ is a closed convex set ($h : \mathbb{R}^n \to \mathbb{R}$ being a convex function).

In the following algorithm (which is analogous to the CS Restart Algorithm for (LRC), except that Procedure DC* is used instead of Procedure DC), $C_\varepsilon = \{x|\ h(x) \leq -\varepsilon\}$ and

$$X_\varepsilon = \bigcup \{\sigma(E)|\ E \in \mathscr{E}\}$$

where $\mathscr{E}$ is the collection of all edges of D and for every $E \in \mathscr{E},\ \sigma(E)$ is the set of extreme points of the line segment $E \cap C_\varepsilon$.

PA Algorithm for (LRC)

Initialization. Solve the linear program $\min\{cx|\ x \in D\}$ to obtain a basic optimal solution w. If $h(w) \geq -\varepsilon$, terminate: w is an ε-approximate optimal solution. Otherwise:

(a) If a vertex z^0 of D is available such that $h(z^0) > -\varepsilon$, go to Phase 1.
(b) Otherwise, set $\gamma = \infty, D(\gamma) = D$, and go to Phase 2.

Phase 1. Starting from z^0 search for a feasible point $\overline{x} \in X_\varepsilon$. Let $\gamma = c\overline{x}, D(\gamma) = \{x \in D|\ cx \leq \gamma\}$. Go to Phase 2.

Phase 2. Call Procedure DC* with $D(\gamma), C_\varepsilon$ in place of D, C resp.

(a) If a point $z \in D(\gamma) \setminus C_\varepsilon$ is obtained, set $z^0 \leftarrow z$ and go to Phase 1.
(b) If $D(\gamma) \subset C_\varepsilon$ then terminate: $\overline{x}$ is an ε-approximate optimal solution (when $\gamma < +\infty$), or the problem is infeasible (when $\gamma = +\infty$).

Proposition 8.1 *The PA Algorithm for LRC is finite.* □

8.2 Partly Convex Problems

Among nonconvex global optimization problems a class of particular interest is constituted by *partly convex problems* which have the following general formulation:

$$(P) \qquad \inf\{F(x,y)|\ G_i(x,y) \leq 0\ i = 1,\ldots,m,\ x \in C,\ y \in D\},$$

where C is a compact convex subset of $\mathbb{R}^n$, D a closed convex subset of $\mathbb{R}^p$, while $F(x,y) : C \times D \to \mathbb{R}, G_i(x,y) : C \times D \to \mathbb{R},\ i = 1,\ldots,m$, are lower semi-continuous functions, satisfying the following assumption:

(PCA) For every fixed $x \in \mathbb{R}^n$ the functions $F(x,y), G_1(x,y),\ldots,G_m(x,y)$ are *convex in y.*

When x, y are control variables, this is a problem encountered in numerous applications such as: pooling and blending in oil refinery, optimal design of water distribution, structural design, signal processing, robust stability analysis, design of chips, etc. When y is the control variable and x a parameter, this is a parametric optimization problem quite often considered in perturbation analysis of optimization problems.

In many cases of interest, e.g., when x is a parameter, the problem is so structured that it becomes much easier to solve when x is held temporarily fixed. To exploit this partitioning structure of the variables, a popular method to solve (P) is to proceed by branch and bound (BB), with branching performed in the x-space, i.e., in $\mathbb{R}^n$, according to an exhaustive rectangular successive subdivision rule (e.g., the standard bisection rule). Since in this way the problem is decomposed into a sequence of easier subproblems of smaller dimension, the procedure is referred to as a *decomposition method.*

Let $G(x,y) := (G_1(x,y),\ldots,G_m(x,y))$. Given a partition set $M \subset \mathbb{R}^n$ the subproblem associated with it is

$$(PM) \qquad \gamma(M) = \inf\{F(x,y)|\ G(x,y) \leq 0,\ x \in C \cap M,\ y \in D\}.$$

Consider the Lagrange dual to (P(M)):

$$(DP(M)) \qquad \alpha(M) = \sup_{\lambda \in \mathbb{R}^m_+} \inf_{x \in C \cap M, y \in D} \{F(x,y) + \langle \lambda, G(x,y)\rangle\}.$$

By weak duality property it is well known that

$$\alpha(M) \leq \gamma(M), \tag{8.9}$$

so $\alpha(M)$ yields a lower bound for the optimal value $\gamma(M)$ of the subproblem (P(M)). The problem (DP(M)) is called a *Lagrange relaxation* of (P(M)) and the bound $\alpha(M)$ is often referred to as a *duality bound* or a *Lagrangian bound.*

In general the inequality (8.9) is strict, i.e., the difference $\gamma(M)-\alpha(M)$, referred to as the *duality gap*, or simply the *gap*, is positive and most often rather large. However, it has been observed that, through a suitable (BB) scheme, one can generate a nested sequence of boxes $\{M_k\}$ shrinking to a point $x^* \in C$, such that the gap $\gamma(M_k) - \alpha(M_k)$ decreases as $k \to +\infty$. Under suitable conditions it is possible to eventually close the gap, i.e., to arrange things so that

$$\gamma(M_k)-\alpha(M_k) \searrow 0, \quad \text{as } k \to +\infty,$$

and an optimal solution y^* of the convex subproblem

$$\min\{F(x^*,y)|\ G(x^*,y) \leq 0,\ y \in D\}$$

will give an optimal solution (x^*, y^*) of (P).

This is the essential idea of the global optimization method by reducing the duality gap that was first developed in Ben-Tal et al. (1994) for partly linear problems, i.e., problems (P) in the special case when $F(x,y), G(x,y)$ are affine in y for every fixed value of x.

In fact, the idea of using duality bounds for solving nonconvex global optimization problems dates back at least to Falk (1969) and Falk and Soland (1969) (see also Ekeland and Temam 1976, and especially Shor and Stetsenko (1989) where a comprehensive study of the subject can be found). However, it was not until Ben-Tal et al. (1994) that the method was proposed for partly linear global optimization. Despite a fairly elaborate apparatus used, only partial and specific results were obtained. This motivated quite a few researches on refining the duality bound method and extending it to more general partly convex and nonconvex optimization problems. However, due to serious errors and misleading conclusions in several widespread publications, a rather confused situation resulted in this area that called for clarification (Tuy 2007a).

We first describe a prototype BB decomposition algorithm for the general problem (P) without assumption (PCA). Afterwards this algorithm will be specialized to problem (P) with assumption (PCA).

Since C is compact, it is contained in a rectangle $X \subset \mathbb{R}^n$. The following standard BB algorithm can be stated for solving problem (P):

Prototype BB Decomposition Algorithm

Select an exhaustive subdivision rule (e.g., the standard bisection). Take a number $\alpha \geq \sup\{F(x,y)|\ x \in C,\ y \in D\}$. If no such $\alpha \in R$ is readily available let $\alpha = +\infty$.

Initialization. If some feasible solutions are available, let CBS:=$(\overline{x}^0, \overline{y}^0)$ be the best among them. Set $M_1 := X$, $\mathscr{P}_1 := \mathscr{S}_1 := \{M_1\}$, $k := 1$.

Step 1. For each rectangle $M \in \mathscr{P}_k$ compute a lower bound for $F(x,y)$ over the feasible points in M, i.e., a number $\beta(M)$ such that

$$\beta(M) \leq \inf\{F(x,y)|\ G(x,y) \leq 0,\ x \in M \cap C,\ y \in D\}. \tag{8.10}$$

Step 2. Update the incumbent CBS by setting $(\overline{x}^k, \overline{y}^k)$ equal to the best among all feasible solutions available at the completion of the previous step.

Step 3. Delete every $M \in \mathscr{S}_k$ such that $\beta(M) \geq \min\{\alpha, F(\overline{x}^k, \overline{y}^k)\}$ (with the convention $F(\overline{x}^k, \overline{y}^k) = +\infty$ if $(\overline{x}^k, \overline{y}^k)$ is not defined). Let $\mathscr{R}_k$ be the set of remaining members of $\mathscr{S}_k$.

Step 4. If $\mathscr{R}_k = \emptyset$, then terminate: CBS $= (\overline{x}^k, \overline{y}^k)$ yields a global optimal solution, or else the problem is infeasible (if CBS is not defined).

Step 5. Choose $M_k \in \operatorname{argmin}\{\beta(M) | M \in \mathscr{R}_k\}$. Subdivide M_k according to the chosen subdivision rule. Let $\mathscr{P}_{k+1}$ be the partition of M_k.

Step 6. Let $\mathscr{S}_{k+1} := (\mathscr{R}_k \setminus \{M_k\}) \cup \mathscr{P}_{k+1}$. Increment k and go back to Step 1.

A key operation in this algorithm is the computation of a bound $\beta(M)$ for every partition set M (Step 1). According to the prototype BB Algorithm described in Sect. 6.2 this bound must be valid, which means that it must satisfy not only condition (8.10) but also

$$\{(x, y)|\ G(x, y) \leq 0,\ x \in M \cap C,\ y \in D\} = \emptyset \Rightarrow \beta(M) = +\infty. \tag{8.11}$$

Furthermore, it is convenient to assume that $\beta(M') \geq \beta(M)$ for $M' \subset M$ as we can always replace $\beta(M')$ with $\beta(M)$ if $\beta(M') < \beta(M)$.

However, since it is very difficult to decide whether the nonconvex set in the left-hand side of the implication (8.11) is empty or not, we replace this possibly impractical condition by a simpler one, namely

$$M \cap C = \emptyset \Rightarrow \beta(M) = +\infty. \tag{8.12}$$

Certainly, with this change the convergence criterion (6.5) for the prototype BB Algorithm may no longer be valid. Fortunately, however, as will be seen shortly, it is not difficult to find a suitable alternative convergence criterion.

Usually $\beta(M)$ is taken to be the optimal value of some relaxation of the subproblem $(P(M))$ in (8.10). The most popular relaxations are

- *linear relaxation*: $F(x, y)$ and $G(x, y)$ are replaced by their respective affine minorants, and C, D by outer approximating polyhedrons;
- *convex relaxation,* in particular *SDP (semidefinite programming) relaxation*: a suitable convex program (in particular a SDP) is derived whose optimal value is an underestimator of $\gamma(M) := \inf\{F(x, y)|\ G(x, y) \leq 0,\ x \in M \cap C,\ y \in D\}$;
- *Lagrange relaxation*: the subproblem (P(M)) is replaced by its Lagrange dual $(DP(M))$; in other words, $\beta(M)$ is taken to be the optimal value of $(DP(M))$:

$$\beta(M) = \sup_{\lambda \in \mathbb{R}^m_+} \inf\{F(x, y) + \langle \lambda, G(x, y)\rangle |\ x \in M \cap C,\ y \in D\}. \tag{8.13}$$

Whatever bounds are used, when infinite the above BB procedure generates a filter (an infinite nested sequence of partition sets) $M_k, k \in I \subset \{1, 2, \ldots\}$, such that

$$\beta(M_k) \leq \min(P)\ \forall k \in I, \quad M_k \cap C \neq \emptyset\ \forall k \in I, \quad \bigcap_{k \in I} M_k = \{x^*\}. \tag{8.14}$$

The algorithm is said to be *convergent* if $x^* \in C$ and

$$\min(P) = \min\{F(x^*, y)|\ G(x^*, y) \leq 0,\ y \in D\}, \tag{8.15}$$

so that any optimal solution y^* of this problem yields an optimal solution (x^*, y^*) of (P).

Theorem 8.4 *If $\beta(M_k) = +\infty$ for some k then (P) is infeasible and the algorithm terminates. If $\beta(M_k) < +\infty\ \forall k$ then there is an infinite subsequence $M_k, k \in I \subset \{1, 2, \ldots\}$, satisfying (8.14) and such that $x^* \in C$. If, in addition,*

$$\lim_{k\to\infty} \beta(M_k) = \min\{F(x^*, y)|\ G(x^*, y) \leq 0,\ y \in D\}, \tag{8.16}$$

then (8.15) holds, i.e., the BB decomposition algorithm is convergent.

Proof Since $M_{k+1} \subset M_k$ and hence, $\beta(M_{k+1}) \geq \beta(M_k)$ we have from (8.14)

$$\beta(M_k) \nearrow \beta^* \leq \min(P). \tag{8.17}$$

If $\beta(M_k) = +\infty$, then $\beta(M) = +\infty\ \forall M \in \mathscr{S}_k$, hence by (8.10) the problem is infeasible. If $\beta(M_k) < +\infty\ \forall k$, then by (8.14) $M_k \cap C \neq \emptyset\ \forall k$. The sets $M_k \cap C$, $k \in I$, then form a nested sequence of nonempty compact sets, so by Cantor theorem, $\cap_{k\in I}(M_k \cap C) = (\cap_{k\in I} M_k) \cap C \neq \emptyset$. Therefore, $x^* \in C$. On the other hand, since $\beta(M_k) \leq \min$ (P) by (8.14), it follows that $\beta^* = \lim_{k\to+\infty} \beta(M_k) \leq \min(P)$ (see (8.17)). If (8.16) holds then

$$\begin{aligned}\beta^* &= \min\{F(x^*, y)|\ G(x^*, y) \leq 0,\ y \in D\}\\ &\geq \min\{F(x, y)|\ G(x, y) \leq 0,\ x \in C,\ y \in D\} = \min(P),\end{aligned}$$

hence (8.15), as was to be proved. □

Remark 8.6 The condition $x \in M \cap C$ in (8.12) is essential to ensure that $x^* \in C$. It cannot be replaced by the weaker condition $x \in M$ (as has been proposed in some published algorithms), for then it may happen that $M_k \cap C = \emptyset$ (even though $\beta(M_k) < +\infty$) for some $k \in I$, hence $x^* \notin C$. To preclude this possibility one has to assume, as in the mentioned algorithms, that $M_1 \subset C$, and hence $M_k \subset C\ \forall k$, or else to compute, for each partition set M, a feasible solution $(x, y) \in M \times D$. The latter, however, is not always possible and may quite often be as difficult as solving the problem itself, especially when the constraints are highly nonconvex.

Back to *partly convex problems*, i.e., problem (P) with assumption (PCA), assume now that *Lagrangian bounds* are used throughout the BB decomposition algorithm. So for every partition set M a lower bound $\beta(M)$ is computed according to the formula (8.13).

Theorem 8.5 *Assume the set D is compact. Then the BB decomposition algorithm using Lagrangian bounds throughout is convergent.*

Proof Let us fix $x \in C$. The function $y \mapsto F(x,y) + \langle \lambda, G(x,y)\rangle$ is convex in $y \in D$ and linear in $\lambda \in \mathbb{R}^m_+$. Since this function is lower semi-continuous in y while D is compact we have the minimax equality (Theorem 2.7):

$$\inf_{y\in D} \sup_{\lambda\in\mathbb{R}^m_+} \{F(x,y) + \langle \lambda, G(x,y)\rangle\} = \sup_{\lambda\in\mathbb{R}^m_+} \min_{y\in D}\{F(x,y) + \langle \lambda, G(x,y)\rangle\}. \tag{8.18}$$

Since clearly

$$\sup_{\lambda\in\mathbb{R}^m_+} \{F(x,y) + \langle \lambda, G(x,y)\rangle\} = \begin{cases} F(x,y) \text{ if } G(x,y) \le 0, \\ +\infty \qquad \text{otherwise,} \end{cases} \tag{8.19}$$

we have from (8.18), for every $x \in C$,

$$\inf\{F(x,y)|\, G(x,y) \le 0,\ y \in D\} = \sup_{\lambda\in\mathbb{R}^m_+} \min_{y\in D}\{F(x,y) + \langle \lambda, G(x,y)\rangle\}. \tag{8.20}$$

Consider now a filter $\{M_k\}$ collapsing to a point x^* (to simplify the notation we write M_k instead of M_{k_ν}). As we saw in the proof of Theorem 8.4, $x^* \in C$, while by virtue of (8.20):

$$\inf\{F(x^*,y)|\, G(x^*,y) \le 0,\ y \in D\} = \sup_{\lambda\in\mathbb{R}^m_+} \min_{y\in D}\{F(x^*,y) + \langle \lambda, G(x^*,y)\rangle\}. \tag{8.21}$$

Let us show that this implies

$$\lim_{k\to\infty} \beta(M_k) = \inf\{F(x^*,y)|\, G(x^*,y) \le 0,\ y \in D\}. \tag{8.22}$$

From the obvious inequalities

$$\begin{aligned} \beta(M_k) &\le \inf\{F(x,y)|\, G(x,y) \le 0,\ x \in M_k \cap C, y \in D\} \\ &\le \inf\{F(x^*,y)|\, G(x^*,y) \le 0,\ y \in D\} \end{aligned}$$

it follows that

$$\beta(M_k) \nearrow \beta^* \le \inf\{F(x^*,y)|\, G(x^*,y) \le 0,\ y \in D\}.$$

Arguing by contradiction, suppose that (8.22) does not hold, i.e.,

$$\inf\{F(x^*,y)|\, G(x^*,y) \le 0,\ y \in D\} > \beta^*. \tag{8.23}$$

Then by virtue of (8.20),

$$\sup_{\lambda\in\mathbb{R}^m_+} \min_{y\in D}\{F(x^*,y) + \langle\lambda, G(x^*,y)\rangle\} > \beta^*, \tag{8.24}$$

so there exists $\tilde{\lambda}$ satisfying

$$\min_{y\in D}\{F(x^*,y) + \langle\tilde{\lambda}, G(x^*,y)\rangle > \beta^*.$$

Using the lower semi-continuity of the function $(x,y) \mapsto \{F(x,y) + \langle\tilde{\lambda}, G(x,y)\rangle\}$ we can then find, for every fixed $y \in D$, an open ball U_y in $\mathbb{R}^n$ around x^* and an open ball V_y in $\mathbb{R}^p$ around y such that

$$F(x',y') + \langle\tilde{\lambda}, G(x',y')\rangle > \beta^* \quad \forall x' \in U_y \cap C, \forall y' \in V_y.$$

Since the balls $V_y, y \in D$, form a covering of the compact set D, there is a finite set $E \subset D$ such that the balls $V_y,\ y \in E$, still form a covering of D. If $U = \cap_{y\in E} U_y$, then for every $y \in D$ we have $y \in V_{y'}$ for some $y' \in E$, hence

$$F(x,y) + \langle\tilde{\lambda}, G(x,y)\rangle > \beta^* \quad \forall x \in U \cap C,\ \forall y \in D.$$

But $M_k \subset U$ for all sufficiently large k, because $\cap_k M_k = \{x^*\}$. Then the just established inequality implies that

$$\sup_{\lambda\in\mathbb{R}^m_+} \min\{F(x,y) + \langle\lambda, G(x,y)\rangle \,|\, x \in M_k \cap C, y \in D\} > \beta^*.$$

Hence, $\beta(M_k) > \beta^*$, a contradiction. This completes the proof. □

Remark 8.7 It has been shown in Tuy (2007a) that Theorem 8.5 still holds if rather than requiring that D be compact it is assumed only that there exists a compact set $D^0 \subset D$ such that

$$(\forall x \in C)(\forall\lambda \in \mathbb{R}^m_+) \quad D^0 \cap \mathrm{argmin}_{y\in D}\{F(x,y) + \langle\lambda, G(x,y)\rangle\} \neq \emptyset.$$

However, a counter-example given there also pointed out that Theorem 8.5 will be false if the above condition is required only for $x \in C, \lambda \in \mathbb{R}^m_+$ such that $\mathrm{argmin}_{y\in D}\{F(x,y) + \langle\lambda, Gx,y)\rangle\} \neq \emptyset$.

For certain applications the assumption on compactness of D turned out to be too restrictive. In the next theorem this assumption is replaced by a weaker one, at the expense, however, of requiring the continuity and not merely the lower semi-continuity of the functions $F(x,y)$ and $G_i(x,y), i = 1,\ldots,m$.

Recall that a function $f : Y \to \mathbb{R}$ is said to be *coercive* on Y if $f(y) \to +\infty$ as $y \in Y, \|y\| \to +\infty$, or equivalently, if for any $\eta \in \mathbb{R}$ the set $\{y \in Y | f(y) \le \eta\}$ is bounded.

Theorem 8.6 *Assume that the functions $F(x,y), G_i(x,y)$, $i = 1,\dots,m$, are continuous in x for fixed $y \in D$ and satisfy the condition*

(S) for every $x \in C$ there exists $\lambda \in \mathbb{R}^m_+$ such that the function $F(x,y) + \langle\lambda, G(x,y)\rangle$ is coercive on D.

Then the BB decomposition algorithm using Lagrangian bounds throughout is convergent.

Proof In view of Theorem 8.5 it suffices to consider the case when D is unbounded. Consider a filter $\{M_k\}$ collapsing to a point x^*. As we saw previously, $x^* \in C$, and, as $k \to +\infty$,

$$\beta(M_k) \nearrow \beta^* \le \overline{\beta} := \inf\{F(x^*,y)|\ G(x^*,y) \le 0, y \in D\}.$$

According to (S) there exists $\tilde{\lambda} \in \mathbb{R}^m_+$ satisfying

$$F(x^*,y) + \langle\tilde{\lambda}, G(x^*,y)\rangle \to +\infty \quad \text{as } y \in D, \|y\| \to +\infty. \tag{8.25}$$

By the minimax theorem (Sect. 2.12, Remark 2.2), this ensures that

$$\begin{aligned}\overline{\beta} &= \sup_{\lambda\in\mathbb{R}^m_+} \inf_{y\in D}[F(x^*,y) + \langle\lambda, G(x^*,y)\rangle] \quad \text{(see (8.19))}\\ &= \min_{y\in D} \sup_{\lambda\in\mathbb{R}^m_+} [F(x^*,y) + \langle\lambda, G(x^*,y)\rangle].\end{aligned} \tag{8.26}$$

We show that $\beta^* = \overline{\beta}$. Suppose the contrary, that

$$\beta^* < \overline{\beta}. \tag{8.27}$$

Obviously $\theta\lambda + (1-\theta)\tilde{\lambda} \in \mathbb{R}^m_+$ for every (λ,θ) with $\lambda \in \mathbb{R}^m_+, 0 \le \theta < 1$. Let

$$\Lambda := \{\lambda \in \mathbb{R}^m_+ |\ \inf_{y\in D}[F(x^*,y) + \langle\lambda, G(x^*,y)\rangle] > -\infty\}. \tag{8.28}$$

For every k, since $\beta(M_k) \le \beta^*$, we have

$$\forall(\lambda,\theta) \in \Lambda \times [0,1) \qquad \inf_{x\in C\cap M_k, y\in D}[F(x,y) + \langle\theta\lambda + (1-\theta)\tilde{\lambda},\ G(x,y)\rangle] \le \beta^*.$$

Hence, if ε denotes an arbitrary number satisfying $0 < \varepsilon < \overline{\beta} - \beta^*$, then for every $(\lambda,\theta) \in \Lambda \times [0,1)$ and every k there exists $x^{(k,\lambda,\theta)} \in C \cap M_k,\ y^{(k,\lambda,\theta)} \in D$ satisfying

$$F(x^{(k,\lambda,\theta)}, y^{(k,\lambda,\theta)}) + \langle\theta\lambda + (1-\theta)\tilde{\lambda},\ G(x^{(k,\lambda,\theta)}, y^{(k,\lambda,\theta)})\rangle \le \beta^* + \varepsilon. \tag{8.29}$$

We contend that not for every $(\lambda, \theta) \in \Lambda \times [0, 1)$ the sequence $\{y^{(k,\lambda,\theta)}, k = 1, 2, \ldots\}$ is bounded. Indeed, otherwise we could assume that, for every $(\lambda, \theta) \in \Lambda \times [0, 1)$: $y^{(k,\lambda,\theta)} \to \overline{y}^{(\lambda,\theta)} \in D$, as $k \to +\infty$. Since $x^{(k,\lambda,\theta)} \in M_k$ while $\cap_{k=1}^{+\infty} M_k = \{x^*\}$, we would have $x^{(k,\lambda,\theta)} \to x^*$. Then letting $k \to +\infty$ in (8.29) would yield

$$F(x^*, \overline{y}^{(\lambda,\theta)}) + \langle \theta\lambda + (1-\theta)\tilde{\lambda},\ G(x^*, \overline{y}^{(\lambda,\theta)})\rangle \le \beta^* + \varepsilon,$$

for every $(\lambda, \theta) \in \Lambda \times [0, 1)$, hence

$$\sup_{\lambda\in\Lambda, 0\le\theta<1} \inf_{y\in D}[F(x^*, y) + \langle \theta\lambda + (1-\theta)\tilde{\lambda}, G(x^*, y)\rangle] \le \beta^* + \varepsilon < \overline{\beta}.$$

But for every $\lambda \in \Lambda$ the concave function $\varphi(\theta) = \inf_{y\in D}[F(x^*, y) + \langle \theta\lambda + (1 - \theta)\tilde{\lambda},\ G(x^*, y)\rangle]$ satisfies $\sup_{0\le\theta<1} \varphi(\theta) \ge \varphi(1)$, i.e.,

$$\sup_{0\le\theta<1} \inf_{y\in D}[F(x^*, y) + \langle \theta\lambda + (1-\theta)\tilde{\lambda},\ G(x^*, y)\rangle] \ge \inf_{y\in D}[F(x^*, y) + \langle \lambda,\ G(x^*, y)\rangle].$$

Therefore, we would have

$$\begin{aligned}
&\sup_{\lambda\in\mathbb{R}^+_m} \inf_{y\in D}[F(x^*, y) + \langle \lambda, G(x^*, y)\rangle \\
&\quad = \sup_{\lambda\in\Lambda} \inf_{y\in D}[F(x^*, y) + \langle \lambda,\ G(x^*, y)\rangle] \qquad \text{(because of (8.28))} \\
&\quad \le \sup_{\lambda\in\Lambda, 0\le\theta<1} \inf_{y\in D}[F(x^*, y) + \langle \theta\lambda + (1-\theta)\tilde{\lambda}, G(x^*, y)\rangle] < \overline{\beta},
\end{aligned}$$

contradicting (8.26). Thus, there exists $(\overline{\lambda}, \overline{\theta}) \in \Lambda \times [0, 1)$ such that, by passing to a subsequence if necessary, $\|y^{(k,\overline{\lambda},\overline{\theta})}\| \to +\infty$ $(k \to +\infty)$. Now, for an arbitrary point $y^* \in D$, let $\rho := F(x^*, y^*) + \langle \overline{\theta}\overline{\lambda} + (1-\overline{\theta})\tilde{\lambda},\ G(x^*, y^*)\rangle$. Then there is k_0 such that for all $k \ge k_0$,

$$F(x^{(k,\overline{\lambda},\overline{\theta})}, y^*) + \langle \overline{\theta\lambda} + (1-\overline{\theta})\tilde{\lambda},\ G(x^{(k,\overline{\lambda},\overline{\theta})}, y^*)\rangle < \rho + \varepsilon. \tag{8.30}$$

For simplicity of notation, let us write $\overline{x}^k, \overline{y}^k$ for $x^{(k,\overline{\lambda},\overline{\theta})}, y^{(k,\overline{\lambda},\overline{\theta})}$. For any $l > \|y^*\|$ define

$$\overline{y}^{kl} = \frac{l}{\|\overline{y}^k\|}\overline{y}^k + \left(1 - \frac{l}{\|\overline{y}^k\|}\right) y^*.$$

Clearly, there is $k_1 \ge k_0$ such that for all $k \ge k_1$,

$$0 < l/\|\overline{y}^k\| < 1, \quad l - \|y^*\| \le \|\overline{y}^{kl}\| \le l + \|y^*\|.$$

From the inequalities (8.29) and (8.30) and the convexity of the function $y \mapsto F(\overline{x}^k, y) + \langle \overline{\theta\lambda} + (1-\overline{\theta})\tilde{\lambda}, G(\overline{x}^k, y)\rangle$, we then deduce, for every fixed $l \geq \|y^*\|$,

$$F(\overline{x}^k, \overline{y}^{kl}) + \langle \overline{\theta\lambda} + (1-\overline{\theta})\tilde{\lambda},\ G(\overline{x}^k, \overline{y}^{kl})\rangle \leq \max\{\beta^*, \rho\} + \varepsilon < +\infty.$$

Since $\{\overline{y}^{kl}\}$ is bounded, we can assume $\overline{y}^{kl} \to u^l \in D$ as $k \to +\infty$. This yields

$$\begin{aligned}&\overline{\theta}[F(x^*, u^l) + \langle \overline{\lambda}, G(x^*, u^l)\rangle] \\ &\quad + (1-\overline{\theta})[F(x^*, u^l) + \langle \tilde{\lambda}, G(x^*, u^l)\rangle] \leq \max\{\beta^*, \rho\} + \varepsilon,\end{aligned}$$

where clearly $\|u^l\| \geq l - \|y^*\| \to +\infty$ as $l \to +\infty$. Since $\overline{\lambda} \in \Lambda$, by letting $l \to +\infty$ in the above inequality and taking account of (8.25) and (8.28), we get $+\infty \leq \max\{\beta^*, \rho\} + \varepsilon$, which is a contradiction. Therefore $\beta^* = \overline{\beta}$, as was to be proved. □

Remark 8.8 Condition (S) is equivalent to the following one:

(T) *for every $x \in C$ there exist $\lambda \in \mathbb{R}^m_+$ and a real number η such that the set $\{y \in D|\ F(x, y) + \langle \lambda, G(x, y)\rangle \leq \eta\}$ is nonempty and bounded.*

Indeed, it is straightforward that (S) implies (T). Conversely, if (T) holds then by a known property of proper convex functions (Corollary 2.3) the set $\{y \in D|\ F(x, y) + \langle \lambda, G(x, y)\rangle \leq \eta\}$ is bounded for every $\eta \in \mathbb{R}$, hence (S).

Corollary 8.1 *In Problem (P) assume that $F(x, y) = \langle c, y\rangle, G(x, y) = A(x)y - b$ with $A(x) \in \mathbb{R}^{m\times p}$ for every fixed x, while $D = \mathbb{R}^p_+$ and the following condition is satisfied:*

(S1) For every $x \in C$ there exists $\lambda \in \mathbb{R}^m_+$ such that $A(x)^T\lambda + c > 0$.

Then the BB decomposition algorithm using Lagrange bounds throughout is convergent.

Proof If (S1) holds then $F(x, y) + \langle \lambda, G(x, y)\rangle = \langle c, y\rangle + \langle \lambda, A(x)y - b\rangle = \langle A(x)^T\lambda + c, y\rangle - \langle \lambda, b\rangle \to +\infty$ as $y \in \mathbb{R}^p_+, \|y\| \to \infty$. Therefore, (S) holds. □

Thus, Theorem 8.6 includes the basic result in (Ben-Tal et al. 1994) for partly linear problems as a special case.

8.2.1 Computation of Lagrangian Bounds

As we saw, a key step of the above decomposition scheme is to compute the Lagrangian bounds. But these are in general not easily computable because in most cases the Lagrangian dual to a nonconvex problem is itself a nonconvex problem. Therefore, the first important issue that arises is when Lagrangian bounds can be practically computed.

We now show a class of problems for which the Lagrangian dual is a convex, or even a linear program. This class includes partly linear optimization problems which have the following general formulation:

$$(GPL) \qquad \min\{\langle c(x), y\rangle + \langle c^0, x\rangle \mid A(x)y + Bx \le b, r \le x \le s, y \ge 0\}$$

where $x \in \mathbb{R}^n,\ y \in \mathbb{R}^p,\ c : \mathbb{R}^n \to \mathbb{R}^p,\ c^0 \in \mathbb{R}^n,\quad A : \mathbb{R}^n \to \mathbb{R}^{m\times p},\ B \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m,\ r, s \in \mathbb{R}^n_+$.

Setting $A(x) = [a_{ij}(x)]$, and denoting the i-th row of B by B_i, this problem can also be written in the expanded form as

$$(GPL) \quad \left| \begin{array}{l} \min \sum_{j=1}^{p} y_j c_j(x) + \langle c^0, x\rangle, \\ \text{s.t. } \sum_{j=1}^{p} y_j a_{ij}(x) + \langle B_i, x\rangle \le b_i \quad i = 1, \dots, m, \\ y \ge 0, \qquad r \le x \le s. \end{array} \right.$$

Theorem 8.7 *The Lagrangian dual of* (GPL) *with respect to the nonlinear constraints, i.e.,*

$$\varphi^*_{[r,s]} = \sup_{\lambda \ge 0} \inf\{\langle y, c(x)\rangle + \langle c^0, x\rangle + \langle \lambda, A(x)y + Bx - b\rangle \mid x \in [r, s],\ y \ge 0\}, \tag{8.31}$$

is the convex program

$$\varphi^*_{[r,s]} = \langle c^0, x\rangle + \max_{\lambda,t}[\langle r - s, t\rangle + \langle Br - b, \lambda\rangle], \tag{8.32}$$

$$s.t. \quad g(\lambda) \ge 0,\ \ t + B^T\lambda \ge 0,\ \lambda \ge 0,\ t \ge 0, \tag{8.33}$$

where $g(\lambda) = \min_{j=1,\dots,p} \min_{r \le x \le s}[\langle A_j(x), \lambda\rangle + c_j(x)]$.

Proof For fixed $\lambda \ge 0$ we have

$$\begin{aligned} &\inf\{\langle y, c(x)\rangle + \langle \lambda, A(x)y + Bx - b\rangle \mid x \in [r, s],\ y \ge 0\} \\ &\quad = -\langle b, \lambda\rangle + \inf_{x \in [r,s]} \inf_{y \ge 0}\{\langle Bx, \lambda\rangle + \langle c(x) + (A(x))^T\lambda, y\rangle\} \\ &\quad = -\langle b, \lambda\rangle + h(\lambda), \end{aligned}$$

where

$$h(\lambda) = \begin{cases} \inf_{x \in [r,s]}\langle Bx, \lambda\rangle & \text{if} \quad c(x) + (A(x))^T\lambda \ge 0 \ \ \forall x \in [r, s], \\ -\infty & \text{otherwise.} \end{cases} \tag{8.34}$$

Now observe that for any $q \in \mathbb{R}^n$:

$$\min_{r \le x \le s}\langle q, x\rangle = \langle q, r\rangle + \max\{\langle r - s, t\rangle \mid t \ge 0, t \ge -q\},$$

since $\min\{\langle q, x\rangle \mid r \le x \le s\} = \min\{\langle q, r\rangle + \langle q, x - r\rangle \mid 0 \le x - r \le s - r\} = \langle q, r\rangle + \sum_{q_i < 0} q_i(s_i - r_i) = \langle q, r\rangle + \max\{\langle r - s, t\rangle \mid t \ge 0, t \ge -q\}$. Therefore,

$$\begin{aligned} &\inf_{r \le x \le s}\{\langle Bx, \lambda\rangle + \langle c^0, x\rangle\} \\ &\qquad = \langle B^T\lambda + c^0, r\rangle + \max\{\langle r - s, t\rangle \mid \ t \ge 0, t \ge -B^T\lambda - c^0\}. \end{aligned} \tag{8.35}$$

Denote the j-th column of A by A_j, so that the condition $\langle A(x), \lambda\rangle + c(x) \geq 0\ \forall x \in [r, s]$ means $g(\lambda) \geq 0$, where

$$g(\lambda) := \min_{j=1,\ldots,p} \min_{r\leq x\leq s} [\langle A_j(x), \lambda\rangle + c_j(x)].$$

Since for fixed x the function $\lambda \mapsto \langle A_j(x), \lambda\rangle + c_j(x)$ is affine, $g(\lambda)$ is a concave function and the problem (8.31) reduces to (8.32)–(8.33), which is a convex program. □

Corollary 8.2 *Assume that*

(*) For every j the functions $c_j(x), a_{ij}(x), i = 1, \ldots, m$, are either all increasing on $[r, s]$ or all decreasing on $[r, s]$.

Then the Lagrangian dual of (GPL) is the dual of its LP relaxation:

$$\min \left\{ \sum_{j\in J+} c_j(r)y_j + \sum_{j\in J-} c_j(s)y_j + \langle c^0, x\rangle \right\} \tag{8.36}$$

$$\text{s.t.} \sum_{j\in J_+} a_{ij}(r)y_j + \sum_{j\in J_-} a_{ij}(s)y_j + \langle B_i, x\rangle \leq b_i \quad i = 1, \ldots, m, \tag{8.37}$$

$$y \geq 0,\ r \leq x \leq s, \tag{8.38}$$

where J_+ is the set of all j such that all $c_j(x), a_{ij}(x), i = 1, \ldots, m$, are increasing, and J_- is the set of all j such that all $c_j(x), a_{ij}(x), i = 1, \ldots, m$, are decreasing.

Proof By hypothesis $J_+ \cup J_- = \{1, \ldots, p\}$, so for every $j = 1, \ldots, p$, we have

$$\begin{aligned}
&\min_{r\leq x\leq s} [\langle A_j(x), \lambda\rangle + c_j(x)] \geq 0\\
&\Leftrightarrow \min_{r\leq x\leq s} \left[\sum_{i=1}^{m} \lambda_i a_{ij}(x) + c_j(x)\right] \geq 0\\
&\Leftrightarrow \begin{cases} \sum_{i=1}^m \lambda_i a_{ij}(r) + c_j(r) \geq 0 & \text{if } j \in J_+ , \\ \sum_{i=1}^m \lambda_i a_{ij}(s) + c_j(s) \geq 0 & \text{if } j \in J_- . \end{cases}
\end{aligned} \tag{8.39}$$

In view of (8.32), (8.33), (8.35) the problem (8.31) thus reduces to the linear program

$$\langle c^0, x\rangle + \max\{\langle r - s, t\rangle + \textstyle\sum_{i=1}^m \lambda_i[\langle B_i, r\rangle - b_i]\},$$

$$\text{s.t.} \quad \left| \begin{array}{ll} -\sum_{i=1}^m \lambda_i B_i - t \leq c^0, & \\ -\sum_{i=1}^m \lambda_i a_{ij}(r) \leq c_j(r) & j \in J_+ , \\ -\sum_{i=1}^m \lambda_i a_{ij}(s) \leq c_j(s) & j \in J_- , \\ \lambda \geq 0,\ t \geq 0. & \end{array} \right.$$

whose dual is

$$\langle c^0, c\rangle + \min\left\{\sum_{j\in J_+} c_j(r)y_j + \sum_{j\in J_-} c_j(s)y_j\right\},$$

$$\text{s.t.} \sum_{j\in J_+} a_{ij}(r)y_j + \sum_{j\in J_-} a_{ij}(s)y_j + \langle B_i, r+z\rangle \le b_i \quad i=1,\dots,m,$$

$$y \ge 0, \quad 0 \le z \le s - r.$$

By setting $x = r + z$, the latter problem becomes (8.36)–(8.38). □

Corollary 8.3 *Assume that*

(#) for fixed λ each function $\langle A_j(x)\lambda + c_j(x)$ is quasiconcave.

Then the Lagrangian dual of (GPL) is the linear program

$$\max\left\{\langle r - s, t\rangle + \sum_{i=1}^{m} \lambda_i[\langle B_i, r\rangle - b_i]\right\},$$

$$\textit{s.t.} \left| \begin{array}{l} -\sum_{i=1}^m \lambda_i B_i - t \le 0, \\ -\sum_{i=1}^m \lambda_i a_{ij}(v) \le c_j(v) \quad v \in V,\ j=1,\dots,k, \\ \quad \lambda \ge 0,\ t \ge 0, \end{array}\right.$$

where V denotes the vertex set of the hyperrectangle $[r, s]$.

Proof Indeed, the condition $\min_{r\le x\le s}[\langle A_j(x), \lambda\rangle + c_j(x)] \ge 0$ is then equivalent to saying that $\min_{v\in V}[\langle A_j(v), \lambda\rangle + c_j(v)] \ge 0$. □

Thus, a Lagrangian bound for (*GPL*) is obtained by solving a convex program which reduces to a linear program if assumption (*) or (#) is satisfied. In the case (*) is satisfied, this linear program is merely a LP relaxation of (GPL).

8.2.2 Applications

The usefulness of the above results is illustrated by two examples of applications.

I. *The Pooling and Blending Problem*. It was shown in Ben-Tal et al. (1994) that this problem from petrochemical industry can be given in the form

$$\min\{c^T y |\ A(x)y \le b,\ y \ge 0,\ x \in X\},$$

where X is a hyperrectangle in $\mathbb{R}^n$, and $A(x)$ is an $m \times p$ matrix whose elements $a_{ij}(x)$ are continuous functions of x. Since this is a special case of (GPL) satisfying condition (#) of Corollary 8.3, the Lagrangian bound can be computed by solving a linear program. Furthermore, for this (GPL) condition (S) in Theorem 8.6 now reads

$$(\forall x \in X) \quad (\exists \lambda \in \mathbb{R}^m_+) \quad c^T y + \langle \lambda, A(x)y - b\rangle \to +\infty \text{ as } y \to +\infty,$$

where the last condition holds if and only if $\langle A(x), \lambda\rangle + c > 0$. Therefore, as shown in Corollary 8.1 this problem can be solved by a convergent BB decomposition algorithm using Lagrangian bounds throughout.

II. *The Bilinear Matrix Inequalities Problem.* Many challenging problems of control theory can be reduced to the so-called bilinear matrix inequality feasibility problem (BMI problem for short) with the following general formulation (Tuan et al. 2000a,b):

$$\min\langle c, x\rangle + \langle d, y\rangle \tag{8.40}$$

$$\text{s.t.} G_0 + \sum_{j=1}^{m} y_j G_j \preceq 0 \tag{8.41}$$

$$L_0 + \sum_{i=1}^{n} x_i L_{i0} + \sum_{j=1}^{m} y_j L_{0j} + \sum_{i=1}^{n}\sum_{j=1}^{m} x_j y_j L_{ij} \prec 0 \tag{8.42}$$

$$x \in X = [p, q] \subset \mathbb{R}^n, \; y \in \mathbb{R}^m_+ \tag{8.43}$$

where x, y are the decision variables, $G_0, G_j, L_0, L_{0i}, L_{j0}, L_{ij}$ are symmetric matrices of appropriate sizes, and the notation $G \preceq 0, L \prec 0$ means that G is a semidefinite negative matrix, L is a definite negative matrix.

For ease of notation we write

$$\begin{bmatrix} A \\ B \end{bmatrix}_d \quad \text{for} \quad \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix},$$

and define

$$A_{00}(x) = \begin{bmatrix} G_0 \\ L_0 + \sum_{i=1}^{n} x_i L_{i0} \\ \langle x, c\rangle \end{bmatrix}_d,$$

$$A_{j0}(x) = \begin{bmatrix} G_j \\ L_{0j} + \sum_{i=1}^{n} x_i L_{ij} \\ d_j \end{bmatrix}_d, \quad Q_{00} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}_d.$$

Then, as shown in Tuan et al. (2000a), this problem can be converted to the form

$$\min \left\{ t |\; A_0(x, p, q) + \sum_{j=1}^{m} y_j A_j(x, p, q) \preceq tQ, \; y \geq 0, \; x \in X \right\},$$

where

$$A_j(x,p,q) = \begin{bmatrix} A_{j0}(x) \\ A_{j1}(x,p,q) \end{bmatrix}_d, \; Q = \begin{bmatrix} Q_{00} \\ Q_{01} \end{bmatrix}_d, \; Q_{01} = 0,$$

$$A_{j1}(x,p,q) = \begin{bmatrix} (x_1 - p_1)G_j \\ (q_1 - x_1)G_j \\ \cdots \\ (x_n - p_n)G_j \\ (q_n - x_n)G_j \end{bmatrix}_d \qquad j = 0, 1, \ldots, n.$$

In this form the BMI problem appears to be a problem (GPL). Condition (S) in Theorem 8.6 can be formulated as

$$(\forall x \in X)(\exists Z_1 \succeq 0) \qquad \mathrm{Tr}(Z_1 Q_{00}) = 1, \quad \mathrm{Tr}(Z_1 A_{j0}(x)) > 0 \quad j = 1, \ldots, m.$$

Therefore, by Theorem 8.6, under this assumption the BMI problem can be solved by a convergent branch and bound algorithm using Lagrangian bounds, as proposed in Tuan et al. (2000a). Note that the Lagrangian dual of the problem

$$\max_{Z \succeq 0} \min_{t \in \mathbb{R}, y \geq 0, x \in M} \left\{ t + \mathrm{Tr}\left[Z(A_0(x,p,q) + \sum_{j=1}^{m} y_j A_j(x,p,q) - tQ) \right] \right\}$$

has been shown there to be equivalent to the LMI program

$$\max\{t |\ \mathrm{Tr}(ZA_0(x,p,q)) \geq t,\ \mathrm{Tr}(ZA_j(x,p,q)) \geq 0\ \forall x \in \mathrm{vert}M,\ j = 1, \ldots, m,$$
$$\mathrm{Tr}(ZQ) = 1,\ Z \succeq 0\}$$

where vert$M$ denotes the vertex set of M.

8.3 Duality in Global Optimization

As we saw in Chap. 6 and the preceding sections, Lagrange duality is very useful in global optimization, especially for computing bounds when solving a nonconvex problem by a BB algorithm. However, since the dual problem of a nonconvex optimization problem is also a nonconvex problem, these Lagrange bounds are often not easily computable. In this section we present a different duality theory due to Thach (1991, 1994, 2012) (see also Thach and Thang 2011) which has the attractive feature of being symmetric and may turn certain difficult nonconvex problems into more tractable, sometimes even convex or linear, ones.

8.3.1 Quasiconjugate Function

Basic for this nonconvex duality theory is the concept of quasiconjugation which plays a role similar to that of conjugation in convex duality theory (Sect. 2.10). Given any function $f : \mathbb{R}^n_+ \to \mathbb{R}_+, \not\equiv 0$, the function $f^\natural : \mathbb{R}^n_+ \to [0, +\infty)$ such that

$$f^\natural(p) = \frac{1}{\sup\{f(x)|\ p^T x \le 1,\ x \ge 0\}} \qquad \forall p \in \mathbb{R}^n_+,$$

(where $\frac{1}{+\infty} = 0$ by the usual convention) is called the *quasiconjugate* of $f(x)$.

Example 8.2 Let f be a Cobb–Douglas function on $\mathbb{R}^n_+$, i.e.,

$$f(x) = \prod_{i=1}^n x_i^{\alpha_i},$$

where $\alpha_i > 0, i = 1, \ldots, n$. Then $f^\natural$ is also a Cobb–Douglas function on $\mathbb{R}^n_+$:

$$f^\natural(p) = (\frac{1}{\alpha})^\alpha \prod_{i=1}^n (\frac{p_i}{\alpha_i})^{\alpha_i},\ p \in \mathbb{R}^n_+$$

where $\sum_{i=1}^n \alpha_i = \alpha$.

Proof If $p = (p_1, p_2, \ldots, p_n)$ and $p_i = 0$ for some $i = 1, \ldots, n$, then $p^T x^k \le 1$ for every $x^k = (\overline{x}_1, \overline{x}^2, \ldots, k\overline{x}_i, \ldots, \overline{x}_n)$, where $\overline{x} > 0$ satisfies $p^T\overline{x} \le 1$. Since $f(\overline{x}) > 0$, $f(x^k) = k^{\alpha_i} f(\overline{x}) \to +\infty$ as $k \to +\infty$ it follows that $f^\natural(p) = 0$.

If $p > 0$ then, since $f(x)$ is continuous the problem $\max\{f(x)|\ p^T x \le 1, x \ge 0\}$ has a solution $\overline{x}$. Noting that $f(x) = 0\ \forall x \in \mathbb{R}^n_+ \setminus \mathbb{R}^n_{++}$ we have $\overline{x} > 0$. By KKT condition there exists $\lambda > 0$ such that

$$\nabla f(\overline{x}) - \lambda \nabla(p^T\overline{x} - 1) = 0$$
$$\lambda(p^T\overline{x} - 1) = 0$$

Solving this system yields $\overline{x}_i = \frac{\alpha_i}{\alpha p_i}, i = 1, \ldots, n$ and $\lambda = \alpha \prod_{i=1}^n \overline{x}_i^{\alpha_i}$. Hence, $f^\natural(p) = \frac{1}{f(\overline{x})} = \left(\frac{1}{\alpha}\right)^\alpha \prod_{i=1}^n \left(\frac{p_i}{\alpha_i}\right)^{\alpha_i} \quad \forall p \in \mathbb{R}^n_+$. □

In the next proposition $f^{\natural\natural}$ denotes the quasiconjugate of $f^\natural$.

Theorem 8.8 (i) *The function $f^\natural$ is quasiconcave, nonnegative, and increasing on $\mathbb{R}^n_+$ and satisfies*

$$f^\natural(0) = \inf\{f^\natural(y)|\ y \in \mathbb{R}^n\}. \tag{8.44}$$

(ii) *If* $f : \mathbb{R}^n_+ \to \mathbb{R}_+$ *is quasiconcave, continuous, and increasing on* $\mathbb{R}^n_+$ *then* $f = f^{\natural\natural}$:

$$f(x) = \frac{1}{\sup\{f^{\natural}(p)|\, p^T x \le 1,\ p \ge 0\}}\ . \tag{8.45}$$

Proof (i) We need only prove the quasiconcavity of $f^{\natural}$. Let $p, p' \in \mathbb{R}^n_+$. For every $t \in [0,1]$ and $x \ge 0$ we can write

$$\langle tp + (1-t)p', x\rangle \le 1 \Rightarrow p^T x \le 1 \text{ or } p'^T x \le 1,$$

hence

$$\begin{aligned}&\sup\{f(x)|\ \langle tp + (1-t)p', x\rangle \le 1, x \ge 0\}\\ &\quad \le \max\{\sup\{f(x)|\, p^T x \le 1, x \ge 0\}, \sup\{f(x)|\, p'^T x \le, x \ge 0\},\end{aligned}$$

which implies

$$f^{\natural}(tp + (1-t)p') \ge \min\{f^{\natural}(p), f^{\natural}p'\},$$

as was to be proved.

(ii) Define

$$\overline{f}(x) = \frac{1}{\sup\{f^{\natural}(p) :\ p^T x \le 1, p \ge 0\}}.$$

For $\gamma > 0$ let $F_\gamma,\ \overline{F}_\gamma$ denote the upper levels of f and $f^{\natural}$, respectively, i.e.,

$$F_\gamma = \{x \in \mathbb{R}^n_+ : f(x) \ge \gamma\},\ \overline{F}_\gamma = \{x \in \mathbb{R}^n_+ : \overline{f}(x) \ge \gamma\}.$$

To prove the equality $f(x) = \overline{f}(x)\ \forall x \in \mathbb{R}^n_+$ it suffices to show that $F_\gamma = \overline{F}_\gamma$ for any $\gamma > 0$. Let $\overline{x} \in F_\gamma$. Suppose $\overline{x} \notin \overline{F}_\gamma$, then we have $\overline{f}(\overline{x}) < \gamma$ or

$$\frac{1}{\sup\{f^{\natural}(p) :\ p^T \overline{x} \le 1, p \ge 0\}} < \gamma,$$

hence, there exists $\overline{p} \ge 0, \overline{p}^T \overline{x} \le 1$ such that

$$\begin{aligned}f^{\natural}(\overline{p}) &> \frac{1}{\gamma}\\ &\Leftrightarrow \frac{1}{\sup\{f(x) :\ \overline{p}^T x \le 1, x \ge 0\}} > \frac{1}{\gamma}\\ &\Leftrightarrow \sup\{f(x) :\ \overline{p}^T x \le 1, x \ge 0\} < \gamma\\ &\Leftrightarrow f(x) < \gamma\ \forall x \ge 0 \text{ such that } \overline{p}^T x \le 1.\end{aligned}$$

It follows that $f(\overline{x}) < \gamma$. This conflicts with $\overline{x} \in F_\gamma$. Hence, $F_\gamma \subset \overline{F}_\gamma$.

Conversely, let $\overline{x} \in \overline{F}_\gamma$. We can write

$$
\begin{aligned}
\overline{f}(\overline{x}) &\geq \gamma \\
&\Leftrightarrow \frac{1}{\sup\{f^\natural(p) : p^T\overline{x} \leq 1, p \geq 0\}} \geq \gamma \\
&\Leftrightarrow f^\natural(p) \leq \frac{1}{\gamma} \ \forall p \geq 0 \text{ such that } p^T\overline{x} \leq 1 \\
&\Leftrightarrow \frac{1}{\sup\{f(x) : p^Tx \leq 1, x \geq 0\}} \leq \frac{1}{\gamma} \ \forall p \geq 0 \text{ such that } p^T\overline{x} \leq 1 \\
&\Leftrightarrow \sup\{f(x) : p^Tx \leq 1, x \geq 0\} \geq \gamma \ \forall p \geq 0 \text{ such that } p^T\overline{x} \leq 1. \qquad (8.46)
\end{aligned}
$$

For $\tilde{p} > 0$ such that $\tilde{p}^T\overline{x} \leq 1$, the set $\{x \in \mathbb{R}^n_+ : \tilde{p}^Tx \leq 1\}$ is compact. Since f is continuous, from (8.46) there exists a vector $\tilde{x} \geq 0$, $\tilde{p}^T\tilde{x} \leq 1$ such that $f(\tilde{x}) \geq \gamma$. This implies that $\tilde{x} \in F_\gamma$. Hence, F_γ is nonempty.

Suppose now that $\overline{x} \notin F_\gamma$. The function f is quasiconcave continuous and increasing on $\mathbb{R}^n_+$, so F_γ is a closed convex set satisfying $x' \in F_\gamma$ whenever $x' \geq x \in F_\gamma$. Since F_γ does not intersect with the line segment $[0, \overline{x}]$, by the strictly separation theorem there exist a vector $q \neq 0$ and real number α such that

$$q^Tx > \alpha \ \forall x \in F_\gamma; \qquad (8.47)$$

$$q^Tx < \alpha \ \forall x \in [0; \overline{x}]. \qquad (8.48)$$

From (8.47) we have $q \geq 0$ by Lemma 1. Further, from (8.48) it follows that $\alpha > 0$ and there exists $t > 0$ sufficiently small such that $(q + te)^T\overline{x} \leq \alpha$, where $e = (1, 1, \ldots, 1)$. Setting $\overline{p} = \frac{1}{\alpha}(q + te)$, we have $\overline{p} > 0$ and

$$\overline{p}^Tx \geq \frac{1}{\alpha}p^Tx > 1 \ \forall x \in F_\gamma; \qquad (8.49)$$

$$\overline{p}^T\overline{x} \leq 1. \qquad (8.50)$$

Consequently,

$$
\begin{aligned}
&\overline{p}^Tx > 1 \ \forall x \geq 0 \text{ such that } f(x) \geq \gamma \qquad (8.51) \\
&\Leftrightarrow f(x) < \gamma \ \forall x \geq 0 \text{ such that } \overline{p}^Tx \leq 1. \qquad (8.52)
\end{aligned}
$$

From (8.46) and (8.50) it follows that

$$\sup\{f(x) : \overline{p}^Tx \leq 1, x \geq 0\} \geq \gamma.$$

Since $\{x \in \mathbb{R}^n_+ : \overline{p}^Tx \leq 1\}$ is a compact set and f is continuous, there exists a vector $\hat{x} \geq 0$, $\overline{p}^T\hat{x} \leq 1$ such that $f(\hat{x}) \geq \gamma$. This conflicts with (8.52). Hence, $F_\gamma = \overline{F}_\gamma$. $\square$

8.3.2 Examples

The equation $f^{\natural\natural} = f$ holds for the following functions:

– Generalized Cobb–Douglas function

$$f(x) = \prod_{j=1}^{k} (f_j(x))^{\alpha_j}, \ \alpha_j > 0,$$

where $f_j(x),\ j = 1, 2, \ldots, k$ are positive concave and increasing on $\mathbb{R}^n_+$. This function is quasiconcave on $\mathbb{R}^n_+$.

– Generalized Leontiev function

$$f(x) = \left(\min\{\frac{x_j}{c_j} \,|\, j = 1, 2, \ldots, n\} \right)^{\alpha},$$

where $c_j > 0,\ j = 1, 2, \ldots, n,\ \alpha > 0$. This function is quasiconcave (and is then concave) if and only if $\alpha \leq 1$.

– Constant elasticity of substitution (C.E.S.) function

$$f(x) = (a_1 x_1^{\beta} + a_2 x_2^{\beta} + \cdots + a_n x_n^{\beta})^{\frac{1}{\beta}},$$

where $a_j > 0,\ i = 1, 2, \ldots, n,\ 0 < \beta \leq 1$. This function is quasiconcave on $\mathbb{R}^n_+$.

– Posynomial function

$$f(x) = \sum_{j=1}^{m} c_j \prod_{i=1}^{n} (x_i)^{a_{ij}} \quad \text{with } c_j > 0 \text{ and } a_{ij} \geq 0.$$

This function is quasiconcave on $\mathbb{R}^n_+$.

We now introduce the concept of quasi-supdifferential for quasiconcave functions. This will allow us to develop in the next section an optimality condition in the generalized KKT form for a quasiconcave maximization problem.

Given any function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ a vector $p \in \mathbb{R}^n_+$ will be called a *quasi-supgradient* of f at x if

$$p^T x = 1 \text{ and } f(x) f^{\natural}(p) \geq 1.$$

Since $f(x) f^{\natural}(p) \leq 1$ by definition of $f^{\natural}$ this is equivalent to saying that

$$p^T x = 1 \text{ and } f(x) f^{\natural}(p) = 1.$$

The *quasi-supdifferential* of f at x, written $\partial^\natural f(x)$, is then defined to be the set of all quasi-supgradients of f at x. If $\partial^\natural f(x)$ is nonempty, f is said to be quasi-supdifferentiable at x.

Proposition 8.2 *Let $f(x)$ be a continuous, quasiconcave, and monotonously increasing on $\mathbb{R}^n_+$. If f is differentiable at a point $\overline{x}$ such that $f(\overline{x}) > 0$ and $\nabla f(\overline{x}) \neq 0$ then*

$$\tilde{\nabla} f(\overline{x}) := \frac{1}{\langle \nabla f(\overline{x}), \overline{x}\rangle} \nabla f(\overline{x}) \in \partial^\natural f(\overline{x}), \tag{8.53}$$

where $\nabla f(\overline{x})$ denotes the gradient of f at $\overline{x}$.

Proof Since f is quasiconcave continuous and increasing on $\mathbb{R}^n_+$, the set $C_{\overline{x}} := \{x' \in \mathbb{R}^n_+ |\, f(x') \geq f(\overline{x})\}$ is closed, convex and satisfies $x' \geq x \in C_{\overline{x}} \Rightarrow x' \in C_{\overline{x}}$. For every $u \in \mathbb{R}^n_+$ we have

$$\langle \nabla f(\overline{x}), u\rangle = f'(\overline{x}, u) := \lim_{t \to 0+} \frac{f(\overline{x} + tu) - f(\overline{x})}{t}. \tag{8.54}$$

Since $f(\overline{x} + tu) \geq f(\overline{x})$ this implies $\langle \nabla f(\overline{x}), u\rangle \geq 0\ \forall u \in \mathbb{R}^n_+$, hence $\nabla f(\overline{x}) \geq 0$. For every $x \in C_{\overline{x}}$ if $t \in (0, 1)$ then $\overline{x} + t(x - \overline{x}) \in C_{\overline{x}}$, so $f(\overline{x} + t(x - \overline{x})) \geq f(\overline{x})$, hence, by (8.54) for $u = x - \overline{x}$, $\langle \nabla f(\overline{x}), x - \overline{x}\rangle \geq 0$. Thus $\langle \nabla f(\overline{x}), x\rangle \geq \langle \nabla f(\overline{x}), \overline{x}\rangle\ \ \forall x \in C_{\overline{x}}$.

Setting $p := \frac{1}{\langle \nabla f(\overline{x}), \overline{x}\rangle} \nabla f(\overline{x})$, we can write

$$\begin{aligned}
\langle p, x\rangle &\geq 1 \quad \forall x \in C_{\overline{x}} \\
&\Leftrightarrow f(x) < f(\overline{x}) \quad \forall x \geq 0 \text{ s.t. } \langle p, x\rangle < 1 \\
&\Rightarrow f(x) \leq f(\overline{x}) \quad \forall x \geq 0 \text{ s.t. } \langle p, x\rangle \leq 1 \\
&\Leftrightarrow \sup\{f(x)|\ \langle p, x\rangle \leq 1,\ x \geq 0\} \leq f(\overline{x}) \\
&\Leftrightarrow \frac{1}{\{\sup\{f(x)|\ \langle p, x\rangle \leq 1,\ x \geq 0\}} \geq \frac{1}{f(\overline{x})} \\
&\Leftrightarrow f(\overline{x}) f^\natural(p) \geq 1.
\end{aligned}$$

This together with the obvious equality $\langle p, \overline{x}\rangle = 1$ shows that $p \in \partial^\natural f(\overline{x})$. □

Theorem 8.9 *If f is quasiconcave continuous strictly increasing on $\mathbb{R}^n_+$, then $\partial^\natural f(x)$ is a nonempty convex set for any $x > 0$. Moreover,*

$$p \in \partial^\natural f(x) \Leftrightarrow x \in \partial^\natural f^\natural(p) \tag{8.55}$$

Proof Since f is strictly increasing on $\mathbb{R}^n_+$, it is easily seen that x is a boundary point of the closed convex set $C_x := \{x' \in \mathbb{R}^n_+ |\, f(x') \geq f(x)\}$. Indeed, if there were an open ball B of center x such that $B \subset C_x$ there would exist $\hat{x} \in B$ such that $\hat{x} < x$

which would imply $f(\hat{x}) < f(x)$, conflicting with $\hat{x} \in C_x$. Consequently, there exist a vector $q \neq 0$ and number $\alpha \in \mathbb{R}$ such that

$$\langle q, z\rangle \geq \alpha \quad \forall z \in C_x, \tag{8.56}$$

$$\langle q, x\rangle = \alpha. \tag{8.57}$$

From (8.56) we have $\langle q, x+u\rangle \geq \alpha \ \forall u \in \mathbb{R}^n_+$, hence, in view of (8.57), $\langle q, u\rangle \geq 0 \ \forall u \in \mathbb{R}^n_+$, i.e., $q \in \mathbb{R}^n_+$. Noting that $x > 0$ it then follows that $\alpha > 0$. Setting $p = \frac{1}{\alpha}q$ yields $p \geq 0$ and

$$p^T z \geq 1 \quad \forall z \in \mathbb{R}^n_+ \text{ s.t. } f(z) \geq f(x), \tag{8.58}$$

$$p^T x = 1. \tag{8.59}$$

From (8.58) we have $f(z) < f(x)$ for all $z \geq 0$ such that $p^T z < 1$, hence $f(z) \leq f(x)$ for all $z \geq 0$ such that $p^T z \leq 1$. Thus, $\sup\{f(z)|\, p^T z \leq 1, z \geq 0\} \geq f(x)$ and therefore $f^\natural(p)f(x) \geq 1$. This together with (8.59) shows that $p \in \partial^\natural f(x)$. The convexity of $\partial^\natural f(x)$ then follows from the definition of quasi-supgradient.

The equivalence (8.55) follows from the equality $(f^\natural)^\natural = f$ and the definition of subgradient. □

8.3.3 Dual Problems of Global Optimization

Let $f : \mathbb{R}^n_+ \to \mathbb{R}$ be a nonnegative and quasiconcave continuous strictly increasing function, $X \subset \mathbb{R}^n_+$ a compact convex with $\mathrm{int}X \neq \emptyset$, satisfying the following condition:

$$\forall x \in X \qquad x' \leq x \Rightarrow x' \in X. \tag{8.60}$$

In economics, when X represents the production set (i.e., the set of all feasible production programs) this condition is referred to as a *free disposal* condition. Obviously, it is equivalent to saying that

$$\forall x \in X \qquad [0, x] \subset X.$$

In monotonic optimization (see Chap. 11) a set $X \subset \mathbb{R}^n_+$ satisfying this condition is said to be *normal*.

Consider the optimization problem:

$$(P) \qquad\qquad \max\{f(x)|\, x \in X\}.$$

Since f is continuous and X is compact, this problem has a solution; moreover, its optimal value is positive.

For every vector $\overline{x} \in X$ let $N_X(\overline{x})$ be the normal cone of X at $\overline{x}$ (cf Sect. 1.6):

$$N_X(\overline{x}) = \{p|\, p^T(x - \overline{x}) \leq 0 \;\forall x \in X\}.$$

Theorem 8.10 *A vector $\overline{x} \in X$ is a global optimal solution of (P) if and only if it satisfies the generalized KKT condition*

$$0 \in \partial^{\natural} f(\overline{x}) - N_X(\overline{x}). \tag{8.61}$$

Proof Suppose $\overline{x} \in X$ and (8.61) holds. Then there exists a vector $p \in \partial^{\natural} f(\overline{x})$ such that $p \in N_X(\overline{x})$. Since $p \in \partial^{\natural} f(\overline{x})$ we have $p^T\overline{x} = 1$ and $f^{\natural}(p)f(\overline{x}) = 1$, while $p \in N_X(\overline{x})$ implies that $p^T(x - \overline{x}) \leq 0$, hence $p^T x \leq 1 \;\forall x \in X$. From the definition of $f^{\natural}$ it is immediate that

$$f(x)f^{\natural}(p) \leq 1 \quad \forall (x,p) \in X \times \mathbb{R}^n_+ \text{ s.t. } p^T x \leq 1.$$

Therefore,

$$f^{\natural}(p)f(\overline{x}) = \max\{f^{\natural}(p)f(x) :\; x \in X\} = f^{\natural}(p) \max_{x \in X} f(x),$$

and consequently,

$$f(\overline{x}) = \max_{x \in X} f(x).$$

So if (8.61) holds then $\overline{x}$ solves the problem (P).

Conversely, suppose $\overline{x}$ solves the problem (P). Then, as we saw in the proof of Theorem 8.9, $\overline{x}$ is a boundary point of the convex closed set $C_{\overline{x}} := \{x \in X|\, f(\overline{x}) \leq f(x)\}$ and hence there is a vector $p \in \partial^{\natural} f(\overline{x})$ such that

$$p^T x \geq 1 \;\forall x \in C_{\overline{x}}, \qquad p^T\overline{x} = 1.$$

So $p^T(x - \overline{x}) \leq 0 \;\forall x \in X$, i.e., $p \in N_X(\overline{x})$, hence, $0 \in \partial^* f(\overline{x}) - N_X(\overline{x})$. $\square$

Let $X^{\natural}$ be the lower conjugate of X defined as

$$X^{\natural} = \{p \geq 0|\, p^T x \leq 1 \;\forall x \in X\}. \tag{8.62}$$

Since $X^{\natural} = X^{\circ} \cap \mathbb{R}^n_+$, where X° is the polar of X defined in Sect. 1.8, it follows from the properties of X° that $X^{\natural}$ is a compact convex set with nonempty interior in $\mathbb{R}^n_+$ and, moreover, that X is the lower conjugate of $X^{\natural}$:

$$X = \{x \geq 0 : p^T x \leq 1 \;\forall p \in X^{\natural}\}. \tag{8.63}$$

Define the *dual* of the problem (P) to be the problem

$$(DP) \qquad\qquad \max\{f^{\natural}(p)|\, p \in X^{\natural}\}.$$

We show that $f^{\natural}(p)$ is continuous. For every $\alpha \in \mathbb{R}$ consider the set $C_\alpha := \{p \in \mathbb{R}^n_+ |\, f^{\natural}(p) \geq \alpha\} = \{p \in \mathbb{R}^n_+ |\, f(x) \leq 1/\alpha \ \forall x \in \mathbb{R}^n_+ \text{ s.t. } \langle p, x\rangle \leq 1\}$. Let $p^k \in C_\alpha$ and $p^k \to p$ as $k \to +\infty$. Then $f(x) \leq 1/\alpha$ for all $x \geq 0$ such that $\langle p^k, x\rangle \leq 1$. Observe that $\langle p^k, x\rangle = \langle p, x^k\rangle + \sum_{p_i=0} p_i^k x_i$ with

$$x_i^k = \begin{cases} \frac{p_i^k}{p_i} x_i & \text{if } p_i > 0 \\ x_i & \text{if } p_i = 0. \end{cases}$$

Consequently, if $\langle p^k, x\rangle \leq 1$, then $\langle p, x^k\rangle \leq 1$, hence $f(x^k) \leq 1/\alpha$, and therefore, as $k \to +\infty$, $f(x) \leq 1/\alpha$ by continuity of f. This implies $p \in C_\alpha$, so the set C_α is closed for every $\alpha \in \mathbb{R}$, proving that $f^{\natural}$ is u.s.c. Analogously, we can show that the set $\{p \in \mathbb{R}^n_+ |\, f^{\natural}(p) \leq \alpha$ is closed for every $\alpha \in \mathbb{R}$, so $f^{\natural}(p)$ is also l.s.c, hence $f^{\natural}(p)$ is continuous. Since $X^{\natural}$ is compact, the dual problem has an optimal solution.

Since X and $X^{\natural}$ are conjugates of each other and $(f^{\natural})^{\natural} = f$, the dual of the dual problem coincides with the primal problem. Therefore, *the duality is symmetric.* If $\overline{x}$ solves (P) and $\overline{p}$ solves (DP), then $f(\overline{x}) \leq \sup\{f(x) |\, \overline{p}^T x \leq 1, x \geq 0\} = \frac{1}{f^{\natural}(\overline{p})}$. The difference $1 - f(\overline{x}) f^{\natural}(\overline{p}) \geq 0$ measures in a sense the duality gap.

The next proposition shows that the duality gap is actually zero.

Theorem 8.11 *Let $\overline{x} \in X$ and $\overline{p} \in P$. Then $\overline{x}$ is optimal to (P) and $\overline{p}$ is optimal to (DP) if and only if*

$$f(\overline{x}) f^{\natural}(\overline{p}) = 1. \tag{8.64}$$

Proof Suppose $\overline{x} \in X$ and $\overline{p} \in P$ satisfy the duality equation (8.64). Then

$$f(\overline{x}) f^{\natural}(\overline{p}) = \max\{f(x) f^{\natural}(p) |\, x \in X, p \in P\} = \max_{x\in X} f(x) \max_{p \in P} f^{\natural}(p),$$

hence,

$$f(\overline{x}) = \max_{x \in X} f(x), \; f^{\natural}(\overline{p}) = \max_{p \in P} f^{\natural}(p).$$

Therefore, $\overline{x}$ solves (P) and $\overline{p}$ solves (DP).

Conversely, suppose $\overline{x}$ solves (P) and $\overline{p}$ solves (DP). Since $\overline{x}$ solves (P), $f(\overline{x}) > 0$. By Theorem 8.10 we have

$$0 \in \partial^{\natural} f(\overline{x}) - N_X(\overline{x}),$$

i.e., there is a vector $q \in \mathbb{R}^n_+$ such that

$$q^T \overline{x} = 1, \tag{8.65}$$

$$f(\overline{x}) f^{\natural}(q) = 1, \tag{8.66}$$

$$q^T x \leq 1 \ \forall x \in X. \tag{8.67}$$

From (8.67) it follows that $q \in X^\natural$. Since $\overline{x} \in X$, $\overline{x}^T p \leq 1\ \forall p \in X^\natural$. This together with (8.65) implies that $\overline{x}^T(p - q) \leq 0\ \forall p \in X^\natural$, i.e., $\overline{x} \in N_X(q)$. From (8.65) and (8.66) we have $\overline{x} \in \partial^\natural f^\natural(q)$. Thus,

$$0 \in \partial^\natural f^\natural(q) - N_X^\natural(q).$$

By Theorem 8.10, it follows that q is optimal solution of the problem (DP). Since $\overline{p}$ also solves (DP), we have $f^\natural(q) = f^\natural(\overline{p})$. Therefore, $f^\natural(\overline{p})f(\overline{x}) = 1$. □

Corollary 8.4 *Let $\overline{x} \in X$ and $\overline{p} \in X^\natural$. Then, $\overline{x}$ solves (P) and $\overline{p}$ solves (DP) if and only if*

$$\overline{p} \in \partial^\natural f(\overline{x}) \text{ or } \overline{x} \in \partial^\natural f^\natural(\overline{p})$$

Proof The "if" part is obvious. To prove the "only if" part, suppose $\overline{x}$ solves (P) and $\overline{p}$ solves (DP). By Theorem 8.11 we have

$$f^\natural(\overline{p})f(\overline{x}) = 1.$$

We show that $\overline{p}^T\overline{x} = 1$. Since $\overline{x}$ solves (P), we have $f(\overline{x}) > 0$. Again by Theorem 8.11

$$f^\natural(\overline{p}) = \frac{1}{\sup\{f(x) :\ \overline{p}^T x \leq 1, x \geq 0\}} = \frac{1}{f(\overline{x})}.$$

Hence,

$$\sup\{f(x) :\ \overline{p}^T x \leq 1, x \geq 0\} = f(\overline{x}). \tag{8.68}$$

If $\overline{p}^T\overline{x} = \alpha < 1$ there would exist a real number $t > 0$ such that $\overline{p}^T(\overline{x} + te) = 1$ (as usual $(e = (1, 1, \ldots, 1))$. Since f is strictly increasing, this would imply $f(\overline{x}) < f(\overline{x} + te)$, conflicting with (8.68). Therefore, $\overline{p}^T\overline{x} = 1$. □

8.4 Application

A basic feature of the quasi-gradient duality scheme as expressed in Corollary 8.4 is that a quasi-gradient of the primal objective function at a primal optimal solution is an optimal solution of the dual problem. Based on this property certain nonconvex problems can be converted into more tractable problems (Thach 1993b), sometimes even into convex problems (Thach and Thang 2014).

Below we illustrate this by an economic application discussed in Thach and Thang (2014).

8.4.1 A Feasibility Problem with Resource Allocation Constraint

Let us begin with a feasibility problem. Given m vectors $a^i \in \mathbb{R}^n_+,\ i = 1,\dots,m,$ together with a vector $\alpha \in \mathbb{R}^n_+$ such that

$$a^i_j > 0\ \forall j = 1,\dots,n; \quad \sum_{j=1}^{n} \alpha_j = 1, \tag{8.69}$$

consider the problem of finding a couple $(x, p) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+$ satisfying the following system of equalities and inequalities:

$$x = \sum_{i=1}^{m} \theta_i a^i,\ \theta_i \geq 0\ i = 1, 2, \dots, m,\ \sum_{i=1}^{m} \theta_i = 1 \tag{8.70}$$

$$p_j \geq 0\, j = 1, 2, \dots, n,\ p^T a^i \leq 1\ i = 1, 2, \dots, m \tag{8.71}$$

$$p_j x_j\ =\ \alpha_j\ j = 1, 2, \dots, n. \tag{8.72}$$

This problem is encountered, e.g., in the activity planning of a company that has m factories for producing n different goods. For the production of these goods a given resource of total amount 1 has to be allocated to the factories. Let a^i be the capacity of i-th factory, so i-th factory can produce a^i_j units of j-th good, $j = 1,\dots,n$ when running at full capacity. Let p_j denote the amount of resource to be allocated to the production of a unit of j-th good and x_j be the quantity of j-th good to be produced. Then $p_j x_j$ is the total amount of resource to be allocated to the production of j-th good. The technology or/and the market requires that $p_j x_j = \alpha_j, j = 1,\dots,n$, where $\alpha_1,\dots,\alpha_n$ are given numbers. The problem is to select a feasible activity program, i.e., a vector $(x, p) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+$, satisfying the system (8.70)–(8.72).

Let X denote the set of all $x \in \mathbb{R}^n_+$ satisfying (8.70) and P the set of all $p \in \mathbb{R}^n_+$ satisfying (8.71). It is obvious that both X, P are convex polytopes with nonempty interior. Furthermore, by formulas (8.62)–(8.63),

$$P = X^{\natural}, \qquad X = P^{\natural}. \tag{8.73}$$

A solution to the system (8.70)–(8.72) is a vector $(x, p) \in X \times P$ satisfying the equalities (8.72). Despite the nonconvexity of these equalities, we sill show that by quasi-gradient duality the system (8.70)–(8.72) can be reduced to a convex minimization problem. More specifically, this system has a unique solution (x, p) which can be obtained by solving either of two concave maximization (i.e., convex minimization) problems.

Define the functions

$$f(x) := \prod_{j=1}^{n} x_j^{\alpha_j}, \quad x \in X,$$

$$g(p) := \prod_{j=1}^{n} p_j^{\alpha_j}, \quad p \in P.$$

Proposition 8.3 *We have*

$$g(p) = f^{\natural}(p).\prod_{j=1}^{n}\alpha_j^{\alpha_j} \qquad f(x) = g^{\natural}(x).\prod_{j=1}^{n}\alpha_j^{\alpha_j}.$$

Proof Following Example 8.2, for every $p \geq 0$,

$$f^{\natural}(p) = \prod_{i=1}^{n}\left(\frac{p_j}{\alpha_j}\right)^{\alpha_j} = \frac{\prod_{j=1}^{n} p_j^{\alpha_j}}{\prod_{j=1}^{n}\alpha_j^{\alpha_j}}$$
$$= \frac{g(p)}{\prod_{j=1}^{n}\alpha_j^{\alpha_j}}.$$

Similarly, for every $x \geq 0$,

$$g^{\natural}(x) = \prod_{i=1}^{n}\left(\frac{x_j}{\alpha_j}\right)^{\alpha_j} = \frac{f(x)}{\prod_{j=1}^{n}\alpha_j^{\alpha_j}}. \qquad \square$$

Consider now the two optimization problems

$$\max\ f(x),\ \text{ s.t. } x \in X, \tag{8.74}$$

$$\max\ g(p),\ \text{ s.t. } p \in P. \tag{8.75}$$

These problems are equivalent, respectively, to the following concave maximization (i.e., convex minimization) problems

$$\max\left\{\sum_{j=1}^{n}\alpha_j \log(x_j)|\ \ x \in X\right\},$$

$$\max\left\{\sum_{j=1}^{n}\alpha_j \log(p_j)|\ \ p \in P\right\}.$$

Since X and P contain positive vectors, the optimal values in (8.74) and in (8.75) are positive. Moreover, since the functions $\log(f(x)) = \sum_{j=1}^{n}\alpha_j \log(x_j)$ and $\log(g(p)) = \sum_{j=1}^{n}\alpha_j \log(p_j)$ are strictly concave on the positive orthant, each problem (8.74) and (8.75) has a unique optimal solution.

Noting that $P = X^{\natural}$ while $g(p) = f^{\natural}(p).\prod_{j=1}^{n}\alpha_j^{\alpha_j}$ we see that problem (8.75) is the dual of problem (8.74). In fact, as $f(x) = g^{\natural}(x).\prod_{j=1}^{n}\alpha_j^{\alpha_j}$ and $X = P^{\natural}$ the two problems are dual of each other.

Theorem 8.12 *If $\overline{x}$ solves problem (8.74), then $\overline{p} = \tilde{\nabla} f(\overline{x})$ solves problem (8.75). Conversely, if $\overline{p}$ solves problem (8.75), then $\overline{x} = \tilde{\nabla} g(\overline{p})$ solves problem (8.74).*

Proof This follows from Corollary 8.4 and Proposition 8.2. $\square$

Theorem 8.13 *If $(\overline{x}, \overline{p})$ is a solution of the system (8.70)–(8.72), then $\overline{x}$ is optimal to (8.74) and $\overline{p}$ is optimal to (8.75). Conversely, if $\overline{x}$ is optimal to (8.74), then $(\overline{x}, \overline{p})$ with $\overline{p} = \tilde{\nabla} f(\overline{x})$ is a solution of the system (8.70)–(8.72); if $\overline{p}$ is optimal to (8.75), then $(\overline{x}, \overline{p})$ with $\overline{x} = \tilde{\nabla} g(\overline{p})$ is a solution to the system (8.70)–(8.72).*

Proof Let $(\overline{x}, \overline{p})$ be a solution of the system (8.70)–(8.72). From (8.72) we have $\overline{p}_i = \frac{\alpha_i}{\overline{x}_i}, i = 1, \ldots n$, while $\nabla f(\overline{x}) = (\alpha_1 \overline{x}_1^{\alpha_1 - 1}, \ldots, \alpha_n \overline{x}_n^{\alpha_n - 1})$, hence

$$\overline{p} = \frac{\nabla f(\overline{x})}{\langle \nabla f(\overline{x}), \overline{x} \rangle} = \tilde{\nabla} f(\overline{x}).$$

By Corollary 8.4 $\overline{x}$ solves (8.74) and $\overline{p}$ solves (8.75).

Conversely if $\overline{x}$ solves (8.74) and $\overline{p} = \tilde{\nabla} f(\overline{x})$ then obviously $\overline{p}$ satisfies (8.72), so $(\overline{x}, \overline{p})$ is a solution of the system (8.70)–(8.72). Analogously, if $\overline{p}$ solves (8.75) and $\overline{x} = \tilde{\nabla} g(\overline{p})$ then $\overline{x}$ satisfies (8.72), so $(\overline{x}, \overline{p})$ is a solution of the system (8.70)–(8.72). □

The above theorem shows that the nonconvex system (8.70)–(8.72) has a unique solution $(\overline{x}, \overline{p})$ which can be obtained either by solving the concave maximization problem (8.74) (then $\overline{x}$ is the unique solution of (8.74) and $\overline{p} = \tilde{\nabla} f(\overline{x})$) or by solving the concave maximization problem (8.75) (then $\overline{p}$ is the unique solution of (8.75) and $\overline{x} = \tilde{\nabla} g(\overline{p})$).

8.4.2 Problems with Multiple Resources Allocation

A more general problem than (8.70)–(8.72) occurs when a company has $k \geq 1$ resources to be used for the production of n goods in m factories of capacities $a^i \in \mathbb{R}^n_+$, $i = 1, \ldots, m$, where $\min_{j=1,\ldots,n} a^i_j > 0$. Let μ_r be the cost of r-th resource, $r = 1, \ldots, k$. Without loss of generality we can suppose $\sum_{r=1}^k \mu_r = 1$.

If a fraction α^r_j of r-th resource is allocated to the production of j-th good, and each resource is totally used then obviously

$$0 \leq \alpha^r_j \leq 1, \quad \sum_{j=1}^n \alpha^r_j = 1 \qquad r = 1, \ldots, k. \tag{8.76}$$

A vector $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n_+$ with $\alpha_j = \sum_{r=1}^k \mu_r \alpha^r_j$ (cost of all resources to be used for producing j-th good), $j = 1, \ldots, n$, will then be referred to as a *resource cost allocation vector*. It indicates that an amount of resources of total cost α_j is allocated to the production of j-th good, $j = 1, \ldots, n$. Let Δ be the set of all resource cost allocation vectors, i.e.,

$$\Delta = \left\{ \alpha \in \mathbb{R}^n_+ | \; \alpha = \sum_{r=1}^k \mu_r \alpha^r, \; \sum_{r=1}^k \mu_r = 1, \mu_r \geq 0 \quad r = 1, 2, \ldots, k \right\}. \tag{8.77}$$

A vector $x \in \mathbb{R}^n_+$ satisfying (8.70) will be termed, as previously, a *production program* while a vector $p = (p_1, \ldots, p_n)$ where p_j is the total cost of all resources used for producing a unit of j-th good (i.e., the production cost of a unit of j-th good) will be referred to as a *production unit cost vector*. Clearly p must satisfy (8.71).

The problem is now to find a production program $x \in \mathbb{R}^n_+$ and a production unit cost $p \in \mathbb{R}^n_+$ satisfying $(p_1x_1, \ldots, p_jx_j) \in \Delta$, where Δ is given by (8.77), that is, to solve the following nonlinear system:

$$x = \sum_{i=1}^{m} \theta_i a^i,\ \theta_i \geq 0 \quad i = 1, 2, \ldots, m,\ \sum_{i=1}^{m} \theta_i = 1, \tag{8.78}$$

$$p_j \geq 0\, j = 1, 2, \ldots, n,\ p^T a^i \leq 1 \quad i = 1, 2, \ldots, m, \tag{8.79}$$

$$p_j x_j = \alpha_j \quad j = 1, 2, \ldots, n, \tag{8.80}$$

$$\alpha = (\alpha_1, \ldots, \alpha_n) \in \Delta. \tag{8.81}$$

Clearly the system (8.70)–(8.72) is a special case of the system (8.78)–(8.81) when $k = 1$, i.e., when Δ is the standard simplex in $\mathbb{R}^n_+$ (which is obvious from (8.76)). We now show that solving the system (8.78)–(8.81) reduces to solving either of two suitable concave maximization problems.

Define

$$f_r(x) = \prod_{j=1}^{n} x_j^{\alpha_j^r} \quad r = 1, 2, \ldots, k,$$

$$g_r(p) = \prod_{j=1}^{n} p_j^{\alpha_j^r} \quad r = 1, 2, \ldots, k.$$

As before let X be the set of all x satisfying (8.78) and P the set of all p satisfying (8.79). For each $\mu \in \mathbb{R}^n_+$ consider the two following problems problems:

$$\max \left\{ \prod_{r=1}^{k} f_r(x)^{\mu_r} |\ x \in X \right\}. \tag{8.82}$$

$$\max \left\{ \prod_{r=1}^{k} g_r(p)^{\mu_r} |\ p \in P \right\}. \tag{8.83}$$

Each of these problems is the maximization of a continuous strictly concave function over a polytope and consequently has a unique optimal solution which can be found by methods of convex minimization.

Theorem 8.14 *Let $x \in P, p \in P$. Then $(\overline{x}, \overline{p})$ is a solution of the system (8.78)–(8.81) if and only if there exists $\mu \in \mathbb{R}^k_+$ such that $\overline{x}$ is optimal solution of problem (8.82) and $\overline{p}$ is optimal solution of problem (8.83).*

Proof (i) Suppose $(\overline{x}, \overline{p})$ is a solution of the system (8.78)–(8.81). According to (8.81) there is $\mu \in \mathbb{R}^k_+$ such that

$$\alpha = \sum_{r=1}^{k} \mu_r \alpha^r. \tag{8.84}$$

By Theorem 8.13, $\overline{x}$ is the unique optimal to problem (8.74) and $\overline{p}$ the unique optimal to problem (8.75), where $\alpha_j = \sum_{r=1}^{k} \mu_r \alpha_j^r$. However, from (8.84) it follows that

$$\begin{aligned}
f(x) &= \prod_{j=1}^{n} x_j^{\alpha_j} \\
&= \prod_{j=1}^{n} x_j^{\sum_{r=1}^{k} \mu_r \alpha_j^r} \\
&= \prod_{j=1}^{n} \prod_{r=1}^{k} x_j^{\mu_r \alpha_j^r} \\
&= \prod_{r=1}^{k} \prod_{j=1}^{n} x_j^{\mu_r \alpha_j^r} \\
&= \prod_{r=1}^{k} \left(\prod_{j=1}^{n} x_j^{\alpha_j^r} \right)^{\mu_r} \\
&= \prod_{r=1}^{k} f_r(x)^{\mu_r}.
\end{aligned}$$

So problem (8.74) is exactly problem (8.82). Similarly,

$$g(p) = \prod_{r=1}^{k} g_r(p)^{\mu_r},$$

and problem (8.75) is exactly problem (8.83).

(ii) Conversely, suppose there exists $\mu \in \mathbb{R}^k_+$ such that $\overline{x} \in X$ is optimal to problem (8.82) and $\overline{p}$ is optimal to problem (8.83). By Theorem 8.13 $(\overline{x}, \overline{p})$ is a solution of the system (8.78)–(8.81).

□

Thus any solution $(\overline{x}, \overline{p})$ of the nonconvex system (8.78)–(8.81) can be obtained either by taking $\overline{x}$ to be the optimal solution of problem (8.82) for some $\mu \in \mathbb{R}^k_+$ and $\overline{p} = \tilde{\nabla} f_r(\overline{x})$, or by taking $\overline{p}$ to be the optimal solution of problem (8.83) for some $\mu \in \mathbb{R}^k_+$ and $\overline{x} = \tilde{\nabla} g_r(\overline{p})$.

8.4.3 Optimization Under Resources Allocation Constraints

Consider now the optimization problem

$$\max\ q(x) \text{ s.t.} \tag{8.85}$$

$$x = \sum_{i=1}^{m} \theta_i a^i,\ \theta_i \geq 0\ i = 1, 2, \ldots, m,\ \sum_{i=1}^{m} \theta_i = 1, \tag{8.86}$$

$$p_j \geq 0\ j = 1, 2, \ldots, n,\ p^T a^i \leq 1\ i = 1, 2, \ldots, m, \tag{8.87}$$

$$p_j x_j = \alpha_j\ j = 1, 2, \ldots, n, \tag{8.88}$$

$$\alpha = (\alpha_1, \ldots, \alpha_n) \in \Delta. \tag{8.89}$$

where $q(x)$ is a continuous concave function defined on $\mathbb{R}^n_+$ such that $q(x) > 0\ \forall x > 0$. This is a difficult problem of concave maximization over a highly nonconvex set.

We shall refer to a vector $x \in \mathbb{R}^n_+$ for which there is a vector $p \in \mathbb{R}^n_+$ satisfying (8.86) through (8.89) as a *feasible production program* (under given resources allocation constraints). If $q(x)$ represents a *profit*, this problem amounts to finding a feasible production program with maximal profit.

By X_E denote the set of all feasible production programs x, so the problem (8.85)–(8.89) is

$$\max\{q(x)|\quad x \in X_E\}. \tag{8.90}$$

Recall from Theorem 8.14 that $x \in X_E$ if and only there exists $\mu \in \mathbb{R}^k_+$ such that x is the unique maximizer of the function $\prod_{r=1}^{k} f_r^{\mu_r}(x)$ over X. So problem (8.90) becomes

$$\max\left\{q(x)|\ x \in X, \mu \in \mathbb{R}^k_+,\ \prod_{r=1}^{k} f_r^{\mu_r}(x) = \max_{x' \in X} \prod_{r=1}^{k} f_r^{\mu_r}(x')\right\}. \tag{8.91}$$

By rescaling if necessary we can assume without lost of generality that $\min_{j=1,\ldots,n} a^i_j > 2$ for every $i = 1, \ldots, m$, and consequently, that

$$\min_{j=1,\ldots,n} x_j > 2 \quad \forall x \in X. \tag{8.92}$$

Define now

$$M = \left\{\eta = (\eta_1, \eta_2, \ldots, \eta_k) \in \mathbb{R}^n_+|\ \max_{x \in X} \prod_{r=1}^{k} f_r^{\eta_r}(x) \leq 3\right\},$$

$$h(\eta) = \max\left\{q(x)|\ \prod_{r=1}^{k} f_r^{\eta_r}(x) \geq 3,\ x \in X\right\} \quad \forall \eta \in \mathbb{R}^n_+.$$

(agree that $\max \emptyset = 0$)

Note that for each $\eta \in \mathbb{R}^k_+$ the function $\log\left(\prod_{r=1}^k f_r^{\eta_r}(x)\right) = \sum_{r=1}^k \eta_r \log(f_r(x)) = \sum_{r=1}^k \eta_r \sum_{j=1}^n \alpha_j \log(x_j)$ is concave, so the feasible set of the subproblem that defines $h(\eta)$ is a convex subset of the polytope X. Therefore, the value of $h(\eta)$ is obtained by solving a convex problem (maximizing a concave function over a compact convex set).

Lemma 8.2 *M is a nonempty compact convex set in $\mathbb{R}^k_+$.*

Proof For any $\eta \in \mathbb{R}^n_+$ we have

$$\max_{x\in X} \prod_{r=1}^k f_r^{\eta_r}(x) \le 3 \;\Leftrightarrow\; \max_{x\in X} \sum_{r=1}^k \eta_r \log(f_r(x)) \le \log(3).$$

Since $\max_{x\in X} \sum_{r=1}^k \eta_r \log(f_r(x)) \le \max_{x\in X} \log(f_r(x)) \sum_{r=1}^k \mu_r$, it is easily seen that there always exists $\eta \in \mathbb{R}^k_+$ such that $\max_{x\in X} \prod_{r=1}^k f_r^{\eta_r}(x) \le 3$. This proves that $M \neq \emptyset$.

The function

$$h'(\eta) \;=\; \max_{x\in X} \sum_{r=1}^k \eta_r \log(f_r(x))$$

is pointwise maximum of a family of linear functions of η, so it is a closed convex function on $\mathbb{R}^k_+$. Therefore, M is closed convex set. Furthermore, since $\log(f_r(x) = \sum_{j=1}^n \alpha_j^r \log(x_j) > \log(2)$ by (8.92), we have

$$\begin{aligned} \eta \in M \Leftrightarrow \eta \ge 0,\; & h'(\eta) \le \log(3) \\ \Rightarrow \eta \ge 0,\; & \sum_{r=1}^k \eta_r \log(f_r(x)) \le \log(3) \;\; \forall x \in X \\ \Rightarrow \eta \ge 0,\; & \sum_{r=1}^k \eta_r \log(2) \le \log(3) \end{aligned}$$

So, M is bounded. □

Lemma 8.3 *The function h is continuous and convex on $\mathbb{R}^k_+$.*

Proof The constraint $\prod_{r=1}^k f_r^{\eta_r}(x) \ge 3$ can be written as $\sum_{r=1}^k \eta_r \log(f_r(x)) \ge \log 3$. Since $q(x)$ and $\log(f_r(x))$ are continuous concave, the subproblem that defines $h(\eta)$ is a convex program. Since furthermore $\log f_r(x))$ is strictly concave, it is easily seen that there must exist $x \in X$ satisfying $\sum_{r=1}^k \eta_r \log(f_r(x)) > \log 3$. Therefore, by the strong Lagrange duality theorem, there exists $\lambda \in \mathbb{R}_+$ such that

$$h(\eta) = \max_{x\in X} \left\{ q(x) + \lambda \left[\sum_{r=1}^k \eta_r \log(f_r(x)) - \log 3 \right] \right\}.$$

Thus $h(\eta)$ is the upper envelope of a family of affine functions. The convexity and continuity of $h(\eta)$ on $\mathbb{R}^k_+$ follow. $\square$

Consider now the problem

$$\max\{h(\eta)|\ \eta \in M\} \tag{8.93}$$

which is a convex maximization over a convex compact set.

Theorem 8.4 *The optimal values in problems (8.90) and (8.93) are equal : $q^* = h^*$. Moreover, if η^* solves (8.93) then the unique maximizer x^* of the strictly concave function $\prod_{r=1}^k f_r^{\eta_r^*}(x)$ on X solves (8.90).*

Proof Let x^* be an optimal solution of (8.90), i.e., $x^* \in X_E,\ q(x^*) = q^*$. Since $x^* \in X_E$, there is a vector $\mu \in \mathbb{R}^n_+$ such that x^* is the optimal solution of (8.82). It is easy to see that for some $\gamma > 0$:

$$\prod_{r=1}^{k} f_r^{\gamma\mu_r}(x^*) = 3. \tag{8.94}$$

Indeed, let

$$\omega = \prod_{r=1}^{k} f_r^{\mu_r}(x^*).$$

By (8.92) $x_j^* > 2, j = 1, \ldots, n$, so $\log(\omega) > 0$. Taking

$$\gamma = \frac{\log(3)}{\log(\omega)}$$

yields $\gamma \log(\omega) = \log(3)$, i.e., $\log(\prod_{r=1}^k f_r^{\gamma\mu_r}(x^*)) = \log(3)$, hence (8.94). Now let $\eta = \gamma\mu$. Clearly

$$\max_{x\in X} \prod_{r=1}^{k} f_r^{\eta_r}(x) = \max_{x\in X} \left(\prod_{r=1}^{k} f_r^{\mu_r}(x)\right)^{\gamma} = \left(\prod_{r}^{k} f_r^{\mu_r}(x^*)\right)^{\gamma} = 3,$$

so $\eta \in M$. Hence,

$$\begin{aligned} q^* &= q(x^*) \\ &\le \sup\left\{q(x)|\ \prod_{r=1}^{k} f_r^{\eta_r}(x) \ge 3,\ x \in X\right\} \\ &= h(\eta) \\ &\le h^*. \end{aligned}$$

This implies, in particular, that $h^* > 0$. Conversely, let η^* be an optimal solution of (8.44), i.e., $\eta^* \in M$ and $h(\eta^*) = h^*$. By x^* denote a maximizer of the function $\prod_{r=1}^{k} f_r^{\eta_r^*}(x)$ on X. By Theorem 8.14 $x^* \in X_E$. Since $\eta^* \in M$ we have

$$\begin{aligned}\prod_{r=1}^{k} f_r^{\eta_r^*}(x^*) &= \sup_{x\in X}\prod_{r=1}^{k} f_r^{\eta_r^*}(x)\\ &\leq 3.\end{aligned}$$

But if

$$\prod_{r=1}^{k} f_r^{\eta_r^*}(x^*) \;<\; 3,$$

then

$$\begin{aligned}h(\eta^*) &= \sup\left\{q(x)|\ \prod_{r=1}^{k} f_r^{\eta_r^*}(x) \geq 3,\ x\in X\right\}\\ &= \sup \emptyset\\ &= 0\\ &< h^*,\end{aligned}$$

which is a contradiction. Thus,

$$\prod_{r=1}^{k} f_r^{\eta_r^*}(x^*) \;=\; 3$$

and therefore,

$$\begin{aligned}h^* &= h(\eta^*)\\ &= \sup\left\{q(x)|\ \prod_{r=1}^{k} f_r^{\eta_r^*}(x) \geq 3,\ x\in X\right\}\\ &= q(x^*)\\ &\leq q^*.\end{aligned}$$

Consequently, $q^* = h^*$, and so x^* is an optimal solution of the problem (10.12). □

Thus, by quasi-gradient duality a difficult nonconvex problem in an n-dimensional space like (8.85)–(8.89) can be converted into a k-th dimensional problem (8.93) with $k \leq n$, which is a much more tractable problem of convex maximization over a compact convex set. Especially when k is small, as is often the case in practice, the converted problem can be solved efficiently even by outer approximation method as proposed in Thach et al. (1996).

8.5 Noncanonical DC Problems

In many dc optimization problems the constraint set is defined by a mixture of inequalities of the following types:

- convex;
- reverse convex;
- piecewise linear;
- quadratic (possibly indefinite).

Problems of this form often occur, for instance, in continuous location theory (Tuy 1996; Tuy et al. 1995a). To handle these problems it is not always convenient to reduce them to the canonical form. In this section we present an outer approximation method based on a "visibility" assumption borrowed from location theory. This method is practical when the number of nonconvex variables is small.

Consider the problem

$$\min\{f(x)| \quad x \in S\}, \tag{8.95}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function, and S is a compact set in $\mathbb{R}^n$.

A feasible point x is said to be *visible from a given point* $w \in \mathbb{R}^n \setminus S$ if x is the unique feasible point in the line segment $[w, x]$, i.e., if $[w, x] \cap S = \{x\}$ (Fig. 7.3). Our basic assumption is the following:

Visibility Assumption: The feasible set S is such that there exists an efficient finite procedure to compute, for any given point $w \in \mathbb{R}^n \setminus S$ and any $z \neq w$, the visible feasible point $\pi_w(z)$ on the line segment $[w, z]$, if it exists.

Note that since S is compact, either the line through w, z does not meet S, or $\pi_w(z)$ exists and is the nearest feasible point to w in this line. To simplify the notation we shall omit the subscript w and write $\pi(z)$ when w is clear from the context.

Proposition 8.4 *If $S = \{x| \ g_i(x) \leq 0 \ \ i = 1, \ldots, m\}$ and each function $g_i(x)$ is either convex, concave, piecewise linear, or quadratic, then S satisfies the visibility assumption.*

Proof It suffices to check that if $g(x)$ belongs to one of the mentioned types and $g(w) > 0$ then, for any given $z \neq w$ one can easily determine whether $g(w + \lambda(z - w)) \leq 0$ for some $\lambda > 0$ and if so, compute the smallest such λ. In particular, if $g(x)$ is quadratic (definite or indefinite), then $g(w + \lambda(z - w))$ is a quadratic function of λ, so computing $\sup\{\lambda| \ g(w + \lambda(z - w)) \leq 0\}$ is an elementary problem. □

Since $f(x)$ is convex one can easily compute $w \in \operatorname{argmin}\{f(x)| \ x \in \mathbb{R}^n\}$. If it so happens that $w \in S$, we are done: w solves the problem. Otherwise, $w \notin S$ and we have

$$f(w) < \min\{f(x)| \ x \in S\}. \tag{8.96}$$

From the convexity of $f(x)$ it then follows that $f(w + \lambda(z - w)) < f(z)$ for any $z \in S, \lambda \in (0, 1)$, so the visible feasible point in $[w, z]$ achieves the minimum of $f(x)$ over the line segment $[w, z]$. Therefore, assuming the availability of a point w satisfying (8.96) we can reformulate (8.95) as the following:

Optimal Visible Point Problem: Find the feasible point visible from w with minimal function value $f(x)$.

Aside from the basic visibility condition and (8.96), we will further assume that:

(a) the set S is robust, i.e., $S = \mathrm{cl}(\mathrm{int}S)$;
(b) $f(x)$ is strictly convex and has bounded level sets.
(c) a number $\alpha \in \mathbb{R}$ is known such that $f(x) < \alpha\ \forall x \in S$.

Let P_0 be a polytope containing S. If we denote $\eta = \min\{f(x)|\ x \in S\}$, $G = \{x \in P_0|\ f(x) \le \eta\}$, then G is a closed convex set and the problem is to find a point of $\Omega := S \cap G$. Applying the outer approximation method to this problem, we construct a sequence of polytopes $P_1 \supset P_2 \supset \ldots \supset G$. At iteration k let $\overline{x}^k$ be the incumbent feasible point and $z^k \in \mathrm{argmax}\{f(x)|\ x \in P_k\}$. The distinguished point associated with P_k is defined to be $x^k = \pi(z^k)$. Clearly, if $x^k \in G$, then x^k is an optimal solution. If $f(x^k) < f(\overline{x}^k)$, then $\overline{x}^{k+1} = x^k$; otherwise $\overline{x}^{k+1} = \overline{x}^k$. Furthermore, $P_{k+1} = P_k \cap \{x|\ l(x) \le 0\}$ where $l(x) \le 0$ is a cut which separates x^k from the closed convex set G. We can thus state the following OA algorithm for solving the problem (8.95):

Optimal Visible Point Algorithm

Initialization. Take a polytope (simplex or rectangle) P_1 known to contain one optimal solution. Let $\overline{x}^1$ be the best feasible solution available, $\gamma_1 = f(\overline{x}^1)$ (if no feasible solution is known, set $\overline{x}^1 = \emptyset, \gamma_1 = \alpha$). Let V_1 be the vertex set of P_1. Set $k = 1$.

Step 1. If $\overline{x}^k \in \mathrm{argmin}\{f(x)|\ x \in P_k\}$, then terminate: $\overline{x}^k$ is an optimal solution.

Step 2. Compute $z^k \in \mathrm{argmax}\{f(x)|\ x \in V_k\}$.

Step 3. Find $\pi(z^k)$. If $\pi(z^k)$ exists and $f(\pi(z^k)) < f(\overline{x}^k)$, set $\overline{x}^{k+1} = \pi(z^k)$ and $\gamma_{k+1} = f(\overline{x}^{k+1})$; otherwise set $\overline{x}^{k+1} = \overline{x}^k, \gamma_{k+1} = \gamma_k$.

Step 4. Let $y^k \in [w, z^k] \cap \{x|\ f(x) = \gamma_{k+1}\}$. Compute $p^k \in \partial f(y^k)$ and let

$$P_{k+1} = P_k \cap \{x|\ \langle p^k, x - y^k\rangle \le 0\}.$$

Step 5. Compute the vertex set V_{k+1} of P_{k+1}. Set $k \leftarrow k+1$ and go back to Step 1.

Figure 8.3 illustrates the algorithm on a simple example in $\mathbb{R}^2$ where the feasible domain consists of two disjoint parts. Starting from a rectangle P_1 we arrive after two iterations at a polygon P_3. The feasible solution $\overline{x}^2$ obtained at the second iteration is already an optimal solution.

Theorem 8.12 *Either the above algorithm terminates by an optimal solution, or it generates an infinite sequence $\{\overline{x}^k\}$ every accumulation point of which is an optimal solution.*

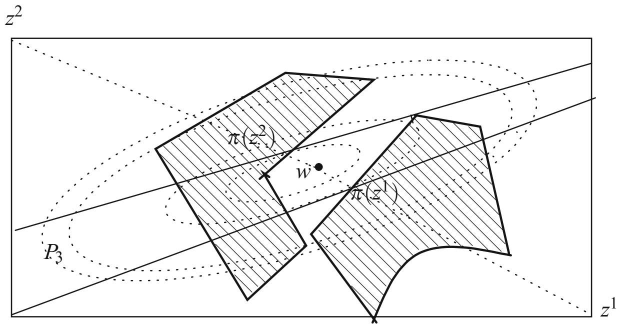


Fig. 8.3 The optimal visible point problem

Proof Let $\overline{\gamma} = \lim_{k\to+\infty}\gamma_k$, $\overline{G} = \{x|\, f(x) \le \overline{\gamma}\}$. By Theorem 6.5, $z^k - y^k \to 0$ and any accumulation point $\overline{z}$ of $\{z^k\}$ belongs to $\overline{G}$. Suppose now that there exists a feasible point x^* better than $\overline{z}$, e.g., such that $f(\overline{z}) - f(x^*) > \delta > 0$. By robustness of S, there exists a point $x' \in \mathrm{int}S$ so close to x^* that $f(x') < f(\overline{z}) - \delta$ and, consequently, a ball $W \subset S$ around x' such that $f(x) < f(\overline{z}) - \delta\ \forall x \in W$. Let $\hat{x}$ be the point where the halfline from w through x' meets $\partial\overline{G}$. Since $f(x)$ is strictly convex, its maximum over $P_1 \supset \overline{G}$ is achieved at a vertex, hence $f(\hat{x}) < f(z^1)$ and there exists in P_1 a sequence $\{x^q\} \to \hat{x}$ such that $f(x^q) \searrow f(\hat{x}) = \overline{\gamma}$. Since $f(z^k) = \max\{f(x)|\ x \in P_k\} \searrow \overline{\gamma}$, for each q there exists k_q such that $f(z^{k_q}) < f(x^q)$, hence $x^q \notin P_k$, and consequently, $\langle p^{k_q}, x^{k_q} - y^{k_q}\rangle > 0$. By passing to subsequences if necessary we may assume $z^{k_q} \to \overline{z}, y^{k_q} \to \overline{y}, p^{k_q} \to \overline{p}$. Since $z^{k_q} - y^{k_q} \to 0$, we must have $\overline{y} = \overline{z} \in \overline{G}$. Furthermore, from $p^k \in \partial f(y^k)$ it follows that $\overline{p} \in \partial f(\overline{y})$, hence $p \neq 0$ for the relation $0 \in \partial f(\overline{y})$ would imply that $\overline{y}$ is a minimizer of $f(x)$ over $\mathbb{R}^n$, which is not the case. Letting $q \to \infty$ in the inequality $\langle p^{k_q}, x^{k_q} - y^{k_q}\rangle > 0$ then yields $\langle \overline{p}, \hat{x} - \overline{y} \ge 0$. This together with the equality $f(\hat{x}) = f(\overline{y}) = \overline{\gamma}$ implies, by the strict convexity of $f(x)$, that $\hat{x} = \overline{y}$, and consequently, $z^{k_q} \to \hat{x}$. For sufficiently large q the halfline from w through z^{k_q} must then meet $W \subset S$, i.e., $\pi(z^{k_q})$ exists and $f(\pi(z^{k_q})) < f(\overline{z}) - \delta$, conflicting with $f(\pi(z^k)) \ge \overline{\gamma} = f(\overline{z})\ \forall k$. Thus, $\overline{z}$ is a global optimal solution, and hence so is any accumulation point of the sequence $\{\overline{x}^k\}$. □

Remark 8.9 Assumptions (a)–(c) are essential for the implementability and convergence of the algorithm. If $f(x)$ is not strictly convex, one could add a small perturbation $\varepsilon\|x\|^2$ to make it strictly convex, though this may not be always computationally practical. The robustness assumption can be bypassed by requiring only an ε-approximate optimal solution as in the case of (*CDC*).

Remark 8.10 Suppose $S = \{x|\ g_i(x) \le 0,\ i = 1,\dots,m\}$ and we know, for every $i \in I \subset \{1, 2, \dots, m\}$, a *concave* function $\tilde{g}_i(x)$ such that $\tilde{g}_i(x) \le g_i(x)$ for all $x \in P_1$. Then along with z^k we can compute the value

$$\varepsilon_k = \max_{i\in I}\min_{x\in P_k} \tilde{g}_i(x).$$

(Of course this involves minimizing the concave functions $\tilde{g}_i(x)$ over P_k but not much extra computational effort is needed, since we have to compute the vertex set V_k of P_k anyway). Let $g(x) = \max_{i=1,\ldots,m} g_i(x)$. Clearly,

$$\begin{aligned}\varepsilon_k \le \max_{i=1,\ldots,m} \min_{x\in P_k} g_i(x) &\le \min_{x\in P_k} \max_{i=1,\ldots,m} g_i(x)\\ &= \min_{x\in P_k} g(x).\end{aligned} \tag{8.97}$$

But it is easy to see that $\{x \in P_1 | \, f(x) \le f(\overline{x}^k)\} \subset P_k$. Indeed, the inclusion is obvious for $k = 1$; assuming that it holds for some $k \ge 1$, we have, for any $x \in P_k$ satisfying $f(x) \le f(\overline{x}^{k+1})$, $\langle p^k, x - y^k\rangle \le f(x) - f(y^k) = f(x) - f(\overline{x}^{k+1}) \le 0$, hence $x \in P_{k+1}$, so that the inclusion also holds for $k + 1$. From the just established inclusion and (8.97) we deduce

$$\min\{g(x)| \, x \in P_1, f(x) \le f(\overline{x}^k)\} \ge \varepsilon_k,$$

implying that there is no $x \in P_1$ such that $g(x) < \varepsilon_k$, $f(x) \le f(\overline{x}^k)$. In other words

$$f(\overline{x}^k) \le \inf\{f(x)| \, g(x) < \varepsilon_k\},$$

so $\overline{x}^k$ is an ε_k-approximate optimal solution.

8.6 Continuous Optimization Problems

The most challenging of all nonconvex optimization problems is the one in which no convexity or reverse convexity is explicitly present, namely:

$$(COP) \qquad\qquad \min\{f(x)| \quad x \in S\},$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a continuous function and S is a compact set in $\mathbb{R}^n$ given by a system of continuous inequalities of the form

$$g_i(x) \le 0, \; i = 1, \ldots, m. \tag{8.98}$$

Despite its difficulty, this problem started to be investigated by a number of authors from the early seventies (Dixon and Szego (1975, 1978) and the references therein). However, most methods in this period dealt with unconstrained minimization of smooth or Lipschitzian functions and were able to handle only problems of just one or two dimensions. Attempts to solve constrained problems of higher dimensions by deterministic methods have begun only in recent years.

As shown in Chap. 4 the core of (COP) is the subproblem of *transcending an incumbent*:

(SPα) Given a number $\alpha \in f(S)$ (the best feasible value of the objective function known at a given stage), find a point $x \in S$ with $f(x) < \alpha$, or else prove that $\alpha = \min\{f(x)| \, x \in S\}$.

In the so-called *modified function approaches*, the subproblem (SPα) for $\alpha = f(\bar{x})$ is solved by replacing the original function with a properly modified function such that a local search procedure, started from $\bar{x}$ and applied to the modified function will lead to a better feasible solution than $\bar{x}$, when $\bar{x}$ is not yet a global minimizer. However, a modified function satisfying the required conditions is very difficult to construct. In its two best known versions, this modified function (*tunneling function* of Levy and Montalvo (1988), or *filled function* of Ge and Qin (1987)) depends on parameters whose correct values, in many cases, can be determined only by trial and error.

The following approach to (*COP*) Thach and Tuy (1990) is based on the transformation of (SPα) into the minimization of a dc function $\varphi(\alpha, x)$, called *a relief indicator*.

Assume that the problem is *regular*, i.e.,

$$\min\{f(x)|\, x \in S\} = \inf\{f(x)|\, x \in \text{int}S\}. \tag{8.99}$$

Since $f(x)$ is continuous this assumption is satisfied provided the constraint set S is *robust* :

$$S = \text{cl}(\text{int}S),$$

whereas usual int and cl denote the interior and the closure, respectively. For $\alpha \in (-\infty, +\infty]$ define

$$F(\alpha, x) = \sup\{f(x) - \alpha, g_1(x), \ldots, g_m(x)\}. \tag{8.100}$$

Proposition 8.5 *Under the assumption (8.99) there holds*

$$\alpha = \min_{x\in S} f(x) \Leftrightarrow 0 = \min_{x\in\mathbb{R}^n} F(\alpha, x).^{1} \tag{8.101}$$

Proof If $\alpha = \min_{x\in S} f(x)$ and $\bar{x} \in S$ satisfies $f(\bar{x}) = \alpha$, then $g_i(\bar{x}) \le 0$ for $i = 1, \ldots, m$, hence $F(\alpha, \bar{x}) = 0$; on the other hand, $f(x) \ge \alpha\ \forall x \in S$, hence $F(\alpha, x) \ge 0\ \forall x \in \mathbb{R}^n$, i.e., $\min_{x\in\mathbb{R}^n} F(\alpha, x) = 0$. Conversely, if the latter equality holds then for all $x \in \text{int}S$, i.e., for all x satisfying $g_i(x) < 0,\ i = 1, \ldots, m$, we have $f(x) \ge \alpha$, hence $\min_{x\in S} f(x) \ge \alpha$ in view of (8.99). Furthermore, there is $\bar{x}$ satisfying $F(\alpha, \bar{x}) = 0$, hence $g_i(\bar{x}) \le 0,\ i = 1, \ldots, m$ (so $\bar{x} \in S$) and $f(\bar{x}) = \alpha$, i.e., $\alpha = \min_{x\in S} f(x)$. □

Thus, solving (*COP*) can be reduced, essentially, to finding the unconstrained minimum of the function $F(\alpha, x)$: if this minimum is zero, and $F(\alpha, \bar{x}) = 0$ then $\bar{x}$ solves (*COP*); if it is negative, then any x such that $F(\alpha, x) < 0$, will be a feasible point satisfying $f(x) < \alpha$; finally, if it is positive then $S = \emptyset$, i.e., the problem is infeasible.

[1] If $f, g_1, \ldots, g_m$ are convex, condition (8.99) is implied by $\text{int}S \neq \emptyset$ and by writing (8.101) in the form $0 \in \partial F(\alpha, \bar{x})$ we recover the classical optimality criterion in convex programming.

However, finding the global minimum of $F(\alpha, x)$ may be as hard as solving the original problem. Therefore, we look for some function $\varphi(\alpha, x)$ which could do essentially the same job as $F(\alpha, x)$ while being easier to handle. The key for this is the representation of closed sets by dc inequalities as discussed in Chap. 3.

Define

$$\tilde{S}_\alpha = \{x \in \mathbb{R}^n |\, F(\alpha, x) \leq 0\} = \{x \in S |\, f(x) \leq \alpha\}. \tag{8.102}$$

To make the problem tractable, assume that for every $\alpha \in (-\infty, +\infty]$ a function $r(\alpha, x)$ (*separator* for (f, S)) is available which is l.s.c. in x and satisfies:

(a) $r(\alpha, x) = 0 \quad \forall x \in \tilde{S}_\alpha$;
(b) $0 < r(\alpha, y) \leq d(y, \tilde{S}_\alpha) := \min\{\|x - y\| \,|\, x \in \tilde{S}_\alpha\} \quad \forall y \notin \tilde{S}_\alpha$;
(c) For fixed x, $r(\alpha_1, x) \geq r(\alpha_2, x)$ if $\alpha_1 < \alpha_2$.

This assumption is fulfilled in many cases of interest, as shown in the following examples:

Example 8.3 If all functions $f, g_1, \ldots, g_m$ are Lipschitzian with constants $L_0, L_1, \ldots, L_m$ then a separator is

$$r(\alpha, x) = \max\left\{0, \frac{f(x) - \alpha}{L_0}, \frac{g_i(x)}{L_i}, i = 1, \ldots, m\right\}.$$

Indeed, for any $x \in \tilde{S}_\alpha$:

$$f(y) - \alpha \leq f(y) - f(x) \leq L_0\|x - y\|$$
$$g_i(y) \leq g_i(y) - g_i(x) \leq L_i\|x - y\| \quad i = 1, \ldots, m$$

hence $r(\alpha, y) \leq \|x - y\|$, from this it is easy to verify (a)–(c).

Example 8.4 If f is twice continuously differentiable, with $\|f''(x)\| \leq K$ for every x, and $S = \{x |\, \|x\| \leq c\}$ then a separator is

$$r(\alpha, x) = \max\{\rho(\alpha, x), \|x\| - c\},$$

where $\rho(\alpha, x)$ denotes the unique nonnegative root of the equation

$$\frac{K}{2}t^2 + \|f''(x)\|t = \max\{0, f(x) - \alpha\}.$$

Indeed, we need only verify condition (b) since the other conditions are obvious. For any $x \in \tilde{S}_\alpha$, we have by Taylor' s formula:

$$|f(x) - f(y) - f'(y)(x - y)| \leq \frac{M}{2}\|x - y\|^2,$$

hence

$$|f(x) - f(y)| \leq \|f'(y)\|\|x - y\| + \frac{M}{2}\|x - y\|^2$$
$$f(x) \geq f(y) - \frac{M}{2}\|x - y\|^2 - \|f'(y)\|\|x - y\|.$$

Now $\frac{M}{2}t^2 + \|f'(y)\|t$ is a monotonic increasing function of t. Therefore, since $f(x) \leq \alpha$ implies that

$$\frac{M}{2}\|x-y\|^2 + \|f'(y)\|\,\|x-y\| \geq f(y) - \alpha,$$

it follows from the definition of $\rho(\alpha, y)$ that $\rho(\alpha, y) \leq \|x-y\|$. On the other hand, since $\|x\| \leq c$, we have $\|y\| - c \leq \|y\| - \|x\| \leq \|x-y\|$. Hence, $r(\alpha, y) \leq \|x-y\|\ \forall x \in \tilde{S}_\alpha$, proving (b).

By Proposition 4.13, given a separator $r(\alpha, x)$, if we define

$$h(\alpha, x) = \sup_{y \notin \tilde{S}_\alpha} \{r^2(\alpha, y) + 2xy - \|y\|^2\}$$

then $h(\alpha, x)$ is a closed convex function in x such that

$$\tilde{S}_\alpha = \{x \mid h(\alpha, x) - \|x\|^2 \leq 0\}. \tag{8.103}$$

Since for any $y \in S$ satisfying $f(y) = \alpha$ we have $r(\alpha, y) = 0$, hence $r^2(\alpha, y) + 2xy - \|y\|^2 - \|x\|^2 = -\|x-y\|^2 \leq 0 \quad \forall x$, the representation (8.103) still holds with

$$h(\alpha, x) = \sup_{y \notin S_\alpha} \{r^2(\alpha, y) + 2xy - \|y\|^2\} \tag{8.104}$$

where

$$S_\alpha = \{x \in S \mid f(x) < \alpha\}. \tag{8.105}$$

Proposition 8.6 *Assume $S \neq \emptyset$ and let $\varphi(\alpha, x) = h(\alpha, x) - \|x\|^2$. If*

$$\varphi(\alpha, \overline{x}) = 0 = \min_{x \in \mathbb{R}^n} \varphi(\alpha, x). \tag{8.106}$$

then $\overline{x}$ is a global optimal solution of (COP). Otherwise, the minimum of $\varphi(\alpha, x)$ is negative and any x such that $\varphi(\alpha, x) < 0$ satisfies $x \in S$, $f(x) < \alpha$.

Proof Since $F(\alpha, x) \leq 0 \Leftrightarrow \varphi(\alpha, x) \leq 0$ (see (8.102) and (8.103)), it suffices to show further that $F(\alpha, x) < 0 \Leftrightarrow \varphi(\alpha, x) < 0$. But if $F(\alpha, x) < 0$ then $x \in \mathrm{int}S_\alpha$ and denoting by $\delta > 0$ the radius of a ball around x contained in S_α, we have for every $y \notin S_\alpha : d^2(y, S_\alpha) \leq \|x-y\|^2 - \delta^2$, hence $r^2(\alpha, y) - \|x-y\|^2 \leq -\delta^2$, i.e., $\varphi(\alpha, x) \leq -\delta^2 < 0$. Conversely, if $\varphi(\alpha, x) < 0$ then for every $y \notin S_\alpha : r^2(\alpha, y) + 2xy - \|y\|^2 < \|x\|^2 - \theta$, with $\theta > 0$, hence $r^2(\alpha, y) < \|x-y\|^2 - \theta$ and we must have $F(\alpha, x) < 0$ (for otherwise $x \notin \mathrm{int}S_\alpha$ and by taking a sequence $y^n \to x$, such that $y^n \notin S_\alpha\ \forall n$ and letting $n \to \infty$ in the inequality $r^2(\alpha, y^n) < \|x - y^n\|^2 - \theta$ we would arrive at a contradiction: $0 \leq -\theta$). $\square$

The function $\varphi(\alpha, x)$ is called a *relief indicator*. By Proposition 8.6 it gives information about the "altitude" $f(x)$ of every point x relative to the level α and can be used as a substitute for $F(\alpha, x)$.

The knowledge of a relief indicator $\varphi(\alpha, x)$ reduces each subproblem $SP(\alpha)$ to an unconstrained dc optimization problem, namely:

$$\text{minimize } h(\alpha, x) - \|x\|^2 \quad s.t. \quad x \in \mathbb{R}^n. \tag{8.107}$$

We shall shortly discuss an algorithm for solving this dc optimization problem. Assuming such an algorithm available, the global optimum of (COP) can be computed by the following iterative scheme:

Relief Indicator Method
Start with an initial value $\alpha_1 \in f(S)$, if available, or $\alpha_1 = +\infty$ if no feasible solution is known. At iteration $k = 1, 2, \ldots$ solve

$$\min\{\varphi(\alpha_k, x) |\, x \in \mathbb{R}^n\}, \tag{8.108}$$

to obtain its optimal solution x^k and optimal value μ_k. If $\mu_k \geq 0$ then terminate: x^k is an optimal solution of (COP) (if $\mu_k = 0$), or (COP) is infeasible (if $\mu_k > 0$). Otherwise, x^k is a feasible solution such that $f(x^k) < \alpha_k$: then set $\alpha_{k+1} = f(x^k)$ and go to the next iteration.

The convergence of this iterative scheme is immediate. Indeed, if $\overline{x}$ is a accumulation point of the sequence $\{x^k\}$ and $\overline{\alpha} = f(\overline{x})$, then $\overline{x} \in S$ and $\varphi(\overline{\alpha}, \overline{x}) \leq 0$. On the other hand, since $x^k \notin S_{\overline{\alpha}}$ it follows that $\varphi(\overline{\alpha}, \overline{x}) \geq r^2(\overline{\alpha}, x^k) + 2\overline{x}x^k - \|x^k\|^2 - \|\overline{x}\|^2 \geq -\|x^k - \overline{x}\|^2 \to 0$ as $x^k \to \overline{x}$. Hence, $\varphi(\overline{\alpha}, \overline{x}) = 0$, so by Proposition 8.6 $\overline{x}$ is a global optimal solution of (COP). □

8.6.1 Outer Approximation Algorithm

The implementation of the above conceptual scheme requires the availability of a procedure for solving subproblems of the form (8.108). One possibility is to rewrite (8.108) as a concave minimization problem:

$$\min\{t - \|x\|^2 |\, h(\alpha_k, x) \leq t\}, \tag{8.109}$$

and to solve the latter by the simple OA algorithm (Sect. 6.2). Since $\alpha_{k+1} < \alpha_k$ we have $h(\alpha_k, x) \leq h(\alpha_{k+1}, x)$, so if G_k denotes the constraint set of (8.109) then $G_{k+1} \subset G_k \quad \forall k$. This allows the procedures of solving (8.109) for successive $k = 1, 2, \ldots$ to be integrated into a unified OA procedure for solving the single concave minimization problem

$$\min\{t - \|x\|^2 | \quad h(\overline{\alpha}, x) \leq t\}, \tag{8.110}$$

where $\overline{\alpha}$ is the (unknown) optimal value of (COP).

Specifically, denote by G the constraint set of (8.110). Let P_k be a polytope outer approximating $G_k \supset G$ and let (x^k, t^k) be an optimal solution of the relaxed problem

$$\min\{t - \|x\|^2 |\ x \in P_k\}.$$

If $t^k - \|x^k\|^2 \geq 0$ then the optimal value of (8.110) is $\mu \geq t^k - \|x^k\|^2 \geq 0$ and there are two possibilities:

(1) if $\alpha_k = +\infty$, then (COP) is infeasible (otherwise, one would have $\mu < 0$, conflicting with $0 \leq t^k - \|x^k\|^2 \leq \mu$);
(2) if $\alpha_k < +\infty$, then x^k is an optimal solution of (COP) (because $0 \leq t^k - \|x^k\|^2 \leq \mu \leq 0$, hence $\mu = 0$).

So the only case that needs further investigation is when $t^k - \|x^k\|^2 < 0$. Define

$$\alpha_{k+1} = \begin{cases} \min\{\alpha_k, f(x^k)\} & \text{if } x^k \in S \\ \alpha_k & \text{otherwise} \end{cases} \tag{8.111}$$

Lemma 8.4 *The affine function*

$$l_k(x) := 2\langle x^k, x\rangle - \|x^k\|^2 + r^2(\alpha_{k+1}, x^k)$$

strictly separates (x^k, t^k) *from* G_{k+1}, *i.e., satisfies*

$$l_k(x^k) > t^k, \quad l_k(x) - t \leq 0\ \forall x \in G_{k+1}.$$

Proof Since $t^k < \|x^k\|^2$ we have $l_k(x^k) - t^k > l_k(x^k) - \|x^k\|^2 = r^2(\alpha_{k+1}, x^k) \geq 0$. On the other hand, $x \in G_{k+1}$ implies $h(\alpha_{k+1}, x) - t \leq 0$, but since $x^k \notin S_{\alpha_{k+1}}$ it follows from (8.104) that $l_k(x) \leq h(\alpha_{k+1}, x)$, and hence $l_k(x) - t \leq h(\alpha_{k+1}, x) - t \leq 0$. □

In view of this Lemma if we define $P_{k+1} = P_k \cap \{x |\ l_k(x) \leq t\}$ then $(x^k, t^k) \notin P_{k+1}$ while $P_{k+1} \supset G_{k+1} \supset G$, so the sequence $P_1 \supset P_2 \supset \cdots \supset G$ realizes an OA scheme for solving (8.110). We are thus led to the following:

Relief Indicator OA Algorithm for (COP)
Initialization. Take a simplex or rectangle $P_1 \subset \mathbb{R}^n$ known to contain an optimal solution of the problem. Let x^1 be a point in P_1 and let $\alpha_1 = f(x^1)$ if $x^1 \in S$, or $\alpha_1 = +\infty$ otherwise. Set $k = 1$.

Step 1. If $x^k \in S$, then improve x^k by local search methods (provided this can be done at reasonable cost). Denote by $\overline{x}^k$ the best feasible solution so far obtained. If $\overline{x}^k$ exists, let $\alpha_{k+1} = f(\overline{x}^k)$, otherwise let $\alpha_{k+1} = +\infty$. Define

$$l_k(x) = 2\langle x^k, x\rangle - \|x^k\|^2 + r^2(\alpha_{k+1}, x^k).$$

Step 2. Solve

$$(SP_k) \qquad \min\{t - \|x\|^2 |\ x \in P_1,\ l_i(x) \leq t,\ i = 1, \ldots, k\}$$

to obtain (x^{k+1}, t^{k+1}).

Step 3. If $t^{k+1} - \|x^{k+1}\|^2 > 0$, then terminate: $S = \emptyset$ ((COP) is infeasible).
If $t^{k+1} - \|x^{k+1}\|^2 = 0$, then terminate ($S = \emptyset$ if $\alpha_{k+1} = +\infty$; $\overline{x}^k$ solves (COP) if $\alpha_{k+1} < +\infty$).

Step 4. If $t^{k+1} - \|x^{k+1}\|^2 < 0$, then set $k \leftarrow k + 1$ and go back to Step 1.

Lemma 8.5 *After finitely many steps the above algorithm either gives evidence that (COP) is infeasible or generates a feasible solution, if there is one.*

Proof Let $\tilde{x} \in \mathrm{int} S$ (see assumption (8.99)), so $\varphi(\alpha, \tilde{x}) < 0$ for $\alpha = +\infty$. If $x^k \notin S\ \forall k$, then $\alpha_k = +\infty$, $t^k < \|x^k\|^2\ \forall k$, and

$$t^k - \|x^k\|^2 \le \min_{x\in P_1} \varphi(\alpha_{k-1}, x) \le \varphi(\alpha_{k-1}, \tilde{x}), \quad \forall k$$

while for all $i < k$:

$$t^k - \|x^k\|^2 \ge l_i(x^k) - \|x^k\|^2 \ge -\|x^i - x^k\|^2.$$

Thus, $\|x^i - x^k\|^2 \ge -\varphi(\alpha_k, \tilde{x}) > 0 \quad \forall i < k$, conflicting with the boundedness of $\{x^k\} \subset P_1$. Therefore, $x^k \in S$ for all sufficiently large k. In particular, any accumulation point of the sequence $\{x^k\}$ is feasible. □

Proposition 8.7 *Whenever infinite the Relief Indicator Algorithm generates an infinite sequence $\{x^k\}$ every accumulation point of which is an optimal solution of (COP).*

Proof Let $\overline{x} = \lim_{s\to+\infty} x^{k_s}$, (as we just saw, $\overline{x} \in S$). Define

$$\tilde{l}_k(x) = l_k(x) - \|x\|^2.$$

For any $q > s$ we have $\tilde{l}_{k_s}(x^{k_q}) = l_{k_s}(x^{k_q}) - \|x^{k_q}\|^2 \le t^{k_q} - \|x^{k_q}\|^2 < 0$, hence, by making $q \to +\infty$: $\tilde{l}_{k_s}(\overline{x}) \le 0$. But clearly, $\tilde{l}_{k_s}(x^{k_s}) = \tilde{l}_{k_s}(\overline{x}) - \|x^{k_s} - \overline{x}\|^2 \le -\|x^{k_s} - \overline{x}\|^2 \to 0$ as $s \to +\infty$, while $\tilde{l}_k(x^k) = r^2(\alpha_{k+1}, x^k) > 0$. Therefore, $\tilde{l}_{k_s}(x^{k_s}) \to 0$ and also $t^{k_s} - \|x^{k_s}\|^2 \to 0$, as $s \to +\infty$. If $\overline{t} = \lim t^{k_s}$ then $\overline{t} - \|\overline{x}\|^2 = 0$. Since $t^k - \|x^k\|^2 = \min\{t - \|x\|^2 |\, x \in P_1, l_i(x) \le t,\ i = 1, 2, \ldots, k-1\} \le \min\{t - \|x\|^2 |\, x \in P_1, h(\alpha_k, x) \le t\} \quad \forall k$, it follows that $\overline{t} - \|\overline{x}\|^2 = \min\{t - \|x\|^2 |\, x \in P_1, h(\overline{\alpha}, x) \le t\}$ for $\alpha = \lim \alpha_k$. That is, $\varphi(\overline{\alpha}, \overline{x}) = 0$, and hence, $\overline{x}$ solves (COP). □

Remark 8.11 Each subproblem (SP_k) is a linearly constrained concave minimization problem which can be solved by vertex enumeration as in simple outer approximation algorithms. Since (SP_{k+1}) differs from (SP_k) by just one additional linear constraint, if V_k denotes the vertex set of the constraint polytope of (SP_k) then V_1 can be considered known, while V_{k+1} can be derived from V_k using an on-line vertex enumeration subroutine (Sect. 6.2). If the algorithm is stopped when $t^k - \|x^k\|^2 > -\varepsilon^2$ then $\overline{x}^k$ yields an ε-approximate optimal solution.

Example 8.5 Solve the problem

$$\begin{aligned} &\text{Minimize} \quad f(x) = x_1^2 + x_2^2 - \cos 18x_1 - \cos 18x_2 \\ &\text{subject to } g(x) = -[(x_1 - 0.5)^2 + (x_2 - 0.415331)^2]^{\frac{1}{2}} + 0.65 \le 0 \\ &\qquad\qquad 0 \le x_1 \le 1,\ 0 \le x_2 \le 1. \end{aligned}$$

The functions f and g are Lipshitzian on $S = [0, 1] \times [0, 1]$, with Lipschitz constants 28.3 and 1, respectively. Following Example 6.2 a separator is

$$r(\alpha, x) = \max\{0, \frac{f(x) - \alpha}{28.3}, g(x)\}.$$

With tolerance $\varepsilon = -0.01$ the algorithm terminates after 16 iterations, yielding an approximate optimal solution $\overline{x} = (0, 1)$ with objective function value -0.6603168.

8.6.2 Combined OA/BB Algorithm

An alternative method for solving (8.110) is by combining outer approximation with branch and bound. This gives rise to a relief indicator OA/BB algorithm (Tuy 1992b, 1995a) for problems of the following form:

$$(CO/SNP) \qquad \min\{f(x) |\ G(x, y) \le 0,\ g_i(x) \le 0,\ i = 1, \ldots, m\},$$

where $f, g_1, \ldots, g_m$ are as previously, while $G : \mathbb{R}^n \times \mathbb{R}^q \to \mathbb{R}$ is a convex function. To take advantage of the convexity of G in this problem it is convenient to treat the constraint $G(x, y) \le 0$ separately. Assuming robustness of the constraint set of (CO/SNP), and the availability of a separator $r(\alpha, x)$ for (f, S) (with S being the set determined by the system $g_i(x) \le 0,\ i = 1, \ldots, m)$ we can transform (CO/SNP) into the concave minimization problem

$$\min\{t - \|x\|^2 |\ G(x, y) \le 0,\ h(\overline{\alpha}, x) \le t\}, \tag{8.112}$$

where $h(\alpha, x)$ is defined as in (8.104) and $\overline{\alpha}$ is the (unknown) optimal value of $f(x)$. Since for every fixed x this is a convex problem, the most convenient way is to branch upon x and compute bounds by using relaxed subproblems of the form

$$\begin{aligned} LP_k(M) \qquad \min\{t - \psi_M(x) |\ x \in M,\ l_i(x) \le t\ (i \in I_k) \\ L_j(x, y) \le 0\ (j \in J_k)\} \end{aligned}$$

where M is the partition set, $\psi_M(x)$ is the affine function that agrees with $\|x\|^2$ at the vertices of M, $\{(x, t) |\ l_i(x) \le t\ (i \in I_k)\}$ the outer approximating polytope of the convex set $C_{\overline{\alpha}} := \{(x, t) |\ h(\overline{\alpha}, x) \le t\}$ and $\{(x, y) |\ L_j(x, y) \le 0\ (j \in J_k)\}$ the outer approximating polytope of the convex set $\Omega = \{(x, y) |\ G(x, y) \le 0\}$, at iteration k.

BB/OA Algorithm for (CO/SNP)

Initialization. Take an n-simplex M_1 in $\mathbb{R}^n$ known to contain an optimal solution of (CO/SNP). If a feasible solution is available then let $(\overline{x}^0, \overline{y}^0)$ be the best one, $\alpha_0 = f(\overline{x}^0)$; otherwise, let $\alpha_0 = +\infty$. Set $I_1 = J_1 = \emptyset$, $\mathscr{R} = \mathscr{N} = \{M_1\}$. Set $k = 1$.

Step 1. For each $M \in \mathscr{N}$ compute the optimal value $\beta(M)$ of the linear program $\mathrm{LP}_k(M)$.

Step 2. Update $(\overline{x}^k, \overline{y}^k)$ and α_k. Delete every $M \in \mathscr{R}$ such that $\beta(M) > 0$. Let $\mathscr{R}'$ be the remaining collection of simplices. If $\mathscr{R}' = \emptyset$ then terminate: (CO/SNP) is infeasible.

Step 3. Select $M_k \in \mathrm{argmin}\{\beta(M) | \ M \in \mathscr{R}'\}$. Let (x^k, y^k, t^k) be a basic optimal solution of $\mathrm{LP}_k(M_k)$. If $\beta(M_k) = 0$ then terminate: (CO/SNP) is infeasible (if $\alpha_k = +\infty$), or $(\overline{x}^k, \overline{y}^k)$ solves (CO/SNP) (if $\alpha_k = +\infty$).

Step 4. If $\|x^k\|^2 \le t^k$ and $G(x^k, y^k) \le 0$ then set $I_{k+1} = I_k$, $J_{k+1} = J_k$. Otherwise:

(a) if $\|x^k\|^2 - t^k > G(x^k, y^k)$ then set $J_{k+1} = J_k$,

$$I_{k+1} = I_k \cup \{k\}, l_k(x) = 2\langle x^k, x\rangle - \|x^k\|^2 + r^2(\alpha_k, x^k);$$

(b) if $\|x^k\|^2 - t^k \le G(x^k, y^k)$ then set $I_{k+1} = I_k$,

$$J_{k+1} = J_k \cup \{k\},\ L_k(x) = \langle u^k, x - x^k\rangle + \langle v^k, y - y^k\rangle + G(x^k, y^k),$$

where $(u^k, v^k) \in \partial G(x^k, y^k)$.

Step 5. Perform a bisection or an ω-subdivision (subdivision via $\omega^k = x^k$) of M_k, according to a normal subdivision rule.
Let $\mathscr{N}^*$ be the partition of M_k. Set $\mathscr{S} \leftarrow \mathscr{N}^*, \mathscr{R} \leftarrow \mathscr{N}^* \cup (\mathscr{R}' \setminus \{M_k\})$, increase k by 1 and return to *Step 1.*

The convergence of this algorithm follows from that of the general algorithm for (*COP*).

8.7 Exercises

1 Solve by polyhedral annexation:

$$\min\ -(x_1-1)^2 - x_2^2 - (x_3-1)^2 \quad \text{s.t.}$$

$$4x_1 - 5x_2 + 4x_3 \ge 4$$

$$6x_1 - x_2 - x_3 \ge 4.1$$

$$x_1 + x_2 - x_3 \le 1$$

$$12x_1 + 5x_2 + 12x_3 \leq 34.8$$
$$12x_1 + 12x_2 + 7x_3 \leq 29.1$$
$$x_1, x_2, x_3 \geq 0$$

(optimal solution $x^* = (1, 0, 0)$)

2 Consider the problem:

$$\min\{cx|\ x \in \Omega, g(x) \leq 0, h(x) \leq 0, \langle g(x), h(x)\rangle = 0\}$$

where $g, h : \mathbb{R}^n \to \mathbb{R}^n$ are two convex maps and Ω a convex subset of $\mathbb{R}^n$ (convex program with an additional complementarity condition). Show that this problem can be converted into a (GCP) (a map $g : \mathbb{R}^n \to \mathbb{R}^m$ is said to be convex if its components $g_i(x),\ i = 1, \ldots, m$ are all convex functions)

3 Show that an ε-optimal solution of (LRC) can be computed according to the following scheme (Nghia and Hieu 1986):

Initialization

Solve the linear program

$$\beta_1 = \min\{cx|\ x \in D\}.$$

Let w be a basic optimal solution of this program. If $h(w) \geq 0$, terminate: w solves (LRC). Otherwise, set $\alpha_1 = +\infty, k = 1$.

Iteration k

(1) If $\alpha_k - \beta_k \leq \varepsilon$, terminates: $\overline{x}$ is an ε-optimal solution.
(2) Solve

$$\eta_k = \max\left\{h(z)|\ z \in D,\ cz \leq \frac{\alpha_k + \beta_k}{2}\right\}$$

to obtain z^k.
(3) If $\eta_k < 0$, and $\alpha_k = \infty$, terminate: the problem is infeasible.
(4) If $\eta_k < 0$, and $\alpha_k < \infty$, set $\alpha_{k+1} = \alpha_k,\ \beta_{k+1} = \dfrac{\alpha_k + \beta_k}{2}$ and go to iteration $k + 1$.
(5) If $\eta_k \geq 0$, then starting from z^k perform a number of simplex pivots to move to a point x^k such that $cx^k \leq cz^k$ and x^k is on the intersection of an edge of D with the surface $h(x) = 0$ (see Phase 1 of FW/BW Procedure). Set $\overline{x} = x^k, \alpha_{k+1} = cx^k, \beta_{k+1} = \beta_k$ and go to iteration $k + 1$.

4 Solve $\min\{-y^3 + 4.5y^2 - 6y|\quad 0 \leq y \leq 3\}$.

5 Solve

$$\begin{aligned}
&\min\ 4x_1^4 + 2x_2^2 - 4x_1^2 :\\
&x_1^2 - 2x_1 - 2x_2 - 1 \le 0\\
&-1 \le x_1 \le 1; \quad -1 \le x_2 \le 1.
\end{aligned}$$

(optimal solution $x^* = (0,707, 0.000)$).

6 Solve

$$\min\{-2x_1 + x_2 | \quad x_1^2 + x_2^2 \le 1,\ x_1 - x_2^2 \le 0\}.$$

(optimal solution $x^* = (0.618034, -0.786151)$)

7 Solve

$$\min\{x_1^2 + x_2^2 + x_1x_2 + x_1 + x_2 | \quad x_1^2 + x_2^2 \ge 0.5;\ x_1, x_2 \ge 0\}.$$

8 Solve

$$\begin{aligned}
&\min\ 2x_1^2 + x_1x_2 - 10x_1 + 2x_2^2 - 20 :\\
&x_1 + x_2 \le 11; \quad 1.5x_1 + x_2 \le 11.4; \quad x_1, x_2 \ge 0
\end{aligned}$$

9 Solve

$$\min\left\{-x_1 + \frac{1}{2}x_2 + \frac{2}{3}x_2^2 - \frac{1}{2}x_1^2 | \quad 2x_1 - x_2 \le 3;\ x_1, x_2 \ge 0\right\}.$$

10 Solve the previous problem by reduction to quasiconcave minimization via dualization (see Sect. 8.3).

11 Let C and D two compact convex sets in $\mathbb{R}^n$ containing 0 in their interiors. Consider the pair of dual problems

$$\begin{aligned}
&(P) \qquad\qquad \max\nolimits_{x,r}\{r | \ x + rC \subset D\}\\
&(Q) \qquad\qquad \min\nolimits_{x,r}\{r | \ C \subset x + rD\}.
\end{aligned}$$

Show that 1) $(\overline{x}, \overline{r})$ solves (P) if and only if $(-\overline{x}/\overline{r}, 1/\overline{r})$ solves Q; 2) When D is a polytope, problem (P) merely reduces to a linear program; 3) When C is a polytope, problem (Q) reduces to minimizing a dc function.

12 Solve $\min\{e^{-x^2}\sin x - |y| \mid \ 0 \le x \le 10,\ 0 \le y \le 10\}$. (Hint: Lipschitz constant =2)

Chapter 9
Parametric Decomposition

A partly convex problem as formulated in Sect. 8.2 of the preceding chapter can also be viewed as a convex optimization problem depending upon a parameter $x \in \mathbb{R}^n$. The dimension n of the parameter space is referred to as the *nonconvexity rank* of the problem. Roughly speaking, this is the number of "nonconvex variables" in the problem. As we saw in Sect. 8.2, a partly convex problem of rank n can be efficiently solved by a BB decomposition algorithm with branching performed in $\mathbb{R}^n$. This chapter discusses decomposition methods for partly convex problems with small nonconvexity rank n. It turns out that these problems can be solved by streamlined decomposition methods based on parametric programming.

9.1 Functions Monotonic w.r.t. a Convex Cone

Let X be a convex set in $\mathbb{R}^n$ and K a closed convex cone in $\mathbb{R}^n$. A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *monotonic* on X with respect to K (or K*-monotonic* on X for short) if

$$x, x' \in X,\ x' - x \in K \quad \Rightarrow \quad f(x') \geq f(x). \tag{9.1}$$

In other words, for every $x \in X$ and $u \in K$ the univariate function $\theta \mapsto f(x + \theta u)$ is nondecreasing in the interval $(0, \theta^*)$ where $\theta^* = \sup\{\theta |\ x + \theta u \in X\}$. When $X = \mathbb{R}^n$ we simply say that $f(x)$ is K-monotonic.

Let L be the lineality space of K, and Q an $r \times n$ matrix of rank r such that $L = \{x |\ Qx = 0\}$. From (9.1) it follows that

$$x, x' \in X,\ Qx = Qx' \quad \Rightarrow \quad f(x) = f(x'), \tag{9.2}$$

i.e., $f(x)$ is constant on every set $\{x \in X |\ Qx = \text{const}\}$. Denote $Q(X) = \{y \in \mathbb{R}^r |\ y = Qx, x \in X\}$.

H. Tuy, *Convex Analysis and Global Optimization*, Springer Optimization and Its Applications 110, DOI 10.1007/978-3-319-31484-6_9

Proposition 9.1 *Let $K \subset \mathrm{rec}X$. A function $f(x)$ is K-monotonic on X if and only if*

$$f(x) = F(Qx) \quad \forall x \in X, \tag{9.3}$$

where $F : Q(X) \to \mathbb{R}$ satisfies

$$y, y' \in Q(X),\ y' - y \in Q(K) \quad \Rightarrow \quad F(y') \geq F(y). \tag{9.4}$$

Proof If $f(x)$ is K-monotonic on X then we can define a function $F : Q(X) \to \mathbb{R}$ by setting, for every $y \in Q(X)$:

$$F(y) = f(x) \text{ for any } x \in X \text{ such that } y = Qx. \tag{9.5}$$

This definition is correct in view of (9.2). Obviously, $f(x) = F(Qx)$. For any $y, y' \in Q(X)$ such that $y' - y \in Q(K)$ we have $y = Qx, y' = Qx'$, with $Q(x' - x) = Qu,\ u \in K$, i.e., $Qx' = Q(x + u),\ u \in K$. Here $x + u \in X$ because $u \in \mathrm{rec}X$, hence by (9.1), $f(x') = f(x + u) \geq f(x)$, i.e., $F(y') \geq F(y)$. Conversely, if $f(x) = F(Qx)$ and $F : Q(X) \to \mathbb{R}$ satisfies (9.4), then for any $x, x' \in X$ such that $x' - x \in K$ we have $y = Qx, y' = Qx'$, with $y' - y \in Q(K)$, hence $F(y') \geq F(y)$, i.e., $f(x') \geq f(x)$. □

The above representation of K-monotonic functions explains their role in global optimization. Indeed, using this representation, any optimization problem in $\mathbb{R}^n$ of the form

$$\min\{f(x)| \quad x \in D\} \tag{9.6}$$

where $D \subset X$, and $f(x)$ is K-monotonic on X, can be rewritten as

$$\min\{F(y)| \quad y \in Q(D)\} \tag{9.7}$$

which is a problem in $\mathbb{R}^r$. Assume further, that $f(x)$ is quasiconcave on X, i.e., by (9.5), $F(y)$ is quasiconcave on $Q(X)$.

Proposition 9.2 *Let $K \subset \mathrm{rec}X$. A quasiconcave function $f : X \to \mathbb{R}$ on X is K-monotonic on X if and only if for every $x^0 \in X$ with $f(x^0) = \gamma$:*

$$x^0 + K \subset X(\gamma) := \{x \in X | f(x) \geq \gamma\}.$$

Proof If $f(x)$ is K-monotonic on X, then for any $x \in X(\gamma)$ and $u \in K$ we have $x + u \in X$ (because $K \subset \mathrm{rec}X$) and condition (9.1) implies that $f(x + u) \geq f(x) \geq \gamma$, hence $x + u \in X(\gamma)$, i.e., u is a recession direction for $X(\gamma)$. Conversely, if for any $\gamma \in f(X)$, K is contained in the recession cone of $X(\gamma)$, then for any $x', x \in X$ such that $x' - x \in K$ one has, for $\gamma = f(x) : x \in X(\gamma)$, hence $x' \in X(\gamma)$, i.e., $f(x') \geq f(x)$. □

Proposition 9.1 provides the foundation for reducing the dimension of problem (9.6), whereas Proposition 9.2 suggests methods for restricting the global search domain. Indeed, since local information is generally insufficient to verify the global optimality of a solution, the search for a global optimal solution must be carried out, in principle, over the entire feasible set. If, however, the objective function $f(x)$ in (9.6) is K-monotonic, then, once a solution $x^0 \in D$ is known, one can ignore the whole set $D \cap (x^0 + K) \subset \{x \in D|\ f(x) \geq f(x^0)\}$, because no better feasible solution than x^0 can be found in this set. Such kind of information is often very helpful and may drastically simplify the problem by limiting the global search to a restricted region of the feasible domain. In the next sections, we shall discuss how efficient algorithms for K-monotonic problems can be developed based on these observations.

Below are some important classes of quasiconcave monotonic functions encountered in the applications.

Example 9.1 $f(x) = \varphi(x_1, \ldots, x_p) + d^T x$,

with $\varphi(y_1, \ldots, y_p)$ a continuous concave function of $y = (y_1, \ldots, y_p)$.

Here $F(y) = \varphi(y_1, \ldots, y_p) + y_{p+1}, Qx = (x_1, \ldots, x_p, d^T x)^T$. Monotonicity holds with respect to the cone $K = \{u|\ u_i = 0(i = 1, \ldots, p),\ d^T u \geq 0\}$. It has long been recognized that concave minimization problems with objective functions of this form can be efficiently solved only by taking advantage of the presence of the linear part in the objective function (Rosen 1983; Tuy 1984).

Example 9.2 $f(x) = -\sum_{i=1}^{p}[\langle c^i, x\rangle]^2 + \langle c^{p+1}, x\rangle$.

Here $F(y) = -\sum_{i=1}^{p} y_i^2 + y_{p+1}, Qx = (\langle c^1, x\rangle, \ldots, \langle c^{p+1}, x\rangle)^T$. Monotonicity holds with respect to $K = \{u|\ \langle c^i, u\rangle = 0(i = 1, \ldots, p),\ \langle c^{p+1}, u\rangle \geq 0\}$. For $p = 1$ such a function $f(x)$ cannot in general be written as a product of two affine functions and finding its minimum over a polytope is a problem known to be NP-hard (Pardalos and Vavasis 1991), however, quite practically solvable by a parametric algorithm (see Sect. 9.5).

Example 9.3 $f(x) = \prod_{i=1}^{r}[\langle c^i, x\rangle + d_i]^{\alpha_i}$ with $\alpha_i > 0,\ i = 1, \ldots, r$.

This class includes functions which are products of affine functions. Since $\log f(x) = \sum_{i=1}^{r} \alpha_i \log[\langle c^i, x\rangle + d_i]$ for every x such that $\langle c^i, x\rangle + d_i > 0$, it is easily seen that $f(x)$ is quasiconcave on the set $X = \{x|\ \langle c^i, x\rangle + d_i \geq 0, i = 1, \ldots, r\}$. Furthermore, $f(x)$ is monotonic on X with respect to the cone $K = \{u|\ \langle c^i, u\rangle \geq 0,\ i = 1, \ldots, r\}$. Here $f(x)$ has the form (9.3) with $F(y) = \prod_{i=1}^{r}(y_i + d_i)^{\alpha_i}$. Problems of minimizing functions $f(x)$ of this form (with $\alpha_i = 1,\ i = 1, \ldots, r$) under linear constraints are termed linear multiplicative programs and will be discussed later in this chapter (Sect. 9.5).

Example 9.4 $f(x) = -\sum_{i=1}^{r} \theta_i e^{\langle c^i, x\rangle}$ with $\theta_i > 0, i = 1, \ldots, r$.

Here $F(y) = -\sum_{i=1}^{r} \theta_i e^{y_i}$ is concave in y, so $f(x)$ is concave in x. Monotonicity holds with respect to the cone $K = \{x|\ \langle c^i, x\rangle \leq 0,\ i = 1, \ldots, r\}$. Functions of this type appear when dealing with geometric programs with negative coefficients ("signomial programs," cf Example 5.6).

Example 9.5 $f(x)$ quasiconcave and differentiable, with

$$\nabla f(x) = \sum_{i=1}^{r} p_i(x)c^i, \quad c^i \in \mathbb{R}^n,\ i = 1,\dots,r.$$

For every $x, x' \in X$ satisfying $\langle c^i, x' - x\rangle = 0, i = 1,\dots,r$, there is some $\theta \in (0,1)$ such that $f(x') - f(x) = \langle \nabla f(x + \theta(x' - x)), x' - x\rangle = \sum_{i=1}^r p_i(x + \theta(x' - x))\langle c^i, x' - x\rangle = 0$, hence $f(x)$ is monotonic with respect to the space $L = \{x|\ \langle c^i, x\rangle = 0,\ i = 1,\dots,r\}$. If, in addition, $p_i(x) \geq 0\ \ \forall x\ (i = 1,\dots,r)$ then monotonicity holds with respect to the cone $K = \{u|\ \langle c^i, u\rangle \geq 0, i = 1,\dots,r\}$.

Example 9.6 $f(x) = \min\{y^TQx|\ y \in E\}$
where E is a convex subset of $\mathbb{R}^m$ and Q is an $m \times n$ matrix of rank r.

This function appears when transforming *rank r bilinear programs* into concave minimization problems (see Sect. 10.4). Here monotonicity holds with respect to the subspace $K = \{u|\ Qu = 0\}$. If $E \subset \mathbb{R}^m_+$ then $Qu \geq 0$ implies $y^TQ(x + u) \geq y^TQx$, hence $f(x + u) \geq f(x)$, and so $f(x)$ is monotonic with respect to the cone $K = \{u|\ Qu \geq 0\}$.

Example 9.7 $f(x) = \sup\{d^Ty|\ \ Ax + By \leq q\}$
where $A \in \mathbb{R}^{p\times n}$ with rank$A = r,\ B \in \mathbb{R}^{p\times m},\ y \in \mathbb{R}^m$ and $q \in \mathbb{R}^p$.

This function appears in bilevel linear programming, for instance, in the max-min problem (Falk 1973b):

$$\min_x \max_y \{c^Tx + d^Ty|\ \ Ax + By \leq q, x \geq 0, y \geq 0\}.$$

Obviously, $Au \leq 0$ implies $A(x + u) \leq Ax$, hence $\{y|\ \ Ax + By \leq q\} \subset \{y|\ \ A(x + u) + By \leq q\}$, hence $f(x + u) \geq f(x)$. Therefore, monotonicity holds with respect to the cone $K = \{u|\ Au \leq 0\}$.

Note that by Corollary 1.15, rank$Q = n - \dim L = \dim K^{\circ}$. A K- monotonic function $f(x)$ on X, with rank$Q = r$ is also said to possess the *rank r monotonicity property*.

9.2 Decomposition by Projection

Let D be a compact subset of a closed convex set X in $\mathbb{R}^n$ and let $f : \mathbb{R}^n \to \mathbb{R}$ be a quasiconcave function, monotonic on X with respect to a polyhedral convex cone K contained in the recession cone of X. Assume that the lineality space of K is $L = \{x|\ Qx = 0\}$, where Q is an $r \times n$ matrix of rank r. Consider the *quasiconcave monotonic problem*

$$(QCM) \qquad \min\{f(x)|\quad x \in D\} \tag{9.8}$$

In this and the next two sections, we shall discuss decomposition methods for solving this problem, under the additional assumption that D is a polytope. The idea of decomposition is to transform the original problem into a sequence of subproblems involving each a reduced number of nonconvex variables. This can be done either by projection, or by dualization (polyhedral annexation), or by parametrization (for $r \leq 3$). We first present decomposition methods by projection.

As has been previously observed, by setting

$$F(y) = f(x) \text{ for any } x \in X \text{ satisfying } y = Qx, \tag{9.9}$$

we unambiguously define a quasiconcave function $F : Q(X) \to \mathbb{R}$ such that problem (QCM) is equivalent to

$$\min F(y) \quad \text{s.t.} \quad y = (y_1, \ldots, y_r) \in G \tag{9.10}$$

where $G = Q(D)$ is a polytope in $\mathbb{R}^r$. The algorithms presented in Chaps. 5 and 6 can of course be adapted to solve this quasiconcave minimization problem.

We first show how to compute $F(y)$. Since $\operatorname{rank} Q = r$, by writing $Q = [Q_B, Q_N]$, $x = \begin{bmatrix} x_B \\ x_N \end{bmatrix}$, where Q_B is an $r \times r$ nonsingular matrix, the equation $Qx = y$ yields

$$x = \begin{bmatrix} Q_B^{-1} \\ 0 \end{bmatrix} y + \begin{bmatrix} -Q_B^{-1} Q_N x_N \\ x_N \end{bmatrix}$$

Setting $Z = \begin{bmatrix} Q_B^{-1} \\ 0 \end{bmatrix}$ and $u = \begin{bmatrix} -Q_B^{-1} Q_N x_N \\ x_N \end{bmatrix}$ we obtain

$$x = Zy + u \quad \text{with } Qu = -Q_N x_N + Q_N x_N = 0. \tag{9.11}$$

Since $Qx = Q(Zy)$ with $Zy = x - u \in X - L = X$, it follows that $f(x) = f(Zy)$. Thus, the objective function of (9.10) can be computed by the formula

$$F(y) = f(Zy). \tag{9.12}$$

We shall assume that $f(x)$ is continuous on some open set in $\mathbb{R}^n$ containing D, so $F(y)$ is continuous on some open set in $\mathbb{R}^r$ containing the constraint polytope $G = Q(D)$ of (9.10). The peculiar form of the latter set (image of a polytope D under the linear map Q) is a feature which requires special treatment, but does not cause any particular difficulty.

9.2.1 Outer Approximation

To solve (9.10) by outer approximation, the key point is: given a point $\overline{y} \in Q(\mathbb{R}^n) \subset \mathbb{R}^r$, determine whether $\overline{y} \in G = Q(D)$, and if $\overline{y} \notin G$ then construct a linear inequality (cut) $l(y) \leq 0$ to exclude $\overline{y}$ without excluding any point of G.

Observe that $\overline{y} \notin Q(D)$ if and only if the linear system $Qx = \overline{y}$ has no solution in D, so the existence of $l(y)$ is ensured by the separation theorem or any of its equivalent forms (such as the Farkas–Minkowski Lemma). However, since we are not so much interested in the existence as in the effective construction of $l(y)$, the best way is to use the duality theorem of linear programming (which is another form of the separation theorem). Specifically, assuming $D = \{x|\ Ax \leq b,\ x \geq 0\}$, with $A \in \mathbb{R}^{m\times n},\ b \in \mathbb{R}^m$, consider the dual pair of linear programs

$$\min\{\langle h, x\rangle|\ \ Qx = \overline{y}, Ax \leq b, x \geq 0\} \tag{9.13}$$

$$\max\{\langle \overline{y}, v\rangle - \langle b, w\rangle|\ Q^T v - A^T w \leq h,\ w \geq 0\}. \tag{9.14}$$

where h is any vector chosen so that (9.14) is feasible (for example, $h = 0$) and can be solved with least effort.

Proposition 9.3 *Solving (9.14) either yields a finite optimal solution or an extreme direction* $(\overline{v}, \overline{w})$ *of the cone* $Q^T v - A^T w \leq 0,\ w \geq 0$, *such that* $\langle \overline{y}, \overline{v}\rangle - \langle b, \overline{w}\rangle > 0$. *In the former case,* $\overline{y} \in G = Q(D)$; *in the latter case, the affine function*

$$l(y) = \langle y, \overline{v}\rangle - \langle b, \overline{w}\rangle \tag{9.15}$$

satisfies

$$l(\overline{y}) > 0, \quad l(y) \leq 0 \quad \forall y \in G. \tag{9.16}$$

Proof Immediate from the duality theorem of linear programming. □

On the basis of this proposition a finite OA procedure for solving (9.10) can be carried out in the standard way, starting from an initial r-simplex $P_1 \supset G = Q(D)$ with a known or readily computable vertex set. In most cases, one can take this initial simplex to be

$$P_1 = \left\{ y|\ \ y_i \geq \alpha_i,\ (i = 1, \ldots, r),\ \sum_{i=1}^{r}(y_i - \alpha_i) \leq \beta \right\} \tag{9.17}$$

where

$$\alpha_i = \min\{y_i|\ \ y = Qx, x \in D\}, i = 1, \ldots, r, \tag{9.18}$$

$$\beta = \max\left\{\sum_{i=1}^{r}(y_i - \alpha_i)|\ \ y = Qx, x \in D\right\}. \tag{9.19}$$

Remark 9.1 A typical case of interest is the *concave program with few nonconvex variables* (Example 7.1):

$$\min\{\varphi(y) + dz|\ Ay + Bz \le b,\ y \ge 0,\ z \ge 0\} \tag{9.20}$$

where the function $\varphi : \mathbb{R}^p \to \mathbb{R}$ is concave and $(y, z) \in \mathbb{R}^p \times \mathbb{R}^q$. By rewriting this problem in the form (9.10):

$$\min\{\varphi(y) + t|\ dz \le t,\ Ay + Bz \le b,\ y \ge 0,\ z \ge 0\}$$

we see that for checking the feasibility of a vector $(\overline{y}, \overline{t}) \in \mathbb{R}^p_+ \times \mathbb{R}$ the best choice of h in (9.13) is $h = d$. This leads to the pair of dual linear programs

$$\min\{dz|\ Bz \le b - A\overline{y},\ z \ge 0\} \tag{9.21}$$

$$\max\{\langle A\overline{y} - b, v\rangle|\ \ -B^T v \le d,\ v \ge 0\}. \tag{9.22}$$

We leave it to the reader to verify that:

1. If $\overline{v}$ is an optimal solution to (9.22) and $\langle A\overline{y} - b, \overline{v}\rangle \le \overline{t}$ then $(\overline{y}, \overline{t})$ is feasible;
2. If $\overline{v}$ is an optimal solution to (9.22) and $\langle A\overline{y} - b, \overline{v}\rangle > \overline{t}$ then the cut $\langle Ay - b, \overline{v}\rangle \le t$ excludes $(\overline{y}, \overline{t})$ without excluding any feasible point;
3. If (9.22) is unbounded, i.e., $\langle A\overline{y} - b, \overline{v}\rangle > 0$ for some extreme direction $\overline{v}$ of the feasible set of (9.22), then the cut $\langle Ay - b, \overline{v}\rangle \le 0$ excludes $(\overline{y}, \overline{t})$ without excluding any feasible point.

   Sometimes the matrix B has a favorable structure such that for each fixed $y \ge 0$ the problem (9.20) is solvable by efficient specialized algorithms (see Sect. 9.8). Since the above decomposition preserves this structure in the auxiliary problems, the latter problems can be solved by these algorithms.

9.2.2 Branch and Bound

A general rule in the branch and bound approach to nonconvex optimization problems is to branch upon the nonconvex variables, except for rare cases where there are obvious reasons to do otherwise. Following this rule the space to be partitioned for solving (QCM) is $Q(\mathbb{R}^n) = \mathbb{R}^r$. Given a partition set $M \subset \mathbb{R}^r$ a basic operation is to compute a lower bound for $F(y)$ over the feasible points in M. To this end, select an appropriate affine minorant $l_M(y)$ of $F(y)$ over M and solve the linear program

$$\min\{l_M(y)|\ Qx = y, x \in D, y \in M\}.$$

If $\beta(M)$ is the optimal value of this linear program then obviously

$$\beta(M) \leq \min\{F(y)|\ y \in Q(D) \cap M\}.$$

According to condition (6.4) in Sect. 6.2 this can serve as a valid lower bound, provided, in addition, $\beta(M) < +\infty$ whenever $Q(D) \cap M \neq \emptyset$ – which is the case when M is bounded as, e.g., when using simplicial or rectangular partitioning.

Simplicial Algorithm
If $M = [u^1, \ldots, u^{r+1}]$ is an r-simplex in $Q(\mathbb{R}^n)$ then $l_M(x)$ can be taken to be the affine function that agrees with $F(y)$ at $u^1, \ldots, u^{r+1}$. Since $l_M(y) = \sum_{i=1}^{r+1} t_i F(u^i)$ for $y = \sum_{i=1}^{r+1} t_i u^i$ with $t = (t_1, \ldots, t_{r+1}) \geq 0, \sum_{i=1}^{r+1} t_i = 1$, the above linear program can also be written as

$$\min \left\{ \sum_{i=1}^{r+1} t_i F(u^i)|\ x \in D, Qx = \sum_{i=1}^{r+1} t_i u^i, \sum_{i=1}^{r+1} t_i = 1, t \geq 0 \right\}.$$

Simplicial subdivision is advisable when $f(x) = f_0(x) + dx$, where $f_0(x)$ is concave and monotonic with respect to the cone $Qx \geq 0$. Although $f(x)$ is then actually monotonic with respect to the cone $Qx \geq 0, dx \geq 0$, by writing (QCM) in the form

$$\min\{F_0(y) + dx|\ Qx = y, x \in D\}$$

one can see that it can be solved by a simplicial algorithm operating in $Q(\mathbb{R}^n)$ rather than $Q(\mathbb{R}^n) \times \mathbb{R}$. As initial simplex M_1 one can take the simplex (9.17).

Rectangular Algorithm
When $F(y)$ is concave separable, i.e.,

$$F(y) = \sum_{i=1}^{r} F_i(y_i),$$

where each $F_i(\cdot), i = 1, \ldots, r$ is a concave univariate function, this structure can be best exploited by rectangular subdivision. In that case, for any rectangle $M = [p, q] = \{y|\ p \leq y \leq q\}$, $l_M(y)$ can be taken to be the uniquely defined affine function which agrees with $F(y)$ at the corners of M. This function is $l_M(y) = \sum_{i=1}^{r} l_{i,M}(y_i)$ where each $l_{i,M}(t)$ is the affine univariate function that agrees with $F_i(t)$ at the endpoints p_i, q_i of the segment $[p_i, q_i]$. The initial rectangle can be taken to be $M_1 = [\alpha_1, \beta_1] \times \ldots \times [\alpha_r, \beta_r]$, where α_i is defined as in (9.18) and

$$\beta_i = \sup\{y_i|\ y = Qx, x \in D\}.$$

Conical Algorithm
In the general case, since the global minimum of $F(y)$ is achieved on the boundary of $Q(D)$, conical subdivision is most appropriate. Nevertheless, instead of bounding, we use a selection operation, so that the algorithm is a branch and select rather than branch and bound procedure.

The key point in this algorithm is to decide, at each iteration, which cones in the current partition are 'non promising' and which one is the 'most promising' (selection rule).

Let $\bar{y}$ be the current best solution (CBS) at a given iteration and $\gamma := F(\bar{y})$. Assume that all the cones are vertexed at 0 and $F(0) > \gamma$, so that

$$0 \in \operatorname{int}\Omega_\gamma,$$

where $\Omega_\gamma = \{y \in \Omega |\ F(y) \geq \gamma\}$. Let $l_M(y) = 1$ be the equation of the hyperplane in $\mathbb{R}^r$ passing through the intersections of the edges of M with the surface $F(y) = \gamma$. Denote by $\omega(M)$ a basic optimal solution and by $\mu(M)$ the optimal value of the linear program

$$LP(M, \gamma) \qquad\qquad \max\{l_M(y)|\ \ y \in Q(D) \cap M\}.$$

Note that if $u^1, \ldots, u^r$ are the directions of the edges of M and $U = (u^1, \ldots, u^r)$ is the matrix of columns $u^1, \ldots, u^r$ then $M = \{y|\ y = Ut, t = (t_1, \ldots, t_r)^T \geq 0\}$, so LP(M,$\gamma$) can also be written as

$$\max\{l_M(Ut)|\ Ax \leq b,\ Qx = Ut,\ x \geq 0,\ t \geq 0\}.$$

The deletion and selection rule are based on the following:

Proposition 9.4 *If $\mu(M) \leq 1$ then M cannot contain any feasible point better than CBS. If $\mu(M) > 1$ and $\omega(M)$ lies on an edge of M then $\omega(M)$ is a better feasible solution than CBS.*

Proof Clearly $D \cap M$ is contained in the simplex $Z := M \cap \{y|\ l_M(y) \leq \mu(M)\}$. If $\mu(M) \leq 1$, then the vertices of Z other than 0 lie inside Ω_γ, hence the values of $F(y)$ at these vertices are at least equal to γ. But by quasiconcavity of $F(y)$, its minimum over Z must be achieved at some vertex of Z, hence must be at least equal to γ. This means that $F(y) \geq \gamma$ for all $y \in Z$, and hence, for all $y \in D \cap M$. To prove the second assertion, observe that if $\omega(M)$ lies on an edge of M and $\mu(M) > 1$, then $\omega(M) \notin \Omega_\gamma$, i.e., $F(\omega(M)) < \gamma$; furthermore, $\omega(M)$ is obviously feasible. □

The algorithm is initialized from a partition of $\mathbb{R}^r$ into $r + 1$ cones vertexed at 0 (take, e.g., the partition induced by the r-simplex $[e^0, e^1, \ldots, e^r]$ where e^i, for $i = 1, \ldots, r$, is the i-th unit vector of $\mathbb{R}^r$ and $e^0 = -\frac{1}{r}(e^1 + \ldots + e^r)$). On the basis of Proposition 9.4, at iteration k a cone M will be fathomed if $\mu(M) \leq 1$, while the most promising cone M_k (the candidate for further subdivision) is the unfathomed cone that has largest $\mu(M)$. If $\omega^k = \omega(M_k)$ lies on any edge of M^k it yields, by Proposition 9.4, a better feasible solution than CBS and the algorithm can go to the next iteration with this improved CBS. Otherwise, M_k can be subdivided via the ray through ω^k.

As proved in Chap. 7 (Theorem 7.3), the above described conical algorithm converges, in the sense that any accumulation point of the sequence of CBS that it generates is a global optimal solution. Note that no bound is needed in this branch and select algorithm. Although a bound for $F(y)$ over the feasible points in M can be easily derived from the constructions used in Proposition 9.4, it would require the computation of the values of $F(y)$ at r additional points (the intersections of the edges of M with the hyperplane $l_M(y) = \mu(M)$), which may entail a nonnegligible extra cost.

9.3 Decomposition by Polyhedral Annexation

The decomposition method by outer approximation is conceptually simple. However, a drawback of this method is its inability to be restarted. Furthermore, the construction of the initial polytope involves solving $r + 1$ linear programs. An alternative decomposition method which allows restarts and usually performs better is by dualization through polyhedral annexation .

Let $\overline{x}$ be a feasible solution and $C := \{x \in X|\ f(x) \geq f(\overline{x})\}$. By quasiconcavity and continuity of $f(x)$, C is a closed convex set. By translating if necessary, assume that 0 is a feasible point such that $f(0) > f(\overline{x})$, so $0 \in D \cap \mathrm{int}C$. The PA method (Chap. 8, Sect. 8.1) uses as a subroutine Procedure DC* for solving the following subproblem:

Find a point $x \in D \setminus C$ (i.e., a better feasible solution than $\overline{x}$) or else prove that $D \subset C$ (i.e., $\overline{x}$ is a global optimal solution).

The essential idea of Procedure DC* is to build up, starting with a polyhedron $P_1 \supset C^{\circ}$ (Lemma 8.1), a nested sequence of polyhedrons $P_1 \supset P_2 \supset \ldots \supset C^{\circ}$ that yields eventually either a polyhedron $P_k \subset D^{\circ}$ (in that case $C^{\circ} \subset D^{\circ}$, hence $D \subset C$) or a point $x^k \in D \setminus C$. In order to exploit the monotonicity structure, we now try to choose the initial polyhedron P_1 so that $P_1 \subset L^{\perp}$ (orthogonal complement of L). If this can be achieved, then the whole procedure will operate in $L^{\perp}$, which allows much computational saving since $\dim L^{\perp} = r < n$.

Let $c^1, \ldots, c^r$ be the rows of Q (so that $L^{\perp}$ is the subspace generated by these vectors) and let $c^0 = -\sum_{i=1}^{r} c^i$. Let $\pi : L^{\perp} \to \mathbb{R}^r$ be the linear map such that $y = \sum_{i=1}^{r} t_i c^i \Leftrightarrow \pi(y) = t$. For each $i = 0, 1, \ldots, r$ since $0 \in \mathrm{int}C$, the halfline from 0 through c^i intersects C. If this halfline meets ∂C, we define α_i to be the positive number such that $z^i = \alpha_i c^i \in \partial C$; otherwise, we set $\alpha_i = +\infty$. Let $I = \{i|\ \alpha_i < +\infty\}, S_1 = \mathrm{conv}\{0, z^i, i \in I\} + \mathrm{cone}\{c^i, i \notin I\}$.

Lemma 9.1

(i) *The polar P_1 of $S_1 + L$ is an r-simplex in $L^{\perp}$ defined by*

$$\pi(P_1) = \left\{ t \in \mathbb{R}^r |\ \sum_{i=1}^{r} t_i \langle c^i, c^j \rangle \leq \frac{1}{\alpha_j},\ j = 0, 1, \ldots, r \right\}, \tag{9.23}$$

(as usual $\frac{1}{+\infty} = 0$).

(ii) *$C^\circ \subset P_1$ and if V is the vertex set of any polytope P such that $C^\circ \subset P \subset P_1$, then*

$$\max\{\langle v, x\rangle |\quad x \in D\} \le 1 \quad \forall v \in V \quad \Rightarrow \quad C^\circ \subset D^\circ. \tag{9.24}$$

Proof Clearly $S_1 \subset L^\perp$ and $0 \in \mathrm{ri}S_1$, so $0 \in \mathrm{int}(S_1 + L)$, hence P_1 is compact. Since $(S_1 + L)^\circ = S_1^\circ \cap L^\perp$, any $y \in P_1$ belongs to $L^\perp$, i.e., is of the form $y = \sum_{i=1}^r t_i c^i$ for some $t = (t_1, \ldots, t_r)$. But by Proposition 1.28, $y \in S_1^\circ$ is equivalent to $\langle c^j, y\rangle \le \frac{1}{\alpha_j}$ $\forall j = 0, 1, \ldots, r$. Hence P_1 is the polyhedron defined by (9.23). We have $0 < \frac{1}{\alpha_j}$ $\forall j = 0, 1, \ldots, r$, so $\dim P_1 = r$, (Corollary 1.16) and since P_1 is compact it must be an r-simplex. To prove (ii) observe that $S_1 + L \subset C \cap L^\perp + L = C$, so $C^\circ \subset (S_1 + L)^\circ = P_1$. If $\max\{\langle v, x\rangle |\ x \in D\} \le 1 \quad \forall v \in V$, then $\max\{\langle y, x\rangle |\ x \in D\} \le 1$ $\forall y \in P$, i.e., $P \subset D^\circ$, hence $C^\circ \subset D^\circ$ because $C^\circ \subset P$. □

Thus Procedure DC* can be applied, starting from P_1 as an initial polytope. By incorporating Procedure DC* (initialized from P_1) into the PA Algorithm for (*BCP*) (Sect. 7.1) we obtain the following decomposition algorithm for (*QCM*):

PA Algorithm for (QCM)

Initialization. Let $\bar{x}$ be a feasible solution (the best available), $C = \{x \in X |\ f(x) \ge f(\bar{x})\}$. Choose a feasible point x^0 of D such that $f(x^0) > f(\bar{x})$ and set $D \leftarrow D - x^0$, $C \leftarrow C - x^0$. Let $\tilde{P}_1 = \pi(P_1)$ be the simplex in $\mathbb{R}^r$ defined by (9.23) (or $\tilde{P}_1 = \pi(P_1)$ with P_1 being any polytope in $L^\perp$ such that $C^\circ \subset P_1$). Let V_1 be the vertex set of $\tilde{P}_1$. Set $k = 1$.

Step 1. For every new $t = (t_1, \ldots, t_r) \in V_k$ solve the linear program

$$\max\left\{\sum_{i=1}^r t_i\langle c^i, x\rangle |\ \ x \in D\right\} \tag{9.25}$$

to obtain its optimal value $\mu(t)$ and a basic optimal solution $x(t)$.

Step 2. Let $t^k \in \mathrm{argmax}\{\mu(t) |\ t \in V_k\}$. If $\mu(t^k) \le 1$ then terminate: $\bar{x}$ is an optimal solution of (*QCM*).

Step 3. If $\mu(t^k) > 1$ and $x^k = x(t^k) \notin C$, then update the current best feasible solution and the set C by resetting $\bar{x} = x^k$.

Step 4. Compute

$$\theta_k = \sup\{\theta |\ f(\theta x^k) \ge f(\bar{x})\}$$

and define

$$\tilde{P}_{k+1} = \tilde{P}_k \cap \left\{t |\ \sum_{i=1}^r t_i\langle x^k, c^i\rangle \le \frac{1}{\theta_k}\right\}.$$

Step 5. From V_k derive the vertex set V_{k+1} of $\tilde{P}_{k+1}$. Set $k \leftarrow k + 1$ and go back to Step 1.

Remark 9.2 Reported computational experiments (Tuy and Tam 1995) seem to indicate that the PA method is quite practical for problems with fairly large n provided r is small (typically $r \leq 6$). For $r \leq 3$, however, specialized parametric methods can be developed which are more efficient. These will be discussed in the next section.

Case When $K = \{u|\ Qu \geq 0\}$

A case when the initial simplex P_1 can be constructed so that its vertex set is readily available when $K = \{u|\ Qu \geq 0\}$, with an $r \times n$ matrix Q of rank r. Let $c^1, \ldots, c^r$ be the rows of Q. Using formula (9.11) one can compute a point w satisfying $Qw = -e$, where $e = (1, \ldots, 1)^T$, and a value $\alpha > 0$ such that $\hat{w} = \alpha w \in C$ (for the efficiency of the algorithm, α should be taken to be as large as possible).

Lemma 9.2 *The polar P_1 of the set $S_1 = \hat{w} + K$ is an r- simplex with vertex set $\{0, -c^1/\alpha, \ldots, -c^r/\alpha\}$ and satisfying condition (ii) in Lemma 9.1.*

Proof Since $K \subset \mathrm{rec}C$ and $\hat{w} \in C$ it follows that $S_1 = \hat{w} + K \subset C$, and hence $P_1 \supset C^{\circ}$. Furthermore, clearly $S_1 = \{x|\ \langle c^i, x\rangle \geq -\alpha,\ i = 1, \ldots, r\}$ because $x \in S_1$ if and only if $x - \hat{w} \in K$, i.e., $\langle c^i, x - \hat{w}\rangle \geq 0$ $i = 1, \ldots, r$, i.e., $\langle c^i, x\rangle \geq \langle c^i, \hat{w}\rangle = -\alpha,\ i = 1, \ldots, r$. From Proposition 1.28 we then deduce that $S_1^{\circ} = \mathrm{conv}\{0, -c^1/\alpha, \ldots, -c^r/\alpha\}$. The rest is immediate. □

Also note that in this case $K^{\circ} = \mathrm{cone}\{-c^1, \ldots, -c^r\}$ and so $\pi(K^{\circ}) = \mathbb{R}^r_-$ while $\pi(P_1) = \{t \in \mathbb{R}^r_-|\ \sum_{i=1}^r t_i \geq -1/\alpha\}$.

9.4 The Parametric Programming Approach

In the previous sections, it was shown how a quasiconcave monotonic problem can be transformed into an equivalent quasiconcave problem of reduced dimension. In this section an alternative approach will be presented in which a quasiconcave monotonic problem is solved through the analysis of an associated linear program depending on a parameter (of the same dimension as the reduced problem in the former approach).

Consider the general problem formulated in Sect. 8.2:

$$(QCM) \qquad \min\{f(x)|\quad x \in D\} \tag{9.26}$$

where D is a compact set in $\mathbb{R}^n$, not necessarily even convex, and $f(x)$ is a quasiconcave monotonic function on a closed convex set $X \supset D$ with respect to a polyhedral cone K with lineality space $L = \{x|\ Qx = 0\}$. To this problem we associate the following program where v is an arbitrary vector in K° :

$$LP(v) \qquad\qquad \min\{\langle v, x\rangle |\quad x \in D\}. \tag{9.27}$$

Denote by K_b° any compact set such that $\text{cone}(K_b^\circ) = K^\circ$.

Theorem 9.1 *For every $v \in -K^\circ$, let x^v be an arbitrary optimal solution of $LP(v)$. Then*

$$\min\{f(x)|\, x \in D\} = \min\{f(x^v)|\, v \in -K_b^\circ\}. \tag{9.28}$$

Proof Let $\gamma = \min\{f(x^v)|\, v \in -K_b^0\}$. Clearly $\gamma = \min\{f(x^v)|\, v \in -K^\circ \setminus \{0\}\}$. Denote by E the convex hull of the closure of the set of all $x^v, v \in -K^\circ \setminus \{0\}$. Since D is compact, so is E and it is easy to see that the convex set $G = E + K$ is closed. Indeed, for any sequence $x^\nu = u^\nu + v^\nu \to x^0$ with $u^\nu \in E, v^\nu \in K$, one has, by passing to subsequences if necessary, $u^\nu \to u^0 \in E$, hence $v^\nu = x^\nu - u^\nu \to x^0 - u^0 \in K$, which implies that $x^0 = u^0 + (x^0 - u^0) \in E + K = G$. Thus, G is a closed convex set. By K-monotonicity we have $f(y+u) \geq f(y)$ for all $y \in E, u \in K$, hence $f(x) \geq \gamma$ for all $x \in E + K = G$. Suppose now that $\gamma > \min\{f(x)|\, x \in D\}$, so that for some $z \in D$ we have $f(z) < \gamma$, i.e., $z \in D \setminus G$. By Proposition 1.15 on the normals to a convex set there exists $x^0 \in \partial G$ such that $v := x^0 - z$ satisfies

$$\langle v, x - x^0\rangle \geq 0 \quad \forall x \in G, \quad \langle v, z - x^0\rangle < 0. \tag{9.29}$$

For every $u \in K$, since $x^0 + u \in G + K = G$, we have $\langle v, u\rangle \geq 0$, hence $v \in -K^\circ \setminus \{0\}$. But, since $z \in D$, it follows from the definition of x^v that $\langle v, x^v\rangle \leq \langle v, z\rangle < \langle v, x^0\rangle$ by (9.29), hence $\langle v, x^v - x^0\rangle < 0$, conflicting with the fact $x^v \in E \subset G$ and (9.29). □

The above theorem reduces problem (QCM) to minimizing the function $\varphi(v) := f(x^v)$ over K_b°. Though $K^\circ \subset L^\perp$, and $\dim L^\perp = r$ may be small, this is still a *difficult problem* even when D is convex, because $\varphi(v)$ is generally highly nonconvex. The point, however, is that in many important cases the problem (9.28) is more amenable to computational analysis than the original problem. For instance, as we saw in the previous section, when D is a polytope the PA Method solves (QCM) through solving a suitable sequence of linear programs (9.25) which can be recognized to be identical to (9.27) with $v = -\sum_{i=1}^r t_i c^i$.

An important consequence of Theorem 9.1 can be drawn when:

$$\left|\begin{array}{l} D \text{ is a polyhedron;} \\ K = \{x|\ \langle c^i, x\rangle \geq 0, i = 1, \dots, p;\ \langle c^i, x\rangle = 0, i = p+1, \dots, r\} \end{array}\right. \tag{9.30}$$

($c^1, \ldots, c^r$ being linearly independent vectors of $\mathbb{R}^n$). Let

$$\begin{aligned}
&\Lambda = \{\lambda \in \mathbb{R}^p_+ |\ \lambda_p = 1\},\\
&W = \{\alpha = (\alpha_{p+1}, \ldots, \alpha_r) |\ \underline{\alpha}_i \le \alpha_i \le \overline{\alpha}_i,\ i = p+1, \ldots, r\},\\
&\underline{\alpha}_i = \inf_{x \in D} \langle c^i, x\rangle; \quad \overline{\alpha}_i = \sup_{x \in D} \langle c^i, x\rangle \ \ i = p+1, \ldots, r.
\end{aligned}$$

Corollary 9.1 *For each* $(\lambda, \alpha) \in \Lambda \times W$ *let* $x^{\lambda,\alpha}$ *be an arbitrary basic optimal solution of the linear program*

$$LP(\lambda,\alpha) \qquad \left| \begin{array}{l} \min \sum_{i=1}^{p-1} \lambda_i \langle c^i, x\rangle + \langle c^p, x\rangle \\ \text{s.t. } \ x \in D,\ \langle c^i, x\rangle = \alpha_i,\ i = p+1, \ldots, r. \end{array} \right.$$

Then

$$\min\{f(x)|\ x \in D\} = \min\{f(x^{\lambda,\alpha})|\ \lambda \in \Lambda,\ \alpha \in W\}. \tag{9.31}$$

Proof Clearly

$$\min\{f(x)|\ x \in D\} = \min_{\alpha} \inf\{f(x)|\ x \in D \cap H_\alpha\},$$

where $H_\alpha = \{x|\ \langle c^i, x\rangle = \alpha_i,\ i = p+1 \ldots, r\}$. Let $\tilde{K} = \{x|\ \langle c^i, x\rangle \ge 0,$ $i = 1, \ldots, p\}$. Since $f(x)$ is quasiconcave $\tilde{K}$-monotonic on $X \cap H_\alpha$, we have by Theorem 9.1 :

$$\min\{f(x)|\ x \in D \cap H_\alpha\} = \min\{f(x^{v,\alpha})|\ v \in -\tilde{K}^\circ_b\}, \tag{9.32}$$

where $x^{v,\alpha}$ is any basic optimal solution of the linear program

$$\min\{\langle v, x\rangle|\ x \in D \cap H_\alpha\}. \tag{9.33}$$

Noting that $\tilde{K}^\circ = -\text{cone}\{c^1, \ldots, c^p\}$ we can take $\tilde{K}^\circ_b = -[c^1, \ldots, c^p]$. Also we can assume $v \in -\text{ri}\tilde{K}^\circ_b$ since for any $v \in -\tilde{K}^\circ_b$ there exists $v' \in -\text{ri}\tilde{K}^\circ_b$ such that $x^{v',\alpha}$ is a basic optimal solution of (9.33). So $v = \sum_{i=1}^p t_i c^i$ with $t_p > 0$ and the desired result follows by setting $\lambda_i = \frac{t_i}{t_p}$. □

As mentioned above, by Theorem 9.1 one can solve (QCM) by finding first the vector $\overline{v} \in -K^\circ_b$ that minimizes $f(x^v)$ over all $v \in K^\circ_b$. In certain methods, as, for instance, in the PA method when D is polyhedral, an optimal solution $\overline{x} = x^{\overline{v}}$ is immediately known once $\overline{v}$ has been found. In other cases, however, $\overline{x}$ must be sought by solving the associated program $\min\{\langle \overline{v}, x\rangle|\ x \in D\}$. The question then arises as to whether any optimal solution of the latter program will solve (QCM).

Theorem 9.2 *If, in addition to the already stated assumptions, $D \subset \text{int}X$, while the function $f(x)$ is continuous and strictly quasiconcave on X, then for any optimal solution $\overline{x}$ of (QCM) there exists a $\overline{v} \in K^{\circ} \setminus \{0\}$ such that $\overline{x}$ is an optimal solution of $LP(\overline{v})$ and any optimal solution to $LP(\overline{v})$ will solve (QCM).*

(A function $f(x)$ is said to be strictly quasiconcave on X if for any two $x, x' \in X$, $f(x') < f(x)$ always implies $f(x') < f(\theta x + (1-\theta)x')$ $\forall \theta \in (0,1)$)

Proof It suffices to consider the case when $f(x)$ is not constant on D. Let $\overline{x}$ be an optimal solution of (QCM) and $X(\overline{x}) = \{x \in X|\, f(x) \geq f(\overline{x})\}$. By quasiconcavity and continuity of $f(x)$, this set $X(\overline{x})$ is closed and convex. By optimality of $\overline{x}$, $D \subset X(\overline{x})$ and again by continuity of $f(x)$, $\{x \in X|\, f(x) > f(\overline{x})\} \subset \text{int}X(\overline{x})$. Furthermore, since $f(x)$ is not constant on D one can find $x' \in D$ such that $f(x') > f(\overline{x})$, so if $\overline{x} \in \text{int}X(\overline{x})$ then for $\theta > 0$ small enough, $x^{\theta} = \overline{x} - \theta(x' - \overline{x}) \in X(\overline{x})$, $f(x^{\theta}) < f(x')$, hence, by strict quasiconcavity of $f(x)$, $f(x^{\theta}) < f(\overline{x})$, conflicting with the definition of $X(\overline{x})$. Therefore, $\overline{x}$ lies on the boundary of $X(\overline{x})$. By Theorem 1.5 on supporting hyperplanes there exists then a vector $\overline{v} \neq 0$ such that

$$\langle \overline{v}, x - \overline{x} \rangle \geq 0 \quad \forall x \in X(\overline{x}). \tag{9.34}$$

For any $u \in K$ we have $\overline{x} + u \in D + K \subset X(\overline{x})$, so $\langle \overline{v}, -u \rangle \geq 0$, and consequently, $\overline{v} \in -K^{\circ} \setminus \{0\}$. If $\tilde{x}$ is any optimal solution of $LP(\overline{v})$, then $\langle \overline{v}, \tilde{x} - \overline{x} \rangle = 0$, and in view of (9.34), $\tilde{x}$ cannot be an interior point of $X(\overline{x})$ for then $\langle \overline{v}, x - \overline{x} \rangle = 0 \quad \forall x \in \mathbb{R}^n$, conflicting with $\overline{v} \neq 0$. So $\tilde{x}$ is a boundary point of $X(\overline{x})$, and hence $f(\tilde{x}) = f(\overline{x})$, because $\{x \in X|\, f(x) > f(\overline{x})\} \subset \text{int}X(\overline{x})$. Thus, any optimal solution $\tilde{x}$ to LP$(\overline{v})$ is optimal to (QCM). □

Remark 9.3 If $f(x)$ is differentiable and pseudoconcave then it is strictly quasiconcave (see, e.g., Mangasarian 1969). In that case, since $X(\overline{x}) = \{x \in X|f(x) \geq f(\overline{x})\}$ one must have $\langle \nabla f(\overline{x}), x - \overline{x} \rangle \geq 0$ $\forall x \in X(\overline{x})$ and since one can assume that there is $x \in X$ satisfying $f(x) > f(\overline{x})$ the strict quasiconcavity of $f(x)$ implies that $\nabla f(\overline{x}) \neq 0$ (Mangasarian 1969). So $\nabla f(\overline{x})$ satisfies (9.34) and one can take $\overline{v} = \nabla f(\overline{x})$. This result was established in Sniedovich (1986) as the foundation of the so-called C-programming.

Corollary 9.2 *Let $F : \mathbb{R}^q \to \mathbb{R}$ be a continuous, strictly quasiconcave and K-monotonic function on a closed convex set $Y \subset \mathbb{R}^q$, where $K \subset \mathbb{R}^q$ is a polyhedral convex cone contained in the recession cone of Y. Let $g : D \to \mathbb{R}^q$ be a map defined on a set $D \subset \mathbb{R}^n$ such that $g(D) \subset \text{int}Y$. If $F(g(x))$ attains a minimum on D then there exists $t \in -K^{\circ}$ such that any optimal solution of*

$$\min\{\langle t, g(x) \rangle | \quad x \in D\} \tag{9.35}$$

will solve

$$\min\{F(g(x)) | \quad x \in D\}. \tag{9.36}$$

Proof If $\bar{x}$ is a minimizer of $F(g(x))$ over D, then $\bar{y} = g(\bar{x})$ is a minimizer of $F(y)$ over $G = g(D)$, and the result follows from Theorem 9.2. □

In particular, when $K \supset \mathbb{R}^q_+$ and D is a convex set while $g_i(x), i = 1, \ldots, q$ are convex functions, problem (9.35) is a parametric convex program ($t \geq 0$ because $-K^\circ \subset \mathbb{R}^q_+$).

Example 9.8 Consider the problem (Tanaka et al. 1991):

$$\min_x \left\{ \left(\sum_{i=1}^n f_{1,i}(x_i) \right) \cdot \left(\sum_{i=1}^n f_{2,i}(x_i) \right) \mid \; x_i \in X_i, \; i = 1, \ldots, n \right\} \tag{9.37}$$

where $X_i \subset \mathbb{R}_{++}$ and $f_{1,i}, f_{2,i}$ are positive-valued functions ($i = 1 \ldots, n$). Applying the Corollary for $g(x) = (g_1(x), g_2(x)), \; g_1(x) = \sum_{i=1}^n f_{1,i}(x_i), \; g_2(x) = \sum_{i=1}^n f_{2,i}(x_i)$, $F(y) = y_1 y_2$ for every $y = (y_1, y_2) \in \mathbb{R}^2_{++}$, we see that if this problem is solvable then there exists a number $t > 0$ such that any optimal solution of

$$\min_x \left\{ \sum_{i=1}^n (f_{1,i}(x_i) + t f_{2,i}(x_i)) | \; x_i \in X_i, \; i = 1, \ldots, n \right\} \tag{9.38}$$

will solve (9.37). For fixed $t > 0$ the problem (9.38) splits into n one-dimensional subproblems

$$\min\{f_{1,i}(x_i) + t f_{2,i}(x_i) | \; x_i \in X_i\}.$$

In the particular case of the problem of optimal ordering policy for jointly replenished products as treated in the mentioned paper of Tanaka et al. $f_{1,i}(x_i) = a/n + a_i/x_i, f_{2,i}(x_i) = b_i x_i/2$ ($a, a_i, b_i > 0$) and X_i is the set of positive integers, so each of the above subproblems can easily be solved.

9.5 Linear Multiplicative Programming

In this section we discuss parametric methods for solving problem (QCM) under assumptions (9.30). Let Q be the matrix of rows $c^1, \ldots, c^r$. In view of Proposition 9.1, the problem can be reformulated as

$$\min\{F(\langle c^1, x\rangle, \ldots, \langle c^r, x\rangle) | \; x \in D\}, \tag{9.39}$$

where D is a polyhedron in $\mathbb{R}^n$, and $F : \mathbb{R}^r \to \mathbb{R}$ is a quasiconcave function on a closed convex set $Y \supset Q(D)$, such that for any $y, y' \in Y$:

$$\left. \begin{array}{l} y_i \geq y'_i \; (i = 1, \ldots, p) \\ y_i = y'_i \; (i = p+1, \ldots, r) \end{array} \right\} \Rightarrow F(y) \geq F(y'). \tag{9.40}$$

A typical problem of this class is the *linear multiplicative program* :

$$(LMP) \qquad \min\left\{\prod_{j=1}^{r}\langle c^i, x\rangle \mid \quad Ax = b, x \geq 0\right\} \tag{9.41}$$

which corresponds to the case $F(y) = \prod_{i=1}^{r} y_i$. Aside from ($LMP$), a wide variety of other problems can be cast in the form (9.39), as shown by the many examples in Sect. 7.1. In spite of that, it turns out that solving (9.39) with different functions $F(y)$ reduces to comparing the objective function values on a finite set of extreme points of D dependent upon the vectors $c^1, \ldots, c^r$ but not on the specific form of $F(y)$.

Due to their relevance to applications in various fields (multiobjective programming, bond portfolio optimization, VLSI design, etc…), the above class of problems, including (LMP) and its extensions, has been the subject of quite a few research. Parametric methods for solving (LMP) date back to Swarup (1966a), Swarup (1966b), Swarup (1966c), Forgo (1975), and Gabasov and Kirillova (1980). However intensive development in the framework of global optimization began only with the works of Konno and Kuno (1990, 1992). The basic idea of the parametric method can be described as follows. By Corollary 9.1 one can solve (LMP) by solving the problem

$$\min\{f(x^{\lambda,\alpha}) \mid \ \lambda \in \Lambda, \alpha \in W\}, \tag{9.42}$$

where, in the notation in Corollary 9.1, $x^{\lambda,\alpha}$ is an arbitrary optimal solution of the linear program

$$LP(\lambda,\alpha) \qquad \left|\begin{array}{l} \min\ \sum_{i=1}^{p} \lambda_i \langle c^i, x\rangle \\ \text{s.t.}\ \ x \in D,\ \langle c^j, x\rangle = \alpha_j,\ j = p+1, \ldots, r. \end{array}\right.$$

For solving (9.42) the parametric method exploits the property of $LP(\lambda,\alpha)$ that the parameter space is partitioned in finitely many polyhedrons over each of which the minimum of the function $\varphi(\lambda,\alpha) = f(x^{\lambda,\alpha})$ can be computed relatively easily. Specifically, by writing the constraints in this linear program in the form

$$Ax = b + S\alpha, \quad x \geq 0, \tag{9.43}$$

where S is a suitable diagonal matrix, denote the collection of all basis matrices of this system by $\mathscr{B}$.

Lemma 9.3 *Each basis B of the system (9.43) determines a set $\Pi_B \times \Delta_B \subset \Lambda \times W$ which is a polyhedron whenever nonempty, and an affine map $x^B : \Delta_B \to \mathbb{R}^n$, such that $x^B(\alpha)$ is a basic optimal solution of $LP(\lambda,\alpha)$ for all $(\lambda,\alpha) \in \Pi_B \times \Delta_B$. The collection of polyhedrons $\Pi_B \times \Delta_B$ corresponding to all bases $B \in \mathscr{B}$ covers all (λ,α) such that $LP(\lambda,\alpha)$ has an optimal solution.*

Proof Let B be any basis of this system. If $A = (B, N), x = \begin{pmatrix} x_B \\ x_N \end{pmatrix}$, so that $Bx_B + Nx_N = b + S\alpha$, then

$$x_B = B^{-1}(b + S\alpha) - B^{-1}Nx_N \tag{9.44}$$

and the associated basic solution of (9.43) is

$$x_B = B^{-1}(b + S\alpha),\ x_N = 0. \tag{9.45}$$

This basic solution is feasible for all $\alpha \in W$ such that $B^{-1}(b + S\alpha) \geq 0$ (nonnegativity condition) and is dual feasible for all $\lambda \in \Lambda$ such that $\sum_{i=1}^{p} \lambda_i[(c_N^i)^T - (c_B^i)^T B^{-1}N] \geq 0$ (optimality criterion), where $c^i = \begin{pmatrix} c_B^i \\ c_N^i \end{pmatrix}$. If we denote the vector (9.45) by $x^B(\alpha)$ and define the polyhedrons

$$\varPi_B = \left\{ \lambda \in \Lambda |\ \sum_{i=1}^{p} \lambda_i \big[(c_N^i)^T - (c_B^i)^T B^{-1}N\big] \geq 0 \right\}$$
$$\varDelta_B = \{\alpha \in W |\ B^{-1}(b + S\alpha) \geq 0\}$$

then whenever both $\varPi_B$ and Λ_B are nonempty the affine map $\alpha \in \varDelta_B \mapsto x^B(\alpha) \in \mathbb{R}^n$ is such that $x^B(\alpha)$ is a basic optimal solution of LP(λ, α) for all $\lambda \in \varPi_B,\ \alpha \in \varDelta_B$. The last assertion of the Lemma is obvious from the fact that if LP(λ, α) has a basic optimal solution, then $(\lambda, \alpha) \in \varPi_B \times \varDelta_B$ for the corresponding basis B. □

Each polyhedron $\varPi_B \times \varDelta_B$ is called a *cell*. Thus, by taking $x^{\lambda,\alpha} = x^B(\alpha)$ for every $(\lambda, \alpha) \in \varPi_B \times \varDelta_B$, the function $\varphi(\lambda, \alpha) = f(x^{\lambda,\alpha})$ is constant in $\lambda \in \varPi_B$ for every fixed $\alpha \in \varDelta_B$ and quasiconcave in $\alpha \in \varDelta_B$ for every fixed $\lambda \in \varPi_B$. Hence its minimum over the cell $\varPi_B \times \varDelta_B$ is equal to the minimum of $f(x^B(\alpha))$ over $\varDelta_B$ and is achieved at a vertex of the polytope $\varDelta_B$. Therefore, if the collection of cells can be effectively computed then the problem (9.42) can be solved through scanning the vertices of the cells. Unfortunately, however, when $r > 3$ the cells are very complicated and there is in general no practical method for computing these cells, except in rare cases where additional structure (like network constraints, see Sect. 9.8) may simplify this computation. Below we briefly examine the method for problems with $r \leq 3$. For more detail the reader is referred to the cited papers of Konno and Kuno.

9.5.1 Parametric Objective Simplex Method

When $p = r = 2$ the problem (9.39) is

$$\min\{F(\langle c^1, x\rangle, \langle c^2, x\rangle)|\ x \in D\} \tag{9.46}$$

where D is a polyhedron contained in the set $\{x|\ \langle c^1, x\rangle \geq 0, \langle c^2, x\rangle \geq 0\}$, and $F(y)$ is a quasiconcave function on the orthant $\mathbb{R}^2_+$ such that

$$y, y' \in \mathbb{R}^2_+, y \geq y' \Rightarrow F(y) \geq F(y'). \tag{9.47}$$

As seen earlier, aside from the special case when $F(y) = y_1 y_2$ (linear multiplicative program), the class (9.46) includes problems with quite different objective functions, such as

$$(\langle c^1, x\rangle)^{q_1}(\langle c^2, x\rangle)^{q_2};\ \ -q_1 e^{\langle c^1,x\rangle} - q_2 e^{\langle c^2,x\rangle}\ (q_1 > 0, q_2 > 0)$$

corresponding to different choices of $F(y)$:

$$y_1^{q_1} y_2^{q_2};\ -q_1 e^{y_1} - q_2 e^{y_2}.$$

By Corollary 9.1 (with $p = r = 2$) the parametric linear program associated with (9.46) is

$$LP(\lambda) \qquad \min\{\langle c^1 + \lambda c^2, x\rangle|\ x \in D\}, \quad \lambda \geq 0. \tag{9.48}$$

Since the parameter domain Λ is the nonnegative real line, the cells of this parametric program are intervals and can easily be determined by the standard *parametric objective simplex method* (see, e.g., Murty 1983) which consists in the following.

Suppose a basis B_k is already available such that the corresponding basic solution x^k is optimal to LP(λ) for all $\lambda \in \Pi_k := [\lambda_{k-1}, \lambda_k]$ but not for λ outside this interval. The latter means that if λ is slightly greater than λ_k then at least one of the inequalities in the optimality criterion

$$(c^1 + \lambda c^2)_{N_k} - (c^1 + \lambda c^2)_{B_k} B_k^{-1} N_k \geq 0$$

becomes violated. Therefore, by performing a number of simplex pivot steps we will pass to a new basis B_{k+1} with a basic solution x^{k+1} which will be optimal to LP(λ) for all λ in a new interval $\Pi_{k+1} \subset [\lambda_k, +\infty)$. In this way, starting from the value $\lambda_0 = 0$ one can determine the interval $\Pi_1 = [\lambda_0, \lambda_1]$ then pass to the next $\Pi_2 = [\lambda_1, \lambda_2]$, and so on, until the whole halfline is covered. Note that the optimum objective value function $\xi(\lambda)$ in the linear program LP(λ) is a concave piecewise affine function of λ with Π_k as linearity intervals. The points $\lambda_0 = 0, \lambda_1, \ldots, \lambda_l$ are the *breakpoints* of $\xi(\lambda)$. An optimal solution to (9.46) is then x^{k*} where

$$k* \in \operatorname{argmin}\{F(\langle c^1, x^k\rangle, \langle c^2, x^k\rangle)|\ k = 1, 2, \ldots, l\}. \tag{9.49}$$

Remark 9.4 Condition (9.47) is a variant of (9.40) for $p = r = 2$. Condition (9.40) with $p = 1, r = 2$ always holds since it simply means $F(y) = F(y')$ whenever

$y = y'$. Therefore, without assuming (9.47), Corollary 9.1 and hence the above method still applies. However, in this case, $\Lambda = \mathbb{R}$, i.e., the linear program $LP(\lambda)$ should be considered for all $\lambda \in (-\infty, +\infty)$.

9.5.2 Parametric Right-Hand Side Method

Consider now problem (9.39) under somewhat weaker assumptions than previously, namely: $F(y)$ is not required to be quasiconcave, while condition (9.47) is replaced by the following weaker one:

$$y_1 \geq y'_1,\ y_2 = y'_2 \quad \Rightarrow \quad F(y) \geq F(y'). \tag{9.50}$$

The latter condition implies that for fixed y_2 the univariate function $F(., y_2)$ is monotonic on the real line. It is then not hard to verify that Corollary 9.1 still holds for this case. In fact this is also easy to prove directly.

Lemma 9.4 *Let* $\underline{\alpha} = \min\{\langle c^2, x\rangle \mid x \in D\}$, $\overline{\alpha} = \max\{\langle c^2, x\rangle \mid x \in D\}$ *and for every* $\alpha \in [\underline{\alpha}, \overline{\alpha}]$ *let* x^α *be a basic optimal solution of the linear program*

$$LP(\alpha) \qquad \min\{\langle c^1, x\rangle \mid\ x \in D,\ \langle c^2, x\rangle = \alpha\} \qquad .$$

Then

$$\min\{F(\langle c^1, x\rangle, \langle c^2, x\rangle) \mid\ x \in D\} = \min\{F(\langle c^1, x^\alpha\rangle, \langle c^2, x^\alpha\rangle) \mid\ \underline{\alpha} \leq \alpha \leq \overline{\alpha}\}.$$

Proof For every feasible solution x of $LP(\alpha)$ we have $\langle c^1, x - x^\alpha\rangle \geq 0$ (because x^α is optimal to LP(α)) and $\langle c^2, x - x^\alpha\rangle = 0$ (because x^α is feasible to LP(α)). Hence, from (9.50) $F(\langle c^1, x\rangle, \langle c^2, x\rangle) \geq F(\langle c^1, x^\alpha\rangle, \langle c^2, x^\alpha\rangle)$ and the conclusion follows. □

Solving (9.39) thus reduces to solving LP(α) parametrically in α. This can be done by the following standard *parametric right-hand side method* (see, e.g., Murty 1983). Assuming that the constraints in LP(α) have been rewritten in the form

$$Ax = b + \alpha b_0,\ x \geq 0,$$

let B_k be a basis of this system satisfying the dual feasibility condition

$$c^1_{N_k} - c^1_{B_k} B_k^{-1} N_k \geq 0.$$

Then the basic solution defined by this basis, i.e., the vector $x^k = u^k + \alpha v^k$ such that $x^k_{B_k} = B_k^{-1}(b + \alpha b_0), x^k_{N_k} = 0$ [see (9.45)], is dual feasible. This basic solution x^k is optimal to LP(α) if and only if it is feasible, i.e., if and only if it satisfies the nonnegativity condition

$$B_k^{-1}(b + \alpha b_0) \geq 0.$$

Let $\Delta_k := [\alpha_{k-1}, \alpha_k]$ be the interval formed by all values α satisfying the above inequalities. Then x^k is optimal to LP(α) for all α in this interval, but for any value α slightly greater than α_k at least one of the above inequalities becomes violated, i.e., at least one of the components of x^k becomes negative. By performing then a number of dual simplex pivots, we can pass to a new basis B_{k+1} with a basic solution x^{k+1} which will be optimal for all α in some interval $\Delta_{k+1} = [\alpha_k, \alpha_{k+1}]$. Thus, starting from the value $\alpha_0 = \underline{\alpha}$, one can determine an initial interval $\Delta_1 = [\underline{\alpha}, \alpha_1]$, then pass to the next interval $\Delta_2 = [\alpha_1, \alpha_2]$, and so on, until the whole interval $[\underline{\alpha}, \overline{\alpha}]$ is covered. The generated points $\alpha_0 = \underline{\alpha}, \alpha_1, \ldots, \alpha_h = \overline{\alpha}$ are the breakpoints of the optimum objective value function $\eta(\alpha)$ in LP(α) which is a convex piecewise affine function. By Lemma 9.4 an optimal solution to (9.39) is then x^{k*} where

$$k^* \in \operatorname{argmin}\{F(\langle c^1, x^k\rangle, \langle c^2, x^k\rangle) |\ k = 1, 2, \ldots, h\}. \tag{9.51}$$

Remark 9.5 In both objective and right-hand side methods the sequence of points x^k on which the objective function values have to be compared depends upon the vectors c^1, c^2 but not on the specific function $F(y)$. Computational experiments (Konno and Kuno 1992), also (Tuy and Tam 1992) indicate that solving a problem (9.39) requires no more effort than solving just a few linear programs of the same size. Two particular problems of this class are worth mentioning:

$$\min\{c^1x.c^2x| \quad x \in D\} \tag{9.52}$$

$$\min\{-x_1^2 + c_0x| \quad x \in D\}, \tag{9.53}$$

where D is a polyhedron [contained in the domain $c^1x > 0, c^2x > 0$ for (9.52)]. These correspond to functions $F(y) = y_1y_2$ and $F(y) = -y_1^2 + y_2$, respectively and have been shown to be NP-hard (Matsui 1996; Pardalos and Vavasis 1991), although, as we saw above, they are efficiently solved by the parametric method.

9.5.3 Parametric Combined Method

When $r = 3$ the parametric linear program associated to (9.39) is difficult to handle because the parameter domain is generally two-dimensional. In some particular cases, however, a problem (9.39) with $r = 3$ may be successfully tackled by the parametric method. As an example consider the problem

$$\min\{c^0x + c^1x.c^2x| \quad x \in D\} \tag{9.54}$$

which may not belong to the class (QCM) because we cannot prove the quasiconcavity of the function $F(y) = y_1 + y_2y_3$. It is easily seen, however, that this problem is equivalent to

$$\min\{\langle c^0 + \alpha c^1, x\rangle| \quad x \in D,\ \langle c^2, x\rangle = \alpha,\ \ \underline{\alpha} \le \alpha \le \overline{\alpha}\} \tag{9.55}$$

where $\underline{\alpha} = \min_{x\in D}\langle c^2, x\rangle$, $\overline{\alpha} = \max_{x\in D}\langle c^2, x\rangle$. By writing the constraints $x \in D$, $\langle c^2, x\rangle = \alpha$ in the form

$$Ax = b + \alpha b_0. \tag{9.56}$$

we can prove the following:

Lemma 9.5 *Each basis B of the system (9.56) determines an interval $\Delta_B \subset [\underline{\alpha}, \overline{\alpha}]$ and an affine map $x^B : \Delta_B \to \mathbb{R}^n$ such that $x^B(\alpha)$ is a basic optimal solution of the problem*

$$P(\alpha) \qquad\qquad \min\{\langle c^0 + \alpha c^1, x\rangle | \quad x \in D,\ \langle c^2, x\rangle = \alpha\}$$

for all $\alpha \in \Delta_B$. The collection of all intervals Δ_B corresponding to all bases B such that $\Delta_B \neq \emptyset$ covers all α for which $P(\alpha)$ has an optimal solution.

Proof Let B be a basis, $A = (B, N), x = \begin{pmatrix} x_B \\ x_N \end{pmatrix}$, so that the associated basic solution is

$$x_B = B^{-1}(b + \alpha b_0), \quad x_N = 0. \tag{9.57}$$

This basic solution is feasible for all $\alpha \in [\underline{\alpha}, \overline{\alpha}]$ satisfying the nonnegativity condition

$$B^{-1}(b + \alpha b_0) \geq 0. \tag{9.58}$$

and is dual feasible for all α satisfying the dual feasibility condition (optimality criterion)

$$(c^0 + \alpha c^1)_N - (c^0 + \alpha c^1)_B B^{-1} N \geq 0. \tag{9.59}$$

Conditions (9.58) and (9.59) determine each an interval. If the intersection Δ_B of the two intervals is nonempty, then the basic solution (9.57) is optimal for all $\alpha \in \Delta_B$. The rest is obvious. □

Based on this Lemma, one can generate all the intervals Δ_B corresponding to different bases B by a procedure combining primal simplex pivots with dual simplex pivots (see, e.g., Murty 1983). Specifically, for any given basis B_k, denote by $u^k + \alpha v^k$ the vector (9.57) which is the basic solution of P(α) corresponding to this basis, and by $\Delta_k = [\alpha_{k-1}, \alpha_k]$ the interval of optimality of this basic solution. When α is slightly greater than α_k either the nonnegativity condition [see (9.58)] or the dual feasibility condition [see (9.59)] becomes violated. In the latter case we perform a primal simplex pivot step, in the former case a dual simplex pivot step, and repeat as long as necessary, until a new basis B_{k+1} with a new interval Δ_{k+1} is obtained. Thus, starting from an initial interval $\Delta_1 = [\underline{\alpha}, \alpha_1]$ we can pass from one interval to the next, until we reach $\overline{\alpha}$. Let then $\Delta_k, k = 1, \ldots, l$ be the collection of all intervals. For each interval Δ_k let $u^k + \alpha v^k$ be the associated basic solution which is optimal to P(α) for all $\alpha \in \Delta_k$ and let $x^k = u^k + \xi_k v^k$, where ξ_k is a minimizer of the quadratic function $\varphi(\alpha) := \langle c^0 + \alpha c^1, u^k + \alpha v^k \rangle$ over Δ_k. Then from (9.55) and by Lemma 9.5, an optimal solution of (9.54) is x^{k*} where

$$k^* \in \operatorname{argmin}\{c^0 x^k + (c^1 x^k)(c^2 x^k)| \ k = 1, 2, \ldots, l\}.$$

Note that the problem (9.54) is also NP-hard (Pardalos and Vavasis 1991).

9.6 Convex Constraints

So far we have been mainly concerned with solving quasiconcave monotonic minimization problems under linear constraints. Consider now the problem

$$(QCM) \qquad \min\{f(x)| \quad x \in D\}. \tag{9.60}$$

formulated in Sect. 9.2 [see (9.8)], where the constraint set D is compact convex but not necessarily polyhedral. Recall that by Proposition 9.1 where $Qx = (c^1 x, \ldots, c^r x)^T$ there exists a continuous quasiconcave function $F : \mathbb{R}^r \to \mathbb{R}$ on a closed convex set $Y \supset Q(D)$ satisfying condition (9.40) and such that

$$f(x) = F(\langle c^1, x\rangle, \ldots, \langle c^r, x\rangle). \tag{9.61}$$

It is easy to extend the branch and bound algorithms in Sect. 8.2 and the polyhedral annexation algorithm in Sect. 8.3 to (QCM) with a compact convex constraint set D, and in fact, to the following more general problem:

$$(GQCM) \qquad \min\{F(g(x))| \quad x \in D\} \tag{9.62}$$

where D and $F(y)$ are as previously (with $g(D)$ replacing $Q(D)$), while $g = (g_1, \ldots, g_r)^T : \mathbb{R}^n \to \mathbb{R}^r$ is a map such that $g_i(x), i = 1, \ldots, p$, are convex on a closed convex set $X \supset D$ and $g_i(x) = \langle c^i, x\rangle + d_i, \ i = p+1, \ldots, r$.

For the sake of simplicity we shall focus on the case $p = r$, i.e., we shall assume $F(y) \geq F(y')$ whenever $y, y' \in Y, y \geq y'$. This occurs, for instance, for the convex multiplicative programming problem

$$\min\left\{\prod_{i=1}^{r} g_i(x)| \ x \in D\right\} \tag{9.63}$$

where $F(y) = \prod_{i=1}^{r} y_i,\ D \subset \{x|\ g_i(x) \geq 0\}$. Clearly problem (GQCM) can be rewritten as

$$\min\{F(y)|\ \ g(x) \leq y, x \in D\} \tag{9.64}$$

which is to minimize the quasiconcave $\mathbb{R}^r_+$-monotonic function $F(y)$ on the closed convex set

$$E = \{y|g(x) \leq y, x \in D\} = g(D) + \mathbb{R}^r_+. \tag{9.65}$$

9.6.1 Branch and Bound

If the function $F(y)$ is concave (and not just quasiconcave) then simplicial subdivision can be used. In this case, a lower bound for $F(y)$ over the feasible points y in a simplex $M = [u^1, \ldots, u^{r+1}]$ can be obtained by solving the convex program

$$\min\{l_M(y)|\ \ y \in E \cap M\},$$

where $l_M(y)$ is the affine function that agrees with $F(y)$ at the vertices of M (this function satisfies $l_M(y) \leq F(y)\ \forall y \in M$ in view of the concavity of $F(y)$). Using this lower bounding rule, a simplicial algorithm for (GQCM) can be formulated following the standard branch and bound scheme. It is also possible to combine branch and bound with outer approximation of the convex set E by a sequence of nested polyhedrons $P_k \supset E$.

If the function $F(y)$ is concave separable, i.e., $F(y) = \sum_{j=1}^{r} F_j(y_j)$, where each $F_j(.)$ is a concave function of one variable, then rectangular subdivision can be more convenient. However, in the general case, when $F(y)$ is quasiconcave but not concave, an affine minorant of a quasiconcave function over a simplex (or a rectangle) is in general not easy to compute. In that case, one should use conical rather than simplicial or rectangular subdivision.

9.6.2 Polyhedral Annexation

Since $F(y)$ is monotonic with respect to the orthant $\mathbb{R}^r_+$, we can apply the PA Algorithm using Lemma 9.2 for constructing the initial polytope P_1.

Let $\overline{y} \in E$ be an initial feasible point of (9.64) and $C = \{y|\ F(y) \geq F(\overline{y})\}$. By translating, we can assume that $0 \in E$ and $0 \in \text{int}C$, i.e., $F(0) > F(\overline{y})$. Note that instead of a linear program like (9.25) we now have to solve a convex program of the form

$$\max\{\langle t, y\rangle|\ \ y \in E\} = \max\{\langle t, g(x)\rangle|\ \ x \in D\} \tag{9.66}$$

where $t \in (\mathbb{R}^r_+)^\circ = -\mathbb{R}^r_+$. Therefore, to construct the initial polytope P_1 (see Lemma 9.2) we compute $\alpha > 0$ such that $F(\alpha w) = F(\overline{y})$ (where $w = (-1,\ldots,-1)^T \in \mathbb{R}^r$) and take the initial polytope P_1 to be the r-simplex in $\mathbb{R}^r$ with vertex set $V_1 = \{0, -e^1/\alpha, \ldots, -e^r/\alpha\}$. We can thus state:

PA Algorithm (for (*GQCM*))

Initialization. Let $\overline{y}$ be a feasible solution (the best available), $C = \{y|\, F(y) \geq F(\overline{y})\}$. Choose a point $y^0 \in E$ such that $F(y^0) > F(\overline{y})$ and set $E \leftarrow E - y^0,\ C \leftarrow C - y^0$. Let P_1 be the simplex in $\mathbb{R}^r$ with vertex set V_1. Set $k = 1$.

Step 1. For every new $t = (t_1, \ldots, t_r)^T \in V_k \setminus \{0\}$ solve the convex program (9.66), obtaining the optimal value $\mu(t)$ and an optimal solution $y(t)$.

Step 2. Let $t^k \in \operatorname{argmax}\{\mu(t)|\ t \in V_k\}$. If $\mu(t^k) \leq 1$, then terminate: $\overline{y}$ is an optimal solution of (*GQCM*).

Step 3. If $\mu(t^k) > 1$ and $y^k := y(t^k) \notin C$ (i.e., $F(y^k) < F(\overline{y})$), then update the current best feasible solution and the set C by resetting $\overline{y} = y^k$.

Step 4. Compute

$$\theta_k = \sup\{\theta|\ \ F(\theta y^k) \geq F(\overline{y})\} \tag{9.67}$$

and define

$$P_{k+1} = P_k \cap \{t|\ \langle t, y^k\rangle \leq \frac{1}{\theta_k}\}.$$

Step 5. From V_k derive the vertex set V_{k+1} of P_{k+1}. Set $k \leftarrow k+1$ and go back to Step 1.

To establish the convergence of this algorithm we need two lemmas.

Lemma 9.6 $\mu(t^k) \searrow 1$ *as* $k \to +\infty$.

Proof Observe that $\mu(t) = \max\{\langle t, y\rangle |\ y \in g(D)\}$ is a convex function and that $y^k \in \partial\mu(t^k)$ because $\mu(t) - \mu(t^k) \geq \langle t, y^k\rangle - \langle t^k, y^k\rangle = \langle t - t^k, y^k\rangle\ \ \forall t$. Denote $l_k(t) = \langle t, y^k\rangle - 1$. We have $l_k(t^k) = \langle t^k, y^k\rangle - 1 = \mu(t^k) - 1 > 0$, and for $t \in [g(D)]^\circ$: $l_k(t) = \langle t, y^k\rangle - 1 \leq \mu(t) - 1 \leq 0$ because $y^k \in g(D) + \mathbb{R}^r_+$. Since $g(D)$ is compact, there exist $t^0 \in \operatorname{int}[g(D]^\circ$ (Proposition 1.21), i.e., t^0 such that $l(t^0) = \mu(t^0) - 1 < 0$ and $s^k \in [t^0, t^k)] \setminus \operatorname{int}[g(D)]^\circ$ such that $l(s^k) = 0$. By Theorem 6.1 applied to the set $[g(D)]^\circ$, the sequence $\{t^k\}$ and the cuts $l_k(t)$, we conclude that $t^k - s^k \to 0$. Hence every accumulation point t^* of $\{t^k\}$ satisfies $\mu(t^*) = 1$, i.e., $\mu(t^k) \searrow \mu(t^*) = 1$. □

Denote by C_k the set C at iteration k.

Lemma 9.7 $C_k^\circ \subset P_k$ *for every* k.

Proof The inclusion $C_1^\circ \subset P_1$ follows from the construction of P_1. Arguing by induction on k, suppose that $C_k^\circ \subset P_k$ for some $k \geq 1$. Since $\theta_k y^k \in C_{k+1}$, for all $t \in C_{k+1}^\circ$ we must have $\langle t, \theta_k y^k \rangle \leq 1$, and since $C_k \subset C_{k+1}$, i.e., $C_{k+1}^\circ \subset C_k^\circ \subset P_k$ it follows that $t \in P_k \cap \{t | \langle t, \theta_k y^k \rangle \leq 1\} = P_{k+1}$. □

Proposition 9.5 *Let $\overline{y}^k$ be the incumbent at iteration k. Either the above algorithm terminates by an optimal solution of (GQCM) or it generates an infinite sequence $\{\overline{y}^k\}$ every accumulation point of which is an optimal solution.*

Proof Let y^* be any accumulation point of the sequence $\{\overline{y}^k\}$. Since by Lemma 9.6 $\max\{\mu(t)| \ t \in P_k\} \searrow 1$ it follows from Lemma 9.7 that $\max\{\mu(t)| \ t \in \cap_{k=1}^\infty C_k^\circ\} \leq 1$. This implies that $\cap_{k=1}^\infty C_k^\circ \subset E^\circ$, and hence $E \subset \cup_{k=1}^\infty C_k$. Thus, for any $y \in E$ there is k such that $y \in C_k$, i.e., $F(y) \geq F(\overline{y}^k) \geq F(\overline{y})$, proving the optimality of $\overline{y}$. □

Incidentally, the above PA Algorithm shows that

$$\min\{F(y)| \ \ y \in E\} = \min\{F(g(x(t)))| \ \ t \in \mathbb{R}_-^r\}, \tag{9.68}$$

where $x(t)$ is an arbitrary optimal solution of (9.66), i.e., of the convex program

$$\min\{\langle -t, g(x)\rangle| \ \ x \in D\}.$$

This result could also be derived from Theorem 9.1.

Remark 9.6 All the convex programs (9.66) have the same constraints. This fact should be exploited for an efficient implementation of the above algorithm. Also, in practice, these programs are usually solved approximately, so $\mu(t)$ is only an ε-optimal value, i.e., $\mu(t) \geq \max\{\langle t, y\rangle | \ y \in E\} - \varepsilon$ for some tolerance $\varepsilon > 0$.

9.6.3 Reduction to Quasiconcave Minimization

An important class of problems (GQCM) is constituted by problems of the form:

$$\text{minimize} \sum_{i=1}^{p} \prod_{j=1}^{q_i} g_{ij}(x) \quad \text{s.t} \quad x \in D, \tag{9.69}$$

where D is a compact convex set and $g_{ij} : D \to \mathbb{R}_{++}, i = 1, \ldots, p, j = 1, \ldots, q_i$, are continuous convex positive-valued functions on D. This class includes *generalized convex multiplicative programming problems* (Konno and Kuno 1995; Konno et al. 1994) which can be formulated as:

$$\min \left\{ g(x) + \prod_{j=1}^{q} g_j(x) | \ x \in D \right\}. \tag{9.70}$$

or, alternatively as

$$\min \left\{ g(x) + \sum_{i=1}^{p} g_i(x)h_i(x) | \ x \in D \right\}, \tag{9.71}$$

where all functions $g(x), g_i(x), h_i(x)$ are convex positive-valued on D. Another special case worth mentioning is the problem of minimizing the scalar product of two vectors:

$$\min \left\{ \sum_{i=1}^{n} x_i y_i | \ (x, y) \in D \subset \mathbb{R}^n_{++} \right\}. \tag{9.72}$$

Let us show that any problem (9.69) can be reduced to concave minimization under convex constraints. For this we rewrite (9.69) as

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{p} (\prod_{j=1}^{q_i} y_{ij})^{1/q_i} \\ \text{s.t.} & [g_{ij}(x)]^{q_i} \le y_{ij}, \ i = 1, \ldots, p; j = 1, \ldots, q_i \\ & y_{ij} \ge 0, \quad x \in D. \end{array} \tag{9.73}$$

Since $\varphi_j(y) = y_{ij}, j = 1, \ldots, q_i$ are linear, it follows from Proposition 2.7 that each term $(\prod_{j=1}^{q_i} y_{ij})^{1/q_i}$, $i = 1, \ldots, p$, is a concave function of $y = (y_{ij}) \in \mathbb{R}^{q_1 + \ldots, q_p}$, hence their sum, i.e., the objective function $F(y)$ in (9.73), is also a concave function of y. Furthermore, $y \ge y'$ obviously implies $F(y) \ge F(y')$, so $F(y)$ is monotonic with respect to the cone $\{y | y \ge 0\}$. Finally, since every function $g_{ij}(x)$ is convex, so is $[g_{ij}(x)]^{q_i}$. Consequently, (9.73) is a concave programming problem. If $q = q_1 + \ldots + q_p$ is relatively small this problem can be solved by methods discussed in Chaps. 5 and 6.

An alternative way to convert (9.69) into a concave program is to use the following relation already established in the proof of Proposition 2.7 [see (2.6)]:

$$\prod_{j=1}^{q_i} g_{ij}(x) = \frac{1}{q_i} \min_{\xi^i \in T_i} \sum_{j=1}^{q_i} \xi_{ij} g_{ij}^{q_i}(x),$$

with $\xi^i = (\xi_{i1}, \ldots, \xi_{iq_i}), T_i = \{\xi^i : \prod_{j=1}^{q_i} \xi_{ij} \ge 1\}$. Hence, setting $\varphi_i(\xi^i, x) = \frac{1}{q_i} \sum_{j=1}^{q_i} \xi_{ij} g_{ij}^{q_i}(x)$, the problem (9.69) is equivalent to

$$\begin{aligned}
&\min_{x\in D}\sum_{i=1}^{p}\min_{\xi^i\in T_i}\varphi_i(\xi^i,x)\\
&=\min_{x\in D}\min\{\sum_{i=1}^{p}\varphi_i(\xi^i,x)\ :\ \xi^i\in T_i, i=1,\dots,p\}\\
&=\min\{\min_{x\in D}\sum_{i=1}^{p}\varphi_i(\xi^i,x)\ :\ \xi^i\in T_i, i=1,\dots,p\}\\
&=\min\{\varphi(\xi^1,\dots,\xi^p)|\ \xi^i\in T_i, i=1,\dots,p\}
\end{aligned}\tag{9.74}$$

where

$$\varphi(\xi^1,\dots,\xi^p)=\min_{x\in D}\sum_{i=1}^{p}\varphi_i(\xi^i,x)=\min\left\{\sum_{i=1}^{p}\sum_{j=1}^{q_i}\frac{1}{q_i}\xi_{ij}g_{ij}^{q_i}(x)|\ x\in D\right\}.$$

Since each function $g_{ij}^{q_i}(x)$ is convex, $\varphi(\xi^1,\dots,\xi^p)$ is the optimal value of a convex program. Furthermore, the objective function of this convex program is a linear function of $(\xi^1,\dots,\xi^p)$ for fixed x. Therefore, $\varphi(\xi^1,\dots,\xi^p)$ is a concave function (pointwise minimum of a family of linear functions). Finally, each $T_i, i=1,\dots,p$, is a closed convex set in $\mathbb{R}^{q_i}$ because

$$T_i=\left\{\xi^i|\ \sum_{j=1}^{q_i}\log\xi_{ij}\geq 0\right\}.$$

Thus, problem (9.69) is equivalent to (9.74) which seeks to minimize the concave function $\varphi(\xi^1,\dots,\xi^p)$ over the convex set $\prod_{i=1}^{p}T_i$.

Example 9.9 As an illustration, consider the problem of determining a rectangle of minimal area enclosing the projection of a given compact convex set $D\subset\mathbb{R}^n (n>2)$ onto the plane $\mathbb{R}^2$. This problem arises from applications in packing and optimal layout (Gale 1981; Haims and Freeman 1970; Maling et al. 1982), especially when two-dimensional layouts are restricted by $n-2$ factors. When the vertices of D are known, there exists an efficient algorithm based on computational geometry (Bentley and Shamos 1978; Graham 1972; Toussaint 1978). In the general case, however, the problem is more complicated.

Assume that D has full dimension in $\mathbb{R}^n=\mathbb{R}^2\times\mathbb{R}^{n-2}$. The projection of D onto $\mathbb{R}^2$ is the compact convex set pr$D=\{y\in\mathbb{R}^2|\ \exists z\in\mathbb{R}^{n-2},\ (y,z)\in D\}$. For any vector $x\in\mathbb{R}^2$ such that $\|x\|=1$ define

$$g_1(x)=\max\{\langle x,y\rangle|\ (y,z)\in D\}$$
$$g_2(x)=\min\{\langle x,y\rangle|\ (y,z)\in D\}.$$

Then the width of pr$D$ in the direction x is

$$g(x) = g_1(x) - g_2(x).$$

On the other hand, since $Hx = (-x_2, x_1)^T$ is the vector orthogonal to x such that $\|H(x)\| = 1$ if $\|x\| = 1$, the width of pr$D$ in the direction orthogonal to x is $g(Hx) = g(-x_2, x_1)$. Hence the product $g(x).g(Hx)$ measures the area of the smallest rectangle containing pr$D$ and having a side parallel to x. The problem can then be formulated as

$$\min\{g(x).g(Hx)|\, x \in \mathbb{R}^2_+,\ \|x\| = 1\}. \tag{9.75}$$

Lemma 9.8 *The function* $g(x)$ *is convex and satisfies* $g(\alpha x) = \alpha g(x)$ *for any* $\alpha \geq 0$.

Proof Clearly $g_1(x)$ is convex as the pointwise maximum of a family of affine functions, while $g_2(x)$ is concave as the pointwise minimum of a family of affine functions. It is also obvious that $g_i(\alpha x) = \alpha g_i(x), i = 1, 2$. Hence, the conclusion. □

Since $H : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear map, it follows that $g(Hx)$ is also convex and that $g(H(\alpha x)) = \alpha g(Hx)$ for every $\alpha \geq 0$. Exploiting this property, a successive underestimation method for solving problem (9.75) has been proposed in Kuno (1993). In view of the above established property, problem (9.75) is actually equivalent to the convex multiplicative program

$$\min\{g(x) \cdot g(Hx)| \quad x \in \mathbb{R}^2_+, \|x\| \leq 1\}, \tag{9.76}$$

and so can be transformed into the concave minimization problem

$$\begin{array}{ll} \min & \sqrt{t_1 t_2} \\ \text{s.t.} & g(x) \leq t_1,\ g(Hx) \leq t_2 \\ & x \in \mathbb{R}^2_+,\ \|x\| \leq 1, \end{array}$$

or, alternatively,

$$\min\{\varphi(\xi_1, \xi_2)|\, \xi_1\xi_2 \geq 1, \xi \in \mathbb{R}^2_+\}. \tag{9.77}$$

where

$$\varphi(\xi_1, \xi_2) = \min\{\xi_1 g(x) + \xi_2 g(Hx)|\, x \geq 0, \|x\| \leq 1\}. \tag{9.78}$$

In the special case when D is a *polytope* given by a system of linear inequalities $Ay + Bz \leq b$ the problem can be solved by an efficient parametric method (Kuno 1993). For any point $x \in \mathbb{R}^2_+$ with $\|x\| = 1$, let $\sigma(x) = (\lambda, 1 - \lambda)$ be the point where the halfline from 0 through x intersects the line segment joining $(1, 0)$ and $(0, 1)$. Since obviously, $\|\sigma(x)\|^2 = \lambda^2 + (1-\lambda)^2$, using Lemma 9.8 problem (9.76) becomes

$$\min\{F(\lambda)|\ \lambda \in [0, 1]\}, \tag{9.79}$$

where

$$F(\lambda) = \frac{g(\lambda, 1-\lambda).g(\lambda-1, \lambda)}{\lambda^2 + (1-\lambda)^2}.$$

Furthermore, $g(\lambda, 1-\lambda) = g_1(\lambda, 1-\lambda) - g_2(\lambda, 1-\lambda)$ with

$$g_1(\lambda, 1-\lambda) = \max\{\lambda y_1 + (1-\lambda)y_2|\ Ay + Bz \le b\}$$
$$g_2(\lambda, 1-\lambda) = \min\{\lambda y_1 + (1-\lambda)y_2|\ Ay + Bz \le b\}.$$

So each $g_i(\lambda, 1-\lambda), i = 1, 2$ is the optimal value of a linear parametric program. Let $0, \lambda_1^{1i}, \ldots, \lambda_{p_{1i}}^{1i}, 1$ be the sequence of breakpoints of $g_i(\lambda, 1-\lambda), i = 1, 2$. Analogously, let $0, \lambda_1^{2i}, \ldots, \lambda_{p_{2i}}^{2i}, 1$ be the sequence of breakpoints of $g_i(\lambda-1, \lambda)$. Finally let

$$0 = \lambda_0, \lambda_1, \lambda_2, \ldots, \lambda_{N-1}, 1 = \lambda_N \tag{9.80}$$

be the union of all these four sequences rearranged in the increasing order.

Proposition 9.6 *A global minimum of the function $F(\lambda)$ over $[0, 1]$ exists among the points $\lambda_s, s = 0, \ldots, N$.*

Proof For any interval $[\lambda_s, \lambda_{s+1}]$, there exist two vectors $u = x^1 - x^2, v = x^3 - v^4$, where $x^i, i = 1, \ldots, 4$ are vertices of D, such that for every $x \in \mathbb{R}^2_+$ of length $\|x\| = 1$, with $\sigma(x) = (\lambda, 1-\lambda), \lambda \in (\lambda_s, \lambda_{s+1})$, we have $g(x) = \langle x, u\rangle,\ g(Hx) = \langle Hx, v\rangle$. Then

$$F(\lambda) = (\|u\| \cos\alpha) \times (\|v\| \cos\beta),$$

where $\alpha = \text{angle}(x, u)$ (angle between the vectors x and u) and $\beta = \text{angle}(Hx, v)$. Noting that $\text{angle}(x, u) + \text{angle}(u, v) + \text{angle}(v, y) = \text{angle}(x, y)$, we deduce

$$\beta = \alpha + \omega - \pi/2, \tag{9.81}$$

where $\omega = \text{angle}(u, v)$. Thus, for all $\lambda \in (\lambda_s, \lambda_{s+1})$:

$$F(\lambda) = (\|u\| \cos\alpha)(\|v\| \sin(\alpha + \omega)).$$

Setting $G(\alpha) = \cos\alpha \sin(\alpha + \omega)$, we have $G'(\alpha) = -\sin\alpha \sin(\alpha + \omega) + \cos\alpha \cos(\alpha + \omega) = \cos(2\alpha + \omega)$. If $F(.)$ attains a minimum at λ, i.e., $G(.)$ attains a minimum at α, then one must have

$$\cos(2\alpha + \omega) = 0, \text{ hence } 2\alpha + \omega = \pi/2 \pm \pi. \tag{9.82}$$

But from (9.81) it follows that $\alpha + \omega = \beta + \pi/2$, hence $2\alpha + \omega = \alpha + \beta + \pi/2$. In view of (9.82) this in turn implies that

$$\alpha + \beta = \pm\pi,$$

contradicting the fact that $0 < |\alpha| < \pi/2$, $0 < |\beta| < \pi/2$ (because $\langle x, u\rangle > 0$, $\langle Hx, v\rangle > 0$). Therefore, the minimum of $F(\lambda)$ over an interval $[\lambda_s, \lambda_{s+1}]$ cannot be attained at any λ such that $\lambda_s < \lambda < \lambda_{s+1}$, but must be attained at either λ_s or λ_{s+1}. Hence the global minimum over the entire segment [0, 1] is attained at one point of the sequence $\lambda_0, \lambda_1, \ldots, \lambda_N$. □

Corollary 9.3 *An optimal solution of (9.79) is*

$$x^* = \frac{(\lambda_*, 1 - \lambda_*)}{\sqrt{\lambda_*^2 + (1 - \lambda_*)^2}}, \quad \lambda_* \in \operatorname{argmin}\{F(\lambda_s) : s = 0, 1, \ldots, N\}.$$

Proof Indeed, $x^* = \sigma(x^*)/\|\sigma(x^*)\|$ and $\sigma(x^*) = (\lambda_*, 1 - \lambda_*)$.

9.7 Reverse Convex Constraints

In the previous sections we were concerned with low rank nonconvex problems with nonconvex variables in the objective function. We now turn to low rank nonconvex problems with nonconvex variables in the constraints. Consider the following *monotonic reverse convex problem*

$$(MRP) \qquad \min\{\langle c, x\rangle | \quad x \in D, \; h(x) \le 0\},$$

where $c \in \mathbb{R}^n, D$ is a closed convex set in $\mathbb{R}^n$, and the function $h : X \to \mathbb{R}$ is quasiconcave and K-monotonic on a closed convex set $X \supset D$. In this section we shall present decomposition methods for (MRP) by specializing the algorithms for reverse convex programs studied in Chaps. 6 and 7.

As in Sect. 9.2 assume that $K \subset \text{rec}X$, with lineality $L = \{x | \; Qx = 0\}$, where Q is an $r \times n$ matrix with r linearly independent rows $c^1, \ldots, c^r$. By quasiconcavity of $h(x)$ the set $C = \{x \in X | \; h(x) \ge 0\}$ is convex and by Proposition 9.2 $K \subset \text{rec}C$. Assume further that $D = \{x | \; Ax \le b, x \ge 0\}$ is a polytope with a vertex at 0 and that:

(a) $\min\{\langle c, x\rangle | \; x \in D, h(x) \le 0\} > 0$, so that $h(0) > 0$;
(b) The problem is regular, i.e.,

$$D \setminus \text{int}C = \text{cl}(D \setminus C).$$

9.7.1 Decomposition by Projection

By writing, as in Sect. 7.2 [see formulas (9.11) and (9.12)], $Q = [Q_B, Q_N]$, $x = \begin{bmatrix} x_B \\ x_N \end{bmatrix}$, where Q_B is an $r \times r$ nonsingular matrix, and setting

$$Z = \begin{bmatrix} Q_B^{-1} \\ 0 \end{bmatrix}, \quad \varphi(y) = h(Zy)$$

we define a function $\varphi(y)$ such that $\varphi(y) = h(x)$ for all $x \in X$ satisfying $y = Qx$. The latter also implies that $\varphi(y)$ is quasiconcave on $Q(X)$.

Denote by $\psi(y)$ the optimal value of the linear program

$$\min\{\langle c, x\rangle \,|\, x \in D,\ Qx = y\}.$$

Then (MRP) can be rewritten as a problem in y :

$$\min\{\psi(y) \,|\, y \in Q(D),\ \varphi(y) \leq 0\}. \tag{9.83}$$

Since it is well known that $\psi(y)$ is a convex piecewise affine function, (9.83) is a reverse convex program in the r-dimensional space $Q(\mathbb{R}^n)$. If r is not too large this program can be solved in reasonable time by outer approximation or branch and bound.

9.7.2 Outer Approximation

Following the OA Algorithm for (CDC) (Sect. 6.3), we construct a sequence of polytopes $P_1 \supset P_2 \supset \ldots$ outer approximating the convex set

$$\Omega = \{y \in Q(D) \,|\ \psi(y) \leq \tilde{\gamma}\},$$

where $\tilde{\gamma}$ is the optimal value of (MRP). Also we construct a sequence $+\infty \geq \gamma_1 \geq \gamma_2 \geq \ldots$ such that

$$\gamma_k < +\infty \quad \Rightarrow \quad \exists \overline{y}^k \in Q(D),\ \varphi(\overline{y}^k) \leq 0, \psi(\overline{y}^k) = \gamma_k\ ;$$
$$\{y \in Q(D) \,|\ \psi(y) \leq \gamma_k\} \subset P_k.$$

At iteration k we select (by vertex enumeration)

$$y^k \in \operatorname{argmin}\{\varphi(y) \,|\ y \in P_k\}$$

(note that this requires $\varphi(y)$ to be defined on P_1). If $\varphi(y^k) \geq 0$ then (MRP) is infeasible when $\gamma_k = +\infty$, or $\overline{y}^k$ is optimal when $\gamma_k < +\infty$. If $\varphi(y^k) < 0$, then $y^k \notin \Omega$ and we construct an affine inequality strictly separating y^k from Ω. The polytope P_{k+1} is defined by appending this inequality to P_k.

The key issue in this OA scheme is, given a point y^k such that $\varphi(y^k) < 0$, to construct an affine function $l(y)$ satisfying

$$l(y^k) > 0, \qquad l(y) \leq 0 \quad \forall y \in \Omega. \tag{9.84}$$

Note that we assume (a) and $0 \in D$, so

$$\varphi(0) > 0, \quad \psi(0) < \min\{\psi(y) |\, y \in Q(D),\ \varphi(y) \leq 0\}. \tag{9.85}$$

Now, since $\varphi(y^k) < 0 < \varphi(0)$ and $\gamma_k - \psi(y^k) \leq 0 < \gamma_k - \psi(0)$, one can compute $z^k \in [0, y^k]$ satisfying $\min\{\varphi(z^k), \gamma_k - \psi(z^k)\} = 0$. Consider the linear program

$$(SP(z^k)) \qquad \max\{\langle z^k, v\rangle - \langle b, w\rangle | Q^T v - A^T w \leq c, v \geq 0\}.$$

Proposition 9.7

(i) *If SP(z^k) has a (finite) optimal solution (v^k, w^k) then $z^k \in Q(D)$ and (9.84) is satisfied by the affine function*

$$l(y) := \langle v^k, y - z^k\rangle. \tag{9.86}$$

(ii) *If SP(z^k) has no finite optimal solution, so that the cone $Q^T v - A^T w \leq d,\ w \geq 0$ has an extreme direction (v^k, w^k) such that $\langle z^k, v^k\rangle - \langle b, w^k\rangle > 0$, then $z^k \notin Q(D)$ and (9.84) is satisfied by the affine function*

$$l(y) := \langle y, v^k\rangle - \langle b, w^k\rangle. \tag{9.87}$$

Proof If SP(z^k) has an optimal solution (v^k, w^k) then its dual

$$(SP^*(z^k)) \qquad \min\{\langle c, x\rangle |\, Ax \leq b,\ Qx = z^k,\ x \geq 0\}$$

is feasible, hence $z^k \in Q(D)$. It is easy to see that $v^k \in \partial\psi(z^k)$. Indeed, for any y, since $\psi(y)$ is the optimal value in SP*(y), it must also be the optimal value in $SP(y)$; i.e., $\psi(y) \geq \langle y, v^k\rangle - \langle b, w^k\rangle$, hence $\psi(y) - \psi(z^k) \geq \langle y, v^k\rangle - \langle b, w^k\rangle - \langle z^k, v^k\rangle + \langle b, w^k\rangle = \langle v^k, y - z^k\rangle$, proving the claim. For every $y \in \Omega$ we have $\psi(y) \leq \tilde{\gamma} \leq \psi(z^k)$, so $l(y) := \langle v^k, y - z^k\rangle \leq \psi(y) - \psi(z^k) \leq 0$. On the other hand, $y^k = \theta z^k = \theta z^k + (1-\theta)0$ for some $\theta \geq 1$, hence $l(y^k) = \theta l(z^k) + (1-\theta)l(0) > 0$ because $l(0) \leq \psi(0) - \psi(z^k) < 0$, while $l(z^k) = 0$.

To prove (ii), observe that if SP(z^k) is unbounded, then SP*(z^k) is infeasible, i.e., $z^k \notin Q(D)$. Since the recession cone of the feasible set of SP(z^k) is the cone $Q^T v - A^T w \leq 0, v \geq 0$, SP($z^k$) may be unbounded only if an extreme direction

(v^k, w^k) of this cone satisfies $\langle z^k, v^k\rangle - \langle b, w^k\rangle > 0$. Then the affine function (9.87) satisfies $l(z^k) > 0$ while for any $y \in Q(D)$, $\mathrm{SP}^*(y)$ (hence $\mathrm{SP}(y)$) must have a finite optimal value, and consequently, $\langle y, v^k\rangle - \langle b, w^k\rangle \leq 0$, i.e., $l(y) \leq 0$. □

The convergence of an OA method for (*MRP*) based on the above proposition follows from general theorems established in Sect. 6.3.

Remark 9.7 In the above approach we only used the constancy of $h(x)$ on every manifold $\{x|\ Qx = y\}$. In fact, since $\varphi(y') \geq \varphi(y)$ for all $y' \geq y$, problem (*MRP*) can also be written as

$$\min\{\psi(y)|\ x \in D,\ Qx \leq y,\ \varphi(y) \leq 0\}. \tag{9.88}$$

Therefore, $\mathrm{SP}(z^k)$ could be replaced by

$$\max\{\langle z^k, v\rangle - \langle b, w\rangle |\ Q^T v - A^T w \geq 0,\ v \geq 0,\ w \geq 0\}.$$

9.7.3 Branch and Bound

We describe a conical branch and bound procedure. A simplicial and a rectangular algorithm for the case when $\varphi(y)$ is separable can be developed similarly.

From assumptions (a), (b) we have that

$$\left|\begin{array}{l} 0 \in Q(D), \quad \varphi(0) > 0, \\ \psi(0) < \min\{\psi(y)|\ y \in Q(D),\ \varphi(y) \leq 0\}. \end{array}\right. \tag{9.89}$$

A conical algorithm for solving (9.83) starts with an initial cone M_0 contained in $Q(X)$ and containing the feasible set $Q(D) \cap \{y|\ \varphi(y) \leq 0\}$.

Given any cone $M \subset Q(\mathbb{R}^n)$ of base $[u^1, \ldots, u^k]$ let $\theta_i u^i$ be the point where the i-th edge of M intersects the surface $\varphi(y) = 0$. If U denotes the matrix of columns $u^1, \ldots, u^k$ then $M = \{t \in \mathbb{R}^k|\ Ut \geq 0, t \geq 0\}$ and $\sum_{i=1}^k t_i/\theta_i = 1$ is the equation of the hyperplane passing through the intersections of the edges of M with the surface $\varphi(y) = 0$. Denote by $\beta(M)$ and $\bar{t}(M)$ the optimal value and a basic optimal solution of the linear program

$$LP(M) \qquad \min\left\{\langle c, x\rangle |\ x \in D,\ Qx = \sum_{i=1}^k t_i u^i,\ \sum_{i=1}^k \frac{t_i}{\theta_i} \geq 1, t \geq 0\right\}.$$

Proposition 9.8 *$\beta(M)$ gives a lower bound for $\psi(y)$ over the set of feasible points in M. If $\omega(M) = U\bar{t}(M)$ lies on an edge of M then*

$$\beta(M) = \min\{\psi(y)|\ y \in Q(D) \cap M, \varphi(y) \leq 0\}. \tag{9.90}$$

Proof The first part of the proposition follows from the inclusion $\{y \in Q(D) \cap M|\ \varphi(y) \leq 0\} \subset \{y \in Q(D) \cap M|\ \sum_{i=1}^{k} t_i/\theta_i \geq 1\}$ and the fact that $\min\{\psi(y)|\ y \in Q(D) \cap M, \varphi(y) \leq 0\} = \min\{\langle c, x\rangle|\ x \in D, Qx = y, y \in M, \varphi(y) \leq 0\}$. On the other hand, if $\omega(M)$ lies on an edge of M it must lie on the portion of this edge outside the convex set $\varphi(y) \geq 0$, hence $\varphi(\omega(M)) \leq 0$. Consequently, $\omega(M)$ is feasible to the minimization problem in (9.90) and hence is an optimal solution of it. □

It can be proved that a conical branch and bound procedure using ω-subdivision or ω-bisection (see Sect. 7.1) with the above bounding is guaranteed to converge.

Example 9.10

$$
\begin{aligned}
\min\ & 2x_1 + x_2 + 0.5x_3 \\
\text{s.t.}\ & -2x_1 + x_3 - x_4 \leq 2.5 \\
& x_1 - 3x_2 + x_4 \leq\ 2 \\
& x_1 + x_2 \leq 2 \\
& x_1,\ x_2,\ x_3,\ x_4 \geq 0 \\
& h(x) := -(3x_1 + 6x_2 + 8x_3)^2 - (4x_1 + 5x_2 + x_4)^2 + 154 \leq\ 0.
\end{aligned}
$$

Here the function $h(x)$ is concave and monotonic with respect to the cone $K = \{y|\ c^1y = 0,\ c^2y = 0\}$, where $c^1 = (3, 6, 8, 0), c^2 = (4, 5, 0, 1)$ (cf Example 7.2). So Q is the matrix of two rows c^1, c^2 and $Q(D) = \{y = Qx, x \in D\}$ is contained in the cone $M_0 = \{y|y \geq 0\}$ (D denotes the feasible set). Since the optimal solution $0 \in \mathbb{R}^4$ of the underlying linear program satisfies $h(0) > 0$, conditions (9.89) hold, and $M_0 = \text{cone}\{e^1, e^2\}$ can be chosen as an initial cone. The conical algorithm terminates after 4 iterations, yielding an optimal solution of $(0, 0, 1.54, 2.0)$ and the optimal value 0.76. Note that branching is performed in $\mathbb{R}^2$ (t-space) rather than in the original space $\mathbb{R}^4$.

9.7.4 Decomposition by Polyhedral Annexation

From $L = \{x|\ Qx = 0\} \subset K \subset C$ we have

$$C^\circ \subset K^\circ \subset L^\perp, \tag{9.91}$$

where $L^\perp$ (the orthogonal complement of L) is the space generated by the rows $c^1, \ldots, c^r$ of Q (assuming rank$Q = r$). For every $\gamma \in \mathbb{R} \cup +\{\infty\}$ let $D_\gamma = \{x \in D|\ \langle c, x\rangle \leq \gamma\}$. By Theorem 7.6 the value $\overline{\gamma}$ is optimal if and only if

$$D_{\overline{\gamma}} \setminus \text{int}C \neq \emptyset, \quad D_{\overline{\gamma}} \subset C. \tag{9.92}$$

Following the PA Method for (LRC) (Sect. 8.1), to solve (MRP) we construct a nested sequence of polyhedrons $P_1 \supset P_2 \supset \ldots$ together with a nonincreasing sequence of real numbers $\gamma_1 \geq \gamma_2 \geq \ldots$ such that $D_{\gamma_k} \setminus \mathrm{int}C \neq \emptyset$, $C^\circ \subset P_k \subset L^\perp\ \forall k$, and eventually $P_l \subset [D_{\gamma_l}]^\circ$ for some l : then $D_{\gamma_l} \subset P_l^\circ \subset C$, hence γ_l is optimal by the criterion (9.92).

To exploit monotonicity [which implies (9.91)], the initial polytope P_1 is taken to be the r-simplex constructed as in Lemma 9.1 (note that $0 \in \mathrm{int}C$ by assumption (a), so this construction is possible). Let $\pi : L^\perp \to \mathbb{R}^r$ be the linear map $y = \sum_{i=1}^r t_i c^i \mapsto \pi(y) = t$. For any set $E \subset \mathbb{R}^r$ denote $\tilde{E} = \pi(E) \subset \mathbb{R}^r$. Then by Lemma 9.1 $\tilde{P}_1$ is the simplex in $\mathbb{R}^r$ defined by the inequalities

$$\sum_{i=1}^r t_i \langle c^i, c^j \rangle \leq \frac{1}{\alpha_j} \quad j = 0, 1, \ldots, r$$

where $c^0 = -\sum_{i=1}^r c^i, \alpha_j = \sup\{\alpha |\ \alpha c^j \in C\}$. In the algorithm below, when a better feasible solution x^k than the incumbent has been found, we can compute a feasible point $\hat{x}^k$ at least as good as x^k and lying on an edge of D (see Sect. 5.7). This can be done by carrying out a few steps of the simplex algorithm. We shall refer to $\hat{x}^k$ as a point obtained by local improvement from x^k.

PA Algorithm for (*MRP*)

Initialization. Choose a vertex x^0 of D such that $h(x^0) > 0$ and set $D \leftarrow D - x^0$, $C \leftarrow C - x^0$. Let $\bar{x}^1$ = best feasible solution available, $\gamma_1 = \langle c, \bar{x}^1 \rangle$ ($\bar{x}^1 = \emptyset, \gamma_k = +\infty$ if no feasible solution is available). Let P_1 be the simplex constructed as in Lemma 9.1, V_1 the vertex set of $\tilde{P}_1$. Set $k = 1$.

Step 1. For every $t = (t_1, \ldots, t_r)^T \in V_k$ solve the linear program

$$\max \left\{ \sum_{i=1}^r t_i \langle c^i, x \rangle |\ x \in D, \langle c, x \rangle \leq \gamma_k \right\}$$

to obtain its optimal value $\mu(t)$. (Only the new $t \in V_k$ should be considered if this step is entered from Step 3). If $\mu(t) \leq 1\ \forall t \in V_k$, then terminate: if $\gamma_k < \infty$, $\bar{x}^k$ is an optimal solution; otherwise (MRP) is infeasible.

Step 2. If $\mu(t^k) > 1$ for some $t^k \in V_k$, then let x^k be a basic optimal solution of the corresponding linear program. If $x^k \notin C$, then let $\hat{x}^k$ be the solution obtained by local improvement from x^k, reset $\bar{x}^k = \hat{x}^k$, $\gamma_k = \langle c, \hat{x}^k \rangle$, and return to Step 1.

Step 3. If $x^k \in C$ then set $\bar{x}^{k+1} = \bar{x}^k$, $\gamma_{k+1} = \gamma_k$, compute $\theta_k = \sup\{\theta |\ \theta x^k \in C\}$ and define

$$\tilde{P}_{k+1} = \tilde{P}_k \cap \left\{ t |\ \sum_{i=1}^r t_i \langle x^k, c^i \rangle \leq \frac{1}{\theta_k} \right\}.$$

From V_k derive the vertex set V_{k+1} of $\tilde{P}_{k+1}$. Set $k \leftarrow k + 1$ and return to Step 1.

Proposition 9.9 *The above PA Algorithm for* (MRP) *is finite.*

Proof Immediate, since this is a mere specialization of the PA Algorithm for $(LRCP)$ in Sect. 6.4 to (MRP). □

Remark 9.8 If $K = \{x|Qx \geq 0\}$ then P_1 can be constructed as in Lemma 9.2. The above algorithm assumes regularity of the problem (assumption (b)). Without this assumption, we can replace C by $C_\varepsilon = \{x \in X|\ h(x) \leq \varepsilon\}$. Applying the above algorithm with $C \leftarrow C_\varepsilon$ will yield an ε-approximate optimal solution (cf Sect. 7.3), i.e., a point x^ε satisfying $x^\varepsilon \in D, h(x^\varepsilon) \leq \varepsilon$, and $\langle c, x^\varepsilon\rangle \leq \min\{\langle c, x\rangle|\ x \in D, h(x) \leq 0\}$.

9.8 Network Constraints

Consider a special concave programming problem of the form:

$$
\begin{aligned}
(SCP) \qquad & \min\ g(y_1, \ldots, y_r) + \langle c, x\rangle \\
& \text{s.t.}\ \ y \in Y && (9.93) \\
& \qquad Qx = y,\ Bx = d,\ x \geq 0. && (9.94)
\end{aligned}
$$

where $g : \mathbb{R}^r_+ \rightarrow \mathbb{R}^+$ is a concave function, Y a polytope in $\mathbb{R}^r_+$, $c, x \in \mathbb{R}^n, d \in \mathbb{R}^m, Q$ an $r \times n$ matrix of rank r, and B an $m \times n$ matrix. In an economic interpretation, $y = (y_1, \ldots, y_r)^T$ may denote a production program to be chosen from a set Y of technologically feasible production programs, x a distribution–transportation program to be determined so as to meet the requirements (9.94). The problem is then to find a production–distribution–transportation program, satisfying conditions (9.94), with a minimum cost. A variant of this problem is the classical *concave production–transportation problem*

$$
(PTP) \quad \left|
\begin{array}{l}
\min\ g(y) + \displaystyle\sum_{i=1}^{r}\sum_{j=1}^{m} c_{ij}x_{ij} \\
\text{s.t.}\ \ \displaystyle\sum_{j=1}^{m} x_{ij} = y_i \quad i = 1, \ldots, r \\
\qquad \displaystyle\sum_{i=1}^{r} x_{ij} = d_j \quad j = 1, \ldots, m \\
\qquad x_{ij} \geq 0, \quad i = 1, \ldots, r,\ j = 1, \ldots, m
\end{array}
\right.
$$

where y_i is the production level to be determined for factory i and x_{ij} the amount to be sent from factory i to warehouse j in order to meet the demands $d_1, \ldots, d_m$ of the warehouses. Another variant of SCP is the *minimum concave cost network flow problem* which can be formulated as

$$(MCCNFP) \qquad \left| \begin{array}{l} \min \sum_{i=1}^{r} g_i(x_i) + \sum_{i=r+1}^{n} c_i x_i \\ \text{s.t.} \quad Ax = d, \ x \geq 0. \end{array} \right.$$

where A is the node-arc incidence matrix of a given directed graph with n arcs, d the vector of node demands (with supplies interpreted as negative demands), and x_i the flow value on arc i. Node j with $d_j > 0$ are the *sinks*, nodes j with $d_j < 0$ are the *sources*, and it is assumed that $\sum_j d_j = 0$ (balance condition). The cost of shipping t units along arc i is a nonnegative concave nondecreasing function $g_i(t)$ for $i = 1, \ldots, r$, and a linear function $c_i t$ (with $c_i \geq 0$) for $i = r+1, \ldots, n$. To cast ($MCCNFP$) into the form (SCP) it suffices to rewrite it as

$$\left| \begin{array}{l} \min \sum_{i=1}^{r} g_i(y_i) + \sum_{i=r+1}^{n} c_i x_i \\ \text{s.t.} \quad x_i = y_i \ (i = 1, \ldots, r), \quad Ax = d, \ x \geq 0. \end{array} \right. \tag{9.95}$$

Obviously, Problem (SCP) is equivalent to

$$\begin{aligned} &\min\ g(y) + t \\ &\text{s.t.} \quad \langle c, x \rangle \leq t, \ Qx = y, \ Bx = d, \ x \geq 0, \ y \in Y. \end{aligned} \tag{9.96}$$

Since this is a concave minimization problem with few nonconvex variables, it can be handled by the decomposition methods discussed in the previous Sects. 7.2 and 7.3.

Note that in both (PTP) and ($MCCNFP$) the constraints (9.94) have a nice structure such that when y is fixed the problem reduces to a special linear transportation problem (for (PTP)) or a linear cost flow problem (for ($MCCNFP$)), which can be solved by very efficient specialized algorithms. Therefore, to take full advantage of the structure of (9.94) it is important to preserve this structure under the decomposition.

9.8.1 Outer Approximation

This method is initialized from a polytope $P_1 = Y \times [\alpha, \beta]$,where α, β are, respectively, a lower bound and an upper bound of $\langle c, x \rangle$ over the polytope $Bx = d, x \geq 0$. At iteration k, we have a polytope $P_k \subset P_1$ containing all (y, t) for which there exists $x \geq 0$ satisfying (9.96). Let

$$(y^k, t^k) \in \operatorname{argmin}\{t + g(y) |\ (y, t) \in P_k\}.$$

Following Remark 7.1 to check whether (y^k, t^k) is feasible to (9.96) we solve the pair of dual programs

$$\min\{\langle c, x \rangle |\ Qx = y^k, \ Bx = d, \ x \geq 0\} \tag{9.97}$$

$$\max\{\langle y^k, u\rangle + \langle d, v\rangle |\ Q^T u + B^T v \leq c\} \tag{9.98}$$

obtaining an optimal solution x^k of (9.97) and an optimal solution (u^k, v^k) of the dual (9.98). If $\langle c, x^k\rangle \leq t^k$ then (x^k, y^k) solves (SCP). Otherwise, we define

$$P_{k+1} = P_k \cap \{(y,t)|\ \langle y, u^k\rangle + \langle d, v^k\rangle \leq t\} \tag{9.99}$$

and repeat the procedure with $k \leftarrow k+1$.

In the particular case of (PTP) the subproblem (9.97) is a linear transportation problem obtained from (PTP) by fixing $y = y^k$, while in the case of $(MCCNFP)$ the subproblem (9.97) is a linear min-cost network flow problem obtained from $(MCCNFP)$ by fixing $x_i = y_i^k$ for $i = 1,\dots,r$ (or equivalently, by deleting the arcs $i = 1,\dots,r$ and accordingly modifying the demands of the nodes which are their endpoints).

9.8.2 Branch and Bound Method

Assume that in problem (SCP) the function $g(y)$ is concave separable, i.e., $g(y) = \sum_{i=1}^r g_i(y_i)$. For each rectangle $M = [a, b]$ let

$$l_M(y) = \sum_{i=1}^R l_{i,M}(y_i)$$

where $l_{i,M}(t)$ is the affine function that agrees with $g_i(t)$ at the endpoints a_i, b_i of the segment $[a_i, b_i]$, i.e.,

$$l_{i,M}(t) = g_i(a_i) + \frac{g_i(b_i) - g_i(a_i)}{b_i - a_i}(t - a_i). \tag{9.100}$$

(see Proposition 7.4). Then a lower bound $\beta(M)$ for the objective function in (SCP) over M is provided by the optimal value in the linear program

$$(LP(M)) \qquad \min\left\{\sum_{i=1}^r l_{i,M}(y_i) + \langle c, x\rangle |\ Qx = y, Ax = d, x \geq 0\right\}.$$

Using this lower bound and a rectangular ω-subdivision rule, a convergent rectangular algorithm has been developed in Horst and Tuy (1996) which is a refined version of an earlier algorithm by Soland (1974) and should be able to handle even large scale (PTP) provided r is relatively small. Also note that if the data are all integral, then it is well known that a basic optimal solution of the linear transportation problem $LP(M)$ always exists with integral values. Consequently, in this case, starting from an initial rectangle with integral vertices the algorithm will generate only subrectangles with integral vertices, and so it will necessarily be finite.

In practice, aside from production and transportation costs there may also be shortage/holding costs due to a failure to meet the demands exactly. Then we have the following problem (cf Example 5.1) which includes a formulation of the *stochastic transportation–location problem* as a special case:

$$(GPTP)\quad \left|\begin{array}{l} \min\ \langle c, x\rangle + \sum_{i=1}^{r} g_i(y_i) + \sum_{j=1}^{m} h_j(z_j) \\ \text{s.t.}\ \displaystyle\sum_{j=1}^{m} x_{ij} = y_i \quad (i = 1, \dots, r) \\ \qquad \displaystyle\sum_{i=1}^{r} x_{ij} = z_j \quad (j = 1, \dots, m) \\ \qquad 0 \le y_i \le s_i \quad \forall i, \\ \qquad x_{ij} \ge 0 \quad \forall i, j. \end{array}\right.$$

where $g_i(y_i)$ is the cost of producing y_i units at factory i and $h_j(z_j)$ is the shortage/holding cost to be incurred if the warehouse j receives $z_j \neq d_j$ (demand of warehouse j). It is assumed that g_i is a concave nondecreasing function satisfying $g_i(0) = 0 \le g_i(0_+)$ (so possibly $g_i(0_+) > 0$ as, e.g., for a fixed charge), while $h_j : [0, s] \to \mathbb{R}_+$ is a convex function ($s = \sum_{i=1}^{r} s_i$), such that $h_j(d_j) = 0$, and the right derivative of h_j at 0 has a finite absolute value ξ_j. The objective function is then a d.c. function, which is a new source of difficulty. However, since the problem becomes convex when y is fixed, it can be handled by a branch and bound algorithm in which branching is performed upon y (Holmberg and Tuy 1995).

For any rectangle $[a, b] \subset Y := \{y \in \mathbb{R}^r_+ \mid 0 \le y_i \le s_i\ i = 1, \dots, r\}$ let as previously $l_{i,M}(t)$ be defined by (9.100) and $l_M(y) = \sum_{i=1}^{r} l_{i,M}(y_i)$. Then a lower bound of the objective function over the feasible points in M is given by the optimal value $\beta(M)$ in the convex program

$$CP(M)\quad \left|\begin{array}{l} \min\ \langle c, x\rangle + l_M(y) + \sum_{j=1}^{m} h_j(z_j) \\ \text{s.t.}\ \displaystyle\sum_{j=1}^{m} x_{ij} = y_i \quad (i = 1, \dots, r) \\ \qquad \displaystyle\sum_{i=1}^{r} x_{ij} = z_j \quad (j = 1, \dots, m) \\ \qquad a_i \le y_i \le b_i \quad \forall i, \\ \qquad x_{ij} \ge 0 \quad \forall i, j. \end{array}\right.$$

Note that if an optimal solution $(\overline{x}(M), \overline{y}(M), \overline{z}(M))$ of this convex program satisfies $l_M(\overline{y}(M)) = g(\overline{y}(M))$ then $\beta(M)$ equals the minimum of the objective function over the feasible solutions in M. When y is fixed, $(GPTP)$ is a convex program in x, z. If $\varphi(y)$ denotes the optimal value of this program then the problem amounts to minimizing $\varphi(y)$ over all feasible (x, y, z).

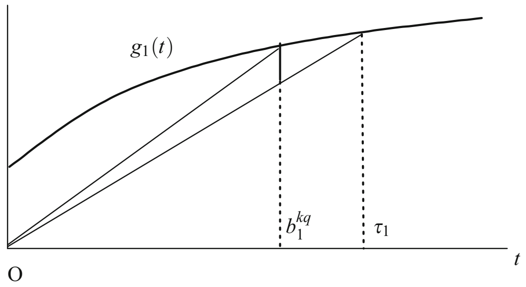


Fig. 9.1 Consequence of concavity of $g_1(t)$

BB Algorithm for (*GPTP*)

Initialization. Start with a rectangle $M_1 \subset Y$. Let y^1 be the best feasible solution available, INV = $\varphi(y^1)$. Set $\mathscr{P}_1 = \mathscr{S}_1 = \{M_1\}, k = 1$.

Step 1. For each $M \in \mathscr{R}_1$ solve the associated convex program CP(M) obtaining its optimal value $\beta(M)$ and optimal solution $(\overline{x}(M), \overline{y}(M), \overline{z}(M))$.

Step 2. Update INV and y^k.

Step 3. Delete all $M \in \mathscr{S}_k$ such that $\beta(M) \geq INV - \varepsilon$. Let $\mathscr{R}_k$ be the collection of remaining rectangles.

Step 4. If $\mathscr{R}_k = \emptyset$, then terminate: y^k is a global ε-optimal solution of (*GPTP*).

Step 5. Select $M_k \in \operatorname{argmin}\{\beta(M) |\ M \in \mathscr{R}_k\}$. Let $\overline{y}^k = \overline{y}(M_k)$,

$$i_k \in \arg\max_i \{g_i(\overline{y}^k) - l_{i,M_K}(\overline{y}^k)\}. \tag{9.101}$$

If $g_{i_k}(\overline{y}^k_{i_k}) - l_{i_k,M_k}(\overline{y}^k) = 0$, then terminate: $\overline{y}^k$ is a global optimal solution.

Step 6. Divide M_k via $(\overline{y}^k, i_k)$ (see Sect. 5.6). Let $\mathscr{P}_{k+1}$ be the partition of M_k and $\mathscr{S}_{k+1} = \mathscr{R}_k \setminus (\{M_k\}) \cup \mathscr{P}_{k+1}$. Set $k \leftarrow k + 1$ and return to Step 1.

To establish the convergence of this algorithm, recall that $\overline{y}^k = \overline{y}(M_k)$. Similarly denote $\overline{x}^k = \overline{x}(M_k), \overline{z}^k = \overline{z}(M_k)$, so $(\overline{x}^k, \overline{y}^k, \overline{z}^k)$ is an optimal solution of $CP(M_k)$. Suppose the algorithm is infinite and let $\overline{y}$ be any accumulation point of $\{\overline{y}^k\}$, say $\overline{y} = \lim_{q\to\infty} \overline{y}^{k_q}$. Without loss of generality we may assume $\overline{x}^{k_q} \to \overline{x}, \overline{z}^{k_q} \to \overline{z}$ and also $i_{k_q} = 1\ \forall q$, so that

$$1 \in \arg\max_i \{g_i(\overline{y}^{k_q}) - l_{i,M_{k_q}}(\overline{y}^{k_q})\}. \tag{9.102}$$

Clearly if for some q we have $\overline{y}_1^{k_q} = 0$, then $g_i(\overline{y}^{k_q}) = l_{1,M_{k_q}}(\overline{y}^{k_q})$, and the algorithm terminates in Step 5. Therefore, since the algorithm is infinite, we must have $\overline{y}_1^{k_q} > 0\ \forall q$.

Lemma 9.9 *If* $g_1(0_+) > 0$ *then*

$$\overline{y}_1 = \lim_{q\to\infty} \overline{y}_1^{k_q} > 0. \tag{9.103}$$

In other words, $\overline{y}_1$ *is always a continuity point of* $g_1(t)$.

Proof Let $M_{k_q,1} = [a_1^q, b_1^q]$, so that $\overline{y}_1^{k_q} \in [a_1^q, b_1^q]$ and $[a_1^q, b_1^q], q = 1, 2, \ldots$ is a sequence of nested intervals. If $0 \notin [a_1^{q_0}, b_1^{q_0}]$ for some q_0, then $0 \notin [a_1^q, b_1^q]\ \forall q \geq q_0$, and $\overline{y}_1^{k_q} \geq a_1^{q_0} > 0\ \forall q \geq q_0$, hence $\overline{y}_1 \geq a_1^{q_0} > 0$, i.e., (9.103) holds. Thus, it suffices to consider the case

$$a_1^q = 0,\ \overline{y}_1^{k_q} > 0 \quad \forall q. \tag{9.104}$$

Recall that ξ_j denotes the absolute value of the right derivative of $h_j(t)$ at 0. Let

$$\pi = \max_j \xi_j - \min_j c_{1j}. \tag{9.105}$$

Since $\sum_{j=1}^n \overline{x}_{1j}^{k_q} = \overline{y}_1^{k_q} > 0$, there exists j_0 such that $\overline{x}_{1j_0}^{k_q} > 0$. Let $(\tilde{x}^{k_q}, \overline{y}^{k_q}, \overline{z}^{k_q})$ be a feasible solution to $CP(M)$ obtained from $(\overline{x}^{k_q}, \overline{y}^{k_q}, \overline{z}^{k_q})$ by subtracting a very small positive amount $\eta < \overline{x}_{1j_0}^{k_q}$ from each of the components $\overline{x}_{1j_0}^{k_q}, \overline{y}_1^{k_q}, \overline{z}_{j_0}^{k_q}$, and letting unchanged all the other components. Then the production cost at factory 1 decreases by $l_{1,M_{k_q}}(\overline{y}_1^{k_q}) - l_{1,M_{k_q}}(\tilde{y}_1^{k_q}) = g_1(b_1^{k_q})\eta/b_1^{k_q} > 0$, the transportation cost in the arc $(1, j_0)$ decreases by $c_{1j_0}\eta$, while the penalty incurred at warehouse j_0 either decreases (if $\overline{z}_{j_0}^{k_q} > d_j$), or increases by $h_{j_0}(\overline{z}_{j_0}^{k_q} - \eta) - h_{j_0}(\overline{z}_{j_0}^{k_q}) \leq \xi_{j_0}\eta$. Thus the total cost decreases by at least

$$\delta = \left[\frac{g_1(b_1^{k_q})}{b_1^{k_q}} + c_{1j_0} - \xi_{j_0}\right]\eta. \tag{9.106}$$

If $\pi \leq 0$ then $c_{1j_0} - \xi_{j_0} \geq 0$, hence $\delta > 0$ and $(\tilde{x}^{k_q}, \tilde{y}^{k_q}, \tilde{z}^{k_q})$ would be a better feasible solution than $(\overline{x}^{k_q}, \overline{y}^{k_q}, \overline{z}^{k_q})$, contradicting the optimality of the latter for $CP(M_{k_q})$. Therefore, $\pi > 0$. Now suppose $g_1(0_+) > 0$. Since $g_1(t)/t \to +\infty$ as $t \to 0_+$ (in view of the fact $g_1(0_+) > 0$), there exists $\tau_1 > 0$ satisfying

$$\frac{g_1(\tau_1)}{\tau_1} > \pi.$$

Observe that since $M_{k_q,1} = [0, b_1^{k_q}]$ is divided via $\overline{y}_1^{k_q}$, we must have $[0, b_1^{k'_q}] \subset [0, \overline{y}_1^{k_q}]$ for all $q' > q$, while $[0, \overline{y}_1^{k_q}] \subset [0, b_1^{k_q}]$ for all q. With this in mind we will show that

$$b_1^{k_q} \geq \tau_1 \quad \forall q. \tag{9.107}$$

By the above observation this will imply that $\overline{y}_1^{k_q} \geq \tau_1 \forall q$, thereby completing the proof. Suppose (9.107) does not hold, i.e., $b_1^{k_q} < \tau_1$ for some q. Then $\overline{y}_1^{k_q} \leq b_1^{k_q} < \tau_1$, and since $g_1(t)$ is concave it is easily seen by Fig. 9.1 that $g_1(b_1^{k_q}) \geq b_1^{k_q} g_1(\tau_1)/\tau_1$, i.e.,

$$\frac{g_1(b_1^{k_q})}{b_1^{k_q}} \geq \frac{g_1(\tau_1)}{\tau_1}.$$

This, together with (9.106) implies that

$$\delta \geq \left(\frac{g_1(\tau_1)}{\tau_1} - \pi\right)\overline{x}_{1j_0}^{k_q} > 0$$

hence $(\tilde{x}^{k_q}, \tilde{y}^{k_q}, \tilde{z}^{k_q})$ is a better feasible solution than $(\overline{x}^{k_q}, \overline{y}^{k_q}, \overline{z}^{k_q})$. This contradiction proves (9.107). □

Theorem 9.3 *The above algorithm for* (*GPTP*) *can be infinite only if* $\varepsilon = 0$ *and in this case it generates an infinite sequence* $\overline{y}^k = \overline{y}(M_k), k = 1, 2, \ldots$ *every accumulation point of which is an optimal solution.*

Proof For $\varepsilon = 0$ let $\overline{y} = \lim_{q\to\infty} \overline{y}^{k_q}$ be an accumulation point of $\{\overline{y}^{k_q}\}$ with, as previously, $i_{k_q} = 1\ \forall q$, and $M_{k_q,1} = [a_1^q, b_1^q]$. Since $\{a_1^q\}$ is nondecreasing, $\{b_1^q\}$ is nonincreasing, we have $a_1^q \to a_1^*, b_1^q \to b_1^*$. For each q either of the following situations occurs:

$$a_1^q < a^{q+1} < b_1^{q+1} < \overline{y}_1^{k_q} < b_1^q \tag{9.108}$$

$$a_1^q < \overline{y}_1^{k_q} < a_1^{q+1} < b_1^{q+1} < b_1^q \tag{9.109}$$

If (9.108) occurs for infinitely many q then $\lim_{q\to\infty} \overline{y}_1^{k_q} = \lim_{q\to\infty} b_1^q = b_1^*$. If (9.109) occurs for infinitely many q then $\lim_{q\to\infty} \overline{y}_1^{k_q} = \lim_{q\to\infty} a_1^q = a_1^*$. Thus, $\overline{y}_1 \in \{a_1^*, b_1^*\}$. By Lemma 9.9, the concave function $g_1(t)$ is continuous at $\overline{y}_1$. Suppose, for instance, that $\overline{y}_1 = a_1^*$ (the other case is similar). Since $\overline{y}_1^{k_q} \to a_1^*$ and $a_1^q \to a_1^*$, it follows from the continuity of $g_1(t)$ at a_1^* that $g_1(\overline{y}_1^{k_q}) - l_{1,M_{k_q}}(\overline{y}_1^{k_q}) \to 0$ and hence, by (9.102),

$$g_i(\overline{y}_i^{k_q}) - l_{i,M_{k_q}}(\overline{y}_i^{k_q}) \to 0 \quad i = 1, \ldots, r$$

as $q \to \infty$. This implies that $g(\overline{y}^{k_q}) - l_{M_{k_q}}(\overline{y}^{k_q}) \to 0$ and hence, that

$$\beta(M_{k_q}) \to \langle c, \overline{x}\rangle + g(\overline{y}) + h(\overline{z})$$

as $q \to \infty$. Since $\beta(M_{k_q}) \leq \gamma :=$ optimal value of (*GPTP*) it follows that $\langle c, \overline{x}\rangle + g(\overline{y}) + h(\overline{z})$ is the sought minimum of (*GPTP*). □

Computational experiments have shown that the above algorithm for (*GPTP*) can solve problems with up to $r = 100$ supply points and $n = 500$ demand points in reasonable time on a Sun SPARCstation SLC (Holmberg and Tuy 1995). To apply this algorithm to the case when $h_j(z_j) = +\infty$ for $z_j \neq d_j$ (e.g., the fixed charge problem) it suffices to redefine $h_j(z_j) = \xi_j(|d_j - x_j|)$ where ξ_j is finite but sufficiently large.

9.9 Some Polynomial-Time Solvable Problems

As mentioned previously, even quadratic programming with one negative eigenvalue is NP-hard (Pardalos and Vavasis 1991). There are, nevertheless, genuine concave programs which are polynomial-time solvable. In this section we discuss this topic, to illustrate the efficiency of the parametric approach for certain classes of low rank nonconvex problems.

9.9.1 The Special Concave Program CPL(1)

Consider the special concave program under linear constraints:

$$(CPL(1)) \quad \left| \begin{array}{l} \min\ g(y) + \langle c, x\rangle \\ \text{s.t.}\ \ \sum_{j=1}^{n} x_j = y, \quad 0 \le x_j \le d_j\ \forall j. \end{array} \right.$$

where $g(y)$ is a concave function, $c \in \mathbb{R}^n$ and $d \in \mathbb{R}^n_+$. Let $s = \sum_{j=1}^{n} d_j$. To solve this concave program by the parametric right-hand side method presented in Sect. 8.5, we first compute the breakpoints $y_0 = 0 \le y_1 \le \ldots \le y_N = s$ of the optimal value function of the associated parametric linear program

$$(LP(y)) \qquad \min\{\langle c, x\rangle |\ \sum_{j=1}^{n} x_j = y,\ 0 \le x_j \le d_j\ \forall j\}, \quad 0 \le y \le s. \qquad .$$

If $\overline{x}^i$ is a basic optimal solution of LP(y_i) then, by Lemma 9.4, an optimal solution of CPL(1) is given by $(\overline{x}^{i_*}, y_{i_*})$ where

$$i_* \in \arg\min\{g(y_i) + c\overline{x}^i |\ i = 0, 1, \ldots, N\}. \tag{9.110}$$

Thus, all is reduced to solving LP(y) parametrically in $y \in [0, s]$. But this is very easy because LP(y) is a continuous knapsack problem. In fact, by reindexing we can assume that the c_j are ordered by increasing values:

$$c_1 \ \le\ c_2 \ \le \ldots\ \le\ c_n.$$

Let k be an index such that $\sum_{j=1}^{k-1} d_j \le y \le \sum_{j=1}^{k} d_j$.

Lemma 9.10 *An optimal solution of LP*(y) *is* $\overline{x}$ *with*

$$\left| \begin{array}{l} \overline{x}_j = d_j \quad \forall j = 1, \ldots, k-1 \\ \overline{x}_k = y - \sum_{j=1}^{k-1} d_j, \quad \overline{x}_j = 0 \quad \forall j = k+1, \ldots, n. \end{array} \right.$$

Proof This well-known result (see, e.g., Dantzig 1963) is rather intuitive because for any feasible solution x of LP(y), by noting that $y = \sum_{j=1}^{n} x_j$ we can write

$$\begin{aligned}\langle c, \overline{x}\rangle &= \sum_{j=1}^{k-1} c_j d_j + c_k \sum_{j=1}^{n} x_j - c_k \sum_{j=1}^{k-1} d_j \\ &= \sum_{j=1}^{k-1} c_j d_j - c_k \sum_{j=1}^{k-1}(d_j - x_j) + c_k \sum_{j=k}^{n} x_j \\ &\le \sum_{j=1}^{k-1} c_j d_j - \sum_{j=1}^{k-1} c_j(d_j - x_j) + c_k \sum_{j=k}^{n} x_j \le \langle c, x\rangle. \square\end{aligned}$$

Proposition 9.10

(i) *The breakpoints of the optimal value function of* $LP(y)$, $0 \le y \le s$, *are among the points:*

$$y_0 = 0,\ y_k = y_{k-1} + d_k \quad k = 1, \ldots, n$$

(ii) *For every* $k = 0, 1, \ldots, n$ *a basic optimal solution of* $TP(y_k)$ *is the vector* $\overline{x}^k$ *such that*

$$\overline{x}^k_j = d_j,\ j = 1, \ldots, k; \quad \overline{x}^k_j = 0,\ j = k+1, \ldots n. \tag{9.111}$$

Proof From Lemma 9.10, for any $y \in [y_k, y_{k+1}]$ a basic optimal solution of $LP(y)$ is

$$x^y = \overline{x}^k + \theta(\overline{x}^{k+1} - \overline{x}^k) \text{ with } \theta = \frac{y - y_k}{d_{k+1}}.$$

Hence each segment $[y_k, y_{k+1}]$ is a linear piece of the optimal value function $\varphi(y)$ of $LP(y)$, which proves (i). Part (ii) is immediate. $\square$

As a result, the concave program $CPL(1)$ can be solved by the following:

Parametric Algorithm for CPL(1)
Initialization. Order the coefficients c_j by increasing values:

$$c_1 \le c_2 \le \ldots \le c_n.$$

Step 2. For $k = 0, 1, \ldots, n$ compute

$$f_k = g\left(\sum_{j=1}^{k} d_j\right) + \sum_{j=1}^{k} c_j d_j.$$

Step 3. Find $f_{k^*} = \min\{f_0, f_1, \ldots, f_n\}$. Then $\overline{x}^{k^*}$ is an optimal solution.

Proposition 9.11 *The above algorithm requires $O(n \log_2 n)$ elementary operations and n evaluations of $g(\cdot)$*

Proof The complexity dominant part of the algorithm is Step 1 which requires $O(n \log_2 n)$ operations, using, e.g., the procedure Heapsort (Ahuja et al. 1993). Also an evaluation of $g(\cdot)$ is needed for each number f_k. □

Thus, assuming the values of the nonlinear function $g(\cdot)$ provided by an oracle the above algorithm for $CPL(1)$ runs in *strongly polynomial time*.

9.10 Applications

9.10.1 Problem PTP(2)

For convenience we shall refer to problem (PTP) with r factories as PTP(r). So $PTP(2)$ is the problem

$$PTP(2) \quad \left| \begin{array}{ll} \min\ g(y) + \sum_{i=1}^{2} \sum_{j=1}^{m} c_{ij} x_{ij} & \\ \text{s.t.} \ \sum_{j=1}^{m} x_{1j} = y & \\ \quad x_{1j} + x_{2j} = d_j & j = 1, \ldots, m, \\ \quad 0 \le x_{ij} \le d_j & i = 1, 2,\ j = 1, \ldots, m \end{array} \right.$$

Substituting $d_j - x_{1j}$ for x_{2j} and setting $x_j = x_{1j}, c_j = c_{1j} - c_{2j}$ we can convert this problem into a CPL(1):

$$\begin{array}{l} \min\ g(y) + \langle c, x \rangle \\ \text{s.t.}\ \ \sum_{j=1}^{m} x_j = y, \quad 0 \le x_j \le d_j\ \forall j. \end{array}$$

Therefore,

Proposition 9.12 *$PTP(2)$ can be solved by an algorithm requiring $O(m \log_2 m)$ elementary operations and m evaluations of $g(\cdot)$.*

9.10.2 Problem FP(1)

We shall refer to Problem (*MCCNFP*) with a single source and r nonlinear arc costs as FP(r). So $FP(1)$ is the problem of finding a feasible flow with minimum cost in a given network $\mathbf{G}$ with m nodes (indexed by $j = 1, \dots, m$), and n arcs (indexed by $i = 1, \dots, n$), where node j has a demand d_j (with $d_m < 0, d_j \geq 0$ for all other $j, \sum_{j=1}^{m} d_j = 0$), arc 1 has a concave nonlinear cost function $g_1(t)$ while each arc $i = 2, \dots, n$ has a linear cost $g_i(t) = c_i t$ (with $c_i \geq 0$). If A_j^+ and A_j^- denote the sets of arcs entering and leaving node j, respectively, then the problem is

$$FP(1) \quad \left| \begin{array}{l} \min g_1(x_1) + \sum_{i=2}^{n} c_i x_i \\ \text{s.t.} \quad \sum_{i \in A_j^+} x_i - \sum_{i \in A_j^-} x_i = d_j, \quad j = 1, 2, \dots, m, \\ \qquad x \geq 0. \end{array} \right.$$

Proposition 9.13 *$FP(1)$ can be reduced to a $PTP(2)$ by means of $O(m \log_2 m + n)$ elementary operations.*

Proof To simplify the language we call the single arc with nonlinear cost the *black* arc. All the other arcs are called *white*, and the unit cost c_i attached to a white arc is called its length. A directed path through white arcs only is called a *white path*; its length is then the sum of the lengths of its arcs. A white path from a node j to a node j' is called *shortest* if its length is smallest among all white paths from j to j'. Denote by Q_{0j} the shortest white path *from the source to sink* j, by Q_{1j} the shortest white path *from the head of the black arc to sink* j, and by P the shortest white path *from the source to the tail of the black arc*. Let c_{0j}, c_{1j}, p be the lengths of these paths, respectively (the length being $+\infty$, or an arbitrarily large positive number, if the path does not exist). Assuming $\{j | d_j > 0\} = \{1, \dots, m_1\}$, let $s = \sum_{j=1}^{m_1} d_j$. Consider now a problem $PTP(2)$ defined for two factories F_0, F_1, and m_1 warehouses $W_1, \dots, W_{m_1}$ with demands $d_1, \dots, d_{m_1}$, where the cost of producing y units at factory F_0 and $s - y$ units at factory F_1 is $g_1(y) + py$ while the unit transportation cost from F_i to W_j is c_{ij} $(i = 0, 1)$, i.e.,

$$\left| \begin{array}{l} \min g_1(y) + py + \sum_{i=0}^{1} \sum_{j=1}^{m_1} c_{ij} x_{ij} \\ \text{s.t.} \quad \sum_{j=1}^{m_1} x_{0j} = y \\ \qquad x_{0j} + x_{1j} = d_j \quad j = 1, \dots, m_1 \\ \qquad 0 \leq x_{0j} \leq d_j \quad j = 1, \dots, m_1 \end{array} \right. \tag{9.112}$$

(the corresponding reduced network is depicted in Fig. 9.2). Clearly from an optimal solution $[\overline{x}_{ij}]$ of (9.112) we can derive an optimal solution of $FP(1)$ by setting, in $FP(1)$,

$$x_1 = \sum_{j=1}^{m_1} \overline{x}_{0j}$$

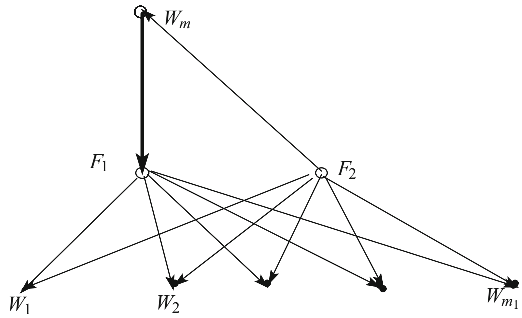


Fig. 9.2 Reduced network

and solving the corresponding linear min-cost network flow problem. Thus, solving $FP(1)$ is reduced to solving the just defined $PTP(2)$. To complete the proof, it remains to observe that the computation of $c_{ij}, i = 0, 1, \ j = 1, \ldots, m_1$, and p needed for the above transformation, requires solving two single-source multiple-sink shortest problems with nonnegative arc costs, hence a time of $O(m \log_2 m + n)$ (using Fredman and Tarjan's (1984) implementation of Dijkstra' s algorithm).

It follows from Propositions 9.11 and 9.13 that $FP(1)$ can be solved by a strongly polynomial algorithm requiring $O(m \log_2 m + m \log_2 m + n) = O(m \log_2 m + n)$ elementary operations and m evaluations of the function $g_1(.)$.

Remark 9.9 A polynomial (but not strongly polynomial) algorithm for $FP(1)$ was first given by Guisewite and Pardalos (1992). Later a strongly polynomial algorithm based on solving linear min-cost network flow problems with a parametric objective function was obtained by Klinz and Tuy (1993). The algorithms presented in this section for $PTP(1)$ and $FP(1)$ were developed by Tuy et al. (1993a) and subsequently extended in a series of papers by Tuy et al. (1993b, 1994a, 1995b, 1996a) to prove the strong polynomial solvability of a class of network flow problems including $PTP(r)$ and $FP(r)$ with fixed r.

9.10.3 Strong Polynomial Solvability of PTP(r)

The problem PTP(r) as was formulated at the beginning of Sect. 8.8 is

$$PTP(r) \quad \left| \begin{array}{l} \min\ g(\sum_{j=1}^m x_{1j}), \ldots, \sum_{j=1}^m x_{rj}) + \sum_{i=1}^r \sum_{j=1}^m c_{ij}x_{ij} \\ \text{s.t.}\ \ \sum_{i=1}^r x_{ij} = d_j \quad j = 1, \ldots, m \\ \qquad x_{ij} \geq 0, \quad i = 1, \ldots, r,\ j = 1, \ldots, m \end{array} \right.$$

where $g(y') \geq g(y)$ whenever $y' \geq y$.

Following the approach in Sect. 8.5 we associate with PTP(r) the parametric problem

$$P(t)\quad \left|\begin{array}{l} \min \sum_{i=1}^{r} t_i \sum_{j=1}^{m} x_{ij} + \sum_{i=1}^{r}\sum_{j=1}^{m} c_{ij}x_{ij} \\ \text{s.t.}\quad \sum_{i=1}^{r} x_{ij} = d_j \quad j=1,\dots,m \\ \qquad x_{ij} \geq 0, \quad i=1,\dots,r,\ j=1,\dots,m \end{array}\right.$$

where $t \in \mathbb{R}^r_+$. The parametric domain $\mathbb{R}^r_+$ is partitioned into a finite collection $\mathscr{P}$ of polyhedrons ("cells") such that for each cell $\Pi \in \mathscr{P}$ there is a basic solution x^{Π} which is optimal to P(t) for all $t \in \Pi$. Then an optimal solution of PTP(r) is x^{Π^*} where

$$\Pi^* \in \operatorname{argmin}\left\{ g\left(\sum_{j=1}^{m} x^{\Pi}_{1j},\dots,\sum_{j=1}^{m} x^{\Pi}_{rj}\right) + \sum_{i,j} c_{ij}x^{\Pi}_{ij} \,|\, \Pi \in \mathscr{P}\right\}. \tag{9.113}$$

Following Tuy (2000b) we now show that the collection $\mathscr{P}$ can be constructed and its cardinality is bounded by a polynomial in m.

Observe that the dual of $P(t)$ is

$$P^*(t)\quad \left|\begin{array}{l} \max \sum_{j=1}^{m} d_j u_j \\ \text{s.t.}\quad u_j \leq t_i + c_{ij}, \quad i=1,\dots,r, j=1,\dots,m \end{array}\right.$$

Also for any fixed $t \in r^r_+$ a basic solution of $P(t)$ is a vector x^t such that for every $j = 1,\dots,m$ there is an i_j satisfying

$$x^t_{ij} = \begin{cases} d_j & i = i_j \\ 0 & i \neq i_j \end{cases} \tag{9.114}$$

By the duality of linear programming x^t is a basic optimal solution of $P(t)$ if and only if there exists a feasible solution $u = (u_1,\dots,u_m)$ of $P^*(t)$ satisfying

$$u_j \begin{cases} = t_i + c_{ij} & i = i_j \\ \leq t_i + c_{ij} & i \neq i_j \end{cases}$$

or, alternatively, if and only if for every $j = 1,\dots,m$:

$$i_j \in \operatorname{argmin}_{i=1,\dots,r}\{t_i + c_{ij}\}. \tag{9.115}$$

Now let I_*^2 be the set of all pairs (i_1, i_2) such that $i_1 < i_2 \in \{1, \dots, r\}$. Define a cell to be a polyhedron $\Pi \subset \mathbb{R}^r_+$ which is the solution set of a linear system formed by taking, for every pair $(i_1, i_2) \in I_*^2$ and every $j = 1, \dots, m$, one of the following inequalities:

$$t_{i_1} + c_{i_1 j} \le t_{i_2} + c_{i_2 j}, \quad t_{i_1} + c_{i_1 j} \ge t_{i_2} + c_{i_2 j}. \tag{9.116}$$

Then for every $j = 1, \dots, m$ the order of magnitude of the sequence

$$t_i + c_{ij}, \quad i = 1, \dots, r$$

remains unchanged as t varies over a cell Π. Hence the index i_j satisfying $PTP(3)$ remains the same for all $t \in \Pi$; in other words, x^t (basic optimal solution of $P(t)$) equals a constant vector x^Π for all $t \in \Pi$. Let $\mathscr{P}$ be the collection of all cells defined that way. Since every $t \in \mathbb{R}^r_+$ satisfies one of the inequalities (9.116) for every $(i_1, i_2) \in I_*^2$ and every $j = 1, \dots, m$, the collection $\mathscr{P}$ covers all of $\mathbb{R}^r_+$. Let us estimate an upper bound of the number of cells in $\mathscr{P}$.

Observe that for any fixed pair $(i_1, i_2) \in I_*^2$ we have $t_{i_1} + c_{i_1 j} \le t_{i_2} + c_{i_2 j}$ if and only if $t_{i_1} - t_{i_2} \le c_{i_2 j} - c_{i_1 j}$. Let us sort the numbers $c_{i_2 j} - c_{i_1 j}, j = 1, \dots, m$, in increasing order

$$c_{i_2 j_1} - c_{i_1 j_1} \le c_{i_2 j_2} - c_{i_1 j_2} \le \dots \le c_{i_2 j_m} - c_{i_1 j_m}. \tag{9.117}$$

and let $\nu_{i_1, i_2}(j)$ be the position of $c_{i_2 j} - c_{i_1 j}$ in this ordered sequence.

Proposition 9.14 *A cell Π is characterized by a map $k_\Pi : I_*^2 \to \{1, \dots, m, m+1\}$ such that Π is the solution set of the linear system*

$$t_{i_1} + c_{i_1 j} \le t_2 + c_{i_2 j} \quad \forall (i_1, i_2) \in I_*^2 \text{ s.t. } \nu_{i_1, i_2}(j) \ge k_\Pi(i_1, i_2) \tag{9.118}$$

$$t_{i_1} + c_{i_1 j} \ge t_2 + c_{i_2 j} \quad \forall (i_1, i_2) \in I_*^2, \text{ s.t. } \nu_{i_1, i_2}(j) < k_\Pi(i_1, i_2) \tag{9.119}$$

Proof Let $\Pi \subset \mathbb{R}^r_+$ be a cell. For every pair (i_1, i_2) with $i_1 < i_2$ denote by $J_\Pi^{i_1, i_2}$ the set of all $j = 1, \dots, m$ such that the left inequality (9.116) holds for all $t \in \Pi$ and define

$$k_\Pi(i_1, i_2) = \begin{cases} \min\{\nu_{i_1, i_2}(j) | \, j \in J_\Pi^{i_1, i_2}\} & \text{if } J_\Pi^{i_1, i_2} \neq \emptyset \\ m + 1 & \text{otherwise} \end{cases}$$

It is easy to see that Π is then the solution set of the system (9.118) and (9.119). Indeed, let $t \in \Pi$. If $\nu_{i_1, i_2}(j) \ge k_\Pi(i_1, i_2)$, then $k_\Pi(i_1, i_2) \neq m + 1$, so $k_\Pi(i_{1,2}) = \nu_{i_1, i_2}(l)$ for some $l \in J_\Pi^{i_1, i_2}$. Then $t_{i_1} + c_{i_1 l} \le t_{i_1} + c_{i_2}$, hence $t_{i_1} - t_{i_2} \le c_{i_2 l} - c_{i_1 l}$ and since the relation $\nu_{i_1, i_2}(j) \ge \nu_{i_1, i_2}(l)$ means that $c_{i_2 j} - c_{i_1 j} \ge c_{i_2 l} - c_{i_1 l}$ it follows that $t_{i_1} - t_{i_2} \le c_{i_2 j} - c_{i_1 j}$, i.e., $t_{i_1} + c_{i_1 j} \le t_{i_2} + c_{i_2 j}$. Therefore (9.118) holds. On the other hand, if $\nu_{i_1, i_2}(j) < k_\Pi(i_1, i_2)$ then by definition of a cell $j \notin J_\Pi^{i_1, i_2}$, hence (9.119) holds, too [since from the definition of a cell, any $t \in \Pi$ must satisfy one of the inequalities (9.116)]. Thus every $t \in \Pi$ is a solution of the system (9.118) and (9.119). Conversely, if t satisfies (9.118) and (9.119) then for every $(i_1, i_2) \in I^2_*$, t satisfies the left inequality (9.116) for $j \in J^{i_1,i_2}_\Pi$ and the right inequality for $j \notin J^{i_1,i_2}_\Pi$, hence $t \in \Pi$. Therefore, each cell Π is determined by a map $k_\Pi : I^2_* \to \{1, \ldots, m+1\}$. Furthermore, it is easy seen that $k_\Pi \neq k_{\Pi'}$ for two different cells Π, Π'. Indeed, if $\Pi \neq \Pi'$ then at least for some $(i_1, i_2) \in I^2_*$ and some $j = 1, \ldots, m$, one has $j \in J^{i_1,i_2}_\Pi \setminus J^{i_1,i_2}_{\Pi'}$. Then $k_\Pi(i_1, i_2) \leq \nu_{i_1,i_2}(j)$ but $k_{\Pi'}(i_1, i_2) > \nu_{i_1,i_2}(j)$. □

Corollary 9.4 *The total number of cells is bounded above by* $(m+1)^{r(r-1)/2}$.

Proof The number of cells does not exceed the number of different maps $k : I^2_* \to \{1, \ldots, m+1\}$ and there are $(m+1)^{r(r-1)/2}$ such maps. □

To sum up, the proposed algorithm for solving $PTP(r)$ involves the following steps:

(1) Ordering the sequence $c_{i_2 j} - c_{i_1 j}, j = 1, \ldots, m$ for every pair $(i_1, i_2) \in I^2_*$ so as to determine $\nu_{i_1,i_2}(j), j = 1, \ldots, m, (i_1, i_2) \in I^2_*$.
(2) Computing the vector x^Π for every cell $\Pi \in \mathscr{P}$ ($\mathscr{P}$ is the collection of cells determinedly the maps $k_\Pi : I^2_* \to \{1, \ldots, m+1\}$.
(3) Computing the values $f(x^\Pi)$ and select Π^* according to (9.113).
   Steps (1) and (2) require obviously a number of elementary operations bounded by a polynomial in m, while step (3) requires $m^{r(r-1)/2}$ evaluations of $f(x)$.

9.11 Exercises

9.1 Show that a differentiable function $f(x)$ is K-monotonic on an open convex set $X \subset \mathbb{R}^n$ if and only if $-\nabla f(x) \in K^\circ \quad \forall x \in X$.

9.2 Consider the problem $\min\{f(x) | \, x \in D\}$ where D is a polyhedron in $\mathbb{R}^n$ with a vertex at 0, and $f : \mathbb{R}^n \to \mathbb{R}$ is a concave function such that for any real number $\gamma \leq f(0)$ the level set $C_\gamma = \{x | \, f(x) \geq \gamma\}$ has a polar contained in the cone generated by two linearly independent vectors $c^1, c^2 \in \mathbb{R}^n$. Show that this problem reduces to minimizing a concave function of two variables on the projection of D on $\mathbb{R}^2$. Develop an OA method for solving the problem.

9.3 Consider the problem

$$\min\{g(t) + \langle d, y\rangle | \, ta + By \leq c, t \geq 0, y \geq 0\}$$

where $t \in \mathbb{R}, y \in \mathbb{R}^n, a \in \mathbb{R}^m, B \in \mathbb{R}^{m\times n}, c \in \mathbb{R}^m_+$. Develop a parametric algorithm for solving the problem and compare with the polyhedral annexation method.

9.4 Show that a linearly constrained problem of the form

$$\min\{F(l(x))|\ x \in D\}$$

where D is a polyhedron in $\mathbb{R}^n$, $l(x)$ is an affine function, and $F(t)$ is an arbitrary quasiconcave real-valued function on an interval $\Delta \supset l(D)$ reduces to at most two linear programs. If $F(.)$ is monotonic on $l(D)$ then the problem is in fact equivalent to even a mere linear program.

9.5 Solve the plant location problem

$$\begin{aligned}
\min\ & \sum_{i=1}^{3} f_i(x_i) + \sum_{i=1}^{3}\sum_{j=1}^{5} d_{ij}y_{ij} \\
\text{s.t.}\ & x_1 + x_2 + x_3 = 203 \\
& y_{1j} + y_{2j} + y_{3j} = b_j \quad j = 1, \ldots, 5 \\
& y_{i1} + \ldots + y_{i5} = x_i \quad i = 1, 2, 3 \\
& x_i \geq 0; \quad y_{ij} \geq 0 \quad i = 1, 2, 3; j = 1, \ldots, 5.
\end{aligned}$$

where $\quad f_i(t) = 0$ if $t = 0$, $f_i(t) = r_i + s_i t$ if $t > 0$

$$r_1 = 1, s_1 = 1.7; \quad r_2 = 88, s_2 = 8.4; \quad r_3 = 39; s_3 = 4.7$$

$$b = (62, 65, 51, 10, 15); \quad d = [d_{ij}] = \begin{bmatrix} 6 & 66 & 68 & 81 & 4 \\ 40 & 20 & 34 & 83 & 27 \\ 90 & 22 & 82 & 17 & 8 \end{bmatrix}$$

9.6 Suppose that details of n different kinds must be manufactured. The production cost of t items of i-th kind is a concave function $f_i(t)$. The number of items of i-th kind required is $a_i > 0$, but one item of i-th kind can be replaced by one item of any j-th kind with $j > i$. Determine the number x_i of items of i-th kind, $i = 1, \ldots, n$, to be manufactured so as to satisfy the demands with minimum cost. By setting $\alpha_i = \sum_{j=1}^{i} a_j$, this problem can be formulated as:

$$\begin{aligned}
& \min \sum_{i=1}^{n} f_i(x_i) \quad \text{s.t.} \\
& x_1 \leq \alpha_1,\ x_1 + x_2 \leq \alpha_2, \ldots, x_1 + \ldots + x_n \leq \alpha_n \\
& x_1, \ldots, x_n \geq 0
\end{aligned}$$

(Zukhovitzki et al. 1968). Solve this problem.

(*Hint*: Any extreme point x of the feasible polyhedron can be characterized by the condition: for each $i = 1, \ldots, n$, either $x_i = 0$ or $x_1 + \ldots + x_i = \alpha_i$).

9.7 Develop an algorithm for solving the problem

$$\min\{c^0x + (c^1x)^{1/2} \times (c^2x)^{1/3} | \quad x \in D\}$$

where $D \subset \mathbb{R}^n_+$ is a polytope, $c^1, c^2 \in \mathbb{R}^n_+$, $c^0 \in \mathbb{R}^n_-$ and cx denotes the inner product of c and x. Apply the algorithm to a small numerical example with $n = 2$.

9.8 Solve the numerical Example 9.10 (page 317).

9.9 Solve by polyhedral annexation:

$$\begin{aligned}
&\min\ (0.5x_1 + 0.75x_2 + 0.75x_3 + 1.25x_4) \quad \text{s.t.}\\
&0.25x_2 + 0.25x_3 + 0.75x_4 \leq 2\\
&-0.25x_2 - 0.25x_3 + 0.25x_4 \leq 0\\
&x_1 - 0.5x_2 + 0.5x_3 + 1.5x_4 \leq 1.5\\
&h(x) + 4x_1 - x_2 = 0\\
&x = (x_1, \ldots, x_4) \geq 0
\end{aligned}$$

where $h(x)$ is the optimal value of the linear program

$$\begin{aligned}
&\min\ (-4y_1 + y_2) \quad \text{s.t.}\\
&y_1 + y_2 \leq 1.5 - 0.5x_2 - 0.5x_3 - 1.5x_4\\
&y_2 \leq x_1 + x_2; \quad y_1, y_2 \geq 0
\end{aligned}$$

(Hint: Exploit the monotonicity property of $h(x)$)

9.10 Solve the problem

$$\begin{aligned}
&\min \sum_{i=1}^{n} \frac{x_i}{y_i + \eta} \quad \text{s.t.}\\
&\sum_{i=1}^{n} x_i = a, \quad \sum_{i=1}^{n} y_i = b\\
&0 \leq x_i \leq \gamma, \quad 0 \leq y_i \leq 1 \quad i = 1, \ldots, n
\end{aligned}$$

where $0 < \eta, 0 < a \leq n\gamma, 0 < b \leq n$.

Hint: Reduce the problem to the concave minimization problem

$$\min\left\{h(x) | \sum_{i=1}^{n} x_i = a, 0 \leq x_i \leq \gamma\right\}$$

where

$$h(x) = \min \left\{ \sum_{i=1}^{n} \frac{x_i}{y_i + \eta} \mid \sum_{i=1}^{n} y_i = b, 0 \leq y_i \leq 1 \quad i = 1, \ldots, n \right\}$$

and use the results on the special concave minimization CPL(1).

9.11 Solve the problem in Example 5.5 (page 142) where the constraints (5.24) are replaced by

$$\sum_{i=1}^{n} x_i = a, \quad \sum_{i=1}^{n} y_i = b.$$

Chapter 10
Nonconvex Quadratic Programming

10.1 Some Basic Properties of Quadratic Functions

As pointed out in the preceding chapters, in deterministic global optimization it is of utmost importance to take advantage of the particular mathematical structure of the problem under study. There are in general two aspects of nonconvexity deserving special attention. First, the rank of nonconvexity, i.e., roughly speaking, the number of nonconvex variables. Second, the degree of nonconvexity, i.e., the extent to which the variables are nonconvex. In Chap. 9 we have discussed decomposition methods for handling low rank nonconvex problems. The present Chap. 10 is devoted to nonconvex optimization problems which involve only linear or quadratic functions, i.e., in a sense functions with lowest degree of nonconvexity.

The importance of quadratic optimization models is due to several reasons:

- Quadratic functions are the simplest smooth d.c. functions whose derivatives are readily available and easy to manipulate.
- Any twice differentiable function can be approximated by a quadratic function in the neighborhood of a given point, so in a sense quadratic models are the most natural.
- Numerous applications in economics, engineering, and other fields lead to quadratic nonconvex optimization problems. Certain combinatorial optimization problems can also be studied as quadratic optimization problems. For instance, a 0–1 constraint of the form $x_i \in \{0, 1\},\ i = 1, \ldots, p$ can be written as the quadratic constraint $\sum_{i=1}^{p} x_i(x_i - 1) \geq 0,\ 0 \leq x_i \leq 1$.

Before discussing specialized approaches to quadratic optimization it is convenient to recall some of the most elementary properties of quadratic functions.

1. Let Q be a symmetric matrix. Then

$$Q = U\Lambda U^T, \quad \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n),$$

H. Tuy, *Convex Analysis and Global Optimization*, Springer Optimization and Its Applications 110, DOI 10.1007/978-3-319-31484-6_10

where $UU^T = I$, $U = (u^1, \ldots, u^n)$ is the matrix of normalized eigenvectors of Q, and $\lambda_1, \ldots, \lambda_n$ are the eigenvalues. In particular, Q is positive definite (semidefinite) if and only if $\lambda_i > 0$ ($\lambda_i \geq 0$, respectively) for all $i = 1, \ldots, n$. The number $\rho(Q) = \max_{i=1,\ldots,n} |\lambda_i|$ is called the *spectral radius* of Q.

2. For any symmetric matrix Q the matrix $Q+rI$ is positive semidefinite if $r \geq \rho(Q)$ (Proposition 4.11); in other words, the function $f(x) = \langle x, Qx\rangle + r\|x\|^2$ is convex if $r \geq \rho(Q)$.
3. Any system of quadratic inequalities

$$f_i(x) = \frac{1}{2}\langle x, Q^i x\rangle + \langle c^i, x\rangle + d_i \leq 0 \quad i = 1, \ldots, m \tag{10.1}$$

is equivalent to a single inequality

$$g(x) - r\|x\|^2 \leq 0,$$

where $g(x)$ is convex and $r \geq \max_{i=1,\ldots,m} \frac{1}{2}\rho(Q^i)$ (Corollary 4.3).

4. The convex envelope of the function $f(x, y) = xy$ on a rectangle $\{(x, y) \in \mathbb{R}^2 | \, p \leq x \leq q, r \leq y \leq s\}$ is the function

$$\max\{rx + py - pr, sx + qy - qs\}$$

(Corollary 4.6).

5. A convex minorant of a quadratic function $f(x) = \frac{1}{2}\langle x, Qx\rangle + \langle c, x\rangle$ on a rectangle $M = \{x | \, p \leq x \leq q\}$ is

$$\varphi_M(x) = f(x) + r\sum_{i=1}^{n}(x_i - p_i)(x_i - q_i),$$

where r is any positive number satisfying $r \geq \frac{1}{2}\rho(Q)$.

Indeed, the function $f(x) + r\sum_{i=1}^n x_i^2$ is convex for $r \geq \frac{1}{2}\rho(Q)$, while $x \in M$ implies that $\sum_{i=1}^n (x_i - p_i)(x_i - q_i) \leq 0$.

We next mention some less known properties which are important in the computational study of quadratic problems. Recall that the rank of a quadratic form is equal to the rank of its matrix.

Proposition 10.1 *Let Q be an $n \times n$ matrix of rank k. Then*

$$Q = \sum_{i=1}^{k}(a^i)(b^i)^T, \quad i.e., \; \langle x, Qy\rangle = \sum_{i=1}^{k}\langle a^i, x\rangle\langle b^i, y\rangle,$$

where a^i, b^i are $n \times 1$ vectors.

Proof Let $Q = (a^1, \ldots, a^n)$ with $a^1, \ldots, a^k$ linearly independent. Then every a^j is a linear combination of $a^1, \ldots, a^k$, i.e., $a^j = \sum_{i=1}^k b^i_j a^i$ $(j = 1, \ldots, n)$, for some $b^i_j \in \mathbb{R}, i = 1, \ldots, k$. Hence, letting $b^i = (b^i_1, \ldots, b^i_n)^T$, we have $Q = \sum_{i=1}^k (a^i)(b^i)^T$. □

Consequence *A quadratic function of rank k is of the form*

$$f(x) = \frac{1}{2}\sum_{i=1}^{k}\langle a^i, x\rangle\langle b^i, x\rangle + \langle c, x\rangle. \qquad \square$$

Proposition 10.2 *Let $f(x) : \mathbb{R}^n \to \mathbb{R}$ be a quadratic function, $H \subset \mathbb{R}^n$ an affine manifold. Any local minimizer $\overline{x}$ of $f(x)$ on H is a global minimizer on H. If $f(x)$ is bounded below on H it is convex on H.*

Proof The first part follows from Proposition 4.18. If a quadratic function $f(x)$ is bounded below on an affine manifold H it must be so on every line $\Gamma \subset H$, hence convex on this line, and consequently convex on the whole H. □

Proposition 10.3 *A quadratic function $f(x)$ is strictly convex on an affine manifold H if and only if it is coercive on H, i.e., $f(x) \to +\infty$ as $x \in H$, $\|x\| \to +\infty$ or equivalently, for any real number η the set $\{x \in H | f(x) \leq \eta\}$ is compact.*

Proof If a quadratic function $f(x)$ is strictly convex on H, then it is strictly convex on every line in H, so for any $\eta \in \mathbb{R}$ the set $\{x \in H | f(x) \leq \eta\}$ (which is closed and convex) cannot contain any halfline, hence must be compact, i.e., $f(x)$ is coercive on H. Conversely, if a quadratic function $f(x)$ is not strictly convex on H, it is not strictly convex on some line $\Gamma \subset H$, so it is concave on Γ and hence cannot be coercive. □

10.2 The S-Lemma

Quadratic optimization with a single quadratic constraint began to be studied as early as in the middle of the last century in connection with nonlinear control theory. A famous result obtained by Yakubovich in these days and named by him the S-Lemma (Yakubovich 1977) turned out to play a crucial role in the subsequent development of quadratic optimization.

There are numerous proofs and extensions of the S-Lemma. Below, following (Tuy and Tuan 2013) we present an extension and a proof of this proposition based on a special minimax theorem which is more suitable for studying quadratic inequalities than classical minimax theorems.

Theorem 10.1 *Let C be a closed subset of $\mathbb{R}^n$, D an interval (closed or open) of $\mathbb{R}$ and $L(x, y) : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ a continuous function. Assume the following conditions:*

(i) *for every $x \in \mathbb{R}^n$ the function $L(x, .)$ is concave;*
(ii) *for every $y \in D$ any local minimizer of $L(., y)$ on C is a global minimizer of $L(., y)$ on C;*
(iii) *there exist $y^* \in D$ such that $L(x, y^*) \to +\infty$ as $x \in C$, $\|x\| \to +\infty$.*

Then we have the minimax equality

$$\inf_{x\in C} \sup_{y\in D} L(x, y) = \sup_{y\in D} \inf_{x\in C} L(x, y). \tag{10.2}$$

Furthermore, if $\inf_{x\in C} \sup_{y\in D} L(x, y) < +\infty$ *there exists $\overline{x} \in C$ satisfying*

$$\sup_{y\in D} L(\overline{x}, y) = \min_{x\in C} \sup_{y\in D} L(x, y). \tag{10.3}$$

Proof Define $\gamma := \inf_{x\in C} \sup_{y\in D} L(x, y)$, $\eta := \sup_{y\in D} \inf_{x\in C} L(x, y)$. We can assume $\eta < +\infty$, otherwise the equality (10.2) is trivial. Take an arbitrary real number $\alpha > \eta$ and for every $y \in D$ let $C_\alpha(y) := \{x \in C|\ L(x, y) \le \alpha\}$. We first prove that for any segment $\Delta = [a, b] \subset D$ the following holds:

$$\cap_{y\in\Delta} C_\alpha(y) \ne \emptyset. \tag{10.4}$$

It can always be assumed that $y^* \in (a, b)$. Since $\inf_{x\in C} L(x, y^*) \le \eta < \alpha$ there exists $x \in C$ satisfying $L(x, y^*) < \alpha$, and by continuity there exist u, v such that $u < y^* < v$ and $L(x, y) < \alpha\ \forall y \in [u, v]$. Therefore, setting

$$s = \sup\{y \in [u, b]|\ \cap_{u\le z\le y} C_\alpha(z) \ne \emptyset\}. \tag{10.5}$$

we have $s \ge v$. By definition of s there exists a sequence $y_k \nearrow s$, such that $\overline{C}_k := \cap_{u\le y\le y_k} C_\alpha(y) \ne \emptyset$. In view of assumption (iii) $C_\alpha(y^*)$ is compact, so $\overline{C}_k$, $k = 1, 2, \ldots$, form a nested sequence of nonempty closed subsets of the compact set $C_\alpha(y^*)$. Therefore, there exists $x^0 \in \cap_{k=1}^\infty \overline{C}_k$, i.e., satisfying $L(x^0, y) \le \alpha\ \forall y \in [u, s)$. Since $L(x^0, y_k) \le 0\ \forall k$, letting $k \to +\infty$ yields $L(x^0, s) \le \alpha$. So $L(x^0, y) \le \alpha\ \forall y \in [u, s]$. We contend that $s = b$. Indeed, if $s < b$ we cannot have $L(x^0, s) < \alpha$ for then by continuity there would exist $q > s$ satisfying $L(x^0, y) < \alpha\ \forall y \in [s, q)$, conflicting with (10.5). Therefore, if $s < b$ then necessarily $L(x^0, s) = \alpha$. Furthermore, for any ball W centered at x^0 we cannot have $\alpha = \min_{x\in C\cap W} L(x, s)$ for then assumption (ii) would imply that $\alpha = L(x^0, s) = \min_{x\in C} L(x, s)$, conflicting with $\alpha > \inf_{x\in C} L(x, s)$. So there exists $x^k \in C$ such that $x^k \to x^0$ and $L(x^k, s) < \alpha$. If for some $y \in (s, b]$ and some k we had $L(x^k, y) > \alpha$, then, since $L(x^k, s) < \alpha$ assumption (i) would imply that $L(x^k, z) < \alpha\ \forall z \in [u, s]$ and, moreover, by continuity $L(x^k, q) < \alpha\ \forall y \in [s, q]$ for some $q > s$, conflicting with (10.5). Therefore, $L(x^k, y) \le \alpha\ \forall k$, hence letting $x^k \to x^0$ yields $L(x^0, y) \le \alpha,\ \forall y \in (s, b]$. This proves that $s = b$ and so

$$\cap_{u\le y\le b} C_\alpha(y) \ne \emptyset. \tag{10.6}$$

In an analogous manner, setting

$$t = \inf\{y \in [u, b] |\ \cap_{y \le z \le b} C_\alpha(z) \neq \emptyset\}.$$

we can show that $t = a$, i.e., $\cap_{a \le y \le b} C_\alpha(y) \neq \emptyset$, proving (10.4).

It is now easy to complete the proof. By assumption (iii) the set $C_\alpha(y^*)$ is compact. Since any finite set $E \subset D$ is contained in some segment Δ it follows from (10.4) that the family $C_\alpha(y) \cap C_\alpha(y^*),\ y \in D$, has the finite intersection property. So there exists $x \in C$ satisfying $L(x, y) \le \alpha\ \forall y \in D$. Taking $\alpha = \eta + 1/k$ we see that for each $k = 1, 2, \ldots,$ there exists $x^k \in C$ satisfying $L(x^k, y) \le \eta + 1/k\ \forall y \in D$. Since, in particular, $L(x^k, y^*) \le \eta + 1/k < \eta + 1$, i.e., $x^k \in C_{\eta+1}(y^*)\ \forall k$, while the set $C_{\eta+1}(y^*)$ is compact, the sequence $\{x^k\}$ has a cluster point $\bar{x} \in C$ satisfying $L(\bar{x}, y) \le \eta\ \forall y \in D$. Noting that $\eta \le \gamma$ this implies (10.3). □

In the above theorem D is an interval of $\mathbb{R}$. The following proposition deals with the case $D \subset \mathbb{R}^m$ with $m \ge 1$:

Corollary 10.1 *Let C be a closed subset of $\mathbb{R}^n$, D a convex subset of $\mathbb{R}^m$, and $L(x, y) : \mathbb{R}^n \times D \to \mathbb{R}$ be a continuous function. Assume the following conditions:*

(i) *for every $x \in \mathbb{R}^n$ the function $L(x, .)$ is affine;*
(ii) *for every $y \in D$ any local minimizer of $L(x, y)$ on C is a global minimizer of $L(x, y)$ on C;*
(iii*) *there exists $y^* \in D$ satisfying $L(x, y^*) \to +\infty$ as $x \in C,\ \|x\| \to +\infty$ and such that, moreover*

$$\inf_{x \in C} L(x, y^*) = \max_{y \in D} \inf_{x \in C} L(x, y) \quad \& \quad L(x^*, y^*) = \min_{x \in C} L(x, y^*) \tag{10.7}$$

for a unique $x^* \in C$.

Then we have the minimax equality

$$\min_{x \in C} \sup_{y \in D} L(x, y) = \max_{y \in D} \inf_{x \in C} L(x, y). \tag{10.8}$$

Proof Consider an arbitrary $\alpha > \eta := \sup_{y \in D} \inf_{x \in C} L(x, y)$ and for every $y \in D$ let $C_\alpha(y) := \{x \in C |\ L(x, y) \le \alpha\}$. By assumption (iii*) $\inf_{x \in C} L(x, y^*) = \eta$, hence $C_\eta(y^*) = \{x \in C |\ L(x, y^*) = \eta\} = \{x^*\}$.

For every $y \in D$ let $D_y := [y^*, y] = \{(1 - t)y^* + ty |\ 0 \le t \le 1\} \subset \mathbb{R}$ and $L_y(x, t) := L(x, (1 - t)y^* + ty)$, which is a function on $C \times D_y$. It can easily be verified that all conditions of Theorem 1 are satisfied for C, D_y and $L_y(x, t)$. Since $\alpha > \eta \ge \sup_{t \in D_y} \inf_{x \in C} L(x, t)$, it follows that

$$\emptyset \neq C_\alpha(y^*) \cap C_\alpha(y) \subset C_\alpha(y^*) \subset C.$$

Furthermore, since by assumption $L(x, y^*) \to +\infty$ as $x \in C, \|x\| \to +\infty$, the set $C_\alpha(y^*)$ is compact and so is $C_\alpha(y^*) \cap C_\alpha(y)$ because $C_\alpha(y)$ is closed. Letting $\alpha \to \eta+$ then yields

$$\emptyset \neq C_\eta(y^*) \cap C_\eta(y) \subset C_\eta(y^*) = \{x^*\}.$$

Thus,

$$L(x^*, y) \leq \eta \quad \forall y \in D$$

and hence,

$$\inf_{x \in C} \sup_{y \in D} L(x, y) \leq \eta$$

completing the proof. □

Remark 10.1 (a) When C is an affine manifold and $L(x, y)$ is a quadratic function in x it is superfluous to require the uniqueness of x^* in assumption (iii*). In fact, in that case the uniqueness of x^* follows from the condition $L(x, y^*) \to +\infty$ as $x \in C, \ \|x\| \to +\infty$ (which implies that $x \mapsto L(x, y^*)$ is strictly convex on the affine manifold C).

(b) Conditions (iii*) is satisfied, provided there exists a unique pair $(x^*, y^*) \in C \times D$ such that

$$L(x^*, y^*) = \max_{y \in D} \min_{x \in C} L(x, y) \text{ and } L(x, y^*) \to +\infty \text{ as } x \in C, \ \|x\| \to +\infty.$$

Indeed, let $\overline{y}$ satisfy $\min_{x \in C} L(x, \overline{y}) = \max_{y \in D} \min_{x \in C} L(x, y)$, and $L(x, \overline{y}) \to +\infty$ as $x \in C, \|x\| \to +\infty$. Since $L(., \overline{y})$ is convex, there is $\overline{x}$ satisfying $L(\overline{x}, \overline{y}) = \min_{x \in C} L(x, \overline{y})$. Then $L(\overline{x}, \overline{y}) = \max_{y \in D} \min_{x \in C} L(x, y)$, so $(x^*, y^*) = (\overline{x}, \overline{y})$ by uniqueness condition. Hence $\min_{x \in C} L(x, y^*) = \max_{y \in D} \min_{x \in C} L(x, y)$, i.e., (iii*) holds.

Assumption (ii) in Theorem 10.1 or Corollary 10.1 is obviously satisfied if $L(., y)$ is a convex function and C is a convex set (then Theorem 10.1 reduces to a classical minimax theorem). However, as will be seen shortly, this assumption is also satisfied in many other cases of interest in quadratic optimization.

Corollary 10.2 *Let $f(x)$ and $g(x)$ be quadratic functions on $\mathbb{R}^n$, $L(x, y) = f(x) + yg(x)$ for $x \in \mathbb{R}^n, y \in \mathbb{R}$. Then for every line Γ in $\mathbb{R}^n$ there holds the minimax equality*

$$\inf_{x \in \Gamma} \sup_{y \in \mathbb{R}_+} L(x, y) = \sup_{y \in \mathbb{R}_+} \inf_{x \in \Gamma} L(x, y). \tag{10.9}$$

If $g(x)$ is not affine on Γ there also holds the minimax equality

$$\inf_{x \in \Gamma} \sup_{y \in \mathbb{R}} L(x, y) = \sup_{y \in \mathbb{R}} \inf_{x \in \Gamma} L(x, y). \tag{10.10}$$

Proof If $f(x)$ and $g(x)$ are both concave on Γ and one of them is not affine then clearly $\inf\{f(x)|\ g(x) \leq 0, x \in \Gamma\} = -\infty$, hence $\inf_{x\in\Gamma}\sup_{y\in\mathbb{R}_+} L(x,y) = \inf\{f(x)|\ g(x) \leq 0, x \in \Gamma\} = -\infty$, and (10.9) is trivial. If $f(x)$ and $g(x)$ are both affine on Γ, (10.9) is a consequence of linear programming duality. Otherwise, either $f(x)$ or $g(x)$ is strictly convex on Γ. Then there exists $y^* \in \mathbb{R}_+$ such that $f(x) + y^*g(x)$ is strictly convex on Γ, hence such that $L(x, y^*) \to +\infty$ as $x \in \Gamma$, $\|x\| \to +\infty$. Since by Proposition 10.2 condition (ii) in Theorem 10.1 is satisfied, (10.9) follows by Theorem 10.1 with $C = \Gamma, D = \mathbb{R}_+$.

If $g(x)$ is not affine on Γ, there always is $y^* \in \mathbb{R}$ such that $f(x)+y^*g(x)$ is strictly convex on Γ ($y^* > 0$ when $g(x)$ is strictly convex on Γ and $y^* < 0$ when $g(x)$ is strictly concave on Γ). Therefore all conditions of Theorem 10.1 are satisfied again for $C = \Gamma$ and $D = \mathbb{R}$ and (10.10) follows. □

We can now state a generalized version of the S-Lemma which is in fact valid for arbitrary lower semi-continuous functions $f(x)$ and $g(x)$, not necessarily quadratic.

Theorem 10.2 (Generalized S-Lemma (Tuy and Tuan 2013)) *Let $f(x)$ and $g(x)$ be quadratic, functions on $\mathbb{R}^n$, W a closed subset of $\mathbb{R}^n$, D a closed subset of $\mathbb{R}$ and $L(x,y) = f(x) + yg(x)$ for $y \in \mathbb{R}$. Assume $\gamma := \inf_{x\in W}\sup_{y\in D} L(x,y) > -\infty$ and the following conditions are satisfied:*

(i) *Either : $D = \mathbb{R}_+$ and there exists $x^* \in W$ such that $g(x^*) < 0$*
Or : $D = \mathbb{R}$ and there exist $a, b \in W$ such that $g(a) < 0 < g(b)$.
(ii) *For any two points $a, b \in W$ such that $g(a) < 0 < g(b)$ we have*

$$\sup_{y\in D}\inf_{x\in\{a,b\}} L(x,y) \geq \gamma.$$

Then there exists $\overline{y} \in D$ satisfying

$$\inf_{x\in W} L(x,\overline{y}) = \gamma, \tag{10.11}$$

and so $\inf_{x\in W}\sup_{y\in D} L(x,y) = \max_{y\in D}\inf_{x\in W} L(x,y)$.

Proof We first show that for any real $\alpha < \gamma$ and any finite set $E \subset W$ there exists $y \in D$ satisfying

$$\inf_{x\in E} L(x,y) \geq \alpha. \tag{10.12}$$

Since $x \in E \subset W, g(x) = 0$ implies $L(x,y) = f(x) + yg(x) = f(x) \geq \gamma > \alpha\ \forall y \in D$, it suffices to prove (10.12) for any finite set $E \subset \{x \in W|\ g(x) \neq 0\}$. Then $E = E_1 \cup E_2$ where $E_1 := \{x \in E|\ g(x) < 0\}$, $E_2 = \{x \in E|\ g(x) > 0\}$. For every $x \in E$ define $\theta(x) := \frac{\alpha - f(x)}{g(x)}$. Clearly

- for every $x \in E_1$: $\theta(x) > 0$ and $[\ L(x,y) \geq \alpha \Leftrightarrow y \leq \theta(x)\ ]$;
- for every $x \in E_2$: $L(x,y) \geq \alpha \Leftrightarrow y \geq \theta(x)$.

Consequently, if $E_1 \neq \emptyset$ then $\min_{x\in E_1} \theta(x) > 0$ and we have $L(x, y) \geq \alpha\ \forall x \in E_1$ when $0 \leq y \leq \min_{x\in E_1} \theta(x)$; if $E_2 \neq \emptyset$, then we have $L(x, y) \geq \alpha\ \forall x \in E_2$ when $y \geq \max_{x\in E_2} \theta(x)$. Further, if $E_i \neq \emptyset, i = 1, 2$, then by assumption (ii) with $a \in \mathrm{argmin}_{x\in E_1}\theta(x)$, $b \in \mathrm{argmax}_{x\in E_2}\theta(x)$ there exists $y \in D$ satisfying $\inf_{x\in\{a,b\}} L(x, y) \geq \alpha$, i.e., such that

$$\max_{x\in E_2} \theta(x) = \theta(b) \leq y \leq \theta(a) = \min_{x\in E_1} \theta(x),$$

hence

$$L(x, y) \geq \alpha\ \forall x \in E_2, \quad L(x, y) \geq \alpha\ \forall x \in E_1.$$

Thus in any case for any finite set $E \subset W$ there exists $y \in D$ satisfying (10.12).

Setting $D(x) := \{y \in D|\ L(x, y) \geq \alpha\}$ we then have $\cap_{x\in E} D(x) \neq \emptyset$ for any finite set $E \subset W$. In other words, the collection of closed sets $D(x), x \in W$, has the finite intersection property. If assumption (i) holds then

- either $D = \mathbb{R}_+$ and $f(x^*) + yg(x^*) \to -\infty$ as $y \to +\infty$, so the set $D(x^*) = \{y \in \mathbb{R}_+|\ L(x^*, y) \geq \alpha\}$ is compact
- or $D = \mathbb{R}$ and $f(a) + yg(a) \to -\infty$ as $y \to +\infty$, $f(b) + yg(b) \to -\infty$ as $y \to -\infty$.

So the set $D(a) = \{y \in \mathbb{R}|\ L(a, y) \geq \alpha\}$ is bounded above, while $D(b)$ is bounded below; hence the set $D(a) \cap D(b)$ is compact.

Consequently, $\cap_{x\in W} D(x) \neq \emptyset$. For every $\alpha_k = \gamma - 1/k,\ k = 1, 2, \ldots$ there exists $y_k \in \cap_{x\in W}\{y \in D|\ L(x, y) \geq \alpha_k\}$, i.e., $y^k \in D$ such that $L(x, y_k) \geq \alpha_k\ \forall x \in W$. Obviously every y_k is contained in the compact set $\cap_{x\in W}\{y \in D|\ L(x, y) \geq \alpha_1\}$, so there exists $\overline{y} \in D$ such that, up to a subsequence $y_k \to \overline{y}$ $(k \to +\infty)$. Then $L(x, \overline{y}) \geq \gamma\ \forall x \in W$, hence, $\inf_{x\in W} L(x, \overline{y}) = \gamma$. This proves (10.11). □

Theorem 10.2 yields *the most general version of S-Lemma*. In fact, if W is an affine manifold in $\mathbb{R}^n$ and $D = \mathbb{R}_+$ condition (ii) is automatically fulfilled because for every line Γ connecting a and b in W we have, by Corollary 10.2,

$$\begin{aligned} \sup_{y\in\mathbb{R}_+} \inf_{x\in\{a,b\}} L(x, y) &\geq \sup_{y\in\mathbb{R}_+} \inf_{x\in\Gamma} L(x, y) = \inf_{x\in\Gamma} \sup_{y\in\mathbb{R}_+} L(x, y) \\ &\geq \inf_{x\in W} \sup_{y\in\mathbb{R}_+} L(x, y) = \gamma. \end{aligned}$$

Also using Corollary 10.2, condition (ii) is automatically fulfilled if W is an affine manifold, $D = \mathbb{R}$ and $g(x)$ is a quadratic function $g(x) = \langle x, Q^1 x\rangle + \langle c^1, x\rangle + d_1$ with either Q^1 nonsingular or with $c^1 = 0$ and $d_1 = 0$ ($g(x)$ is homogeneous). Indeed, either of the latter conditions implies that $g(x)$ is not affine on any line Γ in W and hence, by Corollary 10.2,

$$\sup_{y\in\mathbb{R}} \inf_{x\in\{a,b\}} L(x, y) \geq \sup_{y\in\mathbb{R}} \inf_{x\in\Gamma} L(x, y) = \inf_{x\in\Gamma} \sup_{y\in\mathbb{R}} L(x, y) \geq \inf_{x\in W} \sup_{y\in\mathbb{R}} L(x, y) = \gamma.$$

Corollary 10.3 *Let $f(x)$ and $g(x)$ be two quadratic functions on $\mathbb{R}^n$.*

(i) *Let H be an affine manifold in $\mathbb{R}^n$ and assume there exists $x^* \in H$ such that $g(x^*) < 0$. Then the system $x \in H,\ f(x) < 0,\ g(x) \leq 0$ is inconsistent if and only if there exists $\overline{y} \in \mathbb{R}_+$ such that $f(x) + \overline{y}g(x) \geq 0\ \forall x \in H$.*

(ii) *Let H be an affine manifold in $\mathbb{R}^n$ and assume that $g(x)$ is either homogeneous or $g(x) = \langle x, Q^1x\rangle + \langle c^1, x\rangle + d_1$ with Q^1 nonsingular and that, moreover, $g(x)$ takes both positive and negative values on H. Then the system $x \in H,\ f(x) < 0,\ g(x) = 0$ is inconsistent if and only if there exists $\overline{y} \in \mathbb{R}$ such that $f(x) + \overline{y}g(x) \geq 0\ \forall x \in H$.*

(iii) *Let H be a subspace of $\mathbb{R}^n$ and assume both $f(x),\ g(x)$ are homogeneous. Then the system $x \in H \setminus \{0\},\ f(x) \leq 0,\ g(x) \leq 0$ is inconsistent if and only if there exists $(y_1, y_2) \in \mathbb{R}^2_+$ such that $y_1f(x) + y_2g(x) > 0\ \forall x \in H \setminus \{0\}$.*

Proof Since the 'if' part in each of the three assertions is obvious, it suffices to prove the 'only if' part. Let $L(x, y) = f(x) + yg(x)$.

(i) This is a special case of Theorem 10.2 with $D = \mathbb{R}_+, W = H$. In fact,

$$\gamma := \inf\{f(x)|\ g(x) \leq 0,\ x \in H\} = \inf_{x\in H} \sup_{y\in\mathbb{R}_+} L(x, y),$$

so inconsistency of the system $\{x \in H,\ f(x) < 0,\ g(x) \leq 0\}$ means $\gamma \geq 0$. Therefore, by Theorem 10.2 there exists $\overline{y} \in \mathbb{R}_+$ such that $f(x) + \overline{y}g(x) \geq \gamma \geq 0,\ \forall x \in H$.

(ii) This is a special case of Theorem 10.2 with $D = \mathbb{R},\ W = H$. In fact, inconsistency of the system $\{x \in H,\ f(x) \leq 0,\ g(x) = 0\}$ means $\gamma := \inf\{f(x)|\ g(x) = 0,\ x \in H\} = \inf_{x\in H} \sup_{y\in\mathbb{R}} L(x, y) \geq 0$.

(iii) With H being a subspace and both $f(x),\ g(x)$ homogeneous, the system $\{x \in H \setminus \{0\},\ f(x) \leq 0,\ g(x) \leq 0\}$ is inconsistent only if so is the system $\{x \in H \setminus \{0\},\ f(x)/\langle x, x\rangle \leq 0,\ g(x)/\langle x, x\rangle \leq 0\}$.

Setting $z = x/\sqrt{\langle x, x\rangle}$ we have the implication: $[z \in H,\ \langle z, z\rangle = 1,\ g(z) \leq 0] \Rightarrow f(z) > 0$. Since the set $\{x \in H|\ \langle x, x\rangle = 1\}$ is compact it follows that $\gamma_1 := \inf\{f(z)|\ g(z) \leq 0,\ \langle z, z\rangle = 1\} > 0$ and noting that $f(0) = g(0) = 0$ we can write $\inf\{f(x) - \gamma_1\langle x, x\rangle|\ g(x) \leq 0, x \in H\} \geq 0$. If there exists $x \in H$ satisfying $g(x) < 0$ then by (i) there exists $y_2 \in \mathbb{R}_+$ satisfying $f(x) - \gamma_1\langle x, x\rangle + y_2g(x) \geq\ \ \forall x \in H$, hence $f(x) + y_2g(x) > 0\ \forall x \in H \setminus \{0\}$.

Similarly, with the roles of $f(x)$ and $g(x)$ interchanged, if there exists $x \in H$ satisfying $f(x) < 0$, then there exists $y_1 \in \mathbb{R}_+$ satisfying $y_1f(x) + g(x) \geq 0\ \forall x \in H \setminus \{0\}$.

Finally, in the case $g(x) \geq 0\ \forall x \in H$ and $f(x) \geq 0\ \forall x \in H$, since there cannot exist $x \in H \setminus \{0\}$ such that $f(x) = 0$ and $g(x) = 0$ we must have $f(x) + g(x) > 0\ \forall x \in H \setminus \{0\}$, completing the proof. □

Assertion (i) of Corollary 10.3 is a result proved in Polik and Terlaky (2007) and later termed the *'generalized S-Lemma'* in Jeyakumar et al. (2009). Assertion (ii) for homogeneous $g(x)$ is the *S-Lemma with equality constraint* while Assertion

(iii) is the *S-Lemma with nonstrict inequalities*. So Theorem 10.2 is indeed the most general version of the S-Lemma. Condition (i) is a kind of *generalized Slater condition* and the proof given above is a simple elementary proof for the S-Lemma in its full generality.

Remark 10.2 By Proposition 10.2 the function $f(x) + \bar{y}g(x)$ in assertions (i) and (ii) of the above corollary is convex on H.

10.3 Single Constraint Quadratic Optimization

Let H be an affine manifold in $\mathbb{R}^n$, and $f(x), g(x)$ two quadratic functions on $\mathbb{R}^n$ defined by

$$f(x) := \frac{1}{2}\langle x, Q^0 x\rangle + \langle c^0, x\rangle, \quad g(x) = \frac{1}{2}\langle x, Q^1 x\rangle + \langle c^1, x\rangle + d_1, \tag{10.13}$$

where $Q^i, i = 0, 1$, are symmetric $n \times n$ matrices and $c^i \in \mathbb{R}^n,\ d_1 \in \mathbb{R}$. Consider the quadratic optimization problems:

$$(QPA) \qquad \min\{f(x)|\ x \in H,\ g(x) \leq 0\},$$

$$(QPB) \qquad \min\{f(x)|\ x \in H,\ g(x) = 0\}.$$

Define $D = \mathbb{R}_+,\ G = \{x \in \mathbb{R}^n|\ g(x) \leq 0\}$ for (QPA), and $D = \mathbb{R},\ G = \{x \in \mathbb{R}^n|\ g(x) = 0\}$ for (QPB) and let $L(x, y) = f(x) + yg(x)$, where $x \in \mathbb{R}^n, y \in \mathbb{R}$. Clearly

$$\sup_{y\in D} L(x, y) = \begin{cases} f(x) & \text{if } x \in G \\ +\infty & \text{otherwise} \end{cases}$$

so the optimal value of problem (QPA) or (QPB) is

$$v(QPA) = \inf_{x\in H} \sup_{y\in\mathbb{R}_+} L(x, y), Q \qquad v(QPB) = \inf_{x\in H} \sup_{y\in\mathbb{R}} L(x, y).$$

Strong duality is said to hold for (QPA) or (QPB) if

$$v(QPA) = \sup_{y\in\mathbb{R}_+} \inf_{x\in H} L(x, y) \qquad (\text{case } D = \mathbb{R}_+) \tag{10.14}$$

$$v(QPB) = \sup_{y\in\mathbb{R}} \inf_{x\in H} L(x, y) \qquad (\text{case } D = \mathbb{R}). \tag{10.15}$$

The problem on the right-hand side of (10.14) or (10.15) is called the *dual problem* of (QPA) or (QPB), respectively. By Proposition 10.2, if $\inf_{x\in H} L(x, y) > -\infty$ the function $L(., y)$ is convex on H. Therefore, with the usual convention

$\inf \emptyset = -\infty$, *the supremum in the dual problem can be restricted to those* $y \in D$ *for which the function* $L(., y)$ *is convex on* H. As will be seen shortly (Remark 10.4 below), in view of this fact many properties are obvious which otherwise have been proved sometimes by unnecessarily involved arguments. Actually, it is this hidden convexity that is responsible for the striking analogy between certain properties of quadratic inequalities and corresponding ones of convex inequalities.

To investigate conditions guaranteeing strong duality we introduce the following assumptions:

(S) There exists $x^* \in H$ satisfying $g(x^*) < 0$.
(T) There exists $y^* \in D$ such that $L(x, y^*) = f(x) + y^* g(x)$ is strictly convex on H.

Theorem 10.3 (Duality Theorem)

(i) *If assumption (S) is satisfied then (10.14) holds and if, furthermore,* $v(QPA) > -\infty$ *(which is the case if assumption (T), too, is satisfied with* $D = \mathbb{R}_+$*), then there exists* $\overline{y} \in \mathbb{R}_+$ *such that* $L(x, \overline{y})$ *is convex on* H *and*

$$v(QPA) = \min_{x \in H} L(x, \overline{y}) = \max_{y \in \mathbb{R}_+} \inf_{x \in H} L(x, y). \tag{10.16}$$

A point $\overline{x} \in H$ *is then an optimal solution of* (QPA) *if and only if it satisfies*

$$\langle \nabla f(\overline{x}) + \overline{y} \nabla g(\overline{x}),\ u \rangle = 0\ \forall u \in E, \quad \overline{y} g(\overline{x}) = 0, \quad g(\overline{x}) \leq 0, \tag{10.17}$$

where E *is the subspace parallel to the affine manifold* H.

(ii) *If* H *is a subspace and* $v(QPB) > -\infty$ *while*

- *either* $g(x)$ *is homogeneous and takes both positive and negative values on* H,
- *or* $g(x)$ *and* $f(x)$ *are both homogeneous,*

then (10.15) holds and there exists $\overline{y} \in \mathbb{R}$ *such that* $L(x, \overline{y})$ *is convex on* H *and*

$$v(QPB) = \min_{x \in H} L(x, \overline{y}) = \max_{y \in \mathbb{R}} \inf_{x \in H} L(x, y). \tag{10.18}$$

A point $\overline{x} \in H$ *is then an optimal solution of* (QPB) *if and only if it satisfies*

$$\langle \nabla f(\overline{x}) + \overline{y} \nabla g(\overline{x}),\ u \rangle = 0\ \forall u \in H, \quad g(\overline{x}) = 0. \tag{10.19}$$

(iii) *If assumption (T) is satisfied then (10.15) holds, where the supremum can be restricted to those* $y \in D$ *for which* $L(., y)$ *is convex on* H. *If, furthermore, the dual problem has an optimal solution* $\overline{y} \in D$ *then* $\overline{x} \in H$ *is an optimal solution of the primal problem if and only if*

$$\langle \nabla f(\overline{x}) + \overline{y} \nabla g(\overline{x}),\ u \rangle = 0\ \forall u \in E, \tag{10.20}$$

where E *is the subspace parallel to* H.

Proof (i) The existence of $\overline{y}$ follows from Corollary 10.3, (i). Then $\overline{x}$ is an optimal solution of (QPA) if and only if

$$\max_{y \in \mathbb{R}_+} L(\overline{x}, y) \leq L(\overline{x}, \overline{y}) \leq \min_{x \in H} L(x, \overline{y}),$$

hence, if and only if conditions (10.17) hold by noting that $L(., \overline{y})$ is convex on H.

(ii) When $g(x)$ is homogeneous and takes both positive and negative values on the subspace H the existence of $\overline{y}$ follows from Corollary 10.3, (ii). When both $f(x), g(x)$ are homogeneous, if $g(x) \geq 0\ \forall x \in H$ then $\inf\{f(x) |\ g(x) \leq 0, x \in H\} = v(QPB)$, so by Corollary 10.3, (iii), $v(QPB) > -\infty$ implies the existence of $(y_1, y_2) \in \mathbb{R}_+$, such that $y_1 f(x) + y_2 g(x) > 0\ \forall x \in H \setminus \{0\}$. If $y_1 = 0$ then $y_2 g(x) > 0\ \forall x \in H \setminus \{0\}$, hence $g(x) \geq 0\ \forall x \in H$, hence $g(x) = 0\ \forall x \in H$, a contradiction. Therefore, $y_1 > 0$ and (10.18) holds with $\overline{y} = y_2/y_1 \in \mathbb{R}_+$. Analogously, if $g(x) \leq 0\ \forall x \in H$, then $\inf\{f(x) |\ -g(x) \leq 0, x \in H\} = v(QPB)$, so (10.18) holds for some $\overline{y} \in \mathbb{R}_-$. Since $L(x, \overline{y})$ is convex on H, conditions (10.19) are necessary and sufficient for $\overline{x} \in H$ to be an optimal solution.

(iii) If assumption (T) is satisfied the conclusion follows from Theorem 10.1 where $C = H$ and $D = \mathbb{R}$ for (QPB). If, in addition to (T), the dual problem has an optimal solution $\overline{y} \in D$ then $\overline{x} \in H$ solves the primal problem if and only if it solves the convex optimization problem $\min_{x \in H} L(x, \overline{y})$, and hence, if and only if it satisfies (10.20). □

Remark 10.3 In the usual case $H = \mathbb{R}^n$ it is well known that by setting $Q(y) = Q^0 + yQ^1$, $c(y) = c^0 + yc^1$, $d(y) = yd_1$ the dual problem of (QPA) can be written as the *SDP*

$$\max \left\{ t |\ \begin{bmatrix} Q(y) & c(y) \\ c(y)^T & 2(d(y) - t) \end{bmatrix} \succeq 0,\ y \in D \right\}. \tag{10.21}$$

In the general case $H \neq \mathbb{R}^n$, as noticed above the supremum in (10.14) can be restricted to the set of those $y \in \mathbb{R}_+$ for which $L(x, y)$ is a convex function on H, i.e., $\inf_{x \in H} L(x, y)$ is a convex problem. Since the problem (10.14) amounts to maximizing the concave function $\varphi(y) = \inf_{x \in H} L(x, y)$ on $\mathbb{R}_+$, and the value of $\varphi(y)$ at each y is obtained by solving a convex problem, it is easy to solve the dual problem by convex programming methods.

Corollary 10.4 *In (QPA) or (QPB) if $f(x)$ or $g(x)$ is strictly convex on $\mathbb{R}^n$ then strong duality holds.*

Proof If $f(x)$ or $g(x)$ is strictly convex on $\mathbb{R}^n$ there exists $y^* \in \mathbb{R}_+$ such that $f(x) + y^* g(x)$ is strictly convex on $\mathbb{R}^n$, and hence, by Proposition 10.2, coercive on H. So the function $L(x, y) = f(x) + yg(x)$ satisfies all conditions of Theorem 10.1 with $C = H, D = \mathbb{R}_+$ (for (QPA)) or $D = \mathbb{R}$ (for (QPB)). By this theorem, strong duality follows. □

As a consequence, quadratic minimization over an ellipsoid, i.e., the problem

$$\min\{f(x) \mid \langle Qx, x\rangle \leq r^2,\ x \in \mathbb{R}^n\} \tag{10.22}$$

with $Q \succ 0$ (positive definite) can be solved efficiently by solving a *SDP* which is its dual. This is in contrast with linearly constrained quadratic minimization which has been long known to be NP-hard (Sahni 1974).

Remark 10.4 As a simple example illustrating the relevance of condition (T), consider the problem $\min\{x_1^4+ax_1^2+bx_1\}$. Upon the substitution $x_1^2 \to x_2$, it becomes a quadratic minimization problem of two variables x_1, x_2:

$$\min\{x_2^2 + ax_2 + bx_1 \mid x_1^2 - x_2 = 0\}. \tag{10.23}$$

Since $(x_2^2 + ax_2 + bx_1) + (x_1^2 - x_2) = x_1(x_1 + b) + x_2(x_2 + a - 1) \to +\infty$ as $\|(x_1, x_2)\| \to +\infty$ condition (T) is satisfied. Hence, by Theorem 10.3, (iii), (10.23) is equivalent to a convex problem—a result proved by an involved argument in Shor and Stetsenko (1989). Incidentally, note that the convex hull of the nonconvex set

$$\{(1, x_1, x_1^2, \ldots, x_1^n)^T \mid x_1 \in [a, b]\} \subset \mathbb{R}^{n+1}$$

is exactly described by *SDP* constraints (Hoang et al. 2008), so for any n-order univariate polynomial $P_n(x_1)$ the optimization problem $\min_{x_1 \in [a,b]} P_n(x_1)$ can be fully characterized by means of *SDP*s.

Corollary 10.5 *In* (*QPA*) *where* $H = \mathbb{R}^n$ *assume condition (S). Then* $\overline{x} \in \mathbb{R}^n$ *solves* (*QPA*) *if and only if there exists* $\overline{y} \in \mathbb{R}_+$ *satisfying*

$$Q^0 + \overline{y}Q^1 \succeq 0 \tag{10.24}$$

$$(Q^0 + \overline{y}Q^1)\overline{x} + c^0 + \overline{y}c^1 = 0 \tag{10.25}$$

$$\langle Q^1\overline{x} + 2c^1, \overline{x}\rangle + 2d_1 \leq 0. \tag{10.26}$$

When Q^0 *is not positive definite, the last condition is replaced by*

$$\langle Q^1\overline{x} + 2c^1, \overline{x}\rangle + 2d_1 = 0 \tag{10.27}$$

Proof By Theorem 10.3, (i), $\overline{x}$ is an optimal solution of (*QPA*) if and only if there exists $\overline{y} \in \mathbb{R}_+$ such that $L(x, \overline{y}) = f(x) + \overline{y}g(x)$ is convex and $\overline{x}$ is a minimizer of this convex function. That is, if and only if there exists $\overline{y} \in \mathbb{R}_+$ satisfying (10.24) and (10.17), i.e., (10.24)–(10.26).

If the inequality (10.26) holds strictly, i.e.,

$$\langle Q^1\overline{x} + 2c^1, \overline{x}\rangle + 2d_1 < 0$$

then $g(\overline{x}) < 0$, hence $g(x) \leq 0$ for all x in some neighborhood of $\overline{x}$. Then $\overline{x}$ is a local minimizer, hence a global minimizer, of $f(x)$ over $\mathbb{R}^n$, which happens only if Q^0 is positive definite. □

Corollary 10.6 *With f, g, H defined as in (10.13) above, assume that $g(x)$ is homogeneous and takes both positive and negative values on H. Then*

$$x \in H, \ g(x) = 0 \quad \Rightarrow \quad f(x) \geq 0 \tag{10.28}$$

if and only if there exists $\overline{y} \in \mathbb{R}$ such that $f(x) + \overline{y}g(x) \geq 0 \ \forall x \in H$.

Proof This is a proposition often referred to as the *S-Lemma with equality constraint.* By Corollary 10.3, (ii), it extends to the case when $g(x) = \langle x, Q^1 x\rangle + \langle c^1, x\rangle + d_1$ with Q^1 nonsingular and, moreover, $g(x)$ takes both positive and negative values on H. □

Note that when $H = \mathbb{R}^n$ and $f(x), g(x)$ are homogeneous with associated matrices Q^0 and Q^1, $g(x)$ takes on both positive and negative values if Q^1 is indefinite, while $f(x) + \overline{y}g(x) \geq 0 \ \forall x \in \mathbb{R}^n$ if and only if $Q^0 + \overline{y}Q^1 \succeq 0$. So in this special case Corollary 10.6 gives the following *nonstrict Finsler Theorem* (Finsler 1937):

If Q^1 is indefinite then $\langle Q^1 x, x\rangle = 0$ implies $\langle Q^0 x, x\rangle \geq 0$ if and only if $Q^0 + yQ^1 \succeq 0$ for some $y \in \mathbb{R}$.

Also note:

Corollary 10.7 (Strict Finsler's Theorem (Finsler 1937)) *With f, g homogeneous and $H = \mathbb{R}^n$ we have*

$$x \neq 0, \ g(x) = 0 \quad \Rightarrow \quad f(x) > 0 \tag{10.29}$$

if and only if there exists $\overline{y} \in \mathbb{R}$ such that $f(x) + \overline{y}g(x) > 0 \ \forall x \neq 0$.

Proof This follows directly from Corollary 10.3, (iii) (*S-Lemma with nonstrict inequalities*). Since $f(x)$ and $g(x)$ are both homogeneous, there exists $(y_1, y_2) \in \mathbb{R}^2_+$ such that $y_1 f(x) + y_2 g(x) > 0 \ \forall x \in H \setminus \{0\}$. From (10.29) we must have $y_1 > 0$, so $\overline{y} = y_2/y_1$. □

Corollary 10.8 *Let f, g, H be defined as in (10.13) above. Then just one of the following alternatives holds:*

(i) *There exists $x \in H$ satisfying $f(x) < 0, \ g(x) < 0$;*
(ii) *There exists $y = (y_1, y_2) \in \mathbb{R}^2_+ \setminus \{(0, 0)\}$ satisfying*

$$y_1 f(x) + y_2 g(x) \geq 0 \ \forall x \in H.$$

Proof Clearly (i) and (ii) cannot hold simultaneously. If (i) does not hold, two cases are possible: 1) $f(x) \geq 0$ for every $x \in H$ such that $g(x) < 0$; 2) $g(x) \geq 0$ for every $x \in H$ such that $f(x) < 0$. In case (1), if 0 is a local minimum of $g(x)$ on H, then $g(x) \geq 0 \ \forall x \in H$, so $y_1 f(x) + y_2 g(x) \geq 0 \ \forall x \in H$ with $y_1 = 0, y_2 = 1$; if, on the contrary, 0 is not a local minimum of $g(x)$ on H, then every $x \in H$ satisfying $g(x) = 0$ is the limit of a sequence $x^k \in H$ satisfying $g(x^k) < 0, k = 1, 2, \ldots,$ so $\inf\{f(x)|\ g(x) \leq 0,\ x \in H\} = \inf\{f(x)|\ g(x) < 0,\ x \in H\} \geq 0$, hence by Theorem 10.3, (i), there exists $y_2 \in \mathbb{R}_+$ such that $f(x) + y_2 g(x) \geq 0\ \forall x \in H$. The proof in case (2) is similar, with the roles of $f(x), g(x)$ interchanged. □

Corollary 10.9 *Let f, g, H be defined as above, let A, B be two closed subsets of an affine manifold H such that $A \cup B = H$. Assume condition (S). If*

$$f(x) \geq 0 \quad \forall x \in A, \quad \textit{while} \quad g(x) \geq 0 \quad \forall x \in B \tag{10.30}$$

then there exists $\overline{y} \in \mathbb{R}_+$ such that $f(x) + \overline{y}g(x) \geq 0\ \forall x \in H$.

Proof Condition (S) means that $H \setminus B \neq \emptyset$, hence $A \neq \emptyset$. Clearly $\{x \in H|\ g(x) < 0\} \subset A$. If $x^0 \in H$ satisfies $g(x^0) = 0$, then for every k there exists $x^k \in H$ satisfying $\|x^k - x^0\| \leq 1/k,\ g(x^k) < 0$ for otherwise x^0 would be a local minimum of $g(x)$ on H, hence a global minimum on H, i.e., $g(x) \geq g(x^0) = 0\ \forall x \in H$, contradicting (S). Therefore, $\{x \in H|\ g(x) \leq 0\} = \mathrm{cl}\{x \in H|\ g(x) < 0\} \subset A$. Assumption (10.30) then implies $f(x) \geq 0$ for every $x \in H$ such that $g(x) \leq 0$, hence $\inf\{f(x)|\ g(x) \leq 0\} \geq 0$, and the existence of $\overline{y}$ follows from Theorem 10.3, (i). □

The above results, and especially the next one, are very similar to corresponding properties of convex inequalities.

Corollary 10.10 (i) *With $f(x), g(x), H$ as in (10.13) the set $C = \{t \in \mathbb{R}^2|\ \exists x \in H$ s.t. $f(x) \leq t_1,\ g(x) \leq t_2\}$ is convex.*
(ii) *With $f(x), g(x)$ homogeneous and H a subspace, the set $G = \{t \in \mathbb{R}^2|\ \exists x \in H$ s.t. $f(x) = t_1,\ g(x) = t_2\}$ is convex.*

Proof (i) [1] We first show that for any two points $a, b \in H$ the set $C_{ab} := \{t \in \mathbb{R}^2|\ \exists x \in [a, b] \text{ s.t. } f(x) \leq t_1, g(x) \leq t_2\}$ is closed and convex. Let $t^k \in C_{ab},\ k = 1, 2, \ldots,$ be a sequence converging to $\overline{t}$. For each k let $x^k \in [a, b]$ satisfy $f(x^k) \leq t_1^k,\ g(x^k) \leq t_2^k$. Then $x^k \to \overline{x} \in [a, b]$ (up to a subsequence) and by continuity of f, g we have $f(\overline{x}) \leq \overline{t}_1,\ g(\overline{x}) \leq \overline{t}_2$, so $\overline{t} \in C_{ab}$. Therefore, C_{ab} is a closed set. Further, consider any point $t = (t_1, t_2) \in \mathbb{R}^2 \setminus C_{ab}$. Since C_{ab} is a closed set there exists a point $t' \notin C_{ab}$ such that $t'_1 > t_1,\ t'_2 > t_2$. Then

$$\nexists x \in [a, b] \quad f(x) < t'_1,\ g(x) < t'_2.$$

By Corollary 10.8 there exists $(y_1, y_2) \in \mathbb{R}^2_+ \setminus \{0, 0\}$ satisfying $y_1(f(x) - t'_1) + y_2(g(x) - t'_2) \geq 0\ \forall x \in [a, b]$. Setting $L_t := \{s \in \mathbb{R}^2|\ y_1 s_1 + y_2 s_2 \geq y_1 t'_1 + y_2 t'_2\}$ it is easily checked that $C_{ab} \subset L_t$ while $t \notin L_t$. So for every $t \notin C_{ab}$ there exists a halfplane L_t separating t from C_{ab}. Consequently, $C = \cap_{t \notin C} L_t$, proving the convexity of C_{ab}.

[1] This proof corrects a flaw in the original proof given in Tuy and Tuan (2013).

Now if t, s are any two points of C then there exist $a, b \in H$ such that

$$f(a) \leq t_1,\ g(a) \leq t_2, \quad f(b) \leq s_1,\ g(b) \leq s_2.$$

This means $t, s \in C_{ab}$, and hence $[t, s] \subset C_{ab} \subset C$. Thus $t, s \in C \Rightarrow [t, s] \subset C$, proving the convexity of C.

(ii) Turning to the set G with $f(x), g(x)$ homogeneous and H a subspace, observe that if there exists $x \in H$ satisfying $a_1 g(x) - a_2 f(x) = 0,\ a_1 f(x) > 0$, then $\frac{a_1}{f(x)} = \frac{a_1 f(x)}{[f(x)]^2} > 0$, so setting $\frac{a_1}{f(x)} = r^2$ yields $a_1 = r^2 f(x) = f(rx)$ and $a_2 = \frac{a_1}{f(x)} g(x) = r^2 g(x) = g(rx)$, hence $a \in G$ because $rx \in H$. Therefore, if $a \notin G$ then $\min\{-a_1 f(x) |\ a_1 g(x) - a_2 f(x) = 0,\ x \in H\} = 0$. By Corollary 10.3, (ii), there exists $\bar{y} \in \mathbb{R}$ satisfying $-a_1 f(x) + \bar{y}(a_1 g(x) - a_2 f(x)) \geq 0\ \forall x \in H$. Then $G \subset L_a := \{t \in \mathbb{R}^2 |\ -a_1 t_1 + \bar{y}(a_1 t_2 - a_2 t_1) \geq 0\}$, while $a \notin L_a$ (because $-a_1^2 < 0$). Thus, every $a \notin G$ can be separated from G by a halfspace, proving the convexity of the set G. □

Assertion (i) of Corollary 10.10 is analogous to a well-known proposition for convex functions. It has been derived from Corollary 10.3, (i), but the converse is also true. In fact, suppose for given $f(x), g(x), H$ the set $C := \{t \in \mathbb{R}^2 |\ \exists x \in H \text{ s.t. } f(x) \leq t_1,\ g(x) \leq t_2\}$ is convex. Clearly if the system $x \in H,\ f(x) \leq 0,\ g(x) \leq 0$ is inconsistent then $(0, 0) \notin C$, hence, by the separation theorem, there exists $(u, y) \in \mathbb{R}^2_+$, such that $ut_1 + yt_2 \geq 0\ \forall (t_1, t_2) \in C$, i.e., $uf(x) + yg(x) \geq 0\ \forall x \in H$. If, furthermore, there exists $x^* \in H$ such that $g(x^*) < 0$ then $(f(x^*), g(x^*)) \in C$, so $uf(x^*) + yg(x^*) \geq 0$. We cannot have $u = 0$ for this would imply $yg(x^*) \geq 0$, hence $y = 0$, conflicting with $(u, y) \neq 0$. Therefore $u > 0$ and setting $\bar{y} = y/u$ yields $f(x) + \bar{y}g(x) \geq 0\ \forall x \in H$. So the fact that C is convex is equivalent to Corollary 10.3, (i).

Assertion (ii) of Corollary 10.10 (which still holds with H replaced by a regular cone) includes a result of Dines (1941) which was used by Yakubovich (1977) in his original proof of the S-Lemma. Since this result has been derived from Corollary 10.3, (ii), i.e., the S-Lemma with equality constraints, Dines' theorem is actually equivalent to the latter.

It has been long known that the general linearly constrained quadratic programming problem is NP-hard (Sahni 1974). Recently, two special variants of this problem have also been proved to be NP-hard:

1. The linearly constrained quadratic program with just one negative eigenvalue (Pardalos and Vavasis 1991):

$$\min\{-x_1^2 + \langle c, x\rangle |\ Ax \leq b, x \geq 0\}.$$

2. The linear multiplicative program (Matsui 1996):

$$\min\{(\langle c, x\rangle + \gamma)(\langle d, x\rangle + \delta) |\ Ax \leq b, x \geq 0\}$$

Nevertheless, quite practical algorithms exist which can solve each of the two mentioned problems in a time usually no longer than the time needed for solving a few linear programs of the same size (Konno et al. 1997). In this section we are going to discuss another special quadratic minimization problem which can also be solved by efficient algorithms (Karmarkar 1990; Ye 1992). This problem has become widely known due to several interesting features. Referring to its role in nonlinear programming methods it is sometimes called the *Trust Region Subproblem*. It consists of minimizing a nonconvex quadratic function over a Euclidean ball, i.e.,

$$(TRS) \qquad \min\left\{\frac{1}{2}\langle x, Qx\rangle + \langle c, x\rangle \middle| \quad \langle x, x\rangle \le r^2\right\}, \tag{10.31}$$

where Q is an $n \times n$ symmetric matrix and $c \in \mathbb{R}^n, r \in \mathbb{R}$. A seemingly more general formulation, in which the constraint set is an ellipsoid $\langle x, Hx\rangle \le r^2$ (with H being a positive definite matrix), can be reduced to the above one by means of a nonsingular linear transformation.

Let $u^1, \ldots, u^n$ be a full set of orthonormal eigenvectors of Q, and let $\lambda_1 \ldots, \lambda_n$ be the corresponding eigenvalues. Without loss of generality one can assume that $\lambda_1 = \ldots = \lambda_k < \lambda_{k+1} \le \ldots \le \lambda_n$, and $\lambda_1 < 0$. (The latter inequality simply means that the problem is nonconvex.) By Proposition 4.18, the global minimum is achieved on the sphere $\langle x, x\rangle = r^2$, so the problem is equivalent to

$$\min\left\{\frac{1}{2}\langle x, Qx\rangle + \langle c, x\rangle \middle| \quad \langle x, x\rangle = r^2\right\}.$$

Proposition 10.4 *A necessary and sufficient condition for x^* to be a global optimal solution of* (TRS) *is that there exist $\mu^* \ge 0$ satisfying*

$$Q + \mu^* yI \text{ is positive semidefinite} \tag{10.32}$$

$$Qx^* + \mu^* x^* + c = 0 \tag{10.33}$$

$$\|x^*\| = r. \tag{10.34}$$

Proof Clearly (TRS) is a problem (QPA) where $H = \mathbb{R}^n, Q_0 = Q, Q_1 = I$ and condition (S) is satisfied, so the result follows from Corollary 10.5. □

Remark 10.5 It is interesting to note that when this result first appeared it came as a surprise. In fact, at the time (TRS) was believed to be a nonconvex problem, so the fact that a global optimal solution can be fully characterized by conditions like (10.32)–(10.34) seemed highly unusual. However, as we saw by Theorem 10.3, x^* is a global optimal solution of (TRS) as a (QPA) if and only if there exists $\mu^* \ge 0$ such that the function $\frac{1}{2}\langle x, Qx\rangle + \mu^*\langle c, x\rangle$ is convex (which is expressed by (10.32)) and x^* is a minimizer of this convex function (which is expressed by the equalities (10.33)–(10.34)—standard optimality condition in convex programming).

Also note that by a result of Martinez (1994) (*TRS*) has at most one local nonglobal optimal solution. No wonder that good local methods such as the *DCA* method (Tao 1997; Tao and Hoai-An (1998) and the references therein) can often give actually a global optimal solution.

In view of Proposition 10.4, solving (*TRS*) reduces to finding $\mu^* \geq 0$ satisfying (10.32)–(10.34). Condition (10.33) implies that $\lambda_1 \leq \mu^* < +\infty$. For $-\lambda_1 < \mu$ since $Q + \mu I$ is positive definite, the equation $Qx + \mu x + c = 0$ has a unique solution given by

$$x(\mu) = -(Q + \mu I)^{-1} c.$$

Proposition 10.5 *The function* $\|x(\mu)\|$ *is monotonically strictly increasing in the interval* $(-\lambda_1, +\infty)$. *Furthermore,*

$$\lim_{\mu \to \infty} \|x(\mu)\| = 0,$$

$$\lim_{\mu \to -\lambda_1} \|x(\mu)\| = \begin{cases} \|\bar{x}\| & \text{if } \langle c, u^i \rangle = 0 \ \forall i = 1, \ldots, k \\ +\infty & \text{otherwise,} \end{cases}$$

where

$$\bar{x} = -\sum_{i=k+1}^{n} \frac{\langle c, u^i \rangle u^i}{\lambda_i - \lambda_1}.$$

Proof From $-\mu x(\mu) = Qx(\mu) + c$, we have for every $i = 1, \ldots, n$:

$$-\mu \langle x(\mu), u^i \rangle = \langle x, Qu^i \rangle + \langle c, u^i \rangle = \lambda_i \langle x, u^i \rangle + \langle c, u^i \rangle,$$

hence, for $\mu > \lambda_1$:

$$\langle x(\mu), u^i \rangle = \frac{\langle c, u^i \rangle}{\lambda_i - \mu}.$$

Since $\{u^1, \ldots, u^n\}$ is an orthonormal base of $\mathbb{R}^n$, we then derive

$$x(\mu) = \sum_{i=1}^{n} \frac{\langle c, u^i \rangle u^i}{\lambda_i - \mu}, \tag{10.35}$$

and consequently

$$\|x(\mu)\|^2 = \sum_{i=1}^{n} \frac{(\langle c, u^i \rangle)^2}{(\lambda_i - \mu)^2}. \tag{10.36}$$

This expression shows that $\|x(\mu)\|$ is a strictly monotonically increasing function of μ in the interval $(-\lambda_1, +\infty)$. Furthermore, $\lim_{\mu\to-\infty} \|x(\mu)\| = 0$. If $\langle c, u^i\rangle = 0\ \forall i = 1,\dots,k$, then from (10.35) $x(\mu) = \sum_{i=k+1}^{n}\langle c, u^i\rangle u^i/(\lambda_i - \mu)$, hence $x(\mu) \to \bar{x}$ as $\mu \to -\lambda_1$. Otherwise, since $\langle c, u^i\rangle \neq 0$ for at least some $i = 1,\dots,k$ and $\lambda_1 = \dots, \lambda_k$ we have $\|x(\mu)\| \to +\infty$ as $\mu \to -\lambda_1$. □

From the above results we derive the following method for solving (*TRS*):

- If $\langle c, u^i\rangle \neq 0$ for at least one $i = 1,\dots,k$ or if $\|\bar{x}\| \geq r$ then there exists a unique value $\mu^* \in [-\lambda_1, +\infty)$ satisfying $\|x(\mu^*)\| = r$. This value μ^* can be determined by binary search. The global optimal solution of (*TRS*) is then $x^* = -(Q + \mu^* I)^{-1}c$ and is unique.
- If $\langle c, u^i\rangle = 0\ \forall i = 1,\dots,k$ and $\|\bar{x}\| < r$ then $\mu^* = -\lambda_1$ and an optimal solution is

$$x^* = \sum_{i=1}^{k}\alpha_i u^i + \bar{x},$$

where $\alpha_1,\dots,\alpha_k$ are any real numbers such that $\sum_{i=1}^k \alpha_i^2 = r^2 - \|\bar{x}\|^2$. Indeed, it is easily verified that $\|x^*\|^2 = r^2$ and also $Qx^* + c = \sum_{i=1}^k \alpha_i Qu^i + Q\bar{x} + c = -\sum_{i=1}^k \alpha_i\lambda_1 u^i - \lambda_1\bar{x} = -\lambda_1(\sum_{i=1}^k \alpha_i u^i + \bar{x}) = -\lambda_1 x^*$. Since there are many different choices for $\alpha_1,\dots,\alpha_k$ satisfying the required condition, the optimal solution in this case is not unique.

10.4 Quadratic Problems with Linear Constraints

10.4.1 Concave Quadratic Programming

The Concave Quadratic Programming Problem

$$(CQP) \qquad \min\{f(x) \mid Ax \leq b, x \geq 0\},$$

where $f(x)$ is a concave quadratic function, is a special case of (*BCP*)—the Basic Concave Programming Problem (Chap. 5). Algorithms for this problem have been developed in Konno (1976b), Rosen (1983), Rosen and Pardalos (1986), Kalantari and Rosen (1987), Phillips and Rosen (1988), and Thakur (1991), ... Most often the quadratic function $f(x)$ is considered in the separable form, which is always possible via an affine transformation of the variables (cf Sect. 8.1). Also, particular attention is paid to the case when the quadratic function is of low rank, i.e., involves a restricted number of nonconvex variables and, consequently, is amenable to decomposition techniques (Chap. 7).

Of course, the algorithms presented in Chaps. 5–6–7 for (*BCP*) can be adapted to handle (*CQP*). In doing this one should note two points where substantial simplifications and improvements are possible due to the specific properties of quadratic functions.

The first point concerns the computation of the γ-extension (Remark 5.2). Since $f(x)$ is concave quadratic, for any point $x^0 \in \mathbb{R}^n$ and any direction $u \in \mathbb{R}^n$, the function $\theta \mapsto f(x^0 + \theta u)$ is concave quadratic in θ. Therefore, for any $x^0, x \in \mathbb{R}^n$ and any real number γ such that $f(x^0) > \gamma$, the γ-extension of x with respect to x^0, i.e., the point $x^0 + \theta(x - x^0)$ corresponding to the value

$$\theta = \sup\{\lambda | f(x^0 + \lambda(x - x^0)) \geq \gamma\},$$

can be found by solving the quadratic equation in $\theta : f(x^0 + \theta(x - x^0)) = \gamma$ (note that the level set $f(x) \geq \gamma$ is bounded).

The second point is the construction of the cuts to reduce the search domain. It turns out that when the objective function $f(x)$ is quadratic, the usual concavity cuts can be substantially strengthened by the following method (Konno 1976a, see also Balas and Burdet 1973). Assume that $D \subset \mathbb{R}^n_+$, $x^0 = 0$ is a vertex of D and write the objective function as

$$f(x) = \langle x, Qx \rangle + 2\langle c, x \rangle,$$

where Q is a negative semidefinite matrix. Consider a γ-valid cut for (f, D) constructed at x^0:

$$\sum_{i=1}^{n} \frac{x_i}{t_i} \geq 1. \tag{10.37}$$

So if we define $M(t) = \{x \in D | \sum_{i=1}^n x_i/t_i) \geq 1\}$, then $\{x \in D | f(x) < \gamma\} \subset M(t)$. Note that by construction $f(t_i e^i) \geq \gamma$, $i = 1, \ldots, n$, where as usual e^i is the ith unit vector. We now construct a new concave function $\varphi_t(x)$ such that a γ-valid concavity cut for (φ_t, D) will also be γ-valid for (f, D) but in general deeper than the above cut. To this end define the bilinear function

$$F(x, y) = \langle x, Qy \rangle + \langle c, x \rangle + \langle c, y \rangle.$$

Observe that

$$f(x) = F(x, x), \quad F(x, y) \geq \min\{f(x), f(y)\}. \tag{10.38}$$

Indeed, the first equality is evident. On the other hand,

$$\begin{aligned} F(x, y) - F(x, x) &= \langle x, Q(y - x) \rangle + \langle c, y - x \rangle \\ F(x, y) - F(y, y) &= \langle y, Q(x - y) \rangle + \langle c, x - y \rangle, \end{aligned}$$

hence $[F(x,y)-F(x,x)]+[F(x,y)-F(y,y)] = -\langle x-y, Q(x-y)\rangle \geq 0$ because Q is negative semidefinite. Thus at least one of the differences $F(x,y)-F(x,x)$, $F(x,y)-F(y,y)$ must be nonnegative, whence the inequality in (10.38). The argument also shows that if Q is negative definite then the inequality in (10.38) is strict for $x \neq y$ because then $\langle x-y, Q(x-y)\rangle < 0$.

Lemma 10.1 *The function*

$$\varphi_t(x) = \min\{F(x,y)|\ y \in M(t)\}$$

is concave and satisfies

$$\varphi_t(x) \leq f(x) \quad \forall x \in M(t) \tag{10.39}$$

$$\min\{\varphi_t(x)|\ x \in M(t)\} = \min\{f(x)|\ x \in M(t)\}. \tag{10.40}$$

Proof The concavity of $\varphi_t(x)$ is obvious from the fact that it is the lower envelope of the family of affine functions $x \mapsto F(x,y), y \in M(t)$. Since $\varphi_t(x) \leq F(x,x) = f(x)\ \forall x \in M(t)$ (10.39) holds. Furthermore, $\varphi_t(x) \geq \min\{F(x,y)|\ y \in M(t), y = x\} = \min\{F(x,x)|\ x \in M(t)\}$, hence

$$\varphi_t(x) \geq \min\{f(x)|\ x \in M(t)\},$$

and (10.40) follows from (10.39). □

Proposition 10.6 *Let* $\varphi(t_ie^i) = F(t_ie^i, y^i)$, *i.e.,* $y^i \in \operatorname{argmin}\{F(t_ie^i, y)|\ y \in M(t)\}$. *If* $f(y^i) \geq \gamma,\ i = 1,\dots,n$ *then the vector* $t^* = (t_1^*,\dots,t_n^*)$ *such that*

$$t_i^* = \max\{\lambda|\ \varphi_t(\lambda e^i) \geq \gamma\},\ i = 1,\dots,n \tag{10.41}$$

defines a γ*-valid cut for* (f,D) *at* x^0*:*

$$\sum_{i=1}^{n} \frac{x_i}{t_i^*} \geq 1, \tag{10.42}$$

which is no weaker than (10.37) (i.e., such that $t_i^* \geq t_i,\ i = 1,\dots,n$*).*

Proof The cut (10.42) is γ-valid for (φ_t, D), hence for (f,D), too, in view of (10.39). Since $\varphi_t(t_ie^i) \geq \min\{f(t_ie^i), f(y^i)\}$ by (10.38), and since $f(t_ie^i) \geq \gamma$, it follows from the inequality $f(y^i) \geq \gamma$ that $\varphi_t(t_ie^i) \geq \gamma$, and hence $t_i^* \geq t_i$ by the definition of t_i^*. □

Note that if $\varphi_t(t_ie^i) > \min\{f(t_ie^i), f(y^i)\}$ (which occurs when Q is negative definite) then $t_i^* > t_i$ provided that $t_ie^i \notin D$ (for $y^i \in M(t)$ implies $y^i \in D$, hence $y^i \neq t_ie^i$). This means that the cut (10.42) in this case is strictly deeper than (10.37).

The computation of t_i^* in (10.41) reduces to solving a linear program. Indeed, if Q_i denotes the ith row of Q, then $F(\lambda e^i, y) = \lambda\langle Q_i + c, y\rangle + \lambda c_i$, so for $D = \{x|\ Ax \le b, x \ge 0\}$ we have $\varphi_t(\lambda e^i) = g_i(\lambda) + \lambda c_i$, where

$$g_i(\lambda) = \min\left\{\lambda\langle Q_i + \langle c, y\rangle|\ \sum_{i=1}^n \frac{y_i}{t_i} \ge 1, Ay \le b, y \ge 0\right\}.$$

By the duality theorem of linear programming,

$$g_i(\lambda) = \max\left\{-\langle b, u\rangle + \mu|\ -A^T u + \mu\left(\frac{1}{t_1}, \dots, \frac{1}{t_n}\right)^T \le (Q_i + c)^T, u \ge 0, \mu \ge 0\right\}.$$

Thus, $t_i^* = \max\{\lambda|\ g_i(\lambda) + \lambda c_i \ge \gamma\}$, hence

$$t_i^* = \max\left\{\lambda|\ -\langle b, u\rangle + \mu + \lambda c_i \ge \gamma, Q - A^T u + \mu\left(\frac{1}{t_1}, \dots, \frac{1}{t_n}\right)^T \le (Q_i + c)^T,\right.$$
$$\left. u \ge 0, \mu \ge 0\right\}. \tag{10.43}$$

To sum up, a cut (10.37) can be improved by the following procedure:

1. Compute $y^i \in \operatorname{argmin}\{F(t_i e^i, y)|\ y \in M(t)\},\ i = 1, \dots, n$.
2. If $f(y^i) < \gamma$ for some i, then construct a γ'-valid cut for (f, D),with the new incumbent $\gamma' := \min\{f(y^i)|\ i = 1, \dots, n\}$.
3. Otherwise, compute $t_i^*,\ i = 1, \dots, n$ by (10.43), and construct the cut (10.42).

   Clearly, if a further improvement is needed the above procedure can be repeated with $t \leftarrow t^*$.

10.4.2 Biconvex and Convex–Concave Programming

A problem that has earned a considerable interest due to its many applications is the *bilinear program* whose general form is

$$\min\{\langle c, x\rangle + \langle x, Ay\rangle + \langle d, y\rangle : x \in X, y \in Y\},$$

where X, Y are polyhedrons in $\mathbb{R}^p, \mathbb{R}^q$, respectively, and A is a $p \times q$ matrix. By setting $\varphi(x) = \langle c, x\rangle + \min\{\langle x, Ay\rangle + \langle d, y\rangle : y \in Y\}$, the problem becomes $\min\{\varphi(x) : x \in X\}$, which is a concave program, because $\varphi(x)$ is a concave function (pointwise minimum of a family of affine functions). A more difficult problem, which can also be shown to be more general, is the following *jointly constrained biconvex program*:

$$(JCBP) \qquad \min\{f(x) + \langle x, y\rangle + g(y) : (x, y) \in G\},$$

where $f(x), g(y)$ are convex functions on $\mathbb{R}^n$ and G is a compact convex subset of $\mathbb{R}^n \times \mathbb{R}^n$. Since $\langle x, y\rangle = (\|x+y\|^2 - \|x-y\|^2)/4$, the problem can be solved by rewriting it as the reverse convex program:

$$\min\{f(x) + g(y) + (\|x+y\|^2 - t)/4 : (x, y) \in G, \|x-y\|^2 \geq t\}.$$

An alternative branch and bound method was proposed in Al-Khayyal and Falk (1983), where branching is performed by rectangular partition of $\mathbb{R}^n \times \mathbb{R}^n$. To compute a lower bound for $F(x, y) := f(x) + \langle x, y\rangle + g(y)$ over any rectangle $M = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n : r \leq x \leq s; u \leq y \leq v\}$ observe the following relations for every $i = 1, \ldots, n$:

$$x_i y_i \geq u_i x_i + r_i y_i - r_i u_i \quad \forall x_i \geq r_i, y_i \geq u_i$$

$$x_i y_i \geq v_i x_i + s_i y_i - s_i v_i \quad \forall x_i \leq s_i, y_i \leq v_i$$

(these relations hold because $(x_i - r_i)(y_i - u_i) \geq 0$ and $(x_i - s_i)(y_i - v_i) \geq 0$, respectively). Therefore, an underestimator of xy over M is the convex function

$$\left.\begin{array}{l} \varphi(x, y) = \sum_{i=1}^n \varphi_i(x, y), \\ \varphi_i(x, y) = \max\{u_i x_i + r_i y_i - r_i u_i, v_i x_i + s_i y_i - s_i v_i\} \end{array}\right| \tag{10.44}$$

and a lower bound for $\min\{F(x, y) : (x, y) \in M\}$ is provided by the value

$$\beta(M) := \min\{f(x) + \varphi(x, y) + g(y) : (x, y) \in G \cap M\} \tag{10.45}$$

(the latter problem is a convex program and can be solved by standard methods). At iteration k of the branch and bound procedure, a rectangle M_k in the current partition is selected which has smallest $\beta(M_k)$. Let (x^k, y^k) be the optimal solution of the bounding subproblem (10.45) for M_k, i.e., the point $(x^k, y^k) \in G \cap M_k$, such that

$$\beta(M_k) = f(x^k) + \varphi(x^k, y^k) + g(y^k).$$

If $x^k y^k = \varphi(x^k, y^k)$ then the procedure terminates: (x^k, y^k) is a global optimal solution of (JCBP). Otherwise, select

$$i_k \in \arg \max_{i=1,\ldots,n} \{x_i^k y_i^k - \varphi_i(x^k, y^k)\}.$$

If $M_k = [r^k, s^k] \times [u^k, v^k]$, then subdivide $[r^k, s^k]$ via (x^k, i_k) and subdivide $[u^k, v^k]$ via (y^k, i_k) to obtain a partition of M_k into four smaller rectangles which will replace M_k in the next iteration.

By simple computation, one can easily verify that $\varphi(x, y) = xy$, hence $f(x) + \varphi(x, y) + g(y) = F(x, y)$ at every corner (x, y) of the rectangle M. Using this property and the basic rectangular subdivision theorem it can then be proved that the above branch and bound procedure converges to a global optimal solution of $(JCBP)$.

Another generalization of the bilinear programming problem is the *convex–concave programming* problem studied in Muu and Oettli (1991):

$$(CCP) \qquad \min\{F(x,y) : (x,y) \in G \cap (X \times Y)\},$$

where $G \subset \mathbb{R}^p \times \mathbb{R}^q$ is a closed convex set, X a rectangle in $\mathbb{R}^p$, Y a rectangle in $\mathbb{R}^q$, and $F(x,y)$ is convex in $x \in \mathbb{R}^p$ but concave in $y \in \mathbb{R}^q$. The method of Muu–Oettli starts from the observation that, for any rectangle $M \subset Y$, if $\beta(M)$ is the optimal value and (x^M, y^M, u^M) an optimal solution of the problem

$$\min\{F(x,y) : x \in X, y \in M, u \in M, (x,u) \in G\}, \tag{10.46}$$

then (x^M, u^M) is feasible to (CCP) and

$$\beta(M) = F(x^M, y^M) \leq \min\{F(x,y) : (x,y) \in G \cap (X \times M)\} \leq F(x^M, u^M).$$

In particular, $\beta(M)$ yields a lower bound for $F(x,y)$ on $G \cap (X \times M)$, which becomes the exact minimum if $y^M = u^M$. This suggests a branch and bound procedure in which branching is by rectangular subdivision of Y and $\beta(M)$ is used as a lower bound for every subrectangle M of Y. At iteration k of this procedure, a rectangle M_k in the current partition is selected which has smallest $\beta(M_k)$. Let $(x^k, y^k, u^k) = (x^{M_k}, y^{M_k}, u^{M_k})$. If $y^k = u^k$ then the procedure terminates: (x^k, y^k) solves (CCP). Otherwise, select

$$i_k \in \arg \max_{i=1,\dots,q} \{|y_i^k - u_i^k|\}$$

and subdivide M_k via $(\frac{y^k+u^k}{2}, i_k)$. In other words, subdivide M_k according to the adaptive rule via y^k, u^k (see Theorem 6.4, Chap. 6).

The convergence of the procedure follows from the fact that, by virtue of Theorem 5.4, a nested sequence $\{M_{k(s)}\}$ is generated such that $|y^{M_{k(s)}} - u^{M_{k(s)}}| \to 0$ as $s \to +\infty$.

This method is implementable if the bounding subproblems (10.46) can be solved efficiently, which is the case, e.g., for the d.c. problem

$$\min\{g(x) - h(y) : (x,y) \in G \cap (X \times Y)\}.$$

In that case, the subproblem (10.46):

$$\min\{g(x) - h(y) : x \in X, y \in M, u \in M, (x,u) \in G\}$$

is equivalent to

$$\min\{g(x) : x \in X, u \in M, (x,u) \in G\} - \max\{h(y) : y \in M\},$$

which reduces to solving a standard convex program and computing the maximum of the convex function $h(y)$ at 2^q corners of the rectangle M.

By bilinear programming, one usually means a problem of the form

$$(BLP) \qquad \min\{\langle x, Qy\rangle + \langle c, x\rangle + \langle d, y\rangle \mid x \in D,\ y \in E\},$$

where D, E are polyhedrons in $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$, respectively, Q is an $n_1 \times n_2$ matrix and $c \in \mathbb{R}^{n_1}, d \in \mathbb{R}^{n_2}$. This is a rather special (though perhaps the most popular) variant of bilinear problem, since in the general case a bilinear problem may involve bilinear constraints as well.

The fact that the constraints in (BLP) are linear and separated in x and y allow this problem to be converted into a linearly constrained concave minimization problem. In fact, setting

$$\varphi(x) = \min\{\langle x, Qy\rangle + \langle d, y\rangle \mid y \in E\}$$

we see that (BLP) is equivalent to

$$\min\{\varphi(x) + \langle c, x\rangle \mid x \in D\}.$$

Since for each fixed y the function $x \mapsto \langle x, Qy\rangle + \langle d, y\rangle$ is affine, it follows that $\varphi(x)$ is concave. Therefore, the converted problem is a concave program.

Aside from the bilinear program, a number of other nonconvex problems can also be reduced to concave or quasiconcave minimization. Some of these problems (such as minimizing a sum of products of convex positive-valued functions) have been discussed in Chap. 9 (Subsect. 9.6.3).

10.4.3 Indefinite Quadratic Programs

Using an appropriate affine transformation any indefinite quadratic minimization problem under linear constraints can be written as:

$$(IQP) \qquad \min\{\langle x, x\rangle - \langle y, y\rangle \mid Ax + By + c \leq 0\},$$

where $x \in \mathbb{R}^{n_1}, y \in \mathbb{R}^{n_2}$. Earlier methods for dealing with this problem (Kough 1979; Tuy 1987b) were essentially based on Geoffrion-Benders' decomposition ideas and used the representation of the objective function as a difference of two convex quadratic separable functions. Kough's method, which closely follows Geoffrion's scheme, requires the solution, at each iteration, of a relaxed master problem which is equivalent to a number of linearly constrained concave programs, and, consequently, may not be easy to handle. Another approach is to convert (IQP) to a single concave minimization under convex constraints and to use for solving it a decomposition technique which extends to convex constraints the decomposition by outer approximation presented in Chap. 7 for linear constraints. In more detail, this approach can be described as follows.

For simplicity assume that the set $Y = \{y|\ (\exists x)\ Ax + By + c \leq 0\}$ is bounded. Setting

$$\varphi(y) = \inf\{\langle x, x\rangle|\ Ax + By + c \leq 0\}$$

it is easily seen that the function $\varphi(y)$ is convex (see Proposition 2.6) and that $\mathrm{dom}\varphi \supset Y$. Then (IQP) can be rewritten as

$$\min\{t - \langle y, y\rangle|\ \varphi(y) \leq t, y \in Y\}. \tag{10.47}$$

Here the objective function $t - \langle y, y\rangle$ is concave, while the constraint set $D := \{(y, t) \in \mathbb{R}^{n_2} \times \mathbb{R}|\ \varphi(y) \leq t, y \in Y\}$ is convex. So (10.47) is a concave program under convex constraints, which can be solved, e.g., by the outer approximation method. Since, however, the convex feasible set in this program is defined implicitly via a convex program, the following key question must be resolved for carrying out this method (Chap. 7, Sect. 7.1): given a point $(\overline{y}, \overline{t})$ determine whether it belongs to D and if it does not, construct a cutting plane (in the space $\mathbb{R}^{n_2} \times \mathbb{R}$) to separate it from D.

Clearly, $(\overline{y}, \overline{t}) \in D$ if and only if $\overline{y} \in Y$, and $\varphi(\overline{y}) \leq t$.

- One has $\overline{y} \in Y$ if and only if the linear program $\min\{0|\ Ax \leq -B\overline{y} - c\}$ is feasible, hence, by duality, if and only if the linear program

$$\max\{\langle B\overline{y} + \langle c, u\rangle|\ A^T u = 0, u \geq 0\} \tag{10.48}$$

  has optimal value 0. So, if $u = 0$ is an optimal solution of the linear program (10.48), then $\overline{y} \in Y$; otherwise, in solving this linear program an extreme direction v of the cone $A^T u = 0, u \geq 0$ is obtained such that $\langle B\overline{y} + c, v\rangle > 0$ and the cut

$$\langle By + \langle c, v\rangle \leq 0 \tag{10.49}$$

  separates $(\overline{y}, \overline{t})$ from D.
- To check the condition $\varphi(\overline{y}) \leq \overline{t}$ one has to solve the convex program

$$\min\{\langle x, x\rangle|\ Ax + B\overline{y} + c \leq 0\}. \tag{10.50}$$

Lemma 10.2 *The convex program (10.50) has at least one Karush-Kuhn-Tucker vector λ and any such vector satisfies*

$$B^T\lambda \in \partial\varphi(\overline{y}).$$

Proof Since $\langle x, x\rangle \geq 0$ one has $\varphi(\overline{y}) > -\infty$, so the existence of λ follows from a classical theorem of convex programming (see, e.g., Rockafellar 1970, Corollary 28.2.2). Then

$$\varphi(\overline{y}) = \inf\{\langle x, x\rangle + \langle \lambda, Ax + B\overline{y} + c\rangle \,|\, x \in \mathbb{R}^{n_1}\} \quad \text{(KKT vector)}$$

and for any $y \in \mathbb{R}^{n_2}$:

$$\varphi(y) \geq \inf\{\langle x, x\rangle + \langle \lambda, Ax + By + c\rangle \,|\, x \in \mathbb{R}^{n_1}\} \quad (\text{because } \lambda \geq 0).$$

Hence $\varphi(y) - \varphi(\overline{y}) \geq \langle \lambda, By - B\overline{y}\rangle = \langle B^T\lambda, y - \overline{y}\rangle$, i.e., $B^T\lambda \in \partial\varphi(\overline{y})$. □

Corollary 10.11 *If $\varphi(\overline{y}) > \overline{t}$, then a cut separating $(\overline{y}, \overline{t})$ from D is given by*

$$\langle \lambda, B(y - \overline{y})\rangle + \varphi(\overline{y}) - t \leq 0. \tag{10.51}$$

Thus, given any point $(\overline{y}, \overline{t})$ one can check whether $(\overline{y}, \overline{t}) \in D$ and if $(\overline{y}, \overline{t}) \notin D$ then construct a cut separating $(\overline{y}, \overline{t})$ from D (this cut is (10.49) if $\overline{y} \notin Y$, or (10.51) if $\varphi(\overline{y}) > \overline{t}$).

With the above background, a standard outer approximation procedure can be developed for solving the indefinite quadratic problem (*IQP*) (Tuy 1987b). At iteration k of this procedure a polytope P_k is in hand which is an outer approximation of the feasible set D. Let $V(P_k)$ be the vertex set of P_k. By solving the subproblem $\min\{t - \langle y, y\rangle \,|\, y \in V(P_k)\}$ a solution (y^k, t^k) is obtained. If $(y^k, t^k) \in D$ the procedure terminates. Otherwise a hyperplane is constructed which cuts off this solution and determines a new polytope P_{k+1}. Under mild assumptions this procedure converges to a global optimal solution.

In recent years, BB methods have been proposed for general quadratic problems which are applicable to (*IQP*) as well. These methods will be discussed in the next section and also in Chap. 11 devoted to monotonic optimization.

10.5 Quadratic Problems with Quadratic Constraints

We now turn to the general nonconvex quadratic problem :

$$(GQP) \qquad \min\{f_0(x) \,|\, f_i(x) \leq 0, i = 1, \ldots, m;\ x \in X\},$$

where X is a closed convex set and the functions $f_i,\ i = 0, 1, \ldots, m$ are indefinite quadratic:

$$f_i(x) = \frac{1}{2}\langle x, Q^i x\rangle + \langle c^i, x\rangle + d_i, \quad i = 0, 1, \ldots, m.$$

By virtue of property (iii) listed in Sect. 8.1 (GQP) can be rewritten as

$$\min\{\varphi_0(x) - r\|x\|^2 |\ \varphi_i(x) - r\|x\|^2 \leq 0, i = 1, \ldots, m;\ x \in X\}, \tag{10.52}$$

where $r \geq \max_{i=0,1,\ldots,m} \frac{1}{2}\rho(Q^i)$ and $\varphi_i(x) = f_i(x) + r\|x\|^2, i = 0, 1, \ldots, m$, are convex functions. Thus (GQP) is a d.c. optimization problem to which different approaches are possible. Aside from methods derived by specializing the general methods presented in the preceding chapters, there exist other methods designed specifically for quadratic problems. Below we shall focus on BB (branch and bound) methods.

10.5.1 Linear and Convex Relaxation

For low rank nonconvex problems, i.e., for problems which are convex when a relatively small number of variables are kept fixed, and also for problems whose quadratic structure is implicit, outer approximation (or combined OA/BB methods) may in many cases be quite practical. Two such procedures have been discussed earlier: the *Relief Indicator Method* (Chap. 7, Sect. 7.5), designed for continuous optimization and the *Visible Point Method* (Chap. 7, Sect. 7.4) which seems to work well for problems in low dimension with many nonconvex quadratic or piecewise affine constraints (such as those encountered in continuous location theory (Tuy et al. 1995a)).

For solving large-scale quadratic problems with little additional special structure branch and bound methods seem to be most suitable. Two fundamental operations in a BB algorithm are bounding and branching. As we saw in Chap. 5, these two operations must be consistent in order to ensure convergence of the procedure. Roughly speaking, the latter means that in an infinite subdivision process, the current lower bound must tend to the exact global minimum.

A standard bounding method consists in considering for each partition set a relaxed problem obtained from the original one by replacing the objective function with an affine or convex minorant and enclosing the constraint set in a polyhedron or a convex set.

a. Lower Linearization

Consider a class of problems which can be formulated as:

$$(\text{SNP}) \qquad \left| \begin{array}{ll} \min & g_0(x) + h_0(y) \\ \text{s.t.} & g_i(x) + h_i(y) \leq 0, i = 1, \ldots, m \\ & (x, y) \in \Omega, \end{array} \right.$$

where $g_0, g_i : \mathbb{R}^n \to \mathbb{R}$ are continuous functions, $h_0, h_i : \mathbb{R}^p \to \mathbb{R}$ are affine functions, while Ω is a polytope in $\mathbb{R}^n \times \mathbb{R}^p$. Obviously any quadratically constrained quadratic program is of this form. It turns out that for this problem a bounding method can be defined consistent with simplicial subdivision provided the nonconvex functions involved satisfy a lower linearizability condition to be defined below.

I. Simplicial Subdivision

A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *lower linearizable* on a compact set $C \subset \mathbb{R}^n$ with respect to simplicial subdivision if for every n-simplex $M \subset C$ there exists an affine function $\varphi_M(x)$ such that

$$\varphi_M(x) \le f(x)\ \forall x \in M, \quad \sup_{x\in M}[f(x) - \varphi_M(x)] \to 0 \text{ as } \mathrm{diam} M \to 0. \tag{10.53}$$

The function $\varphi_M(x)$ will be called a lower linearization of $f(x)$ on M. Most continuous functions are locally lower linearizable in the above sense. For example, if $f(x)$ is concave, then $\varphi_M(x)$ can be taken to be the affine function that agrees with $f(x)$ at the vertices of M; if $f(x)$ is Lipschitz, with Lipschitz constant L, then $\varphi_M(x)$ can be taken to be the affine function that agrees with $\tilde{f}(x) := f(x_M) - L\|x - x_M\|$ at the vertices of M, where x_M is an arbitrary point in M (in fact, $\tilde{f}(x)$ is concave, so $\varphi_M(x)$ is a lower linearization of $\tilde{f}(x)$, hence, also a lower linearization of $f(x)$ because by Lipschitz condition $\tilde{f}(x) \le f(x)$ and $|f(x) - \tilde{f}(x)| \le 2\|x - x_M\|\ \forall x \in M$). Thus special cases of problem (*SNP*) with locally lower linearizable $g_i, i = 0, 1, \dots, m$, include *reverse convex programs* ($g_0(x)$ linear, $g_i(x), i = 1, \dots, m$, concave), as well as *Lipschitz optimization problems* ($g_0(x), g_i(x), i = 1, \dots, m$, Lipschitz).

Now assume that every function g_i is lower linearizable and let $\psi_{M,i}(x)$ be a lower linearization of $g_i(x)$ on a given simplex M. Then the following linear program is a relaxation of (*SNP*) on M;

$$SP(M) \qquad \left|\begin{array}{l} \min\ \psi_{M,0}(x) + h_0(y) \\ \text{s.t.}\ \ \psi_{M,i}(x) + h_i(y) \le 0, i = 1, \dots, m. \\ \qquad (x, y) \in \Omega,\ x \in M. \end{array}\right.$$

Therefore, the optimal value $\beta(M)$ of this linear program is a lower bound for the objective function value on the feasible solutions (x, y) such that $x \in M$. Incorporating this lower bounding operation into an exhaustive simplicial subdivision process of the x-space yields the following algorithm:

BB Algorithm for (SNP)

Initialization. Take an n-simplex $M_1 \subset \mathbb{R}^n$ containing the projection of Ω on $\mathbb{R}^n$. Let $(\overline{x}, \overline{y})$ be the best feasible solution available, $\gamma = g_0(\overline{x})$ ($\gamma = +\infty$ if no feasible solution is known). Set $\mathscr{R} = \mathscr{S} = \{M_1\},\ k = 1$.

Step 1. For each $M \in \mathscr{S}$ compute the optimal value $\beta(M)$ of the linear program $SP(M)$.

Step 2. Update $(\bar{x}, \bar{y})$ and γ. Delete every $M \in \mathscr{R}$ such that $\beta(M) \geq \gamma$. Let $\mathscr{R}'$ be the remaining collection of simplices.

Step 3. If $\mathscr{R}' = \emptyset$, then terminate: $(\bar{x}, \bar{y})$ solves (SNP) (if $\gamma < +\infty$) or (SNP) is infeasible (if $\gamma = +\infty$).

Step 4. Let $M_k \in \text{argmin}\{\beta(M) : M \in \mathscr{R}'\}$, and let (x^k, y^k) be a basic optimal solution of $SP(M_k)$.

Step 5. Perform a bisection or a radial subdivision of M_k, following an exhaustive subdivision rule. Let $\mathscr{S}^*$ be the partition of M_k.

Step 6. Set $\mathscr{S} \leftarrow \mathscr{S}^*$, $\mathscr{R} \leftarrow \mathscr{S}^* \cup (\mathscr{R}' \setminus \{M_k\})$, increase k by 1 and return to Step 1.

The convergence of this algorithm is easy to establish. Indeed, if the algorithm is infinite, it generates a filter $\{M_{k_\nu}\}$ collapsing to a point x^*. Since $\beta(M_{k_\nu}) \leq \min(\text{SNP})$ and (10.53) implies that $\psi_{M_{k_\nu},i}(x^*) = g_i(x^*), i = 0, 1, \ldots, m$, it follows that if $\beta(M_{k_\nu}) = g_0(x^{k_\nu}) + h_0(y^{k_\nu})$, then $x^{k_\nu} \to x^*, y^{k_\nu} \to y^*$ such that (x^*, y^*) solves (SNP).

Remark 10.6 If the lower linearizations are such that for any simplex M each function $\psi_{M,i}(x)\}$ matches $g_i(x)$ at every corner of M then as soon as x^k in Step 3 coincides with a corner of M_k the solution (x^k, y^k) will be feasible, hence optimal to the problem. Therefore, in this case an adaptive subdivision rule to drive x^k eventually to a corner of M_k should ensure a much faster convergence (Theorem 6.4).

II. Rectangular Subdivision

A function $f(x)$ is said to be lower linearizable on a compact set $C \subset \mathbb{R}^n$ with respect to rectangular subdivision if for any rectangle $M \subset C$ there exists an affine function $\varphi_M(x)$ satisfying (10.53). For instance, if $f(x)$ is separable: $f(x) = \sum_{i=1}^n f_i(x)$ and each univariate function $f_i(t)$ has a lower linearization $\varphi_{M,i}(x)$ on the segment $[a_i, b_i]$ then a lower linearization of $f(x)$ on the rectangle $C = [a, b]$ is obviously $\varphi_M(x) = \sum_{i=1}^n \varphi_{M,i}(x_i)$.

For problems (SNP) where the functions $g_i(x), i = 0, 1, \ldots, m$ are lower linearizable with respect to rectangular subdivision, a rectangular algorithm can be developed similar to the simplicial algorithm.

b. Tight Convex Minorants

A convex function $\varphi(x)$ is said to be a *tight convex minorant* of a function $f(x)$ on a rectangle M (or a simplex M) if

$$\varphi(x) \leq f(x) \ \forall x \in M, \quad \min_{x \in M}[f(x) - \varphi(x)] = 0. \tag{10.54}$$

If $f(x)$ is continuous on M, then its convex envelope $\varphi(x) := \text{conv}f(x)$ over M must be a tight convex minorant, for if $\min_{x\in M}[f(x) - \varphi(x)] = \alpha > 0$ then the function $\varphi(x) + \alpha$ would give a larger convex minorant.

Example 10.1 The function $\max\{py + rx - pr, qy + sx - qs\}$ is a tight convex minorant of $xy \in \mathbb{R}^2$ on $[p, q] \times [r, s]$. It matches xy for $x = p$ or $x = q$ (and y arbitrary in the interval $r \leq y \leq s$).

Proof This function is the convex envelope of the product xy on the rectangle $[p, q]\times [r, s]$ (Corollary 4.6). The second assertion follows from the inequalities $xy - (py + rx - pr) = (x - p)(y - r) \geq 0,\ xy - (qy + sx - qs) = (q - x)(s - y) \geq 0$ which hold for every $(x, y) \in [p, q] \times [r, s]$. □

Example 10.2 Let $f(x) = \langle x, Qx\rangle$. The function

$$\varphi_M(x) = f(x) + r\sum_{i=1}^{n}(x_i - p_i)(x_i - q_i),$$

where $r \geq \rho(Q)$ is a tight convex minorant of $f(x)$ on the rectangle $M = [p, q]$. It matches $f(x)$ at each corner of the rectangle $[p, q]$, i.e., at each point x such that $x_i \in \{p_i, q_i\}, i = 1, \dots, n$.

Proof By (v), Sect. 8.1, $\varphi_M(x)$ is a convex minorant of $f(x)$ and clearly $\varphi_M(x) = f(x)$ when $x_i \in \{p_i, q_i\},\ i = 1, \dots, n$. □

Note that a d.c. representation of $f(x)$ is $f(x) = g(x) - r\|x\|^2$ with $g(x) = f(x) + r\|x\|^2$, and $r \geq \rho(Q)$. Since the convex envelope of $-\|x\|^2$ on $[p, q]$ is $l(x) = -\sum_{i=1}^{n}[(p_i + q_i)x_i - p_iq_i]$, a tight convex minorant of $f(x)$ is $g(x) + rl(x)$. It is easily verified that $g(x) + rl(x) = \varphi_M(x)$.

Example 10.3 Let Q_i be the row i and Q_{ij} be the element (i, j) of a symmetric $n \times n$ matrix Q. For any rectangle $M = [p, q] \subset \mathbb{R}^n$ define

$$k_i = \min\{Q_ix|\ x \in [p, q]\} = \textstyle\sum_{j=1}^{n} \min\{p_jQ_{ij}, q_jQ_{ij}\}, \tag{10.55}$$

$$K_i = \max\{Q_ix|\ x \in [p, q]\} = \textstyle\sum_{j=1}^{n} \max\{p_jQ_{ij}, q_jQ_{ij}\}. \tag{10.56}$$

Then a tight convex minorant of the function $f(x) = \langle x, Qx\rangle$ on $M = [p, q]$ is

$$\varphi_M(x) = \sum_{i=1}^{n} \varphi_{M,i}(x), \tag{10.57}$$

$$\varphi_{M,i}(x) = \max\{p_iQ_ix + k_ix_i - p_ik_i, q_iQ_ix + K_ix_i - q_iK_i\}. \tag{10.58}$$

Actually, $\varphi_M(x) = f(x)$ at every corner x of the rectangle $[p, q]$.

Proof We can write $f(x) = \sum_{i=1}^{n} f_i(x)$ with $f_i(x) = x_i y_i, y_i = Q_i x$ so the result simply follows from the fact that $\varphi_{M,i}(x)$ is a convex minorant of $f_i(x)$ and $\varphi_{M,i}(x) = f_i(x)$ when $x_i \in \{p_i, q_i\}$ (Example 10.1). □

Consider now the nonconvex optimization problem

$$\min\{f_0(x) | f_i(x) \leq 0, \ i = 1, \ldots, m, \ x \in D\}, \tag{10.59}$$

where $f_0, f_i : \mathbb{R}^n \to \mathbb{R}$ are nonconvex functions and D a convex set in $\mathbb{R}^n$. Given any rectangle $M \subset \mathbb{R}^n$ one can compute a lower bound for $f_0(x)$ over the feasible points x in M by solving the convex problem

$$\min\{\varphi_{M,0}(x) | \ x \in M \cap D, \ \varphi_{M,i}(x) \leq 0, \ i = 1, \ldots, m\}$$

obtained by replacing $f_i, i = 0, 1, \ldots, m$ with their convex minorants $\varphi_{M,i}$ on M. Using this bounding a BB algorithm for (10.59) with an exhaustive rectangular subdivision can then be developed.

Theorem 10.4 *Assume that $f_i, i = 0, 1, \ldots,$ are continuous functions and for every partition set M each function $\varphi_{M,i}$ is a tight convex minorant of f_i on M such that for any filter $\{M_k\}$ and any two sequences $x^k \in M_k, z^k \in M_k$:*

$$\varphi_{M_k,i}(x^k) - \varphi_{M_k,i}(z^k) \to 0 \textit{ whenever } x^k - z^k \to 0.$$

Then the corresponding BB algorithm converges.

Proof If the algorithm is infinite, it generates a filter $\{M_{k_\nu}\}$ such that $\cap_\nu M_{k_\nu} = \{\overline{x}\}$. Let us write $\varphi_{k,i}$ for $\varphi_{M_k,i}$ and denote by $z^{k,i}$ a point of M_k where $\varphi_{k,i}(z^{k,i}) = f_i(z^{k,i})$. Then for every $i = 0, 1, \ldots, m$: $z^{k_\nu,i} \to \overline{x}$. Furthermore, denote by x^k an optimal solution of the relaxed program associated with M_k, so that $\varphi_{k,0}(x^k) = \beta(M_k)$. Then $x^{k_\nu} \to \overline{x}$. Hence $f_i(\overline{x}) = \lim f_i(z^{k_\nu,i}) = \lim \varphi_{k_\nu,i}(z^{k_\nu,i}) = \lim \varphi_{k_\nu,i}(x^{k_\nu}) \ \forall i = 0, 1, \ldots, m$. Since $\varphi_{k_\nu,i}(x^{k_\nu})) \leq 0 \ \forall i = 1, \ldots, m$ it follows that $f_i(\overline{x}) \leq 0 \ \forall i = 1, \ldots, m$, i.e., $\overline{x}$ is feasible. On the other hand, for any k: $\beta(M_k) \leq \gamma := \min\{f_0(x) | f_i(x) \leq 0, i = 1, \ldots, m\}$, so $\varphi_{k_\nu,0}(x^{k_\nu}) \leq \gamma$, hence $f_0(\overline{x}) \leq \gamma$. Therefore, $f_0(\overline{x}) = \gamma$ and $\overline{x}$ is an optimal solution, proving the convergence of the algorithm. □

It can easily be verified that the convex minorants in Examples 8.1–8.3 satisfy the conditions in the above theorem. Consequently, BB algorithms using these convex minorants for computing bounds are guaranteed to converge. An algorithm of this type has been proposed in Al-Khayyal et al. (1995) (see also Van Voorhis 1997) for the quadratic program (10.59) where $f_0(x) = \langle x, Q^0 x\rangle + \langle c^0, x\rangle, f_h(x) = \langle x, Q^h x\rangle + \langle c^h, x\rangle + d_h, \ h = 1, \ldots, m$ and D is a polytope. For each rectangle M the convex minorant of $\langle x, Q^h x\rangle$ over M is $\varphi_M^h(x)$ computed as in Example 10.3. To obtain an upper bound one can use, in addition, a concave majorant for each function f_h. By similar arguments to the above, it is easily seen that a concave majorant of $\langle x, Q^h x\rangle$ is $\psi^h_M(x) = \sum_{i=1}^n \psi^h_i(x)$ with $\psi^h_i(x) = \min\{p_i Q^h_i x + K^h_i x_i - p_i K^h_i,\ q_i Q^h_i x + k^h_i x_i - q_i k^h_i\}$. So the relaxed problem is the linear program

$$\begin{array}{ll} \min & \varphi^0_M(x) + \langle c^0, x\rangle \\ \text{s.t.} & t^h + \langle c^h, x\rangle + d_h \le 0 \quad h = 1, \dots, m \\ & \varphi^h_M(x) \le t_h \le \psi^h_M(x) \quad h = 1, \dots, m \\ & x \in M \cap D \qquad\qquad . \end{array}$$

Tight convex minorants of the above type are also used in Androulakis et al. (1995) for the general nonconvex problem (10.59). By writing each function $f_h(x), h = 0, 1, \dots, m$, in (10.59) as a sum of bilinear terms of the form $b_{ij} x_i x_j$ and a nonconvex function with no special structure, a convex minorant $\varphi_{M,h}(x)$ of $f_h(x)$ on a rectangle M can easily be computed from Examples 10.1 and 10.2. Such a convex minorant is tight because it matches $f_h(x)$ at any corner of M; furthermore, as is easily seen, $\varphi_{M,h}(x), h = 0, 1, \dots, m$, satisfy the conditions formulated in Theorem 10.4. The convergence of the BB algorithm in Androulakis et al. (1995) then follows.

The reader is referred to the cited papers (Al-Khayyal et al. 1995; Androulakis et al. 1995) for details and computational results concerning the above algorithms. In fact the convex minorants in Examples 10.1–10.3 could be obtained from the general rules for computing convex minorants of factorable functions (see Chap. 4, Sect. 4.7). In other words the above algorithms could be viewed as specific realizations of the general scheme of McCormick (1982) for *factorable programming*.

Remark 10.7 (i) If the convex minorants used are of the types given in Examples 8.1–8.3 then for any rectangle M each function $\varphi_{M,i}$ matches f_i at the corners of M. In that case, the convergence of the algorithm would be faster by using the ω-subdivision, viz : at iteration k divide $M_k = [p^k, q^k]$ via (ω^k, j_k), where ω^k is an optimal solution of the relaxed problem at iteration k and j_k is a properly chosen index ($j_k \in \text{argmax}\{\eta_{k,j} |\ j = 1, \dots, n\}$ where $\eta_{k,j} = \min\{\omega^k_j - p^k_j, q^k_j - \omega^k_j\}$; see Theorem 6.4.

(ii) A function $f(x)$ is said to be *locally convexifiable* on a compact set $C \subset \mathbb{R}^n$ with respect to rectangular subdivision if there exists for each rectangle $M \subset C$ a convex minorant $\varphi_M(x)$ satisfying (10.53) (this convex minorant is then said to be asymptotically tight). By an argument analogous to the proof of Theorem 10.4 it is easy to show that if the functions $f_i(x), i = 0, 1, \dots, m$ in (10.59) are locally convexifiable and the convex minorants $\varphi_{M,i}$ used for bounding are asymptotically tight, while the subdivision is exhaustive then the BB algorithm converges. For the applications, note that if two functions $f_1(x), f_2(x)$ are locally convexifiable then so will be their sum $f_1(x) + f_2(x)$.

c. Decoupling Relaxation

A lower bounding method which can be termed *decoupling relaxation* consists in converting the problem into an equivalent one where the nonconvexity is concentrated in a number of "coupling constraints." When these coupling constraints are omitted, the problem becomes linear, or convex, or easily solvable. The method proposed in Muu and Oettli (1991), Muu (1993), and Muu (1996), for convex–concave problems, and also the reformulation–linearization method of Sherali and Tuncbilek (1992, 1995), can be interpreted as different variants of variable decoupling relaxations.

d. Convex–Concave Approach

A real-valued function $f(x, y)$ on $X \times Y \subset \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ is said to be convex–concave if it is convex on X for every fixed $y \in Y$ and concave on Y for every fixed $x \in X$. Consider the problem

$$(\text{CCAP}) \qquad \min\{f(x, y) |\ (x, y) \in C,\ x \in X,\ y \in Y\}, \tag{10.60}$$

where $X \subset \mathbb{R}^{n_1}, Y \subset \mathbb{R}^{n_2}$ are polytopes, C is a compact convex set in $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ and the function $f(x, y)$ is convex–concave on $X \times Y$. Clearly, linearly (or convexly) constrained quadratic programs and more generally, d.c. programming problems with convex constraints, are of this form. Since by fixing y the problem is convex, a branch and bound method for solving $(CCAP)$ should branch upon y. Using, e.g., rectangular partition of the y-space, one has to compute, for a given rectangle $M \subset \mathbb{R}^{n_2}$, a lower bound $\beta(M)$ for

$$(SP(M)) \qquad \min\{f(x, y) |\ (x, y) \in C, x \in X, y \in Y \cap M\}. \tag{10.61}$$

The difficulty of this subproblem is due to the fact that convexity and concavity are coupled in the contraints. To get round this difficulty the idea is to decouple x and y by reformulating the problem as

$$\min\{f(x, y) |\ (x, u) \in C, x \in X, u \in Y, y \in M, u = y\}$$

and omitting the constraint $u = y$. Then, obviously, the optimal value in the relaxed problem

$$(RP(M)) \qquad \min\{f(x, y) |\ (x, u) \in C, x \in X, u \in Y, y \in M\} \tag{10.62}$$

is a lower bound of the optimal value in $(CCAP)$. Let $V(M)$ be the vertex set of M. By concavity of $f(x, y)$ in y we have

$$\min\{f(x, y) |\ y \in M\} = \min\{f(x, y) | y \in V(M)\},$$

so the problem (10.62) is equivalent to

$$\min_{y \in V(M)} \min\{f(x, y) |\ (x, u) \in C, x \in X, u \in Y\}.$$

Therefore, one can take

$$\beta(M) = \min_{y \in V(M)} \min\{f(x,y)|\ (x,u) \in C \cap (X \times Y)\}. \tag{10.63}$$

For each fixed y the problem $\min\{f(x,y)|\ (x,u) \in C \cap (X \times Y)\}$ is convex (or linear if $f(x,y)$ is affine in x, and C is a polytope), so (10.63) can be solved by standard algorithms. If (x^M, y^M, u^M) is an optimal solution of $RP(M)$ then along with the lower bound $f(x^k, y^k) = \beta(M_k)$, we also obtain a feasible solution (x^M, u^M) which can be used to update the current incumbent. Furthermore, when $y^M = u^M$, $\beta(M)$ equals the exact minimum in (10.63).

At iteration kth of the branch and bound procedure let M_k be the rectangle such that $\beta(M_k)$ is smallest among all partition sets still of interest. Denote $(x^k, y^k, u^k) = (x^{M_k}, y^{M_k}, u^{M_k})$. If $u^k = y^k$ then, as mentioned above, $\beta(M_k) = f(x^k, y^k)$ while (x^k, y^k) is a feasible solution. Since $\beta(M_k) \leq \min(\text{CCAP})$ it follows that (x^k, y^k) is an optimal solution of $(CCAP)$. Therefore, if $u^k = y^k$ the procedure terminates. Otherwise, M_k must be further subdivided. Based on the above observation, one should choose a subdivision method which, intuitively, could drive $\|y^k - u^k\|$ to zero more rapidly than an exhaustive subdivision process. A good choice confirmed by computational experience is an adaptive subdivision via (y^k, u^k), i.e., to subdivide M_k via the hyperplane $y_{i_k} = (y^k_{i_k} + u^k_{i_k})/2$ where

$$i_k \in \operatorname{argmax}_{i=1,\dots,n_2} |y^k_i - u^k_i|.$$

Proposition 10.7 *The rectangular algorithm with the above branching and bounding operations is convergent.*

Proof If the algorithm is infinite, it generates at least a filter of rectangles, say $M_{k_q}, q = 1, 2, \dots$. To simplify the notation we write M_q for M_{k_q}. Let $M_q = [a^q, b^q]$, so that by a basic property of rectangular subdivision (Lemma 5.4), one can suppose, without loss of generality, that $i_q = 1, [a^q, b^q] \to [a^*, b^*], (y^q_1 + u^q_1)/2 \to y^*_1 \in \{a^*_1, b^*_1\}$. The latter relation implies that $|y^q_1 - u^q_1| \to 0$, hence $y^q - u^q \to 0$ because $|y^q_1 - u^q_1| = \max_{i=1,\dots,n_2} |y^q_i - u^q_i|$ by the definition of $i_q = 1$. Let $x^* = \lim x^q, y^* = \lim y^q (= \lim u^q)$. Then $\beta(M_q) = f(x^q, y^q) \to f(x^*, y^*)$ and since $\beta(M_q) \leq \min(\text{CCAP})$ for all q, while (x^*, y^*) is feasible, it follows that (x^*, y^*) is an optimal solution. □

This method is practicable if n_2 is relatively small. Often, instead of rectangular subdivision, it may be more convenient to use simplicial subdivision. In that case, a simplex M_k is subdivided via $\omega^k = (y^k + u^k)/2$ according to the ω-subdivision rule. The method can also be extended to problems in which convex and nonconvex variables are coupled in certain joint constraints, e.g., to problems of the form

$$\min\{f(x)|\ (x,y) \in C,\ g(x,y) \leq 0,\ x \in X, y \in Y\}, \tag{10.64}$$

where $f(x)$ is a convex function, X, Y are as previously, and $g(x, y)$ is a convex–concave function on $X \times Y$. For any rectangle M in the y-space, one can take

$$\begin{aligned}\beta(M) &= \min\{f(x)|\ (x, u) \in C,\ g(x, y) \leq 0,\ x \in X, u \in Y, y \in M\}\\ &= \min\{f(x)|\ (x, u) \in C \cap (X \times Y), g(x, y) \leq 0, y \in V(M)\}.\end{aligned}$$

Computing $\beta(M)$ again amounts to solving a convex subproblem.

e. Reformulation–Linearization

Let X, Y be two polyhedrons in $\mathbb{R}^{n_1}, \mathbb{R}^{n_2}, n_1 \leq n_2$, defined by reformulation–linearization

$$X = \{x \in \mathbb{R}^{n_1}|\ \langle a^i, x\rangle \leq \alpha_i,\ \ i = 1, \ldots, h_1\} \tag{10.65}$$

$$Y = \{y \in \mathbb{R}^{n_2}|\ \langle b^j, y\rangle \leq \beta_j,\ \ j = 1, \ldots, h_2\}, \tag{10.66}$$

and let

$$f_0(x, y) = \langle x, Q^0 y\rangle + \langle c^0, x\rangle + \langle d^0, y\rangle \tag{10.67}$$

$$f_i(x, y) = \langle x, Q^i y\rangle + \langle c^i, x\rangle + \langle d^i, y\rangle + g_i,\ i = 1, \ldots, m, \tag{10.68}$$

where for each $i = 0, 1, \ldots, m$ Q^i is an $n_1 \times n_2$ matrix, $c^i \in \mathbb{R}^{n_1}, d^i \in \mathbb{R}^{n_2}$ and $g_i \in \mathbb{R}$. Consider the general bilinear program

$$\min\{f_0(x, y)|\, f_i(x, y) \leq 0\ (i = 1, \ldots, m),\ x \in X,\ y \in Y\}. \tag{10.69}$$

Assume that the linear inequalities defining X and Y are such that both systems

$$\langle a^i, x\rangle - \alpha_i \geq 0 \quad i = 1, \ldots, h_1 \tag{10.70}$$

$$\langle b^j, y\rangle - \beta_j \geq 0 \quad j = 1, \ldots, h_2 \tag{10.71}$$

are inconsistent, or equivalently,

$$\begin{aligned}&\forall x \in \mathbb{R}^{n_1}\ \ \min_{i=1,\ldots,h_1}(\langle a^i, x\rangle - \alpha_i) < 0,\\ &\forall y \in \mathbb{R}^{n_2}\ \ \min_{j=1,\ldots,h_2}(\langle b^j, y\rangle - \beta_j) < 0.\end{aligned} \tag{10.72}$$

This condition is fulfilled if the inequalities defining X and Y include inequalities of the form

$$-\infty < p \leq x_{i_0} \leq q < +\infty \tag{10.73}$$

$$-\infty < r \leq y_{j_0} \leq s < +\infty \tag{10.74}$$

for some $i_0 \in \{1, \ldots, h_1\}, p < q$, and some $j_0 \in \{1, \ldots, h_2\}, r < s$; in particular, it is fulfilled if X and Y are bounded.

Lemma 10.3 *Under condition (10.72) the system $x \in X, y \in Y$ is equivalent to*

$$(\langle a^i, x\rangle - \alpha_i)(\langle b^j, y\rangle - \beta_j) \geq 0 \quad \forall i = 1, \ldots, h_1, j = 1, \ldots, h_2. \tag{10.75}$$

Proof It suffices to show that (10.75) implies $(x, y) \in X \times Y$. Let (x, y) satisfy (10.75). By (10.72) there exists for this x an index $j_0 \in \{1, \ldots, h_2\}$ such that $\langle b^{j_0}, y\rangle - \beta_{j_0} < 0$. Then for any $i = 1, \ldots, h_1$, since

$$(\langle a^i, x\rangle - \alpha_i)(\langle b^{j_0}, y\rangle - \beta_{j_0}) \geq 0,$$

it follows that $\langle a^i, x\rangle - \alpha_i \leq 0$. Similarly, for any $j = 1, \ldots, h_2 : \langle b^j, y\rangle - \beta_j \geq 0$. □

Thus, a system of two sets of linear constraints in x and in y like (10.65) and (10.66) can be converted to an equivalent system of formally bilinear constraints. Now for each bilinear function $P(x, y)$ denote by $[P(x, y]_L$ the affine function of x, y, w obtained by replacing each product $x_l y_{l'}$ by $w_{ll'}$. For instance, if $P(x, y) = (\langle a, x\rangle - \alpha)(\langle b, y\rangle - \beta)$ where $a \in \mathbb{R}^{n_1}, b \in \mathbb{R}^{n_2}, \alpha, \beta \in \mathbb{R}$ then

$$[(\langle a, x\rangle - \alpha)(\langle b, y\rangle - \beta)]_L = \sum_{l=1}^{n_1}\sum_{l'=1}^{n_2} a_l b_{l'} w_{ll'} - \alpha\langle b, y\rangle - \beta\langle a, x\rangle + \alpha\beta.$$

Let $[f_i(x, y)]_L = \varphi_i(x, y, w)$, $i = 0, 1, \ldots, m$ and for every (i, j): $[(\langle a^i, x\rangle - \alpha_i)(\langle b^j, y\rangle - \beta_j)]_L = L_{ij}(x, y, w)$. Then the bilinear program (10.69) can be written as

$$\left|\begin{array}{l} \min\ \varphi_0(x, y, w) \\ \text{s.t.}\ \ \varphi_i(x, y, w) \leq 0 \quad i = 1, \ldots, m, \\ \qquad L_{ij}(x, y, w) \geq 0 \quad i = 1, \ldots, h_1, j = 1, \ldots, h_2 \\ \qquad w_{ll'} = x_l y_{l'} \quad l = 1, \ldots, n_1, l' = 1, \ldots, n_2. \end{array}\right. \tag{10.76}$$

A new feature of the latter problem is that the only nonlinear elements in it are the coupling constraints $w_{ll'} = x_l y_{l'}$. Let (LP) be the linear program obtained from (10.76) by omitting these coupling constraints:

$$\text{(LP)} \qquad \left|\begin{array}{l} \min\ \varphi_0(x, y, w) \\ \text{s.t.}\ \ \varphi_i(x, y, w) \leq 0 \quad i = 1, \ldots, m, \\ \qquad L_{ij}(x, y, w) \geq 0 \quad i = 1, \ldots, h_1, j = 1, \ldots, h_2. \end{array}\right.$$

Proposition 10.8 *The optimal value in the linear program (LP) is a lower bound of the optimal value in (10.69). If an optimal solution $(\overline{x}, \overline{y}, \overline{w})$ of (LP) satisfies $\overline{w}_{ll'} = \overline{x}_l \overline{y}_{l'}\ \forall l, l'$ then it solves (10.69).*

Proof This follows because (LP) is a relaxation of the problem (10.76) equivalent to (10.69). □

Remark 10.8 The replacement of the linear constraints $x \in X, y \in Y$, by constraints $L_{ij}(x, y, w) \geq 0, i = 1, \ldots, h_1, j = 1, \ldots, h_2$, is essential. For example, to compute a lower (or upper) bound for the product $xy \in \mathbb{R} \times \mathbb{R}$ over the rectangle $p \leq x \leq q,\ r \leq y \leq s$, one should minimize (maximize, resp.) the function w under the constraints

$$[(x-p)(y-r)]_L \geq 0, \quad [(q-x)(s-y)]_L \geq 0$$
$$[(x-p)(s-y)]_L \geq 0, \quad [(q-x)(y-r)]_L \geq 0.$$

This yields the well-known bounds given in Example 8.1:

$$\max\{rx + py - pr, sx + qy - qs\} \leq w \leq \min\{sx + py - ps, rx + qy - qr\}.$$

Notice that trivial bounds $-\infty < w < +\infty$ would have been obtained, were the original linear constraints used instead of the constraints $L_{ij}(x, y, w) \geq 0$. Thus, although it is more natural to linearize nonlinear constraints, it may sometimes be useful to tranform linear constraints into formally nonlinear constraints.

From the above results we now derive a BB algorithm for solving (10.69). For simplicity, consider the special case of (10.69) when $x = y$, i.e., the general nonconvex quadratic program

$$\min\{f_0(x) | f_i(x) \leq 0\ (i = 1, \ldots, m),\ x \in X\}, \tag{10.77}$$

where X is a polyhedron in $\mathbb{R}^n$ and

$$f_0(x) = \frac{1}{2}\langle x, Q^0 x\rangle + \langle c^0, x\rangle,\ f_i(x) = \frac{1}{2}\langle x, Q^i x\rangle + \langle c^i, x\rangle + d_i. \tag{10.78}$$

Given any rectangle M in $\mathbb{R}^n$, the set $X \cap M$ can always be described by a system

$$\langle a^i, x\rangle \leq \alpha_i \quad i = 1, \ldots, h \tag{10.79}$$

which includes among others all the inequalities $p_i \leq x_i \leq q_i, i = 1, \ldots, n$ and satisfies (10.72), i.e., $\min_{i=1,\ldots,h}(\langle a^i, x\rangle - \alpha_i) < 0\ \forall x \in \mathbb{R}^n$. For every (i, j) with $1 \leq i \leq j \leq h$ denote by

$$L_{ij}(x, w) = [(\langle a^i, x\rangle - \alpha_i)(\langle a^j, x\rangle - \alpha_j)]_L$$

the function that results from $(\langle a^i, x\rangle - \alpha_i)(\langle a^j, x\rangle - \alpha_j)$ by the substitution $w_{ll'} = x_l x_{l'},\ 1 \leq l \leq l' \leq n$. Also let

$$\varphi_0(x,w) = \left[\frac{1}{2}\langle x, Q^0 x\rangle\right]_L + \langle c^0, x\rangle$$

$$\varphi_i(x,w) = \left[\frac{1}{2}\langle x, Q^i x\rangle\right]_L + \langle c^i, x\rangle + d_i \quad i = 1,\ldots,m.$$

Then a lower bound of $f_0(x)$ over all feasible points x in $X \cap M$ is given by the optimal value $\beta(M)$ in the linear program

$$LP(M) \qquad \left| \begin{array}{ll} \min & \varphi_0(x,w) \\ \text{s.t.} & \varphi_i(x,w) \le 0 \;\; i = 1,\ldots,m \\ & L_{ij}(x,w) \ge 0 \;\; 1 \le i \le j \le h. \end{array} \right.$$

To use this lower bounding for algorithmic purpose, one has to construct a subdivision process consistent with this lower bounding, i.e., such that the resulting BB algorithm is convergent. The next result gives the foundation for such a subdivision process.

Proposition 10.9 *Let $M = [p,q]$. If (x,w) is a feasible point of $LP(M)$ such that x is a corner of the rectangle M then (x,w) satisfies the coupling constraints $w_{ll'} = x_l x_{l'}$ for all l, l'.*

Proof Since x is a corner of $M = [p,q]$ one has, for some $I \subset \{1,\ldots,n\}$: $x_i = p_i$ for $i \in I$ and $x_j = q_j$ for $j \notin I$. The feasibility of (x,w) implies that $[(x_l - p_l)(q_{l'} - x_{l'})]_L \ge 0$, $[(x_l - p_l)(x_{l'} - p_{l'})]_L \ge 0$ for any l, l', i.e.,

$$p_l x_{l'} + p_{l'} x_l - p_l p_{l'} \le w_{ll'} \le q_{l'} x_l + p_l x_{l'} - p_l q_{l'}.$$

For $l \in I$ we have $x_l = p_l$, so

$$x_l x_{l'} + p_{l'} p_l - p_l p_{l'} \le w_{ll'} \le q_{l'} p_l + x_l x_{l'} - p_l q_{l'},$$

hence $w_{ll'} = x_l x_{l'}$. On the other hand, $[(q_l - x_l)(q_{l'} - x_{l'})]_k \ge 0$, $[(q_l - x_l)(x_{l'} - p_{l'})]_k \ge 0$, i.e.,

$$-q_l q_{l'} + q_{l'} x_l + q_l x_{l'} \le w_{ll'} \le q_l x_{l'} - q_l p_{l'} + p_{l'} x_l.$$

For $l \notin I$ we have $x_l = q_l$, so

$$-q_l q_{l'} + q_{l'} q_l + x_l x_{l'} \le w_{ll'} \le x_l x_{l'} - q_l p_{l'} + p_{l'} q_l,$$

hence $w_{ll'} = x_l x_{l'}$. Thus, in any case, $w_{ll'} = x_l x_{l'}$. □

Based on this property the following rectangular BB procedure can be proposed for solving (10.77). At iteration k, let $M_k = [p^k, q^k]$ be the partition set with smallest lower bound among all partition sets still of interest and let (x^k, w^k) be the optimal solution of the bounding problem, so that $\beta(M_k) = \varphi(x^k, w^k)$. If $w^k_{ll'} = x^k_l x^k_{l'}$ for all l, l' (in particular, if x^k is a corner of M_k) then x^k is an optimal solution of (10.77) and the procedure terminates. Otherwise, bisect M_k via (x^k, i_k) where

$$i_k \in \operatorname{argmax}\{\eta^k_i | \ i = 1, \ldots, n\}, \quad \eta^k_i = \min\{x^k_i - p^k_i, q^k_i - x^k_i\}.$$

Proposition 10.10 *The rectangular algorithm with bounding and branching as above defined converges to an optimal solution of (10.77).*

Proof The subdivision ensures that if the algorithm is infinite it generates a filter of rectangles $\{M_\nu = [p^\nu, q^\nu]\}$ such that $(x^\nu, w^\nu) \to (x^*, w^*)$, and x^* is a corner of $[p^*, q^*] = \cap_\nu [p^\nu, q^\nu]$ (Theorem 6.3). Clearly (x^*, w^*) is feasible to $LP(M_*)$ for $M_* = [p^*, q^*]$, so by Proposition 10.9, $w^*_{ll'} = x^*_l x^*_{l'}$ for all l, l', and hence, x^* is a feasible solution to (10.77). Since $\varphi_0(x^*, w^*) = f_0(x^*)$ is a lower bound for $f_0(x)$ over the feasible solutions, it follows that x^* is an optimal solution. □

Example 10.4 (Sherali and Tuncbilek 1995) Minimize $-(x_1 - 12)^2 - x_2^2$ subject to

$$-6x_1 + 8x_2 \le 48, \ 3x_1 + 8x_2 \le 120, \ 0 \le x_1 \le 24, \ x_2 \ge 0.$$

Setting $w_{11} = x_1^2, \ w_{12} = x_1 x_2, \ w_{22} = x_2^2$ the first bounding problem is to minimize the linear function $-w_{11} + 24x_1 - w_{22} - 144$ subject to $(5 \times 6)/2 = 15$ linear constraints corresponding to different pairwise products of the factors

$$(48 + 6x_1 - 8x_2) \ge 0, \ (120 - 3x_1 - 8x_2) \ge 0, \ (24 - x_1) \ge 0, \ x_1 \ge 0, \ x_2 \ge 0.$$

The solution to this linear program is $(\overline{x}_1, \overline{x}_2, \overline{w}_{11}, \overline{w}_{12}, \overline{w}_{22}) = (8, 6, 192, 48, 72)$, with objective function value -216. By bisecting the feasible set X into $M_1^- = \{x \in X | \ x_1 \le 8\}$ and $M_1^+ = \{x \in X | \ x_1 \ge 8\}$ and solving the corresponding bounding problems one obtains a lower bound of -180 for both. Furthermore, the solution $(x_1^*, x_2^*, w_{11}^*, w_{12}^*, w_{22}^*) = (0, 6, 0, 0, 36)$ to the bounding problem on M_1^- is such that $(x_1^*, x_2^*) = (0, 6)$ is a vertex of M_1^-. Hence, by Proposition 10.9, $(0, 6)$ is an optimal solution of the original problem and -180 is the optimal value. This result could also be checked by solving the problem as a concave minimization over a polytope (the vertices of this polytope are $(0, 0)$, $(0, 6)$, $(8, 12)$, $(24, 6)$, and $(24, 0)$).

Remark 10.9 Since the number of constraints $L_{ij}(x, w) \ge 0$ is often quite large, it may sometimes be more practical to use only a selected subset of these constraints, even if this may result in a poorer bound. On the other hand, it is always possible to add some convex constraints of the form $x_i^2 \le w_{ii}$ to obtain a convex rather than linear relaxation of the original problem. This may give, of course, a tighter bound at the expense of more computational effort.

10.5.2 Lagrangian Relaxation

Consider again the general problem

$$f_X^\natural := \inf\{f_0(x) | f_i(x) \le 0 \ (i = 1, \dots, m),\ x \in X\}, \tag{10.80}$$

where X is a convex set in $\mathbb{R}^n$, $f_0(x)$, $f_i(x)$ are arbitrary real-valued functions. The function

$$L(x, u) = f_0(x) + \sum_{i=1}^m u_i f_i(x), \quad u \in \mathbb{R}^m_+$$

is called the *Lagrangian* of problem (10.80) and the problem

$$\psi_X^* := \sup_{u \in \mathbb{R}^m_+} \psi(u), \quad \psi(u) = \inf_{x \in X} L(x, u) \tag{10.81}$$

the *Lagrange dual problem* . Clearly $\psi(u) \le f_X^\natural\ \forall u \in \mathbb{R}^m_+$, hence

$$\psi_X^* \le f_X^\natural, \tag{10.82}$$

i.e., ψ_X^* always provides a lower bound for $f_X^\natural$. The difference $f_X^\natural - \psi_X^* \ge 0$ is referred to as the *Lagrange duality gap*. When the gap is zero, i.e., strong duality holds, then ψ_X^* is the exact optimal value of (10.80).

Note that the function $\psi(u)$ is concave on every convex subset of its domain as it is the lower envelope of a family of affine functions in u. Furthermore:

Proposition 10.11 *Assume X is compact while $f_0(x)$ and $f_i(x)$ are continuous on X. Then $\psi(u)$ is defined and concave on $\mathbb{R}^m_+$ and if $\psi(\overline{u}) = L(\overline{x}, \overline{u})$ then the vector $-(f_i(\overline{x}), i = 1, \dots, m)$ is a subgradient of $-\psi(u)$ at $\overline{u}$.*

Proof It is enough to prove the last part of the proposition. For any $u \in \mathbb{R}^m_+$ we have

$$\psi(u) - \psi(\overline{u}) \le L(\overline{x}, u) - L(\overline{x}, \overline{u}) = \sum_{i=1}^m (u_i - \overline{u}_i) f_i(\overline{x}).$$

The rest is obvious. □

The next three propositions indicate important cases when strong duality holds.

Proposition 10.12 *Assume $f_0(x)$ and $f_i(x)$ are convex and finite on X. If the optimal value in (10.80) is not $-\infty$, and the Slater regularity condition holds, i.e., there exists at least one feasible solution x^0 of the problem such that*

$$x^0 \in \text{ri}X, \quad f_i(x^0) < 0 \quad \forall i \in I,$$

where $I = \{i = 1, \dots, m | f_i(x)$ *is not affine*$\}$, *then*

$$\psi_X^* = f_X^\natural.$$

Proof This is a classical result from convex programming theory which can also be directly derived from Theorem 2.3 on convex inequalities. □

Proposition 10.13 *In (10.80) where* $X = \mathbb{R}^n$, *assume that the concave function* $u \mapsto \inf_{x\in\mathbb{R}^n} L(x,u)$ *attains its maximum at a point* $\overline{u} \in \mathbb{R}^m_+$ *such that* $L(x,\overline{u})$ *is strictly convex. Then strong duality holds, so*

$$\psi_X^* = f_X^\natural.$$

Proof This follows from Corollary 10.1 with $C = \mathbb{R}^n$, $D = \mathbb{R}^m_+$ and Remark 10.1(a) by noting that by Proposition 10.3 $L(x,\overline{u})$ is coercive on $\mathbb{R}^n$. □

Example 10.5 Consider the following minimization of a concave quadratic function under linear constraints

$$\min_{x\in\mathbb{R}^{10}} \{-50||x||^2 + c^T x \mid Ax \le b,\ x_i \in [0,1],\ i = 1,2,\dots,10\}, \tag{10.83}$$

$$A = \begin{bmatrix} -2 & -6 & -1 & 0 & -3 & -3 & -2 & -6 & -2 & -2 \\ 6 & -5 & 8 & -3 & 0 & 1 & 3 & 8 & 9 & -3 \\ -5 & 6 & 5 & 3 & 8 & -8 & 9 & 2 & 0 & -9 \\ 9 & 5 & 0 & -9 & 1 & -8 & 3 & -9 & -9 & -3 \\ -8 & 7 & -4 & -5 & -9 & 1 & -7 & -1 & 3 & -2 \end{bmatrix}$$

$$c = (48,42,48,45,44,41,47,42,45,46)^T,\ b = (-4,22,-3,-23,-12)^T.$$

The box constraint $0 \le x_i \le 1, i = 1,\dots,10$, can be written as 10 quadratic constraints $x_i^2 - x_i \le 0,\ i = 1,2,\dots,10$, so actually there are in total 15 constraints in (10.83).

By solving a SDP, it can be checked that $\max_{y\in\mathbb{R}^{15}_+} \inf_{x\in\mathbb{R}^{10}} L(x,y)$ is attained at $\overline{y}$ such that $L(x,\overline{y})$ is strictly convex and attains its minimum at

$$\overline{x} = (1,1,0,1,1,1,0,1,1,1)^T$$

which is feasible to (10.83). The optimal value is -47 with zero duality gap by Proposition 10.13.

Proposition 10.14 *In (10.80) where* $X = \mathbb{R}^n$ *assume there exists* $\overline{y} \in \mathbb{R}^m_+$ *such that that the function* $L(x,\overline{y}) := f(x) + \sum_{i=1}^m \overline{y}_i g_i(x)$ *is convex and has a minimizer* $\overline{x}$ *satisfying* $\overline{y}_i g_i(\overline{x}) = 0,\ g_i(\overline{x}) \le 0\ \forall i = 1,\dots,m$. *Then strong duality holds, so*

$$\psi_X^* = f_X^\natural.$$

Proof It is readily verified that

$$L(\bar{x}, y) \leq L(\bar{x}, \bar{y}) \leq L(x, \bar{y}) \quad \forall x \in \mathbb{R}^n,\ \forall y \in \mathbb{R}^m_+,$$

so $(\bar{x}, \bar{y})$ is a saddle point of $L(x, y)$ on $\mathbb{R}^n \times \mathbb{R}^m_+$, and hence, by Proposition 2.35,

$$\max_{y \in \mathbb{R}^m_+} \inf_{x \in \mathbb{R}^n} L(x, y) = \min_{x \in \mathbb{R}^n} \sup_{y \in \mathbb{R}^m_+} L(x, y). \qquad \square$$

Example 10.6 Consider the following nonconvex quadratic problem which arises from the so-called optimal beamforming in wireless communication (Bengtsson and Ottersen 2001):

$$\min_{x_i \in \mathbb{R}^{N_i},\ i=1,2,\ldots,d} \sum_{i=1}^{d} ||x_i||^2 : \ -x_i^T R_{ii} x_i + \sum_{j \neq i} x_j^T R_{ij} x_j + \sigma_i \leq 0,\ i = 1, 2, \ldots, d, \tag{10.84}$$

where

$$0 < \sigma_i, \quad 0 \preceq R_{ij} \in \mathbb{R}^{N_i \times N_i},\ i, j = 1, 2, \ldots, d, \tag{10.85}$$

so each x_i is a vector in $\mathbb{R}^{N_i}$ while each R_{ij} is a positive semidefinite matrix of dimension $N_i \times N_i$.

Applying Proposition 10.14 yields the following result (Tuy and Tuan 2013):

Corollary 10.12 *Strong Lagrangian duality holds for problem (10.84), i.e.,*

$$\min_{x_i \in \mathbb{R}^{N_i},\ i=1,2,\ldots,d} \sup_{y \in \mathbb{R}^d_+} L(x, y) = \max_{y \in \mathbb{R}^d_+} \min_{x_i \in \mathbb{R}^{N_i},\ i=1,2,\ldots,d} L(x, y), \tag{10.86}$$

where $L(x, y) = \sum_{i=1}^{d} x_i^T (I + \sum_{j \neq i} y_j R_{ji} - y_i R_{ii}) x_i + \sum_{i=1}^{d} \sigma_i y_i.$

Proof It is convenient to precede the proof by a lemma. As usual the inequalities $y > 0$ or $y < x$ for vectors y and x are component-wise understood.

Lemma 10.4 *For a given matrix $R \in \mathbb{R}^{d \times d}$ with positive diagonal entries $r_{ii} > 0$ and nonpositive off-diagonal entries $r_{ij} \leq 0$, $i \neq j$, assume that there is $\bar{y} > 0$ such that*

$$R^T \bar{y} > 0. \tag{10.87}$$

Then the inverse matrix R^{-1} is positive, i.e., all its entries are positive.

Proof Define a positive diagonal matrix $D := \text{diag}[r_{ii}]_{i=1,2,\ldots,d}$ and nonnegative matrix $G = D - R$, so $R = D - G$. Then, (10.87) means $\bar{\sigma} := D\bar{y} - G^T \bar{y} > 0$ and $0 < D^{-1} G^T \bar{y} = \bar{y} - D^{-1} \bar{\sigma} < \bar{y}$. Due to (10.87), without loss of generality we can assume $r_{ij} < 0$ so GD^{-1} is irreducible. Let $\lambda_{\max}[.]$ stand for the maximal eigenvalue of a matrix. As $D^{-1}G^T$ is obviously nonnegative, by Perron–Frobenius theorem (Gantmacher 1960, p.69), we can write

$$\begin{aligned}\lambda_{\max}[GD^{-1}] = \lambda_{\max}[D^{-1}G^T] &= \min_{y>0} \max_{i=1,2,\dots,d} \frac{(D^{-1}G^T y)_i}{y_i} \\ &\leq \max_{i=1,2,\dots,d} \frac{(D^{-1}G^T \overline{y})_i}{\overline{y}_i} < 1.\end{aligned}$$

Again by Perron–Frobenius theorem, the matrix $I - GD^{-1}$ is invertible; moreover its inverse $(I - GD^{-1})^{-1}$ is positive. Then $R^{-1} = D^{-1}(I - GD^{-1})^{-1}$ is obviously positive. □

Turning to the proof of Corollary 10.12, let

$$\eta := \sup_{y \in \mathbb{R}^d_+} \min_{x_i \in \mathbb{R}^{N_i},\ i=1,2,\dots,d} \left[\sum_{i=1}^{d} x_i^T \left(I + \sum_{j \neq i} y_j R_{ji} - y_i R_{ii} \right) x_i + \sum_{i=1}^{d} \sigma_i y_i \right],$$

which means

$$\eta = \sup_{y \in \mathbb{R}^d_+} \left\{ \sum_{i=1}^{d} \sigma_i y_i \,\middle|\, I + \sum_{j \neq i} y_j R_{ji} - y_i R_{ii} \succeq 0,\ i = 1, 2, \dots, d \right\}, \tag{10.88}$$

so at the optimum $\overline{y} \in \mathbb{R}^d_+$ of (10.88) the function $L(x, \overline{y})$ is convex in $x = (x_1, \dots, x_d) \in \mathbb{R}^{N_1} \times \dots \times \mathbb{R}^{N_d}$.

For proving (10.86) it suffices to show that the conditions of Proposition 10.14 are fulfilled, i.e., there exist $\overline{y} \in \mathbb{R}^d_+$ and $(\overline{x}_1, \dots, \overline{x}_d) \in \arg \min_{x_i \in \mathbb{R}^{N_i}, i=1,\dots,d} L(x, \overline{y})$ such that $(\overline{x}_1, \dots, \overline{x}_d)$ is feasible to (10.84) and moreover

$$\overline{y}_i \left(\overline{x}_i^T R_{ii} \overline{x}_i - \sum_{j \neq i} \overline{x}_j^T R_{ij} \overline{x}_j - \sigma_i \right) = 0,\ i = 1, 2, \dots, d. \tag{10.89}$$

Note that an optimal solution $\overline{y} \in \mathbb{R}^d_+$ of (10.88) must satisfy

$$I + \sum_{j \neq i} \overline{y}_j R_{ji} - \overline{y}_i R_{ii} \not\succ 0,\ i = 1, 2, \dots, d. \tag{10.90}$$

Indeed, if for some i,

$$I + \sum_{j \neq i} \overline{y}_j R_{ji} - \overline{y}_i R_{ii} \succ 0$$

we can take $\tilde{y}_i > \overline{y}_i$ such that

$$I + \sum_{j \neq i} \overline{y}_j R_{ji} - \overline{y}_i R_{ii} \succeq 0.$$

Then setting

$$\tilde{y}_j = \begin{cases} \overline{y}_j & j \neq i \\ \tilde{y}_i & j = i \end{cases}$$

yields a feasible solution $\tilde{y} \in \mathbb{R}^d_+$ to (10.88) satisfying $\sum_{i=1}^{d} \sigma_i \overline{y}_i < \sum_{i=1}^{d} \sigma_i \tilde{y}_i$, so $\overline{y}$ cannot be an optimal solution of (10.88).

Thus, at optimality each matrix

$$I + \sum_{j \neq i} \overline{y}_j R_{ji} - \overline{y}_i R_{ii}$$

must be singular. We have $\overline{y}_i > 0, i = 1, 2, \ldots, d$ and there are $\tilde{x}_i \in \mathbb{R}^{N_i}$, $||\tilde{x}_i|| = 1$, $i = 1, 2, \ldots, d$ such that

$$\tilde{x}_i^T \left(I + \sum_{j \neq i} \overline{y}_j R_{ji} - \overline{y}_i R_{ii} \right) \tilde{x}_i = 0, i = 1, 2, \ldots, d, \tag{10.91}$$

hence, by setting $R = [r_{ij}]_{i,j=1,2,\ldots,d}$ with $r_{ii} := \tilde{x}_i^T R_{ii} \tilde{x}_i$ and $r_{ij} := -\tilde{x}_j^T R_{ij} \tilde{x}_j$, $i \neq j$,

$$R^T \overline{y} = (1, 1, \ldots, 1)^T > 0.$$

By Lemma 10.4, $\mathbb{R}^{-1}$ is positive so $p = \mathbb{R}^{-1}\sigma > 0$, which is the solution of the linear equations

$$p_i(\tilde{x}_i^T R_{ii} \tilde{x}_i) - \sum_{j \neq i} p_j(\tilde{x}_j^T R_{ij} \tilde{x}_j) = \sigma_i, \ \ i = 1, 2, \ldots, d. \tag{10.92}$$

With such p it is obvious that $\{\overline{x}_i := \sqrt{p_i} \tilde{x}_i, \ \ i = 1, 2, \ldots, d\}$ is feasible to (10.84) and satisfies (10.89), completing the proof of (10.86). □

Remark 10.10 The main result in Bengtsson and Ottersen (2001) states that the convex relaxation of (10.84), i.e., the convex program

$$\min_{0 \preceq X_i \in \mathbb{R}^{N_i \times N_i}, \ i=1,2,\ldots,d} \sum_{i=1}^{d} \mathrm{Trace}(X_i) : \ \mathrm{Trace}(X_i R_{ii})$$
$$\geq \sum_{j \neq i} \mathrm{Trace}(X_j R_{ij}) + \sigma_i, \ \ i = 1, 2, \ldots, d, \tag{10.93}$$

attains its optimal value at a rank-one feasible solution $X_i = x_i x_i^T$, so (10.84) and (10.93) are equivalent. It can be shown that the Lagrangian dual of the convex problem (10.93) is precisely $\max_{y\in\mathbb{R}^d_+}\min_{x_i\in\mathbb{R}^{N_i}} L(x,y)$. In other words the above result of Bengtsson-Ottersen is just equivalent to Corollary 10.12 which, besides, has been obtained here by a more straightforward and rigorous argument.

Also note that the computation of the optimal solution of the nonconvex program (10.84), which is based on the optimal solution $\overline{y}$ of the convex Lagragian dual (10.88) and $\sqrt{p_i}\tilde{x}_i,\ i = 1,2,\dots,d$ by (10.91) and (10.92), is much more efficient than by using the optimal solutions of the convex program (10.93). This is because the convex program (10.93) has multiple optimal solutions, so its rank-one optimal solution is not quite easily computed.

We have shown above several examples where strong duality holds, so that the optimal value of the dual equals exactly the optimal value of the original problem. In the general case, a gap exists which in a sense reflects the extent to which the given problem fails to be convex and regular. Quite often, however, the gap can be reduced by gradually restricting the set X. More specifically, it may happen that:

(*) *Whenever a nested sequence of rectangles* $M_1 \supset M_2 \supset \dots$ *collapses to a singleton:* $\cap_{k=1}^{\infty} M_k = \{\overline{x}\}$ *then* $f^{\natural}_{M_k} - \psi^*_{M_k} \to 0$.

Under these circumstances, a BB algorithm for (10.80) can be developed in the standard way, such that:

- Branching follows a selected exhaustive rectangular subdivision rule.
- Bounding is by Lagrangian relaxation: for each rectangle M the value ψ^*_M is taken to be a lower bound for $f^{\natural}_M$.
- The candidate for further subdivision at iteration k is $M_k \in \operatorname{argmin}\{\psi^*_M|\ M \in \mathscr{R}_k\}$ where $\mathscr{R}_k$ denotes the collection of all partition sets currently still of interest.
- Termination criterion: At each iteration k a new feasible solution is computed and the incumbent x^k is set equal to the best among all feasible solutions so far obtained. If $f(x^k) = \psi^*_{M_k}$, then terminate (x^k is an optimal solution).

Theorem 10.5 *Under assumption (*) the above BB algorithm converges. In other words, whenever infinite, the algorithm generates a sequence* $\{M_k\}$ *such that* $\psi^*_{M_k} \nearrow f^{\natural} := f^{\natural}_X$.

Proof By the selection of M_k we have $\psi^*_{M_k} \le f^{\natural}$ for all k. By exhaustiveness of the subdivision, if the procedure is infinite it generates at least one filter $\{M_{k_\nu}\}$ collapsing to a singleton $\{\overline{x}\}$. Since $\psi^*_{M_k} \le f^{\natural} \le f^{\natural}_{M_k}$ and since obviously $\psi^*_{M_{k+1}} \ge \psi^*_{M_k}$ for all k, it follows from Assumption (*) that $\psi^*_{M_k} \nearrow f^{\natural}$. □

10.5.3 Application

a. Bounds by Lagrange relaxation

Following Ben-Tal et al. (1994) let us apply the above method to the problem:

$$(PX) \qquad f^{\natural}_X = \min\{c^T y|\ A(x)y \le b, y \ge 0, x \in X\}, \tag{10.94}$$

where X is a rectangle in $\mathbb{R}^n$, and $A(x)$ is an $m \times p$ matrix whose elements $a_{ij}(x)$ are continuous functions of x. Clearly this problem includes as a special case the problem of minimizing a linear function under bilinear constraints. Since for fixed x the problem is linear in y we form the dual problem with respect to the constraints $A(x)y \leq b$, i.e.,

$$\max_{u \geq 0} \min\{\langle c, y\rangle + \langle u, A(x)y - b\rangle |\, y \geq 0, x \in X\}. \tag{10.95}$$

Lemma 10.5 *The problem (10.95) can be written as*

$$(DPX) \qquad \max_{u \geq 0}\{-\langle b, u\rangle |\, \langle A(x), u\rangle + c \geq 0 \ \forall x \in X\}. \tag{10.96}$$

Proof For each fixed $u \geq 0$ we have

$$\begin{aligned} &\min\{\langle c, y\rangle + \langle u, A(x)y - b\rangle |\, x \in X,\ y \geq 0\} \\ &= \min_{x \in X} \min_{y \geq 0}\{\langle c, y\rangle + \langle u, A(x)y - b\rangle\} \\ &= -\langle b, u\rangle + \min_{x \in X}\{\langle c, y\rangle + \langle u, A(x)y\rangle |\, y \geq 0\} \\ &= -\langle b, u\rangle + h(u), \end{aligned}$$

where

$$h(u) = \begin{cases} 0 & \text{if } \langle A(x), u\rangle + c \geq 0 \ \forall x \in X \\ -\infty & \text{otherwise.} \end{cases}$$

The last equality is due to the fact that the infimum of a linear function of y over the orthant $y \geq 0$ is 0 or $-\infty$ depending on whether the coefficients of this linear function are all nonnegative or not. The conclusion follows. □

Denote the column j of $A(x)$ by $A_j(x)$ and for any given rectangle $M \subset X$ define

$$g_M(u) = \min_{j=1,\ldots,p} \min_{x \in M}[\langle A_j(x), u\rangle + c_j].$$

Since for each fixed x the function $u \mapsto \langle A_j(x), u\rangle + c_j$ is affine, $g_M(u)$ is a concave function and the optimal value of the dual (DPM) is

$$\psi^*_M = \max\{-\langle b, u\rangle |\, g_M(u) \geq 0,\ u \geq 0\}.$$

Consider now a filter $\{M_k\}$ of subrectangles of X such that $\cap_{k=1}^{\infty} M_k = \{\overline{x}\}$. For brevity write $f^\natural_k, \psi^*_k, \psi^*_{\overline{x}}$ for $f^\natural_{M_k}, \psi^*_{M_k}, \psi^*_{\{\overline{x}\}}$, respectively. We have

$$f^\natural_k \leq \min\{\langle c, y\rangle |\, A(\overline{x})y \leq b, y \geq 0\} \text{ (since } \overline{x} \in M_k \ \forall k)$$

$$\min\{\langle c, y\rangle |\, A(\overline{x})y \leq b, y \geq 0\} = \max\{-\langle b, u\rangle |\, \langle A(\overline{x}), u\rangle + c \geq 0, u \geq 0\}$$

(dual linear programs)

$$\psi^*_{\overline{x}} = \max\{-\langle b, u\rangle |\, \langle A(\overline{x}), u\rangle + c \geq 0, u \geq 0\} \text{ (definition of } \psi^*_{\{\overline{x}\}}),$$

hence,

$$f^\natural_k \leq \psi^*_{\overline{x}}. \tag{10.97}$$

Also, setting

$$L = \{u \in \mathbb{R}^m_+ |\ \langle A(\overline{x}), u\rangle + c \geq 0\},$$
$$L_k = \{u \in \mathbb{R}^m_+ |\ \langle A(x), u\rangle + c \geq 0\ \forall x \in M_k\}$$

we can write

$$\psi_k^* = \max\{-\langle b, u\rangle | u \in L_k\}, \quad \psi_{\overline{x}}^* = \max\{-\langle b, u\rangle |\ u \in L\}$$
$$L_1 \subset \ldots \subset L_k \subset \ldots \subset L, \quad \psi_1^* \leq \ldots \leq \psi_k^* \leq \ldots \leq \psi_{\overline{x}}^*.$$

Proposition 10.15 *Assume*

$$\textbf{B1}. \qquad (\forall x \in X) \quad (\exists u \in \mathbb{R}^m_+) \quad \langle A(x), u\rangle + c > 0.$$

If $\{M_k\}$ *is any filter of subrectangles of* X *such that* $\cap_{k=1}^{\infty} M_k = \{\overline{x}\}$*, then*

$$f_k^{\natural} - \psi_k^* \to 0.$$

Proof For all k we have, by (10.97), $\psi_k^* \leq f^{\natural} \leq f_k^{\natural} \leq \psi_{\overline{x}}^*$, so it suffices to show that

$$\psi_k^* \nearrow \psi_{\overline{x}}^* \quad (k \to +\infty).$$

Denote $\sigma(x, u) = \langle A(x), u\rangle + c \in \mathbb{R}^m$. Assumption B1 ensures that int $L \neq \emptyset$ and $\sigma(\overline{x}, u)$ is never identical to zero. Therefore, if $u \in \mathrm{int}L$ then $\sigma(\overline{x}, u) > 0$ (i.e., $\sigma_i(\overline{x}, u) > 0\ \forall i = 1, \ldots, m$) and since $\sigma(x, u)$ is continuous in x one must have $\sigma(x, u) > 0$ for all x in a sufficiently small ball around $\overline{x}$, hence for all $x \in M_k$ with sufficiently large k (since $\max_{x \in M_k} \|x - \overline{x}\| \to 0$ as $k \to +\infty$). Thus, if $u \in \mathrm{int}L$ then $u \in L_k$ for all sufficiently large k. Keeping this in mind let $\varepsilon > 0$ be an arbitrary positive number and $\overline{u}$ an arbitrary element of L. Since L is a polyhedron with nonempty interior, there exists a sequence $\{u^t\} \subset \mathrm{int}L$ such that $\overline{u} = \lim_{t \to +\infty} u^t$. Hence, there exists $u^t \in \mathrm{int}L$ such that $\langle b, u^t\rangle - \langle b, \overline{u}\rangle \leq \varepsilon$. Since $u^t \in \mathrm{int}L$, for all sufficiently large k we have, by the above, $u^t \in L_k$ and, consequently, $\langle b, u^t\rangle \geq -\psi_k^*$. Thus $-\psi_k^* \leq \langle b, \overline{u}\rangle + \varepsilon$ for all sufficiently large k. Since $\varepsilon > 0$ and $\overline{u} \in L$ are arbitrary, this shows that $\psi_k^* \to \psi_{\overline{x}}^*$, completing the proof. □

Thus, under Assumption B1, Condition (*) is fulfilled. It then follows from Theorem 10.5 that Problem (*PX*) can be solved by a convergent BB procedure using the bounds by Lagrangian relaxation and an exhaustive rectangular subdivision process.

For the implementation of this BB algorithm note that the dual problem (*DPX*) is a semi-infinite linear program which in the general case may not be easy to solve. However, if we assume, additionally, that for $u \geq 0$:

B2. *Every function* $\langle A_j(x), u\rangle + c_j,\ j = 1, \ldots p,$ *is quasiconcave*

then the condition $\langle A_j(x), u\rangle + c_j \geq 0\ \ \forall x \in M$ is equivalent to $\langle A_j(v), u\rangle + c_j \geq 0\ \ \forall v \in V(M)$, where $V(M)$ denotes the vertex set of the rectangle M. Therefore, under Assumption B2, (DPM) merely reduces to the linear program

$$\max_{u\geq 0}\{-\langle b, u\rangle |\ \langle A_j(v), u\rangle + c_j \geq 0 \quad v \in V(M), j = 1, \ldots, p\}. \tag{10.98}$$

If $M = [r, s]$ then the set $V(M)$ consists of all points $v \in \mathbb{R}^n$ such that $v_j \in \{r_j, s_j\},\ j = 1, \ldots, n$. Thus,under Assumptions B1 and B2 the problem (PX) can be solved by the following algorithm:

BB Algorithm for (PX)

Initialization. Let $(\bar{x}^1, \bar{y}^1)$ be an initial feasible solution and $\gamma_1 = \langle c, \bar{y}^1\rangle$ (found, e.g., by local search). Set $M_0 = X, \mathscr{S}_1 = \mathscr{P}_1 = \{M_0\}$. Set $k = 1$.

Step 1. For each rectangle $M \in \mathscr{P}_k$ solve

$$\max\{-\langle b, u\rangle |\ \langle A_j(v), u\rangle + c_j \geq 0\ v \in V(M),\ j = 1, \ldots, p,\ u \geq 0\}$$

to obtain ψ^*_M.

Step 2. Delete every rectangle $M \in \mathscr{S}_k$ such that $\psi^*_M \geq \gamma_k - \varepsilon$. Let $\mathscr{R}_k$ be the collection of remaining rectangles.

Step 3. If $\mathscr{R}_k = \emptyset$ then terminate: $(\bar{x}^k, \bar{y}^k)$ is an ε-optimal solution.

Step 4. Let $M_k \in \operatorname{argmin}\{\psi^*_M |\ M \in \mathscr{R}_k\}$. Compute a new feasible solution by local search from the center of M_k (or by solving a linear program, see Remark 10.11 below). Let $(\bar{x}^{k+1}, \bar{y}^{k+1})$ be the new incumbent, $\gamma_{k+1} = \langle c, \bar{y}^{k+1}\rangle$.

Step 5. Bisect M_k upon its longest side. Let $\mathscr{P}_{k+1}$ be the partition of M_k.

Step 6. Set $\mathscr{S}_{k+1} = (\mathscr{R}_k \setminus \{M_k\}) \cup \mathscr{P}_{k+1}, k \leftarrow k + 1$ and return to Step 1.

Proposition 10.16 *The above algorithm can be infinite only if* $\varepsilon = 0$. *In the latter case,* $\psi^*_{M_k} \nearrow \min(PX)$, *and any filter of rectangles* $\{M_{k_\nu}\}$ *determines an optimal solution* $(\bar{x}, \bar{y})$, *such that* $\{\bar{x}\} = \cap_{\nu=1}^{\infty} M_{k_\nu}$ *and* $\bar{y}$ *is an optimal solution of* $P_{\{\bar{x}\}}$.

Proof Immediate from Theorem 10.5 and Proposition 10.15. □

Remark 10.11 (i) If x^k is any point of M_k then an optimal solution y^k of the linear program

$$\min\{\langle c, y\rangle |\ A(x^k)y \leq b, y \geq 0\}$$

yields a feasible solution (x^k, y^k).

(ii) For definiteness we assumed that a rectangular subdivision is used but the above development is also valid for simplicial subdivisions (replace everywhere "rectangle" by "simplex"). For many problems, simplicial subdivision is even more efficient because in that case $|V(M)| = n+1$ instead of $|V(M)| = 2^n$ as above, so that (10.98) will involve much less constraints.

Example 10.7 Consider the Pooling and Blending Problem formulated in Example 4.4. Setting $x_{il} = q_{il}\sum_j y_{lj}$ and $f_{lk} = \sum_i C_{ik}q_{il}$, it can be reformulated as

$$\begin{array}{l}
\min -\sum_j \gamma_j \quad \text{s.t.} \\
\sum_l (d_j - \sum_i c_i q_{il})y_{lj} + \sum_i (d_i - c_i)z_{ij} = \gamma_j \\
\sum_l \sum_j q_{il}y_{lj} + \sum_j z_{ij} \le A_i, \quad \sum_j y_{lj} \le S_l, \quad \sum_l y_{lj} + \sum_i z_{ij} \le D_j \\
\sum_l (\sum_i C_{ik}q_{il} - P_{jk})y_{lj} + \sum_i (C_{ik} - P_{jk})z_{ij} \le 0 \\
q_{il} \ge 0,\ \sum_i q_{il} = 1,\ y_{lj} \ge 0,\ z_{ij} \ge 0.
\end{array}$$

In this form it appears as a problem of type (PX) since for fixed q it is linear in (y, z, γ), while the set $X = \{q|\ q_{il} \ge 0,\ \sum_i q_{il} = 1\}$ is a simplex. It is readily verified that conditions B1 and B2 are satisfied, so the above method applies.

b. Bounds via Semidefinite Programming

Lagrange relaxation has also been used to derive good and cheaply computable lower bounds for the optimal value of problem (10.80) where X is a polyhedron and

$$f_0(x) = \frac{1}{2}\langle x, Q^0x\rangle + \langle c^0, x\rangle + d_0,\ f_i(x) = \frac{1}{2}\langle x, Q^ix\rangle + \langle c^i, x\rangle + d_i.$$

The Lagrangian is

$$L(x, u) = \frac{1}{2}\langle x, Q(u)x\rangle + \langle c(u), x\rangle + d(u)$$

with

$$Q(u) = Q^0 + \sum_{i=1}^m u_iQ^i,\ c(u) = c^0 + \sum_{i=1}^m u_ic^i,\ d(u) = d_0 + \sum_{i=1}^m u_id_i.$$

Define

$$D = \{u \in \mathbb{R}^m_+ |\ Q(u) \succ 0\}, \quad \overline{D} = \{u \in \mathbb{R}^m_+ |\ Q(u) \succeq 0\},$$

where for a matrix A the notation $A \succ 0$ ($A \succeq 0$) means that A is positive definite (positive semidefinite, resp.). Obviously

$$Q(u) \succ 0 \Leftrightarrow \min\{\langle x, Q(u)x\rangle |\ \|x\| = 1\} > 0$$

$$Q(u) \succeq 0 \Leftrightarrow \min\{\langle x, Q(u)x\rangle |\ \|x\| = 1\} \ge 0$$

so $D, \overline{D}$ are convex sets (when nonempty) and $\overline{D}$ is the closure of D.

If $u \in D$ then finding the value

$$\psi(u) = \inf_{x\in X}\{L(x, u)\}$$

is a convex minimization problem which can be solved by efficient algorithms. But if $u \notin \overline{D}$, then computing $\psi(u)$ is a difficult global optimization problem. Therefore, instead of $\psi^* = \sup_{u\in\mathbb{R}^m_+} \psi(u)$ Shor (1991) proposes to consider the weaker estimate

$$\overline{\psi}^* = \sup_{u\in\overline{D}} \psi(u) \le \psi^*.$$

Note that in the special case $X = \mathbb{R}^n$, we have $\overline{\psi}^* = \psi^*$ because then $u \notin \overline{D}$ implies $\psi(u) = -\infty$. For simplicity below we assume $X = \mathbb{R}^n$.

Proposition 10.17 *If $\psi^* = \sup_{u\in\overline{D}} \psi(u)$ is attained on D then $\psi^* = f^\natural$.*

Proof Let $\psi(\overline{u}) = L(\overline{x}, \overline{u}) = \sup_{u\in\overline{D}} \psi(u)$. By Proposition 10.11 the vector $-(f_i(\overline{x}), i = 1, \dots, m)$ is a subgradient of $-\psi(u)$ at $\overline{u}$. If the concave function $\psi(u)$ attains a maximum at a point $\overline{u} \in D$ then since $\overline{u}$ is an interior point of D one must have $-f_i(\overline{x}) \ge 0,\ i = 1, \dots, m$. That is, $\overline{x}$ is feasible and since $\psi^* = L(\overline{x}, \overline{u}) \le f^\natural$ one must have $\psi^* = f^\natural$. □

Thus, to compute a lower bound, one way is to solve the problem

$$\sup\{\psi(u)|\ Q(u) \succeq 0,\ u = (u_1, \dots, u_m) \ge 0\}, \tag{10.99}$$

where $\psi(u)$ is the concave function

$$\psi(u) = \inf_{x\in\mathbb{R}^n} \frac{1}{2}\langle x, Q(u)x\rangle + \langle c(u), x\rangle + d(u).$$

Remark 10.12 The problem (10.99) is a concave maximization (i.e., convex minimization) under convex constraints. It can also be written as

$$\max\ t \quad \text{s.t.} \quad \psi(u) - t \ge 0,\ Q(u) \succeq 0,\ u = (u_1, \dots, u_m) \ge 0.$$

Since $\psi(u) - t \ge 0$ if and only if

$$\frac{1}{2}\langle x, Q(u)x\rangle + \langle c(u), x\rangle + d(u) - t \ge 0 \quad \forall x \in \mathbb{R}^n$$

it is easy to check that (10.99) is equivalent to

$$\max\left\{t \mid \begin{bmatrix} Q(u) & c(u) \\ (c(u))^T & 2(d(u) - t)\end{bmatrix} \succeq 0, \quad u = (u_1, \dots, u_m) \ge 0\right\}.$$

Thus (10.99) is actually a linear program under LMI constraints . For this problem several efficient algorithms are currently available (see, e.g., Vandenberge and Boyd (1996) and the references therein). However, it is not known whether the bounds computed that way are consistent with a rectangular or simplicial subdivision process and can be used to produce convergent BB algorithms.

We will have to get back to quadratic optimization under quadratic constraints in Chap. 11 on polynomial optimization. There a SIT Algorithm for this class of problems will be presented which is the most suitable when a feasible solution is available and we want to check whether it is already optimal and if not, to find a better feasible solution.

10.6 Exercises

1 Let $D = \{x \in \mathbb{R}^n |\ \langle a^i, x\rangle \le \alpha_i \quad i = 1,\dots,m\}$ be a polytope.

1. Show that there exist $\theta_{ij} \ge 0$ such that $\langle x, x\rangle = -\sum_{i,j=1}^m \theta_{ij}\langle a^i, x\rangle\langle a^j, x\rangle$. (Hint: use (2) of Exercise 11, Chap. 1, to show that the convex hull of $a^1,\dots,a^m$ contains 0 in its interior; then for every $i = 1,\dots,n$, the ith unit vector e^i is a nonnegative combination of $a^1,\dots,a^m$).
2. Deduce that for any quadratic function $f(x)$ there exists ρ such that for all $r > \rho$ the quadratic function $f(x) + r\sum_{i,j=1}^m \theta_{ij}(\langle a^i, x\rangle - \alpha_i)(\langle a^j, x\rangle - \alpha_j)$ is convex, where θ_{ij} are defined as in (1).

2 Consider the problem $\min\{f(x)|\ x \in X\}$ where X is a polyhedron, and $f(x)$ is an indefinite quadratic function. Denote the optimal value of this problem by $f_X^\natural$. Given any rectangle $M = [p, q]$, assume that $X \cap M = \{x \in \mathbb{R}^n |\ \langle a^i, x\rangle \le \alpha_i \quad i = 1,\dots,m\}$. Show that for sufficiently large r the function $f(x) + r\sum_{i,j=1}^m \theta_{ij}(\langle a^i, x\rangle - \alpha_i)(\langle a^j, x\rangle - \alpha_j)$ is a convex minorant of $f(x)$ over M and a lower bound for $f_M^\natural = \min\{f(x)|\ x \in X \cap M\}$ is given by the unconstrained minimum of the convex function $f(x) + r\sum_{i,j=1}^m \theta_{ij}(\langle a^i, x\rangle - \alpha_i)(\langle a^j, x\rangle - \alpha_j)$ over $\mathbb{R}^n$. (Hint: see Exercise 1).

3 Solve the indefinite quadratic program:

$$\begin{aligned} \min\ & 3(x - 0.5)^2 - 2y^2 \\ \text{s.t.}\ & -x + 2y \le 1,\ 0.5 \le x \le 1,\ 0.75 \le y \le 2. \end{aligned}$$

4 Solve the generalized linear program

$$\begin{aligned} \min\ & 3x_1 + 4x_2 \\ \text{s.t.}\ & y_{11}x_1 + 2x_2 + x_3 = 5, \quad y_{12}x_1 + x_2 + x_4 = 3 \\ & x_{\ge}0,\ x_2 \ge 0; \quad y_{11} - y_{12} \le 1; \quad 0 \le y_{11} \le 2,\ 0 \le y_{12} \le 2. \end{aligned}$$

5 Solve the problem

$$\min\{-0.25x_1 - 0.5x_2 + x_1x_2 |\ x_1 \ge 0.5,\ 0 \le x_2 \le 0.5\}.$$

6 Solve the problem

$$\begin{aligned} \min\{-x_1 + x_1x_2 - x_2 |\ & -6x_1 + 8x_2 \le 3;\ 3x_1 - x_2 \le 3; \\ & 0 \le x_1 \le 5;\ 0 \le x_2 \le 5\}. \end{aligned}$$

7 Solve the problem

$$\min\ [x_1, x_2]\begin{bmatrix} 1 & 2 \\ -1 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + [1, 1]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \text{s.t.}$$

$$[x_1, x_2]\begin{bmatrix} -1 & 1 \\ 3 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leq 6,\ x_1 + x_2 \leq 4,\ 0 \leq x_1 \leq 4,\ 0 \leq x_2 \leq 4.$$

8 Consider a quadratic program $\min\{f(x)|\ x \in D\}$, where D is a polytope in $\mathbb{R}^n, f(x) = \frac{1}{2}\langle x, Qx\rangle + \langle c, x\rangle$ and Q is a symmetric matrix of rank $k \ll n$. Show that there exist k linearly independent vectors $a^1, \ldots, a^k$ such that for fixed values of $t_1, \ldots, t_k$ the problem $\min\{f(x)|\ x \in D, \langle a^i, x\rangle = t_i,\ i = 1, \ldots\}$ becomes a linear program. Develop a decomposition method for solving this problem (see Chap. 7).

9 Let $G = G(V, E)$ be an undirected graph, where $V = \{1, 2, \ldots, n\}$ denotes the set of vertices, E the set of edges. A subset $V' \subset V$ is said to be *stable* if no two distinct members of V' are adjacent, i.e., if $(i, j) \notin E$ for every $i, j \in V'$ such that $i \neq j$. Let $\alpha(G)$ be the cardinality of the largest stable set in G. Show that $\alpha(G)$ is the optimal value in the quadratic problem;

$$\max_x \left\{\sum_{i=1}^n x_i \mid x_i x_j = 0\ \forall (i, j) \in E),\ x_i(x_i - 1) = 0\ \forall i = 1, \ldots, n)\right\}.$$

1. Using the Lagrangian relaxation method show that an upper bound for $\alpha(G)$ is given by

$$\psi^* = \inf\{\psi(\lambda)|\ Q(\lambda) > 0\}, \quad \psi(\lambda) = \max_x \left[2\sum_{i=1}^n x_i - \langle x, Q(\lambda)x\rangle\right],$$

   where $Q(\lambda) = [Q_{ij}]$ is a symmetric matrix such that $Q_{ij} = 0$ if $i \neq j, (i, j) \notin E$ and $Q_{ij} = \lambda_{ij}$ for every other (i, j).
2. Show that $\psi(\lambda) = \varphi(\lambda)$ where

$$\varphi(\lambda) = \left\{\min_x [\langle x, Q(\lambda)x\rangle|\ \sum_{i=1}^n x_i = 1]\right\}^{-1}$$

   (Shor and Stetsenko 1989). Hint: show that for any positive definite matrix A and vector $b \in \mathbb{R}^n$: $\max_x[2\langle b, x\rangle - \langle x, Ax\rangle] = \{\min_x[\langle x, Ax\rangle|\ \langle b, x\rangle = 1]\}^{-1}$.

10 Let A_0, A_1, A_2, B be symmetric matrices of order n, and a, b two positive numbers such that $a < b$. Consider the problem

$$\gamma^* := \inf\{\gamma|\ A_0 + uA_1 + vA_2 + \gamma B < 0,$$
$$u = \gamma x,\ \langle x, v\rangle \leq \gamma,\ a \leq x \leq b,\ \gamma > 0\}.$$

Show that $\dfrac{4ab}{(a+b)^2}\alpha \le \gamma^* \le \alpha$, where α is the optimal value of the problem

$$\begin{aligned}\min\{\gamma|\ & zA_0 + uA_1 + vA_2 + \gamma B < 0\\ & 4u + (a+b)^2 v \le 4(a+b)\gamma,\ \gamma a \le u \le \gamma b, \gamma \ge 0\}.\end{aligned}$$

Hint: Observe that in the plane (x, v) the line segment $\{(x, v)|\ 4\gamma x + (a+b)^2 v = 4(p+q)\gamma,\ a \le x \le b\}$ is tangent to the arc $\{(x, v)|\ \langle x, v\rangle = \gamma,\ a \le x \le b\}$ at point $(\frac{a+b}{2}, \frac{2\gamma}{a+b})$, and is the chord of the arc $\{(x, v)|\ \langle x, v\rangle = \frac{4ab}{(a+b)^2}\gamma,\ a \le x \le b\}$.

11 Show that the water distribution network problem formulated in Exercise 8, Chap. 4, is of type (PX) and can be solved by a BB method using Lagrange relaxation for lower bounding.

12 Show that the objective function $f(q, H, J)$ in the water distribution network problem (Exercise 8, Chap. 4) is convex in J and concave in (q, H). Given any rectangle $M = \{(q, H)|\ \underline{q}^M \le q \le \overline{q}^M,\ \underline{H}^M \le H \le \overline{H}^M\}$ we have

$$\begin{aligned} f(\underline{q}^M, \underline{H}^M, J) &\qquad\qquad \le f(q, H, J)\\ KL_i(\underline{q}_i^M/c)^\lambda \le KL_i(q_i/c)^\lambda &\le KL_i(\overline{q}_i^M/c)^\lambda,\ i = 1, \dots, s\end{aligned}$$

for all $(q, H) \in M$. Therefore, a lower bound for the objective function value over all feasible solutions (q, H, J) with $(q, H) \in M$ can be computed by solving a convex program. Derive a BB method for solving the problem based on this lower bounding.

Chapter 11
Monotonic Optimization

By their very nature, many nonconvex functions encountered in mathematical modeling of real world systems in a broad range of activities, including economics and engineering, exhibit monotonicity with respect to some variables (partial monotonicity) or to all variables (total monotonicity). In this chapter we will present a mathematical framework for solving optimization problems that involve increasing functions, and more generally, functions representable as differences of increasing functions. Monotonic optimization was first considered in Rubinov et al. (2001) and extensively developed in the two main papers Tuy (2000a) and Tuy et al. (2006).

11.1 Basic Concepts

11.1.1 Increasing Functions, Normal Sets, and Polyblocks

For any two vectors $x, y \in \mathbb{R}^n$ we write $x \leq y$ ($x < y$, resp.) and say that y *dominates* x (y *strictly dominates* x, resp.) if $x_i \leq y_i$ ($x_i < y_i$ resp.) for all $i = 1, \ldots, n$. If $a \leq b$ then the box $[a, b]$ is the set of all x such that $a \leq x \leq b$. We also write $(a, b] := \{x|\ a < x \leq b\}$, $[a, b) := \{x|\ a \leq x < b\}$. As usual e is the vector of all ones and e^i the i-th unit vector of the space under consideration, i.e., the vector such that $e^i_i = 1, e^i_j = 0\ \forall j \neq i$.

A function $f : \mathbb{R}^n_+ \to \mathbb{R}$ is said to be *increasing* if $f(x) \leq f(x')$ whenever $0 \leq x \leq x'$; *strictly increasing* if, in addition, $f(x) < f(x')$ whenever $x < x'$. A function f is said to be *decreasing* if $-f$ is increasing. Many functions encountered in various applications are increasing in this sense. Outstanding examples are the production functions and the utility functions in mathematical economics (under the assumptions that all goods are useful), polynomials (in particular quadratic functions) with nonnegative coefficients, and posynomials

H. Tuy, *Convex Analysis and Global Optimization*, Springer Optimization and Its Applications 110, DOI 10.1007/978-3-319-31484-6_11

$$\sum_{j=1}^{m} c_j \prod_{i=1}^{n} x_i^{a_{ij}} \quad \text{with } c_j \geq 0 \text{ and } a_{ij} \geq 0.$$

(in particular, Cobb–Douglas functions $f(x) = \prod_i x_i^{a_i},\ a_i \geq 0$).

The following obvious proposition shows that the class of increasing functions includes a large variety of functions:

Proposition 11.1 (i) *If f_1, f_2 are increasing functions, then for any nonnegative numbers λ_1, λ_2 the function $\lambda_1 f_1 + \lambda_2 f_2$ is increasing.*

(ii) *The pointwise supremum of a bounded above family $(f_\alpha)_{\alpha \in A}$ of increasing functions and the pointwise infimum of a bounded below family $(f_\alpha)_{\alpha \in A}$ of increasing functions are increasing.*

Nontrivial examples of increasing functions given by this proposition are functions of the form $f(x) = \sup_{y \in a(x)} g(y)$, where $g : \mathbb{R}^n_+ \to \mathbb{R}$ is an arbitrary function and $a : \mathbb{R}^n_+ \to \mathbb{R}^n_+$ is a set-valued map with bounded images such that $a(x') \subset a(x)$ for $x' \geq x$.

A set $G \subset \mathbb{R}^n_+$ is called *normal* if $[0, x] \subset G$ whenever $x \in G$; in other words, if $0 \leq x' \leq x \in G \Rightarrow x' \in G$. The empty set, the singleton $\{0\}$, and $\mathbb{R}^n_+$ are special normal sets which we shall refer to as trivial normal sets in $\mathbb{R}^n_+$. If G is a normal set then $G \cup \{x \in \mathbb{R}^n_+ |\ x_i = 0\}$ for some $i = 1, \ldots, n$ is still normal.

Proposition 11.2 *For any increasing function $g(x)$ on $\mathbb{R}^n_+$ and any $\alpha \in \mathbb{R}$ the set $G = \{x \in \mathbb{R}^n_+ |\ g(x) \leq \alpha\}$ is normal and it is closed if $g(x)$ is lower semi-continuous. Conversely, for any closed normal set $G \subset \mathbb{R}^n_+$ with nonempty interior there exists a lower semi-continuous increasing function $g : \mathbb{R}^n_+ \to \mathbb{R}$ and $\alpha \in \mathbb{R}$ such that $G = \{x \in \mathbb{R}^n_+ |\ g(x) \leq \alpha\}$.*

Proof We need only prove the second assertion. Let G be a closed normal set with nonempty interior. For every $x \in \mathbb{R}^n_+$ define $g(x) = \inf\{\lambda > 0 |\ x \in \lambda G\}$. Since $\text{int}G \neq \emptyset$ there is $u > 0$ such that $[0, u] \subset G$. The halfline $\{\lambda x |\ \lambda \geq 0\}$ intersects $[0, u] \subset G$, hence $0 \leq g(x) < +\infty$. Since for every $\lambda > 0$ the set λG is normal, if $0 \leq x \leq x' \in \lambda G$, then $x \in \lambda G$, too, so $g(x') \geq g(x)$, i.e., $g(x)$ is increasing. We show that $G = \{x \in T^n_+ |\ g(x) \leq 1\}$. In fact, if $x \in G$, then obviously $g(x) \leq 1$. Conversely, if $x \notin G$, then since G is closed there exists $\alpha > 1$ such that $x \notin \alpha G$; hence, since G is normal, $x \notin \lambda G\ \forall \lambda \leq \alpha$, which means that $g(x) \geq \alpha > 1$. Consequently, $G = \{x \in \mathbb{R}^n_+ |\ g(x) \leq 1\}$. It remains to prove that $g(x)$ is lower semi-continuous. Let $\{x^k\} \subset \mathbb{R}^n_+$ be a sequence such that $x^k \to x^0$ and $g(x^k) \leq \alpha\ \forall k$. Then for any given $\alpha' > \alpha$ we have $\inf\{\lambda |\ x^k \in \lambda G\} < \alpha'\ \forall k$, i.e., $x^k \in \alpha' G\ \forall k$; hence $x^0 \in \alpha' G$ in view of the closedness of the set $\alpha' G$. This implies $g(x^0) \leq \alpha'$ and since α' can be taken arbitrarily close to α, we must have $g(x^0) \leq \alpha$, as was to be proved. □

A point $y \in \mathbb{R}^n$ is called an *upper boundary point* of a normal set G if $y \in \text{cl}G$ and there is no point $x \in G$ such that $x = a + \lambda(y - a)$ with $\lambda > 1$. The set of all upper boundary points of G is called its *upper boundary* and is denoted $\partial^+ G$.

A point $y \in \mathrm{cl}G$ is called a *Pareto point* of G if $x \in G, x \geq y \Rightarrow x = y$. Since such a point necessarily belongs to $\partial^+ G$, it is also called an *upper extreme point* of G.

Proposition 11.3 *If $G \subset [a, b]$ is a compact normal set with nonempty interior, then for every point $z \in [a, b] \setminus \{a\}$ the line segment joining a to z meets $\partial^+ G$ at a unique point $\pi_G(z)$ defined by*

$$\pi_G(z) = a + \lambda(z - a), \quad \lambda = \max\{\alpha > 0 |\ a + \alpha(z - a) \in G\}. \tag{11.1}$$

Proof If $u \in \mathrm{int}G$ then $[a, u] \subset G$ and the set $\{\alpha > 0 |\ a + \alpha(z - a) \in G\}$ is nonempty. So the point $\pi_G(z)$ defined by (11.1) exists and is unique. Clearly $\pi_G(z) \in G$ and for $\alpha > \lambda$ we have $a + \lambda(z - a) \notin G$, hence $\pi_G(z)$ is in fact the point where the line segment joining a to z meet $\partial^+ G$. □

A set $H \subset \mathbb{R}^n_+$ is called *conormal* (or sometimes, *reverse normal)* if $x + \mathbb{R}^n_+ \subset H$ whenever $x \in H$. It is called *conormal* in a box $[a, b]$ if $b \geq x' \geq x \geq a,\ x \in H$ implies $x' \in H$, or equivalently, if $x \notin H$ whenever $a \leq x \leq x' \leq b, x' \notin H$. Clearly a set H is conormal if and only if the set $H^\flat := \mathbb{R}^n_+ \setminus H$ is normal.

The following propositions are analogous to Propositions 11.2 and 11.3 and can be proved by similar arguments:

1. For any increasing function $h(x)$ on $\mathbb{R}^n_+$ and any $\alpha \in \mathbb{R}$ the set $H = \{x \in \mathbb{R}^n_+ |\ h(x) \geq \alpha\}$ is normal and is closed if $h(x)$ is upper semi-continuous. Conversely, for any closed conormal set $H \subset \mathbb{R}^n_+$ such that $\mathbb{R}^n_+ \setminus H$ has a nonempty interior there exist an upper semi-continuous increasing function $h : \mathbb{R}^n_+ \to \mathbb{R}$ and $\alpha \in \mathbb{R}$ such that $H = \{x \in \mathbb{R}^n_+ |\ h(x) \geq \alpha\}$.

   A point $y \in [a, b]$ is called a *lower boundary point* of a conormal set $H \subset [a, b]$ if $y \in \mathrm{cl}H$ and there is no point $x \in H$ such that $x = a + \lambda(y - a)$ with $\lambda < 1$. The set of all lower boundary points of H is called its *lower boundary* and is denoted $\partial^- H$. A point $y \in \mathrm{cl}H$ is called a *Pareto point* or a *lower extreme point* of H if $x \in H, x \leq y \Rightarrow x = y$.
2. $H \subset [a, b]$ is a closed conormal set with $b \in \mathrm{int}H$, then for every point $z \in [a, b] \setminus H$ the line segment joining z to b meets $\partial^- H$ at a unique point $\rho_H(z)$ defined by

$$\rho_H(z) = b - \lambda(z - b), \quad \lambda = \max\{\alpha > 0 |\ b - \alpha(z - b) \in H\}. \tag{11.2}$$

   Given a set $A \subset \mathbb{R}^n_+$ the whole orthant $\mathbb{R}^n_+$ is a normal set containing A. The intersection of all normal sets containing A, i.e., the smallest normal set containing A, is called the *normal hull* of A and denoted $A^\rceil$. The *conormal hull* of A, denoted $\lfloor A$, is the smallest conormal set containing A.

Proposition 11.4 (i) *The normal hull of a set $A \subset [a, b]$ is the set $A^\rceil = \cup_{z \in A}[a, z]$. If A is compact then so is $A^\rceil$.*

(ii) *The conormal hull of a set $A \subset [a, b]$ is the set $\lfloor A = \cup_{z \in A}[z, b]$. If A is compact then so is $\lfloor A$.*

Proof It suffices to prove (i), because the proof of (ii) is similar. Let $P = \cup_{z\in A}[a, z]$. Clearly P is normal and $P \supset A$, hence $P \subset A^{\rceil}$. Conversely, if $x \in P$ then $x \in [a, z]$ for some $z \in A \subset A^{\rceil}$, hence $x \in A^{\rceil}$ by normality of $A^{\rceil}$, so $P \subset A^{\rceil}$, proving the equality $P = A^{\rceil}$. If A is compact then for any sequence $\{x^k\} \subset A^{\rceil}$ we have $x^k \in [a, z^k]$ with $z^k \in A$ so $x^k \subset A$, hence, up to a subsequence, $x^k \to x^0 \in A \subset A^{\rceil}$, proving the compactness of $A^{\rceil}$. □

Proposition 11.5 *A compact normal set $G \subset [a, b]$ has at least one upper extreme point and is equal to the normal hull of the set V of its upper extreme points.*

Proof For each $i = 1, \dots, n$ an extreme point of the line segment $\{x \in G|\ x = \lambda e^i, \lambda \geq 0\}$ is an upper extreme point of G. So $V \neq \emptyset$. We show that $G = V^{\rceil}$. Since $V \subset \partial^+ G$, we have $V^{\rceil} \subset (\partial^+ G)^{\rceil} \subset G$, so it suffices to show that $G \subset V^{\rceil}$. If $y \in G$ then $\pi_G(y) \in \partial^+ G$, so $y \in (\partial^+ G)^{\rceil}$. Define $x^1 \in \text{argmax}\{x_1|\ x \in G, x \geq y\}$, and $x^i \in \text{argmax}\{x_i|\ x \geq x^{i-1}\}$ for $i = 2, \dots, n$. Then $v := x^n \in G$ and $v \geq x$ for all $x \in G$ satisfying $x \geq y$. Therefore, $x \in G, x \geq v \Rightarrow x = v$. This means that $y \leq v \in V$, hence $y \in V^{\rceil}$. Consequently, $G \subset V^{\rceil}$, as was to be proved. □

A *polyblock* P is the normal hull of a finite set $V \subset [a, b]$ called its *vertex set*. By the above proposition $P = \cup_{z\in V}[a, z]$. A point $z \in V$ is called a *proper vertex* of P if there is no $z' \in V \setminus \{z\}$ such that $z' \geq z$. It is easily seen that a proper vertex of a polyblock is nothing but a Pareto point of it. The vertex set of a polyblock P is denoted by vert$P$, the set of proper vertices by pvert$P$. A vertex which is not proper is called *improper*. Clearly a polyblock is fully determined by its proper vertex; more precisely, $P = (\text{pvert}P)^{\rceil}$, i.e., a polyblock is the normal hull of its proper vertex set (Fig. 11.1).

Similarly, a *copolyblock* (or *reverse polyblock*) Q is the conormal hull of a finite set $T \subset [a, b]$, i.e., $Q = \cup_{z\in T}[z, b]$. The set T is called the vertex set of the copolyblock. A *proper vertex* of a copolyblock Q is a vertex z for which there is no vertex $z' \neq z$ such that $z' \leq z$. A vertex which is not proper is said to be improper. A copolyblock is fully determined by its proper vertex set, i.e., a copolyblock is the conormal hull of its proper vertex set.

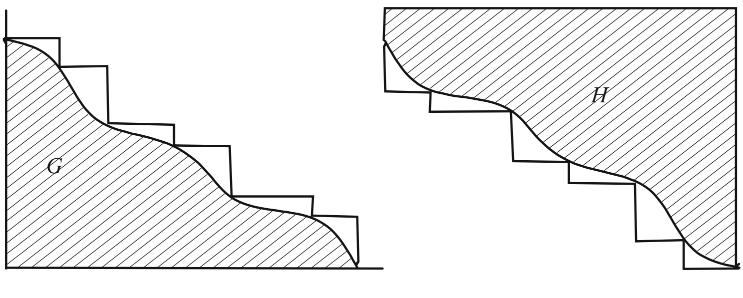


Fig. 11.1 Polyblock, Copolyblock

Proposition 11.6 (i) *The intersection of finitely many polyblocks is a polyblock.* (ii) *The intersection of finitely many copolyblocks is a copolyblock.*

Proof If V_1, V_2 are the vertex sets of two polyblocks P_1, P_2, respectively, then

$$P_1 \cap P_2 = (\cup_{z\in V_1}[a,z]) \cap (\cup_{y\in V_2}[a,y]) = \cup_{(z,y)\in V_1\times V_2}[a,z]\cap[a,y]$$

$$= \cup_{(z,y)\in V_1\times V_2}[a, z\wedge y]$$

where $u = z\wedge y$ means $u_i = \min\{z_i, y_i\}$ $\forall i = 1,\dots,n$. Similarly if T_1, T_2 are the vertex sets of two copolyblocks Q_1, Q_2, respectively, then $Q_1 \cap Q_2 = \cup_{(z,y)\in T_1\times T_2}[z,b]\cap[y,b] = \cup_{(z,y)\in T_1\times T_2}[z\vee y, b]$ where $v = z\vee y$ means $v_i = \max\{z_i, y_i\}$ $\forall i = 1,\dots,n$. □

Proposition 11.7 *If $y \in [a,b)$ then the set $[a,b]\setminus(y,b]$ is a polyblock with proper vertices*

$$u^i = b + (y_i - b_i)e^i,\ i = 1,\dots,n. \tag{11.3}$$

Proof We have $(y,b] = \cup_{i=1}^n\{x\in[a,b] |\, x_i > y_i\}$, so $[a,b]\setminus(y,b] = \cup_{i=1}^n(x \in [a,b] |\, x_i \le y_i\} = \cup_{i=1}^n[a,u^i]$. □

11.1.2 DM Functions and DM Constraints

A function $f : \mathbb{R}^n_+ \to \mathbb{R}$ is said to be a *dm (difference-of-monotonic) function* if it can be represented as a difference of two increasing functions: $f(x) = f_1(x) - f_2(x)$, where f_1, f_2 are increasing functions.

Proposition 11.8 *If $f_1(x), f_2(x)$ are dm then*

(i) *for any $\lambda_1, \lambda_2 \in \mathbb{R}$ the function $\lambda_1 f_1(x) + \lambda_2 f_2(x)$ is dm;*
(ii) *the functions* $\max\{f_1(x), f_2(x)\}$ *and* $\min\{f_1(x), f_2(x)\}$ *are dm.*

Proof (i) is trivial. To prove (ii) let $f_i = f_i(x) - q_i(x), i = 1,2$, with f_i, q_i increasing functions. Since $f_1 = p_1 + q_2 - (q_1+q_2), f_2 = p_2 + q_1 - (q_1+q_2)$ we have $\max\{f_1, f_2\} = \max\{p_1+q_2, p_2+q_1\} - (q_1+q_2)$, $\min\{f_1, f_2\} = \min\{p_1+q_2, p_2+q_1\} - (q_1+q_2)$. □

Proposition 11.9 *The following functions are dm on $[a,b]\subset\mathbb{R}^n_+$:*

(i) *any polynomial, and more generally, any synomial $P(x) = \sum_\alpha c_\alpha x^\alpha$, where $\alpha = (\alpha_1,\dots,\alpha_n) \ge 0$ and $x^\alpha = x_1^{\alpha_1}x_2^{\alpha_2}\cdots x_n^{\alpha_n}$, $c_\alpha \in \mathbb{R}$.*
(ii) *any convex function $f : [a,b]\subset\mathbb{R}^n_+ \to \mathbb{R}$.*
(iii) *any Lipschitz function $f : [a,b] \to \mathbb{R}$ with Lipschitz constant K.*
(iv) *any C_1-function $f : [a,b]\subset\mathbb{R}^n_+ \to \mathbb{R}$.*

Proof (i) $P(x) = P_1(x) - P_2(x)$ where $P_1(x)$ is the sum of terms $c_\alpha x^\alpha$ with $c_\alpha > 0$ and $P_2(x)$ is the sum of terms $c_\alpha x^\alpha$ with $c_\alpha < 0$.

(ii) Since the set $\cup_{a<x<b}\partial f(x)$ is nonempty and bounded (Theorem 2.6), we can take $M > K := \sup\{\|p\|| \ p \in \partial f(x), x \in [a,b]$. Then the function $g(x) = f(x) + M\sum_{i=1}^n x_i$ satisfies, for $a < x \leq x' < b$:

$$g(x') - g(x) = f(x') - f(x) + M\sum_{i=1}^n (x'_i - x_i)$$

$$\geq \langle p, x' - x\rangle + M\sum_{i=1}^n (x'_i - x_i),$$

where $p \in \partial f(x)$. Hence $g(x') - g(x) \geq (M - K)\sum_{i=1}^n (x'_i - x_i) \geq 0$, so $g(x)$ is increasing.

(iii) The function $g(x) = f(x) + M\sum_{i=1}^n x_i$ where $M > K$ is increasing on $[a,b]$ because one has, for $a \leq x \leq x' \leq b$:

$$g(x') - g(x) = M\sum_{i=1}^n (x'_i - x_i) + f(x') - f(x)$$

$$\geq M\sum_{i=1}^n (x'_i - x_i) - K|x' - x| \geq 0.$$

(iv) For $M > K := \sup_{x\in[a,b]} \|\nabla f(x)\|$ the function $g(x) = f(x) + M\sum_{i=1}^n x_i$ is increasing. In fact, by the mean-value theorem, for $a \leq x \leq x' \leq b$ there exists $\theta \in (0,1)$ such that

$$f(x') - f(x) = \langle \nabla f(x + \theta(x' - x)), x' - x\rangle,$$

hence $g(x') - g(x) \geq M\sum_{i=1}^n (x'_i - x_i) - K\sum_{i=1}^n (x'_i - x_i) \geq 0.$ □

As a consequence of Proposition 11.9, (i), the class of dm functions on $[a,b]$ is dense in the space $C_{[a,b]}$ of continuous functions on $[a,b]$ with the supnorm topology.

A function $f(x)$ is said to be *locally increasing* (*locally dm*, resp.) on $\mathbb{R}^n_+$ if for every $x \in \mathbb{R}^n_+$ there is a box $[a,b]$ containing x in its interior such that $f(x)$ is increasing (dm, resp.) on $[a,b]$.

Proposition 11.10 *A locally increasing function on $\mathbb{R}^n_+$ is increasing on $\mathbb{R}^n_+$. A locally dm function on a box $[a,b]$ is dm on this box.*

Proof Suppose $f(x)$ is locally increasing and let $0 \leq x \leq x'$. Since $[x,x']$ is compact there exists a finite covering of it by boxes such that $f(x)$ is increasing on each of these boxes. Then the segment Δ joining x to x' can be divided into a finite number of intervals: $x^1 = x \leq x^2 \leq \cdots \leq x^m = x'$ such that $f(x^i) \leq f(x^{i+1}), i = 1,\ldots,m-1$. Hence, $f(x) \leq f(x^2) \leq \cdots \leq f(x')$.

Suppose now that $f(x)$ is locally dm on $[a,b]$. By compactness the box $[a,b]$ can be covered by a finite number of boxes $[p^1,q^1],\cdots,[p^m,q^m]$ such that $f(x)=u_i(x)-v_i(x)\ \forall x\in[p^i,q^i]$, with u_i,v_i increasing functions on $[p^1,q^i]$. For $x\in[p^i,q^i]\cap[p^j,q^j]$ we have $u_i(x)-v_i(x)=u_j(x)-v_j(x)=f(x)$, so we can define $u(x)=u_i(x)+\sum_{j\neq i}\max\{0,u_j(q^i)-u_i(q^i)\}$, $v(x)=v_i(x)+\sum_{j\neq i}\max\{0,u_j(q^i)-u_i(q^i)\}$ for every $x\in[p^i,q^i]$. These two functions are increasing and obviously $f(x)=u(x)-v(x)$. □

11.1.3 Canonical Monotonic Optimization Problem

A monotonic optimization problem is an optimization problem of the form

$$\max\{f(x)|\ g_i(x)\leq 0,\ i=1,\ldots,m,\ x\in[a,b]\}, \tag{11.4}$$

where the objective function $f(x)$ and the constraint functions $g_i(x)$ are dm functions on the box $[a,b]\subset\mathbb{R}^n$. For the purpose of optimization an important property of the class of dm functions is that, like the class of dc functions, it is closed under linear operations and operations of taking pointwise maximum and pointwise minimum. Exploiting this property we have the following proposition:

Theorem 11.1 *Any monotonic optimization problem can be reduced to the canonical form*

$$(\mathrm{MO})\qquad \max\{f(x)|\ g(x)\leq 0\leq h(x),\ x\in[a,b]\} \tag{11.5}$$

where f,g,h are increasing functions.

Proof Consider a monotonic optimization problem in the general form (11.4). Since the system of inequalities $g_i(x)\leq 0, i=1,\ldots,m$, is equivalent to the single inequality $g(x):=\max_{i=1,\ldots,m}g_i(x)\leq 0$, where $g(x)$ is dm, we can rewrite the problem as

$$\max\{f_1(x)-f_2(x)|\ g_1(x)-g_2(x)\leq 0,\ x\in[a,b]\},$$

where f_1,f_2,g_1,g_2 are increasing functions. Introducing two additional variables z,t we can further write

$$\begin{aligned}\max\{f_1(x)+z|\ &f_2(x)+z\leq 0,\ g_1(x)+t\leq 0\leq g_2(x)+t,\\ &z\in[-f_2(b),-f_2(a)],\ t\in[-g_2(b),-g_1(a)],\ x\in[a,b]\}.\end{aligned}$$

Finally, replacing the system of two inequalities $f_2(x)+z\leq 0, g_1(x)+t\leq 0$ by the single inequality $\max\{f_2(x)+z,g_1(x)+t\}\leq 0$ and changing the notations yield the canonical form (MO). □

Remark 11.1 A minimization problem such as

$$\min\{f(x)|\ g(x) \le 0 \le h(x),\ x \in [a, b]\}$$

can be converted to the form (MO). In fact, by setting $x = a + b - y, \tilde{f}(y) = -f(a + b - y), \tilde{g}(y) = -g(a + b - y)$, it can be rewritten as

$$\max\{\tilde{f}(y)|\ \tilde{h}(y) \le 0 \le \tilde{g}(y),\ y \in [a, b]\}.$$

For simplicity we will assume in the following that $f(x), h(x)$ are upper semi-continuous, while $g(x)$ is lower semi-continuous. Then the set $G := \{x \in [a, b]|\ g(x) \le 0\}$ is a compact normal set, $H := \{x \in [a, b]|\ h(x) \ge 0\}$ is a compact conormal set and the problem (MO) can be rewritten as

$$\max\{f(x)|\ x \in G \cap H\} \tag{11.6}$$

Proposition 11.11 *If $G \cap H \neq \emptyset$ the problem (MO) has at least an optimal solution which is a Pareto point of G.*

Proof Since the feasible set $G \cap H$ is compact, while the objective function $f(x)$ is upper semi-continuous, the problem has an optimal solution $\overline{x}$. If $\overline{x}$ is not a Pareto point of G then the set $G_{\overline{x}} := \{x \in G|\ x \ge \overline{x}\}$ is a normal set in the box $[\overline{x}, b]$ and any Pareto point $\tilde{x}$ of this set is clearly a Pareto point of G and $f(\tilde{x}) = f(\overline{x})$, i.e., $\tilde{x}$ is also optimal. □

Proposition 11.12 *Let $G \subset [a, b]$ be a compact normal set and $z \in [a, b] \setminus G,\ y = \pi_G(z)$. Then the polyblock P with proper vertices*

$$u^i = b + (y_i - b)e^i,\ i = 1, \dots, n \tag{11.7}$$

strictly separates z from G, i.e., $P \supset G, z \in P \setminus G$.

Proof Recall that $\pi_G(z)$ denotes the point where the line segment from a to z meets the upper boundary $\partial^+ G$ of G, so $y < z$. By Proposition 11.7 $P_y = [a, b] \setminus (y, b]$. □

The next corollary shows that with respect to compact normal sets polyblocks behave like polytopes with respect to compact convex sets.

Corollary 11.1 *Any compact normal set is the intersection of a family of polyblocks. In other words, a compact normal set can be approximated as closely as desired by a polyblock.*

Proof Clearly, if $G \subset [a, b]$ then $P_0 = [a, b]$ is a polyblock containing G. Therefore, the family I of polyblocks containing G is not empty and $G \subset \cap_{i \in I} P_i$. If there were $x \in \cap_{i \in I} P_i \setminus G$ then by the above proposition there would exist a polyblock $P \supset G$ such that $x \notin P$, a contradiction. □

On the basis of these properties, one method for solving problem (MO) is to generate inductively a nested sequence of polyblocks outer approximating the feasible set:

$$[a,b] := P_0 \supset P_1 \supset \cdots \supset G \cap H,$$

in such a way that

$$\max\{f(x)|\ x \in P_k\} \searrow \max\{f(x)|\ x \in G \cap H\}.$$

This can be done as follows. Start with the polyblock $P_1 = [a,b]$. At iteration k a polyblock $P_k \supset G$ has been constructed with proper vertex set T_k. If $T'_k = T_k \cap H = \emptyset$ stop: the problem is infeasible. Otherwise let $z^k \in \operatorname{argmax}\{f(x)|\ x \in T'_k\}$. Since $f(x)$ is an increasing function, $f(z^k)$ is the maximum of $f(x)$ over $P_k \cap H \supset G \cap H$. If $z^k \in G$, then stop: $z^k \in G \cap H$, i.e., z^k is feasible, so $f(z^k)$ is also the maximum of $f(x)$ over $G \cap H$, i.e., z^k solves the problem. If, on the other hand, $z^k \notin G$, then compute $x^k = \pi_G(z^k)$ and let $P_{k+1} = ([a,b] \setminus (x^k, b]) \cap P_k$. By Proposition 11.7 $[a,b] \setminus (x^k, b]$ is polyblock, so the polyblock P_{k+1} satisfies $G \subset P_{k+1} \subset P_k \setminus \{z^k\}$. The procedure can then be repeated at the next iteration.

This method is analogous to the outer approximation method for minimizing a convex function over a convex set. Although it can be proved that, under mild conditions, as $k \to +\infty$ the sequence x^k generated by this method converges to a global optimal solution of (MO), the convergence is generally slow. To speed up convergence the method should be improved in two points: (1) bounding: using valid cuts reduce the current box $[p,q]$ in order to have tighter bounds; (2) computing the set T_{k+1} (proper vertex set of polyblock P_{k+1}): not from scratch, but by deriving it from T_k.

11.2 The Polyblock Algorithm

11.2.1 Valid Reduction

At any stage of the solving process, some information may have been gathered about the global optimal solution. To economize the computational effort this information should be used to reduce the search domain by removing certain unfit portions where the global optimal solution cannot be expected to be found. The subproblem which arises is the following.

Given a number $\gamma \in f(G \cap H)$ (the current best feasible objective function value) and a box $[p,q] \subset [a,b]$ check whether the box $[p,q]$ contains at least one solution to the system

$$\{x \in [p,q]|\ g(x) \le 0 \le h(x),\ f(x) \ge \gamma\} \neq \emptyset \tag{11.8}$$

and to find a smaller box $[p',q'] \subset [p,q]$ still containing at least one solution of (11.8). Such a box $[p',q']$ is called a *γ-valid reduction* of $[p,q]$ and is denoted $\mathrm{red}_\gamma[p,q]$.

Observe that if $g(q) \geq 0$ then for every $x \in [p,q]$ satisfying (11.8) the line segment joining x to q intersects the surface $g(.) = 0$ at a point $x' \in [p,q]$ satisfying

$$g(x') = 0 \leq h(x'), \quad f(x') \geq f(x) \geq \gamma.$$

Therefore, the box $[p',q']$ should contain all $x \in [p,q]$ satisfying $g(x) = 0 \leq h(x), f(x) \geq \gamma$, i.e., satisfying

$$g(x) \leq 0 \leq h_\gamma(x) := \min\{g(x), h(x), f(x) - \gamma\}. \tag{11.9}$$

Lemma 11.1 (i) *If $g(p) > 0$ or $h_\gamma(q) < 0$ then there is no $x \in [p,q]$ satisfying (11.9).*

(ii) *If $g(p) \leq 0$ then the box $[p,q']$ where $q' = p + \sum_{i=1}^n \alpha_i(q_i - p_i)e^i$, with*

$$\alpha_i = \sup\{\alpha |\ 0 \leq \alpha \leq 1,\ g(p + \alpha(q_i - p_i)e^i) \leq 0\}, i = 1\ldots,n, \tag{11.10}$$

still contains all feasible solutions to (11.9) in $[p,q]$.

(iii) *If $g(p) \leq 0 \leq h_\gamma(q)$ then the box $[p',q']$ where $p' = q' - \sum_{i=1}^n \beta_i(q'_i - p_i)e^i$, with*

$$\beta_i = \sup\{\beta |\ 0 \leq \beta \leq 1,\ h_\gamma(q' - \beta(q'_i - p_i)e^i) \geq 0\}, i = 1,\ldots,n, \tag{11.11}$$

still contains all feasible solutions to (11.9) in $[p,q]$.

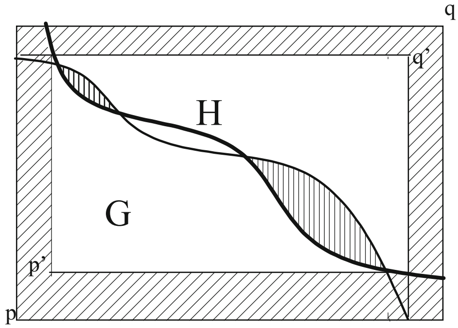
Proof It suffices to prove (ii) because (iii) can be proved analogously, while (i) is obvious. Since $q'_i = \alpha q_i + (1-\alpha)p_i$ with $0 \leq \alpha \leq 1$, it follows that $p_i \leq q'_i \leq q_i\ \forall i = 1,\ldots,n$, i.e., $[p,q'] \subset [p,q]$. Recall that

$$G := \{x \in [p,q]|\ g(x) \leq 0\}.$$

For any $x \in G \cap [p,q]$ we have by normality $[p,x] \subset G$, so $x^i = p + (x_i - p_i)e^i \in G, i = 1,\ldots,n$. But $x_i \leq q_i$, so $x^i = p + \alpha(q_i - p_i)e^i$ with $0 \leq \alpha \leq 1$. This implies that $\alpha \leq \alpha_i$, i.e., $x^i \leq p + \alpha_i(q_i - p_i)e^i, i = 1,\ldots,n$, and consequently $x \leq q'$, i.e., $x \in [p,q']$. Thus, $G \cap [p,q] \subset G \cap [p,q']$, which completes the proof because the converse inclusion is obvious by the fact $[p,q'] \subset [p,q]$. □

Clearly the box $[p,q']$ defined in (1) is obtained from $[p,q]$ by cutting off the set $\cup_{i=1}^n\{x|\ x_i > q'_i\}$, while the box $[p',q]$ defined in (2) is obtained from $[p,q]$ by cutting off the set $\cup_{i=1}^n\{x_i|\ x_i < p'_i\}$. The cut $\cup_{i=1}^n\{x_i|x_i > q'_i\}$ is referred to as an *upper γ-valid cut* with vertex q' and the cut $\cup_{i=1}^n\{x_i|\ x_i < p'_i\}$ as a *lower γ-valid cut* with vertex p', applied to the box $[p,q]$ (Fig. 11.2). We now deduce a closed formula for $\mathrm{red}_\gamma[p,q]$.

Fig. 11.2 Reduced box



Given any continuous increasing function $\varphi(\alpha) : [0, 1] \to \mathbb{R}$ such that $\varphi(0) < 0 < \varphi(1)$ let $N(\varphi)$ denote the largest value $\overline{\alpha} \in (0, 1)$ satisfying $\varphi(\overline{\alpha}) = 0$. For every $i = 1, \ldots, n$ let $q^i := p + (q_i - p_i)e^i$.

Proposition 11.13 *A γ-valid reduction of a box $[p, q]$ is given by the formula*

$$\mathrm{red}_\gamma[p, q] = \begin{cases} \emptyset, & \text{if } g(p) > 0 \text{ or } h_\gamma(q') < 0, \\ [p', q'], & \text{if } g(p) \le 0 \;\&\; h_\gamma(q') \ge 0. \end{cases} \tag{11.12}$$

where

$$q' = p + \sum_{i=1}^{n} \alpha_i(q_i - p_i)e^i, \quad p' = q' - \sum_{i=1}^{n} \beta_i(q' - p_i)e^i, \tag{11.13}$$

$$\alpha_i = \begin{cases} 1, & \text{if } g(q^i) \le 0 \\ N(\varphi_i), & \text{if } g(q^i) > 0. \end{cases} \qquad \varphi_i(t) = g(p + t(q_i - p_i)e^i).$$

$$\beta_i = \begin{cases} 1, & \text{if } h_\gamma(q'^i) \ge 0 \\ N(-\psi_i), & \text{if } h_\gamma(q'^i) < 0. \end{cases} \qquad \psi_i(t) = h_\gamma(q' - t(q'_i - p_i)e^i).$$

Proof Straightforward from Lemma 11.1 . □

Remark 11.2 The reduction operation which is meant to enhance the efficiency of the search procedure is worthwhile only if the cost of this operation does not offset its advantage. Therefore, although the exact values of $N(\varphi_i), N(-\psi_i)$ can be computed easily using, e.g., a procedure of Bolzano bisection, in practical implementation of the above formula it is enough to compute only approximate values of $N(\varphi_i), N(-\psi_i)$ by applying just a few iterations of the Bolzano bisection procedure.

Proposition 11.14 *Let $E := \{x \in G \cap H |\, f(x) \ge \gamma\}$, and let P be a polyblock containing E with proper vertex V. Let V' be the set obtained from V by deleting* *every* $z \in V$ *satisfying* $red_\gamma[a,z] = \emptyset$ *and replacing every other* $z \in V$ *with the highest corner* z' *of the box* $\mathrm{red}_\gamma[a,z] = [a,z']$. *Then the polyblock* P' *generated by* V' *satisfies*

$$\{x \in G \cap H | f(x) \geq \gamma\} \subset P' \subset P.$$

Proof Since $E \cap [a,z] = \emptyset$ for every deleted z, while $E \cap [a,z] \subset \mathrm{red}_\gamma[a,z] := [a,z']$ for every other z, the conclusion follows. □

We shall refer to the polyblock P' with vertex set V' as the γ-valid reduction of the polyblock P and denote it by $\mathrm{red}_\gamma P$.

11.2.2 Generating the New Polyblock

The next question after reducing the search domain is how to derive the new polyblock P_{k+1} that should better approximate $G \cap H$ than the current polyblock P_k. The general subproblem to be solved is the following.

Given a polyblock $P \subset [a,b]$ *with* $V = \mathrm{pvert}P$, *a vertex* $v \in V$ *and a point* $x \in [a,v)$ *such that* $G \cap H \subset P \setminus (x,b]$, *determine a new polyblock* P' *satisfying* $P \setminus (x,b] \subset P' \subset P \setminus \{v\}$.

For any two z, y let $J(z,y) = \{j|\ z_j > y_j\}$ and if $a \leq x \leq z \leq b$ then define $z^i = z + (x_i - z_i)e^i, i = 1,\ldots,n$.

Proposition 11.15 *Let* P *be a polyblock with proper vertex* $V \subset [a,b]$ *and let* $x \in [a,b]$ *satisfy* $V_* := \{z \in V|\ x < z\} \neq \emptyset$.

(i) $P' := P \setminus (x,b]$ *is a polyblock with vertex set*

$$T' = (V \setminus V_*) \cup \{z^i = z + (x_i - z_i)e^i |\ z \in V_*, i = 1,\ldots,n\}. \tag{11.14}$$

(ii) *The proper vertex set of* P' *is obtained from* T' *by removing improper elements according to the rule:*

For every pair $z \in V_*, y \in V_*^+ := \{y \in V|\ x \leq y\}$ *compute* $J(z,y) = \{J|\ z_j > y_j\}$ *and if* $J(z,y) = \{i\}$ *then* z^i *is an improper element of* T'.

Proof Since $[a,z] \cap (x,b] = \emptyset$ for every $z \in V \setminus V_*$ it follows that $P \setminus (x,b] = P_1 \cup P_2$, where P_1 is the polyblock with vertex set $V \setminus V_*$ and $P_2 = (\cup_{z \in V_*}[a,z]) \setminus (x,b] = \cup_{z \in V_*}([a,z] \setminus (x,b])$. Noting that by Proposition 11.7 $[a,b] \setminus (x,b]$ is a polyblock with vertices $u^i = +(x_i - b_i)e^i, i = 1,\ldots,n$, we can then write $[a,z] \setminus (x,b] = [a,z] \cap ([a,b] \setminus (x,b]) = [a,z] \cap (\cup_{i=1}^n [a,u^i]) = \cup_{i=1}^n [a,z] \cap [a,u^i] = \cup_{i=1}^n [a, z \wedge u^i]$, where $z \wedge u^i$ denotes the vector whose every j-th component equals $\min\{z_j, u^i_j\}$. Hence, $P_2 = \cup\{[a, z \wedge u^i]|\ z \in V_*, i = 1,\ldots,n\}$. Since $z \wedge u^i = z^i$ for $z \in V_*$, this shows that the vertex set of $P \setminus (x,b]$ is the set T' given by (11.14).

It remains to show that every $y \in V \setminus V_*$ is proper, while a $z^i = z + (x_i - z_i)e^i$ with $z \in V_*$ is improper if and only if $J(z, y) = \{i\}$ for some $y \in V_*^+$.

Since every $y \in V \setminus V_*$ is proper in V, while $z^i \le z \in V$ for every z^i it is clear that every $y \in V \setminus V_*$ is proper. Therefore, an improper element must be some z^i such that $z^i \le y$ for some $y \in T'$. Two cases are possible: (1) $y \in V$; (2) $y \in T' \setminus V$. In the former case, since obviously $x \le z^i$ we must have $x \le y$; furthermore, $z_j = z^i_j \le y_j \ \forall j \neq i$, hence, since $z \not\le y$, it follows that $z_i > y_i$, i.e., $J(z, y) = \{i\}$. In the latter case, $z^i \le y^l$ for some $y \in V_*$ and some $l \in \{1, \dots, n\}$. Then it is clear that we cannot have $y = z$, so $y \neq z$ and $z^i_j \le y^l_j \ \forall j = 1, \dots, n$. It $l = i$ then this implies $z_j = z^i_j \le y^i_j = y_j \ \forall j \neq i$, hence, since $z \not\le y$ it follows that $z_i > y_i$ and $J(z, y) = \{i\}$. On the other hand, if $l \neq i$ then from $z^i_j \le y^l_j \ \forall j = 1, \dots, n$, we have $z_j \le y_j \ \forall j \neq i$ and again, since $z \not\le y$ we must have $z_i > y_i$, so $J(z, y) = \{i\}$. Thus an improper z^i must satisfy $J(z, y) = \{i\}$ for some $y \in V_*^+$. Conversely, if $J(z, y) = \{i\}$ for some $y \in V_*^+$ then $z_j \le y_j \ \forall j \neq i$, hence, since $z^i_i = x_i$ and $x_i = y_i$ if $y \in V_*$ or else $x_i \le y_i$ if $y \in V_*^+$, it follows that $z^i \le y^i$ if $y \in V_*$ or else $z^i \le y$ if $y \in V_*^+$; so in either case z^i is improper. This completes the proof of the proposition. □

11.2.3 Polyblock Algorithm

By shifting the origin to $-\alpha e$ if necessary we may always assume that

$$x \le a + \alpha e \quad \forall x \in H. \tag{11.15}$$

where $\alpha > 0$ is chosen not too small compared to $\|b - a\|$ (e.g., $\alpha = \frac{1}{4}\|b - a\|$). Furthermore, since the problem is obviously infeasible if $b \notin H$ whereas b is an obvious optimal solution if $b \in G \cap H$, we may also assume that

$$b \in H \setminus G, \quad \text{i.e., } h(b) \ge 0, \ g(b) \ge 0. \tag{11.16}$$

Algorithm POA (Polyblock Outer Approximation Algorithm, or briefly, Polyblock Algorithm)

Initialization Let P_1 be an initial polyblock containing $G \cap H$ and let V_1 be its proper vertex set. For example, $P_1 = [a, b]$, $V_1 = \{b\}$. Let $\overline{x}^1$ be the best feasible solution available and $CBV = f(\overline{x}^1)$ (if no feasible solution is available, set $CBV = -\infty$). Set $k = 1$.

Step 1. For $\gamma = CBV$ let $\overline{P}_k = \text{red}_\gamma P_k$ and $\overline{V}_k = \text{pvert}\overline{P}_k$. Reset $b \leftarrow \vee\{x \in \overline{V}_k\}$ where $x = \vee\{a \in A\}$ means $x_i = \max_{a \in A} a_i \ \forall i = 1, \dots, n$.

Step 2. If $\overline{V}_k = \emptyset$ terminate: if $CBV = -\infty$ the problem is infeasible; if $CBV > -\infty$ the current best solution $\overline{x}^k$ is an optimal solution.

Step 3. If $\overline{V}_k \neq \emptyset$ select $z^k \in \text{argmax}\{f(x) | \ x \in \overline{V}_k\}$. If $g(z^k) \le 0$ then terminate: z^k is an optimal solution. Otherwise, go to Step 4.

Step 4. Compute $x^k = \pi_G(z^k)$, the last point of G in the line segment from a to x^k. Determine the new current best feasible $\overline{x}^{k+1}$ and the new current best value CBV. Compute the proper vertex set V_{k+1} of $P_{k+1} = \overline{P}_k \setminus (x^k, b]$ according to Proposition 11.15.

Step 5. Increment k and return to Step 1.

Proposition 11.16 *When infinite Algorithm POA generates an infinite sequence $\{x^k\}$ every cluster point of which is a global optimal solution of (MO).*

Proof First note that by replacing P_k with $\overline{P}_k = \text{red}_\gamma P_k$ in Step 1, all $z \in V_k$ such that $f(z) < \gamma$ are deleted, so that at any iteration k all feasible solutions z with $f(z) \geq CBV$ are contained in the polyblock $\overline{P}_k$. This justifies the conclusions when the algorithm terminates at Step 2 or Step 3. Therefore it suffices to consider the case when the algorithm is infinite.

Condition (12.16) implies that

$$\min_i (z_i - a_i) \geq \frac{\alpha}{\|z - a\|}\|z - a\| \geq \rho\|z - a\| \quad \forall z \in H, \tag{11.17}$$

where $\rho = \frac{\alpha}{\|b-a\|}$. We contend that $z^k - x^k \to 0$ as $k \to +\infty$. Suppose the contrary, that there exist $\eta > 0$ and an infinite sequence k_l such that $\|z^{k_l} - x^{k_l}\| \geq \eta > 0\ \forall l$. For $\mu > l$ we have $z^{k_\mu} \notin (x^{k_l}, z^{k_l}]$ because $P_{k_{m}u} \subset P_{k_l} \setminus (x^{k_l}, b]$. Hence, $\|z^{k_m u} - z^{k_l}\| \geq \min_{i=1,\ldots,n} |z_i^{k_l} - x_i^{k_l}|$. On the other hand, $\min_{i=1,\ldots,n}(z_i^{k_l} - a_i)] \geq \rho\|z^{k_l} - a\|$ by (11.17) because $z^{k_l} \in H$, while x^{k_l} lies on the line segment joint a to z^{k_l}, so $z_i^{k_l} - x_i^{k_l} = \frac{z_i^{k_l} - a_i}{\|z^{k_l} - a\|}\|z^{k_l} - x^{k_l}\| \geq \rho\|z^{k_l} - x^{k_l}\|\ \forall i$, i.e., $\min_{i=1,\ldots,n} |z_i^{k_l} - x_i^{k_l}| \geq \rho\|z^{k_l} - x^{k_l}\|$. Therefore, $\|z^{k_m u} - z^{k_l}\| \geq \min_{i=1,\ldots,n} |z_i^{k_l} - x_i^{k_l}| \geq \rho\|z^{k_l} - x^{k_l}\| \geq \rho\mu$, conflicting with the boundedness of the sequence $\{z^{k_l}\} \subset [a, b]$.

Thus $z^k - x^k \to 0$ and by boundedness we may assume that, up to subsequences, $x^k \to \overline{x}, z^k \to \overline{x}$. Then, since $z^k \in H, x^k \in G\ \forall k$, it follows that $\overline{x} \in G \cap H$, i.e., $\overline{x}$ is feasible. Furthermore, $f(z^k) \geq f(z)\ \forall z \in_k \supset G \cap H$, hence, by letting $k \to +\infty$, $f(\overline{x}) \geq f(x) \quad \forall x \in G \cap H$, i.e., $\overline{x}$ is a global optimal solution. □

Remark 11.3 In practice we must stop the algorithm at some iteration k. The question arises as to when to stop, i.e., how large k should be, to obtain a z^k sufficiently close to an optimal solution.

If the problem (MO) does not involve conormal constraints, i.e., has the form

$$\text{(MO1)} \qquad \max\{f(x) |\ g(x) \leq 0,\ x \in [a, b]\},$$

then x^k is always feasible, so when $f(z^k) - f(x^k) < \varepsilon$, then $f(x^k) + \varepsilon > f(z^k) \geq \max\{f(x) |\ g(x) \leq 0,\ x \in [a, b]\}$, i.e., x^k will be an ε-optimal solution of the problem. In other words, the Polyblock Algorithm applied to problem (MO1) provides an ε-*optimal solution* in finitely many steps.

In the general case, if both normal and conormal constraints are present, x^k may not be feasible, and the Polyblock Algorithm may not provide an ε-optimal solution in finitely many steps. The most that can be said is that, since $g(z^k) - g(x^k) \to 0$ and $g(x^k) \leq 0$, while $h(z^k) \geq 0\ \forall k$, $f(z^k) \to \max$ (MO) as $k \to +\infty$, we will have, for k sufficiently large, $g(z^k) \leq \varepsilon, h(z^k) \geq 0$, $f(z^k) \geq \max$ (MO) $- \varepsilon$.

A point $\overline{x} \in [a, b]$ satisfying

$$g(\overline{x}) - \varepsilon \leq 0 \leq h(\overline{x}),\ f(\overline{x}) \geq \max\,(\text{MO}) - \varepsilon$$

is called an *ε-approximate optimal solution* of (MO). Thus, given any tolerance $\varepsilon > 0$, Algorithm POA can only provide an *ε-approximate optimal solution z^k* in finitely many steps.

11.3 Successive Incumbent Transcending Algorithm

Very often in practice we are not so much interested in solving a problem to optimality as in obtaining quickly a feasible solution better than a given one. In that case, the Polyblock Algorithm is a good choice only if the problem does not involve conormal constraints, i.e., if its feasible set is a normal set. But if the problem involves both normal and conormal sets, i.e., if the feasible set is the intersection of a normal and a conormal set, then the Polyblock Algorithm does not serve our purpose. As we saw at the end of the previous section, for such a problem the Polyblock Algorithm is incapable of computing a feasible solution in finitely many iterations. More importantly, when the problem happens to have an isolated feasible or almost feasible solution with an objective function value significantly better than the optimal value (as, e.g., point x^* in the problem depicted in Fig. 11.3), then the ε-approximate solution produced by the algorithm may be quite far off the optimal solution and so can hardly be accepted as an adequate approximation of the optimal solution.

The situation is very much like to what happens for dc optimization problems with a nonconvex feasible set. To overcome or at least alleviate the above difficulties a robust approach to (MO) can be developed which is similar to the robust approach for dc optimization (Chap. 7, Sect. 7.5), with two following basic features :

1. Instead of an (ε, η)-approximate optimal solution an *essential η-optimal solution* is computed which is always feasible and provides an adequate approximation of the optimal solution;
2. The algorithm generates a sequence of better and better nonisolated feasible solutions converging to a nonisolated feasible solution which is the best among all nonisolated feasible solutions. In that way, for any prescribed tolerance $\eta > 0$ an essential η-optimal solution is obtained in finitely many iterations.

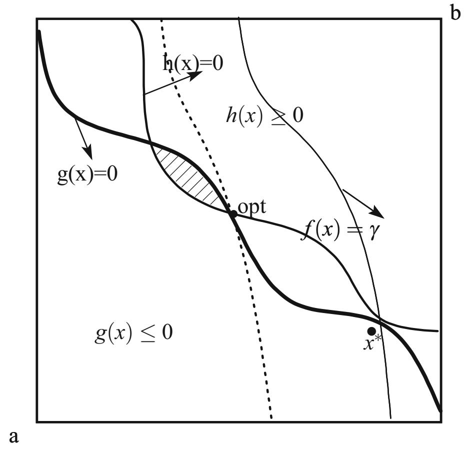


Fig. 11.3 Inadequate approximate optimal solution x^*

11.3.1 Essential Optimal Solution

When the problem has an isolated optimal solution this optimal solution is generally difficult to compute and difficult to implement when computable.

To get round the difficulty a common practice is to assume that the feasible set D of the problem under consideration is *robust*, i.e., it satisfies

$$D = \mathrm{cl}(\mathrm{intD}), \tag{11.18}$$

where cl and int denote the closure and the interior, respectively. Condition (11.18) rules out the existence of an isolated feasible solution and ensures that for $\varepsilon > 0$ sufficiently small an ε-approximate optimal solution will be close enough to the exact optimum. However, even in this case the trouble is that in practice we often do not know what does it mean precisely *sufficiently small* $\varepsilon > 0$. Moreover, condition (11.18) is generally very hard to check. Quite often we have to deal with feasible sets which are not known a priori to contain isolated points or not.

Therefore, from a practical point of view a method is desirable which could help to discard isolated feasible solutions from consideration without having to preliminarily check their presence. This method should employ a more adequate concept of approximate optimal solution than the commonly used concept of ε-approximate optimal solution.

A point $x \in \mathbb{R}^n$ is called an *essential feasible* solution of (MO) if it satisfies

$$x \in [a, b], \quad g(x) < 0 \leq h(x).$$

Given a tolerance $\eta > 0$ an essential feasible solution $\overline{x}$ is said to be an *essential η-optimal solution* of (MO) if it satisfies $f(\overline{x}) \geq f(x) - \eta$ for all essential feasible solutions x of (MO).

Given a tolerance $\varepsilon > 0$, a point $x \in \mathbb{R}^n$ is said to be an *ε-essential feasible solution* if it satisfies

$$x \in [a, b], \ g(x) + \varepsilon \leq 0 \leq h(x)$$

Finally, given two tolerances $\varepsilon > 0, \eta > 0$, an ε-essential feasible solution $\overline{x}$ is said to be (ε, η)*-optimal* if it satisfies $f(\overline{x}) \geq f(x) - \eta$ for all ε-essential optimal solutions.

Since g, h are continuous increasing functions, it is readily seen that if x is an essential feasible point then for all $t > 0$ sufficiently small $x' = x + t(b - x)$ is still an essential feasible point. Therefore, an essential feasible solution is always a nonisolated feasible solution.

The above discussion suggests that instead of trying to find an optimal solution to (MO), it would be more practical and reasonable to look for an essential η-optimal solution. The *robust approach* to problem (MO) to be presented below embodies this point of view.

11.3.2 Successive Incumbent Transcending Strategy

A key step towards finding a global optimal solution of problem (MO) is to deal with the following subproblem of incumbent transcending:

($SP\gamma$) *Given a real number γ, check whether* (MO) *has a nonisolated feasible solution x satisfying $f(x) > \gamma$, or else establish that no nonisolated feasible solution x of* (MO) *exists such that $f(x) > \gamma$.*

Since f, g are increasing, if $f(b) \leq \gamma$, then $f(x) \leq \gamma \ \forall x \in [a, b]$, so there is no feasible solution with objective function value exceeding γ; if $g(a) > 0$, then $g(x) > 0 \ \forall x \in [a, b]$, so the problem is infeasible.

If $f(b) > \gamma$ and $g(b) < 0$ then b is an essential feasible solution with objective function value exceeding γ.

Therefore, barring these trivial cases, we can assume that

$$f(b) > \gamma, \quad g(a) \leq 0 \leq g(b). \tag{11.19}$$

To seek an answer to ($SP\gamma$) consider the **master problem**:

$$\min\{g(x) | \ h(x) \geq 0, f(x) \geq \gamma, \ x \in [a, b]\}.$$

Setting $k_\gamma(x) := \min\{h(x), f(x) - \gamma\}$ and noting that $k_\gamma(x)$ is an increasing function, we write the master problem as

$$(Q_\gamma) \qquad \min\{g(x) | \ k_\gamma(x) \geq 0, \ x \in [a, b]\}.$$

For our purpose an important feature of this problem is that it can be solved efficiently, while solving it furnishes an answer to question (IT), as will shortly be seen.

By $\min(Q_\gamma)$ denote the optimal value of (Q_γ). The robust approach to (MO) is founded on the following relationship between problems (MO) and (Q_γ):

Proposition 11.17 *Assume (11.19).*

(i) *If* $\min(Q_\gamma) < 0$ *there exists an essential feasible solution* x^0 *of (MO) satisfying* $f(x^0) \geq \gamma$.
(ii) *If* $\min(Q_\gamma) \geq 0$, *then there is no essential feasible solution* x *of (MO) satisfying* $f(x) \geq \gamma$.

Proof (i) If $\min(Q_\gamma) < 0$ there exists an $x^0 \in [a,b]$ satisfying $g(x^0) < 0 \leq h(x^0)$, $f(x^0) \geq \gamma$. This point x^0 is an essential feasible solution of (MO), satisfying $f(x^0) \geq \gamma$.
(ii) If $\min(Q_\gamma) \geq 0$ any $x \in [a,b]$ such that $g(x) < 0$ is infeasible to (Q_γ). Therefore, any essential feasible solution of (MO) must satisfy $f(x) < \gamma$. □

On the basis of this proposition, to solve the problem (MO) one can proceed according to the following conceptual SIT (Successive Incumbent Transcending) scheme:

Select tolerance $\eta > 0$. Start from $\gamma = \gamma_0$ with $\gamma_0 = f(a)$.
Solve the master problem (Q_γ).

– If $\min(Q_\gamma) < 0$ an essential feasible $\bar{x}$ with $f(\bar{x}) \geq \gamma$ is found. Reset $\gamma \leftarrow f(\bar{x}) - \eta$ and repeat.
– Otherwise, $\min(Q_\gamma) \geq 0$: no essential feasible solution x for (MO) exists such that $f(x) \geq \gamma$. Hence, if $\gamma = f(\bar{x}) - \eta$ for some essential feasible solution $\bar{x}$ of (MO) then $\bar{x}$ is an essential η-optimal solution; if $\gamma = \gamma_0$ the problem (MO) is essentially infeasible (has no essential feasible solution).

Below we indicate how this conceptual scheme can be practically implemented.

a. Solving the Master Problem

Recall that the master problem (Q_γ) is

$$\min\{g(x) | \; x \in [a,b], k_\gamma(x) \geq 0\}, \tag{11.20}$$

where $k_\gamma(x) := \min\{h(x), f(x) - \gamma\}$ is an increasing function. Since the feasible set of (Q_γ) is robust (has no isolated point) and easy to handle (a feasible solution can be computed at very cheap cost), this problem can be solved efficiently by a rectangular RBB (reduce-bound-and-branch) algorithm characterized by three basic operations (see Chap. 6, Sect. 6.2):

– *Reducing*: using valid cuts reduce each partition set $M = [p,q] \subset [a,b]$ without losing any feasible solution currently still of interest. The resulting box $[p', q']$ is referred to as a *valid reduction* of M.

– *Bounding*: for each partition set $M = [p, q]$ compute a lower bound $\beta(M)$ for $g(x)$ over the feasible solutions of (Q_γ) contained in the valid reduction of M;
– *Branching*: use an *adaptive subdivision rule* to further subdivide a selected box in the current partitioning of the original box $[a, b]$.

b. Valid Reduction

At a given stage of the RBB algorithm for (Q_γ) let $[p, q]$ be a box generated during the partitioning process and still of interest. The search for a nonisolated feasible solution x of (MO) in $[p, q]$ such that $f(x) \geq \gamma$ can then be restricted to the set

$$\{x \in [p, q]|\ g(x) \leq 0 \leq k_\gamma(x)\}.$$

A γ*-valid reduction* of $[p, q]$, written $\text{red}_\gamma[p, q]$, is a box $[p', q'] \subset [p, q]$ that still contains all the just mentioned set.

For any box $[p, q] \subset \mathbb{R}^n$ define $q^i := p+(q_i-p_i)e^i,\ i = 1, \ldots, n$. As in Sect. 10.2, for any increasing function $\varphi(t) : [0, 1] \to \mathbb{R}$ denote by $N(\varphi)$ the largest value of $t \in [0, 1]$ such that $\varphi(t) = 0$. By Proposition 11.13 we can state:

A γ-valid reduction of a box $[p, q]$ is given by the formula

$$\text{red}_\gamma[p, q] = \begin{cases} \emptyset, & \text{if } g(p) > 0 \text{ or } k_\gamma(q) < 0, \\ [p', q'] & \text{if } g(p) \leq 0\ \&\ k_\gamma(q) \geq 0. \end{cases}$$

where

$$q' = p + \sum_{i=1}^{n} \alpha_i(q_i - p_i)e^i, \quad p' = q' - \sum_{i=1}^{n} \beta_i(q'_i - p_i)e^i, \tag{11.21}$$

$$\alpha_i = \begin{cases} 1, & \text{if } g(q^i) \leq 0 \\ N(\varphi_i), & \text{if } g(q^i) > 0. \end{cases} \qquad \varphi_i(t) = g(p + t(q_i - p_i)e^i).$$

$$\beta_i = \begin{cases} 1, & \text{if } k_\gamma(q'^i) \geq 0 \\ N(-\psi_i), & \text{if } k_\gamma(q'^i) < 0. \end{cases} \qquad \psi_i(t) = k_\gamma(q' - t(q'_i - p_i)e^i).$$

In the context of RBB algorithm for (Q_γ), a smaller $\text{red}_\gamma[p, q]$ allows a tighter bound to be obtained for $g(x)$ over $[p, q]$ but requires more computational effort. Therefore, as already mentioned in Remark 11.2, in practical implementation of the above formulas for α_i, β_i it is enough to compute approximate values of $N(\varphi_i)$ or $N(-\psi_i)$ by using, e.g., just a few iterations of a Bolzano bisection procedure.

c. Bounding

Let $M = [p, q]$ be a valid reduction of a box in the current partition. Since $g(x)$ is an increasing function an obvious lower bound for $g(x)$ over the box $[p, q]$, is the number

$$\beta(M) = g(p). \tag{11.22}$$

As simple as it is, this bound combined with an adaptive branching rule suffices to ensure convergence of the algorithm.

However, to enhance efficiency better bounds are often needed.

For example, applying two or more iterations of the Polyblock procedure for computing $\min\{g(x)|\ x \in [p, q]\}$ may yield a reasonably good lower bound. Alternatively, based on the observation that $\min\{g(x) + tk_\gamma(x)|\ x \in [p, q]\} \leq \min\{g(x)|\ x \in [p, q], k_\gamma(x) \geq 0\}\ \forall t \in \mathbb{R}_+$, one may take $\beta(M) = g(p) + tk_\gamma(p)$ for some $t > 1$. As shown in (Tuy 2010), if either of the functions $f(x), g(x) + tk_\gamma(x)$ is coercive, i.e., tends to $+\infty$ as $\|x\| \to +\infty$, then this bound approaches the exact minimum of $g(x)$ over $[p, q]$ as $t \to +\infty$.

Also one may try to exploit any partial convexity present in the problem. For instance, if a convex function $u(x)$ and a concave function $v(x)$ are available such that $u(x) \leq g(x),\ v(x) \leq k_\gamma(x)\ \forall x \in [p, q]$, then a lower bound is given by the optimal value of the convex minimization problem

$$\min\{u(x)|\ v(x) \geq 0, x \in [p, q]\}$$

which can be solved by many currently available algorithms.

Sometimes it may be useful to write $g(x)$ as a sum of several functions such that a lower bound can be easily computed for each of these functions: $\beta(M)$ can then be taken to be the sum of the separate lower bounds. Various methods and examples of computing good bounds are considered in (Tuy-AlKhayyal-Thach 2005).

d. Branching

Branching is performed according to an adaptive rule.

Specifically let $M_k = [p^k, q^k]$ be the valid reduction of the partition set chosen for further subdivision at iteration k, and let $z^k \in M_k$ be a point such that $g(z^k)$ provides a lower bound for $\min\{g(x)|\ k_\gamma(x) \geq 0, x \in M_k\}$. (for instance, $z^k = p^k$). Obviously, if $k_\gamma(z^k) \geq 0$ then $g(z^k)$ is the exact minimum of the latter problem and so if M_k is the partition set with smallest lower bound among all partition sets M currently still of interest then z^k is a global optimal solution of (Q_γ). Based on this observation one can use an adaptive branching method to eventually drive z^k to the feasible set of (Q_γ). This method proceeds as follows. Let y^k be the point where the line segment $[z^k, q^k]$ meets the surface $k_\gamma(x) = 0$, i.e., $y^k = N(\xi)$ for $\xi(t) = k_\gamma(z^k + t(q^k - z^k)), 0 \leq t \leq 1$. Let w^k be the midpoint of the segment

$[z^k, y^k] : w^k = (z^k + y^k)/2$ and $j_k \in \text{argmax}\{|z_i^k - y_i^k| \ i = 1, \ldots, n\}$. Then divide the box $M_k = [p^k, q^k]$ into two smaller boxes M_{k-}, M_{k+} via the hyperplane $x_{j_k} = w_{j_k}^k$:

$$M_{k-} = [p^k, q^{k-}], \quad M_{k+} = [p^{k+}, q^k], \quad \text{with}$$
$$q^{k-} = q^k - (q_{j_k}^k - w_{j_k}^k)e^{j_k}, \ p^{k+} = p^k + (w_{j_k}^k - p_{j_k}^k)e^{j_k}.$$

In other words, subdivide M_k according to the adaptive subdivision rule via (z^k, y^k).

Proposition 11.18 *Whenever infinite the above adaptive* RBB *algorithm for* (Q_γ) *generates a sequence of feasible solutions* y^k *at least a cluster point* $\overline{y}$ *of which is an optimal solution of* (Q_γ).

Proof Since the algorithm uses an adaptive subdivision rule via (z^k, y^k), by Theorem 6.4 there exists an infinite sequence $K \subset \{1, 2, \ldots\}$ such that $\|z^k - y^k\| \to 0$ as $k \in K, k \to +\infty$. From the fact

$$g(z^k) \leq \min\{g(x)| \ x \in [a, b], k_\gamma(x) \geq 0\}, k_\gamma(y^k) = 0,$$

we deduce that, up to a subsequence, $z^k \to \overline{y}, \ y^k \to \overline{y}$ with $g(\overline{y}) = \min\{g(x)| \ x \in [a, b], k_\gamma(x) \geq 0\}$. □

Incorporating the above RBB procedure for solving the master problem into the SIT (Successive Incumbent Transcending) Scheme yields the following robust algorithm for solving (MO):

SIT Algorithm for (MO)

Let $\gamma_0 = f(b)$. Select tolerances $\varepsilon > 0, \eta > 0$.

Initilization. Start with $\mathscr{P}_1 = \{M_1\}, M_1 = [a, b], \mathscr{R}_1 = \emptyset$. If an essential feasible solution $\overline{x}$ of (MO) is available let $\gamma = f(\overline{x}) - \eta$. Otherwise, let $\gamma = \gamma_0$. Set $k = 1$.

Step 1. For each box $M \in \mathscr{P}_k$:

- Compute its valid reduction $\text{red}_\gamma M$;
- Delete M if $\text{red}_\gamma M = \emptyset$. Replace M by $\text{red}_\gamma M$ if otherwise;
- Letting $\text{red}_\gamma M = [p, q]$ compute a lower bound $\beta(M)$ for $g(x)$ over the set $\{x \in [p, q]| \ k_\gamma(x) \geq 0\}$ such that $\beta(M) = g(z)$ for some $z \in [p, q]$.
- Delete M if $\beta(M) > -\varepsilon$.

Step 2. Let $\mathscr{P}'_k$ be the collection of boxes that results from $\mathscr{P}_k$ after completion of Step 1 and let $\mathscr{R}'_k = \mathscr{R}_k \cup \mathscr{P}'_k$.

Step 3. If $\mathscr{R}'_k = \emptyset$, then terminate: if $\gamma = \gamma_0$ the problem (MO) is ε-essential infeasible; if $\gamma = f(\overline{x}) - \eta$ then $\overline{x}$ is an essential (ε, η)-optimal solution of (MO).

Step 4. If $\mathscr{R}'_k \neq \emptyset$, let $[p^k, q^k] = M_k \in \text{argmin}\{\beta(M)| \ M \in \mathscr{R}'_k\}$ and let $z^k \in M_k$ be the point satisfying $g(z^k) = \beta(M_k)$. If $k_\gamma(z^k) \geq 0$ let $y^k = z^k$ and go to Step 5. If $k_\gamma(z^k) < 0$, let y^k be the point where the line segment joining z^k

to q^k meets the surface $k_\gamma(x) = 0$. If $g(y^k) < \varepsilon$ go to Step 5. If $g(y^k) \geq \varepsilon$ go to Step 6.

Step 5. y^k is an essential feasible solution of (MO) with $f(y^k) \geq \gamma$. Reset $\overline{x} \leftarrow y^k$ and go back to Step 0.

Step 6. Divide M_k into two subboxes according to an adaptive branching via (z^k, y^k). Let $\mathscr{P}_{k+1}$ be the collection of these two subboxes of M_k. Let $\mathscr{R}_{k+1} = \mathscr{R}'_k \setminus \{M_k\}$. Increment k and return to Step 1.

Proposition 11.19 *The above algorithm terminates after finitely many steps, yielding either an essential* (ε, η)*-optimal solution of* (MO)*, or an evidence that the problem is essentially infeasible.*

Proof Since Step 1 deletes every box M with $\beta(M) > -\varepsilon$, the fact $\mathscr{R}_k = \emptyset$ in Step 3 implies that $g(x) + \varepsilon > 0\ \forall x \in [p, q] \cap \{x|\ k_\gamma(x) \geq 0\}$, hence there is no ε-essential feasible solution x of (MO) with $f(x) \geq \gamma$, justifying the conclusion in Step 3. It remains to show that Step 3 occurs for sufficiently large k. Suppose the contrary, that the algorithm is infinite. Since each occurrence of Step 5 increases the current value of $f(\overline{x})$ by at least $\eta > 0$ while $f(x)$ is bounded above on the feasible set, Step 5 cannot occur infinitely often. Therefore, for all k sufficiently large Step 6 occurs, i.e., $g(y^k) \geq \varepsilon$. By Proposition 11.18 there exists a subsequence $\{k_s\}$ such that $\|z^{k_s} - y^{k_s}\| \to 0$ as $s \to +\infty$. Also, by compactness, one can assume that $y^{k_s} \to \overline{y}$, $z^{k_s} \to \overline{y}$. Then $g(\overline{y}) \geq \varepsilon$, while $g(z^k) = \beta(M_k) \leq -\varepsilon$, hence $g(\overline{y}) \leq -\varepsilon$, a contradiction. □

Remark 11.4 So far we assumed that the problem involves only inequality constraints. In the case there are also equality constraints, e.g.,

$$c_j(x) = 0,\ j = 1, \ldots, s. \tag{11.23}$$

observe that the linear constraints can be used to eliminate certain variables, so we can always assume that all the constraints (11.23) are *nonlinear*. Since, however, in the most general case one cannot expect to compute a solution to a system of nonlinear equations in finitely many steps, one should be content with replacing the system (11.23) by the approximate system

$$-\delta \leq c_j(x) \leq \delta, \quad j = 1, \ldots, s\,,$$

where $\delta > 0$ is the tolerance. The method presented in the previous sections can then be applied to the resulting approximate problem. An essential ε-optimal solution to the above defined approximate problem is called an *essential* (δ, ε)*-optimal solution* of (MO).

e. Application to Quadratic Programming

Consider the quadratically constrained quadratic optimization problem

$$\text{(QQP)} \qquad \min\{f_0(x) | f_k(x) \geq 0, k = 1, \ldots, m, \ x \in C\}$$

where $C \subset \mathbb{R}^n$ is a polyhedron and each $f_k(x), k = 0, 1 \ldots, m$, is a quadratic function. Noting that every quadratic polynomial can be written as a difference of two quadratic polynomials with positive coefficients we can rewrite (QQP) as a monotonic optimization problem of the form

$$(P) \qquad \min\{f(x) | \ g(x) \leq 0 \leq h(x), \ x \in [a, b]\}$$

where $f(x), g(x), h(x)$ are quadratic functions with positive coefficients (hence, increasing functions). The SIT method applied to the monotonic optimization problem (P) then yields a robust algorithm for solving (QQP). For details of implementation, we refer the reader to the article (Tuy and Hoai-Phuong 2007).

Sometimes it may be more convenient to convert (QQP) to the form

$$(Q) \qquad \min\{f(x) | \ g(x) \geq 0, \ x \in [a, b]\}$$

where $g(x) = \min_{k=1,\ldots,m}\{u_k(x) - v_k(x)\}$ and f, u_k, v_k are quadratic functions with positive coefficients. Then the SIT Algorithm can be applied, using the master problem

$$(Q_\gamma) \qquad \max\{g(x) | f(x) \leq \gamma, \ x \in [a, b]\}.$$

11.4 Discrete Monotonic Optimization

Consider the discrete monotonic optimization problem

$$\max\{f(x) | \ g(x) \leq 0 \leq h(x), \ (x_1, \ldots, x_s) \in S, \ x \in [a, b]\},$$

where $f, g, h : \mathbb{R}^n_+ \to \mathbb{R}$ are increasing functions and S is a given finite subset of $\mathbb{R}^s_+, s \leq n$. Setting

$$G = \{x | \ g(x) \leq 0\}, \ H = \{x | \ h(x) \geq 0\}$$

$$S^* = \{x \in [a, b] | \ (x_1, \ldots, x_s) \in S\},$$

we can write the problem as

$$\text{(DMO)} \qquad \max\{f(x) | \ x \in [a, b], \ x \in (G \cap S^*) \cap H\}.$$

Proposition 11.20 *Let* $\tilde{G} = (G \cap S^*)^{\rceil}$, *the normal hull of the set* $G \cap S^*$. *Then problem* (DMO) *is equivalent to*

$$\max\{f(x)|\, x \in \tilde{G} \cap H\}. \tag{11.24}$$

Proof Since $G \cap S^* \subset \tilde{G}$ we need only show that for each point $\overline{x} \in \tilde{G}$ there exists $z \in G \cap S^*$ satisfying $f(z) \geq f(\overline{x})$. By Proposition 11.4 $\tilde{G} = \cup_{z \in G \cap S^*}[a, z]$, so $\overline{x} \in [a, z]$ for some $z \in G \cap S^*$. Since $\overline{x} \leq z$ and $f(x)$ is increasing we have $f(\overline{x}) \leq f(z)$. □

As a consequence, solving problem (DMO) reduces to solving (11.24) which is a monotonic problem without explicit discrete constraint. The difficulty, now, is how to handle the polyblock $\tilde{G}$ which is defined only implicitly as the normal hull of $G \cap S^*$.

Given a box $[p, q] \subset [a, b]$ and a point $x \in [p, q]$, we define the *lower S-adjustment* of x to be the point

$$\lfloor x \rfloor_S = \tilde{x}, \;\text{ with }\; \tilde{x}_i = \max\{y_i|\, y \in S^* \cup \{p\},\; y_i \leq x_i\},\; i = 1, \ldots, n \tag{11.25}$$

and the *upper S-adjustment* of x to be the point

$$\lceil x \rceil_S = \hat{x}, \;\text{ with }\; \hat{x}_i = \min\{y_i|\, y \in S^* \cup \{q\},\; y_i \geq x_i\},\; i = 1, \ldots, n. \tag{11.26}$$

In the frequently encountered special case when $S = S_1 \times \cdots \times S_s$ and every S_i is a finite set of real numbers we have

$$\tilde{x}_i = \max\{\xi|\, \xi \in S_i \cup \{p_i\},\; \xi \leq x_i\},\; i = 1, \ldots, n.$$

$$\hat{x}_i = \min\{\xi|\, \xi \in S_i \cup \{q_i\},\; \xi \geq x_i\},\; i = 1, \ldots, n.$$

For example, if each S_i is the set of integers and $p_i, q_i \in S_i$ then $\tilde{x}_i$ is the largest integer no larger than x_i, while $\hat{x}_i$ is the smallest integer no less than x_i.

The box obtained from $\text{red}_\gamma[p, q]$ upon the above S-adjustment is referred to as the S-adjusted γ-valid reduction of $[p, q]$ and denoted $\text{red}^S_\gamma[p, q]$. However rather than compute first $\text{red}_\gamma[p, q]$ and then S-adjust it, one should compute directly $\text{red}^S_\gamma[p, q]$ using the formulas $\text{red}^S_\gamma[p, q] = [\tilde{p}, \tilde{q}]$ with

$$\tilde{q}_i = \max\{x_i|\, x \in \text{red}_\gamma[p, q] \cap (G \cap S^*)\}, i = 1, \ldots, n,$$

$$\tilde{p}_i = \min\{x_i|\, x \in \text{red}_\gamma[p, \tilde{q}] \cap (H \cap S^*)\}, i = 1, \ldots, n.$$

With the above modifications to accommodate the discrete constraint the Polyblock Algorithm and the SIT Algorithm for (MO) can be extended to solve (DMO). For every box $[p, q]$ let $\text{red}^S_\gamma[p, q]$ denote the S-adjusted γ-valid reduction of $[p, q]$, i.e., the box obtained from $[p', q'] := \mathrm{red}_\gamma[p, q]$ by replacing p' with $\lceil p' \rceil_S$ and q' with $\lfloor q' \rfloor_S$.

Discrete Polyblock Algorithm

Initialization. Take an initial polyblock $P_0 \supset G \cap H$, with proper vertex set T_0. Let $\hat{x}^0$ be the best feasible solution available (the current best feasible solution), $\gamma_0 = f(\hat{x}^0), S_0^* = \{x \in S^* | \, f(x) > \gamma_0\}$. If no feasible solution is available, let $\gamma_0 = -\infty$. Set $k := 0$.

Step 1. Delete every $v \in T_k$ such that $\mathrm{red}^S_{\gamma_k}[a, v] = \emptyset$. For every $v \in T_k$ such that $\mathrm{red}^S_{\gamma_k}[a, v] \neq \emptyset$ denote the highest vertex of the latter box again by v and if $f(v) \leq \gamma_k$ then drop v. Let $\tilde{T}_k$ be the set that results from T_k after performing this operation for every $v \in T_k$ (so $\tilde{T}_k \subset H$). Reset $T_k := \tilde{T}_k$.

Step 2. If $T_k = \emptyset$ terminate: if $\gamma_k = -\infty$ the problem is infeasible; if $\gamma_k = f(\hat{x}^k), \hat{x}^k$ is an optimal solution.

Step 3. If $T_k \neq \emptyset$ select $v^k \in \mathrm{argmax}\{f(v) | \, v \in T_k\}$. If $v^k \in G \cap S^*$ terminate: v^k is an optimal solution.

Step 4. If $v^k \in G \setminus S^*$ compute $\tilde{v}^k = \lfloor v^k \rfloor_{S_k}$ (using formula (11.25) for $S^* = S_k^*$). If $v^k \notin G$ compute $x^k = \pi_G(v^k)$ and define $\tilde{v}^k = x^k$ if $x^k \in S_k^*$, $\tilde{v}^k = \lfloor x^k \rfloor_{S_k}$ if $x^k \notin S_k^*$.

Step 5. Let $T_{k,*} = \{z \in T_k | \, z > \tilde{v}^k\}$. Compete

$$T'_k = (T_k \setminus T_{k,*}) \bigcap \{z^{k,i} = z + (\tilde{v}^k_i - z_i)e^i | \, z \in T_{k,*}, i = 1, \ldots, n\}.$$

Let T_{k+1} be the set obtained from T'_k by removing every z^i such that $\{j | \, z_j > y_j\} = \{i\}$ for some $z \in T_{k,*}$ and $y \in T_k$.

Step 6. Determine the new current best feasible solution $\hat{x}^{k+1}$. Set $\gamma_{k+1} = f(\hat{x}^{k+1})$, $S^*_{k+1} = \{x \in S^* | f(x) > \gamma_{k+1}\}$. Increment k and return to Step 1.

Proposition 11.21 *If $s = n$ the* DPA *(Discrete Polyblock) algorithm is finite. If $s < n$ and there exists a constant $\alpha > 0$ such that*

$$\min_{i=s+1,\ldots,n} (x_i - a_i) \geq \alpha \quad \forall x \in H, \tag{11.27}$$

then either the DPA *algorithm is finite or it generates an infinite sequence $\{v^k\}$ converging to an optimal solution.*

Proof A each iteration k a pair $v^k, \tilde{v}^k$ is generated such that $v^k \notin (G \cap S_k^*)^\rfloor, \tilde{v}^k \in (G \cap S_k^*)^\rfloor$ and the rectangle $(\tilde{v}^k, b]$ contains no point of P_l with $l > k$ and hence no $\tilde{v}^l$ with $l > k$. Therefore, there can be no repetition in the sequence $\{\tilde{v}^0, \tilde{v}^1, \ldots, \tilde{v}^k, \ldots\} \subset S^*$. This implies finiteness of the algorithm in the case $s = n$ because then $S^* = S$ is then finite. In the case $s < n$, if the sequence $\{v^k\}$ is infinite, it follows from the fact $(\tilde{v}^k_1, \ldots, \tilde{v}^k_s) \in S^*$ that for some sufficiently large k_0,

$$\tilde{v}^k_i = \tilde{v}_i, \quad i = 1, \ldots, s, \quad \forall k \geq k_0.$$

On the other hand, since $v^k \in H$ we have from (11.27) that

$$\min_{i=s+1,\ldots,n} (v^k_i - a_i) \geq \alpha \quad \forall k. \tag{11.28}$$

We show that $v^k - x^k \to 0$ as $k \to +\infty$. Suppose the contrary, that there exist $\eta > 0$ and an infinite sequence k_l such that $\|v^{k_l} - x^{k_l}\| \geq \eta > 0\ \forall l$. For all $\mu > l$ we have $v^{k_\mu} \notin (\tilde{v}^{k_l}, v^{k_l}]$ because $P_{k_\mu} \subset P_{k_l} \setminus (\tilde{v}^{k_l}, b]$. Since $\tilde{v}^k_i = x^k_i\ \forall i > s$, we then derive

$$\|v^{k_\mu} - v^{k_l}\| \geq \min_{i=s+1,\dots,n} |v^{k_l}_i - \tilde{v}^{k_l}_i| = \min_{i=s+1,\dots,n} |v^{k+l}_i - x^{k+l}_i|.$$

By (11.28) $\min\{v^{k_l}_i - a_i |\ i = s+1,\dots,n\} \geq \alpha$ because $v^{k_l} \in H$, while x^{k_l} lies on the line segment joining a to v^{k_l}. Thus

$$v^{k_l}_i - x^{k_l}_i = \frac{v^{k_l}_i - a_i}{\|v^{k_l} - a\|}\|v^{k_l} - x^{k_l}\| \geq \frac{\alpha\eta}{\|b-a\|} \quad \forall i > s.$$

Consequently,

$$\|v^{k_\mu} - v^{k_l}\| \geq \min_{i=s+1,\dots,n} |v^{k_l}_i - x^{k_l}_i\| \geq \frac{\alpha\eta}{\|b-a\|},$$

conflicting with the boundedness of the sequence $\{v^{k_l}\} \subset [a,b]$. Therefore, $\|v^k - x^k\| \to 0$ and by boundedness we may assume that, up to a subsequence, $x^k \to \bar{x}$, $v^k \to \bar{x}$. Then, since $v^k \in H, x^k \in G\ \forall k$, it follows that $\bar{x} \in G \cap H$ and hence, $\overline{v} \in G \cap H \cap S^*$ for $\overline{v}$ defined by $\overline{v}_i = \tilde{v}_i\ (i = 1,\dots,s), \overline{v}_i = \bar{x}_i\ (i = s+1,\dots,n)$. Furthermore, $f(v^k) \geq f(v)\ \forall v \in P_k \supset G \cap H \cap S_k$; hence, letting $k \to +\infty$ yields $f(\bar{x}) \geq f(x)\ \forall x \in G \cap H \cap S_{\overline{\gamma}}$ for $\overline{\gamma} = \lim_{k\to+\infty} \gamma_k$. The latter in turn implies that $\overline{v} \in \operatorname{argmax}\{f(x)|\ x \in G \cap H \cap S\}$, so $\overline{v}$ is an optimal solution. □

Discrete SIT Algorithm

Let $\gamma_0 = f(b)$. Select tolerances $\varepsilon > 0, \eta > 0$.

Initialization. Start with $\mathscr{P}_1 = \{M_1\}, M_1 = [a,b], \mathscr{R}_1 = \emptyset$. If an essential feasible solution $\bar{x}$ of (DMO) is available let $\gamma = f(\bar{x}) - \eta$. Otherwise, let $\gamma = \gamma_0$. set $k = 1$.

Step 1. For each box $M \in \mathscr{P}_k$:

- Compute its S-adjusted γ-valid reduction $\mathrm{red}^S_\gamma M$.
- Replace M by $\mathrm{red}^S_\gamma M$ (in particular, delete it if $\mathrm{red}^S_\gamma M = \emptyset$);
- Let $\mathrm{red}^S_\gamma M = [p^M, q^M]$, take $g(p^M)$ to be $\beta(M)$, a lower bound for $g(x)$ over the box M.
- Delete M if $g(p^M) > -\varepsilon$.

Step 2. Let $\mathscr{P}'_k$ be the collection of boxes that results from $\mathscr{P}_k$ after completion of Step 1 and let $\mathscr{R}'_k = \mathscr{R}_k \cup \mathscr{P}'_k$.

Step 3. If $\mathscr{R}'_k = \emptyset$ terminate: in the case $\gamma = \gamma_0$ the problem (DMO) is ε-essential infeasible; otherwise, $\gamma = f(\bar{x}) - \eta$ then $\bar{x}$ is an essential (ε,η)-optimal solution of (DMO).

Step 4. If $\mathscr{R}'_k \neq \emptyset$, let $[p^k, q^k] = M_k \in \operatorname{argmin}\{\beta(M)|\ M \in \mathscr{R}'_k\}$. If $k_\gamma(p^k) \geq 0$, let $y^k = p^k$ and go to Step 5. If $k_\gamma(p^k) < 0$ let y^k be the upper S-adjustement of the point where the line segment joining p^k to q^k meets the surface $k_\gamma(x) = 0$. If $g(y^k) \leq -\varepsilon$ go to Step 5; otherwise go to Step 6.

Step 5. y^k is an essential feasible solution of (DMO) with $f(y^k) \geq \gamma$. If $\gamma = \gamma_0$ define $\bar{x} = y^k, \gamma = f(\bar{x}) - \eta$ and go back to Step 0. If $\gamma = f(\bar{x}) - \eta$ and $f(y^k) > f(\bar{x})$ reset $\bar{x} \leftarrow y^k$ and go to Step 0.

Step 6. Divide M_k into two subboxes according to an adaptive branching via (p^k, y^k). Let $\mathscr{P}_{k+1}$ be the collection of these two subboxes of M_k. Let $\mathscr{R}_{k+1} = \mathscr{R}'_k \setminus \{M_k\}$. Increment k and return to Step 1.

Proposition 11.22 *The Discrete SIT Algorithm terminates after finitely many iterations, yielding either an essential* (ε, η)*-optimal solution, or an evidence that problem (DMO) is* ε*-essential infeasible.*

Proof Analogous to the proof of Proposition 11.19. □

11.5 Problems with Hidden Monotonicity

Many problems encountered in various applications have a hidden monotonic structure which can be disclosed upon suitable variable transformations. In this section we study two important classes of such problems.

11.5.1 Generalized Linear Fractional Programming

This is a class of problems with hidden monotonicity in the objective function. It includes problems of either of the following forms:

$$(P) \qquad \max\{\Phi\left(\frac{f_1(x)}{g_1(x)}, \ldots, \frac{f_m(x)}{g_m(x)}\right) \mid x \in D\}$$

$$(Q) \qquad \min\{\Phi\left(\frac{f_1(x)}{g_1(x)}, \ldots, \frac{f_m(x)}{g_m(x)}\right) \mid x \in D\}$$

where D is a nonempty polytope in $\mathbb{R}^n$,$f_1, \ldots, f_m, g_1, \ldots, g_m$ are linear affine functions on $\mathbb{R}^n$ such that

$$-\infty < a_i := \min_{x \in D} \frac{f_i(x)}{g_i(x)} < +\infty, \quad i = 1, \ldots, m, \tag{11.29}$$

while $\Phi : \mathbb{R}^m \to \mathbb{R}$ is a continuous function, *increasing* on $\mathbb{R}^m_{a+} := \{y \in \mathbb{R}^m | \, y_i \geq a_i, i = 1, \ldots, m\}$, i.e., satisfying

$$a_i \leq y'_i \leq y_i \;\; (i = 1, \ldots, m) \Rightarrow \Phi(y') \leq \Phi(y). \tag{11.30}$$

This class of problems include, aside from linear multiplicative programs (see Chap. 8, Sect. 8.5), diverse generalized linear fractional programs, namely:

1. *Maximin and Minimax*

$$\max\{\min\left(\frac{f_1(x)}{g_1(x)},\dots,\frac{f_m(x)}{g_m(x)}\right) \mid x \in D\} \tag{11.31}$$

$$(\Phi(y) = \min\{y_1,\dots,y_m\})$$

$$\min\{\max\left(\frac{f_1(x)}{g_1(x)},\dots,\frac{f_m(x)}{g_m(x)}\right) \mid x \in D\} \tag{11.32}$$

$$(\Phi(y) = \max\{y_1,\dots,y_m\})$$

2. *Maxsum and Minsum*

$$\max\{\sum_{i=1}^m \frac{f_1(x)}{g_1(x)} \mid x \in D\} \quad (\Phi(y) = \sum_{i=1}^m y_i) \tag{11.33}$$

$$\min\{\sum_{i=1}^m \frac{f_1(x)}{g_1(x)} \mid x \in D\} \quad (\Phi(y) = \sum_{i=1}^m y_i) \tag{11.34}$$

3. *Maxproduct and Minproduct*

$$\max\{\prod_{i=1}^m \frac{f_1(x)}{g_1(x)} \mid x \in D\} \quad (\Phi(y) = \prod_{i=1}^m y_i) \tag{11.35}$$

$$\min\{\prod_{i=1}^m \frac{f_1(x)}{g_1(x)} \mid x \in D\} \quad (\Phi(y) = \prod_{i=1}^m y_i) \tag{11.36}$$

(For the last two problems it is assumed that $a_i > 0, i = 1,\dots,m$ in condition (11.29))

Standard linear fractional programs which are special cases of problems (P), (Q) when $m = 1$ were first studied by Charnes and Cooper as early as in 1962. Problems (11.31) and (11.33) received much attention from researchers (Schaible 1992 and the references therein); problems (11.33) and (11.34) were studied in Falk and Palocsay (1992), Konno and Abe (1999), Konno et al. (1991), and problems (11.35), (11.36) in Konno and Abe (1999) and Konno and Yajima (1992).

For a review of fractional programing and extensions up to 1993, the reader is referred to Schaible (1992), where various potential and actual applications of fractional programming are also described. In general, fractional programming arises in situations where one would like to optimize functions of such ratios as return/investment, return/risk, cost/time, output/input, etc. Let us also mention the book (Stancu-Minasian 1997) which is a comprehensive presentation of the theory, methods and applications of fractional programming until this date.

From a computational point of view, so far different methods have been proposed for solving different variants of problems (P) and (Q). A general unifying idea of parametrization underlies the methods developed by Konno and his associates (Konno et al. 1991; Konno and Yajima 1992; Konno and Yamashita (1997)). Each of these methods is devised for a specific class of problems, and most of them work quite well for problem instances with $m \leq 3$.

In the following we present a unified approach to all variants of problems (P) and (Q) which has been developed in Hoai-Phuong and Tuy (2003), based on an application of monotonic optimization.

a. Conversion to Monotonic Optimization

First note that in the study of problems (11.31)–(11.34) it is usually assumed that

$$\min_{x\in D} f_i(x) \geq 0, \quad \min_{x\in D} g_i(x) > 0, \quad i = 1,\ldots,m, \tag{11.37}$$

a condition generally weaker than (11.29). Further, without loss of generality we can always assume in problems (P) and (Q) that $\Phi : \mathbb{R}^n_+ \to \mathbb{R}_{++}$ and

$$\min\{g_i(x), f_i(x)\} > 0 \quad \forall x \in D, \quad i = 1,\ldots,m. \tag{11.38}$$

In fact, in view of (11.29) $g_i(x)$ does not vanish on D, hence has a constant sign on D and by replacing f_i, g_i with their negatives if necessary, we can assume $g_i(x) > 0\ \forall x \in D, i = 1,\ldots,m$. Then setting $\tilde{f}_i(x) = f_i(x) - a_i g_i(x)$ we have, by (11.29), $\tilde{f}_i(x) > 0\ \forall x \in D, i = 1,\ldots,m$, and since $\frac{\tilde{f}_i(x)}{g_i(x)} + a_i = \frac{f_i(x)}{g_i(x)}$, the problem (P) is equivalent to

$$\max\{\Phi\left(\frac{\tilde{f}_i(x)}{g_i(x)},\ldots,\frac{\tilde{f}_m(x)}{g_m(x)}\right) \mid x \in D\},$$

where $\tilde{f}_i(x) > 0\ \forall x \in D, i = 1,\ldots,m$. Setting $\tilde{\Phi}(y_1,\ldots,y_m) = \Phi(y_1 + a_1,\ldots, y_m + a_m)$ one has, in view of (11.30), an increasing function $\tilde{\Phi} : \mathbb{R}^m_+ \to \mathbb{R}$, so by adding a positive constant one can finally assume $\tilde{\Phi} : \mathbb{R}^m_+ \to \mathbb{R}_{++}$. Analogously for problem (Q).

Now define the sets

$$G = \{y \in \mathbb{R}^m_+ \mid y_i \leq \tfrac{f_i(x)}{g_i x)}, (i = 1,\ldots,m), \quad x \in D\} \tag{11.39}$$

$$H = \{y \in \mathbb{R}^m_+ \mid y_i \geq \tfrac{f_i(x)}{g_i x)}, (i = 1,\ldots,m), \quad x \in D\}. \tag{11.40}$$

Clearly, G is a normal set, H a conormal set and both are contained in the box $[0, b]$ with

$$b_i = \max_{x\in D} \frac{f_i(x)}{g_i(x)}, \quad i = 1,\ldots,m.$$

Proposition 11.23 *Problems (P) and (Q) are equivalent, respectively, to the following problems:*

$$\text{(MP)} \qquad \max\{\Phi(y) \mid y \in G\}$$

$$\text{(MQ)} \qquad \min\{\Phi(y) \mid y \in H\}.$$

More precisely, if $\overline{x}$ solves (P) ((Q), resp.) then $\overline{y}$ with $\overline{y}_i = \frac{f_i(\overline{x})}{g_i(\overline{x})}$ solves (MP) ((MQ), *resp.). Conversely, if $\overline{y}$ solves* (MP) ((MQ), *resp.) and $\overline{x}$ satisfies $\overline{y}_i \leq \frac{f_i(\overline{x})}{g_i(\overline{x})}$ ($\overline{y}_i \geq \frac{f_i(\overline{x})}{g_i(\overline{x})}$, resp.) then $\overline{x}$ solves (P) ((Q), resp.).*

Proof Suppose $\overline{x}$ solves (P). Then $\overline{y} = (\frac{f_1(\overline{x}}{g_i(\overline{x})}, \ldots, \frac{f_m(\overline{x})}{g_m(\overline{x})})$ satisfies $\Phi(\overline{y}) \geq \Phi(y)$ for all y satisfying $y_i = \frac{f_i(x)}{g_i(x)}, i = 1, \ldots, m$, with $x \in D$. Since Φ is increasing this implies $\Phi(\overline{y}) \geq \Phi(y)\ \forall y \in G$, hence $\overline{y}$ solves (MP). Conversely, if $\overline{y}$ solves (MP) then $\overline{y}_i \leq \frac{f_i(\overline{x})}{g_i(\overline{x})}, i = 1 \ldots, m$, for some $\overline{x} \in D$, and since Φ is increasing, $\Phi(\frac{f_1(\overline{x})}{g_1(\overline{x})}, \ldots, \frac{f_m(\overline{x})}{g_m(\overline{x})}) \geq \Phi(\overline{y}) \geq \Phi(y)$ for every $y \in G$, i.e., for every y satisfying $y_i = \frac{f_i(x)}{g_i(x)}, i = 1, \ldots, m$ with $x \in D$. This means $\overline{x}$ solves (P). $\square$

Thus generalized linear fractional problems (P), (Q) in $\mathbb{R}^n$ can be converted into monotonic optimization problems (MP), $M(Q)$ respectively, in $\mathbb{R}^m_+$ with m usually much smaller than n.

b. Solution Method

We only briefly discuss a solution method for problem (MP) (a similar method works for (MQ)):

$$\text{(MP)} \qquad\qquad \max\{\Phi(y) |\ y \in G\}$$

where (see (11.39)) $G = \{y \in \mathbb{R}^m_+ |\ y_i \leq \frac{f_i(x)}{g_i(x)}\ \forall i = 1, \ldots, m,\ x \in D\}$ is contained in the box $[0, b]$ with b satisfying

$$b_i = \max_{x \in D} \frac{f_i(x)}{g_i(x)} \quad i = 1, \ldots, m. \tag{11.41}$$

In view of assumption (11.38) we can select a vector $a \in \mathbb{R}^m_{++}$ such that

$$a_i = \min_{x \in D} \frac{f_i(x)}{g_i(x)} > 0 \quad i = 1, \ldots, m, \tag{11.42}$$

where it can be assumed that $a_i < b_i, i = 1, \ldots, m$ because if $a_i = b_i$ for some i then $\frac{f_i(x)}{g_i(x)}\ \forall x \in D$ and $\Phi(.)$ reduces to a function of $m - 1$ variables. So with a, b defined as in (11.41) and (11.42) the problem to be solved is

$$\max\{\Phi(y) |\ y \in G \cap [a, b]\}, \quad [a, b] \subset \mathbb{R}^n_{++}. \tag{11.43}$$

Let γ be the optimal value of (MP) ($\gamma > 0$ since $\Phi : \mathbb{R}^n_+ \to \mathbb{R}_{++}$). Given a tolerance $\varepsilon > 0$, we say that a vector $\overline{y} \in G$ is an *ε-optimal solution* of (MP) if $\gamma \leq (1 + \varepsilon)\Phi(\overline{y})$, or equivalently, if $\Phi(y) \leq (1 + \varepsilon)\Phi(\overline{y})\ \forall y \in G$. The vector $\overline{x} \in D$ satisfying $\overline{y} = \frac{f_i(\overline{x})}{g_i(x)}$ is then called an *ε-optimal solution* of (P).

The POA algorithm specialized to problem (11.43) reads as follows:
POA algorithm (tolerance $\varepsilon > 0$)
Initilization. Start with $T_1 = \tilde{T}_1 = \{b\}$. Set $k = 1$.

Step 1. Select $z^k \in \operatorname{argmax}\{\Phi(z)|\, z \in \tilde{T}_k\}$. Compute $y^k = \pi_G(z^k)$, the point where the halfline from a through z^k meets the upper boundary $\partial^+ G$ of G. Let $x^k \in D$ satisfy $y_i^k \le \frac{f_i(x^k)}{g_i(x^k)}, i = 1, \dots, m$. Define $\tilde{y}^k \in G$ by $\tilde{y}_i^k = \frac{f_i(x^k)}{g_i(x^k)},\ i = 1, \dots, m$ (so $\tilde{y}^k$ is a feasible solution). Determine the new current best solution $\overline{y}^k$ (with corresponding $\overline{x}^k \in D$) and the new current best value $CBV = \Phi(\overline{y}^k)$ by comparing $\tilde{y}^k$ with $\overline{y}^{k-1}$ (for $k = 1$ set $\overline{y}^1 = \tilde{y}^1, \overline{x}^1 = x^1$).

Step 2. Select a set $T'_k \subset T_k$ such that $z^k \in T'_k \subset \{z|\ z > y^k\}$ and let T^*_k be the set obtained from T_k by replacing each $z \in T'_k$ with the points $z^{*i} = z - (z_i - y_i^k)e^i,\ i = 1, \dots, m$.

Step 3. From T^*_k remove all improper elements, all $z \in \tilde{T}_k$ such that $\Phi(z) < (1+\varepsilon)CBV$ and also all $z \not\geq a$. Let T_{k+1} be the set of remaining elements. If $T_{k+1} = \emptyset$, terminate: $\overline{y}^k$ is an ε-optimal solution of (MP), and the corresponding $\overline{x}^k$ an ε-optimal solution of (P).

Step 4. If $T_{k+1} \neq \emptyset$, set $k \leftarrow k + 1$ and return to Step 1.

c. Implementation Issues

A basic operation in Step 1 is to compute the point $\pi_G(z)$, where the halfline from a through z meets the upper boundary of the normal set G. Since $G \subset [0, b]$ we have $\pi_G(z) = \lambda z$ where

$$\begin{aligned}\lambda &= \max\{\alpha|\ \alpha \le \min_{i=1,\dots,m} \frac{f_i(x)}{z_i g_i(x)}, x \in D\}\\ &= \max_{x\in D} \min_{i=1,\dots,m} \frac{f_i(x)}{z_i g_i(x)}.\end{aligned} \tag{11.44}$$

So computing $\pi_G(z)$ amounts to solving a problem (11.31). Since this is computationally expensive, instead of computing $\pi_G(z) = \lambda z$ one should compute only an approximate value of $\pi_G(z)$, i.e., a vector $y \approx \pi_G(z)$, such that

$$y = (1+\eta)\hat{\lambda} z, \text{ with } \hat{\lambda} \le \lambda \le (1+\eta)\hat{\lambda}, \tag{11.45}$$

where $\eta > 0$ is a tolerance. Of course, to guarantee termination of the above algorithm, the tolerance η must be suitably chosen. This is achieved by choosing $\eta > 0$ such that

$$\Phi((1+\eta)\hat{\lambda} z) < (1+\varepsilon/2)\Phi(\hat{\lambda} z), \tag{11.46}$$

i.e., $\Phi(y) \le (1+\varepsilon/2)\Phi(\pi_G(z))$. Indeed, since $\hat{\lambda} z \le \pi_G(z) \le y := (1+\eta)\hat{\lambda} z$, while $\Phi(y)$ is continuous, it follows that, for sufficiently large k, $\Phi(z^k) - \Phi(y^k) \le \varepsilon/2\Phi(\pi_G(z^k))$. Then, noting that $\Phi(\tilde{y}^k) \ge \Phi(\pi_G(z^k))$ because $\Phi(y)$ is increasing, we will have

$$\begin{aligned}
(1+\varepsilon)CBV &\ge (1+\varepsilon)\Phi(\tilde{y}^k) \ge (1+\varepsilon)\Phi(\pi_G(z^k)) \\
&\ge (1+\varepsilon/2)\Phi(\pi_G(z^k)) + (\varepsilon/2)\Phi(\pi_G(z^k)) \\
&\ge \Phi(y^k) + (\varepsilon/2)\Phi(\pi_G(z^k)) \ge \Phi(y^k) + [\Phi(z^k) - \Phi(y^k)] \\
&= \Phi(z^k),
\end{aligned}$$

and so the termination criterion will hold.

Consider the subproblem

$$\mathrm{LP}(\alpha) \qquad R(\alpha) := \max_{x \in D} \min_{i=1,\dots,m} (f_i(x) - \alpha z_i g_i(x))$$

which is equivalent to the linear program

$$\max\{t \mid t \le f_i(x) - \alpha z_i g_i(x), i = 1,\dots,m,\ x \in D\}.$$

As has been proved in Hoai-Phuong and Tuy (2003), one way to ensure (11.46) is to compute $y \approx \pi_G(z)$ by the following subroutine:

Procedure $\pi_G(z)$

Step 1. Set $s = 1, \alpha_1 = 0$. Solve $\mathrm{LP}(\alpha_1)$, obtaining an optimal solution x^1 of it. Define $\tilde{\alpha}_1 = \min_i \frac{f_i(x^1)}{z_i g_i(x^1)}$.

Step 2. Let $\alpha_{s+1} = \tilde{\alpha}_s(1+\eta)$. Solve $\mathrm{LP}(\alpha_{s+1})$, obtaining an optimal solution x^{s+1} of it. If $R(\alpha_{s+1}) \le 0$, and $\Phi(\alpha_{s+1} z) \le (1+\varepsilon/2)\Phi(\tilde{\alpha}_s z)$, terminate: $y = \alpha_{s+1} z$, $\tilde{y}_i^s = f_i(x^{s+1})/g_i(x^{s+1}), i = 1,\dots,m$.

Step 3. If $R(\alpha_{s+1}) \le 0$ but $\Phi(\alpha_{s+1} z) > (1+\varepsilon/2)\Phi(\tilde{\alpha}_s z)$ reset $\eta \leftarrow \eta/2$, $\tilde{\alpha}_{s+1} = \min_i \frac{f_i(x^{s+1})}{z_i g_i(x^{s+1})}$, increment s and go back to Step 2.

Step 4. If $R(\alpha_{s+1}) > 0$, set $\tilde{\alpha}_{s+1} = \min_i \frac{f_i(x^{s+1})}{z_i g_i(x^{s+1})}$, increment s and go back to Step 2.

11.5.2 Problems with Hidden Monotonicity in Constraints

This class includes problems of the form

$$\min\{f(x) \mid \Phi(C(x)) \ge 0,\ x \in D\}, \tag{11.47}$$

where $f(x)$ is an increasing function, $\Phi : \mathbb{R}^m_+ \to \mathbb{R}$ is an u.s.c. increasing function, D a compact convex set and $C : \mathbb{R}^n \to \mathbb{R}^m_+$ a continuous map.

The monotonic structure in (11.47) is easily disclosed. Since the set $C(D) := \{y = C(x), x \in D\}$ is compact it is contained in some box $[a, b] \subset \mathbb{R}^m_+$. For each $y \in [a, b]$ let

$$R(y) \qquad\qquad \varphi(y) := \min\{f(x)|\ C(x) \geq y,\ x \in D\}.$$

Proposition 11.24 *The problem (11.47) is equivalent to the monotonic optimization problem*

$$\min\{\varphi(y)|\ \Phi(y) \geq 0\} \tag{11.48}$$

in the sense that if y^* solves (11.48) then an optimal solution x^* of $R(y^*)$ solves (11.47) and conversely, if x^* solves (11.47) then $y^* = C(x^*)$ solves (11.48).

Proof That (11.48) is a monotonic optimization problem is clear from the fact that $\varphi(y)$ is an increasing function by the definition of $R(y)$ while $\Phi(y)$ is an increasing function by assumption. The equivalence of the two problems is easily seen by writing (11.47) as

$$\begin{aligned}
&\inf_{x,y}\{f(x)|\ x \in D,\ C(x) \geq y,\ y \in [a, b],\ h(y) \geq 0\}\\
&= \min_{y\in[a,b],h(y)\geq 0}\ \min_{x}\{f(x)|\ x \in D,\ C(x) \geq 0\}\\
&= \min\{\varphi(y)|\ y \in [a, b],\ h(y) \geq 0\}.
\end{aligned}$$

In more detail, let y^* be an optimal solution of (11.48) and let x^* be an optimal solution of $R(y^*)$, i.e., $\varphi(y^*) = f(x^*)$ with $x^* \in D, C(x^*) \geq y^*$. For any $x \in D$ satisfying $\Phi(C(x)) \geq 0$, we have $\Phi(y) \geq 0$, for $y = C(x) \geq \varphi(y) \geq \varphi(y^*) = f(x^*)$, proving that x^* solves (11.47). Conversely, let x^* be an optimal solution of (11.47), i.e., $x^* \in D, \Phi(y^*) \geq 0$ with $y^* = C(x^*)$. For any $y \in [a, b]$ satisfying $\Phi(y) \geq 0$, we have $\varphi(y) = f(x)$ for some $x \in D$ with $C(x) \geq y$, hence $\Phi(C(x)) \geq \Phi(y) \geq 0$, i.e., x is feasible to (11.47), and hence, $\varphi(y) = f(x) \geq f(x^*) \geq \varphi(y^*)$, i.e., y^* is an optimal solution of (11.48). □

Note that the objective function $\varphi(y)$ in the monotonic optimization problem (11.48) is given implicitly as the optimal value of a subproblem $R(y)$. However, in many cases of interest, $R(y)$ is solvable by standard methods and no difficulty can arise from that. Consider, for example, the problem

$$\min\{f(x)|\ x \in D,\ \sum_{i=1}^{m} \frac{u_i(x)}{v_i(x)} \geq 1\},$$

where D is a polytope, $f(x)$ is an increasing function, $u_i(x), v_i(x), i = 1, \ldots, m$, are continuous functions such that $v_i(x) > 0\ \forall x \in D$ and the set $\{y|\ y_i = \frac{u_i(x)}{v_i(x)}, i = 1, \ldots, m,\ x \in D\}$ is contained in a box $[a, b] \subset \mathbb{R}^m_+$.

Setting $C_i(x) = \frac{u_i(x)}{v_i(x)}$, $\Phi(y) = \sum_{i=1}^m y_i - 1$, we have

$$\varphi(y) = \min\{f(x) | \, x \in D, \; y_i v_i(x) \le u_i(x), \; i = 1, \ldots, n\}$$

so the equivalent monotonic optimization problem is

$$\min\{\varphi(y) | \, \sum_{i=1}^m y_i \ge 1, \; y \in [a, b]\}.$$

11.6 Applications

Monotonic optimization has met with important applications in economics, engineering, and technology, especially in risk and reliability theory and in communication and networking systems.

11.6.1 Reliability and Risk Management

As an example of problem with hidden monotonicity in the constraints consider the following problem which arises in reliability and risk management (Prékopa 1995):

$$\min\{\langle c, x\rangle | \, Ax = b, \; x \in \mathbb{R}^n_+, \; P\{Tx \ge \xi(\omega)\} \ge \alpha\}, \tag{11.49}$$

where $C \in \mathbb{R}^n, A \in \mathbb{R}^{p \times n}, b \in \mathbb{R}^p, T \in \mathbb{R}^{m \times n}$ are given deterministic data, ω is a random vector from the probability space $(\Omega, \Sigma, \mathscr{P})$ and $\xi : \Omega \to \mathbb{R}^m$ represent stochastic right-hand side parameters; $P\{S\}$ represent the probability of the event $S \in \Sigma$ under the probability measure $\mathscr{P}$ and $\alpha \in (0, 1)$ is a scalar. This is a linear program with an additional probabilistic constraint which requires that the inequalities $Tx \ge \xi(\omega)$ hold with a probability of at least α.

If the stochastic parameters have a continuous log-concave distribution, then problem (11.49) is guaranteed to have a convex feasible set (Prékopa 1973), and hence may be solvable with standard convex programming techniques (Prékopa 1995). For general distributions, in particular for discrete distributions, the feasible set is nonconvex.

For every $y \in \mathbb{R}^m$ define $f(y) := \inf\{\langle c, x\rangle | \, Ax = b, \; x \in \mathbb{R}^n_+, \; Tx \ge y\}$. Let $F : \mathbb{R}^m \to [0, 1]$ be the cumulative density function of the random vector $\xi(\omega)$, i.e., $F(y) = P\{y \ge \xi(\omega)\}$. Assuming $p \le y \le q$ with $-\infty < f(y) < +\infty \; \forall y \in [p, q]$ problem (11.49) can then be rewritten as

$$\min\{f(y) | \, F(y) \ge \alpha, \; y \in [p, q]\}. \tag{11.50}$$

As is easily seen, $f(y)$ is a continuous convex and increasing function while $F(y)$ is an u.s.c. increasing function, so this is a monotonic optimization problem equivalent to (11.49). For a study of problem (11.49) based on its monotonic reformulation (11.50) see, e.g., Cheon et al. (2006).

11.6.2 Power Control in Wireless Communication Networks

We briefly describe a general monotonic optimization framework for solving resource allocation problems in communication and networking systems, as developed in the monographs (Bjōrnson and Jorswieck 2012; Zhang et al. 2012) (see also Bjōrnson et al. 2012; Utschick and Brehmer 2012; Jorswieck and Larson 2010; Qian and Jun Zhang 2010; Qian et al. 2009, 2012). For details the reader is referred to these works and the references therein.

Consider a wireless network with a set of n distinct links. Each link includes a transmitter node T_i and receiver node R_i. Let G_{ij} be the channel gain between node T_i and node R_j; f_i the transmission power of link i (from T_i to R_i); η_i the received noise power on link i; Then the received signal to interference-plus-noise ratio (SINR) of link i is

$$\gamma_i = \frac{G_{ii}f_i}{\sum_{j \neq i} G_{ij}f_j + \eta_i}$$

If $p = (f_1, \ldots, f_n)$, $\gamma(p) = (\gamma_1(p), \ldots, \gamma_n(p))$ and $U(\gamma(p))$ is the system utility then the problem we are concerned with can be formulated as

$$\left.\begin{array}{ll} \max & U(\gamma(p)), \\ \text{s.t.} & \gamma_i(p) \geq \gamma_{i,\min} \quad i = 1, \ldots, n; \\ & 0 \leq f_i \leq f_i^{\max} \quad i = 1, \ldots, n. \end{array}\right| \tag{11.51}$$

By its very nature $U(\gamma(p))$ is an increasing function of $\gamma(.)$, so this is a problem of Generalized Linear Fractional Programming as discussed in the previous section.

In the more general case of MISO (multi-input-single-output) links, each link includes a multi-antenna transmitter T_i and a single-antenna receiver R_i. The channel gain between node T_i and node R_j is denoted by vector h_{ij}. If w_i is the beamforming vector of T_i which yields a transmit power level of $f_i = \|w_i\|_2^2$, and η_i s the noise power at node R_i, then the SINR of link i is

$$\gamma_i(w_1, \ldots, w_n) = \frac{|h_{ij}^H w_i|^2}{\sum_{j \neq i} |h_{ij}^Y w_j|^2 + \eta_i}.$$

As before, by $U(\gamma)$ denote the system utility function which monotonically increases with $\gamma := (\gamma_1, \ldots, \gamma_n)$. Then the problem is to

$$\left.\begin{array}{ll} \max\ U(\gamma(w)) & \\ \text{s.t.}\quad \gamma_i(w_1,\ldots,w_n) \geq \gamma_{i,\min} & i=1,\ldots,n; \\ \qquad 0 \leq \|w_i\|_2^2 \leq P_i^{\max} & i=1,\ldots,n. \end{array}\right| \tag{11.52}$$

where $\gamma_{i,\min}$ is the minimum SINR requirement of link i and $P_i^{\max}$ is the maximum transmit power of link i.

Again problem (11.52) can be converted into a canonical monotonic optimization problem. In fact, setting

$$G = \{y|\ 0 \leq y_i \leq \gamma_i(w_1,\ldots,w_n), 0 \leq \|w_i\|_2^2 \leq P_i^{\max}, i=1,\ldots,n\}$$
$$H = \{y|\ y_i \geq \gamma_{i,\min}, i=1,\ldots,n\}$$

we can rewrite problem (11.52) as

$$\max\{U(y)|\ y \in G \cap H\},$$

where G is a closed normal set, H a closed conormal set.

Remark 11.5 In Bjōrnson and Jorswieck (2012) and Zhang et al. (2012) the monotonic optimization problem was solved by the Polyblock Algorithm or the BRB Algorithm. Certainly the performance could have been much better with the SIT Algorithm, had it been available at the time. Once again, it should be emphasized that not only is the SIT algorithm more efficient, but it also suits best the user's purpose if we are mostly interested in quickly finding a feasible solution or, possibly, a feasible solution better than a given one.

11.6.3 Sensor Cover Energy Problem

An important optimization problem that arises in wireless communication about sensor coverage energy (Yick et al. 2008) can be described as follows:

Consider a wireless network deployed in a certain geographical region, with n sensors at points $s^i \in \mathbb{R}^p, i = \{1,\ldots,n\}$, and m target nodes at points $t^j \in \mathbb{R}^p, j = \{1,\ldots,m\}$, where p is generally set to 2 or 3. The energy consumption per unit time of a sensor i is a monotonically non decreasing function of its sensing radius r_i. Denoting this function by $E_i(r_i)$ assume, as usual in many studies on wireless sensor coverage, that

$$E_i(r_i) = \alpha_i r_i^{\beta_i} + \gamma, \quad l_i \leq r_i \leq u_i,$$

where $\alpha_i > 0, \beta_i > 0$ are constants depending on the specific device and γ represents the idle-state energy cost. The problem is to determine $r_i \in [l_i, u_i] \subset \mathbb{R}_+, i = 1,\ldots,n$, such that each target node $t^j, j = 1,\ldots,m$, is covered by at least one sensor and the total energy consumption is minimized. That is,

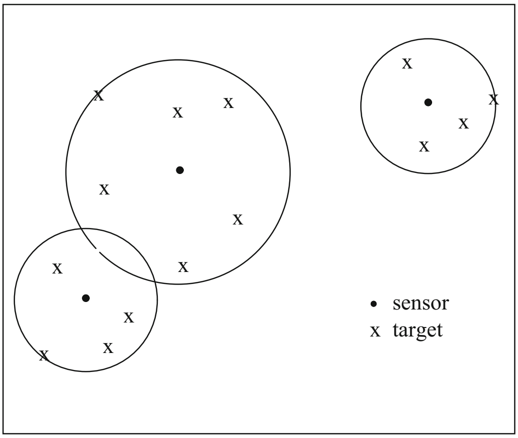


Fig. 11.4 Sensor cover energy problem (n=3, m=14)

$$\text{(SCEP)} \qquad \begin{array}{ll} \min & \sum_{i=1}^{n} E_i(r_i) \\ \text{s.t.} & \max_{1\le i\le n}(r_i - \|s^i - t^j\|) \ge 0 \quad j = 1,\dots,m, \\ & l \le r \le u, \end{array}$$

where the last inequalities are component-wise understood (Fig. 11.4).

Setting $x_i = \alpha_i r_i^{\beta_i}, a_{ij} = \alpha_i \|s^i - t^j\|^{\beta_i}$ and $a_i = \alpha_i l_i^{\beta_i}, b_i = \alpha_i u_i^{\beta_i}$ we can rewrite (SCEP) as

$$\left.\begin{array}{l} \text{minimize} f(x) = \sum_{i=1}^{n} x_i \text{ subject to} \\ g_j(x) := \max_{1\le i\le n}(x_i - a_{ij}) \ge 0, \quad j = 1,\dots,m, \\ a \le x \le b. \end{array}\right| \tag{11.53}$$

Clearly each $g_j(x), j = 1,\dots,m$, is a convex function, so the problem is a linear optimization under a reverse convex constraint and can be solved by methods of reverse convex programming (Chap. 7, Sect. 7.3). In Astorino and Miglionico (2014), the authors use a penalty function to shift the reverse convex constraint to the objective function, transforming (11.53) into an equivalent dc optimization problem which is then solved by the DCA algorithm of P.D. Tao. Aside from the need of a penalty factor which is not simple to determine even approximately, a drawback of this method is that it can at best provide an approximate local optimal solution which is not guaranteed to be global.

A better approach to problem (SCEP) is to view it as a monotonic optimization problem. In fact, each $g_j(x), j = 1,\dots,m$, is an increasing function, i.e., $g_j(x) \ge g_j(x')$ whenever $x \ge x'$. So the feasible set of problem (11.53) :

$$H := \{x \in [a,b] | \; g_j(x) \ge 0, j = 1,\dots,m\}$$

is a conormal set (see Sect. 11.1) and (SCEP) is actually a standard monotonic optimization problem: minimizing the increasing function $f(x) = \sum_{i=1}^{n} x_i$ over the conormal set H.

Let $a^j = (a_{1j}, \ldots, a_{nj}),\ J := \{j = 1, \ldots, m|\ a \leq a^j\}$. We have

$$H = \cap_{j\in J} H_j,$$

where $H_j = \{x \in [a, b]|\ \max_{1\leq i\leq n}(x_i - a_{ij}) \geq 0\}$. Clearly $H_j = [a, b] \setminus C_j$ with $C_j = \{x \in [a, b]|\ \max_{1\leq i\leq n}(x_i - a_{ij}) < 0\}$. For every $j \in J$, since the closure of C_j is $\overline{C}_j = \{x|\ a \leq x \leq a^j\} = [a, a^j]$, the set

$$G := \cup_{j\in J}\overline{C}_j = \cup_{j\in J}[a, a^j]$$

is a polyblock. Let J_1 be the set of all $j \in J$ for which a^j is a proper vertex of the polyblock G, i.e., for which there is no $k \in J, k \neq j$ such that $a^k \geq a^j$. Then $G = \cup_{j\in J_1}[a, a^j]$, hence $\cup_{j\in J} C_j = \cup_{j\in J_1} C_j$. Since

$$H = \cap_{j\in J} H_j = [a, b] \setminus \cup_{j\in J} C_j = [a, b] \setminus \cup_{j\in J_1} C_j$$

it follows that

$$H = \{x \in [a, b]|\ g_j(x) = \max_{1\leq i\leq n}(x_i - a_{ij}) \geq 0,\ \forall j \in J_1\} = \cap_{j\in J_1} H_j.$$

Proposition 11.25 *H is a copolyblock in $[a, b]$ and (SCEP) reduces to minimizing the increasing function $f(x) = \sum_{i=1}^n x_i$ over this copolyblock.*

Proof We have $H_j = [a, b]| \setminus C_j$, where $C_j = \{x \in [a, b]|\ \max_{1\leq i\leq n}(x_i - a_{ij} < 0\} = [a, b] \setminus [a, a^j) = \cup_{1\leq i\leq n}[u^{ij}, b]$, with $u^{ij} = a + (a_{ij} - a_i)e^i$ (Proposition 11.7), so each C_j is a copolyblock with vertex set $\{u^{1j}, \ldots, u^{nj}\}$. Hence $H = \cap_{j\in J_1} H_j$ is a copolyblock (Proposition 11.6). □

Let V denote the proper vertex set of the copolyblock H. Since each $v \in V$ is a local minimizer of $f(x) = \sum_{i=1}^n x_i$ over H, we thus have a simple procedure for computing a local optimal solution of (SCEP). Starting from such a local optimal solution, a BRB Algorithm can then be applied to solve the discrete monotonic optimization

$$\min\{\sum_{i=1}^n x_i|\ x \in V\}$$

which is equivalent to (SCEP). This is essentially the method recently developed in Hoai and Tuy (2016) for solving SCEP. For detail on the implementation of the method and computational experiments with problems of up to 1000 variables, the reader is referred to the cited paper.

11.6.4 Discrete Location

Consider the following discrete location problem:

(DL) Given m balls in $\mathbb{R}^n$ of centers $a^i \in \mathbb{R}^n_{++}$ and radii $\alpha_i (i = 1, \ldots, m)$, and a bounded discrete set $S \subset \mathbb{R}^n_+$, find the largest ball that has center in S and is disjoint from any of these m balls. In other words,

$$\left. \begin{array}{l} \text{maximize } z \text{ subject to} \\ \|x - a^i\| - \alpha_i \geq z, \quad i = 1, \ldots, m, \\ x \in S \subset \mathbb{R}^n_+, \ z \in \mathbb{R}_+. \end{array} \right| \tag{11.54}$$

This problem is encountered in various applications. For example, in location theory (Plastria 1995), it can be interpreted as a "maximin location problem": a^i, $i = 1, \ldots, m$, are the locations of m obnoxious facilities, and $\alpha_i > 0$ is the radius of the polluted region of facility i, while an optimal solution is a location $x \in S$ outside all polluted regions and as far as possible from the nearest of these obnoxious facilities. In engineering design (DL) appears as a variant of the "design centering problem" (Thach 1988; Vidigal and Director 1982), an important special case of which, when $\alpha_i = 0, i = 1, \ldots, m$, is the "largest empty ball problem" (Shi and Yamamoto 1996): given m points $a^1, \ldots, a^m$ in $\mathbb{R}^n_{++}$ and a bounded set S, find the largest ball that has center in $S \subset \mathbb{R}^n_+$ and contains none of these points (Fig. 11.5).

As a first step towards solving (DL) one can study the following feasibility problem:

(Q(r)) Given a number $r \geq 0$, find a point $x(r) \in S$ lying outside any one of the m balls of centers a^i and radii $\theta_i = \alpha_i + r$.

It has been shown in Tuy et al. (2003) that this problem can be reduced to

$$\max\{\|x\|^2 - h(x) |\ x \in S\},$$

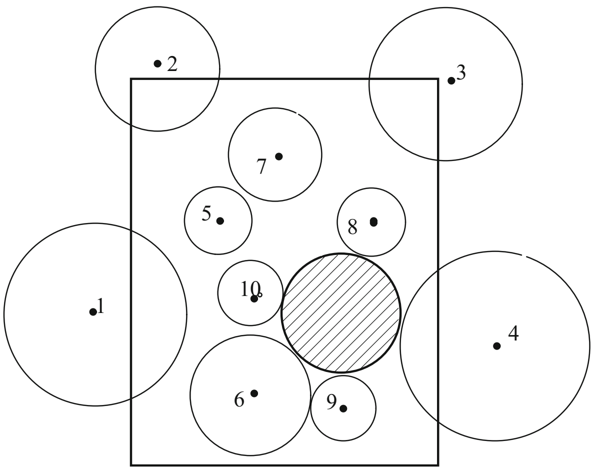


Fig. 11.5 Discrete location problem ($n = 2, m = 10$)

where $h(x) = \max_{i=1,\dots,m}(2\langle a^i, x\rangle + \theta_i^2 - \|a^i\|^2)$. Since both $\|x\|^2$ and $h(x)$ are obviously increasing functions, this is a discrete dm optimization problem that can be solved by a branch-reduce-and-bound method as discussed for solving the Master Problem in Sect. 11.3. Clearly if $\overline{r}$ is the maximal value of r such that (Q(r)) is feasible, then $\overline{x} = x(\overline{r})$ will solve (DL). Noting that $\overline{r} \geq 0$ and for any $r > 0$ one has $\overline{r} \geq r$ or $\overline{r} < r$ according to whether (Q(r)) is feasible or not, the value $\overline{r}$ can be found by a Bolzano binary search scheme: starting from an interval $[0, s]$ containing $\overline{r}$, one reduces it by a half at each step by solving a (Q(r)) with a suitable r. Quite encouraging computational results with this preliminary version of the SIT approach to dm optimization have been reported in Tuy et al. (2003). However, it turns out that more complete and much better results can be obtained by a direct application of the enhanced Discrete SIT Algorithm adapted to problem (11.54). Next we describe this method.

Observe that

$$\begin{aligned}
&\|x - a^i\| - \alpha_i \geq z, \quad i = 1, \dots, m, \\
\Leftrightarrow\ &\|x\|^2 + \|a^i\|^2 - 2\langle a^i, x\rangle \geq (z + \alpha_i)^2, \quad i = 1, \dots, m, \\
\Leftrightarrow\ &\max_{i=1,\dots,m}\{(z + \alpha_i)^2 + 2\langle a^i, x\rangle - \|a^i\|^2\} \leq \|x\|^2.
\end{aligned}$$

Therefore, by setting

$$\varphi(x, z) = \max_{i=1,\dots,m}((z + \alpha_i)^2 + 2\langle a^i, x\rangle - \|a^i\|^2), \tag{11.55}$$

problem (11.54) can be restated as

$$\max\{z|\ \varphi(x, z) - \|x\|^2 \leq 0,\ x \in [a, b] \cap S,\ z \geq 0\}, \tag{11.56}$$

where $\varphi(x, z)$ and $\|x\|^2$ are increasing functions and $[a, b]$ is a box containing S.

To apply the Discrete SIT Algorithm for solving this problem, observe that the value of z is determined when x is fixed. Therefore, branching should be performed upon the variables x only.

Another key operation in the algorithm is bounding: given a box $[p, q] \subset [a, b]$, compute an upper bound $\mu(M)$ for the optimal value of the subproblem

$$\max\{z|\ \varphi(x, z) - \|x\|^2 \leq 0,\ x \in S \cap [p, q],\ 0 \leq z\}. \tag{11.57}$$

If $CBV = r > 0$ is the largest thus far known value of z at a feasible solution (x, z), then only feasible points (x, z) with $z > r$ are still of interest. Therefore, to obtain a tighter value of $\mu(M)$ one should consider, instead of (11.57), the problem

$$(\mathrm{DL}(M, r)) \qquad \max\{z|\ \varphi(x, z) - \|x\|^2 \leq 0,\ x \in S \cap [p, q],\ z \geq r\}.$$

Because $\varphi(x,z)$ and $\|x\|^2$ are both increasing, the constraints of $\mathrm{DL}(M,r)$ can be relaxed to $\varphi(p,z) - \|q\|^2$, so an obvious upper bound is

$$c(p,q) := \max\{z|\ \varphi(p,z) - \|q\|^2 \leq 0\}. \tag{11.58}$$

Although this bound is easy to compute (it is the zero of the increasing function $z \mapsto \varphi(p,z) - \|q\|^2$), it is often not quite efficient. A better bound can be computed by solving a linear relaxation of $\mathrm{DL}(M,r)$ obtained by omitting the discrete constraint $x \in S$ and replacing $\varphi(x,z)$ with a linear underestimator. Since

$$(z+\alpha_i)^2 \geq (r+\alpha_i)^2 + 2(r+\alpha_i)(z-r),$$

a linear underestimator of $\varphi(x,z)$ is

$$\psi(x,z) := \max_{i=1,\ldots,n}\{2(r+\alpha_i)(z-r) + (r+\alpha_i)^2 + 2\langle a^i, x\rangle - \|a^i\|^2\}.$$

On the other hand,

$$\|x\|^2 \leq \sum_{j=1}^{n}[(p_j+q_j)x_j - p_jq_j] \quad \forall x \in [p,q],$$

so by (11.55) the constraint $\varphi(x,z) - \|x\|^2 \leq 0$ can be relaxed to

$$\psi(x,z) - \sum_{j=1}^{n}[(p_j+q_j)x_j - p_jq_j] \leq 0,$$

which is equivalent to the system of linear constraints

$$\max_{i=1,\ldots,m}\{2(r+\alpha_i)(z-r) + (r+\alpha_i)^2 + 2\langle a^i, x\rangle - \|a^i\|^2\} \leq \sum_{j=1}^{n}[(p_j+q_j)x_j - p_jq_j].$$

Therefore, a lower bound can be taken to be the optimal value $\mu(M)$ of the linear programming problem

$$(\mathrm{LP}(M,r)) \quad \left|\begin{array}{ll} \max & 2(r+\alpha_i)(z-r) + (r+\alpha_i)^2 + 2\langle a^i, x\rangle - \|a^i\|^2 \\ \text{s.t.} & \displaystyle\sum_{j=1}^{n}[(p_j+q_j)x_j - p_jq_j] \leq 0,\ i = 1,\ldots,m, \\ & z \geq r,\ p \leq x \leq q. \end{array}\right.$$

The bounds can be further improved by using valid reduction operations. For this, observe that, since $\varphi(x,r) \leq \varphi(x,z)$ for $r \leq z$, the feasible set of $(\mathrm{DL}(M,r))$ is contained in the set

$$\{x|\ \varphi(x,r) - \|x\|^2 \leq 0,\ x \in [p,q]\},$$

so, according to Proposition 11.13, if p', q' are defined by (11.13) with

$$g(x) = \varphi(x, r), \quad h(x) = \|x\|^2,$$

and $\tilde{p} = \lceil p' \rceil_{S^*}, \tilde{q} = \lfloor q' \rfloor_{S^*}$, then the box $[\tilde{p}, \tilde{q}] \subset [p, q]$ is a valid reduction of $[p, q]$.

Thus, the Discrete SIT Algorithm as specialized to problem (DL) becomes the following algorithm:

SIT Algorithm for (DL)

Initialization. Let $\mathscr{P}_1 := \{M_1\}, M_1 := [a, b], \mathscr{R}_1 := \emptyset$. If some feasible solution $(\overline{x}, \overline{z})$ is available, let $r := \overline{z}$ be *CBV*, the current best value. Otherwise, let $r := 0$. Set $k := 1$.

Step 1. Apply S-adjusted reduction cuts to reduce each box $M := [p, q] \in \mathscr{P}_k$. In particular delete any box $[p, q]$ such that $\varphi(p, r) - \|q\|^2 \geq 0$. Let $\mathscr{P}'_k$ be the resulting collection of reduced boxes.

Step 2. For each $M := [p, q] \in \mathscr{P}'_k$ compute $\mu(M)$ by solving LP(M, r). If LP(M, r) is infeasible or $\mu(M) = r$, then delete M. Let $\mathscr{P}^*_k := \{M \in \mathscr{P}'_k | \ \mu(M) > r\}$. For every $M \in \mathscr{P}^*_k$ if a basic optimal solution of LP(M, r) can be S-adjusted to derive a feasible solution (x^M, z^M) with $z^M > r$, then use it to update *CBV*.

Step 3. Let $\mathscr{S}_k := \mathscr{R}_k \cup \mathscr{P}^*_k$. Reset $r := CBV$. Delete every $M \in \mathscr{S}_k$ such that $\mu(M) < r$ and let $\mathscr{R}_{k+1}$ be the collection of remaining boxes.

Step 4. If $\mathscr{R}_{k+1} = \emptyset$, then terminate: if $r = 0$, the problem is infeasible; otherwise, r is the optimal value and the feasible solution $(\overline{x}, \overline{z})$ with $\overline{z} = r$ is an optimal solution.

Step 5. If $\mathscr{R}_{k+1} \neq \emptyset$, let $M_k \in \text{argmax}\{\mu(M) | \ M \in \mathscr{R}_{k+1}\}$. Divide M_k into two boxes according to the standard bisection rule. Let $\mathscr{P}_{k+1}$ be the collection of these two subboxes of M_k.

Step 6. Increment k and return to Step 1.

Theorem 11.2 *"SIT Algorithm for (DL)" solves the problem* (DL) *in finitely many iterations.*

Proof Although the variable z is not explicitly required to take on discrete values, this is actually a discrete variable, since the constraint (11.54) amounts to requiring that $z = \min_{i=1,\ldots,m}(\|x - a^i\| - \alpha_i)$ for $x \in S$. The finiteness of "SIT Algorithm for (DL)" then follows from Theorem 7.10. □

As reported in Tuy et al. (2006), the above algorithm has been tested on a number of problem instances of dimension ranging from 10 to 100 (10 problems for each instance of n). Points a^j were randomly generated in the square $1000 \leq x_i \leq 90{,}000$, $i = 1, \ldots, n$, while S was taken to be the lattice of points with integral coordinates. The computational results suggest that the method is quite practical for this class of discrete optimization problems. The computational cost increases much more rapidly with n (dimension of space or number of variables) than with m (number of balls). Also, when compared with the results preliminarily reported in Tuy et al. (2003), they show that the performance of the method can be drastically improved by using a more suitable monotonic reformulation.

11.7 Exercises

1 Let P be a polyblock in the box $[a,b] \subset \mathbb{R}^n$. Show that the set $Q = \mathrm{cl}([a,b] \setminus P)$ is a copolyblock.

Hint: For each vertex a^i of P the set $[a,b] \setminus [a,a^i)$ is a copolyblock; see Proposition 11.7.

2 Let $h(x)$ be a dm function on $[a,b]$ such that $h(x) \geq \alpha \geq 0\ \forall x \in [a,b]$ and $q : \mathbb{R}_+ \to \mathbb{R}$ a convex decreasing function with finite right derivative at a : $q'^+(a) > -\infty$. Prove that $q(h(x))$ is a dm function on$[a,b]$.

Hint: For any $\theta \in \mathbb{R}_+$ we have $q(t) \leq q(\theta) + q'^+(\theta)(t-\theta)$ with equality holding for $\theta = t$, so $q(t) = \sup_{\theta \in \mathbb{R}_+}\{q(\theta) + q'^+(\theta)(t-\theta)\}$. If $h(x) = u(x) - v(x)$ with $u(x), v(x)$ increasing, show that $g(x) := K(u(x) + v(x)) - q(h(x))$ is increasing.

3 Rewrite the following problem (Exercise 10, Chap. 5) as a canonical monotonic optimization problem (MO):

$$\text{minimise}\ \ 3x_1 + 4x_2\ \ \text{subject to}$$

$$y_{11}x_1 + 2x_2 + x_3 = 5;\quad y_{12}x_1 + x_2 + x_4 = 3$$

$$x_1, x_2 \geq 0;\ y_{11} - y_{12} \leq 1;\ \ 0 \leq y_{11} \leq 2;\ 0 \leq y_{12} \leq 2$$

Solve this problem by the SIT Algorithm for (MO).

4 Consider the monotonic optimization problem $\min\{f(x)|\ g(x) \leq 0, x \in [a,b]\}$ where $f(x), g(x)$ are lower semi-continuous increasing functions. Devise a rectangular branch and bound method for solving this problem using an adaptive branching rule.

Hint: A polyblock (with vertex set V) which contains the normal set $G = \{x \in [a,b]|\ g(x) \leq 0\}$ can be viewed as a covering of G by the set of overlapping boxes $[a,v], v \in V$: $G \subset \cup_{v \in V}[a,v]$, so the POA algorithm is a BB procedure in each iteration of which a set of overlapping boxes $[a,v], v \in V$ is used to cover the feasible set instead of a set of non-overlapping boxes $[p,q] \subset [a,b]$ as in a conventional BB algorithm.

Chapter 12
Polynomial Optimization

In this chapter we are concerned with polynomial optimization which is a very important subject with a multitude of applications: production planning, location and distribution (Duffin et al. 1967; Horst and Tuy 1996), risk management, water treatment and distribution (Shor 1987), chemical process design, pooling and blending (Floudas 2000), structural design, signal processing, robust stability analysis, design of chips (Ben-Tal and Nemirovski 2001), etc.

12.1 The Problem

In its general form the polynomial optimization problem can be formulated as

$$(P) \qquad \min\{f_0(x) | f_i(x) \geq 0 \ (i = 1, \ldots, m), \ a \leq x \leq b\},$$

where $a, b \in \mathbb{R}^n_+$ and $f_0(x), f_1(x), \ldots, f_m(x)$ are polynomials, i.e.,

$$f_i(x) = \sum_{\alpha \in A_i} c_{i\alpha} x^\alpha, \ x^\alpha := x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}, c_{i\alpha} \in \mathbb{R}, \tag{12.1}$$

with $\alpha = (\alpha_1, \ldots, \alpha_n)$ being an n-vector of nonnegative integers. Since the set of polynomials on $[a, b]$ is dense in the space $C[a, b]$ of continuous functions on $[a, b]$ with the supnorm topology, any continuous optimization problem can be, in principle, approximated as closely as desired by a polynomial optimization problem. Therefore, polynomial programming includes an extremely wide class of optimization problems. It is also a very difficult problem of global optimization because many special cases of it such as the nonconvex quadratic optimization problem are known to be NP-hard.

H. Tuy, *Convex Analysis and Global Optimization*, Springer Optimization and Its Applications 110, DOI 10.1007/978-3-319-31484-6_12

Despite its difficulty, polynomial optimization has become a subject of intensive research during the last decades. Among the best so far known methods of polynomial optimization one should mention: an earliest method by Shor and Stetsenko (1989), the RLT method (Reformulation–Lineralization Technique) by Sherali and Adams (1999), a SDP relaxation method by Lasserre (2001). Furthermore, since any polynomial on a compact set can be represented as a dc function, it is natural that dc methods of optimization have been adapted to solve polynomial programs. Particularly a primal-dual relaxed algorithm called GOP developed by Floudas (2000) for biconvex programming has been later extended to a wider class of polynomial optimization problems. Finally, by viewing a polynomial as a dm function, polynomial optimization problems can be efficiently treated by methods of monotonic optimization.

12.2 Linear and Convex Relaxation Approaches

In branch and bound methods for solving optimization problems with a nonconvex constraint set, an important operation is bounding over a current partition set M. A common bounding practice consists in replacing the original problem (P) by a more easily solvable relaxed problem. Solving this relaxed problem restricted to M yields a lower bound for the objective function value over M. Under suitable conditions this relaxation can be proved to lead to a convergent algorithm. Below we present the RLT method (Sherali and Adams 1999) which is an extension of the Reformulation–Linearization technique for quadratic programming (Chap. 6, Sect. 9.5) and the SDP relaxation method (Lasserre 2001, 2002) which uses a specific convex relaxation technique based on Semidefinite Programming.

12.2.1 The RLT Relaxation Method

Letting $l_i(x) := c_0^i + \langle c^i, x\rangle$ be the linear part of $f_i(x)$, the polynomial (12.1) can be written as

$$f_i(x) = l_i(x) + \sum_{\alpha \in A_i^*} c_{i\alpha} x^{\alpha},$$

where $A_i^* = \{\alpha \in A_i | \sum_{j=1}^n \alpha_j \geq 2\}$. Setting $A = \cup_{i=0}^m A_i^*, c_{i\alpha} = 0$ for $\alpha \in A \setminus A_i^*$ we can further write (12.1) as

$$f_i(x) = l_i(x) + \sum_{\alpha \in A} c_{i\alpha} x^{\alpha}. \tag{12.2}$$

The RLT method is based on the simple observation that using the substitution $y_\alpha \leftarrow x^\alpha$ each polynomial $f_i(x)$ can be transformed into the function $L_i(x, y) = l_i(x) + \sum_{\alpha\in A} c_{i\alpha} y_\alpha$, which is linear in $(x, y) \in \mathbb{R}^n \times \mathbb{R}^{|A|}$. Consequently, the polynomial optimization problem (P) can be rewritten as

$$(Q) \qquad \left| \begin{array}{ll} \min & l_0(x) + \sum_{\alpha\in A} c_{0\alpha} y_\alpha, \\ \text{s.t} & l_i(x) + \sum_{\alpha\in A} c_{i\alpha} y_\alpha \geq 0 \ (i = 1, \ldots, m), \\ & a \leq x \leq b, \\ & y_\alpha = x^\alpha, \ \alpha \in A. \end{array} \right.$$

which is a linear program with the additional nonconvex constraints

$$y_\alpha = x^\alpha, \ \alpha \in A. \tag{12.3}$$

The linear program obtained by omitting these nonconvex constraints

$$(LR(P)) \qquad \left| \begin{array}{ll} \min & l_0(x) + \sum_{\alpha\in A} c_{0\alpha} y_\alpha, \\ \text{s.t} & l_i(x) + \sum_{\alpha\in A} c_{i\alpha} y_\alpha \geq 0 \ (i = 1, \ldots, m), \\ & a \leq x \leq b. \end{array} \right.$$

yields a linear relaxation of (P). For any box $M = [p, q] \subset [a, b]$ let (P_M) denotes the problem (P) restricted to the box $M = [p, q]$; then the optimal value of $LR(P_M)$ yields a lower bound $\beta(M)$ for $\min(P_M)$ such that $\beta(M) = +\infty$ whenever the feasible set of (P_M) is empty.

The bounds computed this way may be rather poor. To have tighter relaxation and hence, tighter bounds, one way is to augment the original constraint set with implied constraints formed by taking mixed products of original constraints. For example, for a partition set $[p, q]$, such an implied constraint is $(x_1 - p_1)(x_3 - p_3)(q_2 - x_2) \geq 0$, i.e.,

$$\begin{aligned} &-x_1x_2x_3 + p_1x_2x_3 + p_3x_1x_2 + q_2x_1x_3 - p_1p_3x_2 \\ &-q_2p_3x_1 + p_1q_3x_2 - p_1q_2x_3 \geq 0. \end{aligned}$$

Of course, the more implied constraints are added, the tighter the relaxation is, but the larger the problem becomes. In practice it is enough to add just a few implied constraints.

With bounds computed by RLT relaxation it is not difficult to devise a convergent rectangular branch and bound algorithm for polynomial optimization.

To ease the reasoning in the sequel it is convenient to write a vector $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in A$ as a sequence of nonnegative integers between 0 and n as follows: α_1 numbers 1, then α_2 numbers $2, \ldots$, then α_n numbers n. For example, for $n = 4$ a vector $\alpha = (2, 3, 1, 2)$ is written as $J(\alpha) = 11222344$. By $\mathscr{A}$ denote the set of all $J(\alpha), \alpha \in A$.

Lemma 12.1 *Consider a box* $M = [p, q] \subset [a, b]$ *and let* $(\overline{x}, \overline{y})$ *be a feasible solution to* $LR(P_M)$*. If* $\overline{x}_l = p_l$ *then*

$$\overline{y}_{J \cup \{l\}} = p_l \overline{y}_J \quad \forall J \in \mathscr{A}, l \in \{0, 1, \ldots, n\}. \tag{12.4}$$

Similarly, if $\overline{x}_l = q_l$ *then*

$$\overline{y}_{J \cup \{l\}} = q_l \overline{y}_J \quad \forall J \in \mathscr{A}, l \in \{0, 1, \ldots, n\}. \tag{12.5}$$

Proof Consider first the case $\overline{x}_l = p_l$. For $|J| = 1$ consider any $j \in \{0, 1, \ldots, n\}$. If $(\overline{x}, \overline{y})$ is feasible to $LR(P_M)$ then $(\overline{x}, \overline{y})$ satisfies

$$[(\overline{x}_l - p_l)(\overline{x}_j - p_j)]_L = \overline{y}_{lj} - p_j \overline{x}_l - p_l \overline{x}_j + p_l p_j \geq 0$$
$$[(\overline{x}_l - p_l)(q_j - \overline{x}_j)]_L = -\overline{y}_{lj} + p_l \overline{x}_j + q_j \overline{x}_l - p_l q_j \geq 0,$$

hence, noting that $\overline{x}_l = p_l$,

$$p_j(\overline{x}_l - p_l) + p_l \overline{x}_j \leq \overline{y}_{lj} \leq q_j(\overline{x}_l - p_l) + p_l \overline{x}_j$$

and consequently, $\overline{y}_{lj} = p_l \overline{x}_j$. Thus (12.4) is true for $|J| = 1$.

Arguing by induction, suppose (12.4) true for $|J| = 1, \ldots, t-1$ and consider $|J| = t$ where $2 \leq t \leq |\mathscr{A}|$. For any $j \in \{0, 1, \ldots, n\}$, if $(\overline{x}, \overline{y})$ is feasible to $LR(P_M)$ then

$$[(\overline{x}_j - p_j)(\overline{x}_l - p_l) \textstyle\prod_{r \in J \setminus \{j\}} (\overline{x}_r - p_r)]_L \geq 0$$
$$[(\overline{x}_l - p_l)(q_j - \overline{x}_j) \textstyle\prod_{r \in J \setminus \{j\}} (\overline{x}_r - p_r)]_L \geq 0.$$

By writing $\prod_{r \in J \setminus \{j\}} (\overline{x}_r - p_r) = \prod_{r \in J \setminus \{j\}} \overline{x}_r + f(\overline{x})$ where $f(x)$ is a polynomial in x of degree no more than $t-2$, we then have

$$(\overline{y}_{J \cup \{l\}} - p_l \overline{y}_J) \geq p_j(\overline{y}_{J \setminus \{l\} \cup \{j\}} - p_l \overline{y}_{J \setminus \{j\}})$$
$$+ [p_l \overline{x}_j f(\overline{x}) - \overline{x}_l \overline{x}_j f(\overline{x})]_L + p_j [\overline{x}_l f(\overline{x}) - p_l f(\overline{x})]_L$$

$$(y_{J \cup \{l\}} - p_l \overline{y}_J) \leq q_j(y_{J \cup \{l\} \setminus \{j\}} - p_l \overline{y}_{J \setminus \{j\}})$$
$$+ [p_l \overline{x}_j f(\overline{x}) - \overline{x}_l \overline{x}_j f(\overline{x})]_L + q_j [\overline{x}_l f(\overline{x}) - p_l f(\overline{x})]_L.$$

By induction hypothesis $\overline{y}_{J \cup \{l\} \setminus \{j\}} = p_l \overline{y}_{J \setminus \{j\}}$, $[\overline{x}_l \overline{x}_j f(\overline{x})]_L = p_l [\overline{x}_j f(\overline{x})]_L$, and $[\overline{x}_l f(\overline{x})]_L = p_l [f(\overline{x})]_L$, so the right-hand sides of both the above inequalities are zero. Hence, $\overline{y}_{J \cup \{l\}} = p_l \overline{y}_J$. □

Corollary 12.1 *Let* $M = [p, q] \subset [a, b]$*. If* $(\overline{x}, \overline{y})$ *is a feasible solution of* $LR(P_M)$ *such that* $\overline{x}$ *is a corner of the box* M *then* $\overline{x}$ *is a feasible solution of* (P_M)*, i.e., is a feasible solution of problem (P) in the box M.*

Proof If $\overline{x}$ is a corner of the box $M = [p, q]$ there exists a set $I \subset \{1, \ldots, n\}$ such that $\overline{x}_i = p_i \; \forall i \in I$ and $\overline{x}_j = q_j \; \forall j \in K := \{1, \ldots, n\} \setminus I$. Consider any $\alpha \in A$ and $J = J(\alpha)$. By Lemma 12.1, for every $j \in J \cap K$ we have $\overline{y}_J = q_j \overline{y}_{J\setminus\{j\}} = \overline{x}_j \overline{y}_{J\setminus\{j\}}$, so $\overline{y}_J = (\prod_{j\in J\cap K} \overline{x}_j)\overline{y}_{J\cap I}$. Then $\overline{y}_J = (\prod_{j\in J\cap K} \overline{x}_j)(\prod_{i\in J\cap I} \overline{x}_i)$, i.e., $y_\alpha = x^\alpha$. Hence, $\overline{x} \in M$ and is a feasible solution of (P). □

Observe that the above proposition is an extension of Proposition 9.9 for quadratic programming to polynomial optimization. Just like for quadratic programming, it gives rise to the following branching method in a BB procedure for solving (P). By D denote the feasible set of (P). Let $M_k = [p^k, q^k] \subset [a, b]$ be the partition set selected for further exploration at iteration kth of the BB procedure. Let $\beta(M_k)$ be the lower bound for $f_0(x)$ over $D \cap M_k$ computed by RLT relaxation, i.e.,

$$\beta(M_k) = l_0(x^k) + \sum_{\alpha\in A} c_{0\alpha} y^k_\alpha, \tag{12.6}$$

where $(x^k, y^k) \in \mathbb{R}^n \times \mathbb{R}^{|A|}$ is an optimal solution of problem $LR(P_{M_k})$. If x^k is a corner of M_k, we are done by the above proposition. Otherwise, we bisect M_k via (x^k, i_k) where

$$i_k \in \operatorname{argmax}\{\eta^k_i |\; i = 1, \ldots, n\}, \quad \eta^k_i = \min\{x^k_i - p^k_i, q^k_i - x^k_i\}.$$

Theorem 12.1 *A rectangular BB Algorithm with bounding by RLT relaxation and branching as above either terminates after finitely many iterations, yielding an optimal solution x^k of (P) or it generates an infinite sequence $\{x^k\}$ converging to an optimal solution $\overline{x}$ of (P).*

Proof The proof is analogous to that of Proposition 9.10. Specifically, we need only consider the case when the algorithm is infinite. By Theorem 6.3 the branching rule ensures that a filter of rectangles $M_\nu = [p^\nu, q^\nu]$ is generated such that $(x^\nu, y^\nu) \to (x^*, y^*)$, and x^* is a corner of $M_* := [p^*, q^*] = \cap_\nu [p^\nu, q^\nu)$. Clearly (x^*, y^*) is a feasible solution of $LR(P_{M_*})$, so by above corollary x^* is a feasible solution of (P). Since $\beta(M_\nu) = l_0(x^\nu) + \sum_{\alpha\in A} c_{0\alpha} y^\nu_\alpha \le f_0(x) \; \forall x \in D$ we have $l_0(x^*) + \sum_{\alpha\in A} c_{0\alpha} y^*_\alpha \le f_0(x) \; \forall x \in D$. But $x^* \in D$ implies that $l_0(x^*) + \sum_{\alpha\in A} c_{0\alpha} y^*_\alpha = f_0(x^*)$, hence $f_0(x^*) \le f_0(x) \; \forall x \in D$. □

12.2.2 The SDP Relaxation Method

This method is due essentially to Lasserre (2001, 2002). By p^* denote the optimal value of the polynomial optimization problem (P):

$$p^* := \min\{f_0(x) |\, f_i(x) \ge 0 \; (i = 1, \ldots, m),\; x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n\}.$$

We show that a SDP relaxation of (P) can be constructed whose optimal value approximates p^* as closely as desired. Consider the following matrices called *moment matrices*:

$$M_1(x) = \begin{bmatrix} x_1 \\ \dots \\ x_n \end{bmatrix} \begin{bmatrix} 1 \; x_1 \dots x_n \end{bmatrix} \tag{12.7}$$

$$= \begin{bmatrix} 1 & x_1 & x_2 & \dots & x_n \\ x_1 & x_1^2 & x_1x_2 & \dots & x_1x_n \\ x_2 & x_1x_2 & x_2^2 & \dots & x_2x_n \\ \dots & \dots & \dots & \dots & \dots \\ x_n & x_1x_n & x_2x_n & \dots & x_n^2 \end{bmatrix} \tag{12.8}$$

$$M_2(x) = \begin{bmatrix} x_1 \\ \dots \\ x_n \\ x_1^2 \\ x_1x_2 \\ \dots \\ x_1x_n \\ x_2^2 \\ x_2x_3 \\ \dots \\ x_n^2 \end{bmatrix} \begin{bmatrix} x_1 \; \dots \; x_n \; x_1^2 \; x_1x_2 \; \dots \; x_1x_n \; x_2^2 \; x_2x_3 \; \dots \; x_n^2 \end{bmatrix} \tag{12.9}$$

$$= \begin{bmatrix} 1 & x_1 \dots & x_n & x_1^2 & \dots & x_n^2 \\ x_1 & x_1^2 \dots & x_1x_n & x_1^3 & \dots & x_1x_n^2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_n^2 & x_1x_n^2 \dots & x_n^3 & x_1^2x_n^2 & \dots & x_n^4 \end{bmatrix} \tag{12.10}$$

and so all. We also set $M_0(x) = [1]$. Clearly, each moment matrix $M_k(x)$ is semidefinite positive. Setting $M_{ik}(x) := f_i(x)M_k(x)$ for any given natural k we then have

$$M_{i(k-1)}(x) \succeq 0 \Leftrightarrow f_i(x) \geq 0,$$

so the problem (P) is equivalent to

$$(Q_k) \qquad \min\{f_0(x) | \; M_{i(k-1)}(x) \succeq 0, i = 1, \dots, m\}.$$

Upon the change of variables $x^\alpha \leftarrow y_\alpha$ the objective function in (Q_k) becomes a linear function of $y = \{y_\alpha\}$ while the constraints become linear matrix inequalities in $y = \{y_\alpha\}$. The problem (Q_k) then becomes a SDP which is a relaxation of (P), referred to as the *k-th order SDP relaxation* of (P). We have the following fundamental result:

Theorem 12.2 *Suppose the constraint set* $\{x : f_i(x) \geq 0,\ i = 1, 2, \ldots, m\}$ *is compact. Then*

$$\lim_{k \to +\infty} \inf(Q_k) = \min(Q). \tag{12.11}$$

We omit the proof which is based on a theorem of real algebraic geometry by Putinar (1993) on the representation of a positive polynomial as sum of squares of polynomials. The interested reader is referred to Lasserre (2001, 2002) for more detail.

Remark 12.1 In the particular case of quadratic programming:

$$f_i(x) = \frac{1}{2}\langle Q_i x, x\rangle + \langle c_i, x\rangle + d_i,\ i = 0, 1, \ldots, m,$$

with symmetric matrices Q_i, the first order SDP relaxation is

$$\min_{x,y}\{\frac{1}{2}\langle Q_0, y\rangle + \langle c_0, x\rangle + d_0 |\ \frac{1}{2}\langle Q_i, y\rangle + d_i \geq 0, i = 1, \ldots, m,\ \begin{bmatrix} y & x \\ x^T & x_0 \end{bmatrix} \succeq 0\}$$

which can be shown to be the Lagrangian dual of the problem

$$\max_{t \in \mathbb{R}, u \in \mathbb{R}^m_+} \{t |\ \begin{bmatrix} Q(u) & c(u) \\ (c(u))^T & 2(d(u) - t) \end{bmatrix} \succeq 0\},$$

where $Q(u) = Q_0 + \sum_{i=1}^m u_i Q_i,\ c(u) = c_0 + \sum_{i=1}^m u_i c_i,\ d(u) = \sum_{i=1}^m u_i d_i.$

Thus, for quadratic programming the first order SDP relaxation coincides with the Lagrange relaxation as computed at the end of Chap. 10.

12.3 Monotonic Optimization Approach

Although very ingenious, the RLT and SDP relaxation approaches suffer from two drawbacks.

First, they are very sensitive to the highest degree of the polynomials involved. Even for problems of small size but with polynomials of high degree they require introducing a huge number of additional variables.

Second, they only provide an approximate optimal value rather than an approximate optimal solution of the problem. In fact, in these methods the problem (P) is relaxed to a linear program or a SDP problem in the variables $x \in \mathbb{R}^n, y = \{y_\alpha, \alpha \in A\}$, which will become equivalent to (P) when added the nonconvex constraints

$$y_\alpha = x^\alpha \quad \forall \alpha \in A. \tag{12.12}$$

So if $(\hat{x}, \hat{y})$ is an optimal solution to this relaxed problem then $\hat{x}$ will solve problem (P) if $\hat{y}_\alpha = \hat{x}^\alpha \ \forall \alpha \in A$. The idea, then, is to use a suitable BB procedure with bounding based on this kind of relaxation to generate a sequence $(\hat{x}^k, \hat{y}^k)$ such that $\|\hat{y}^k - \hat{x}^k\| \to 0$ as $k \to +\infty$. Under these conditions, given tolerances $\varepsilon > 0, \eta > 0$ it is expected that for large enough k one would have $\|\hat{y}^k - \hat{x}^k\| \leq \varepsilon$, while $f_0(\hat{x}^k) - \eta \leq p^*$, where p^* is the optimal value of problem (P). Such an $\hat{x}^k$ is accepted as an approximate optimal solution, though it is infeasible and may sometimes be quite far off the true optimal solution.

Because of these difficulties, if we are not so much interested in finding an optimal solution as in finding quickly a feasible solution, or possibly a feasible solution better than a given one, then neither the RLT nor the SDP approach can properly serve our purpose. By contrast, with the monotonic optimization approach to be presented below we will be able to get round these limitations.

12.3.1 Monotonic Reformulation

Recall that a function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *increasing* on a box $[a, b] := \{x \in \mathbb{R}^n |\ a \leq x \leq b\} \subset \mathbb{R}^n$ if $f(x') \leq f(x)$ for all x', x satisfying $a \leq x' \leq x \leq b$. It is said to be a *d.m. function* on $[a, b]$ if $f(x) = f_1(x) - f_2(x)$, where f_1, f_2 are increasing on $[a, b]$.

Clearly a polynomial of n variables with positive coefficients is increasing on the orthant $\mathbb{R}^n_+$. Since every polynomial can be written as a difference of two polynomials with positive coefficients:

$$\sum_\alpha c_\alpha x^\alpha = \sum_{c_\alpha > 0} c_\alpha x^\alpha - [\sum_{c_\alpha < 0} (-c_\alpha) x^\alpha]$$

it follows that any polynomial of n variables is a d.m. function on $\mathbb{R}^n_+$.

As shown in Chap. 11, if $g_1(x), \ldots, g_m(x)$ are d.m. functions on $[a, b]$ then so are the functions

$$\max\{g_1(x), \ldots, g_m(x)\}, \ \min\{g_1(x), \ldots, g_m(x)\}.$$

Therefore, the system of polynomial constraints

$$f_i(x) \geq 0 \quad i = 1, \ldots, m, \ x \in [a, b]$$

on $\mathbb{R}^n_+$ can be replaced by a single d.m. constraint

$$g(x) := \min_{i=1,\ldots,m} f_i(x) \geq 0.$$

If $f_0(x) = f_{01}(x) - f_{02}(x)$ where $f_{01}(x), f_{02}(x)$ are polynomials with positive coefficients, then the problem (P) can be written as

$$\begin{aligned} \text{minimize } & f_{01}(x) + t - f_{02}(b), \\ \text{s.t. } & f_j(x) \geq 0 \quad j = 1, \ldots, m, \\ & f_{02}(x) + t - f_{02}(b) \geq 0, \\ & a \leq x \leq b, \ 0 \leq t \leq f_{02}(b) - f_{02}(a), \end{aligned}$$

where the function $(x, t) \mapsto f_{01}(x) + t - f_{02}(b)$ is a polynomial with positive coefficients.

It then follows that by introducing an additional variable if necessary, and changing the notation, we can convert any polynomial programming problem (P) into the form

$$\min\{f(x) | \ g(x) \geq 0, \ x \in [a, b]\}, \tag{Q}$$

where $f(x)$ is an increasing polynomial (polynomial with positive coefficients), while

$$g(x) = \min_{j=1,\ldots,m} \{u_j(x) - v_j(x)\}, \tag{12.13}$$

and $u_j(x), v_j(x)$ are increasing polynomials.

12.3.2 Essential Optimal Solution

From now on we assume that the original polynomial programming problem (P) has been converted to the form (Q).

In general the feasible solution set D of (Q) is not robust, it may contain isolated points. Therefore, as argued in Sect. 11.3, to avoid possible numerical instability, instead of trying to compute a global optimal solution of (P) one should look for the best nonisolated feasible solution. This motivates the following definitions:

A nonisolated feasible solution x^* of (Q) is called an *essential optimal solution* if $f(x^*) \leq f(x)$ for all nonisolated feasible solutions x of (Q), i.e., if

$$f(x^*) = \min\{f(x) | \ x \in S^*\},$$

where S^* denotes the set of all nonisolated feasible solutions of (Q). Assume

$$\{x \in [a, b] | \ g(x) > 0\} \neq \emptyset. \tag{12.14}$$

For $\varepsilon \geq 0$, an $x \in [a,b]$ satisfying $g(x) \geq \varepsilon$ is called an *ε-essential feasible solution*, and a nonisolated feasible solution $\overline{x}$ of (Q) is called an *essential ε-optimal solution* if it satisfies

$$f(\overline{x}) - \varepsilon \leq \inf\{f(x)|\ g(x) \geq \varepsilon,\ x \in [a,b]\}. \tag{12.15}$$

Clearly for $\varepsilon = 0$ a nonisolated feasible solution which is essentially ε-optimal is optimal.

A basic step towards finding an essential ε-optimal solution is to deal with the following subproblem of incumbent transcending:

$(SP\gamma)$ *Given a real number $\gamma \geq f(a)$ (the incumbent value), find a nonisolated feasible solution $\hat{x}$ of (Q) such that $f(\hat{x}) \leq \gamma$, or else establish that none such $\hat{x}$ exists.*

Since $f(x)$ is increasing, if $\gamma = f(b)$ then an answer to the subproblem $(SP\gamma)$ would give a nonisolated feasible solution or else identify essential infeasibility of the problem (a problem is called *essentially infeasible* if it has no nonisolated feasible solution). If $\overline{x}$ is the best nonisolated feasible solution currently available and $\gamma = f(\overline{x}) - \varepsilon$, then an answer to $(SP\gamma)$ would give a new nonisolated feasible solution $\hat{x}$ with $f(\hat{x}) \leq f(\overline{x}) - \varepsilon$, or else identify $\overline{x}$ as an essential ε-optimal solution.

In connection with question $(SP\gamma)$ consider the **master problem**:

$$(Q_\gamma) \qquad\qquad \max\{g(x)|\, f(x) \leq \gamma,\ x \in [a,b]\}.$$

If $g(a) > 0$ and $\gamma = f(a)$ then a solves (Q_γ). Therefore, barring this trivial case, we can assume that

$$g(a) \leq 0,\ f(a) < \gamma. \tag{12.16}$$

Since g and f are both increasing functions, problem (Q_γ) can be solved very efficiently, while, as it turns out, solving (Q_γ) furnishes an answer to question $(SP\gamma)$. More precisely,

Proposition 12.1 *Under assumption (12.16):*

(i) *Any feasible solution x^0 of (Q_γ) such that $g(x^0) > 0$ is a nonisolated feasible solution of (P) with $f(x^0) \leq \gamma$. In particular, if $\max(Q_\gamma) > 0$ then the optimal solution $\hat{x}$ of (Q_γ) is a nonisolated feasible solution of (P) with $f(\hat{x}) \leq \gamma$.*

(ii) *If $\max(Q_\gamma) < \varepsilon$ for $\gamma = f(\overline{x}) - \varepsilon$, and $\overline{x}$ is a nonisolated feasible solution of (P), then it is an essential ε-optimal solution of (P). If $\max(Q_\gamma) < \varepsilon$ for $\gamma = f(b)$, then the problem (P) is ε-essentially infeasible.*

Proof (i) In view of assumption (12.16), $x^0 \neq a$. Since $f(a) < \gamma$, and $x^0 \neq a$, every x in the line segment joining a to x^0 and sufficiently close to x^0 satisfies $f(x) \leq f(x^0) \leq \gamma,\ g(x) > 0$, i.e., is a feasible solution of (P). Therefore, x^0 is a nonisolated feasible solution of (P) satisfying $f(x^0) \leq \gamma$.

(ii) If $\max(Q_\gamma)) < \varepsilon$ then

$$\varepsilon > \sup\{g(x) | f(x) \leq \gamma,\ x \in [a,b]\},$$

so for every $x \in [a,b]$ satisfying $g(x) \geq \varepsilon$, we must have $f(x) > \gamma = f(\overline{x}) - \varepsilon$, in other words,

$$\inf\{f(x) |\ g(x) \geq \varepsilon,\ x \in [a,b]\} > f(\overline{x}) - \varepsilon.$$

This means that, if $\overline{x}$ is a nonisolated feasible solution, then it is an essential ε-optimal solution, and if $\gamma = f(b)$, then $\{x \in [a,b] |\ g(x) \geq \varepsilon\} = \emptyset$, i.e., the problem is ε-essentially infeasible. □

On the basis of this proposition, the problem (Q) can be solved by proceeding according to the following SIT (Successive Incumbent Transcending) scheme:

1. Select tolerance $\eta > 0$. Start with $\gamma = \gamma_0$ for $\gamma_0 = f(b)$.
2. Solve the master problem (Q_γ).

- If $\max(Q_\gamma) > 0$ an essential feasible $\hat{x}$ with $f(\hat{x}) \leq \gamma$ is found. Reset $\gamma \leftarrow f(\hat{x}) + \eta$ and repeat.
- Otherwise, $\max(Q_\gamma) \leq 0$: no essential feasible solution x for (Q) exists such that $f(x) \leq \gamma$. So if $\gamma = f(\hat{x}) + \eta$ for some essential feasible solution $\hat{x}$ of (Q), $\hat{x}$ is an essential η-optimal solution; if $\gamma = \gamma_0$ problem (Q) is essentially infeasible.

The key for implementing this conceptual scheme is an algorithm for solving the master problem.

12.3.3 Solving the Master Problem

Recall that the master problem (Q_γ) is

$$\max\{g(x) | f(x) \leq \gamma,\ x \in [a,b]\}.$$

We will solve this problem by a special RBB (Reduce-Bound-and-Branch) algorithm involving three basic operations: reducing (the partition sets), bounding, and branching.

- *Reducing*: using valid cuts reduce the size of each current partition set $M = [p,q] \subset [a,b]$. The box $[p',q']$ obtained from M as a result of the cuts is referred to as a *valid reduction* of M and denoted by $\text{red}_\gamma[p,q]$.
- *Bounding*: for each current partition set $M = [p,q]$ estimate a valid upper bound $\beta(M)$ for the maximum of $g(x)$ over the feasible portion of (Q_γ) contained in the box $[p',q'] = \text{red}_\gamma[p,q]$.
- *Branching*: successive partitioning of the initial box $M_0 = [a,b]$ using an exhaustive subdivision rule or any subdivision consistent with the bounding.

a. Valid Reduction

At a given stage of the RBB Algorithm for (Q_γ), let $[p,q] \subset [a,b]$ be a box generated during the partitioning procedure and still of interest. The search for a nonisolated feasible solution of (Q) in $[p,q]$ such that $f(x) \le \gamma$ can then be restricted to the set $B_\gamma \cap [p,q]$, where

$$B_\gamma := \{x | f(x) - \gamma \le 0,\ g(x) \ge 0\}. \tag{12.17}$$

Since $g(x) = \min_{j=1,\dots,m}\{u_j(x) - v_j(x)\}$ with $u_j(x), v_j(x)$ being increasing polynomials (see (12.13)), we can also write

$$B_\gamma = \{x | f(x) \le \gamma,\ u_j(x) - v_j(x) \ge 0\ \ j = 1,\dots,m\}.$$

The reduction operation aims at replacing the box $[p,q]$ with a smaller box $[p',q'] \subset [p,q]$ without losing any point $x \in B_\gamma \cap [p,q]$, i.e., such that

$$B_\gamma \cap [p',q'] = B_\gamma \cap [p,q].$$

The box $[p',q']$ satisfying this condition is referred to as a *valid reduction* of $[p,q]$ and denoted by $\text{red}[p,q]$.

Let e^i be the i-th unit vector of $\mathbb{R}^n$ and for any continuous function $\varphi(t) : [0,1] \to \mathbb{R}$ such that $\varphi(0) < 0 < \varphi(1)$, let $N(\varphi)$ denote the largest value $\bar{t} \in (0,1)$ satisfying $\varphi(\bar{t}) = 0$. Also for every $i = 1,\dots,n$ let $q^i = p + (q_i - p_i)e^i$. By Proposition 10.13 applied to problem (Q_γ) we have

$$\text{red}_\gamma[p,q] = \begin{cases} \emptyset & \text{if } f(p) > \gamma \ \text{ or } \ \min_j\{u_j(q) - v_j(p)\} < 0 \\ [p',q'] & \text{if } f(p) \le \gamma \ \ \& \ \ \min_j\{u_j(q) - v_j(p)\} \ge 0 \end{cases}$$

where

$$q' = p + \sum_{i=1}^n \alpha_i(q_i - p_i)e^i, \quad p' = q' - \sum_{i=1}^n \beta_i(q'_i - p_i)e^i,$$

$$\alpha_i = \begin{cases} 1, & \text{if } f(q^i) - \gamma \le 0 \\ N(\varphi_i), & \text{if } f(q^i) - \gamma > 0. \end{cases} \qquad \varphi_i(t) = f(p + t(q_i - p_i)e^i) - \gamma.$$

$$\beta_i = \begin{cases} 1, & \text{if } \min_j\{u_j(q'^i) - v_j(p)\} \ge 0 \\ N(-\psi_i), & \text{if } \min_j\{u_j(q'^i) - v_j(p)\} < 0. \end{cases}$$

$$\psi_i(t) = \min_j\{u_j(q' - t(q'_i - p_i)e^i) - v_j(p)\}.$$

Remark 12.2 It is easily verified that the box $[p',q'] = \text{red}_\gamma[p,q]$ still satisfies

$$f(p') \le \gamma, \quad \min_j\{u_j(q') - v_j(p')\} \ge 0.$$

b. Valid Bounds

Given a box $M := [p, q]$, which is supposed to have been reduced, we want to compute an upper bound $\beta(M)$ for

$$\max\{g(x) | f(x) \leq \gamma,\ x \in [p, q]\}.$$

Since $g(x) = \min_{j=1,\ldots,m}\{u_j(x) - v_j(x)\}$ and $u_j(x), v_j(x)$ are increasing, an obvious upper bound is

$$\beta(M) = \min_{j=1,\ldots,m}[u_j(q) - v_j(p)]. \tag{12.18}$$

Though very simple, this bound ensures convergence of the algorithm, as will be seen shortly. However, for a better performance of the procedure, tighter bounds can be computed using, for instance, the following:

Lemma 12.2 (i) *If $g(p) > 0$ and $f(p) < \gamma$, then p is a nonisolated feasible solution with $f(p) < \gamma$.*

(ii) *If $f(q) > \gamma$ and $x(M) = p + \lambda(q - p)$ with $\lambda = \max\{\alpha | f(p + \alpha(q - p)) \leq \gamma\}$, $z^i = q + (x_i(M) - q_i)e^i, \quad i = 1, \ldots, n$, then an upper bound of $g(x)$ over all $x \in [p, q]$ satisfying $f(x) \leq \gamma$ is*

$$\beta(M) = \max_{i=1,\ldots,n}\min_{j=1,\ldots,m}\{u_j(z^i) - v_j(p)\}.$$

Proof (i) Obvious.

(ii) Let $M_i = [p, z^i] = \{x | p \leq x \leq z^i\} = \{x \in [p, q] | f_i \leq x_i \leq x_i(M)\}$. From the definition of $x(M)$ it is clear that $f(z) > f(x(M)) = \gamma$ for all $z = p + \alpha(q - p)$ with $\alpha > \lambda$. Since for every $x > x(M)$ there exists $z = p + \alpha(q - p)$ with $\alpha > \lambda$, such that $x \geq z$, it follows that $f(x) \geq f(z) > \gamma$ for all $x > x(M)$. So $\{x \in [p, q] | f(x) \leq \gamma\} \subset [p, q] \setminus \{x | x > x(M)\}$. On the other hand, noting that $\{x | x > x(M)\} = \cap_{i=1}^n \{x | x_i > x_i(M)\}$, we can write $[p, q] \setminus \cap_{i=1}^n \{x | x_i > x_i(M)\} = \cup_{i=1,\ldots,n}\{x \in [p, q] | p_i \leq x_i \leq x_i(M)\} = \cup_{i=1,\ldots,n} M_i$. Since $\beta(M_i) = \min_{j=1,\ldots,m}[u_j(z^i) - v_j(p)] \geq \max\{g(x) | x \in M_i\}$ it follows that $\beta(M) = \max\{\beta(M_i) | i = 1, \ldots, n\} \geq \max\{g(x) | x \in \cup_{i=1,\ldots,n} M_i\} \geq \max\{g(x) | x \in B_\gamma \cap [p, q]\}$. □

Remark 12.3 Each box $[p, z^i]$ can be reduced by the method presented above. If $[p'^i, q'^i] = \mathrm{red}[p, z^i]$, $i = 1, \ldots, n$, then without much extra effort, we can have a more refined upper bound, namely

$$\beta(M) = \max_{i=1,\ldots,n}\{\min_{j=1,\ldots,m}[u_j(q'^i) - v_j(p'^i)]\}.$$

The points z^i constructed as in (ii) determine a set $Z := \cup_{i=1}^n [p, z^i]$ containing $\{x \in [p, q] | f(x) \leq \gamma\}$. Such a set Z is a *polyblock* with vertex set $z^i, i = 1, \ldots, n$

(see Sect. 10.1). The above procedure thus amounts to constructing a polyblock $Z \supset \{x \in [p,q] | f(x) \leq \gamma\}$—which is possible because $f(x)$ is increasing. To have a tighter bound, one can even construct a sequence of polyblocks $Z_1 \supset Z_2, \ldots,$ approximating the set $B_\gamma \cap [p,q]$ more and more closely. For the details of this construction the interested reader is referred to Tuy (2000a). By using polyblock approximation one could compute a bound as tight as we wish; since, however, the computation cost increases rapidly with the quality of the bound, a trade-off must be made, so practically just one approximating polyblock as in the above lemma is used.

Remark 12.4 In many cases, efficient bounds can also be computed by exploiting, additional structure such as partial convexity, aside from monotonicity properties Tuy (2007b).

For example, assuming, without loss of generality, that $p \in \mathbb{R}^n_{++}$, i.e., $f_i > 0$, $i = 1,\ldots,n$, and using the transformation $t_i = \log x_i, i = 1,\ldots,n$, one can write each monomial x^α as $\exp\langle t,\alpha\rangle$, and each polynomial $f_k(x) = \sum_{\alpha\in A_k} c_{k\alpha}x^\alpha$ as a function $\varphi_k(t) = \sum_{\alpha\in A_k} c_{k\alpha}\exp\langle t,\alpha\rangle$. Since, as can easily be seen, $\exp\langle t,\alpha\rangle$ with $\alpha \in \mathbb{R}^m_+$ is an increasing convex function of t, each polynomial in x becomes a d.c. function of t and the subproblem over $[p,q]$ becomes a d.c. optimization problem under d.c. constraints. Here each d.c. function is a difference of two increasing convex functions of t on the box $[\exp p, \exp q]$, where $\exp p$ denotes the vector $(\exp f_1,\ldots,\exp f_n)$. A bound over $[p,q]$ can then be computed by exploiting simultaneously the d.c. and d.m. structures.

c. A Robust Algorithm

Incorporating the above RBB procedure for solving (Q_γ) into the SIT scheme yields the following robust algorithm for solving (Q) :

SIT Algorithm for (Q)

Initialization. If no feasible solution is known, let $\gamma = f(b)$; otherwise, let $\overline{x}$ be the best nonisolated feasible solution available, $\gamma = f(\overline{x}) - \varepsilon$. Let $\mathscr{P}_1 = \{M_1\}, M_1 = [a,b], \mathscr{R}_1 = \emptyset$. Set $k = 1$.

Step 1. For each box $M \in \mathscr{P}_k$:

- Compute its valid reduction red$M$;
- Delete M if red$M = \emptyset$;
- Replace M by red$M$ if red$M \neq \emptyset$;
- If red$M = [p',q']$, then compute an upper bound $\beta(M)$ for $g(x)$ over the feasible solutions in red$M$. ($\beta(M)$ must satisfy $\beta(M) \leq \min_{j=1,\ldots,m}[u_j(q') - v_j(p')]$, see (12.18)). Delete M if $\beta(M) < 0$.

Step 2. Let $\mathscr{P}'_k$ be the collection of boxes that results from $\mathscr{P}_k$ after completion of Step 1. Let $\mathscr{R}'_k = \mathscr{R}_k \cup \mathscr{P}'_k$.

Step 3. If $\mathscr{R}'_k = \emptyset$, then terminate: $\bar{x}$ is an essential ε-optimal solution of (Q) if $\gamma = f(\bar{x}) - \varepsilon$, or the problem (Q) is essentially infeasible if $\gamma = f(b)$.

Step 4. If $\mathscr{R}'_k \neq \emptyset$, let $[p^k, q^k] := M_k \in \operatorname{argmax}\{\beta(M) | \, M \in \mathscr{R}'_k\}$, $\beta_k = \beta(M_k)$.

Step 5. If $\beta_k < \varepsilon$, then terminate: $\bar{x}$ is an essential ε-optimal solution of (Q) if $\gamma = f(\bar{x}) - \varepsilon$, or the problem (Q) is ε-essentially infeasible if $\gamma = f(b)$.

Step 6. If $\beta_k \geq \varepsilon$, compute $\lambda_k = \max\{\alpha | \, f(p^k + \alpha(q^k - p^k)) - \gamma \leq 0\}$ and let

$$x^k = p^k + \lambda_k(q^k - p^k).$$

6.a) If $g(x^k) > 0$, then x^k is a new nonisolated feasible solution of (Q) with $f(x^k) \leq \gamma$: if $g(p^k) < 0$, compute the point $\bar{x}^k$ where the line segment joining p^k to x^k meets the surface $g(x) = 0$, and reset $\bar{x} \leftarrow \bar{x}^k$; otherwise, reset $\bar{x} \leftarrow p^k$. Go to Step 7.

6.b) If $g(x^k) \leq 0$, go to Step 7, with $\bar{x}$ unchanged.

Step 7. Divide M_k into two subboxes by the standard bisection (or any bisection consistent with the bounding $M \mapsto \beta(M)$). Let $\mathscr{P}_{k+1}$ be the collection of these two subboxes of M_k, $\mathscr{R}_{k+1} = \mathscr{R}'_k \setminus \{M_k\}$. Increment k, and return to Step 1.

Proposition 12.2 *The above algorithm terminates after finitely many steps, yielding either an essential ε-optimal solution of (Q), or an evidence that the problem is essentially infeasible.*

Proof Since any feasible solution x with $f(x) \leq \gamma = f(\bar{x}) - \varepsilon$ must lie in some box $M \in \mathscr{R}'_k$ the event $\mathscr{R}'_k = \emptyset$ implies that no such solution exists, hence the conclusion in Step 3. If Step 5 occurs, so that $\beta_k < \varepsilon$, then $\max(Q^*/\gamma) \leq \varepsilon$, hence the conclusion in Step 5 (see Proposition 12.1). Thus the conclusions in Step 3 and Step 5 are correct. It remains to show that either Step 3 ($\mathscr{R}'_k = \emptyset$) or Step 5 ($\beta_k < \varepsilon$) must occur for sufficiently large k. To this end, observe that in Step 6, since $f(p^k) \leq \gamma$ (Remark 12.2), the point x^k exists and satisfies $f(x^k) \leq \gamma$, so if $g(x^k) > 0$, then by Proposition 12.1, x^k is a nonisolated feasible solution with $f(x^k) \leq f(\bar{x}) - \varepsilon$, justifying Step 6a. Suppose now that the algorithm is infinite. Since each occurrence of Step 6a decreases the current best value at least by $\varepsilon > 0$ while $f(x)$ is bounded below it follows that Step 6a cannot occur infinitely often. Consequently, for all k sufficiently large, $\bar{x}$ is unchanged, and $g(x^k) \leq 0$, while $\beta_k \geq \varepsilon$. But, as $k \to +\infty$, we have, by exhaustiveness of the subdivision, $\operatorname{diam} M_k \to 0$, i.e., $\|q^k - p^k\| \to 0$. Denote by $\tilde{x}$ the common limit of q^k and p^k as $k \to +\infty$. Since

$$\varepsilon \leq \beta_k \leq \min_{j=1,\ldots,m} [u_j(q^k) - v_j(p^k)],$$

it follows that

$$\varepsilon \leq \lim_{k \to +\infty} \beta_k \leq \min_{j=1,\ldots,m} [u_j(\tilde{x}) - v_j(\tilde{x})] = g(\tilde{x}).$$

But by continuity, $g(\tilde{x}) = \lim_{k \to +\infty} g(x^k) \leq 0$, a contradiction. Therefore, the algorithm must be finite. $\square$

Remark 12.5 The SIT Algorithm works its way to the optimum through a sequence of better and better nonisolated solutions. If for some reason the algorithm has to be stopped prematurely, some reasonably good feasible solution may have been already obtained. This is a major advantage of it over most existing algorithms, which may be useless when stopped prematurely.

12.3.4 Equality Constraints

So far we assumed (12.14), so that the feasible set has a nonempty interior. Consider now the case when there are *equality constraints*, e.g.,

$$h_l(x) = 0,\ l = 1, \ldots, s\ , \tag{12.19}$$

First, if some of these constraints are linear, they can be used to eliminate certain variables. Therefore, without loss of generality we can assume that all the constraints (12.19) are *nonlinear*. Since, however, in the most general case one cannot expect to compute a solution of a nonlinear system of equations in finitely many steps, one should be content with an approximate system

$$-\delta \leq h_l(x) \leq \delta, \quad l = 1, \ldots, s\ ,$$

where $\delta > 0$ is the tolerance. In other words, a set of constraints of the form

$$\begin{aligned} g_j(x) &\geq 0 \quad j = 1, \ldots, m \\ h_l(x) &= 0 \quad l = 1, \ldots, s \end{aligned}$$

should be replaced by the approximate system

$$\begin{aligned} g_j(x) &\geq 0 \quad j = 1, \ldots, m \\ h_l(x) + \delta &\geq 0,\ l = 1, \ldots, s \\ -h_l(x) + \delta &\geq 0 \quad l = 1, \ldots, s. \end{aligned}$$

The method presented in the previous sections can then be applied to the resulting approximate problem. With $g(x) = \min_{j=1,\ldots,m} g_j(x),\ h(x) = \max_{l=1,\ldots,s} |h_l(x)|$, the required assumption is, instead of (12.14), $\{x \in [a, b] |\ g(x) > 0, -h(x) + \delta > 0\} \neq \emptyset$.

12.3.5 Discrete Polynomial Optimization

The above approach can be extended without difficulty to the class of discrete (in particular, integral) polynomial optimization problems. This can be done along the same lines as the extension of monotonic optimization methods to discrete monotonic optimization.

12.3.6 Signomial Programming

A large class of problems called *signomial programming* or *generalized geometric programming* can be considered which inludes problems of the form (P) where each vector $\alpha = (\alpha_1, \ldots, \alpha_n)$ may involve rational non integral components, i.e., may be an arbitrary vector $\alpha \in \mathbb{R}^n_+$. The above solution method via conversion of polynomial programming into monotonic optimization can be applied without modification to signomial programming.

12.4 An Illustrative Example

We conclude by an illustrative example.

Minimize $(3+x_1x_3)(x_1x_2x_3x_4+2x_1x_3+2)^{2/3}$ subject to

$$\begin{aligned}
&-3(2x_1x_2 + 3x_1x_2x_4)(2x_1x_3 + 4x_1x_4 - x_2)\\
&\quad -(x_1x_3 + 3x_1x_2x_4)(4x_3x_4 + 4x_1x_3x_4 + x_1x_3 - 4x_1x_2x_4)^{1/3}\\
&\quad +3(x_4 + 3x_1x_3x_4)(3x_1x_2x_3 + 3x_1x_4 + 2x_3x_4 - 3x_1x_2x_4)^{1/4} \leq\\
&\quad -309.219315\\
&-2(3x_3 + 3x_1x_2x_3)(x_1x_2x_3 + 4x_2x_4 - x_3x_4)^2\\
&\quad +(3x_1x_2x_3)(3x_3 + 2x_1x_2x_3 + 3x_4)^4 - (x_2x_3x_4 + x_1x_3x_4)(4x_1 - 1)^{3/4}\\
&\quad -3(3x_3x_4 + 2x_1x_3x_4)(x_1x_2x_3x_4 + x_3x_4 - 4x_1x_2x_3 - 2x_1)^4 \leq\\
&\quad -78243.910551\\
&-3(4x_1x_3x_4)(2x_4 + 2x_1x_2 - x_2 - x_3)^2\\
&\quad +2(x_1x_2x_4 + 3x_1x_3x_4)(x_1x_2 + 2x_2x_3 + 4x_2 - x_2x_3x_4 - x_1x_3)^4 \leq 9618\\
&\qquad 0 \leq x_i \leq 5 \; i = 1, 2, 3, 4.
\end{aligned}$$

This problem would be difficult to solve by the RLT or Lasserre's method, since a very large number of variables would have to be introduced.

For $\varepsilon = 0.01$, after 53 cycles of incumbent transcending, the SIT Algorithm found the essential ε-optimal solution

$$x^{\text{essopt}} = (4.994594,\ 0.020149,\ 0.045424,\ 4.928073)$$

with essential ε-optimal value 5.906278 at iteration 667, and confirmed its essential ε-optimality at iteration 866.

12.5 Exercises

1 Solve the problem

$$\begin{aligned}&\text{minimize } 3x_1 + 4x_2, \ \text{s.t.}\\ &x_1^2x_2 + 2x_2 + x_3 = 5; \ \ x_1^2x_2^2 + x_2 + x_4 = 3;\\ &x_1 \geq 0, x_2 \geq 0; \ x_1x_2 - x_1x_2^2 \leq 1; \ 0 \leq x_1x_2 \leq 2; \ 0 \leq x_1x_2^2 \leq 2.\end{aligned}$$

2 Consider the problem

$$\begin{aligned}\text{Minimize} \quad & 4(x_1^2x_3 + 2x_1^2x_2x_3^2x_5 + 2x_1^2x_2x_3)(5x_1^2x_3x_4^2x_5 + 3x_2)^{3/5}\\ & + 3(2x_4^2x_5^2)(4x_1^2x_4 + 4x_2x_5)^{5/3}\end{aligned}$$

subject to

$$\begin{aligned}&-2(2x_1x_5 + 5x_1^2x_2x_4^2x_5)(3x_1x_4x_5^2 + 5 + 4x_3x_5^2)^{1/2} \leq -7684.470329\\ &2(2x_1x_2^2x_3x_4^2)(2x_1x_2x_3x_4^2 + 2x_2x_4^2x_5 - x_1^2x_5^2)^{3/2} \leq 1286590.314422\\ &0 \leq x_i \leq 5 \qquad i = 1, 2, 3, 4, 5.\end{aligned}$$

With tolerance $\varepsilon = 0.01$ solve this problem by the SIT monotonic optimization algorithm.

Hints: the essential ε-optimal solution is

$$\overline{x} = (4.987557, \ 4.984973, \ 0.143546, \ 1.172267, \ 0.958926)$$

with objective function value 28766.057367.

Chapter 13
Optimization Under Equilibrium Constraints

In this closing chapter we study optimization problems with nonconvex constraints of a particular type called equilibrium constraints. This class of constraints arises from numerous applications in natural sciences, engineering, and economics.

13.1 Problem Formulation

Let there be given two closed subsets C, D of $\mathbb{R}^n, \mathbb{R}^m$, respectively, and a function $F : C \times D \to \mathbb{R}$. As defined in Chap. 3, Sect. 3.3, an equilibrium point of the system C, D, F is a point $z \in \mathbb{R}^n$ such that

$$z \in C, \quad F(z, y) \geq 0 \;\forall y \in D.$$

Consider now a real-valued function $f(x, z)$ on $\mathbb{R}^n \times \mathbb{R}^m$ together with a nonempty closed set $U \subset \mathbb{R}^n \times \mathbb{R}^m$, and two a set-valued maps C, D from U to $\mathbb{R}^n$ with closed values $C(x), D(x)$ for every $(x, z) \in U$. The problem we are concerned with and which is commonly referred to as *Mathematical Programming problem with Equilibrium Constraints* can be formulated as follows:

$$\text{(MPEC)} \qquad \begin{array}{ll} \text{minimize} & f(x, z) \\ \text{subject to} & \left| \begin{array}{l} (x, z) \in U, \; z \in C(x), \\ F(z, y) \geq 0 \;\forall y \in D(x). \end{array} \right. \end{array}$$

Setting $S(x) = \{z \in C(x) | \; F(x, z) \geq 0 \;\forall y \in D(x)\}$ the equilibrium constraints can also be written as

$$z \in S(x). \tag{13.1}$$

H. Tuy, *Convex Analysis and Global Optimization*, Springer Optimization and Its Applications 110, DOI 10.1007/978-3-319-31484-6_13

For every fixed $x \in X := \{x \in \mathbb{R}^n | \ (x,z) \in U$ for some $z \in \mathbb{R}^m\}$, $S(x)$ is the set of all equilibrium points of the system $(C(x), D(x), F)$. So the problem can be viewed as a game (called *Stackelberg game*) in which a distinctive player is leader and the remaining players are followers. The leader has an objective function $f(x,z)$ depending on two variables $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^m$. The leader controls the variable x and to each decision $x \in X$ by the leader, the followers react by choosing a value of $z \in C(x)$ achieving equilibrium of the system $(C(x), D(x), F)$. Supposing that $f(x,z)$ is a loss, the leader's objective is to minimize $f(x,z)$ under these conditions. In this interpretation the variable x under the leader's control is also referred to as *upper level variable* and the variable y under the followers' control as *lower level variable.* The most important cases that occur in the applications are the following:

1. S is generated by an optimization problem:
$$S(x) = \text{argmin}\{g(x,y) | \ y \in D(x)\}; \tag{13.2}$$
then $F(z,y) = g(x,y) - g(x,z)$.
2. S is generated by a variational inequality:
$$S(x) = \{z \in C(x) | \ \langle g(z), y - z\rangle \geq 0 \ \forall y \in D(x)\}; \tag{13.3}$$
then $F(z,y) = \langle g(z), y - z\rangle$.
3. S is generated by a complementarity problem:
$$\begin{aligned} S(x) &= \{z \in \mathbb{R}^n_+ | \ \langle F(z), y - z\rangle \geq 0 \ \forall y \in \mathbb{R}^n_+\} \\ &= \{z \in \mathbb{R}^n_+, \ F(z) \geq 0, \ \langle F(z), z\rangle = 0\}; \end{aligned} \tag{13.4}$$
then $C(x) = D(x) = \mathbb{R}^n_+, \ F(z,y) = \langle F(z), y - z\rangle$. Although this case can be reduced to the previous one, it deserves special interest as it can be shown to include the important case when S is generated by a Nash equilibrium problem.

There is no need to dwell on the difficulty of MPEC. Finding just an equilibrium point is already a hard problem, let alone solving an optimization problem with an equilibrium constraint. In spite of that, the problem has attracted a great deal of research, due to its numerous important applications in natural sciences, engineering, and economics.

Quite understandably, in the early period aside from heuristic algorithms for solving certain classes of MPECs (Friesz et al. 1992), many solution methods proposed for MPECs have been essentially local optimization methods. Few papers have been concerned with finding a global optimal solution to MPEC.

Among the most important works devoted to MPECs we should mention the monographs by Luo et al. (1996) and Outrata et al. (1998). Both monographs describe a large number of MPECs arising from applications in diverse fields of natural sciences, engineering, and economics.

In Luo et al. (1996), under suitable assumptions including a constraint qualification, MPEC is converted into a standard optimization problem. First and second order optimality conditions are then derived and based on these a penalty interior point algorithm and a descent algorithm are developed for solving the problem. Particular attention is also given to MPAECs which are mathematical programs with affine equilibrium constraint.

The monograph by Outrata–Kočvara–Zowe presents a nonsmooth approach to MPECs where the equilibrium constraint is given in the form of a variational or quasivariational equality. Specifically, using the implicit function locally defined by the equilibrium constraint the problem is converted into a mathematical program with a nonsmooth objective which can then be solved by the bundle method of nonsmooth optimization.

In the following we will present a global optimization approach to MPEC. Just like in the approaches of Luo–Pang–Ralp and Outrata–Kočvara–Zowe, the basic idea is to convert MPEC into a standard optimization problem. The difference, however, is that the latter will be an essential global optimization problem belonging to a class amenable to efficient solution methods, such as dc or monotonic optimization. As it turns out, most MPECs of interest can be approached in that way. The simplest among these are *bilevel programming* and *optimization over efficient set*, for which the equilibrium constraint is an ordinary or multiobjective optimization problem.

13.2 Bilevel Programming

In its general form the Bilevel Programming Problem can be formulated as follows

$$\text{(GBP)}\qquad \begin{array}{rl} & \min f(x,y) \quad \text{s.t.} \\ & g_1(x,y) \le 0, \quad x \in \mathbb{R}^{n_1}_+, \quad y \text{ solves} \\ R(x): & \min\{d(y)|\ g_2(C(x),y) \le 0,\ y \in \mathbb{R}^{n_2}_+\} \end{array}$$

where it is assumed that

(A1) The map $C : \mathbb{R}^{n_1} \to \mathbb{R}^m$ and the functions $f(x,y), d(y), g_1(x,y),\ g_2(u,y)$ are continuous.

(A2) The set $D := \{(x,y) \in \mathbb{R}^{n_1}_+ \times \mathbb{R}^{n_2}_+ |\ g_1(x,y) \le 0, g_2(C(x),y) \le 0\}$ is nonempty and compact.

Also an assumption common to a wide class of (GBP) is that

(A3) For every fixed y the function $u \mapsto g_2(u,y)$ is decreasing.

Assumptions (A1) and (A2) are quite natural. Assumption (A3) means that the function $g_2(C(x),y)$ decreases when the action $C(x)$ of the leader increases (component-wise). By replacing u with $v = -u$, if necessary, this assumption also holds when $g_2(C(x),y)$ increases with $C(x)$.

If all the functions $F(x,y), g_1(x,y), g_2(C(x),y), d(y)$ are convex, the problem is called a *Convex Bilevel Program (CBP)*; if all these functions are linear affine, the problem is called a *Linear Bilevel Program (LBP)*.

Specifically, the Linear Bilevel Programming problem can be formulated as

$$\text{(LBP)} \qquad \begin{array}{l} \min L(x,y) \quad \text{s.t.} \\ A_1x + B_1y \geq c^1, \quad x \in \mathbb{R}^{n_1}_+, \quad y \text{ solves} \\ \min\{\langle d, y\rangle \,|\, A_2x + B_2y \geq c^2,\ y \in \mathbb{R}^{n_2}_+\} \end{array}$$

where $L(x,y)$ is a linear function, $d \in \mathbb{R}^{n_2}, A_i \in \mathbb{R}^{m_i \times n_1}, B_i \in \mathbb{R}^{m_i \times n_2}, c^i \in \mathbb{R}^{m_i}$, $i = 1, 2$. In this case $g_1(x,y) = \max_{i=1,\dots,m_1}(c^1 - A_1x - B_1y)_i \leq 0, g_2(u,y) = \max_{i=1,\dots,m_2}(c^2 - u - B_2y)_i \leq 0$, while $C(x) = A_2x$.

Clearly all assumptions (A1), (A2), (A3) hold for (LBP) provided the set $D := \{(x,y) \in \mathbb{R}^{n_1}_+ \times \mathbb{R}^{n_2}_+ \,|\, A_1x + B_1y \geq c^1, A_2x + B_2y \geq c^2\}$ is a nonempty polytope. Also for every $x \geq 0$ satisfying $A_1x + B_1y \geq c^1$ the set $\{y \in \mathbb{R}^{n_2}_+ \,|\, A_2x + B_2y \geq c^2\}$ is compact, so $\mathrm{R}(x)$ is then solvable.

Originally formulated and studied as a mathematical program by Bracken and McGill (1973, 1974) bilevel and more generally, multilevel optimization has become a subject of extensive research, owing to numerous applications in diverse fields such as: economic development policy, agriculture economics, road network design, oil industry regulation, international water systems and flood control, energy policy, traffic assignment, structural analysis in mechanics, etc. Multilevel optimization is also useful for the study of hierarchical designs in many complex systems, particularly in biological systems (Baer 1992; Vincent 1990).

Even (LBP) is an NP-hard nonconvex global optimization problem (Hansen et al. 1992). Despite the linearity of all functions involved, this is a difficult problem fraught with pitfalls. Actually, several early published methods for its solution turned out to be non convergent, incorrect or convergent to a local optimum quite far off the true optimum. Some of these errors have been reported and analyzed, e.g., in Ben-Ayed (1993). A popular approach to (LBP) at that time was to use, directly or indirectly, a reformulation of it as a single mathematical program which, in most cases, was a linear program with an additional nonconvex constraint. It was this peculiar additional nonconvex constraint that constituted the major source of difficulty and was actually the cause of most common errors.

Whereas (LBP) has been extensively studied, the general bilevel programming problem has received much less attention. Until two decades ago, most research in (GBP) was limited to convex or quadratic bilevel programming problems, and/or was mainly concerned with finding stationary points or local solutions (Bard 1988, 1998; Bialas and Karwan 1982; Candler and Townsley 1982; Migdalas et al. 1998; Shimizu et al. 1997; Vicentee et al. 1994; White and Annandalingam 1993). Global optimization methods for (GBP) appeared much later, based on converting (GBP) into a monotonic optimization problem (Tuy et al. 1992, 1994b, and also Tuy and Ghannadan 1998).

13.2.1 Conversion to Monotonic Optimization

We show that under assumptions (A1)–(A3) (GBP) can be converted into a single level optimization problem which is a monotonic optimization problem.

In view of Assumptions (A1) and (A2) the set $\{(C(x), d(y))|\ (x, y) \in D\}$ is compact, so without loss of generality we can assume that this set is contained in some box $[a, b] \subset \mathbb{R}^{m+1}_+$:

$$a \leq (C(x), d(y)) \leq b \quad \forall (x, y) \in D. \tag{13.5}$$

Setting

$$U = \{u \in \mathbb{R}^m|\ \exists (x, y) \in D \quad u \geq C(x)\}, \tag{13.6}$$

$$W = \{(u, t) \in \mathbb{R}^m \times \mathbb{R}|\ \exists (x, y) \in D \quad u \geq C(x),\ t \geq d(y)\}, \tag{13.7}$$

we define

$$\theta(u) = \min\{d(y)|\ (x, y) \in D,\ u \geq C(x)\}, \tag{13.8}$$

$$f(u, t) = \min\{F(x, y)|\ (x, y) \in D,\ u \geq C(x),\ t \geq d(y)\}. \tag{13.9}$$

Proposition 13.1 (i) *The function $\theta(u)$ is finite for every $u \in U$ and satisfies $\theta(u') \leq \theta(u)$ whenever $u' \geq u \in U$. In other words, $\theta(u)$ is a decreasing function on U.*

(ii) *The function $f(u, t)$ is finite on the set W. If $(u', t') \geq (u, t) \in W$ then $(u', t') \in W$ and $f(u', t') \geq f(u, t)$. In other words, $f(u, t)$ is a decreasing function on W.*

(iii) *$u \in U$ for every $(u, t) \in W$.*

Proof (i) If $u \in U$ then there exists $(x, y) \in D$ such that $C(x) \leq u$, hence $\theta(u) < +\infty$. It is also obvious that $u' \geq u \in U$ implies that $\theta(u') \leq \theta(u)$.

(ii) By assumption (A2) the set D of all $(x, y) \in \mathbb{R}^{n_1}_+ \times \mathbb{R}^{n_2}_+$ satisfying $g_1(x, y) \leq 0$, $g_2(C(x), y) \leq 0$ is nonempty and compact. Then for every $(u, t) \in W$ the feasible set of the problem determining $f(u, t)$ is nonempty, compact, hence $f(u, t) < +\infty$ because $F(x, y)$ is continuous by (A1). Furthermore, if $(u', t') \geq (u, t) \in W$, then obviously $(u', t') \in W$ and $f(u', t') \leq f(u, y)$.

(iii) Obvious. □

As we just saw, the assumption (A2) is only to ensure the solvability of problem (13.9) and hence, finiteness of the function $f(u, t)$. Therefore, without any harm it can be replaced by the weaker assumption

(A2*) The set $W = \{(u, t) \in \mathbb{R}^m \times \mathbb{R}|\ \exists (x, y) \in D \quad u \geq C(x),\ t \geq d(y)\}$ is compact.

In addition to (A1), (A2)/(A2*), and (A3) we now make a further assumption:

(A4) The function $\theta(u)$ is upper semi-continuous on int$U$ while $f(u, t)$ is lower semi-continuous on int$W$.

Proposition 13.2 *If all the functions $F(x,y), g_1(x,y), g_2(C(x),y), d(y)$ are convex then Assumption (A4) holds, with $\theta(u), f(u,t)$ being convex functions.*

Proof Observe that if $\theta(u) = d(y)$ for x satisfying $g_2(C(x),y) \leq 0,\ C(x) \leq u$, and $\theta(u') = d(y')$ for x' satisfying $g_2(C(x'),y') \leq 0,\ C(x') \leq u'$, then $g_2(C(\alpha x + (1-\alpha)x'), \alpha y + (1-\alpha)y') \leq \alpha g_2(C(x),y) + (1-\alpha)g_2(C(x'),y')$, and $C(\alpha x + (1-\alpha)x') \leq \alpha C(x) + (1-\alpha)C(x') \leq \alpha u + (1-\alpha)u'$, so that $(\alpha x + (1-\alpha)x', \alpha y + (1-\alpha)y')$ is feasible to the problem determining $\theta(\alpha u + (1-\alpha)u')$, hence $\theta(\alpha u + (1-\alpha)u') \leq d(\alpha y + (1-\alpha)y') \leq \alpha d(y) + (1-\alpha)d(y') = \alpha\theta(u) + (1-\alpha)\theta(u')$. Therefore the function $\theta(u)$ is convex, and hence continuous on int$U$. The convexity and hence, the continuity of $f(u,t)$, on int$W$ is proved similarly. □

Proposition 13.3 *(GBP) is equivalent to the monotonic optimization problem*

$$(Q) \qquad \min\{f(u,t)|\ t - \theta(u) \leq 0,\ (u,t) \in W\}$$

in the sense that min(GBP)= min(Q) and if $(\overline{u}, \overline{t})$ solves (Q) then any optimal solution $(\overline{x}, \overline{y})$ of the problem

$$\min\{F(x,y)|\ g_1(x,y) \leq 0, g_2(\overline{u},y) \leq 0, \overline{u} \geq C(x), \overline{t} \geq d(y),\ x \geq 0, y \geq 0\}$$

solves (GBP).

Proof By Proposition 13.1 the function $f(u,t)$ is decreasing while $t - \theta(u)$ is increasing, so (Q) is indeed a monotonic optimization problem. If (x,y) is feasible to (GBP), then $g_1(x,y) \leq 0, g_2(C(x),y) \leq 0$, and taking $u = C(x), t = d(y) = \theta(C(x))$ we have $\theta \geq t$, hence

$$\begin{aligned}
F(x,y) \geq \min\{F(x',y')|\ &g_1(x',y') \leq 0, g_2(C(x'),y') \leq 0,\\
&u \geq C(x'), t \geq d(y'), x' \geq 0, y' \geq 0\}\\
= f(u,t) &\geq \min\{f(u',t')|\ \theta(u') - t' \geq 0\}\\
= \min(Q)&.
\end{aligned}$$

Conversely, if $\theta(u) - t \geq 0$, then the inequalities $u \geq C(x), t \geq d(y)$ imply $d(y) \leq t \leq \theta(u) \leq \theta((x))$, hence

$$\begin{aligned}
f(u,t) &= \min\{F(x,y)|\ g_1(x,y) \leq 0, g_2(C(x),y) \leq 0,\\
&\qquad u \geq C(x), t \geq d(y), x \geq 0, y \geq 0\}\\
&\geq \min\{F(x,y)|g_1(x,y) \leq 0, g_2(C(x),y) \geq 0, \theta(C(x)) \geq d(y)\}\\
&= \min(GBP).
\end{aligned}$$

Consequently, min(GBP)= min(Q). The last assertion of the proposition readily follows. □

13.2.2 Solution Method

For convenience of notation we set

$$z = (u, t) \in \mathbb{R}^{m+1}, \qquad h(z) = t - \theta(u). \tag{13.10}$$

The problem (Q) can then be rewritten as

$$(Q) \qquad\qquad \min\{f(z)|\ h(z) \leq 0,\ z \in W\},$$

where $f(z)$ is a decreasing function and $h(z)$ a continuous increasing function The feasible set of this monotonic optimization problem (Q) is a normal set, hence is robust and can be handled without difficulty. As shown in Chap. 11, such a problem (Q) can be solved either by the polyblock outer approximation method or by the BRB (branch-reduce-and-bound) method. The polyblock method is practical only for problems (Q) with small values of m (typically $m \leq 5$). The BRB method we are going to present is more suitable for problems with fairly large values of m.

As its name indicates, the BRB method proceeds according to the standard branch and bound scheme, with three basic operations: branching, reducing (the partition sets), and bounding.

1. Branching consists in a successive rectangular partition of the initial box $M_0 = [a, b] \subset \mathbb{R}^m \times \mathbb{R}$, following a chosen subdivision rule. As will be explained shortly, branching should be performed upon the variables $u \in \mathbb{R}^m$, so the subdivision rule is induced by a standard bisection upon these variables.
2. Reducing consists in applying valid cuts to reduce the size of the current partition set $M = [p, q] \subset [a, b]$. The box $[p', q']$ that results from the cuts is referred to as a *valid reduction* of M.
3. Bounding consists in estimating a *valid lower bound* $\beta(M)$ for the objective function value $f(z)$ over the feasible portion contained in the valid reduction $[p', q']$ of a given partition set $M = [p, q]$.

Throughout the sequel, for any vector $z = (u, t) \in \mathbb{R}^m \times \mathbb{R}$ we define $\hat{z} = u$.

a. Branching

Since an optimal solution (u, t) of the problem (Q) must satisfy $t = \theta(u)$, it suffices to determine the values of the variables u in an optimal solution. This suggests that instead of branching upon $z = (u, t)$ as usual, we should branch upon u, according to a chosen subdivision rule which may be the standard bisection or the adaptive subdivision rule. As shown by Corollary 6.2, the standard subdivision method is exhaustive, i.e., for any infinite nested sequence $M_k, k = 1, 2, \ldots,$ such that each M_{k+1} is a child of M_k via a bisection we have diam$M_k \to 0$. This property guarantees a rather slow convergence, because the method does not take account of the current problem structure. By contrast the adaptive subdivision method exploits the current problem structure and due to that can ensure a much more rapid convergence.

b. Valid Reduction

At a given stage of the BRB algorithm for (Q) a feasible solution $\overline{z}$ is available which is the best so far known. Let $\gamma = f(\overline{z})$ and let $[p,q] \subset [a,b]$ be a box generated during the partitioning procedure which is still of interest. Since an optimal solution of (Q) is attained at a point satisfying $h(z) := t - \theta(u) = 0$, the search for a feasible solution z of (Q) in $[p,q]$ such that $f(z) \leq \gamma$ can be restricted to the set $B_\gamma \cap [p,q]$, where

$$B_\gamma := \{z | f(z) \leq \gamma,\ h(z) = 0\}. \tag{13.11}$$

The reduction aims at replacing the box $[p,q]$ with a smaller box $[p',q'] \subset [p,q]$ without losing any point $y \in B_\gamma \cap [p,q]$, i.e., such that

$$B_\gamma \cap [p',q'] = B_\gamma \cap [p,q].$$

The box $[p',q']$ satisfying this condition is referred to as a γ-*valid reduction* of $[p,q]$ and denoted by $\mathrm{red}_\gamma[p,q]$.

As usual e^i denotes the ith unit vector of $\mathbb{R}^{m+1}$, i.e., $e^i_i = 1, e^i_j = 0\ \forall j \neq i$.

Given any continuous increasing function $\varphi(\alpha) : [0,1] \to \mathbb{R}$ such that $\varphi(0) < 0 < \varphi(1)$ let $N(\varphi)$ denote the largest value $\overline{\alpha} \in (0,1)$ satisfying $\varphi(\overline{\alpha}) = 0$. For every $i = 1,\dots,n$ let $q^i := p + (q_i - p_i)e^i$.

Proposition 13.4 *A γ-valid reduction of a box $[p,q]$ is given by the formula*

$$\mathrm{red}_\gamma[p,q] = \begin{cases} \emptyset, & \text{if } h(q) < 0 \text{ or } \gamma - f(q) < 0, \\ [p',q'], & \text{if } h(q) \geq 0\ \&\ \gamma - f(q) \geq 0. \end{cases} \tag{13.12}$$

where

$$q' = p + \sum_{i=1}^n \alpha_i (q_i - p_i) e^i, \quad p' = q' - \sum_{i=1}^n \beta_i (q' - p_i) e^i, \tag{13.13}$$

$$\alpha_i = \begin{cases} 1, & \text{if } h(q^i) \geq 0 \\ N(\varphi_i), & \text{if } h(q^i) < 0. \end{cases} \quad \varphi_i(\alpha) = h(p + \alpha(q_i - p_i)e^i).$$

$$\beta_i = \begin{cases} 1, & \text{if } f(q'^i) - \gamma \leq 0 \\ N(\psi_i), & \text{if } f(q'^i) - \gamma > 0. \end{cases} \quad \psi_i(\beta) = \gamma - f(q' - \beta(q'_i - p_i)e^i).$$

Proof Follows from Proposition 11.13 □

Remark 13.1 As can be easily checked, the box $[p',q'] = \mathrm{red}[p,q]$ still satisfies $h(q') \geq 0$, $f(q') \leq \gamma$.

c. Valid Bounds

Let $M := [p, q]$ be a partition set which is supposed to have been reduced, so that, as mentioned in the above remark,

$$h(q) \geq 0, \quad f(q) \leq \gamma. \tag{13.14}$$

Let us now examine how to compute a lower bound $\beta(M)$ for

$$\min\{f(z)|\ z \in [p, q],\ h(z) = 0\}.$$

Since $f(z)$ is decreasing, an obvious lower bound is $f(q)$. We will shortly see that to ensure convergence of the BRB Algorithm, it suffices that the lower bounds satisfy

$$\beta(M) \geq f(q).$$

We shall refer to such a bound as a *valid lower bound.*

Define $h_\gamma(z) := \min\{\gamma - f(z), h(z)\}$. Since $-f(z)$ and $h(z)$ are increasing, so is $h_\gamma(z)$. Let

$$G = \{z \in [p, q]|\ h(z) \leq 0\}, \quad H = \{z \in [p, q]|\ h_\gamma(z) \geq 0\}.$$

From (13.14) $h_\gamma(q) \geq 0$, so if $h(q) = 0$ then obviously q is an exact minimizer of $f(z)$ over the feasible points $z \in [p, q]$ that are at least as good as the current best, and $f(q)$ can be used to update the current best.

Suppose now that $h(q) < 0$, and hence, $h(p) < 0$. For each $z \in [p, q]$ such that $z \in H \setminus G$ let $\pi_p(z)$ be the first point where the line from z to p meets the upper boundary of G, i.e.,

$$\pi_p(z) = z - \overline{\alpha}(z - p) \quad \text{with } \overline{\alpha} = N(-\varphi),\ \varphi(\alpha) = h(z - \alpha(z - p)).$$

Obviously, $h(\pi_p(z)) = 0$.

Lemma 13.1 *If* $s = \pi_p(q), v^i = q - (q_i - s_i)e^i, i{=}1, \ldots, m{+}1,$ *and* $I{=}\{i|\ v^i) \in H\}$ *then a valid lower bound over* $M = [p, q]$ *is*

$$\begin{aligned}\beta(M) &= \min\{f(v^i)|\ i \in I\} \\ &= \min\{f(u, t)|\ g_1(x, y) \leq 0, \\ &\qquad g_2(\hat{u}_i, y) \leq 0, (C(x), d(y)) \leq v^i(i \in I),\ x \geq 0,\ y \geq 0\}.\end{aligned}$$

Proof This follows since the polyblock with vertices $v^i, i \in I$, contains all feasible solutions $z = (u, t)$ still of interest in $[p, q]$. □

Remark 13.2 Each box $M_i = [p, v^i]$ can be reduced by the method presented above. If $[p', v'^i] = \text{red}_\gamma[p, v^i], i = 1, \ldots, m+1$, then without much extra effort, we can have the following more refined lower bound:

$$\beta(M) = \min_{i \in I} f(v'^i), \quad I = \{i | \tilde{f}(v^i) \geq 0\}.$$

The above procedure amounts to constructing a polyblock $Z \supset B_\gamma \cap [p, q] = \{z \in [p, q] | h(z) \leq 0 \leq \tilde{f}(z)\}$, which is possible since $h(z), \tilde{f}(z)$, are increasing functions.

d. Improvements for the Case of CBP

In the case of (CBP) (Convex Bilevel Programming) the function $\theta(u)$ is convex by Proposition 13.1, so the function $h(u, t) = t - \theta(u)$ is concave, and a lower bound can be computed which is rather tight. In fact, since $h(z)$ is concave and since the box $[p, q]$ has been γ-reduced, $h(q^i) = 0, i = 1, \ldots, n$ [see (13.13)], the set $\{z \in [p, q] | \; h(z) = 0\}$ is contained in the simplex $S := [q^1, \ldots, q^n]$. Hence a lower bound of $f(u, t)$ over $M = [p, q]$ is given by

$$\beta(M) = \min\{f(z) | \; z \in S\}. \tag{13.15}$$

Taking account of (13.8)–(13.9) this convex program can be written as

$$\begin{array}{c} \min_{u,t} F(x, y) \quad \text{s.t.} \\ g_1(x, y) \leq 0, \; g_2(C(x), y) \leq 0, \\ C(x) \leq u, \; d(y) \leq t, \; x \geq 0, \; y \geq 0, \\ (u, t) \in [q^1, \ldots, q^n]. \end{array} \tag{13.16}$$

We are now in a position to state the proposed monotonic optimization (MO) algorithm for (Q).

Recall that the problem is $\min\{f(z) | \; z \in [a, b], h(z) = 0\}$ and we assume (A1)–(A4) so always $b \in W, \hat{b} = (b_1, \ldots, b_m) \in U$, and $(\hat{b}, \theta(\hat{b}))$ yields a feasible solution.

MO Algorithm for (GBP)

Initialization. Start with $\mathscr{P}_1 = \{M_1\}, M_1 = [a, b], \mathscr{R}_1 = \emptyset$. Let CBS be the best feasible solution available and CBV (current best value) the corresponding value of $f(z)$. (At least CBV=$f(\hat{b}, \theta(\hat{b}))$. Set $k = 1$.

Step 1. For each box $M \in \mathscr{P}_k$ and for γ =CBV:

(1) Compute the γ-valid reduction $\text{red}_\gamma M$ of M;
(2) Delete M if $\text{red}_\gamma M = \emptyset$;
(3) Replace M by $\text{red}_\gamma M$ if $\text{red}_\gamma M \neq \emptyset$;
(4) If $\text{red}_\gamma M = [p, q]$, then compute a valid lower bound $\beta(M)$ for $f(z)$ over the feasible solutions in M.

Step 2. Let $\mathscr{P}'_k$ be the collection of boxes that results from $\mathscr{P}_k$ after completion of Step 1. Update CBS and CBV. From $\mathscr{R}_k$ remove all $M \in \mathscr{R}_k$ such that $\beta(M) \geq$ CBV and let $\mathscr{R}'_k$ be the resulting collection. Let $\mathscr{S}_k = \mathscr{R}'_k \cup \mathscr{P}'_k$.

Step 3. If $\mathscr{S}_k = \emptyset$ terminate: CBV is the optimal value and CBS (the feasible solution $\bar{z}$ with $f(\bar{z}) =$ CBV) is an optimal solution.

Step 4. If $\mathscr{S}_k \neq \emptyset$ select $M_k \in \mathrm{argmin}\{\beta(M)|\ M \in \mathscr{S}_k\}$. Subdivide $M_k = [p^k, q^k]$ according to the adaptive subdivision rule: if $\beta(M_k) = f(v^k)$ for some $v^k \in M_k$ and $z^k = \pi_{p^k}(v^k)$ is the first point where the line segment from v^k to p^k meets the surface $h(z) = 0$ then bisect M_k via (w^k, j_k) where $w^k = \frac{1}{2}(v^k + z^k)$ and j_k satisfies $\|z^k_{j_k} - v^k_{j_k}\| = \max_j \|z^k_j - v^k_j\|$. Let $\mathscr{P}_{k+1}$ be the collection of newly generated boxes.

Step 5. Set $\mathscr{R}_{k+1} = \mathscr{S}_k \setminus \{M_k\}$, increment k and return to Step 1.

Proposition 13.5 *Whenever infinite the above algorithm generates an infinite sequence $\{z^k\}$ every cluster point of it yields a global optimal solution of (GBP).*

Proof Follows from Proposition 6.2. □

13.2.3 Cases When the Dimension Can Be Reduced

Suppose $m > n_1$ while the map C satisfies

$$x' \geq x \Rightarrow C(x') \geq C(x). \tag{13.17}$$

In that case the dimension of the problem (Q) can be reduced. For this, define

$$\theta(v) = \min\{d(y)|\ g_2(C(v), y) \leq 0,\ y \in \mathbb{R}^{n_2}_+\} \tag{13.18}$$

$$\begin{aligned} f(v,t) = \min\{F(x,y)|\ &G_1(x,y) \leq 0,\ g_2(C(v),y) \leq 0 \\ &v \geq x,\ t \geq d(y),\ x \in \mathbb{R}^{n_1}_+,\ y \in \mathbb{R}^{n_2}_+\}. \end{aligned} \tag{13.19}$$

It is easily seen that all these functions are decreasing. Indeed, $v' \geq v$ implies $C(v') \geq C(v)$ and hence $g_2(C(v'), y) \leq 0$, i.e., whenever v is feasible to (13.18) and $v' \geq v$ then v is also feasible. This shows that $\theta(v') \leq \theta(v)$. Similarly, whenever (v, t) is feasible to (13.19) and $(v', t') \geq (v, t)$ then (v, t) is also feasible, and hence, $f(v', t') \leq f(v, t)$.

With $\theta(v)$ and $f(v, t)$ defined that way, we still have

Proposition 13.6 *Problem (GBP) is equivalent to the monotonic optimization problem*

$$(\tilde{Q}) \qquad \min\{f(v,t)|\ t - \theta(v) \leq 0\}$$

and if $(\overline{v}, \overline{t})$ *solves* $(\tilde{Q})$ *then an optimal solution* $(\overline{x}, \overline{t})$ *of the problem*

$$\min\{F(x,y)|\ g_1(x,y) \le 0,\ g_2(C(x),y) \le 0,$$
$$\overline{v} \ge x,\ \overline{t} \ge d(y),\ x \ge 0,\ y \ge 0\}$$

solves (GBP).

Proof If (x, y) is feasible to (GBP), then $g_1(x,y) \le 0,\ g_2(C(x),y) \le 0$ and taking $v = x,\ t = d(y) = \theta(x)$ we have $t - \theta(v) = 0$, hence, setting $G := \{(v,t)|\ t - \theta(v) \le 0\}$ yields

$$\begin{aligned} F(x,y) \ge \min\{F(x',y')|\ & g_1(x',y') \le 0,\ g_2(C(x'),y) \le 0, \\ & v \ge x',\ t \ge d(y'),\ x' \ge 0,\ y' \ge 0\} \\ = f(v,t) \ge \min\{ & f(v',t')|\ (v',t') \in G\} \\ = \min(\tilde{Q}). & \end{aligned}$$

Conversely, if $t - \theta(v) \le 0$, then the inequalities $v \ge x,\ t \ge d(y)$ imply $d(y) \le t \le \theta)v) \le \theta(x)$, hence

$$\begin{aligned} f(v,t) &= \min\{F(x,y)|\ g_1(x,y) \le 0,\ g_2(C(x),y) \le 0, \\ &\qquad v \ge x,\ t \ge d(y),\ x \ge 0,\ y \ge 0\} \\ &\ge \min\{F(x,y)|\ g_1(x,y) \le 0,\ g_2(C(x),y) \le 0, \\ &\qquad \theta(x) \ge d(y),\ x \ge 0,\ y \ge 0\} \\ &= \min(GDP) \end{aligned}$$

Consequently, $\min(\text{GBP}) = \min(\tilde{Q})$. The rest is clear. □

Thus, under assumption (13.17), (GBP) can be solved by the same method as previously, with $v \in \mathbb{R}^{n_1}$ as main parameter instead of $u \in \mathbb{R}^m$. Since $n_1 < m$ the method works in a space of smaller dimension.

More generally, if C is a linear map with $\text{rank}C = r < m$ such that $E = \text{Ker}C$ satisfies

$$Ex' \ge Ex \Rightarrow C(x') \ge C(x), \tag{13.20}$$

then one can arrange the method to work basically in a space of dimension r rather than m. In fact, writing $E = [E_B, E_N],\ x = \begin{bmatrix} x_B \\ x_N \end{bmatrix}$, where E_B is an $r \times r$ nonsingular matrix, we have for every $v \in Ex$:

$$x = \begin{bmatrix} E_B^{-1} \\ 0 \end{bmatrix} v + \begin{bmatrix} -E_B^{-1} E_N x_N \\ x_N \end{bmatrix}.$$

Hence, setting $Z = \begin{bmatrix} E_B^{-1} \\ 0 \end{bmatrix}$, $z = \begin{bmatrix} -E_B^{-1} E_N x_N \\ x + N \end{bmatrix}$ yields

$$x = Zv + z \quad \text{with} \quad Ez = -E_N x_N + E_N x_N = 0.$$

Since by (13.20) $Ez = 0$ implies $Cz = 0$, it follows that $Cx = C(Zv)$. Now define

$$\theta(v) = \min\{d(y) | \; g_2(C(Zv), y) \le 0, \; y \in \mathbb{R}^{n_2}_+\}, \tag{13.21}$$

$$\begin{aligned} f(v,t) = \min\{F(x,y) | \; & g_1(x,y) \le 0, \; g_2(C(Zv), y) \le 0, \\ & v \ge Ex, \; t \ge d(y), \; x \in \mathbb{R}^{n_1}_+, \; y \in \mathbb{R}^{n_1}_+\}. \end{aligned} \tag{13.22}$$

For any $v' \in \mathbb{R}^r = E(\mathbb{R}^{n_1})$ we have $v' = Ex'$ for some x', so $v' \ge v$ implies $Ex' \ge Ex$, hence by (13.20) $Cx' \ge Cx$, i.e., $C(Zv') \ge C(Zv)$, and therefore, $g_2(C(Zv'), y) \le g_2(C(Zv), y) \le 0$. It is then easily seen that $\theta(v') \le \theta(v)$ and $f(v', t') \ge f(v, t)$ for some $v' \ge v$ and $t' \ge t$, i.e., both $\theta(v)$ and $f(v,t)$ are decreasing. Furthermore, if $(\overline{v}, \overline{t})$ solves the problem

$$\min\{f(v,t) | \; t - \theta(v) \le 0\} \tag{13.23}$$

(so that, in particular, $\overline{t} = \theta(\overline{v})$, and $(\overline{x}, \overline{y})$ solves the problem (13.22) where $v = \overline{v}$, $t = \overline{t} = \theta(\overline{v})$), then, since $\overline{v} = E\overline{x}$, we have $\theta(E\overline{x}) \ge \theta(\overline{v}) \ge d(\overline{y})$. Noting that $Z\overline{v} = ZE\overline{x} = \overline{x}$, this implies that $\overline{y}$ solves $\min\{d(y) | \; g_2(C\overline{x}, y) \le 0\}$, and consequently, $(\overline{x}, \overline{y})$ solves (GBP).

13.2.4 Specialization to (LBP)

In the special case when all functions involved in (GBP) are linear we have the Linear Bilevel Programming problem

$$(\text{LBP}) \qquad \left| \begin{array}{c} \min L(x,y) \qquad \text{s.t.} \\ A_1 x + B_1 y \ge c^1, \quad x \in \mathbb{R}^{n_1}_+, \; y \text{ solves} \\ \min\{\langle d, y\rangle | \; A_2 x + B_2 y \ge c^2, \quad y \in \mathbb{R}^{n_2}_+\} \end{array} \right.$$

where $L(x,y)$ is a linear function, $d \in \mathbb{R}^{n_2}, A_i \in \mathbb{R}^{m_i \times n_1}, B_i \in \mathbb{R}^{m_i \times n_2}, i = 1, 2.$

In this case $C(x) = A_2 x$ while $g_1(x,y) = \max_{i=1,\dots,m_1}(c_1 x - B_1 y)_i \le 0$, $g_2(u,y) = \max_{i=1,\dots,m_2}(c^2 - u - B_2 y)_i \le 0$.

Clearly all assumptions (A1), (A2), and (A3) hold, provided the polytope $D := \{(x,y) \in \mathbb{R}^{n_1}_+ \in \mathbb{R}^{n_2}_+ | \; A_1 x + B_1 y \ge c^1, \; A_2 x + B_2 y \ge c^2\}$ is nonempty. To specialize the MO Algorithm for (GBP) to (*BLP*) define

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad c = \begin{bmatrix} c^1 \\ c^2 \end{bmatrix},$$

$$A_1 = \begin{bmatrix} A_{1,1} \\ \dots \\ A_{1,m_1} \end{bmatrix}, \qquad A_2 = \begin{bmatrix} A_{2,1} \\ \dots \\ A_{2,m_1} \end{bmatrix},$$

$$D = \{(x,y) | \, Ax + By \geq c, \ x \geq 0, \ y \geq 0\}.$$

By Proposition 13.3, (BLP) is equivalent to the monotonic optimization problem $\min\{f(u,t) | \, h(u,t) \leq 0\}$, where for $u \in \mathbb{R}^{m_2}, \ t \in \mathbb{R}$:

$$f(u,t) = \min\{L(x,y) | \, (x,y) \in D, \ u \geq A_2 x, \ t \geq \langle d, y\rangle\},$$
$$h(u,t) = t - \min\{\langle d, y\rangle | \, (x,y) \in D, \ u \geq A_2 x\},$$

with $f(u,t)$ decreasing and $h(u,t)$ increasing.

For convenience set $z = (u,t) \in \mathbb{R}^{m_2+1}$. The basic operations proceed as follows:

– *Initial box:* $M_1 = [a,b] \subset \mathbb{R}^{m_2+1}$ with

$$a_i = \min\{(A_{2,i}, x) | \, (x,y) \in D\}, \quad i = 1, \dots, m_2$$
$$a_{m_2+1} = \min\{\langle d, y\rangle | \, (x,y) \in D\}$$
$$b_i = \max\{\langle A_{2,i}, x\rangle | (x,y) \in D| \quad i = 1, \dots, m_2$$
$$b_{m_2+1} = \max\{\langle d, y\rangle | \, (x,y) \in D\}.$$

– *Reduction operation*: Let $\bar{z} = (\bar{u}, \bar{t})$ = CBS and $\gamma = f(\bar{z})$ =CBV. If $M = [p,q]$ is any subbox of $[a,b]$ still of interest at a given iteration, then the box $\text{red}_\gamma M$ (valid reduction M) is determined as follows:

(1) If $h(z) < 0$ or $f(z) > \gamma$, then $\text{red}_\gamma(M) = \emptyset$ (M is to be deleted);
(2) Otherwise, $\text{red}_\gamma(M) = [p', q']$ with p', q' being determined by formulas (13.12)–(13.13), i.e.,

$$q' = p + \sum_{i=1}^{n} \alpha_i (q_i - p_i) e^i, \quad p' = q' - \sum_{i=1}^{n} \beta_i (q'_i - p_i) e^i,$$

where

$$\alpha_i = \begin{cases} 1, & \text{if } h(q^i) \geq 0 \\ N(\varphi_i), & \text{if } h(q^i) < 0 \end{cases} \quad \varphi_i(\alpha) = h(p + \alpha(q_i - p_i)e^i).$$

$$\beta_i = \begin{cases} 1, & \text{if } f(q'^i) - \gamma \leq 0 \\ N(-\psi_i), & \text{if } f(q'^i) - \gamma > 0 \end{cases} \quad \psi_i(\beta) = f(q' - \beta(q'_i - p_i)e^i) - \gamma.$$

– *Lower bounding*: For any box $M = [p, q]$ with $\mathrm{red}_\gamma(M) = [p', q']$ a lower bound for $f(z)$ over $[p, q]$ is

$$\begin{gathered}\beta(M) = \min_{u,t} L(x, y) \quad \text{s.t.}\\ A_1x + B_1y \geq c^1,\ A_2x + B_2y \geq c^2,\\ A_2x \leq u,\ d(y) \leq t,\ x \geq 0,\ y \geq 0,\\ (u, t) \in [q'^1, \dots, q'^n].\end{gathered}$$

13.2.5 Illustrative Examples

Example 13.1 Solve the problem

$$\begin{gathered}\min(x^2 + y^2) \quad \text{s.t.}\\ x \geq 0, y \geq 0, \quad y \text{ solves}\\ \min(-y) \quad \text{s.t.}\\ 3x + y \leq 15,\\ x + y \leq 7,\\ x + 3y \leq 15.\end{gathered}$$

We have $D = \{(x, y) \in \mathbb{R}^2_+ |\ 3x + y \leq 15; x + y \leq 7, x + 3y \leq 15\}$,

$$\begin{aligned}\theta(u) &= \min\{-y)|\ (x, y) \in D,\ -3x \leq u\},\\ f(u, t) &= \min\{x^2 + y^2 |\ (x, y) \in D, -3x \leq u,\ -y \leq t\}.\end{aligned}$$

Computational results (tolerance 0.0001):
Optimal solution: (4.492188, 1.523438)
Optimal value: 22.500610 (relative error $\leq$ 0.0001)
Optimal solution found and confirmed at iteration 10.

Example 13.2 (Test Problem 7 in Floudas (2000), Chap. 9)

$$\begin{gathered}\min(-8x_1 - 4x_2 + 4y_1 - 40y_2 + 4y_3) \quad \text{s.t.}\\ x_1, x_2 \geq 0, \quad y \text{ solves}\\ \min(y_1 + y_2 + 2_3) \quad \text{s.t.}\\ -y_1 + y_2 + y_3 \leq 1,\\ 2x_1 - y_1 + 2y_2 - 0.5y_3 \leq 1,\\ 2x_2 + 2y_1 - y_2 - 0.5y_3 \leq 1,\\ y_1, y_2, y_3 \geq 0.\end{gathered}$$

$$D = \{(x, y) \in \mathbb{R}^2_+ \times \mathbb{R}^3_+ |\ -y_1 + y_2 + y_3 \leq 1,\ 2x_1 - y_1 + 2y_2 - 0.5y_3 \leq 1,$$
$$2x_2 + 2y_1 - y_2 - 0.5y_3 \leq 1\},$$

$$\theta(u) = \min\{y_1 + y_2 + 2y_3 |\ (x, y) \in D, -x_1 \leq u_1,\ -x_2 \leq u_2\},$$
$$f(u, t) = \min\{-8x_1 - 4x_2 + 4y_1 - 40y_2 + 4y_3 |\ (x, y) \in D,\ -x_1 \leq u_1,$$
$$-x_2 \leq u_2,\ y_1 + y_2 + 2y_3 \leq t\}.$$

Computational results (tolerance 0.01):
Optimal solution $x = (0, 0.896484)$, $y = (0, 0.597656, 0.390625)$
Optimal value: -25.929688 (relative error ≤ 0.01)
Optimal solution found and confirmed at iteration 22.

13.3 Optimization Over the Efficient Set

Given a nonempty set $X \subset \mathbb{R}^n$ and a mapping $C = (C_1, \ldots, C_m) : \mathbb{R}^n \to \mathbb{R}^m$, a point $x^0 \in X$ is said to be *efficient* w.r.t. C if there is no $x \in X$ such that $C(x) \geq C(x^0)$ and $C(x) \neq C(x^0)$. It is said to be *weakly efficient* w.r.t. C if there is no $x \in X$ such that $C(x) > C(x^0)$. These concepts appear in the context of multiobjective programming, when a decision maker has several objectives $C_1(x), \ldots, C_m(x)$ he/she would like to maximize over a feasible set X. Since these objectives may not be consistent with one another, efficient or weakly efficient solutions are used as substitutes for the standard concept of optimal solution.

Denote the set of efficient points by X_E and the set of weakly efficient points by X_{WE}. Obviously, $X_E \subset X_{WE}$. To help the decision maker to select a most preferred solution, the following problems can be considered:

$$\text{(OE)} \qquad \max\{f(x) |\ x \in X,\ g(x) \leq 0,\ x \in X_E\}$$
$$\text{(OWE)} \qquad \max\{f(x) |\ x \in X,\ g(x) \leq 0,\ x \in X_{WE}\}$$

where $f(x), g(x)$ are given functions.

Since $x \in X_{WE}$ if and only if $C(y) - C(x) \leq 0\ \forall y \in X$ the problem (OWE) is a MPEC. It can be proved that (OE), too, is a MPEC, though the proof of this fact is a little more involved (see Philip 1972).

Among numerous studies devoted to different aspects of the two above problems we should mention (Benson 1984, 1986, 1991, 1993, 1995, 1996, 1998, 1999, 2006; Bolintineau 1993a,b; Ecker and Song 1994; Bach-Kim 2000; An-Tao-Muu 2003; An-Tao-Nam-Muu 2010; Muu and Tuyen 2002; Thach et al. 1996). All these approaches are based in one way or another on dc optimization or extensions. Below we present a different approach, first proposed in Luc (2001) and later expanded in Tuy and Hoai-Phuong (2006), which is based on reformulating these problems as monotonic optimization problems.

13.3.1 Monotonic Optimization Reformulation

Assume X is compact and C is continuous. Then the set $C(X)$ is compact and there exists a box $[p, q] \subset \mathbb{R}^m_+$ such that $C(X) \subset [a, b]$.

For every $y \in [a, b]$ define

$$h(y) = \min\{t|\ y_i - C_i(x) \le t,\ i = 1, \dots, m,\ x \in X\}. \tag{13.24}$$

Proposition 13.7 *The function $h : [a, b] \to \mathbb{R}$ is increasing and continuous and there holds*

$$x \in X_{WE} \Leftrightarrow h(C(x)) \ge 0.$$

Proof We can write

$$\begin{aligned} x \in X_{WE} &\Leftrightarrow \forall z \in X\ \exists i\ C_i(x) \le C_i(z) \\ &\Leftrightarrow \max_{y \in Y} \min_i (C_i(x) - y_i) \le 0 \\ &\Leftrightarrow \max_{y \in Y} \min\{t|\ y_i - C_i(x) \le t,\ i = 1, \dots, m\} \ge 0 \\ &\Leftrightarrow h(C(x)) \ge 0. \qquad \square \end{aligned}$$

Since the set $C(X)$ is compact, it is contained in some box $[a, b] \subset \mathbb{R}^m$. For each $y \in [a, b]$ let

$$\varphi(y) := \min\{f(x)|\ x \in X,\ g(x) \le 0,\ C(x) \ge y\}.$$

Clearly for $y' \ge y$ we have $\varphi(y') \le \varphi(y)$, so this is a decreasing function.

Proposition 13.8 *Problem (OWE) is equivalent to the monotonic optimization problem*

$$\min\{\varphi(y)|\ h(y) \ge 0,\ y \in [a, b]\}. \tag{13.25}$$

Proof By Proposition 13.7 problem (OWE) can be written as

$$\min\{f(x)|\ x \in X,\ g(x) \le 0,\ h(C(x)) \ge 0\}. \tag{13.26}$$

This is a problem with hidden monotonicity in the constraints which has been studied in Chap. 11, Sect. 11.5. By Proposition 11.24, it is equivalent to problem (13.25).

13.3.2 Solution Method for (OWE)

For small values of m the monotonic optimization problem (13.25) can be solved fairly efficiently by the copolyblock approximation algorithm analogous to the polyblock algorithm described in Sect. 11.2, Chap. 11. For larger values of m, the SIT Algorithm presented in Sect. 11.3 is more suitable. The latter is a rectangular BRB algorithm involving two specific operations: valid reduction and lower bounding.

a. Valid Reduction

At a given stage of the BRB algorithm for solving (13.25), a feasible solution $\overline{y}$ is known or has been produced which is the best so far available. Let $\gamma = \varphi(\overline{y})$ and let $[p, q] \subset [a, b]$ be a box currently still of interest among all boxes generated in the partitioning process. Since both $\varphi(x)$ and $h(x)$ are increasing, an optimal solution of (13.25) must be attained at a point $y \in [p, q]$ satisfying $h(y) = 0$ (i.e., a lower boundary point of the conormal set $\Omega := \{x \in [a, b] | \ h(y) \geq 0\}$). The search for a better feasible solution than $\overline{y}$ can therefore be restricted to the set

$$\begin{aligned} B_\gamma &:= \{y \in [p, q] | \ h(y) \leq 0 \leq h(y), \ \varphi(y) - \gamma \leq 0\} \\ &= \{y \in [p, q] | \ h(y) \leq 0 \leq h_\gamma(y)\}, \end{aligned} \tag{13.27}$$

where $h_\gamma(y) = \min\{\gamma - \varphi(y), h(y)\}$.

In other words, we can replace the box $[p, q]$ by a smaller one, say $[p', q'] \subset [p, q]$, provided

$$B_\gamma \cap [p', q'] = B_\gamma.$$

A box $[p', q']$ satisfying this condition is referred to as a γ-*valid reduction* of $[p, q]$, and denoted $\text{red}_\gamma[p, q]$.

For any continuous increasing function $\varphi(\alpha) : [0, 1] \to \mathbb{R}$ such that $\varphi(0) < 0 < \varphi(1)$ let $N(\varphi)$ denote the largest value of $\overline{\alpha} \in (0, 1)$ satisfying $\varphi(\overline{\alpha}) = 0$. For every $i = 1, \ldots, n$ let $q^i := p + (q_i - p_i)e^i$, where $e^i_i = 1, \ e^i_j = 0 \ \forall j \neq i$.

Proposition 13.9 *We have*

$$\text{red}_\gamma[p, q] = \begin{cases} \emptyset & \text{if } h(p) > 0 \text{ or } h_\gamma(q')\} < 0, \\ [p', q'], & \text{if } h(p) \leq 0 \ \& \ h_\gamma(q')\} \geq 0. \end{cases}$$

where

$$q' = p + \sum_{i=1}^{n} \alpha_i (q_i - p_i) e^i, \qquad p' = q'_i - \sum_{i=1}^{n} \beta_i (q'_i - p_i) e^i,$$

$$\alpha_i = \begin{cases} 1, & \text{if } h(q^i) \leq 0 \\ N(\varphi_i), & \text{if } h(q^i) > 0. \end{cases} \qquad \varphi_i(t) = h(p + t(q_i - p_i)e^i).$$

$$\beta_i = \begin{cases} 1, & \text{if } h_\gamma(q'^i) \geq 0 \\ N(-\psi_i), & \text{if } h_\gamma(q'^i) < 0. \end{cases} \qquad \psi_i(t) = h_\gamma(q' - t(q'_i - p_i)e^i).$$

Remark 13.3 It is easily verified that the box $[p', q'] = \text{red}_\gamma[p, q]$ still satisfies $\varphi(p') \leq \gamma,\ h(p') \leq 0, h(q') \geq 0$.

b. Lower Bounding

Let $M := [p, q]$ be a partition set which is assumed to have been reduced, so that according to Remark 13.3

$$\varphi(p) \leq \gamma,\ h(p) \leq 0,\ h(q) \geq 0.$$

To compute a lower bound $\beta(M)$ for

$$\min\{\varphi(y) |\ y \in [p, q],\ h(y) = 0\}$$

we note that, since $\varphi(y)$ is increasing, an obvious bound is $\varphi(p)$. It turns out that, as will be shortly seen, the convergence of the BRB algorithm for solving (eqn13.25) will be ensured, provided that

$$\beta(M) \geq \varphi(p). \tag{13.28}$$

A lower bound satisfying this condition will be referred to as a *valid lower bound*.

Define $\theta(y) = \min\{\varphi(y) - \gamma, h(y)\},\ \Delta = \{y \in [p, q] |\ \theta(y) \leq 0\}$.

If $\theta(p) \leq 0 \leq \theta(q)$, then obviously p is an exact minimizer of $\varphi(y)$ over the feasible points in $[p, q]$ at least as good as the current best, and can be used to update the current best solution. Suppose therefore that $\theta(p) \leq 0, h(p) < 0$, i.e., $p \in \Delta \setminus \Omega$.

For each $y \in [p, q]$ such that $h(y) < 0$ let $\pi(y)$ be the first point where the line segment from y to q meet the lower boundary of Ω, i.e.,

$$\pi(y) = y + \lambda(q - y),\ \text{with}$$
$$\lambda = \max\{\alpha |\ h(y + \alpha(q - y)) \leq 0\}.$$

Obviously, $h(\pi(y)) = 0$.

Lemma 13.2 *If* $z = \pi(p), z^i = p + (z_i - p_i)e^i, i = 1, \ldots, m$, *and* $I = \{i |\ z^i \in \Delta\}$, *then a valid lower bound over* $M = [p, q]$ *is*

$$\begin{aligned}\beta(M) &= \min\{\varphi(z^i) |\ i \in I\}\\ &= \min\{f(x) |\ x \in D,\ C(x) \geq z^i, i \in I\}.\end{aligned}$$

Proof Let $M_i = [z^i, q]$. From the definition of z it is easily seen that $h(u) < 0\ \forall u \in [p, z)$, i.e., $[p, q] \cap \Delta \cap \Omega \subset [p, q] \setminus [p, z)$. Noting that $\{u |\ p \leq u < z\} = \cap_{i=1}^m \{u |\ p_i \leq u_i < z_i\}$ we can write $[p, q] \setminus [p, z) = [p, q] \setminus \cap_{i=1}^m \{u |\ u_i < z_i\} = \cup_{i=1}^m \{u \in [p, q] |\ z_i \leq u_i \leq q_i\} = \cup_{i=1}^m M_i$. Thus, if $I = \{i | z^i \in \Delta\}$, then $[p, q] \cap \Delta \cap \Omega \subset \cup_{i=1}^m M_i$. Since $\varphi(z^i) \leq \min\{\varphi(y) |\ y \in M_i\}$, the result follows. □

MO Algorithm for (OWE)

Initialization. Start with $\mathscr{P}_1 = \{M_1\}, M_1 = [a, b], \mathscr{R}_1 = \emptyset$. If a current best feasible solution (CBS) is available let CBV (current best value) denote the value of $f(x)$ at this point. Otherwise, set CBV=$+\infty$. Set $k = 1$.

Step 1. For each bx $M \in \mathscr{P}_k$:

- Compute the γ-valid reduction $\text{red}_\gamma M$ of M for γ=CBV;
- Delete M if $\text{red}_\gamma M = \emptyset$;
- Replace M by $\text{red}_\gamma M$ otherwise;
- If $\text{red}_\gamma M = [p, q]$, then compute a valid lower bound $\beta(M)$ for $\varphi(y)$ over the feasible solutions in M.

Step 2. Let $\mathscr{P}'_k$ be the collection of boxes that results from $\mathscr{P}_k$ after completion of Step 1. From $\mathscr{R}_k$ remove all $M \in \mathscr{R}_k$ such that $\beta(M) \geq$CBV and let $\mathscr{R}'_k$ be the resulting collection. Let $\mathscr{S}_k = \mathscr{R}'_k \cup \mathscr{P}'_k$.

Step 3. If $\mathscr{S}_k = \emptyset$, then terminate: the problem is infeasible if CBV$= +\infty$ or CBV is the optimal value and the feasible solution $\overline{y}$ with $\varphi(\overline{y})$=CBV is an optimal solution if CBV$< +\infty$.
Otherwise, let $M_k \in \text{argmin}\{\beta(M)|\ M \in \mathscr{S}_k\}$.

Step 4. Divide M_k into two subboxes by the standard bisection. Let P_{k+1} be the collection of these two subboxes of M_k.

Step 5. Let $\mathscr{R}_{k+1} = \mathscr{S}_k \setminus \{M_k\}$. Increment k and return to Step 1.

Proposition 13.10 *Either the algorithm is finite, or it generates an infinite filter of boxes $\{M_{k_l}\}$ whose intersection yields a global optimal solution.*

Proof If the algorithm is finite it must generate an infinite filter of boxes $\{M_{k_l}\}$ such that $\cap_{l=1}^{+\infty} M_{k_l} = \{y^*\}$ by exhaustiveness of the standard bisection process. Since $M_{k_l} = [p^{k_l}, q^{k_l}]$ with $h(p^{k_l}) \leq 0 \leq h(q^{k_l})$ and $y^* = \lim p^{k_l} = \lim q^{k_l}$ it follows that $h(y^*) \leq 0 \leq h(q^*)$, so $y*$ is a feasible solution. Therefore, if the problem is infeasible, the algorithm must stop at some iteration where no box remains for consideration while CBV$= +\infty$ (see Step 3). Otherwise,

$$\varphi(p^{k_l}) \leq \beta(M_{k_l}) \leq \varphi(q^{k_l}),$$

whence $\lim_{l\to+\infty} \beta(M_{k_l}) = \varphi(y^*)$. On the other hand, since $\beta(M_{k_l})$ is minimal among the current set $\mathscr{S}_{k_l}$ we have $\beta(M_{k_l}) \leq \min\{\varphi(y)|\ y \in [a, b] \cap \Omega\}$ ad, consequently,

$$\varphi(y^*) \leq \min\{\varphi(y)|\ y \in [a, b] \cap \Omega\}.$$

Since y^* is feasible it must be an optimal solution. $\square$

c. Enhancements When C Is Concave

Tighter Lower Bounds

Recall from (13.25) that (OWE) is equivalent to the monotonic optimization problem

$$\min\{\varphi(y)|\ y \in \Omega\}$$

where $\Omega = \{y \in [a,b]|\ h(y) \geq 0\}$, $h(y) = \min\{t|\ y_i - C_i(x) \leq t, i = 1,\ldots,m, x \in X\}$ and

$$\varphi(y) = \min\{f(x)|\ x \in X,\ g(x) \leq 0,\ C(x) \geq y\}. \tag{13.29}$$

Lemma 13.3 *If the functions $C_i(x), i = 1,\ldots,m$, are concave the set $\Gamma = \{y \in [a,b]|\ h(y) \leq 0\}$ is convex.*

Proof Clearly $\Gamma = \{y \in [a,b]|\ y \leq C(x)$ for some $x \in X\}$. If $y, y' \in \Gamma, 0 \leq \alpha \leq 1$, then $y \leq C(x), y' \leq C(x')$ with $x, x' \in C$ and $\alpha y + (1-\alpha)y' \leq \alpha C(x) + (1-\alpha)C(x') \leq C(\alpha x + (-\alpha)x')$, where $\alpha x + (1-\alpha)x' \in X$ because X is convex by assumption. □

The convexity of Γ permits a simple method for computing a tight lower bound for the minimum of $\varphi(y)$ over the feasible solutions in $[p,q]$.

For each $i = 1,\ldots,m$ let y^i be the last point of Γ on the line segment from p to $p' = p + (q_i - p_i)e^i$, i.e., $y^i = p + \lambda_i(q_i - p_i)e^i$ with

$$\lambda_i = \max\{\alpha|\ p + \alpha(q_i - p_i)e^i \in \Gamma,\ 0 \leq \alpha \leq 1\}.$$

Clearly $h(y^i) = 0$.

Proposition 13.11 *If the functions $C_i(x), i = 1,\ldots,m$ are concave then a lower bound for $\varphi(y)$ over the feasible portion in $M = [p,q]$ is*

$$\begin{aligned}\beta(M) = \min\{f(x)|\ x \in X,\ g(x) \leq 0,\ C(x) \geq p, \\ \sum_{i=1}^{m}(C_i(x) - p_i)/(y_i^i - p_i) \geq 1\}.\end{aligned} \tag{13.30}$$

Proof First observe that by convexity of Γ the simplex $S = [y^1,\ldots,y^m]$ satisfies $\{y \in [p,q]|\ h(y) \leq 0\} \subset S + \mathbb{R}^m_+$ and since $\varphi(y)$ is increasing we have $\varphi(y) \leq \varphi(\pi(y))\ \forall y \in S$, where $\pi(y) = y + \lambda(q-y)$ with $\lambda = \max\{\alpha|\ h(y + \alpha(q-y) \leq 0\}$, i.e., $h(\pi(y)) = 0$. Hence,

$$\begin{aligned}\beta(M) &= \min\{\varphi(y)|\ y \in S\} \\ &\leq \min\{\varphi(y)|\ y \in [p,q], h(y) = 0\},\end{aligned} \tag{13.31}$$

so (13.31) gives a valid lower bound. Further, letting

$$E = \{(x,y)|\ x \in X,\ g(x) \leq 0,\ C(x) \geq y,\ y \in S\}$$
$$E_y = \{x \in X|\ g(x) \leq 0,\ C(x) \geq y\} = \{x|\ (x,y) \in E\} \text{ for every } y \in S,$$

we can write

$$\begin{aligned}&\min_{y\in S} \varphi(y)\\ &\quad = \min_{y\in S} \min\{f(x)|\ x \in E_y\}\\ &\quad = \min\{f(x)|\ (x,y) \in E\}\\ &\quad = \min\{f(x)|\ x \in X,\ g(x) \leq 0,\ C(x) \geq y,\ y \in S\}.\end{aligned}$$

Since $S = \{y \geq p|\ \sum_{i=1}^m (y_i - p_i)/(y_i^i - p_i) = 1\}$, (13.30) follows. □

Remark 13.4 In particular, when X is a polytope, C is a linear mapping, $g(x), f(x)$ are affine functions, as e.g., in problems (OWE) with linear objective criteria, then $\varphi(y)$ and $\beta(M)$ are optimal values of linear programs. Also note that the linear program (13.29) for different y differs only by the right-hand side of the linear constraints and the linear programs (13.30) for different partition sets M differ only by the last constraint and the right-hand side of the constraint $C(x) \geq p$. Exploiting these facts, reoptimization techniques can be used to solve efficiently these linear programs.

Adaptive Subdivision Rule

Let y^M be the optimal solution of problem (13.30) and $z^M = \pi(y^M)$. Since the equality $y^M = z^M$ would imply that $\beta(M) = \min\{\varphi(y)|\ y \in [p,q], h(y) = 0\}$, the following adaptive subdivision rule must be used to accelerate convergence of the algorithm (see Theorem 6.4, Chap. 6):

Determine $v^M = \frac{1}{2}(y^M + z^M)$, $s_M \in \text{argmax}_{i=1,\dots,m} \|y_i^M - z_i^M|$, and subdivide M via (v^M, s_M).

13.3.3 Solution Method for (OE)

We now turn to the constrained optimization problem over the efficient set

$$\text{(OE)} \qquad \alpha := \min\{f(x)|\ x \in X,\ g(x) \leq 0,\ x \in X_E\}$$

where X_E is the set of efficient points of X with respect to $C : \mathbb{R}^n \to \mathbb{R}^m$.

Although efficiency is a much stronger property than weak efficiency, it turns out that the problem (OE) can be solved, roughly speaking, by the same method as (OWE). More precisely, let, as previously, $\Gamma = \{y \in [a,b]|\ x \in X,\ C(x) \geq y\}$. This is a *normal set* in the sense that $y' \leq y \in \Gamma \Rightarrow y' \in \Gamma$ (see Chap. 11);

moreover, it is compact. Recall from Sect. 11.1 (Chap. 11) that a point $z \in \Gamma$ is called an *upper boundary point* of Γ, written $z \in \partial^+\Gamma$, if $(z, b] \subset [a, b] \setminus \Gamma$. A point $z \in \partial^+\Gamma$ is called an *upper extreme point* of Γ, written $z \in \text{ext}^+\Gamma$, if $x \in \Gamma, x \geq z$ implies $x = z$. By Proposition 11.5 a compact normal set such as Γ always has at least one upper extreme point.

Proposition 13.12 $x \in X_E \Leftrightarrow C(x) \in \text{ext}^+\Gamma$.

Proof If $x^0 \in X_E$, then $y^0 = C(x^0) \in \Gamma$ and there is no $x \in X$ such that $C(x) \geq C(x^)$, $C(x) \neq C(x^0)$, i.e., no $y = C(x), x \in X$ such that $y \geq y^0, y \neq y^0$, hence $y^0 \in \text{ext}^+\Gamma$. Conversely, if $y^0 = C(x^0) \in \text{ext}^+\Gamma$ with $x^0 \in X$, then there is no $y = C(x), x \in X$, such that $y \geq y^0$, $y \neq y^0$, i.e., no $x \in X$ with $C(x) \geq C(x^0)$, $C(x) \neq C(x^0)$, hence $x^0 \in X_E$. □

For every $y \in \mathbb{R}^m_+$ define

$$\rho(y) = \min\left\{\sum_{i=1}^{m}(y_i - z_i) \mid z \geq y,\ z \in \Gamma\right\}. \tag{13.32}$$

Proposition 13.13 *The function* $\rho(y) : \mathbb{R}^m_+ \to \mathbb{R}_= \cup \{+\infty\}$ *is increasing and satisfies*

$$\rho(y) = \begin{cases} \leq 0 & \textit{if } y \in \Gamma \\ = 0 & \textit{if } y \in \text{ext}^+\Gamma \\ = +\infty & \textit{if } y \notin \Gamma. \end{cases} \tag{13.33}$$

Proof If $y' \geq y$, then $z \geq y'$ implies that $z \geq y$, hence $\rho(y') \geq \rho(y)$. Suppose $\rho(y) = 0$. Then $y \in \Gamma$ since there exists $z \in \Gamma$ such that $y \leq z$. There cannot be any $z \geq y, z \neq y$, for this would imply $z_> y_i$ for at least one i, hence $\rho(y) \leq \sum_{i=1}^m (y_i - z_i) < 0$, a contradiction. Conversely, if $y \in \text{ext}^+\Gamma$ then for every $z \in \Gamma$ such that $z \geq y$ one must have $z = y$, hence $\sum_{i=1}^m (y_i - z_i) = 0$, i.e., $\rho(y) = 0$. That $\rho(y) \leq 0\ \forall y \in \Gamma$, $\rho(y) = +\infty\ \forall y \neq \Gamma$ is obvious, □

Setting $D = \{x \in X \mid g(x) \leq 0\}$, $\tilde{h}(y) = \min\{h(y), \rho(y)\}$, we can thus formulate the problem (OE) as

$$\min\{f(x) \mid x \in D,\ \tilde{h}(C(x)) \geq 0\}, \tag{13.34}$$

which has the same form as (13.25), with the difference, however, that $\tilde{h}(y)$, though increasing like $h(y)$, is not usc. Upon the same transformation as previously, this problem can be rewritten as

$$\min\{\varphi(y) \mid y \in \Omega,\ \rho(y) = 0\}, \tag{13.35}$$

where $\varphi(y)$ is defined as in (13.29) and $\Omega = \{x \in [a, b] \mid h(y) \geq 0\}$.

Clearly problem (13.35) differs from problem (13.25) only by the additional constraint $\rho(y) = 0$, whose presence is what makes (OE) a bit more complicated than (OWE).

There are two possible approaches for solving (OE). In the first approach proposed by Thach and discussed in Tuy et al. (1996a) the problem (OE) is approximated by an (OWE) differing from (OE) only in that the criterion mapping C is slightly perturbed and replaced by C^ε with

$$C_j^\varepsilon(x) = C_j(x) + \varepsilon \sum_{i=1}^{m} C_i(x), \quad j = 1, \ldots, m,$$

where $\varepsilon > 0$ is sufficiently small. Denote by X_{WE}^ε and X_E^ε resp. the weakly efficient set and the set of efficient set of X w.r.t. $C^\varepsilon(x)$ and consider the problem

$$(\text{OWE}^\varepsilon) \qquad \alpha(\varepsilon) := \min\{f(x)|\ g(x) \leq 0,\ x \in X_{WE}^\varepsilon\}.$$

Proposition 13.14 (i) *For any* $\varepsilon > 0$ *we have* $X_{WE}^\varepsilon \subset X_E$.

(ii) $\alpha(\varepsilon) \to \alpha$ *as* $\varepsilon \searrow 0$.

(iii) *If* X *is a polytope and* C *is linear, then there is* $\varepsilon_0 > 0$ *such that* $X_{WE}^\varepsilon = X_E$ *for all* $\varepsilon \in (0, \varepsilon_0)$.

Proof (i) Let $x^* \in X \setminus X_E$. Then there is $x \in X$ satisfying $C(x) \geq C(x^*)$, $C_i(x) > C_i(x^*)$ for at least one i. This implies $\sum_{i=1}^m C_i(x) > \sum_{i=1}^m C_i(x^*)$, hence $C^\varepsilon(x) > C^\varepsilon(x^*)$, so that $x^* \notin X_{WE}^\varepsilon$. Thus, if $x^* \in X \setminus X_E$ then $x^* \in X \setminus X_{WE}^\varepsilon \subset X_E$.

(ii) Property (i) implies that $\alpha(\varepsilon) \geq \alpha$. Define

$$\Gamma^\varepsilon = \{y \in [a, b]|\ y \leq C^\varepsilon(x),\ x \in X\}, \tag{13.36}$$

$$\varphi_\varepsilon(y) = \min\{f(x)|,\ x \in X,\ g(x) \leq 0,\ C^\varepsilon(x) \geq y\}, \tag{13.37}$$

so that problem (OWE^ε) reduces to

$$\min\{\varphi_\varepsilon(y)|\ y \in \mathrm{cl}([a, b] \setminus \Gamma^\varepsilon\}. \tag{13.38}$$

Next consider an arbitrary point $\overline{x} \in X_E \setminus X_{WE}^\varepsilon$. Since $\overline{x} \notin X_{WE}^\varepsilon$ there exists $x^\varepsilon \in X$ such that $C^\varepsilon(x^\varepsilon) > C^\varepsilon(\overline{x})$. Then x^ε is feasible to problem (OWE^ε) and, consequently, $f(x^\varepsilon) \geq \alpha(\varepsilon) \geq \alpha$. In view of the compactness of X, by passing to a subsequence if necessary, one can suppose that $x^\varepsilon \to x^* \in X$ as $\varepsilon \searrow 0$. Therefore, $f(x^*) \geq \lim \alpha(\varepsilon) \geq \alpha$. But from $C^\varepsilon(x^\varepsilon) > C^\varepsilon(\overline{x})$ we have $C(x^*) \geq C(\overline{x})$ and since $\overline{x} \in X_E$ it follows that $x^* = \overline{x}$, and hence, $f(\overline{x}) \geq \lim \alpha(\varepsilon) \geq \alpha$. This being true for arbitrary $\overline{x} \in X_E \setminus X_{WE}^\varepsilon$ we conclude that $\lim \alpha(\varepsilon) = \alpha$.

(iii) We only sketch the proof, referring the interested reader to Tuy et al. (1996b, Proposition 15.1), for detail. Since $C(x)$ is linear, we can consider the cones $K := \{x|\ C_i(x) \geq 0, i = 1, \ldots, m\}$, $K^\varepsilon := \{x|\ C^\varepsilon(x) \geq 0, i = 1, \ldots, m\}$. Let $x^0 \in X_E$ and denote by F the smallest facet of X that contains x^0. Since

$x^0 \in X_E$ we have $(x^0 + K) \cap F = \emptyset$, and hence, there exits $\varepsilon_F > 0$ such that $(x^0 + K^\varepsilon) \cap F = \emptyset \ \forall \varepsilon \in (0, \varepsilon_F)$. Then we have $(x + K^\varepsilon) \cap F = \emptyset \ \forall x \in F$ whenever $0 < \varepsilon \leq \varepsilon_F$. If $\mathscr{F}$ denotes the set of facets of X such that $F \subset X_E$, then $\mathscr{F}$ is finite and $\varepsilon_0 = \min\{\varepsilon_F |\ F \in \mathscr{F}\} > 0$. Hence, $(x + K^\varepsilon) \cap F = \emptyset \ \forall x \in F, \forall F \in \mathscr{F}$, i.e., $X_E \subset X^\varepsilon_{WE}$ whenever $0 < \varepsilon < \varepsilon_0$. □

As a consequence, an approximate optimal solution of (OE) can be obtained by solving (OWE$^\varepsilon$) for any ε such that $0 < \varepsilon \leq \varepsilon_0$.

An alternative approach to (OE) consists in solving the equivalent monotonic optimization problem (13.25) by a BRB algorithm analogous to the one for solving problem (OWE). To take account of the peculiar constraint $\rho(y) = 0$ some precautions and special manipulations are needed

(i) A "feasible solution" is an $y \in [a, b]$ satisfying $\rho(y) = 0$, i.e., a point of $\mathrm{ext}^+ \Gamma$. The current best solution is the best among all so far known feasible solutions.
(ii) In Step 1, after completion of the reduction operation described above, a special supplementary reduction–deletion operation is needed to identify and fathom any box $[p, q]$ that contains no point of $\mathrm{ext}^+ \Gamma$. This supplementary operation will ensure that every filter of partition sets $[p^{k_\nu}, q^{k_\nu}]$ collapses to a point corresponding to an efficient point.

Define $h^\varepsilon(y) = \min\{t|\ x \in X, y_i - C^\varepsilon_i(x) \leq t, i = 1, \ldots, m\}$, $\Gamma^\varepsilon = \{y \in [a, b]|\ y \leq C^\varepsilon(x),\ x \in X\}$ where $C^\varepsilon_j(x) = C_j(x) + \varepsilon \sum_{i=1}^m C_i(x), \quad j = 1, \ldots, m$, Since $C(x) \geq a \in \mathbb{R}^m_+ \ \forall x \in X$, we have $C(x) \leq C^\varepsilon(x) \ \forall x \in X$, hence $\Gamma \subset \Gamma^\varepsilon$.

Let $[p, q]$ be a box which has been reduced as described in Proposition 13.9, so that $p \in \Gamma, q \notin \Gamma$. Since $p \in \Gamma$ we have $\rho(p) \leq 0$.

Supplementary Reduction–Deletion Rule: Fathom $[p, q]$ in either of the following events:

(a) $\rho(q) \leq 0$ (so that $\rho(y) \leq 0 \ \forall y \in \Gamma \cap [p, q] \supset \partial^+ \Gamma \cap [p, q]$); if $\rho(q) = 0$, then q gives an efficient point and can be used to update the current best.
(b) $h^\varepsilon(q) \leq 0$ for some small enough $\varepsilon > 0$ (then $q \in \Gamma^\varepsilon$, so the box $[p, q]$ contains no point of $\partial^+(\Gamma^\varepsilon)$, and hence no point of $\mathrm{ext}^+ \Gamma$).

Proposition 13.15 *The BRB algorithm applied to problem (13.35) with precaution (i) and the supplementary reduction–deletion rule either yields an optimal solution to (OE) in finitely many iterations or generates an infinite filter of boxes whose intersection is a global optimal solution y^* of (13.35), corresponding to an optimal solution x^* of (OE).*

Proof Because of precautions (i) and (ii) the box selected for further partitioning in each iteration contains either a point of $\mathrm{ext}^+ \Gamma$ or a point of $\mathrm{ext}^+ \Gamma^\varepsilon$. Therefore, any infinite filter of boxes generated by the algorithm collapses either to a point $y^* \in \mathrm{ext}^+ \Gamma$ (yielding then an $x^* \in X_E$) or to a point $y^* \in \mathrm{ext}^+ \Gamma^\varepsilon$ (yielding then $x^* \in X^\varepsilon_{WE} \subset X_E$). □

13.3.4 Illustrative Examples

Example 13.3 *Problem (OWE) with linear input functions.* Solve

$$\min\{f(x)|\ x \in X,\ g(x) \le 0,\ x \in X_{WE}$$

with following data:

$$\begin{aligned} f(x) &= \langle c, x\rangle,\ c = (4, -8, -3, -1, -7, 0, -6, -2, 9, -3), \\ X &= \{x \in \mathbb{R}^{10}_{+}|\ A_1 x \le b_1,\ A_2 x \ge b_2\}, \end{aligned}$$

where

$$\begin{aligned} A_1 &= \begin{bmatrix} 7 & 1 & -3 & -7 & 0 & 9 & 2 & 1 & -5 & 1 \\ -5 & 8 & -1 & 7 & 5 & 0 & -1 & -3 & 4 & 0 \\ 2 & -1 & 0 & -2 & 3 & -2 & 2 & -5 & -1 & -3 \\ -1 & -4 & 2 & 9 & -4 & 3 & -3 & 4 & 0 & -2 \\ -3 & -1 & 0 & 8 & -3 & -1 & -2 & -2 & 5 & -5 \end{bmatrix} \\ b_1 &= (-66, 150, -81, 79, 53) \\ A_2 &= \begin{bmatrix} -2 & 9 & -1 & -2 & 2 & 1 & 4 & -1 & 5 & 2 \\ -2 & 3 & 2 & 4 & 5 & 4 & 1 & -9 & -2 & -1 \\ -4 & -8 & 1 & 1 & -5 & 3 & -2 & 0 & -2 & 9 \\ 2 & 7 & -1 & -2 & -5 & -9 & 4 & -1 & -2 & 0 \\ 2 & 3 & -1 & -1 & 4 & 3 & -1 & 0 & 0 & -6 \\ -6 & 0 & 0 & 0 & -4 & 3 & -2 & -2 & 4 & -6 \\ 2 & 2 & 3 & -5 & 6 & -4 & 0 & 0 & -1 & -4 \\ -1 & 4 & 4 & 6 & 0 & 3 & -4 & 2 & -4 & -1 \end{bmatrix} \\ b_2 &= (126, 12, -52, -23, 2, -23, -28, 90) \\ g(x) &:= Gx \le d \end{aligned}$$

where

$$\begin{aligned} G &= \begin{bmatrix} -1 & 1 & 2 & 4 & -4 & 4 & -1 & -4 & -6 & 3 \\ 4 & 9 & 0 & -1 & -2 & 1 & -6 & 5 & 0 & 0 \end{bmatrix} \\ d &= (-14, 89) \\ C &= \begin{bmatrix} 5 & 1 & 7 & 1 & 4 & 9 & 0 & -4 & -3 & 7 \\ -1 & -2 & -5 & -4 & -1 & -6 & -4 & 0 & -3 & 0 \\ -3 & -3 & 0 & 4 & 0 & 1 & -2 & 1 & 4 & 0 \end{bmatrix} \end{aligned}$$

Computational results:
Optimal solution (tolerance $\eta = 0.01$):

$$x^* = (0, 8.741312, 8.953411, 9.600323, 3.248449, 3.916472, 6.214436, 9.402384, 10, 4.033163)$$

(found at iteration 78 and confirmed at iteration 316)

Optimal value: -107.321065
Maximal number of nodes generated: 32

Example 13.4 *Problem (OE) with linear-fractional multicriteria*
(taken from Muu and Tuyen (2002))
Solve

$$\min\{f(x)|\ x \in X,\ x \in X_E\}$$

with following data:

$$f(x) = -x_1 - x_2$$
$$X = \{x \in \mathbb{R}^2_+ |\ Ax \le b\}$$

where

$$A = \begin{bmatrix} 1 & -2 \\ -1 & -2 \\ -1 & 1 \\ 1 & 0, \end{bmatrix},$$

$$b = (2, -2, 1, 6)^T,$$
$$C_1(x) = \frac{x_1}{x_1 + x_2},$$
$$C_2(x) = \frac{-x_1 + 6}{x_1 - x_2 + 3}.$$

Computational results:
Optimal solution (tolerance $\eta = 0.01$):
$x^* = (1.991962, 2.991962)$
(found at iteration 16 and confirmed at iteration 16),
Optimal value: -4.983923.

13.4 Equilibrium Problem

Given a nonempty compact set $C \subset \mathbb{R}^n$, a nonempty closed convex set $D \subset \mathbb{R}^m$, and a function $F(x, y) : C \times D \to \mathbb{R}$ which is upper semi-continuous in x and quasiconvex in y, the problem is to find an equilibrium of the system (C, D, F), i.e., a point $\overline{x} \in \mathbb{R}^n$ satisfying

$$\overline{x} \in C,\ F(\overline{x}, y) \ge 0\ \forall y \in D. \tag{13.39}$$

(cf. Sect. 3.3).

By Theorem 3.4 we already know that this problem has a solution, provided the following assumption holds:

(A) *There exists a closed set-valued function φ from D to C with nonempty compact values $\varphi(y) \subset C$ $\forall y \in D$, such that*

$$\inf_{y\in D} \inf_{x\in\varphi(y)} F(x, y) \geq 0. \tag{13.40}$$

The next question that arises is how to find effectively the equilibrium point when condition (13.40) is fulfilled.

As it turns out, this is a question of fundamental importance both for the theory and applications of mathematics.

Historically, about 40 years ago the so-called combinatorial pivotal methods were developed for computing fixed point and economic equilibrium, which was at the time a very popular subject. Subsequently, the interest switched to variational inequalities—a new class of equilibrium problems which have come to play a central role in many theoretical and practical questions of modern nonlinear analysis. With the rapid progress in nonlinear optimization, including global optimization, nowadays equilibrium problems and, in particular, variational inequalities can be approached and solved as equivalent optimization problems.

13.4.1 Variational Inequality

Let C be a compact convex set in $\mathbb{R}^n$, $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ a continuous mapping. Consider the variational inequality

$$\text{(VI)} \qquad \text{Find} \quad \overline{x} \in C, \quad \text{s.t.} \quad \langle \varphi(\overline{x}), x - \overline{x} \rangle \geq 0 \quad \forall x \in C.$$

For $x, y \in C$ define

$$F(x, y) = \max\{\langle x, y - z\rangle \mid z \in C\} = \langle x, y\rangle - \min_{z\in C}\langle x, z\rangle. \tag{13.41}$$

Writing $\min_{z\in C}\langle x, z\rangle = \langle x, \overline{z}\rangle$ for $\overline{z} \in C$ we see that the function $F(x, y)$ is continuous in x and convex in y. By Theorem 3.4 there exists $\overline{x} \in C$ such that

$$\inf_{y\in C} F(\overline{x}, y) \geq \inf_{y\in C} F(\varphi(y), y).$$

But $\inf_{y\in C} F(\overline{x}, y) = \inf_{y\in C}\langle \overline{x}, y\rangle - \min_{z\in C}\langle \overline{x}, z\rangle = 0$, so

$$\inf_{y\in C} F(\varphi(y), y) \leq 0. \tag{13.42}$$

On the other hand, $F(\varphi(y), y) = \langle\varphi(y), y\rangle - \min_{z\in C}\langle\varphi(y), z\rangle \geq 0 \;\; \forall y \in C$, hence, taking account of the continuity of $F(\varphi(y), y)$ and the compactness of C:

$$\min_{y\in C} F(\varphi(y), y) = 0. \tag{13.43}$$

If $\overline{y}$ is a global optimal solution of this optimization problem, then $F(\varphi(\overline{y}), \overline{y}) = 0$, i.e., $0 = \langle\varphi(\overline{y}), \overline{y}\rangle - \min_{z\in C}\langle\varphi(\overline{y}), z\rangle$, hence $\langle\varphi(\overline{y}), \overline{y}\rangle = \min_{y\in C}\langle\varphi(\overline{y}), y\rangle$, i.e., $\langle\varphi(\overline{y}), y - \overline{y}\rangle \geq 0 \; \forall y \in C$, which means $\overline{y}$ is a solution to (VI). Conversely, it is easily seen that any solution $\overline{y}$ of (VI) satisfies $F(\varphi(\overline{y}), \overline{y}) = 0$.

Thus, under the stated assumptions the variational inequality (VI) is solvable and any solution of it can be found by solving the global optimization problem

$$\min\{F(\varphi(y), y) |\; y \in C\}, \tag{13.44}$$

where

$$F(\varphi(y), y) = \langle\varphi(y), y\rangle - \min_{z\in C}\langle\varphi(y), z\rangle \geq 0 \;\; \forall y \in C. \tag{13.45}$$

It is important to note that the optimization problem (13.44) should be solved to *global optimality* because $F(\varphi(y), y)$ is generally nonconvex and a local minimizer $\overline{y}$ of it over C may satisfy $F(\varphi(\overline{y}), \overline{y}) > 0$, so may fail to provide a solution to the variational inequality.

A function $g(y)$, such that $g(y) \geq 0 \; \forall y \in C$ and $g(\overline{y}) = 0$ for $\overline{y} \in C$ if and only if $\overline{y}$ solves the variational inequality (VI), is often referred to as a *gap function* associated with (VI).

It follows from the above that the function $F(\varphi(y), y)$ defined by (13.45), which was first introduced by Auslender (1976), is a gap function for (VI).

Thus, once a gap function $g(y)$ is known for the variational inequality (VI), solving the latter reduces to solving the equation

$$y \in C, \;\; g(y) = 0, \tag{13.46}$$

which in turn is equivalent to the *global optimization* problem

$$\min\{g(y) |\; y \in C\}. \tag{13.47}$$

The existence theorem for variational inequalities and the conversion of variational inequalities into equivalent equations date back to Hartman and Stampacchia (1966). The use of gap functions to convert variational inequalities into optimization problems (13.47) appeared subsequently (Auslender 1976). However, in the earlier period, when global optimization methods were underdeveloped, the converted optimization problem was solved mainly by local optimization methods, i.e., by methods seeking a local minimizer, or even a stationary point satisfying necessary optimality conditions. Since the function $g(y)$ may not be convex such points do not always satisfy the required inequality in (VI) for all $y \in C$, so strictly speaking they do not solve the variational inequality. Therefore, additional ad-hoc manipulations have often to be used to guarantee the global optimality of the solution found.

Aside from the function (13.45) introduced by Auslender, many other gap functions have been developed and analyzed in the last two decades: (Fukushima 1992), (Zhu and Marcotte 1994), (Mastroeni 2003),... For a gap function to be practically useful it should possess some desirable properties making it amenable to currently available optimization methods. In particular, it should be differentiable, if differentiable optimization methods are to be used. From this point of view, the following function, discovered independently by Auchumuty (1989) and Fukushima (1992), is of particular interest:

$$g(y) = \max_{z \in C} \left\{ \langle \varphi(y), z - y \rangle - \frac{1}{2} \langle z - y, M(z - y) \rangle \right\}, \tag{13.48}$$

where M is an arbitrary $n \times n$ symmetric positive definite matrix. Let

$$h(z) = \langle \varphi(y), z - y \rangle - \frac{1}{2} \langle z - y, M(z - y) \rangle. \tag{13.49}$$

For every fixed $y \in C$, since $h(z)$ is a strictly concave quadratic function, it has a unique maximizer $u = u(y)$ on the compact convex set C which is also a minimizer of the strictly convex quadratic function $-h(z)$ on C and hence, is completely defined by the condition $0 \in -\nabla h(u) + N_X(u)$ (see Proposition 2.31), i.e.,

$$\langle M(u(y) - y) - \varphi(y),\ z - u(y) \rangle \geq 0 \quad \forall z \in C. \tag{13.50}$$

Proposition 13.16 (Fukushima 1992) *The function (13.48) is a gap function for the variational inequality (VI). Furthermore, a vector $\overline{y} \in C$ solves (VI) if and only if $\overline{y} = u(\overline{y})$, i.e., if and only if $\overline{y}$ is a fixed point of the mapping $u : C \to C$ that carries each point $y \in C$ to the maximizer $u(y)$ of the function $h(z)$ over the set C.*

Proof Clearly $h(y) = 0$, so $g(y) = \max_{z \in C} h(z) \geq h(y) = 0\ \forall y \in C$. We show that $\overline{y} \in C$ solves (VI) if and only if $g(\overline{y}) = 0$. Since $g(y) = \max_{z \in C} h(z)$ we have $g(y) = \langle \varphi(y), u(y) - y \rangle - \frac{1}{2} \langle u(y) - y, M(u(y) - y) \rangle$. It follows that $g(\overline{y}) = 0$ for $\overline{y} \in C$ if and only if $\overline{y} = u(\overline{y}), g(\overline{y}) = \min_{z \in C} g(y)$, and $\langle \varphi(\overline{y}), z - \overline{y} \rangle \geq 0,\ \forall z \in C$ [see (13.50)], consequently, if and only if $\overline{y}$ is a solution of (VI). □

It has also been proved in Fukushima (1992) that under mild assumptions on continuity and differentiability of $\varphi(y)$ the function (13.48) is also continuous and differentiable and that any stationary point of it on C solves (VI). Consequently, with the special gap function (13.48) and under appropriate assumptions (VI) can be solved by using conventional differentiable local optimization methods to find a stationary point of the corresponding problem (13.47).

The next proposition indicates two important cases when the problem (13.44) using the gap function (13.45) can be solved by well-developed global optimization methods.

Proposition 13.17 (i) *If the map $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ is affine, then the gap function (13.45) is a dc function and the corresponding optimization problem equivalent to (VI) is a dc optimization problem.*

(ii) *If $C \subset \mathbb{R}^m_+$ and each function $\varphi_i(y) : \mathbb{R}^n_+ \to \mathbb{R}_+,\ i = 1, \ldots, m$, is an increasing function then the gap function (13.45) is a dm function and the corresponding optimization problem equivalent to (VI) is a dm optimization problem.*

Proof (i) If $\varphi(y)$ is affine, then $\langle \varphi(y), y\rangle$ is a quadratic function, while $\min_{z\in C}\langle \varphi(y), z\rangle$ is a concave function, so (13.45) is a dc function.

(ii) If every function $\varphi_i(y)$ is increasing then for any $y, y' \in \mathbb{R}^m_+$ such that $y \geq y'$ by writing

$$\langle \varphi(y), y\rangle - \varphi(t'), y'\rangle = [\langle \varphi(y), y\rangle - \langle \varphi(y'), y\rangle] + [\langle \varphi(y'), y\rangle - \langle \varphi(y'), y'\rangle],$$

it is easily seen that $\langle \varphi(y), y\rangle$ is an increasing function. Since it is also obvious that $\min_{z\in C}\langle \varphi(y), z\rangle$ is an increasing function, (13.45) is a dm function. □

In the case (i) where, in addition, C is a polytope, (VI) is called an *affine variational inequality*.

13.5 Optimization with Variational Inequality Constraint

A class of MPECs which has attracted much research in recent years includes optimization problems with a variational inequality constraint. These are problems of the form

$$(\text{OVI}) \qquad \begin{array}{ll} \text{minimize} & f(x, y) \\ \text{subject to} & \left| \begin{array}{l} (x, y) \in U,\ y \in C, \\ \langle \varphi(x, y), z - y\rangle \geq 0 \quad \forall z \in C, \end{array} \right. \end{array}$$

where $f(x, y) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a continuous function, $U \subset \mathbb{R}^n \times \mathbb{R}^m$ a nonempty closed set, and $C \subset \mathbb{R}^m$ is a nonempty compact set. For each given $x \in X := \{x \in \mathbb{R}^n |\ (x, y) \in U$ for some $y \in \mathbb{R}^m\}$ the set $S(x) := \{y \in C |\ \langle \varphi(x, y), z - y\rangle \geq 0\ \forall z \in C\}$ is the solution set of a variational inequality. So (OVI) is effectively an MPEC as defined at the beginning of this chapter.

Among the best known solution methods for MPECs let us mention the nonsmooth approach by Outrata et al. (1998), the implicit programming approach and the penalty interior point method by Luo et al. (1996).

Using a gap function $g(y)$ we can convert (OVI) into an equivalent optimization problem as follows:

Proposition 13.18 (OVI) *is equivalent to the optimization problem*

$$\begin{gathered}\min f(x,y) \quad s.t. \\ (x,y) \in U,\ y \in C,. \\ g(y) \leq 0.\end{gathered}$$

Proof Immediate, because by definition of a gap function, $y \in C$ solves (VI) if and only if $g(y) = 0$. □

Of particular interest is (OVI) when (VI) is an *affine variational inequality*, i.e., when φ is an affine mapping and C is a polytope. The problem is then MPEC with an affine equilibrium constraint and is referred to as MPAEC:

$$\text{(MPAEC)} \qquad \begin{array}{ll} \text{minimize} & f(x,y) \\ \text{subject to} & \left| \begin{array}{l} (x,y) \in U,\ y \in C, \\ \langle Ax + By + c, z - y\rangle \leq 0 \ \ \forall z \in C, \end{array}\right. \end{array}$$

where $f(x,y) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a convex function, $U \subset \mathbb{R}^n \times \mathbb{R}^m$ a nonempty closed set, A, B are appropriate real matrices, $c \in \mathbb{R}^n$ and $C \subset \mathbb{R}^n$ a nonempty polytope.

By Proposition 13.18, using a gap function $g(x,y)$ we can reformulate (MPAEC) as the optimization problem

$$\begin{gathered}\min f(x,y) \quad \text{s.t.} \\ (x,y) \in U, \\ g(x,y) \leq 0.\end{gathered} \tag{13.51}$$

If the gap function used is (13.48), i.e.,

$$g(x,y) = \max_{z\in C} \left\{ \langle \varphi(y), z - y\rangle - \frac{1}{2}\langle z - y, M(z-y)\rangle \right\},$$

where M is a symmetric $n \times n$ positive definite matrix, then (13.51) is a differentiable optimization problem for which, as shown by Fukushima (1992), even a stationary point yields a global optimal solution, hence a solution to MPAEC. Exploiting this property which is rather exceptional for gap functions, descent methods and other differentiable (local) optimization methods can be developed for solving MPAEC: see Fukushima (1992), and also Muu-Quoc-An-Tao, (2012) among others.

If the gap function (13.45) is used, we have

$$g(x,y) = \langle Ax + By + c, y\rangle - \min_{z\in C}\langle Ax + By + c, z\rangle,$$

which is a dc function (more precisely, the difference of a quadratic function and a concave function). In this case, (13.51) is a convex minimization under dc constraints, i.e., a tractable global optimization problem.

Since the feasible set of problem (13.51) is nonconvex, an efficient method for solving it, as we saw in Chap. 7, Sect. 7.4, is to proceed according to the following SIT scheme.

For any real number γ denote by (SP_γ) the auxiliary problem of finding a feasible solution (x, y) to (MPAEC) such that $f(x, y) \leq \gamma$. This problem is a dc optimization under convex constraints since it can be formulated as

$$(SP_\gamma) \qquad \begin{array}{c} \min g(x, y) \quad \text{s.t.} \\ (x, y) \in U, \\ f(x, y) \leq \gamma. \end{array} \tag{13.52}$$

Consequently, as argued in Sect. 6.2, it should be much easier to handle than (13.51) whose feasible set is nonconvex. In fact this is the rationale for trying to reduce (MPAEC) to a sequence of subproblems (SP_γ).

Given a tolerance $\eta > 0$, a feasible solution $(\overline{x}, \overline{y})$ of (MPAEC) is said to be an η-optimal solution if $f(\overline{x}, \overline{y}) \leq f(x, y) + \eta$ for every feasible solution (x, y).

Basically, the SIT approach consists in replacing (MPAEC) by a sequence of auxiliary problems (SP_γ) where the parameter γ is gradually adjusted until a suitable value is reached for which by solving the corresponding (SP_γ) an η-optimal solution of (MPAEC) will be found.

Suppose (x^0, y^0) is the best feasible solution available at a given stage and let $\gamma_0 = f(x^0, y^0) - \eta$. Solving the problem (SP_{γ_0}) leads to two alternatives: either a solution (x^1, y^1) is obtained, or an evidence is found that (SP_{γ_0}) is infeasible. In the former alternative, we set $(x^0, y^0) \leftarrow (x^1, y^1)$ and repeat, with $\gamma_1 = f(x^1, y^1) - \eta$ in place of γ_0. In the latter alternative, $f(x^0, y^0) \leq f(x, y) + \eta$ for any feasible solution (x, y) of MPAEC, so (x^0, y^0) is an η-optimal solution as desired and we stop. In that way each iteration improves the objective function value by at least $\eta > 0$, so the procedure must terminate after finitely many iterations, yielding eventually an η-optimal solution.

Formally, we can state the following algorithm.

By translating if necessary it can be assumed that U is a compact convex subset of a hyperrectangle $[a, b] \subset \mathbb{R}^n_+ \times \mathbb{R}^m_+$ and C is a compact convex subset of $\mathbb{R}^m_+$. A vector $u = (x, y) \in U$ is said to be ε-*feasible* to (MPAEC) if it satisfies $g(u) \leq \varepsilon$; a vector $\overline{u} = (\overline{x}, \overline{y}) \in U$ is called an (ε, η)-*optimal solution* if it is ε-feasible and satisfies $f(u) \leq f(\overline{u}) + \eta$ for all ε-feasible solutions.

Algorithm for (MPAEC)

Select tolerances $\varepsilon > 0, \eta > 0$. Let $\gamma_0 = f(a)$.

Step 0. Set $\mathscr{P}_1 = \{M_1\}, M_1 = [a, b], \mathbb{R}_1 = \emptyset, \gamma = \gamma_0$. Set $k = 1$.

Step 1. For each box (hyperrectangle) $M \in \mathscr{P}_k$:

- Reduce M, i.e., find a box $[p, q] \subset M$ as small as possible satisfying $\min\{g(u) \mid f(u) \leq \gamma,\ u \in [p, q]\} = \min\{g(u) \mid f(u) \leq \gamma,\ u \in M\}$. Set $M \leftarrow [p, q]$.
- Compute a lower bound $\beta(M)$ for $g(u)$ over the feasible solutions in $[p, q]$.

Delete every M such that $\beta(M) > \varepsilon$.

Step 2. Let $\mathscr{P}'_k$ be the collection of boxes that results from $\mathscr{P}_k$ after completion of Step 1. Let $\mathscr{R}'_k = \mathscr{R}_k \cup \mathscr{P}'_k$.

Step 3. If $\mathscr{R}'_k = \emptyset$, then *terminate*: if $\gamma = \gamma_0$ the problem (MPAEC) is ε-infeasible (has no ε-feasible solutions); if $\gamma = f(\overline{u}) - \eta$ for some ε-feasible solution $\overline{u}$, then $\overline{u}$ is an (ε, η)-optimal solution of (MPAEC).

Step 4. If $\mathscr{R}'_k \neq \emptyset$, let $M_k \in \operatorname{argmin}\{\beta(M) | \, M \in \mathscr{R}'_k\}$, $\beta_k = \beta(M_k)$. Determine $u^k \in M_k$ and $v^k \in M_k$ such that

$$f(u^k) \leq \gamma, \quad g(v^k) - \beta(M_k) = o(\|u^k - v^k\|). \tag{13.53}$$

If $g(u^k) < \varepsilon$, go to Step 5. If $g(u^k) \geq \varepsilon$, go to Step 6.

Step 5. u^k is an ε-feasible solution satisfying $f(u^k) \leq \gamma$.
If $\gamma = \gamma_0$ reset $\gamma \leftarrow f(\overline{u}) - \eta, \overline{u} = u^k$, and go to Step 6. If $\gamma = f(\overline{u}) - \eta$ and $f(u^k) < f(\overline{u})$ reset $\overline{u} \leftarrow u^k$ and go to Step 6.

Step 6. Divide M_k into two subboxes by the adaptive bisection, i.e., bisect M_k via (w^k, j_k), where $w^k = \frac{1}{2}(u^k + v^k)$, $j_k \in \operatorname{argmax}\{|v^k_j - u^k_j| : j = 1, \ldots, n\}$.
Let $\mathscr{P}_{k+1}$ be the collection of these two subboxes of M_k, $\mathscr{R}_{k+1} = \mathscr{R}'_k \setminus \{M_k\}$.
Increment k, and return to Step 1.

Proposition 13.19 *The above SIT Algorithm terminates after finitely many iterations, yielding either an* (ε, η)*-optimal solution to MPAEC, or an evidence that the problem is infeasible.*

Proof Immediate. □

Remark 13.5 If a feasible solution can be found by some fast local optimization method available, it should be used to initialize the algorithm. Also if for some reason the algorithm has to be stopped prematurely, some reasonably good feasible solution may have been already obtained. This is in contrast with other algorithms which would become useless in that case.

13.6 Exercises

1 Solve the Linear Bilevel Program (Tuy et al. 1993a):

$$\min(-2x_1 + x_2 + 0.5y_1) \quad \text{s.t.}$$

$$x_1, x_2 \geq 0, \quad y = (y_1, y_2) \text{ solves}$$

$$\min(-4y_1 + y_2) \quad \text{s.t.}$$

$$-2x_1 + y_1 - y_2 \leq -2.5$$

$$x_1 - 3x_2 + y_2 \leq 2$$

$$y_1, y_2 \geq 0.$$

Hints: a non-optimal feasible solution is $x = (1,0), y = (0.5,1)$.

2 Solve the Convex Bilevel Program (Floudas 2000, Sect. 9.3):

$$\min((x-5)^2 + (2y+1)^2) \quad \text{s.t.}$$
$$x \geq 0, \quad y \text{ solves}$$
$$\min((y-1)^2 - 1.5xy) \quad \text{s.t.}$$
$$-3x + y \leq -3$$
$$x - 0.5y \leq 4$$
$$x + y \leq 7$$
$$x \geq 0, y \geq 0.$$

Hints: A known feasible solution is $x = 1, y = 0$.

3 Show that any linear bilevel program is a MPAEC.

4 Solve the MPAEC equivalent to the LBP in Exercise 1.

5 Consider the MPAEC

$$\text{minimize} \quad f(x,y) \tag{1}$$
$$\text{subject to} \quad (x,y) \in U,\ y \in C, \tag{2}$$
$$\langle Ax + By + c, z - y\rangle \leq 0 \ \ \forall z \in C \tag{3}$$

where $U \subset \mathbb{R}^{n+m}, C \subset \mathbb{R}^n$ are two nonempty closed convex sets, $f(x,y)$ is convex function on $\mathbb{R}^{n+m}$ while A, B are given appropriate real matrices and $c \in \mathbb{R}^n$.

Show that if A is a symmetric positive semidefinite matrix then this problem reduces to a convex bilevel program.

6 Let $g(x,y) = \max_{y \in C}\{\langle Ax + By + c, y - z\rangle - \frac{1}{2}\langle z - y, G(z-y)\rangle\}$, where G is an arbitrary $n \times n$ symmetric positive definite matrix. Show that:

(i) $g(x,y) \geq 0 \ \ \forall (x,y) \in C \times \mathbb{R}^m$,
(ii) $(x,y) \in U,\ y \in C,\ g(x,y) = 0$ if and only if (x,y) satisfies the conditions (2) and (3) in Exercise 5.

In other words, the MPAEC in Exercise 5 is equivalent to the dc optimization problem

$$\min f(x,y):$$
$$(x,y) \in U,\ y \in C,$$
$$g(x,y) \leq 0.$$

References

Ahuja, R.K., Magnanti, T.L., Orlin, J.B.: Networks Flows: Theory, Algorithms and Applications. Prentice Hall, Englewood Cliffs, NJ (1993)

Alexandrov, A.D.: On surfaces which may be represented by a difference of convex functions. Izvestiya Akademii Nauk Kazakhskoj SSR, Seria Fiziko-Matematicheskikh **3**, 3–20 (1949, Russian)

Alexandrov, A.D.: On surfaces which may be represented by differences of convex functions. Dokl. Akad. Nauk SSR **72**, 613–616 (1950, Russian)

Al-Khayyal, F.A., Falk, J.E.: Jointly constrained biconvex programming. Math. Oper. Res. **8**, 273–286 (1983)

Al-Khayyal, F.A., Larsen, C., Van Voorhis, T.: A relaxation method for nonconvex quadratically constrained quadratic programs. J. Glob. Optim., 215–230 (1995)

Al-Khayyal, F.A., Tuy, H., Zhou, F.: DC optimization methods for multisource location problems. School of Industrial and Systems Engineering. George Institute of Technology, Atlanta (1997). Preprint

Al-Khayyal, F.A., Tuy, H., Zhou, F.: Large-scale single facility continuous location by DC Optimization. Optimization **51**, 271–292 (2002)

An, L.T.H.: Solving large scale molecular distance geometry problems by a smoothing technique via the Gaussian Transform and D.C. Programming. J. Glob. Optim. **27**, 375–397 (2003)

An, L.T.H., Tao, P.D.: The DC (difference of convex) programming and DCA revisited with DC models of real world nonconvex optimization problems. Ann. Oper. Res. **133**, 23–46 (2005)

An, L.T.H., Tao, P.D., Muu, L.D.: Simplicially-constrained DC optimization over the efficient and weakly efficient set. J. Optim. Theory Appl. **117**, 503–531 (2003)

An, L.T.H., Belghiti, M.T., Tao, P.D.: A new efficient algorithm based on DC programming and DCA for clustering. J. Glob. Optim. **37**, 557–569 (2007)

An, L.T.H., Tao, P.D., Nam, N.C., Muu, L.D.: Methods for optimization over the efficient and weakly efficient sets of an affine fractional vector optimisation problem. Optimization **59**, 77–93 (2010)

Androulakis, I.P., Maranas, C.D., Floudas, C.A.: αBB: a global optimization method for general constrained nonconvex problems. J. Glob. Optim. **7**, 337–363 (1995)

Apkarian, P., Tuan, H.D.: Concave programming in control theory. J. Glob. Optim. **15**, 343–370 (1999)

Asplund, E.: Differentiability of the metric projection in finite dimensional Euclidean space. Proc. AMS **38**, 218–219 (1973)

Astorino, A., Miglionico, G.: Optimizing sensor cover energy via DC Programming. Optim. Lett. (2014). doi:10.1007/s11590-014-0778-y

H. Tuy, *Convex Analysis and Global Optimization*, Springer Optimization and Its Applications 110, DOI 10.1007/978-3-319-31484-6

Aubin, J.-P., Ekeland, I.: Applied Nonlinear Analysis. John Wiley & Sons, NewYork (1984)
Auchumuty, G.: Variational principles for variational inequalities. Numer. Funct. Anal. Optim. **10**, 863–874 (1989)
Auslender, A.: Optimisation, Méthodes Numériques. Masson, Paris (1976)
Bach-Kim, N.T.: An algorithm for optimization over the efficient set. Vietnam J. Math. **28**, 329–340 (2000)
Baer, E., et al.: Biological and synthetic hierarchical composites. Phys. Today **46**, 60–67 (1992)
Balas, E.: Intersection cuts—a new type of cutting planes for integer programming. Oper. Res. **19**, 19–39 (1971)
Balas, E.: Integer programming and convex analysis: intersection cuts and outer polars. Math. Program. **2**, 330–382 (1972)
Balas, E., Burdet, C.A.: Maximizing a convex quadratic function subject to linear constraints. Management Science Research Report No 299, Carnegie–Melon University (1973)
Bard, J.F.: Practical Bilevel Optimization Algorithms and Optimization. Kluwer, Dordrecht (1988)
Bard, J. F.: Convex two-level optimization. Math. Program. **40**, 15–27 (1998)
Ben-Ayed, O.: Bilevel linear programming. Comput. Oper. Res. **20**, 485–501 (1993)
Bengtsson, M., Ottersen, B.: Optimum and suboptimum transmit beamforming. In: Handbook of Antennas in Wirelss Coomunications, chap. 18. CRC, Boca Raton (2001)
Benson, H.P.: Optimization over the efficient set. J. Math. Anal. Appl. **98**, 562–580 (1984)
Benson, H.P.: An algorithm for optimizing over the weakly-efficient set. Eur. J. Oper. Res. **25**, 192–199 (1986)
Benson, H.P.: An all-linear programming relaxation algorithm for optimization over the efficient set. J. Glob. Optim. **1**, 83–104 (1991)
Benson, H.P.: A bisection-extreme point search algorithm for optimizing over the efficient set in the linear dependance case. J. Glob. Optim. **3**, 95–111 (1993)
Benson, H.P.: Concave minimization: theory, applications and algorithms. In: Horst, R., Pardalos, P. (eds.) Handbook of Global Optimization, pp. 43–148. Kluwer Academic Publishers, Dordrecht (1995)
Benson, H.P.: Deterministic algorithms for constrained concave minimization: a unified critical survey. Nav. Res. Logist. **43**, 765–795 (1996)
Benson, H.P.: An outer approximation algorithm for generating all efficient extreme points in the outcome set of a multi objective linear programming problem. J. Glob. Optim. **13**, 1–24 (1998)
Benson, H.P.: An outcome space branch and bound outer approximation algorithm for convex multiplicative programming. J. Glob. Optim. **15**, 315–342 (1999)
Benson, H.P.: A global optimization approach for generating efficient points for multiplicative concave fractional programming. J. Multicrit. Decis. Anal. **13**, 15–28 (2006)
Ben-Tal, A., Eiger, G., Gershovitz, V.: Global minimization by reducing the duality gap. Math. Program. **63**, 193–212 (1994)
Ben-Tal, A., Nemirovski, A.: Lectures on Modern Convex Optimization. MPS-SIAM Series on Optimization. SIAM, Philadelphia (2001)
Bentley, J.L., Shamos, M.I.: Divide and conquer for linear expected time. Inf. Proc. Lett. **7**, 87–91 (1978)
Bialas, W.F., Karwan, C.H.: On two-level optimization. IEEE Trans. Auto. Const. **AC-27**, 211–214 (1982)
Bjōrnson, E., Jorswieck, E.: Optimal resource allocation in coordinated multi-cell systems. Foundations and Trends in Communications and Information Theory, vol. 9(2–3), pp. 113–381 (2012). doi:10.1561/0100000069
Bjōrnson, E., Zheng, G., Bengtson, M., Ottersen, B.: Robust monotonic optimization framework for multicell MISO systems. IEEE Trans. Signal Process. **60**(5), 2508–2521 (2012)
Blum, E., Oettli, W.: From optimization and variational inequality to equilibrium problems. Math. Student **63**, 127–149 (1994)
Bolintineau, S.: Minimization of a quasi-concave function over an efficient set. Math. Program. **61**, 83–120 (1993a)

Bolintineau, S.: Optimality condition for minimization over the (weakly or properly) efficient set. J. Math. Anal. Optim. **173**, 523–541 (1993b)

Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge, United Kingdom (2004)

Bracken, J., McGill, J.: Mathematical programs with optimization problems in the constraints. Oper. Res. **21**, 37–44 (1973)

Bracken, J., McGill, J.: A method for solving mathematical programs with nonlinear programs in the constraints. Oper. Res. **22**, 1097–1101 (1974)

Bradley, P.S., Mangasarian, O.L., Street, W.N.: Clustering via concave minimization. Technical report 96–03. Computer Science Department, University of Wisconsin (1997)

Brimberg, J., Love, R.F.: A location problem with economies of scale. Stud. Locat. Anal. (7), 9–19 (1994)

Bui, A.: Etude analytique d'algorithmes distribués de routage. Thèse de doctorat, Université Paris VII (1993)

Bulatov, V.P. Embedding methods in optimization problems. Nauka, Novosibirsk (1977) (Russian)

Bulatov, V.P. Methods for solving multiextremal problems (Global Search). In Beltiukov, B.A., Bulatov, V.P. (eds) Methods of Numerical Analysis and Optimization. Nauka, Novosibirsk (1987) (Russian)

Candler, W., Townsley, R.: A linear two-level programming problem. Comput. Oper. Res. **9**, 59–76 (1982)

Caristi, J.: Fixed point theorems for mappings satisfying inwardness conditions. Trans. Am. Math. Soc. **215**, 241–251 (1976)

Charnes, A., Cooper, W.W.: Programming with linear fractional functionals. Nav. Res. Logist. **9**, 181–186 (1962)

Chen, P.C., Hansen, P., Jaumard, B.: On-line and off-line vertex enumeration by adjacency lists. Oper. Res. Lett. **10**, 403–409 (1991)

Chen, P.-C., Hansen, P., Jaumard, B., Tuy, H.: Weber's problem with attraction and repulsion. J. Regional Sci. **32**, 467–486 (1992)

Chen, P.-C., Hansen, P., Jaumard, B., Tuy, H.: Solution of the multifacility Weber and conditional Weber problems by D.C. Programming, Cahier du GERAD G-92-35. Ecole Polytechnique, Montréal (1992b)

Chen, P.C., Hansen, P., Jaumard, B., Tuy, H.: Solution of the multifacility Weber and conditional Weber problems by DC Programming. Oper. Res. **46**, 548–562 (1998)

Cheon, M.-S., Ahmed, S., Al-Khayyal, F.: A branch-reduce-cut algorithm for the global optimization of probabilistically constrained linear programs. Math. Program. Ser. B **108**, 617–634 (2006)

Cheney, E.W., Goldstein, A.A.: Newton's method for convex programming and Tchebycheff approximation. Numer. Math. **1**, 253–268 (1959)

Cooper, L.: Location-allocation problems. Oper. Res. **11**, 331–343 (1963)

Dantzig, G.B.: Linear Programming and Extensions. Princeton University Press, Princeton, NJ (1963)

Dines, L.L.: On the mapping of quadratic forms. Bull. Am. Math. Soc. **47**, 494–498 (1941)

Dixon, L.C.W., Szego, G.P. (eds.): Towards Global Optimization, vol. I. North-Holland, Amsterdam (1975)

Dixon, L.C.W., Szego, G.P. (eds.): Towards Global Optimization, vol. II. North-Holland, Amsterdam (1978)

Duffin, R.J., Peterson, E.L., Zener, C.: Geometric Programming—Theory and Application. Wiley, New York (1967)

Eaves, B.C., Zangwill, W.I.: Generalized cutting plane algorithms. SIAM J. Control **9**, 529–542 (1971)

Ecker, J.G., Song, H.G.: Optimizing a linear function over an efficient set. J. Optim. Theory Appl. **83**, 541–563 (1994)

Ekeland, I.: On the variational principle. J. Math. Anal. Appl. **47**, 324–353 (1974)

Ekeland, I., Temam, R.: Convex Analysis and Variational Problems. North-Holland, Amsterdam (1976)

Ellaia, R.: Contribution à l'analyse et l'optimisation de différences de fonctions convexes, Thèse de trosième cycle, Université de Toulouse (1984)

Falk, J.E.: Lagrange multipliers and nonconvex problems. SIAM J. Control **7**, 534–545 (1969)

Falk, J.E.: Conditions for global optimality in nonlinear programming, Oper. Res. **21**, 337–340 (1973a)

Falk, J.E.: A linear max-min problem. Math. Program. **5**, 169–188 (1973b)

Falk, J.E., Hoffman, K.L.: A successive underestimation method for concave minimization problems. Math. Oper. Res. **1**, 251–259 (1976)

Falk, J.E., Hoffman, K.L.: Concave minimization via collapsing polytopes. Oper. Res. **34**, 919–929 (1986)

Falk, J.E., Palocsay, S.W.: Optimizing the sum of linear fractional functions. In: Floudas, C., Pardalos, P. (eds) Global Optimization, pp. 221–258. Princeton University Press, Princeton (1992)

Falk, J.E., Soland, R.M.L: An algorithm for separable nonconvex programming problems. Manag. Sci. **15**, 550–569 (1969)

Fan, K.: A minimax inequality and applications. In: Shisha (ed.) Inequalities, vol. 3, pp. 103–113. Academic, New York (1972)

Finsler, P.: Uber das Vorkommen definiter und semidefiniter Formen in Scharen quadratischer Formen. Comment. Math. Helv. **9**, 188–192 (1937)

Floudas, C.A., Aggarwal, A.: A decomposition strategy for global optimum search in the pooling problem. ORSA J. Comput. **2**(3), 225–235 (1990)

Floudas, C.A. : Deterministic Global Optimization, Theory, Methods and Applications. Kluwer, Dordrecht, The Netherlands (2000)

Forgo, F.: The solution of a special quadratic problem (in Hungarian). Szigma 53–59 (1975)

Foulds, L.R., Haugland, D., Jörnsten, K.: A bilinear approach to the pooling problem. Optimization **24**, 165–180 (1992)

Fredman, M.L., Tarjan, R.E.: Fibonacci heaps and their uses in improved network optimization algorithms. J. Assoc. Comput. Mach. **34**, 596–615 (1984)

Friesz, T.L., et al.: A simulated annealing approach to the network design problem with variational inequality constraints. Transp. Sci. **28**, 18–26 (1992)

Fujiwara, O., Khang, D.B.: A two-phase decomposition method for optimal design of looped water distribution networks. Water Resour. Res. **23**, 977–982 (1990)

Fukushima, M.: Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems. Math. Program. **53**, 99–110 (1992)

Gabasov, R., Kirillova, F.M.: Linear Programming Methods. Part 3 (Special Problems). Minsk, Russian (1980)

Gale, P.: An algorithm for exhaustive generation of building floor plans. Commun. ACM **24**, 813–825 (1981)

Gantmacher F.R.: The Theory of Matrices, vol. 2. Chelsea Publishing Company, New York (1960)

Ge, R.P., Qin, Y.F.: A class of filled functions for finding global minimizers of a function of several variables. J. Optim. Theory Appl. **54**, 241–252 (1987)

Glover, F.: Cut search methods in integer programming. Math. Program. **3**, 86–100 (1972)

Glover, F.: Convexity cuts and cut search. Oper. Res. **21**, 123–124 (1973a)

Glover, F.: Concave programming applied to a special class of 0–1 integer programs. Oper. Res. **21**, 135–140 (1973b)

Graham, R.L.: An efficient algorithm for determining the convex hull of a finite planar set. Inf. Proc. Lett. **1**, 132–133 (1972)

Groch, A., Vidigal, L., Director, S.: A new global optimization method for electronic circuit design. IEEE Trans. Circuits Syst. **32**, 160–179 (1985)

Guisewite, G.M., Pardalos, P.M.: A polynomial time solvable concave network flow problem. Network **23**, 143–148 (1992)

Haims, M.J., Freeman, H.: A multistage solution of the template-layout problem. IEEE Trans. Syst. Sci. Cyber. **6**, 145–151 (1970)
Hamami, M., Jacobsen, S.E.: Exhaustive non-degenerate conical processes for concave minimization on convex polytopes. Math. Oper. Res. **13**, 479–487 (1988)
Hansen, P., Jaumard, B., Savard, G.: New branch and bound rules for linear bilinear programming. SIAM J. Sci. Stat. Comput. **13**, 1194–1217 (1992)
Hansen, P., Jaumard, B., Tuy, H.: Global optimization in location. In Dresner, Z. (ed.) Facility Location, pp. 43–68. Springer, NewYork (1995)
Hartman, P.: On functions representable as a difference of convex functions. Pac. J. Math. **9**, 707–713 (1959)
Hartman, P., Stampacchia, G.: On some nonlinear elliptic differential functional equations. Acta Math. **115**, 153–188 (1966)
Hillestad, R.J., Jacobsen, S.E.: Linear programs with an additional reverse convex constraint. Appl. Math. Optim. **6**, 257–269 (1980b)
Hiriart-Urruty, J.-B.: A short proof of the variational principle for approximate solutions of a minimization problem. Am. Math. Mon. **90**, 206–207 (1983)
Hiriart-Urruty, J.-B.: From convex optimization to nonconvex optimization, Part I: necessary and sufficient conditions for global optimality. In: Clarke, F.H., Demyanov, V.F., Giannessi, F. (eds.) Nonsmooth Optimization and Related Topics. Plenum, New York (1989a)
Hiriart-Urruty, J.-B. Conditions nécessaires et suffisantes d'optimalité globale en optimisation de différences de deux fonctions convexes. C.R. Acad. Sci. Paris **309**, Série I, 459–462 (1989b)
Hoai, P.T., Tuy, H.: Monotonic Optimization for Sensor Cover Energy Problem. Institute of Mathematics, Hanoi (2016). Preprint
Hoai-Phuong, N.T., Tuy, H.: A unified monotonic approach to generalized linear fractional programming. J. Glob. Optim. **26**, 229–259 (2003)
Hoang, H.G., Tuan, H.D., Apkarian, P.: A Lyapunov variable-free KYP Lemma for SISO continuous systems. IEEE Trans. Autom. Control **53**, 2669–2673 (2008)
Hoffman, K.L.: A method for globally minimizing concave functions over convex sets. Math. Program. **20**, 22–32 (1981)
Holmberg, K., Tuy, H.: A production-transportation problem with stochastic demands and concave production cost. Math. Program. **85**, 157–179 (1995)
Horst, R., Tuy, H.: Global Optimization (Deterministic Approaches), 3rd edn. Springer, Berlin/Heidelberg/New York (1996)
Horst, R., Thoai, N.V., Benson, H.P.: Concave minimization via conical partitions and polyhedral outer approximation. Math. Program. **50**, 259–274 (1991)
Horst, R., Thoai, N.V., de Vries, J.: On finding new vertices and redundant constraints in cutting plane algorithms for global optimization. Oper. Res. Lett. **7**, 85–90 (1988)
Idrissi, H., Loridan, P., Michelot, C.: Approximation for location problems. J. Optim. Theory Appl. **56**, 127–143 (1988)
Jaumard, B., Meyer, C.: On the convergence of cone splitting algorithms with ω-subdivisions. J. Optim. Theory Appl. **110**, 119–144 (2001)
Jeyakumar, V., Lee, G.M., Li, G.Y.: Alternative theorems for quadratic inequality systems and global quadratic optimization. SIAM J. Optim. **20**, 983–1001 (2009)
Jorswieck, E.A., Larson, E.G.: Monotonic optimization framework for the two-user MISO interference channel. IEEE Trans. Commun. **58**, 2159–2168 (2010)
Kakutani, S.: A generalization of Brouwer's fixed point theorem. Duke Math. J. **8**, 457–459 (1941)
Kalantari, B., Rosen, J.B.: An algorithm for global minimization of linearly constrained concave quadratic problems. Math. Oper. Res. **12**, 544–561 (1987)
Karmarkar, N.: An interior-point approach for NP-complete problems. Contemp. Math. **114**, 297–308 (1990)
Kelley, J.E.: The cutting plane method for solving convex programs. J. SIAM **8**, 703–712 (1960)
Klinz, B., Tuy, H.: Minimum concave-cost network flow problems with a single nonlinear arc cost. In: Pardalos, P., Du, D. (eds.) Network Optimization Problems, pp. 125–143. World Scientific, Singapore (1993)

Knaster, B., Kuratowski, C., Mazurkiewicz, S.: Ein Beweis des Fix-punktsatzes für n-dimensionale Simplexe. Fund. Math. **14**, 132–137 (1929)

Konno, H.: A cutting plane for solving bilinear programs. Math. Program. **11**,14–27 (1976a)

Konno, H.: Maximization of a convex function subject to linear constraints. Math. Program. **11**, 117–127 (1976b)

Konno, H., Abe, N.: Minimization of the sum of three linear fractions. J. Glob. Optim. **15**, 419–432 (1999)

Konno, H., Kuno, T.: Generalized linear and fractional programming. Ann. Oper. Res. **25**, 147–162 (1990)

Konno, H., Kuno, T.: Linear multiplicative programming. Math. Program. **56**, 51–64 (1992)

Konno, H., Kuno, T.: Multiplicative programming problems. In: Horst, R., Pardalos, P. (eds.) Handbook of Global Optimization, pp. 369–405. Kluwer, Dordrecht, The Netherlands (1995)

Konno, H., Yajima, Y.: Minimizing and maximizing the product of linear fractional functions. In: Floudas, C., Pardalos, P. (eds.) Recent Advances in Global Optimization, pp. 259–273. Princeton University Press, Princeton (1992)

Konno, H., Yamashita, Y.: Minimization od the sum and the product of several linear fractional functions. Nav. Res. Logist. **44**, 473–497 (1997)

Konno, H., Yajima, Y., Matsui, T.: Parametric simplex algorithms for solving a special class of nonconvex minimization problems. J. Glob. Optim. **1**, 65–82 (1991)

Konno, H., Kuno, T., Yajima, Y.: Global minimization of a generalized convex multiplicative function. J. Glob. Optim. **4**, 47–62 (1994)

Konno, H., Thach, P.T., Tuy, H.: Optimization on Low Rank Nonconvex Structures. Kluwer, Dordrecht, The Netherlands (1997)

Kough, P.L.: The indefinite quadratic programming problem. Oper. Res. **27**, 516–533 (1979)

Kuno, T.: Globally determining a minimum-area rectangle enclosing the projection of a higher dimensional set. Oper. Res. Lett. **13**, 295–305 (1993)

Kuno, T., Ishihama, T.: A convergent conical algorithm with ω-bisection for concave minimization. J. Glob. Optim. **61**, 203–220 (2015)

Landis, E.M.: On functions representable as the difference of two convex functions. Dokl. Akad. Nauk SSSR **80**, 9–11 (1951)

Lasserre, J.: Global optimization with polynomials and the problem of moments. SIAM J. Optim. **11**, 796–817 (2001)

Lasserre, J.: Semidefinite programming vs. LP relaxations for polynomial programming. Math. Oper. Res. **27**, 347–360 (2002)

Levy, A.L., Montalvo, A.: The tunneling algorithm for the global minimization of functions. SIAM J. Sci. Stat. Comput. **6**, 15–29 (1988)

Locatelli, M.: Finiteness of conical algorithm with ω-subdivisions. Math. Program. **85**, 593–616 (1999)

Luc, L.T.: Reverse polyblock approximation for optimization over the weakly efficient set and the efficient set. Acta Math. Vietnam. **26**, 65–80 (2001)

Luo, Z.Q., Pang, J.S., Ralph, D.: Mathematical Programs with Equilibrium Constraints. Cambridge University Press, Cambridge, United Kingdom (1996)

Maling, K., Nueller, S.H., Heller, W.R.: On finding optimal rectangle package plans. In: Proceedings 19th Design Automation Conference, pp. 663–670 (1982)

Mangasarian, O.L.: Nonlinear Programming. MacGraw-Hill, New York (1969)

Mangasarian, O.L.: Mathematical programming data mining. In: Data Mining and Knowledge Discovery, vol. 1, pp. 183–201 (1997)

Maranas, C.D., Floudas, C.A.: A global optimization method for Weber's problem with attraction and repulsion. In: Hager, W.W., Hearn, D.W., Pardalos, P.M. (eds.) Large Scale Optimization, pp. 259–293. Kluwer, Dordrecht (1994)

Martinez, J.M.: Local minimizers of quadratic functions on Euclidean balls and spheres. SIAM J. Optim. **4**, 159–176 (1994)

Mastroeni, G.: Gap functions for equilibrium problems. J. Glob. Optim. **27**, 411–426 (2003)

Matsui, T.: NP-hardness of linear multiplicative programming and related problems. J. Glob. Optim. **9**, 113–119 (1996)
Mayne, D.Q., Polak, E.: Outer approximation algorithms for non-differentiable optimization. J. Optim. Theory Appl. **42**, 19–30 (1984)
McCormick, G.P.: Attempts to calculate global solutions of problems that may have local minima. In: Lootsma, F. (ed.) Numerical Methods for Nonlinear Optimization, pp. 209–221. Academic, London/New York (1972)
McCormick, G.P.: Nonlinear Programming: Theory, Algorithms and Applications. Wiley, New York (1982)
Megiddo, N., Supowwit, K.J.: On the complexity of some common geometric location problems. SIAM J. Comput. **13**, 182–196 (1984)
Migdalas, A., Pardalos, P.M., Värbrand, P. (eds): Multilevel Optimization: Algorithms and Applications. Kluwer, Dordrecht (1998)
Murty, K.G.: Linear Programming. Wiley, New York (1983)
Murty, K.G., Kabadi, S.N.: Some NP-complete problems in quadratic and nonlinear programming. Math. Program. **39**, 117–130 (1987)
Muu, L.D.: A convergent algorithm for solving linear programs with an additional reverse convex constraint. Kybernetika **21**, 428–435 (1985)
Muu, L.D.: An algorithm for solving convex programs with an additional convex–concave constraint. Math. Program. **61**, 75–87 (1993)
Muu, L.D.: Models and methods for solving convex–concave programs. Habilitation thesis, Institute of Mathematics, Hanoi (1996)
Muu, L.D., Oettli, W.: Method for minimizing a convex concave function over a convex set. J. Optim. Theory Appl. **70**, 377–384 (1991)
Muu, L.D., Tuyen, H.Q.: Bilinear programming approach to optimization over the efficient set of a vector affine fractional problem. Acta Math. Vietnam. **27**, 119–140 (2002)
Muu, L.D., Quoc, T.D., An, L.T. and Tao, P.D.: A new decomposition algorithm for globally solving mathematical programs with equilibrium constraints. Acta Math. Vietnam. **37**, 201–217 (2012)
Nadler, S.B.: Multivalued contraction mappings. Pac. J. Math. **30**, 475–488 (1969)
Nash, J.: Equilibrium points in n-person game. Proc. Natl. Acad. Sci. U.S.A. **36**, 48–49 (1950)
Nghia, N.D., Hieu, N.D.: A Method for solving reverse convex programming problems. Acta Math. Vietnam. **11**, 241–252 (1986)
Ortega, J.M., Rheinboldt, W.C.: Iterative Solution of Nonlinear Equations in Several Variables. Academic, New York (1970)
Outrata, J.V., Kočvara, M., Zowe, J.: Nonsmooth Approach to Optimization Problems with Equilibrium Constraints. Kluwer, Cambridge (1998)
Pardalos, P.M., Vavasis, S.A.: Quadratic programming with one negative eigenvalue is NP-hard. J. Glob. Optim. **21**, 843–855 (1991)
Philip, J. : Algorithms for the vector maximization problem. Math. Program. **2**, 207–229 (1972)
Phillips, A.T., Rosen, J.B.: A parallel algorithm for constrained concave quadratic global minimization. Math. Program. **42**, 421–448 (1988)
Pincus, M.: A close form solution of certain programming problems. Oper. Res. **16**, 690–694 (1968)
Plastria, F.: Continuous location problems. In: Drezner, Z., Hamacher, H.W. (eds.) Facility Location, pp. 225–262. Springer, Berlin (1995)
Polik, I., Terlaky T.: A survey of the S-lemma. SIAM Rev. **49**, 371–418 (2007)
Prékopa, A.: Contributions to the theory of stochastic programming. Math. Program. **4**, 202–221 (1973)
Prékopa, A.: Stochastic Programming. Kluwer, Dordrecht (1995)
Putinar, M.: Positive polynomials on compact semi-algebraic sets. Indiana Univ. Math. J. **42**, 969–984 (1993)
Qian, L.P., Jun Zhang, Y.: S-MAPEL: monotonic optimization for non-convex joint power control and scheduling problems. IEEE Trans. Wirel. Commun. **9**, 1708–1719 (2010)

Qian, L.P., Jun Zhang, Y., Huang, J.W.: MAPEL: achieving global optimality for a nonconvex wireless power control problem. IEEE Trans. Wirel. Commun. **8**, 1553–1563 (2009)

Qian, L.P., Jun Zhang, Y., Huang, J.: Monotonic optimization in communications and networking systems. Found. Trends Netw. **7**, 1–75 (2012)

Roberts, A.W., Varberg, D.E.: Convex Functions. Academic, New York (1973)

Robinson, S.: Generalized equations and their solutions, part 1: basic theory. Math. Program. Stud. **10**,128–141 (1979)

Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)

Rosen, J.B.: Global minimization of a linearly constrained concave function by partition of feasible domain. Math. Oper. Res. **8**, 215–230 (1983)

Rosen, J.B., Pardalos, P.M.: Global minimization of large-scale constrained concave quadratic problems by separable programming. Math. Program. **34**, 163–174 (1986)

Rubinov, A., Tuy, H., Mays, H.: Algorithm for a monotonic global optimization problem. Optimization **49**, 205–221 (2001)

Saff, E.B., Kuijlaars, A.B.J.: Distributing many points on a sphere. Math. Intell. **10**, 5–11 (1997)

Sahni, S.: Computationally related problems. SIAM J. Comput. **3**, 262–279 (1974)

Schaible, S.: Fractional Programming. In: Horst ,R., Pardalos, P. (eds.) Handbook of Global Optimization, pp. 495–608. Kluwer, Dordrecht (1992)

Sherali, H.D., Adams, W.P.: A Reformulation-Lineralization Technique (RLT) for Solving Discrete and Continuous Nonconvex Programming Problems. Kluwer, Dordrecht (1999)

Sherali, H.D., Tuncbilek, C.H.: A global optimization algorithm for polynomial programming problems using a reformulation-linearization technique. J. Glob. Optim. **2**, 101–112 (1992)

Sherali, H.D., Tuncbilek, C.H.: A reformulation-convexification approach to solving nonconvex quadratic programming problems. J. Glob. Optim. **7**, 1–31 (1995)

Shi, J.S., Yamamoto, Y.: A D.C. approach to the largest empty sphere problem in higher dimension. In: Floudas, C., Pardalos, P. (eds.) State of the Art in Global Optimization, pp. 395–411. Kluwer, Dordrecht (1996)

Shiau, T.-H.: Finding the largest l^p-ball in a polyhedral set. Technical Summary Report No. 2788. University of Wisconsin, Mathematics Research Center, Madison (1984)

Shimizu, K., Ishizuka, Y., Bard, J.F.: Nondifferentiable and Two-Level Mathematical Programming. Kluwer, Cambridge (1997)

Shor, N.Z.: Quadratic Optimization Problems. Technicheskaya Kibernetica, **1**, 128–139 (1987, Russian)

Shor, N.Z.: Dual estimates in multiextremal problems. J. Glob. Optim. **2**, 411–418 (1991)

Shor, N.Z., Stetsenko, S.I.: Quadratic Extremal Problems and Nondifferentiable Optimization. Naukova Dumka, Kiev (1989, Russian)

Sion, M.: On general minimax theorems. Pac. J. Math. **8**, 171–176 (1958)

Sniedovich, M.: C-programming and the minimization of pseudolinear and additive concave functions. Oper. Res. Lett. **5**, 185–189 (1986)

Soland, R.M.: Optimal facility location with concave costs. Oper. Res. **22**, 373–382 (1974)

Sperner, E.: Neuer Beweis für die Invarianz der Dimensionszahl und des Gebietes. Abh. Math. Seminar Hamburg **VI** , 265–272 (1928)

Stancu-Minasian, I.M.: Fractional Programming, Theory, Methods and Applications. Kluwer, Dordrecht (1997)

Strekalovski, A.C.: On the global extremum problem. Sov. Dokl. **292**, 1062–1066 (1987, Russian)

Strekalovski, A.C.: On problems of global extremum in nonconvex extremal problems. Izvestya Vuzov, ser. Matematika **8**, 74–80 (1990, Russian)

Strekalovski, A.C.: On search for global maximum of convex functions on a constraint set. J. Comput. Math. Math. Phys. **33**, 349–363 (1993, Russian)

Suzuki, A., Okabe, A.: Using Voronoi diagrams. In: Drezner, Z. (ed.) Facility Location, pp. 103–118. Springer, New York (1995)

Swarup, K.: Programming with indefinite quadratic function with linear constraints. Cahiers du Centre d'Etude de Recherche Operationnelle **8**, 132–136 (1966a)

Swarup, K.: Indefinite quadratic programming. Cahiers du Centre d'Etude de Recherche Operationnelle **8**, 217–222 (1966b)

Swarup, K.: Quadratic programming. Cahiers du Centre d'Etude de Recherche Opérationnelle **8**, 223–234 (1966c)

Tanaka, T., Thach, P.T., Suzuki, S.: An optimal ordering policy for jointly replenished products by nonconvex minimization techniques. J. Oper. Res. Soc. Jpn. **34**, 109–124 (1991)

Tao, P.D.: Convex analysis approach to dc programming: theory, algorithm and applications. Acta Math. Vietnam. **22**, 289–355 (1997)

Tao, P.D.: Large-scale molecular optimization from distance matrices by a dc optimization approach. SIAM J. Optim. **14**, 77–116 (2003)

Tao, P.D., Hoai-An, L.T.: DC optimization algorithms for solving the trust region subproblem. SIAM J. Optim. **8**, 476–505 (1998)

Thach, P.T.: The design centering problem as a d.c. programming problem. Math. Program. **41**, 229–248 (1988)

Thach, P.T.: Quasiconjugates of functions, duality relationship between quasiconvex minimization under a reverse convex constraint and quasiconvex maximization under a convex constraint, and applications. J. Math. Anal. Appl. **159**, 299–322 (1991)

Thach, P.T.: New partitioning method for a class of nonconvex optimization problems. Math. Oper. Res. **17**, 43–69 (1992)

Thach, P.T.: D.c. sets, d.c. functions and nonlinear equations. Math. Program. **58**, 415–428 (1993a)

Thach, P.T.: Global optimality criteria and duality with zero gap in nonconvex optimization problems. SIAM J. Math. Anal. **24**, 1537–1556 (1993b)

Thach, P.T.: A nonconvex duality with zero gap and applications. SIAM J. Optim. **4**, 44–64 (1994)

Thach, P.T.: Symmetric duality for homogeneous multiple-objective functions. J. Optim Theory Appl. (2012). doi: 10.1007/s10957-011-9822-6

Thach, P.T., Konno, H.: On the degree and separability of nonconvexity and applications to optimization problems. Math. Program. **77**, 23–47 (1996)

Thach, P.T., Thang, T.V.: Conjugate duality for vector-maximization problems. J. Math. Anal. Appl. **376**, 94–102 (2011)

Thach, P.T., Thang, T.V.: Problems with resource allocation constraints and optimization over the efficient set. J. Glob. Optim. **58**, 481–495 (2014)

Thach, P.T., Tuy, H.: The relief indicator method for constrained global optimization. Naval Res. Logist. **37**, 473–497 (1990)

Thach, P.T., Konno, H., Yokota, D.: A dual approach to a minimization on the set of Pareto-optimal solutions. J. Optim. Theory Appl. **88**, 689–701 (1996)

Thakur, L.S.: Domain contraction in nonlinear programming: minimizing a quadraticconcave function over a polyhedron. Math. Oper. Res. **16**, 390–407 (1991)

Thieu, T.V.: Improvement and implementation of some algorithms for nonconvex optimization problems. In: Optimization—Fifth French German Conference, Castel Novel 1988. Lecture Notes in Mathematics, vol. 1405, pp. 159–170. Springer, Berlin (1989)

Thieu, T.V.: A linear programming approach to solving a jointly contrained bilinear programming problem with special structure. In: Van tru va He thong, vol. 44, pp. 113–121. Institute of Mathematics, Hanoi (1992)

Thieu, T.V., Tam, B.T., Ban, V.T.: An outer approximation method for globally minimizing a concave function over a compact convex set. Acta Math. Vietnam **8**, 21–40 (1983)

Thoai, N.V.: A modified version of Tuy's method for solving d.c. programming problems. Optimization **19**, 665–674 (1988)

Thoai, N.V.: Convergence of duality bound method in partly convex programming. J. Glob. Optim. **22**, 263–270 (2002)***

Thoai, N.V., Tuy, H.: Convergent algorithms for minimizing a concave function. Math. Oper. Res. **5**, 556–566 (1980)

Toland, J.F.: Duality in nonconvex optimization. J. Math. Anal. Appl. **66**, 399–415 (1978)

Toussaint, G.T.: Solving geometric problems with the rotating calipers. Proc. IEEE MELECON **83** (1978)

Tuan, H.D., Apkarian, P., Hosoe, S., Tuy, H.: D.C. optimization approach to robust control: feasibility problems. Int. J. Control **73**(2), 89–104 (2000a)

Tuan, H.D., Hosoe, S., Tuy, H.: D.C. optimization approach to robust controls: the optimal scaling value problems. IEEE Trans. Autom. Control **45**(10), 1903–1909 (2000b)

Tuy, H.: Concave programming under linear constraints. Sov. Math. Dokl. **5**, 1437–1440 (1964)

Tuy, H.: On a general minimax theorem. Sov. Math. Dokl. **15**, 1689–1693 (1974)

Tuy, H.: A note on the equivalence between Walras' excess demand theorem and Brouwer's fixed point theorem'. In: Los, J., Los, M. (eds.) *Equilibria: How and Why ?*, pp. 61–64. North-Holland, Amsterdam (1976)

Tuy, H.: On outer approximation methods for solving concave minimization problems. Acta Math. Vietnam. **8**, 3–34 (1983)

Tuy, H.: Global minimization of a difference of two convex functions. In: Lecture Notes in Economics and Mathematical Systems, vol. 226, pp. 98–108. Springer, Berlin (1984)

Tuy, H.: Convex Programs with an additional reverse convex constraint. J. Optim. Theory Appl. **52**, 463–486 (1987a)

Tuy, H.: Global minimization of a difference of two convex functions. Math. Program. Study **30**, 150–182 (1987b)

Tuy, H.: Convex Analysis and Global Optimization. Kluwer, Dordrecht (1998)

Tuy, H.: On polyhedral annexation method for concave minimization. In: Leifman, L.J., Rosen, J.B. (eds.) Functional Analysis, Optimization and Mathematical Economics, pp. 248–260. Oxford University Press, Oxford (1990)

Tuy, H.: Normal conical algorithm for concave minimization over polytopes. Math. Program. **51**, 229–245 (1991)

Tuy, H.: On nonconvex optimization problems with separated nonconvex variables. J. Glob. Optim. **2**, 133–144 (1992)

Tuy, H.: D.C. optimization: theory, methods and algorithms. In: Horst, R., Pardalos, P. (eds.) In: Handbook of Global Optimization, pp. 149–216. Kluwer, Dordrecht (1995a)

Tuy, H.: Canonical d.c. programming: outer approximation methods revisited. Oper. Res. Lett. **18**, 99–106 (1995b)

Tuy, H.: A general d.c. approach to location problems. In: Floudas, C., Pardalos, P. (eds.) State of the Art in Global Optimization: Computational Methods and Applications, pp. 413–432. Kluwer, Dordrecht (1996)

Tuy, H.: Monotonic Optimization: Problems and Solution Approaches. SIAM J. Optim. **11**, 464–494 (2000a)

Tuy, H.: Strong polynomial time solvability of a minimum concave cost network flow problem. Acta Math. Vietnam. **25**, 209–217 (2000b)

Tuy, H.: On global optimality conditions and cutting plane algorithms. J. Optim. Theory Appl. **118**, 201–216 (2003)

Tuy, H.: On a decomposition method for nonconvex global optimization. Optim. Lett. **1**, 245–258 (2007a)

Tuy, H.: On dual bound methods for nonconvex global optimization. J. Glob. Optim. **37**, 321–323 (2007b)

Tuy, H.: D(C)-optimization and robust global optimization. J. Glob. Optim. **47**, 485–501 (2010)

Tuy, H., Ghannadan, S.: A new branch and bound method for bilevel linear programs. In: Migdalas, A., Pardalos, P.M., Värbrand, P. (eds.) Multilevel Optimization: Algorithms and Applications, pp. 231–249. Kluwer, Dordrecht (1998)

Tuy, H., Hoai-Phuong, N.T.: Optimization under composite monotonic constraints and constrained optimization over the efficient set. In: Liberti, L., Maculan, N. (eds.) Global Optimization: From Theory to Implementation, pp. 3–32. Springer, Berlin/New York (2006)

Tuy, H., Hoai-Phuong, N.T.: A robust algorithm for quadratic optimization under quadratic constraints. J. Glob. Optim. **37**, 557–569 (2007)

Tuy, H., Horst, R.: Convergence and restart in branch and bound algorithms for global optimization. Application to concave minimization and d.c. optimization problems. Math. Program. **42**, 161–184 (1988)

Tuy, H., Luc, L.T.: A new approach to optimization under monotonic constraints. J. Glob. Optim. **18**, 1–15 (2000)

Tuy, H., Oettli, W.: On necessary and sufficient conditions for global optimization. Matemáticas Aplicadas **15**, 39–41 (1994)

Tuy, H., Tam, B.T.: An efficient solution method for rank two quasiconcave minimization problems. Optimization **24**, 3–56 (1992)

Tuy, H., Tam, B.T.: Polyhedral annexation vs outer approximation methods for decomposition of monotonic quasiconcave minimization. Acta Math. Vietnam. **20**, 99–114 (1995)

Tuy, H., Tuan, H.D.: Generalized S-Lemma and strong duality in nonconvex quadratic programming. J. Glob. Optim. **56**, 1045–1072 (2013)

Tuy, H., Migdalas, A., Väbrand, P.: A global optimization approach for the linear two-level program. J. Glob. Optim. **3**, 1–23 (1992)

Tuy, H., Dan, N.D., Ghannadan, S.: Strongly polynomial time algorithm for certain concave minimization problems on networks. Oper. Res. Lett. **14**, 99–109 (1993a)

Tuy, H., Ghannadan, S., Migdalas, A., Värbrand, P.: Strongly polynomial algorithm for a production-transportation problem with concave production cost. Optimization **27**, 205–227 (1993b)

Tuy, H., Ghannadan, S., Migdalas, A., Värbrand, P.: Strongly polynomial algorithm for two special minimum concave cost network flow problems. Optimization **32**, 23–44 (1994a)

Tuy, H., Migdalas, A., Väbrand, P.: A quasiconcave minimization method for solving linear two-level programs. J. Glob. Optim. **4**, 243–263 (1994b)

Tuy, H., Al-Khayyal, F., Zhou, F.: A d.c. optimization method for single facility location problems. J. Glob. Optim. **7**, 209–227 (1995a)

Tuy, H., Ghannadan, S., Migdalas, A., Värbrand, P.: The minimum concave cost flow problem with fixed numbers of nonlinear arc costs and sources. J. Glob. Optim. **6**, 135–151 (1995b)

Tuy, H., Al-Khayyal, F., Zhou, P.C.: DC optimization method for single facility location problems. J. Glob. Optim. **7**, 209–227 (1995c)

Tuy, H., Ghannadan, S., Migdalas, A., Värbrand, P.: Strongly polynomial algorithm for a concave production-transportation problem with a fixed number of nonlinear variables. Math. Program. **72**, 229–258 (1996a)

Tuy, H., Thach, P.T., Konno, H., Yokota, D.: Dual approach to optimization on the set of pareto-optimal solutions. J. Optim. Theory Appl. **88**, 689–707 (1996b)

Tuy, H., Bagirov, Rubinov, A.M.: Clustering via D.C. optimization. In: Hadjisavvas, N., Pardalos, P.M. (eds.) Advances in Convex Analysis and Optimization, pp. 221–235. Kluwer, Dordrecht (2001)

Tuy, H., Nghia, N.D., Vinh, L.S.: A discrete location problem. Acta Math. Vietnam. **28**, 185–199 (2003)

Tuy, H., Al-Khayyal, F., Thach, P.T.: Monotonic optimization: branch and cut methods. In: Audet, C., Hansen, P., Savard, G. (eds.) Essays and Survey on Global Optimization, pp. 39–78. GERAD/Springer, New York (2005)

Tuy, H., Minoux, M., Hoai-Puong, N.T.: Discrete monotonic optimization with application to a discrete location problem. SIAM J. Optim. **17**, 78–97 (2006)

Utschick, W., Brehmer, J.: Monotonic optimization framework for coordinated beamforming in multi-cell networks. IEEE Trans. Signal Process. **60**, 2496–2507 (2012)

Vandenberge, L., Boyd, S.: Semidefinite programming. SIAM Rev. **38**, 49–95 (1996)

Van Voorhis, T.: The quadratically constrained quadratic programs. Ph.D. thesis, Georgia Institute of Technology, Atlanta (1997)

Veinott, A.F.: The supporting hyperplane method for unimodal programming. Oper. Res. **15**, 147–152 (1967)

Vidigal, L., Director, S.: A design centering algorithm for nonconvex regions of acceptability. IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. **14**, 13–24 (1982)

Vincent, J.F.: Structural Biomaterials. Princeton University Press, Princeton (1990)

Vicentee, L., Savard, G., Judice, J.: Descent approaches for quadratic bilevel programming. J. Optim. Theory Appl. **81**, 379–388 (1994) von Neumann, J.: Zur Theorie der Gesellschaftsspiele. Math. Ann. **100**, 293–320 (1928)

White, D.J., Annandalingam, G.: A penalty function approach for solving bilevel linear programs. J. Glob. Optim. **3**, 397–419 (1993)

Wolberg, W.H., Street, W.N., Mangasarian, O.L.: Machine learning techniques to diagnose breast cancer from fine-needle aspirates. Cancer Lett. **77**, 163–171 (1994)

Wolberg, W.H., Street, W.N., Mangasarian, O.L.: Image analysis and machine learning applied to breast cancer diagnosis and prognosis. Anal. Quant. Cytol. Histol. **17**(2), 77–87 (1995)

Wolberg, W.H., Street, W.N., Heisey, D.M., Mangasarian, O.L.: Computerized breast cancer diagnosis and prognosis from fine-needle aspirates. Arch. Surg. **130**, 511–516 (1995a)

Wolberg, W.H., Street, W.N., Heisey, D.M., Mangasarian, O.L.: Computer-derived nuclear features distinguish malignant from benign beast cytology. Hum. Pathol. **26**, 792–796 (1995b)

Yakubovich, V.A.: The S-procedure in nonlinear control theory. Vestnik Leningrad Univ. **4**, 73–93 (1977)

Ye, Y.: On affine scaling algorithms for nonconvex quadratic programming. Math. Program. **56**, 285–300 (1992)

Yick, J., Mukherjee, B., Ghosal, D.: Wireless sensor network survey. Comput. Network **52**, 2292–2320 (2008)

Zadeh, N.: On building minimum cost communication networks over time. Networks **4**, 19–34 (1974)

Zhang, Y.J., Qian, L.P., Huang, J.W.: Monotonic optimization communication and networking systems. Found. Trends Commun. Inf. Theory **7**(1), 1–75 (2012)

Zheng, Q.: Optimality conditions for global optimization (I) and (II). Acta Math. Appl. Sinica, English Ser. **2**, 66–78, 118–134 (1985)

Zhu, D.L., Marcotte, P.: An extended descent framework for variational inequalities. J. Optim. Theory Appl. **80**, 349–366 (1994)

Zukhovitzki, S.I., Polyak, R.A., Primak, M.E.: On a class of concave programming problems. Ekonomika i Matematitcheskie Methody **4**, 431–443 (1968, Russian)

Zwart, P.B.: Global maximization of a convex function with linear inequality constraints. Oper. Res. **22**, 602–609 (1974)

Index

H. Tuy, *Convex Analysis and Global Optimization*, Springer Optimization and Its Applications 110, DOI 10.1007/978-3-319-31484-6