MATHEMATICS
IN SCIENCE
AND
ENGINEERING
Volume 176

Theory of Multiobjective Optimization

Yoshikazu Sawaragi
Hirotaka Nakayama
Tetsuzo Tanino

THEORY OF MULTIOBJECTIVE OPTIMIZATION

This is Volume 176 in
MATHEMATICS IN SCIENCE AND ENGINEERING
A Series of Monographs and Textbooks
Edited by RICHARD BELLMAN, *University of Southern California*

The complete listing of books in this series is available from the Publisher upon request.

THEORY OF MULTIOBJECTIVE OPTIMIZATION

YOSHIKAZU SAWARAGI

Faculty of Science
The Kyoto Sangyo University
Kyoto, Japan

HIROTAKA NAKAYAMA

Department of Applied Mathematics
Konan University
Kobe, Japan

TETSUZO TANINO

Department of Mechanical Engineering II
Tohoku University
Sendai, Japan

1985

Academic Press, Inc.
(Harcourt Brace Jovanovich, Publishers)
Orlando San Diego New York London
Toronto Montreal Sydney Tokyo

ACADEMIC PRESS, INC.
Orlando, Florida 32887

United Kingdom Edition published by
ACADEMIC PRESS INC. (LONDON) LTD.
24–28 Oval Road, London NW1 7DX

Library of Congress Cataloging in Publication Data

Sawaragi, Yoshikazu, Date
Theory of multiobjective optimization.

Includes index.
1. Mathematical optimization. I. Nakayama,
Hirotaka, Date. II. Tanino, Tetsuzo.
III. Title.
QA402.5.S28 1985 519 84-14501
ISBN 0-12-620370-9 (alk. paper)

PRINTED IN THE UNITED STATES OF AMERICA

85 86 87 88 9 8 7 6 5 4 3 2 1

To our wives
Atsumi, Teruyo, and Yuko

This page intentionally left blank

CONTENTS

PREFACE

This book presents in a comprehensive manner some salient theoretical aspects of multiobjective optimization. The authors had been involved in the special research project Environmental Science, sponsored by the Education Ministry of Japan, for more than a decade since 1970. Through the research activities, we became aware that an important thing is not merely to eliminate pollutants after they are discharged, but how to create a good environment from a holistic viewpoint. What, then, is a good environment? There are many factors: physical, chemical, biological, economic, social, and so on. In addition, to make the matter more difficult, there appear to be many conflicting values.

System scientific methodology seems effective for treating such a multiplicity of values. Its main concern is how to trade off these values. One of the major approaches is multiobjective optimization. Another is multiattribute utility analysis. The importance of these research themes has been widely recognized in theory and practice. Above all, the workshops at South Carolina in 1972 and at IIASA in 1975 have provided remarkable incentives to this field of research. Since then, much active research has been observed all over the world.

Although a number of books in this field have been published in recent years, they focus primarily on methodology. In spite of their importance, however, theoretical aspects of multiobjective optimization have never been dealt with in a unified way.

In Chapter 1 (Introduction), fundamental notions in multiobjective decision making and its historical background are briefly explained. Throughout this chapter, readers can grasp the purpose and scope of this volume.

Chapters 2–6 are the core of the book and are concerned with the mathematical theories in multiobjective optimization of existence, necessary and sufficient conditions of efficient solutions, characterization of efficient solutions, stability, and duality. Some of them are still developing, but we have tried to describe them in a unified way as much as possible.

Chapter 7 treats methodology including utility/value theory, stochastic dominance, and multiobjective programming methods. We emphasized critical points of these methods rather than a mere introduction. We hope that this approach will have a positive impact on future development of these areas.

The intended readers of this book are senior undergraduate students, graduate students, and specialists of decision making theory and mathematical programming, whose research fields are applied mathematics, electrical engineering, mechanical engineering, control engineering, economics, management sciences, operations research, and systems science. The book is self-contained so that it might be available either for reference and self-study or for use as a classroom text; only an elementary knowledge of linear algebra and mathematical programming is assumed.

Finally, we would like to note that we were motivated to write this book by a recommendation of the late Richard Bellman.

NOTATION AND SYMBOLS

$x := y$	x is defined as y
$x \in S$	x is a member of the set S
$x \notin S$	x is not a member of the set S
S^c	complement of the set S
cl S	closure of the set S
int S	interior of the set S
∂S	boundary of the set S
$S \subset T, T \supset S$	S is a subset of the set T
$S \cup T$	union of two sets S and T
$S \cap T$	intersection of two sets S and T
$S \backslash T$	difference between S and T, i.e., $S \cap T^c$
$S + T$	sum of two sets S and T, i.e., $S + T := \{s + t : s \in S$ and $t \in T\}$
$S \times T$	Cartesian product of the sets S and T, i.e., $S \times T := \{(s, t) : s \in S$ and $t \in T\}$
R^n	n-dimensional Euclidean space. A vector x in R^n is written $x = (x_1, x_2, \ldots, x_n)$, but when used in matrix calculations it is represented as a column vector, i.e., $$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$ The corresponding row vector is $x^T = (x_1, x_2, \ldots, x_n)$. When $A = (a_{ij})$ is an $m \times n$ matrix with entry a_{ij} in its ith row and jth column, the

product Ax is the vector $y \in R^m$ with components $y_i = \sum_{j=1}^{m} a_{ij}x_j$, $i = 1, \ldots, m$.

R^p_+ — nonnegative orthant of R^p, i.e.,

$$R^p_+ := \{y \in R^p : y_i \geqq 0 \text{ for } i = 1, \ldots, p\}$$

$\mathring{R}^p_+$ — positive orthant of R^p, i.e., $\mathring{R}^p_+ := \{y \in R^p : y_i > 0$ for $i = 1, \ldots, p\}$

$$S^p := \{y \in R^p_+ : \textstyle\sum_{i=1}^{p} y_i = 1\}$$
$$\mathring{S}^p := \{y \in \mathring{R}^p_+ : \textstyle\sum_{i=1}^{p} y_i = 1\}$$

For x, $y \in R^p$ and a pointed cone K of R^p with int $K \neq \varnothing$,

$x \geqq_K y$	is defined as	$x - y \in K$,
$x \geq_K y$	is defined as	$x - y \in K \backslash \{0\}$,
$x >_K y$	is defined as	$x - y \in \text{int } K$.

In particular, in case of $K = R^p_+$, the subscript K is omitted, namely,

$x \geqq y : x_i \geqq y_i$	for all	$i = 1, \ldots, p$;
$x \geq y : x \geqq y$	and	$x \neq y$;
$x > y : x_i > y_i$	for all	$i = 1, \ldots, p$.

$x \succ y$ — x is preferred to y

$x \sim y$ — x is indifferent to y

$\langle x, y \rangle$ — inner product of the vectors x and y in the Euclidean space

$\|x\|$ — Euclidean norm of the vector x in the Euclidean space

co S — convex hull of the set S

$T(S, y)$ — tangent cone of the set S at y

$P(S)$ — projecting cone of S

$0^+ Y$ — recession cone of the set Y

Y^+ — extended recession cone of the set Y

$\mathscr{E}(Y, D)$ — efficient set of the set Y with respect to the domination structure (cone) D

$\mathscr{P}(Y, D)$ — properly efficient set (in the sense of Benson) of the set Y with respect to the domination cone D

D-epi W — D-epigraph of the (point-to-set) map W

$\delta(\cdot\,\vert\, S)$	indicator function of the set S
$f'(x; d)$	one-sided directional derivative of the function f at x
$\nabla f(x)$	gradient of the function f at x, i.e., if $f: R^n \to R$, $$\nabla f(x) := \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \dots, \frac{\partial f(x)}{\partial x_n}\right)$$
f^*	conjugate map (or function) of the function f
F^*	conjugate map of the point-to-set map F
$F^{**}(f^{**})$	biconjugate map of $F(f)$
$\partial f(x)$	subdifferential of the (vector-valued) function f at x
$\partial F(x; y)$	subdifferential of the point-to-set map F at $(x; y)$
$\mathrm{Max}_D\, Y\ (\mathrm{Min}_D\, Y)$	set of the D-maximal (D-minimal) points of the set Y, i.e., $\mathrm{Max}_D\, Y := \mathscr{E}(Y, -D)$ ($\mathrm{Min}_D\, Y := \mathscr{E}(Y, D)$). In case of $D = R^p_+$, in particular, $\mathrm{Max}\, Y := \mathscr{E}(Y, -R^p_+)$ ($\mathrm{Min}\, Y := \mathscr{E}(Y, R^p_+)$)
w-$\mathrm{Max}_D\, Y$ (w-$\mathrm{Min}_D\, Y$)	set of the weak D-maximal (D-minimal) points of the set Y. In case of $D = R^p_+$, in particular, the subscript D is omitted.
w-Sup Y(w-Inf Y)	set of the weak supremal (infimum) points of the set Y
w-F^*	weak conjugate map of the point-to-set map F
w-F^{**}	weak biconjugate map of the point-to-set map F
w-$\partial F(x)$	weak subdifferential of the point-to-set map F at x
w-L	weak Lagrangian
max Y (**min** Y)	strong maximum (minimum) of the set Y
sup Y (**inf** Y)	strong supremum (infimum) of the set Y
$\mathbf{f}^*$	strong conjugate of the vector-valued function f
$\partial \mathbf{f}(x)$	strong subdifferential of the vector-valued function f at x

This page intentionally left blank

1 INTRODUCTION

Every day we encounter various kinds of decision making problems as managers, designers, administrative officers, mere individuals, and so on. In these problems, the final decision is usually made through several steps, even though they sometimes might not be perceived explicitly. Figure 1.1 shows a conceptual model of the decision making process. It implies that the final decision is made through three major models, the structure model, the impact model, and the evaluation model.

By structure modeling, we mean constructing a model in order to know the structure of the problem, what the problem is, which factors comprise the problem, how they interrelate, and so on. Through the process, the *objective* of the problem and *alternatives* to perform it are specified. Hereafter, we shall use the notation $\mathcal{O}$ for the objective and X for the set of alternatives, which is supposed to be a subset of an n-dimensional vector space. If we positively know a consequence caused by an alternative, the decision making is said to be *under certainty*; whereas if we cannot know a sure result because of some uncertain factor(s), the decision making is said to be *under uncertainty*. Furthermore, if we objectively or subjectively know the probability distribution of the possible consequences caused by an alternative, the decision making is said to be *under risk*. Even though the final objective might be a single entity, we encounter, in general, many subobjectives $\mathcal{O}_i$ on the way to the final objective. In this book, we shall consider decision making problems with multiple objectives. Interpretive structural modeling (ISM) (Warfield [W5]) can be applied effectively in order to obtain a hierarchical structure of the objectives.

In order to solve our decision making problem by some systems–analytical methods, we usually require that degrees of objectives be represented in numerical terms, which may be of multiple kinds even for one objective. In

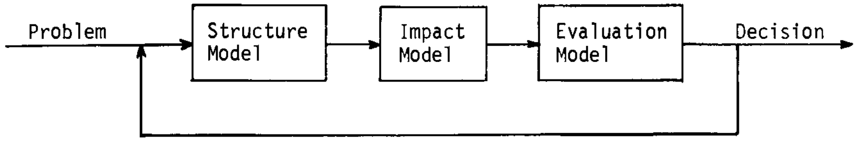

Fig. 1.1. Conceptual model of the decision making process.

order to exclude subjective value judgment at this stage, we restrict these numerical terms to physical measures (for example, money, weight, length, and time). As such a performance index, or criterion, for the objective $\mathcal{O}_i$, an *objective function* $f_i: X \to R^1$ is introduced, where R^1 denotes one-dimensional Euclidean space. The value $f_i(x)$ indicates how much impact is given on the objective $\mathcal{O}_i$ by performing an alternative x. Impact modeling is performed to identify these objective functions from various viewpoints such as physical, chemical, biological, social, economic, and so on. For convenience of mathematical treatment, we assume in this book that a smaller value for each objective function is preferred to a larger one. Now we can formulate our decision making problems as a *multiobjective optimization problem*:

$$\text{(P)} \qquad \text{Minimize} \qquad f(x) = (f_1(x), f_2(x), \ldots, f_p(x)) \qquad \text{over} \quad x \in X.$$

This kind of problem is also called a *vector optimization*. In some cases, some of the objective functions are required to be maintained under given levels prior to minimizing other objective functions. Denoting these objective functions by $g_j(x)$, we require that

$$g_j(x) \leq 0, \qquad j = 1, \ldots, m,$$

which, for convenience, is also supposed to represent some other technical constraints. Such a function $g_j(x)$ is generally called a *constraint function* in this book. According to the situation, we will consider either the problem (P) itself or (P) accompanied by the constraint $g_j(x) \leq 0$ $(j = 1, \ldots, m)$. Of course, an equality constraint $h_k(x) = 0$ can be embedded within two inequalities $h_k(x) \leq 0$ and $-h_k(x) \leq 0$, and, hence, it does not appear explicitly in this book.

Unlike traditional mathematical programming with a single objective function, an optimal solution in the sense of one that minimizes all the objective functions simultaneously does not necessarily exist in multiobjective optimization problems, and, hence, we are troubled with conflicts among objectives in decision making problems with multiple objectives. The final decision should be made by taking the total balance of objectives into account. Therefore, a new problem of value judgment called *value trade-off* arises. Evaluation modeling treats this problem that is peculiar to decision making with multiple objectives. Here we assume a *decision maker* who is responsible for the final decision. In some cases, there may be many decision makers, for which cases the decision making problems are called *group decision problems*. We will consider cases with a single decision maker in this book. The decision maker's value is usually represented by saying whether or not an alternative x is preferred to another alternative x', or equivalently whether or not $f(x)$ is preferred to $f(x')$. In other words, the decision maker's value is represented by some binary relation defined over X or $f(X)$. Since such a binary relation representing the decision maker's preference usually becomes an *order*, it is called a *preference order*. In this book, the decision maker's preference order is supposed to be defined on the so-called *criteria space* Y, which includes the set $f(X)$. Several kinds of preference orders will be possible, sometimes, the decision maker cannot judge whether or not $f(x)$ is preferred to $f(x')$. Roughly speaking, such an order that admits incomparability for a pair of objects is called a *partial order*, whereas the order requiring the comparability for every pair of objects is called a *weak order* (or *total order*). In practice, we often observe a partial order for the decision maker's preference. Unfortunately, however, an optimal solution in the sense of one that is most preferred with respect to the order, whence the notion of *optimality* does not necessarily exist for partial orders. Instead of strict optimality, we introduce in multiobjective optimization the notion of *efficiency*. A vector $f(\hat{x})$ is said to be efficient if there is no $f(x)$ $(x \in X)$ preferred to $f(\hat{x})$ with respect to the preference order. The final decision is usually made among the set of efficient solutions.

One approach to evaluation modeling is to find a scalar-valued function $u(f_1, \ldots, f_p)$ representing the decision maker's preference, which is called a *preference function* in this book. A preference function in decision making under risk is called a *utility function*, whereas the one in decision making under certainty is called a *value function*. The theory regarding existence, uniqueness, and practical representation of such a utility or value function is called the *utility and value theory*. Once we obtain such a preference function, our problem reduces to the traditional mathematical programming:

$$\text{Maximize} \quad u(f_1(x), \ldots, f_p(x)) \quad \text{over} \quad x \in X.$$

Another popular approach is the *interactive programming* that performs the solution search and evaluation modeling. In this approach, the solution is searched without identifying the preference function by eliciting iteratively some local information on the decision maker's preference.

Kuhn and Tucker [K10] first gave some interesting results concerning multiobjective optimization in 1951. Since then, research in this field has made remarkable progresss both theoretically and practically. Some of the earliest attempts to obtain conditions for efficiency were carried out by Kuhn and Tucker [K10], and Arrow *et al.* [A5]. Their research has been inherited by Da Cunha and Polak [D1], Neustadt [N14], Ritter [R4–R6], Smale [S10, S11], Aubin [A7], and others. After the 1970s, practical methodology such as utility and value analysis and interactive programming methods have been actively researched as tools for supporting decision making, and many books and conference proceedings on this topic have been published. (See, for example, Lee [L1], Cochrane and Zeleny [C12], Keeney and Raiffa [K6], Leitmann and Marzollo [L3], Leitmann [L2], Wilhelm [W15], Zeleny [Z4–Z6], Thiriez and Zionts [T14], Zionts [Z7], Starr and Zeleny [S13], Nijkamp and Delft [N18], Cohon [C13], Hwang and Masud [H17], Salkuvadze [S3], Fandel and Gal [F2], Rietveld [R3], Hwang and Yoon [H18], Morse [M5], Goicoeche *et al.* [G8], Hansen [H3], Chankong and Haimes [C6], and Grauer and Wierzbicki [G10].)

On the other hand, duality and stability, which play an important role in traditional mathematical programming, have been extended to multiobjective optimization since the late 1970s. Isermann [I5–I7] developed multiobjective duality in the linear case, while the results for nonlinear cases have been given by Schönfeld [S6], Rosinger [R10], Guglielmo [G12], Tanino and Sawaragi [T9, T11], Mazzoleni [M3], Bitran [B13], Brumelle [B21], Corley [C16], Jahn [J1], Kawasaki [K2, K3], Luc [L10], Nakayama [N5], and others. Stability for multiobjective optimization has been developed by Naccache [N2] and Tanino and Sawaragi [T10].

This book will be mainly concerned with some of the theoretical aspects in multiobjective optimization; in particular, we will focus on existence, necessary/sufficient conditions, stability, Lagrange duality, and conjugate duality for efficient solutions. In addition, some methodology such as utility and value theory and interactive programming methods will also be discussed.

Chapter 2 is devoted to some mathematical preliminaries. The first section gives a brief review of the elements of convex analysis that play an important role not only in traditional mathematical programming but also in multiobjective optimization. The second section describes point-to-set maps that play a very important role in the theory of multiobjective optimization, since the efficient solutions usually constitute a *set*. The concepts of continuity and convexity of point-to-set maps are introduced. These concepts are fundamental for existence and necessary/sufficient conditions for efficient solutions. The third section is concerned with a brief explanation of preference order and domination structures.

Chapter 3 begins with the introduction of several possible concepts for solutions in multiobjective optimization. Above all, efficient solutions will be the subject of primary consideration in subsequent theories. Next, some properties of efficient solutions, such as existence, external stability, connectedness, and necessary/sufficient conditions, will be discussed.

Chapter 4 develops the stability theory in multiobjective optimization. The factors that specify multiobjective optimization problems are the objective functions with some constraints and the preference structure of the decision maker. Therefore, the stability of the solution set for perturbations of those factors is considered by using the concept of continuity of the solution map (which is a point-to-set map). A number of illustrative examples will be shown.

Chapter 5 will be devoted to the duality theory in multiobjective optimization. Duality is a fruitful result in traditional mathematical programming and is very useful both theoretically and practically. Consequently, it is quite interesting to extend the duality theory to the case of multiobjective optimization. In this chapter, the duality of linear multiobjective optimization will be introduced. Next, a more general duality theory for nonlinear cases will be discussed in parallel with the case of ordinary convex programming. Given a convex multiobjective programming problem, some new concepts such as the primal map, the dual map, and the vector-valued Lagrangian will be defined. The Lagrange multiplier theorem, the saddle-point theorem, and the duality theorem will be obtained via geometric consideration.

Chapter 6 will develop the conjugate duality theory in multiobjective optimization. Rockafellar [R8] constructed a general and comprehensive duality theory for an ordinary convex optimization problem by embedding it in a family of perturbed problems and associating a dual problem with it by using conjugate functions. In this chapter, we shall introduce the concept of conjugate maps, which are extensions of conjugate functions, and develop duality for convex multiobjective optimization problems. In addition, recent results on this topic from several approaches will be introduced.

Chapter 7 deals with the methodology for practical implementation of multiobjective optimization. Although various methods for multiobjective optimization have been developed, we shall focus on some of the main approaches (for example, the ultility and value theory and interactive programming methods). However, our aim is not merely to introduce these methods, but to clarify the noticeable points about them. In the first section, the utility and value theory will be treated. There, we shall introduce several results on existence, uniqueness, and representation of preference functions that represent the decision maker's preference. Several practical implications of these methods, in particular, von Neumann–Morgenstern utility functions, will also be discussed along with clarification of some of their controversial points. In the next section, we shall discuss interactive programming methods, showing some typical methods as examples. Desirable properties of this approach will be clarified throughout the discussion.

2 MATHEMATICAL PRELIMINARIES

This chapter is devoted to mathematical preliminaries. First, some fundamental results in convex analysis, mainly according to Rockafellar [R7], are reviewed. Second, the concepts of continuity and convexity of vector-valued point-to-set maps are introduced. Finally, the concepts of preference orders, domination structures, and nondominated solutions are introduced to provide a way of specifying solutions for multiobjective optimization problems.

2.1 Elements of Convex Analysis

As is well known, convexity plays an important role in the theory of optimization with a single objective. A number of books dealing with convex optimization both in finite and infinite dimensional spaces have been published (Rockafellar [R7], Stoer and Witzgall [S15], Holmes [H12], Bazaraa and Shetty [B3], Ekeland and Temam [E1], Ioffe and Tihomirov [I4], Barbu and Precupanu [B1], and Ponstein [P2]).

Since they are also fundamental in the theory of multiobjective optimization, some elementary results are summarized below. All spaces considered are finite-dimensional Euclidean spaces. The fundamental text for this section is "Convex Analysis" by Rockafellar [R7], and the reader might refer to it for details.

2.1.1 Convex Sets

In this subsection, we will first look at convex sets, which are fundamental in convex analysis.

Definition 2.1.1 (*Convex Set*)

A subset X of R^n is said to be convex if

$$\alpha x^1 + (1-\alpha)x^2 \in X \qquad \text{for any} \quad x^1, x^2 \in X \qquad \text{and any} \quad \alpha \in [0,1].$$

Proposition 2.1.1

A subset X of R^n is convex if and only if $x^i \in X$ $(i = 1, \ldots, k)$, $\alpha_i \geqq 0$ $(i = 1, \ldots, k)$, and $\sum_{i=1}^k \alpha_i = 1$ imply $\sum_{i=1}^k \alpha_i x^i \in X$; that is, if and only if it contains all convex combinations of its elements.

Definition 2.1.2 (*Convex Hull*)

The intersection of all the convex sets containing a given subset X of R^n is called the convex hull of X and is denoted by co X.

Proposition 2.1.2

(i) co X is the unique smallest convex set containing X.
(ii) co $X = \{\sum_{i=1}^k \alpha_i x^i : \alpha_i \geqq 0, \sum_{i=1}^k \alpha_i = 1, x^i \in X \ (i = 1, \ldots, k)\}$;

that is, co X consists of all the convex combinations of the elements of X.

Definition 2.1.3 (*Cone and Convex Cone*)

A subset K of R^n is called a cone if $\alpha x \in K$ whenever $x \in K$ and $\alpha > 0$. Moreover, a cone K is said to be a convex cone when it is also convex.

Proposition 2.1.3

A set K in R^n is a convex cone if and only if

(i) $x \in K$ and $\alpha > 0$ imply $\alpha x \in K$,
(ii) $x^1, x^2 \in K$ imply $x^1 + x^2 \in K$.

Definition 2.1.4 (*Pointed Cone and Acute Cone*)

A cone $K \subset R^n$ is said to be pointed if $-x \notin K$ when $x \neq 0$ and $x \in K$. In other words, K is pointed if it does not contain any nontrivial subspaces. It is said to be acute if there is an open halfspace (see Definition 2.1.16)

$$\mathring{H}^+ = \{x \in R^n : \langle x, x^* \rangle > 0\} \qquad x^* \neq 0$$

such that

$$\text{cl}\, K \subset \mathring{H}^+ \cup \{0\}.$$

Proposition 2.1.4

When K is a convex cone, it is acute if and only if $\mathrm{cl}\, K$ is pointed.

Of course, *acuteness* is a stronger concept than *pointedness*. If K is closed and convex, these concepts coincide.

Definition 2.1.5 (*Polar and Strict Polar*)

For a subset X in R^n, its positive polar X° is defined by

$$X^\circ = \{x^* \in R^n : \langle x, x^* \rangle \geqq 0 \text{ for any } x \in X\}.$$

The strict positive polar X^{so} of X is defined by

$$X^{so} = \{x^* \in R^n : \langle x, x^* \rangle > 0 \text{ for any nonzero } x \in X\}.$$

Proposition 2.1.5

Let X, X_1, and X_2 be sets in R^n. Then

(i) the polar X° is a closed convex cone,
(ii) the strict polar X^{so} is a convex cone,
(iii) $X^\circ = (\mathrm{cl}\, X)^\circ$,
(iv) if X is open, $X^{so} \cup \{0\} = X^\circ$,
(v) $X_1 \subset X_2$ implies $X_2^\circ \subset X_1^\circ$ and $X_2^{so} \subset X_1^{so}$, and
(vi) if K is a nonempty convex cone, $K^{\circ\circ} = \mathrm{cl}\, K$.

Proposition 2.1.6

Let K_1 and K_2 be cones in R^n. Then

(i) if K_1 and K_2 are nonempty,

$$(K_1 + K_2)^\circ = K_1^\circ \cap K_2^\circ = (K_1 \cup K_2)^\circ,$$

(ii) $(K_1 \cap K_2)^\circ \supset K_1^\circ + K_2^\circ = \mathrm{co}(K_1^\circ \cup K_2^\circ)$,
(iii) if K_1 and K_2 are closed convex with nonempty intersection,

$$(K_1 \cap K_2)^\circ = \mathrm{cl}(K_1^\circ + K_2^\circ) = \mathrm{cl}\,\mathrm{co}(K_1^\circ \cup K_2^\circ).$$

If $\mathrm{ri}\, K_1 \cap \mathrm{ri}\, K_2 \neq \varnothing$, in addition, then

$$(K_1 \cap K_2)^\circ = K_1^\circ + K_2^\circ.$$

Furthermore, the last equality holds also for convex polyhedral cones (Definition 2.1.7) K_1 and K_2.

Proposition 2.1.7†

Let K be a cone in R^n. Then

(i) $\operatorname{int} K^\circ \neq \emptyset$ if and only if K is acute,
(ii) when K is acute, $\operatorname{int} K^\circ = (\operatorname{cl} K)^{s\circ}$.

Proof

(i) Suppose first that $\operatorname{int} K^\circ \neq \emptyset$. Let $x^* \neq 0$ in $\operatorname{int} K^\circ$ and $\mathring{H}^+ = \{x \in R^n : \langle x, x^* \rangle > 0\}$. We now show that $\operatorname{cl} K \subset \mathring{H}^+ \cup \{0\}$. Suppose, to the contrary, that $\langle x, x^* \rangle \leqq 0$ for some nonzero $x \in \operatorname{cl} K$. Since $x^* \in \operatorname{int} K^\circ$, $x^* - \alpha x \in K^\circ$ for sufficiently small $\alpha > 0$. Since $K^\circ = (\operatorname{cl} K)^\circ$ [Proposition 2.1.5 (iii)],

$$\langle x, x^* - \alpha x \rangle \geqq 0.$$

However, this implies that $\langle x, x^* \rangle \geqq \alpha \langle x, x \rangle > 0$, which contradicts the hypothesis $\langle x, x^* \rangle \leqq 0$. Hence, $\operatorname{cl} K \subset \mathring{H}^- \cup \{0\}$.

Conversely, suppose that K is acute, i.e., $\operatorname{cl} K \subset \mathring{H}^+ \cup \{0\} = \{x \in R^n : \langle x, x^* \rangle > 0\} \cup \{0\}$ for some $x^* \neq 0 \in R^n$. We now prove that $x^* \in \operatorname{int} K^\circ$. If we suppose the contrary, there exist sequences $\{x^{*k}\} \subset R^n$ and $\{x^k\} \subset K$ such that

$$x^{*k} \to x^* \qquad \text{and} \qquad \langle x^k, x^{*k} \rangle < 0 \qquad \text{for all} \quad k.$$

We may assume without loss of generality that $\|x^k\| = 1$ for all k and so $x^k \to x$ for some $x \in \operatorname{cl} K$ with $\|x\| = 1$. Clearly then $\langle x^k, x^{*k} \rangle \to \langle x, x^* \rangle \leqq 0$. However, this contradicts the fact that $\operatorname{cl} K \subset \mathring{H}^+ \cup \{0\}$.

(ii) The inclusive relation $\operatorname{int} K^\circ \subset (\operatorname{cl} K)^{s\circ}$ holds whether K is acute or not. In fact, let $x^* \in \operatorname{int} K^\circ$, and suppose that $x^* \notin (\operatorname{cl} K)^{s\circ}$. Then there exists nonzero $x' \in \operatorname{cl} K$ such that $\langle x', x^* \rangle \leqq 0$. Since $x^* - \alpha x' \in K^\circ = (\operatorname{cl} K)^\circ$ for sufficiently small $\alpha > 0$, then

$$\langle x', x^* \rangle = \langle x', x^* - \alpha x' \rangle + \alpha \langle x', x' \rangle > 0,$$

which is a contradiction.

To prove the converse, let $x^* \notin \operatorname{int} K^\circ$. Then there exist sequences $\{x^{*k}\} \subset R^n$ and $\{x^k\} \subset K$ such that

$$x^{*k} \to x^* \qquad \text{and} \qquad \langle x^k, x^{*k} \rangle < 0 \qquad \text{for all} \quad k.$$

We may assume without loss of generality that $\|x^k\| = 1$ and so $x^k \to x$ for some $x \in \operatorname{cl} K$ with $\|x\| = 1$. Then clearly $\langle x^k, x^{*k} \rangle \to \langle x, x^* \rangle \leqq 0$. This implies that $x^* \notin (\operatorname{cl} K)^{s\circ}$, which is what we wished to prove.

† Yu [Y1].

Proposition 2.1.8†

Let K_1 be a nonempty, closed, pointed convex cone and K_2 be a nonempty, closed, convex cone. Then

$$(-K_1) \cap K_2 = \{0\}$$

if and only if

$$K_1^{so} \cap K_2^{\circ} \neq \varnothing.$$

Proof First suppose that there exists $x^* \in K_1^{so} \cap K_2^{\circ}$. Then, for any nonzero $x \in (-K_1) \cap K_2$,

$$\langle -x, x^* \rangle > 0 \qquad \text{and} \qquad \langle x, x^* \rangle \geqq 0,$$

which is impossible. Therefore $(-K_1) \cap K_2 = \{0\}$. (Note that 0 is always contained in $(-K_1) \cap K_2$ because both K_1 and K_2 are nonempty closed cones.)

Conversely, suppose that $K_1^{so} \cap K_2^{\circ} = \varnothing$. Then, by the separation theorem of convex sets (Theorem 2.1.1), there exist $x \neq 0 \in R^n$ and $\beta \in R$ such that

$$\langle x, x^* \rangle \leqq \beta \qquad \text{for any} \quad x^* \in K_1^{so},$$

$$\langle x, x^* \rangle \geqq \beta \qquad \text{for any} \quad x^* \in K_2^{\circ}.$$

Since both K_1^{so} and K_2 are cones, β should be equal to 0. Hence,

$$x \in [-(K_1^{so})^{\circ}] \cap (K_2^{\circ})^{\circ}.$$

From Propositions 2.1.5 and 2.1.7,

$$(K_1^{so})^{\circ} = (\operatorname{int} K_1^{\circ})^{\circ} = (K_1^{\circ})^{\circ} = \operatorname{cl} K_1 = K_1,$$

and from Proposition 2.1.5 (vi),

$$(K_2^{\circ})^{\circ} = \operatorname{cl} K_2 = K_2.$$

Therefore, $x \in (-K_1) \cap K_2$, i.e., $(-K_1) \cap K_2 \neq \{0\}$.

Definition 2.1.6 (*Recession Cone*)

Let X be a convex set in R^n. Its recession cone 0^+X is defined by

$$0^+X = \{x' \in R^n : x + \alpha x' \in X \text{ for any } x \in X \text{ and } \alpha > 0\}.$$

Of course, 0^+X is a convex cone containing the origin.

† Borwein [B17] and Bitran and Magnanti [B14].

Proposition 2.1.9

If $\{X_i : i \in I\}$ is an arbitrary collection of closed convex sets in R^n whose intersection is not empty, then

$$0^+\left(\bigcap_{i\in I} X_i\right) = \bigcap_{i\in I} 0^+X_i.$$

Proposition 2.1.10

A nonempty closed convex set X in R^n is bounded if and only if its recession cone $0^+X = \{0\}$.

Proposition 2.1.11

Let X_1 and X_2 be nonempty closed convex sets in R^n. If

$$0^+X_1 \cap (-0^+X_2) = \{0\},$$

then the set $X_1 + X_2$ is closed, and

$$0^+(X_1 + X_2) = 0^+X_1 + 0^+X_2.$$

Definition 2.1.7 (*Polyhedral Convex Set and Polyhedral Convex Cone*)

A set X in R^n is said to be a polyhedral convex set if it can be expressed as the intersection of some finite collection of closed halfspaces (see Definition 2.1.16), i.e., if

$$X = \{x : \langle b^i, x\rangle \leqq \beta_i \ (i = 1, \ldots, m)\},$$

where $b^i \in R^n$ and $\beta_i \in R$ $(i = 1, \ldots, m)$. If $\beta_i = 0$ for all $i = 1, \ldots, m$ in the above expression, X is said to be a polyhedral convex cone.

Definition 2.1.8 (*Finitely Generated Convex Set and Finitely Generated Convex Cone*)

A set X in R^n is said to be a finitely generated convex set if there exist vectors $a^1, a^2, \ldots, a^m$ such that for a fixed integer k $(0 \leqq k \leqq m)$, X can be expressed as

$$X = \left\{x : x = \sum_{i=1}^{m} \alpha_i a^i,\ \alpha_i \geqq 0 \ (i = 1, \ldots, m),\ \sum_{i=1}^{k} \alpha_i = 1\right\}.$$

If X can be expressed as

$$X = \left\{x : x = \sum_{i=1}^{m} \alpha_i a^i,\ \alpha_i \geqq 0\ (i = 1, \ldots, m)\right\},$$

then X is said to be a finitely generated convex cone and $\{a^1, \ldots, a^m\}$ is called the set of generators for the cone.

Proposition 2.1.12

A convex set X is polyhedral if and only if it is finitely generated.

Proposition 2.1.13

The polar of a polyhedral convex set is also polyhedral. In particular, if

$$X = \{x : \langle b^i, x\rangle \leqq 0\ (i = 1, \ldots, m)\},$$

then its polar is

$$X^\circ = \left\{x' : x' = \sum_{i=1}^{m} \alpha_i b^i, \alpha_i \leqq 0\ (i = 1, \ldots, m)\right\}$$

(Farkas' lemma).

Proposition 2.1.14

If X is a nonempty polyhedral convex set, then its recession cone 0^+X is also polyhedral. In fact,

(i) if X can be expressed as

$$X = \{x : \langle b^i, x\rangle \leqq \beta_i\ (i = 1, \ldots, m)\},$$

then

$$0^+X = \{x' : \langle b^i, x'\rangle \leqq 0\ (i = 1, \ldots, m)\},$$

(ii) if X can be expressed as

$$X = \left\{x : x = \sum_{i=1}^{m} \alpha_i a^i,\ \alpha_i \geqq 0\ (i = 1, \ldots, m),\ \sum_{i=1}^{k} \alpha_i = 1\right\},$$

then

$$0^+X = \left\{x' : x' = \sum_{i=k+1}^{m} \alpha_i a^i,\ \alpha_i \geqq 0\ (i = k + 1, \ldots, m)\right\},$$

i.e., 0^+X is the convex cone generated by $\{a^{k+1}, \ldots, a^m\}$.

Proposition 2.1.15

Let X be a polyhedral convex set in R^n and let f be a linear vector-valued function from R^n into R^p. Then the set $f(X)$ is a polyhedral convex set in R^p. In fact, if

$$X = \left\{x : x = \sum_{i=1}^{m} \alpha_i a^i,\ \alpha_i \geqq 0\ (i = 1, \ldots, m),\ \sum_{i=1}^{k} \alpha_i = 1\right\},$$

then

$$f(X) = \left\{x : x = \sum_{i=1}^{m} \alpha_i b^i,\ \alpha_i \geqq 0\ (i = 1, \ldots, m),\ \sum_{i=1}^{k} \alpha_i = 1\right\},$$

where $b^i = f(a^i)$ $(i = 1, \ldots, m)$.

Corollary 2.1.1

If X_1 and X_2 are polyhedral convex sets in R^n, then $X_1 + X_2$ is polyhedral.

We have the following extended concept of convexity of sets (Yu [Y1]), which is often useful in convex multiobjective optimization.

Definition 2.1.9 (*Cone Convexity*)

Given a set X and a convex cone K in R^p, X is said to be K-convex if $X + K$ is a convex set.

Remark 2.1.1

A set X is convex if and only if X is $\{0\}$-convex. Moreover, if X is a convex set, it is D-convex for an arbitrary, nonempty convex cone D.

2.1.2 Convex Functions

In the following, we consider an extended, real-valued function f from R^n to $[-\infty, +\infty]$.

Definition 2.1.10 (*Epigraph*)

Let f be a function from $X \subset R^n$ to $[-\infty, +\infty]$. The set

$$\{(x, \alpha) : x \in X,\ \alpha \in R,\ \alpha \geqq f(x)\}$$

is called the epigraph of f and is denoted by epi f.

Definition 2.1.11 (*Convex Function*)

A function f from $X \subset R^n$ to $[-\infty, +\infty]$ is said to be a convex function on X if epi f is convex as a subset of R^{n+1}. A concave function on X is a function whose negative is convex. An affine function on X is a function which is finite, convex, and concave.

Definition 2.1.12 (*Effective Domain*)

The effective domain of a convex function f on X is given by

$$\{x \in X : \exists \alpha \in R \text{ s.t.} (x, \alpha) \in \operatorname{epi} f\} = \{x \in X : f(x) < +\infty\}$$

and is denoted by dom f.

Definition 2.1.13 (*Proper Convex Function*)

A convex function f on X is said to be proper if $f(x) < +\infty$ for at least one $x \in X$, and if $f(x) > -\infty$ everywhere.

Proposition 2.1.16

Let X be a convex set in R^n and let f be a function from X to $(-\infty, +\infty]$. Then f is convex on X if and only if

$$f(\alpha x^1 + (1-\alpha)x^2) \leqq \alpha f(x^1) + (1-\alpha)f(x^2), \qquad 0 \leqq \alpha \leqq 1,$$

for every x^1 and x^2 in X.

Proposition 2.1.17 (*Jensen's Inequality*)

Let X be a convex set in R^n and let f be a function from R^n to $(-\infty, +\infty]$. Then f is convex on X if and only if

$$f(\alpha_1 x^1 + \cdots + \alpha_k x^k) \leqq \alpha_1 f(x^1) + \cdots + \alpha_k f(x^k),$$

whenever $\alpha_i \geqq 0$ $(i = 1, \ldots, k)$, $\sum_{i=1}^k \alpha_i = 1$, and $x^i \in X$ $(i = 1, \ldots, k)$.

Definition 2.1.14 (*Indicator Function*)

Let X be a subset of R^n. The indicator function $\delta(\cdot \mid X)$ of X is defined by

$$\delta(x \mid X) = \begin{cases} 0 & \text{if} \quad x \in X, \\ +\infty & \text{if} \quad x \notin X. \end{cases}$$

Proposition 2.1.18

A subset X of R^n is a convex set if and only if the indicator function $\delta(\cdot \mid X)$ of X is a convex function.

Proposition 2.1.19

For any convex function f on a convex set X and any $\alpha \in [-\infty, +\infty]$, the level sets $\{x \in X : f(x) < \alpha\}$ and $\{x \in X : f(x) \leqq \alpha\}$ are convex.

Proposition 2.1.20

Let $\{f_i\}_{i \in I}$ be a collection of convex functions on a convex set $X \subset R^n$, and let f be a function on X defined by

$$f(x) = \sup_{i \in I} f_i(x), \qquad x \in X.$$

Then f is a convex function on X.

Definition 2.1.15 (*Cone Convex Function*)

Let X be a convex set in R^n, f be a function from X into R^p, and D be a convex cone in R^p. Then, f is said to be D-convex if for any $x^1, x^2 \in X$ and for any α $(0 \leqq \alpha \leqq 1)$,

$$\alpha f(x^1) + (1 - \alpha)f(x^2) - f(\alpha x^1 + (1 - \alpha)x^2) \in D.$$

Proposition 2.1.21

Let X be a convex set in R^n, f be a function from R^n into R^p, and D be a convex cone in R^p. If the function f is D-convex, then the set $f(X)$ is D-convex.

Proposition 2.1.22

Let X be a convex set in R^n and $f = (f_1, \ldots, f_p)$ be a function from R^n into R^p. The function f is R^p_+-convex if and only if each f_i is convex, and in this case $f(X)$ is R^p_+-convex.

2.1.3 Separation Theorems for Convex Sets

This subsection is devoted to the fundamental separation theorems of two convex sets in R^n. The theorems play an essential role in deriving optimal conditions for both single-objective and multiobjective optimization problems.

Definition 2.1.16 (*Hyperplane and Halfspace*)

A subset H of R^n is called a hyperplane if it is represented as

$$H = \{x \in R^n : \langle x, x^* \rangle = \beta\},$$

for some nonzero $x^* \in R^n$ and $\beta \in R$. In this case, the vector x^* is called a normal to the hyperplane H. Moreover, the sets

$$H^-(\text{resp. } \mathring{H}^-) = \{x \in R^n : \langle x, x^* \rangle \leqq (\text{resp.} <) \beta\},$$
$$H^+(\text{resp. } \mathring{H}^+) = \{x \in R^n : \langle x, x^* \rangle \geqq (\text{resp.} >) \beta\}$$

are called closed (resp. open) halfspaces associated with the hyperplane H.

Definition 2.1.17 (*Separation of Sets*)

Let X_1 and X_2 be nonempty sets in R^n. A hyperplane H is said to separate X_1 and X_2 if $X_1 \subset H^+$ and $X_2 \subset H^-$ (or $X_1 \subset H^-$ and $X_2 \subset H^+$). It is said to separate X_1 and X_2 properly if, in addition, $X_1 \cup X_2 \not\subset H$. It is said to separate X_1 and X_2 strongly if there exists a positive number ε such that $X_1 + \varepsilon B \subset H^+$ and $X_2 + \varepsilon B \subset H^-$ (or $X_1 + \varepsilon B \subset H^-$ and $X_2 + \varepsilon B \subset H^+$), where B is the closed unit ball.

Proposition 2.1.23

Let X_1 and X_2 be nonempty sets in R^n. There exists a hyperplane separating X_1 and X_2 properly if and only if there exists a vector x^* such that

(i) $\inf\{\langle x, x^* \rangle : x \in X_1\} \geqq \sup\{\langle x, x^* \rangle : x \in X_2\}$,
(ii) $\sup\{\langle x, x^* \rangle : x \in X_1\} > \inf\{\langle x, x^* \rangle : x \in X_2\}$.

There exists a hyperplane separating X_1 and X_2 strongly if and only if there exists a vector x^* such that

$$\inf\{\langle x, x^* \rangle : x \in X_1\} > \sup\{x, x^* \rangle : x \in X_2\}.$$

Theorem 2.1.1 (*Separation Theorem*)

Let X_1 and X_2 be nonempty convex subsets of R^n. There exists a hyperplane separating X_1 and X_2 properly if and only if $\operatorname{ri} X_1 \cap \operatorname{ri} X_2 = \varnothing$.

Theorem 2.1.2 (*Strong Separation Theorem*)

Let X_1 and X_2 be nonempty, disjoint convex sets. If X_1 is compact and X_2 is closed, then there exists a hyperplane separating X_1 and X_2 strongly.

Theorem 2.1.3

A closed convex set is the intersection of the closed halfspaces that contain it.

Definition 2.1.18 (*Supporting Hyperplane*)

Let X be a subset of R^n. A hyperplane H is said to be a supporting hyperplane to X if

$$X \subset H^- \text{ (or } X \subset H^+) \qquad \text{and} \qquad \text{cl}\, X \cap H \neq \varnothing.$$

Theorem 2.1.4

Let X be a nonempty convex set in R^n, and let x' be a boundary point of X. Then there exists a vector $x^* \in R^n$ such that

$$\langle x', x^* \rangle = \sup\{\langle x, x^* \rangle : x \in X\};$$

that is, there exists a hyperplane supporting X at x'. In this case, x' is said to be a supporting point of X with a (outer) normal (or supporting functional) x^*.

2.1.4 Conjugate Functions

This subsection deals with the concept of conjugacy, which is based on the dual way of describing a function. That is, the epigraph of a convex function (which is a convex set) can be represented as the intersection of all closed halfspaces in R^{n+1} that contain it.

Lemma 2.1.1

Let f be a function from R^n to $[-\infty, +\infty]$. Then the following conditions are equivalent:

(i) f is lower semicontinuous throughout R^n,
(ii) $\{x : f(x) \leqq \alpha\}$ is closed for every $\alpha \in R$, and
(iii) the epigraph of f is a closed set in R^{n+1}.

Definition 2.1.19 (*Closure of a Convex Function*)

Given a function f from R^n to $[-\infty, +\infty]$, the function whose epigraph is the closure in R^{n+1} of epi f is called the lower semicontinuous hull of f. The closure of a convex function f (denoted by cl f) is defined to be the lower semicontinuous hull of f if f does not have the value $-\infty$ anywhere, whereas cl $f(x) \equiv -\infty$ if $f(x) = -\infty$ for some x. A convex function f is said to be closed if cl $f = f$.

Proposition 2.1.24

If f is a proper convex function, then

$$\operatorname{cl} f(x) = \liminf_{x' \to x} f(x').$$

Proposition 2.1.25

A closed convex function f is the pointwise supremum of the collection of all affine functions h such that $h \leqq f$.

Definition 2.1.20 (*Conjugate Function and Biconjugate Function*)

Let f be a convex function from R^n to $[-\infty, +\infty]$. The function f^* on R^n defined by

$$f^*(x^*) = \sup\{\langle x, x^* \rangle - f(x) : x \in R^n\}, \qquad x^* \in R^n$$

is called the conjugate function of f. The conjugate of f^*, i.e., the function f^{**} on R^n defined by

$$f^{**}(x) = \sup\{\langle x, x^* \rangle - f^*(x^*) : x^* \in R^n\}, \qquad x \in R^n,$$

is called the biconjugate function of f.

Proposition 2.1.26

Let f be a convex function. The conjugate function f^* is a closed proper convex function if and only if f is proper. Moreover, $(\operatorname{cl} f)^* = f^*$ and $f^{**} = \operatorname{cl} f$. Thus, the conjugacy operation $f \to f^*$ induces a symmetric one-to-one correspondence in the class of all closed proper convex functions on R^n.

Proposition 2.1.27 (*Fenchel's Inequality*)

Let f be a proper convex function. Then

$$f(x) + f^*(x^*) \geqq \langle x, x^* \rangle \qquad \text{for any} \quad x \text{ and } x^*.$$

Definition 2.1.21 (*Support Function*)

Let X be a convex set in R^n. The conjugate function of the indicator function $\delta(\cdot \mid X)$ of X, which is given by

$$\delta^*(x^* \mid X) = \sup\{\langle x, x^* \rangle : x \in X\},$$

is called the support function of X.

Proposition 2.1.28

If X is a nonempty, closed convex set in R^n, then $\delta^*(\cdot \mid X)$ is a closed, proper convex function that is positively homogeneous. Conversely, if γ is a positively homogeneous, closed, proper convex function, it defines a nonempty closed convex set X' by

$$X' = \{x' \in R^n : \langle x, x' \rangle \leqq \gamma(x) \text{ for all } x \in R^n\}.$$

2.1.5 Subgradients of Convex Functions

Differentiability facilitates the analysis of optimization problems. For convex functions, even when they are not differentiable, we might consider subgradients instead of nonexistent gradients.

Definition 2.1.22 (*One-Sided Directional Derivative*)

Let f be a function from R^n to $[-\infty, +\infty]$, and let x be a point where f is finite. The one-sided directional derivative of f at x with respect to a vector d is defined to be the limit

$$f'(x; d) = \lim_{t \downarrow 0} \{[f(x + td) - f(x)]/t\},$$

if it exists.

Remark 2.1.2

If f is actually differentiable at x, then

$$f'(x; d) = \langle \nabla f(x), d \rangle \qquad \text{for any} \quad d,$$

where $\nabla f(x)$ is the gradient of f at x.

Definition 2.1.23 (*Subgradient*)

Let f be a convex function from R^n to $[-\infty, +\infty]$. A vector $x^* \in R^n$ is said to be a subgradient of f at x if

$$f(x') \geqq f(x) + \langle x^*, x' - x \rangle \qquad \text{for any} \quad x' \in R^n.$$

The set of all subgradients of f at x is called the subdifferential of f at x and is denoted by $\partial f(x)$. If $\partial f(x)$ is not empty, f is said to be subdifferentiable at x.

Remark 2.1.3

When f is finite at x, we have the following geometric interpretation of the subgradient: x^* is a subgradient of f at x if and only if $(x^*, -1) \in R^{n+1}$ is a normal vector to the supporting hyperplane of epi f at $(x, f(x))$.

Proposition 2.1.29

Let f be a convex function from R^n to $[-\infty, +\infty]$. Then

$$f(\hat{x}) = \min\{f(x) : x \in R^n\} \qquad \text{if and only if} \quad 0 \in \partial f(\hat{x}).$$

Proposition 2.1.30

Let f be a convex function and suppose that $f(x)$ is finite. Then x^* is a subgradient of f at x if and only if

$$f'(x;d) \geqq \langle x^*, d \rangle \quad \text{for any} \quad d.$$

Proposition 2.1.31

If f is a proper convex function, then $\partial f(x)$ is nonempty for $x \in \mathrm{ri}(\mathrm{dom}\, f)$, and

$$f'(x;d) = \sup\{\langle x^*, d \rangle : x^* \in \partial f(x)\} = \delta^*(d \mid \partial f(x)).$$

Moreover, $\partial f(x)$ is a nonempty bounded set if and only if $x \in \mathrm{int}(\mathrm{dom}\, f)$, in which case $f'(x;d)$ is finite for every d.

Corollary 2.1.2

If f is a finite convex function on R^n, then at each point x, the subdifferential $\partial f(x)$ is a nonempty, closed, bounded convex set and

$$f'(x;d) = \max\{\langle x^*, d \rangle : x^* \in \partial f(x)\}.$$

Proposition 2.1.32

Let f be a convex function and x be a point at which f is finite. If f is differentiable at x, then $\nabla f(x)$ is the unique subgradient of f at x, such that in particular

$$f(x') \geqq f(x) + \langle \nabla f(x), x' - x \rangle \qquad \text{for any} \quad x'.$$

Conversely, if f has a unique subgradient at x, then f is differentiable at x.

Proposition 2.1.33

Let f be a proper convex function on R^n. Then

$$x^* \in \partial f(x) \qquad \text{if and only if} \quad f(x) + f^*(x^*) = \langle x^*, x \rangle.$$

Moreover, if f is a closed proper convex function, $x^* \in \partial f(x)$ is equivalent to $x \in \partial f^*(x^*)$.

Proposition 2.1.34

Let $f_1, f_2, \ldots, f_m$ be proper convex functions on R^n, and let $f = f_1 + f_2 + \cdots + f_m$. Then

$$\partial f(x) \supset \partial f_1(x) + \partial f_2(x) + \cdots + \partial f_m(x) \qquad \text{for any} \quad x.$$

If the convex sets $\operatorname{ri}(\operatorname{dom} f_i)$ $(i = 1, \ldots, m)$ have a point in common, then actually

$$\partial f(x) = \partial f_1(x) + \partial f_2(x) + \cdots + \partial f_m(x) \qquad \text{for any} \quad x.$$

2.2 Point-To-Set Maps

This section deals with the concept of point-to-set maps and their properties of continuity and convexity.

2.2.1 Point-To-Set Maps and Optimization

A point-to-set map F from a set X into a set Y is a map that associates a subset of Y with each point of X. Equivalently, F can be viewed as a function from the set X into the power set 2^Y. Some authors use other names such as correspondence and multifunction instead of point-to-set map. In addition to their intrinsic mathematical interest, the study of such maps has been motivated by numerous applications in different fields. Readers interested in the history and the state of the art of the study of point-to-set maps may refer to Huard [H15].

The use of point-to-set maps in the theory of optimization can be seen in several areas. The first one is the application of fixed point theorems in the field of economics and game theory. The most prominent result is related to the existence of an equilibrium or a saddle point. (See, for example, Debreu [D3].) The second application is concerned with the representation of iterative solving methods, or algorithms, for mathematical programming problems. (See, for example, Zangwill [Z2].) The third application, which is closely related to the results in this book, is the study of the stability of the optimal value of a program (or the set of optimal solutions) when the problem data depend on a parameter (Berge [B11], Dantzig *et al.* [D2], Evans and Gould [E2], Hogan [H11], and Fiacco [F4]). For other applications, see Huard [H14].

In a multiobjective optimization problem, it is rather difficult to obtain a unique *optimal* solution. Solving the problem often leads to a solution set.

Thus, if the problem has a parameter, the solution set defines a point-to-set map from the parameter space into the objective (or decision) space. Some point-to-set maps of this type will be investigated in this book.

2.2.2 Continuity of Point-To-Set Maps

A point-to-set map from a set X into a set Y is a function from X into the power set 2^Y. We can consider the concept of continuity if a certain topology (or distance) is defined in the power set 2^Y. Moreover, noting that the inclusion relation between subsets of Y induces a partial order in 2^Y, we may also consider the concepts corresponding to the upper and lower semicontinuity of real-valued functions. Thus, in the theory of point-to-set maps, two kinds of continuity have been studied. Though they have been described and analyzed in a number of different settings, we follow the definitions by Hogan [H11] in this book. Several similar definitions of continuity and relationships between them are summed up in Delahare and Denel [D4]. In the following, sets X and Y are assumed to be subsets of some finite-dimensional Euclidean spaces for simplicity.

Definition 2.2.1 *(Lower Semicontinuity, Upper Semicontinuity, and Continuity)*

If F is a point-to-set map from a set X into a set Y, then F is said to be

(1) lower semicontinuous (l.s.c.) at a point $x \in X$ if $\{x^k\} \subset X$, $x^k \to x$, and $y \in F(x)$ all imply the existence of an integer m and a sequence $\{y^k\} \subset Y$ such that $y^k \in F(x^k)$ for $k \geqq m$ and $y^k \to y$;

(2) upper semicontinuous (u.s.c.) at a point $x \in X$ if $\{x^k\} \subset X$, $x^k \to x$, $y^k \in F(x^k)$, and $y^k \to y$ all imply that $y \in F(x)$;

(3) continuous at a point $x \in X$ if it is both l.s.c. and u.s.c. at x; and

(4) l.s.c. (resp. u.s.c. and continuous) on $X' \subset X$ if it is l.s.c. (resp. u.s.c. and continuous) at every $x \in X'$.

Definition 2.2.2 *(Uniform Compactness)*

A point-to-set map F from a set X to a set Y is said to be uniformly compact near a point $x \in X$ if there is a neighborhood N of x such that the closure of the set $\bigcup_{x \in N} F(x)$ is compact.

Figure 2.1a shows a map that is lower semicontinuous but not upper semicontinuous at x. On the other hand, part b of the figure shows a map that is upper semicontinuous but not lower semicontinuous at x.

We frequently encounter the case in which the feasible set is described by inequalities in a mathematical programming problem. Some properties of maps determined by inequalities have been investigated.

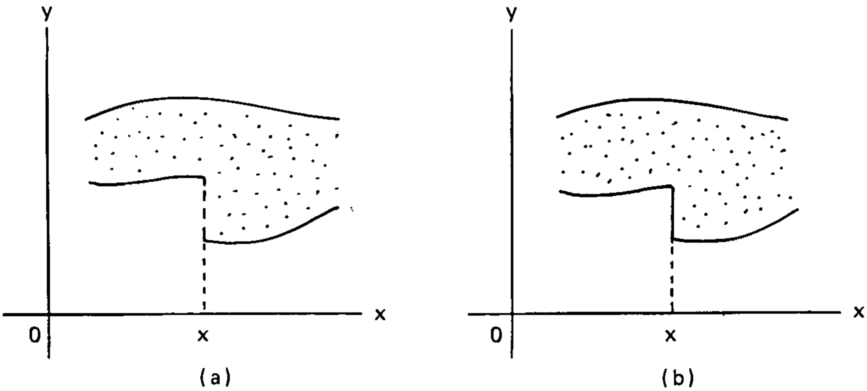

Fig. 2.1. (a) Map that is lower semicontinuous but not upper semicontinuous at x. (b) Map that is upper semicontinuous but not lower semicontinuous at x.

Let

$$X(u) = \{x \in \bar{X} : g(x, u) \leqq 0\},$$

where g is a function from $\bar{X} \times U$ to R^m.

Proposition 2.2.1

If each component of g is lower semicontinuous on $\bar{X} \times u$ and if the set $\bar{X}$ is closed, then X is upper semicontinuous at u.

Proposition 2.2.2

If the set $\bar{X}$ is convex, if each component of g is continuous on $X(\hat{u}) \times \hat{u}$ and convex in x for each fixed $u \in U$, and if there exists an $x \in X$ such that $g(x, u) < 0$, then the map X is lower semicontinuous at $\hat{u}$.

2.2.3 Convexity of Point-To-Set Maps

In this subsection, we will extend the convexity concept of functions to point-to-set maps by generally taking values of subsets of a finite-dimensional Euclidean space.

Definition 2.2.3 (*Cone Epigraph of a Point-To-Set Map*)[†]

Let F be a point-to-set map from R^n into R^p and D be a convex cone in R^p. The set

$$\{(x, y) : x \in R^n,\ y \in R^p,\ y \in F(x) + D\}$$

is called the D-epigraph of F and is denoted by D-epi F.

[†] See Definition 2.1.10.

Definition 2.2.4 *(Cone Convexity and Cone Closedness of a Point-To-Set Map)*

Let F be a point-to-set map from R^n into R^p and let D be a convex cone in R^p; then F is said to be a D-convex (resp. D-closed) if D-epi F is convex (resp. closed) as a subset of $R^n \times R^p$.

Proposition 2.2.3

Let D be a convex cone containing 0 in R^p and let F be a D-convex point-to-set map from R^n into R^p; then F is D-convex if and only if

$$\alpha F(x^1) + (1 - \alpha)F(x^2) \subset F(\alpha x^1 + (1 - \alpha)x^2) + D, \qquad (*)$$

for all $x^1, x^2 \in R^n$ and all $\alpha \in [0, 1]$.

Proof First, assume that F is D-convex, and let $x^1, x^2 \in R^n$, $y^1 \in F(x^1)$, $y^2 \in F(x^2)$, and $\alpha \in [0, 1]$. Since $(x^i, y^i) \in D$-epi F $(i = 1, 2)$, then

$$\alpha(x^1, y^2) + (1 - \alpha)(x^2, y^2) \in D\text{-epi } F,$$

i.e.,

$$\alpha y^1 + (1 - \alpha)y^2 \in F(\alpha x^1 + (1 - \alpha)x^2) + D.$$

Next, suppose conversely that the relationship $(*)$ holds. Let $(x^i, y^i) \in D$-epi F $(i = 1, 2)$ and $\alpha \in [0, 1]$. Since $y^i \in F(x^i) + D$, for $i = 1, 2$,

$$\alpha y^1 + (1 - \alpha)y^2 \in \alpha F(x^1) + (1 - \alpha)F(x^2) + D \subset F(\alpha x^1 + (1 - \alpha)x^2) + D.$$

Hence

$$\alpha(x^1, y^1) + (1 - \alpha)(x^2, y^2) \in D\text{-epi } F,$$

which is what we wished to prove.

Definition 2.2.5 *(Cone-Concavity of a Point-To-Set Map)*

Let F be a point-to-set map from R^n into R^p, and let D be a convex cone in R^p; then F is said to be D-concave if

$$F(\alpha x^1 + (1 - \alpha)x^2) \subset \alpha F(x^1) + (1 - \alpha)F(x^2) + D$$

for all $x^1, x^2 \in R^n$ and all $\alpha \in [0, 1]$.

Remark 2.2.1

Note that the D-concavity of a point-to-set map F is not equivalent to $(-D)$-convexity of F (see Fig. 2.2).

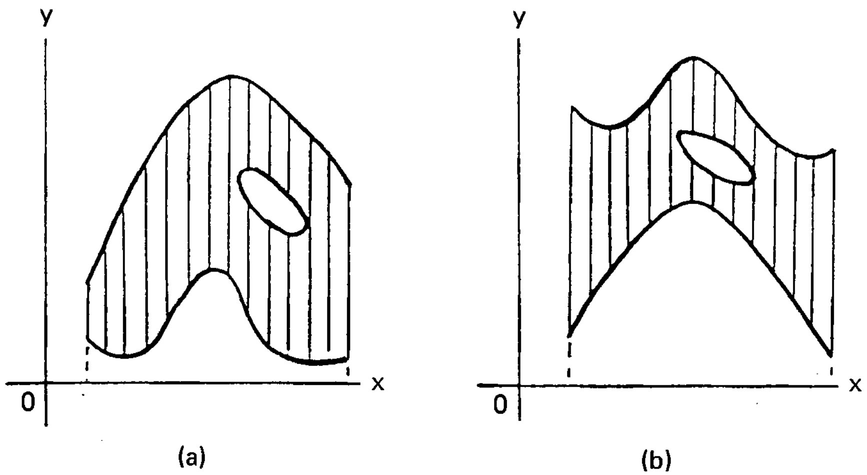

Fig. 2.2. (a) $(-R^1_+)$-convex map and (b) R^1_+-concave map.

2.3 Preference Orders and Domination Structures

In ordinary single-objective optimization problems, the meaning of *optimality* is clear. It usually means the *maximization* or the *minimization* of a certain objective function under given constraints. In multiobjective optimization problems, on the other hand, it is not clear. Let us consider the case in which there are a finite number of objective functions each of which is to be minimized. If there exists a feasible solution, action, or alternative that minimizes all of the objective functions simultaneously, we will have no objection to adopting it as the *optimal* solution. However, we can rarely expect the existence of such an *optimal* solution, since the objectives usually conflict with one another. The objectives, therefore, must be traded off.

Thus, in multiobjective optimization problems, the preference attitudes of the decision maker play an essential role that specifies the meaning of *optimality* or *desirability*. They are very often represented as binary relations on the objective space and are called preference orders. This section is devoted to preference orders and their representation by point-to-set maps (also known as domination structures).

2.3.1 Preference Orders

A preference order represents a preference attitude of the decision maker in the objective space. It is a binary relation on a set $Y = f(X) \subset R^p$, where f is the vector-valued objective function and X the feasible decision set. The basic binary relation $\succ$ means *strict preference*; that is, $y \succ z$ for $y, z \in Y$ implies that the result (objective value) y is preferred to z. From this, we may define two other binary relations $\sim$ and $\succsim$ as

$$y \sim z \quad \text{if and only if} \quad \text{not } y \succ z \quad \text{and} \quad \text{not } z \succ y,$$

$$y \succsim z \quad \text{if and only if} \quad y \succ z \quad \text{or} \quad y \sim z.$$

The relation $\sim$ is called indifference (read $y \sim z$ as y is indifferent to z), and $\succsim$ is called preference–indifference (read $y \succsim z$ as z is not preferred to y).

The binary relations that we use as preference or indifference relations have certain properties. A list of some of these properties are as follows (Fishburn [F7], [F8]): A binary relation R on a set Y is

(1) reflexive if yRy for every $y \in Y$,
(2) irreflexive if not yRy for every $y \in Y$,
(3) symmetric if $yRz \Rightarrow zRy$, for every $y, z \in Y$,
(4) asymmetric if $yRz \Rightarrow$ not zRy, for every $y, z \in Y$,
(5) antisymmetric if $(yRz, zRy) \Rightarrow y = z$, for every $y, z \in Y$,
(6) transitive if $(yRz, zRw) \Rightarrow yRw$, for every $y, z, w \in Y$,
(7) negatively transitive if (not yRz, not zRw) $\Rightarrow$ not yRw, for every $y, z, w \in Y$,
(8) connected or complete if yRz or zRy (possibly both) for every $y, z \in Y$, and
(9) weakly connected if $y \neq z \Rightarrow (yRz$ or $zRy)$ throughout Y.

Lemma 2.3.1

Let R be a binary relation on Y.

(i) If R is transitive and irreflexive, it is asymmetric.
(ii) If R is negatively transitive and asymmetric, it is transitive.
(iii) If R is transitive, irreflexive, and weakly connected, it is negatively transitive.

Definition 2.3.1 (*Strict Partial Order, Weak Order, and Total Order*)

A binary relation R on a set Y is said to be

(i) a strict partial order if R is irreflexive and transitive,
(ii) a weak order if R is asymmetric and negatively transitive, and
(iii) a total order if R is irreflexive, transitive, and weakly connected.

Proposition 2.3.1

If R is a total order, then it is a weak order; and if R is a weak order, then it is a strict partial order.

In this book, the preference order is usually assumed to be at least a strict partial order; that is, irreflexivity of preference (y is not preferred to itself) and transitivity of preference (if y is preferred to z and z is preferred to w, then y is preferred to w) will be supposed.

Given the preference order $\succ$ (or $\succsim$) on a set Y, it is a very interesting problem to find some real-valued function u on Y that which represents the order $\succ$ as

$$u(y) > u(z) \qquad \text{if and only if} \quad y \succ z \qquad \text{for all} \quad y, z \in Y.$$

Definition 2.3.2 (*Preference Function*)[†]

Given the preference order $\succ$ on a set Y, a real-valued function u on Y such that

$$u(y) > u(z) \qquad \text{if and only if} \quad y \succ z \qquad \text{for all} \quad y, z \in Y$$

is called a preference function.

Remark 2.3.1

Note the following relationships:

$$y \sim z \Leftrightarrow \text{not } y \succ z \qquad \text{and} \qquad \text{not } z \succ y \Leftrightarrow u(y) = u(z),$$

$$y \succsim z \Leftrightarrow y \succ z \qquad \text{or} \qquad y \sim z \Leftrightarrow u(y) \geqq u(z).$$

With regard to the existence of the preference function, the following result is fundamental (Fishburn [F5]).

Lemma 2.3.2

If $\succ$ on Y is a weak order, then $\sim$ is an equivalence relation (i.e., reflexive, symmetric, and transitive).

Theorem 2.3.1

If $\succ$ on Y is a weak order and $Y/\sim$ (the set of equivalence classes of Y under $\sim$) is countable, then there is a preference function u on Y.

The utility theory (especially multiattribute utility theory) will be explained in a later chapter. (See Fishburn [F7] and Keeney and Raiffa [K6] for details.)

Given a preference order, what kind of solutions should be searched for?

[†] In some literature preference functions are also called utility functions or value functions. In this book, however, we use the term preference functions to include both of them. See Section 7.1 for more details.

Definition 2.3.3 (*Efficiency*)

Let Y be a feasible set in the objective space R^p, and let $\succ$ be a preference order on Y. Then an element $\hat{y} \in Y$ is said to be an efficient (noninferior) element of Y with respect to the order $\succ$ if there does not exist an element $y \in Y$ such that $y \succ \hat{y}$. The set of all efficient elements is denoted by $\mathscr{E}(Y, \succ)$. That is,

$$\mathscr{E}(Y, \succ) = \{\hat{y} \in Y: \text{there is no } y \in Y \text{ such that } y \succ \hat{y}\}.$$

Our aim in a multiobjective optimization problem might be to find the set of efficient elements (usually not singleton).

Remark 2.3.2

The set of efficient elements in the decision space X may be defined in an analogous way as

$$(X, f^{-1}(\succ)) = \{\hat{x} \in X : \text{there is no } x \in X \text{ such that } f(x) \succ f(\hat{x})\},$$

where $f^{-1}(\succ)$ is the order on X induced by $\succ$ as

$$xf^{-1}(\succ)x' \qquad \text{if and only if} \quad f(x) \succ f(x').$$

2.3.2 Domination Structures

Preference orders (and more generally, binary relationships) on a set Y can be represented by a point-to-set map from Y to Y. In fact, a binary relationship may be considered to be a subset of the product set $Y \times Y$, and so it can be regarded as a graph of a point-to-set map from Y to Y. Namely, we identify the preference order $\succ$ with the graph of the point-to-set map P:

$$P(y) = \{y' \in Y: y \succ y'\}$$

[$P(y)$ is a set of the elements in Y less preferred to y],

$$\text{graph}\, P := \{(y, y') \in Y \times Y: y' \in P(y)\} = \{(y, y') \in Y \times Y: y \succ y'\}.$$

This viewpoint for preference orders is taken, for example, by Ruys and Weddepohl [R11] (though P^{-1} is considered instead of P).

Another way of representing preference orders by point-to-set maps is the concept of domination structures as suggested by Yu [Y1]. For each $y \in Y \subset R^p$, we define the set of domination factors

$$D(y) := \{d \in R^p : y \succ y + d\} \cup \{0\}.$$

This means that deviation $d \in D(y)$ from y is less preferred to the original y. Then the point-to-set map D from Y to R^p clearly represents the given preference order. We call D the domination structure.

Since the feasible set Y is not always fixed, it is more convenient to define the domination structure on a sufficiently large set that includes Y. In this book, D is defined on the whole objective space R^p for simplicity. The point-to-set map D may then have the following properties (the numbers in parentheses correspond to those for the properties of preference orders listed in Subsection 2.3.1): The domination structure $D(\cdot)$ is said to be

(4) asymmetric if $d \in D(y)$, $d \neq 0 \Rightarrow -d \notin D(y + d)$ for all y, i.e., if $y \in y' + D(y')$, $y' \in y + D(y) \Rightarrow y = y'$,
(6) transitive if $d \in D(y)$, $d' \in D(y + d) \Rightarrow d + d' \in D(y)$ for all y, i.e., if $y \in y' + D(y')$, $y' \in y'' + D(y'') \Rightarrow y \in y'' + D(y'')$,
(7) negatively transitive if $d \notin D(y)$, $d' \notin D(y + d) \Rightarrow d + d' \notin D(y)$ for all y.

Efficiency can be redefined with a domination structure.

Definition 2.3.3′ (*Efficiency*)

Given a set Y in R^p and a domination structure $D(\cdot)$, the set of efficient elements is defined by

$$\mathscr{E}(Y, D) = \{\hat{y} \in Y : \text{there is no } y \neq \hat{y} \in Y \text{ such that } \hat{y} \in y + D(y)\}.$$

This set $\mathscr{E}(Y, D)$ is called the efficient set.

Remark 2.3.3

We can induce a domination structure D' on X from a given domination structure D on Y as follows:

$$D'(x) = \{d' \in R^n : f(x + d') \in f(x) + D(f(x))\backslash\{0\}\} \cup \{0\}.$$

If we denote D' simply by $f^{-1}(D)$, the set of efficient solutions in the decision space is

$$\{x : f(x) \in \mathscr{E}(Y, D)\} = \mathscr{E}(X, f^{-1}(D)).$$

For example, if f is linear, [i.e., $f = C$ (a $p \times n$ matrix)] and if $D = R^p_+$, then

$$D'(x) = \{d' : Cd' \geq 0\} \cup \{0\}.$$

The most important and interesting special case of the domination structures is when $D(\cdot)$ is a constant point-to-set map, particularly when $D(y)$ is a constant cone for all y. When $D(y) = D$ (a cone)

(4) asymmetry $\Leftrightarrow [d \in D, d \neq 0 \Rightarrow -d \notin D] \Leftrightarrow D$ is pointed (Definition 2.1.4.),
(5) transitivity $\Leftrightarrow [d, d' \in D \Rightarrow d + d' \in D] \Leftrightarrow D$ is convex.

Thus, pointed convex cones are often used for defining domination structures. We usually write $y \leqq_D y'$ for $y, y' \in R^p$ if and only if $y' - y \in D$ for a convex cone D in R^p. Also $y \leq_D y'$ means that $y' - y \in D$ but $y - y' \notin D$. When D is pointed, $y \leq_D y'$ if and only if $y' - y \in D\backslash\{0\}$. When $D = R^p_+$, it is omitted as $\leqq$ or $\leq$. In other words,

$$y \leqq y' \quad \text{if and only if} \quad y_i \leqq y_i' \quad \text{for all} \quad i = 1, \ldots, p,$$

$$y \leq y' \quad \text{if and only if} \quad y \leqq y' \quad \text{and} \quad y \neq y',$$

i.e.,

$$y_i \leqq y_i' \quad \text{for all} \quad i = 1, \ldots, p,$$

$$y_i < y_i' \quad \text{for some} \quad i \in \{1, \ldots, p\}.$$

Moreover, we write

$$y < y' \quad \text{if and only if} \quad y_i < y_i' \quad \text{for all} \quad i = 1, \ldots, p.$$

Lemma 2.3.3

Let Y be a set and D be a pointed convex cone in R^p. Then

$$y^1 \leqq_D y^2 \quad \text{and} \quad y^2 \leq_D y^3 \quad \text{imply} \quad y^1 \leq_D y^3,$$

$$y^1 \leq_D y^2 \quad \text{and} \quad y^2 \leqq_D y^3 \quad \text{imply} \quad y^1 \leq_D y^3,$$

or, in the form of contraposition,

$$y^1 \nleq_D y^3 \quad \text{and} \quad y^1 \leqq_D y^2 \quad \text{imply} \quad y^2 \nleq_D y^3,$$

$$y^1 \nleq_D y^3 \quad \text{and} \quad y^2 \leqq_D y^3 \quad \text{imply} \quad y^1 \nleq_D y^2.$$

Proof The results are immediate from the fact that

$$D + D\backslash\{0\} = D\backslash\{0\},$$

for a pointed convex cone D.

Some examples of preference orders are the following.

Example 2.3.1

(i) Pareto order: $\succ = \leq$, i.e., $D = R^p_+$.
(ii) Weak Pareto order: $\succ = <$, i.e., $D\backslash\{0\} = \mathring{R}^p_+ = \{y \in R^p : y > 0\}$.

(iii) Lexicographic order: $\succ_l$

$$y \succ_l y' \quad \text{if and only if} \quad \exists k \in \{1, \ldots, p\} \text{ such that } y_i = y'_i \quad \text{for all} \quad i < k \quad \text{and} \quad y_k < y'_k,$$

i.e.,

$$D = \{d \in R^p : \text{there exists a number } k \in \{1, \ldots, p\} \text{ such that } d_i = 0 \quad \text{for all} \quad i < k \quad \text{and} \quad d_k > 0\} \cup \{0\}.$$

Note that this D is a cone, but it is neither open nor closed (Fig. 2.3).

(iv) Order by a polyhedral convex cone: $D = \{d \in R^p : Ad \geqq 0\}$ with an $l \times p$ matrix A. Here we suppose that $\operatorname{Ker} A = \{y : Ay = 0\} = \{0\}$ in order for D to be pointed. Then, this order is essentially reduced to the Pareto order as is shown in the following lemma.

Lemma 2.3.4

Let Y be a set in R^p, A be an $l \times p$ matrix with $\operatorname{Ker} A = \{0\}$, and $D = \{d \in R^p : Ad \geqq 0\}$. Then

$$\mathscr{E}(Y, D) = \{y \in Y : Ay \in \mathscr{E}(AY, R^l_+)\}.$$

Proof Let $\hat{y} \in Y$; $\hat{y} \notin \mathscr{E}(Y, D)$ if and only if there exists $y \in Y$ such that $\hat{y} - y \in D\backslash\{0\}$. Since $\operatorname{Ker} A = \{0\}$, $\hat{y} - y \in D\backslash\{0\}$ if and only if $A\hat{y} \geq Ay$, i.e., if and only if $A\hat{y} \notin \mathscr{E}(AY, R^l_+)$.

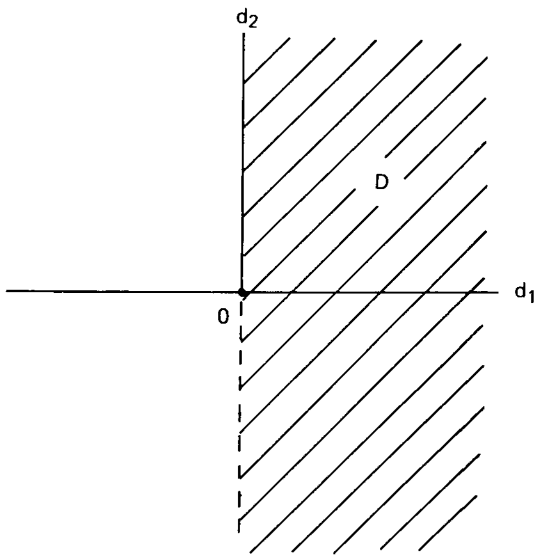

Fig. 2.3. Domination structure of lexicographic order.

3 SOLUTION CONCEPTS AND SOME PROPERTIES OF SOLUTIONS

In this chapter we discuss solution concepts for multiobjective optimization problems and investigate some fundamental properties of solutions. First, efficiency and proper efficiency are introduced as solution concepts. In the second section, existence and external stability of efficient solutions are discussed. The third section is devoted to conditions for the connectedness of efficient sets. Three kinds of characterization for efficient or properly efficient solutions—scalarization, best approximations, and constraint problems—are given in the fourth section. The final section deals with the Kuhn–Tucker conditions for multiobjective problems.

3.1 Solution Concepts

As we have already mentioned, the concept of *optimal* solutions to multiobjective optimization problems is not trivial and in itself debatable. It is closely related to the preference attitudes of the decision makers. The most fundamental solution concept is that of efficient solutions (also called nondominated solutions or noninferior solutions) with respect to the domination structure of the decision maker, which is discussed in the first subsection. We also discuss another slightly restricted concept—proper efficiency—in the second subsection.

3.1.1 Efficient Solutions

In this section we consider the multiobjective optimization problem

(P) minimize $f(x) = (f_1(x), f_2(x), \ldots, f_p(x))$ subject to $x \in X \subset R^n$.

Let

$$Y = f(X) = \{y : y = f(x), x \in X\}.$$

A domination structure representing a preference attitute of the decision maker is supposed to be given as a point-to-set map D from Y into R^p. Though it might be natural to suppose that $D(y) \supset R^p_+$ since a smaller value is preferred for each f_i $(i = 1, \ldots, p)$, we do not assume this in order to deal with more general multiobjective problems.

Definition 3.1.1 (*Efficient Solution*)

A point $\hat{x} \in X$ is said to be an efficient solution to the multiobjective optimization problem (P) with respect to the domination structure D if $f(\hat{x}) \in \mathscr{E}(Y, D)$; that is, if there is no $x \in X$ such that $f(\hat{x}) \in f(x) + D(f(x))$ and $f(x) \neq f(\hat{x})$ (i.e., such that $f(\hat{x}) \in f(x) + D(f(x))\backslash\{0\}$).

The following proposition is immediate.

Proposition 3.1.1

Given two domination structures D_1 and D_2, D_1 is said to be included by D_2 if

$$D_1(y) \subset D_2(y) \qquad \text{for all} \quad y \in Y.$$

In this case,

$$\mathscr{E}(Y, D_1) \supset \mathscr{E}(Y, D_2).$$

Many interesting cases of efficient solutions are obtained when D is a constant point-to-set map whose value is a constant (convex) cone. In such cases, we identify the map (domination structure) with the cone D. Then $\hat{x} \in X$ is an efficient solution to the problem (P) if and only if there is no $x \in X$ such that $f(\hat{x}) - f(x) \in D\backslash\{0\}$; namely, $\hat{x}$ is efficient if and only if $(f(X) - f(\hat{x})) \cap (-D) = \{0\}$. If D is an open cone, it does not contain 0. However, we consider $D(y) = D \cup \{0\}$ and call D itself the domination structure in this book. As a matter of fact it does not matter so much whether D contains 0 or not since the set $D\backslash\{0\}$ is used for the definition of efficient solutions.

Remark 3.1.1

Given a closed convex cone D, some authors call $\hat{x}$ a weakly efficient solution to the problem (P) if $f(\hat{x}) \in \mathscr{E}(Y, \operatorname{int} D)$, i.e., if $(f(X) - f(\hat{x})) \cap (-\operatorname{int} D) = \varnothing$ (Nieuwenhuis [N15, N16], and Corley [C15]). Weakly efficient solutions are often useful, since they are completely characterized by scalarization (see Corollary 3.4.1 later).

The following propositions are often very useful.

Proposition 3.1.2

Let D be a nonempty cone containing 0, then

$$\mathscr{E}(Y, D) \supset \mathscr{E}(Y + D, D)$$

with equality holding if D is pointed and convex.

Proof The result is trivial if Y is empty, so we assume otherwise. First suppose $y \in \mathscr{E}(Y + D, D)$ but $y \notin \mathscr{E}(Y, D)$. If $y \notin Y$, there exist $y' \in Y$ and nonzero $d \in D$ such that $y = y' + d$. Since $0 \in D$, $Y \subset Y + D$. Hence, $y \notin \mathscr{E}(Y + D, D)$, which is a contradiction. If $y \in Y$, we directly have a similar contradiction.

Next suppose that D is pointed and convex, $y \in \mathscr{E}(Y, D)$ but $y \notin \mathscr{E}(Y + D, D)$. Then there exists a $y' \in Y + D$ with $y - y' = d' \in D \backslash \{0\}$. Then $y' = y'' + d''$ with $y'' \in Y$, $d'' \in D$. Hence, $y = y'' + (d' + d'')$ and $d' + d'' \in D$, since D is a convex cone. Since D is pointed, $d' + d'' \neq 0$ and so $y \notin \mathscr{E}(Y, D)$, which leads to a contradiction. This completes the proof of the proposition.

Remark 3.1.2

It is clear that the pointedness of D cannot be eliminated in the inclusion

$$\mathscr{E}(Y, D) \subset \mathscr{E}(Y + D, D).$$

The convexity of D is also essential. In fact, let

$$Y = \{(y_1, y_2) : 0 \leqq y_1 = y_2 \leqq 1\} \subset R^2$$

and

$$D = \{(d_1, d_2) : d_2 \geqq 2d_1 \geqq 0\} \cup \{(d_1, d_2) : d_1 \geqq 2d_2 \geqq 0\},$$

which is pointed but not convex. Then $(1, 1) \in \mathscr{E}(Y, D)$. However,

$$(1, 1) = (0, 0) + (\tfrac{1}{2}, 0) + (\tfrac{1}{2}, 1) \in Y + D + D.$$

Hence $(1, 1) \notin \mathscr{E}(Y + D, D)$ (see Fig. 3.1).

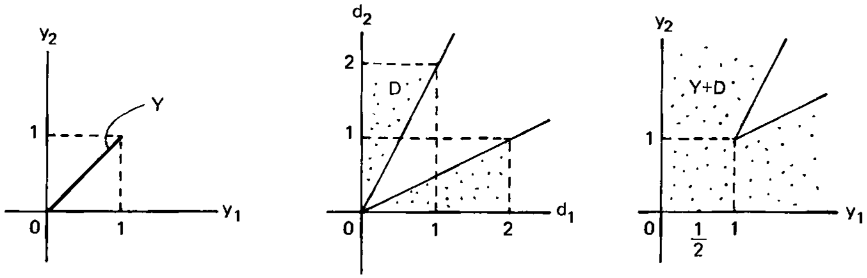

Fig. 3.1. $\mathscr{E}(Y, D) \not\subset \mathscr{E}(Y + D, D)$. (Remark 3.1.2.)

Proposition 3.1.3

Let Y_1 and Y_2 be two sets in R^p, and let D be a constant domination structure on R^p (a constant cone, for example). Then

$$\mathscr{E}(Y_1 + Y_2, D) \subset \mathscr{E}(Y_1, D) + \mathscr{E}(Y_2, D).$$

Proof Let $\hat{y} \in \mathscr{E}(Y_1 + Y_2, D)$. Then $\hat{y} = y^1 + y^2$ for some $y^1 \in Y_1$ and $y^2 \in Y_2$. We show that $y^1 \in \mathscr{E}(Y_1, D)$. If we suppose the contrary, then there exist $y \in Y_1$ and nonzero $d \in D$ such that $y^1 = y + d$. Then $\hat{y} = y^1 + y^2 = y + y^2 + d$ and $y + y^2 \in Y_1 + Y_2$, which contradicts the assumption $\hat{y} \in \mathscr{E}(Y_1 + Y_2, D)$. Similarly we can prove that $y^2 \in \mathscr{E}(Y_2, D)$. Therefore, $\hat{y} \in \mathscr{E}(Y_1, D) + \mathscr{E}(Y_2, D)$.

Remark 3.1.3

The converse inclusion of Proposition 3.1.3 does not always hold. For example, let

$$Y_1 = Y_2 = \{(y_1, y_2) : (y_1)^2 + (y_2)^2 \leqq 1\} \subset R^2$$

and $D = R^2_+$. Then

$$y^1 = (-1, 0) \in \mathscr{E}(Y_1, D) \qquad \text{and} \qquad y^2 = (0, -1) \in \mathscr{E}(Y_2, D).$$

However,

$$\begin{aligned} y^1 + y^2 &= (-1, -1) > (-\sqrt{2}, -\sqrt{2}) \\ &= (-\sqrt{2}/2, -\sqrt{2}/2) + (-\sqrt{2}/2, -\sqrt{2}/2) \in Y_1 + Y_2. \end{aligned}$$

Proposition 3.1.4

Let Y be a set in R^p, D be a cone in R^p, and α be a positive real number. Then

$$\mathscr{E}(\alpha Y, D) = \alpha \mathscr{E}(Y, D).$$

Proof The proof of this proposition is easy and therefore left to the reader.

The most fundamental kind of efficient solution is obtained when D is the nonnegative orthant $R^p_+ = \{y \in R^p : y \geqq 0\}$ and is usually called a Pareto optimal solution or noninferior solution.

Definition 3.1.2 (*Pareto Optimal Solution*)

A point $\hat{x} \in X$ is said to be a Pareto optimal solution (or noninferior solution (Zadeh [Z1])) to the problem (P) if there is no $x \in X$ such that $f(x) \leq f(\hat{x})$.

A number of theoretical papers concerning multiobjective optimization are related to the Pareto optimal solution. In some cases a slightly weaker solution concept than Pareto optimality is often used. It is called weak Pareto optimality, which corresponds to the case in which the domination cone $D\backslash\{0\}$ is equal to the positive orthant $\mathring{R}^p_+ = \{y \in R^p : y > 0\}$.

Definition 3.1.3 (*Weak Pareto Optimal Solution*)[†]

A point $\hat{x} \in X$ is said to be a weak Pareto optimal solution to the problem (P) if there is no $x \in X$ such that $f(x) < f(\hat{x})$.

3.1.2 Properly Efficient Solutions

This subsection is devoted to another slightly strengthened solution concept, proper efficiency. As can be understood from the definitions (particularly from Geoffrion's definition), proper efficiency eliminates unbounded trade-offs between the objectives. It was originally introduced by Kuhn and Tucker [K10], and later followed by Klinger [K7], Geoffrion [G5], Tamura and Arai [T3], and White [W9] for usual vector optimization problems with the domination cone R^p_+. Borwein [B16, B17], Benson [B7, B9], and Henig [H7] dealt with more general closed convex cones as the domination cone. Hence, in this subsection, the domination cone D is assumed to be a nontrivial closed convex cone in R^p unless otherwise noted (Definition 3.1.8 later).

Definition 3.1.4 (*Tangent Cone*)

Let $S \subset R^p$ and $y \in S$. The tangent cone to S at y, denoted by $T(S, y)$, is the set of limits of the form $h = \lim t_k(y^k - y)$, where $\{t_k\}$ is a sequence of nonnegative real numbers and $\{y^k\}$ is a sequence in S with limit y.

[†] cf. Remark 3.1.1.

Remark 3.1.4

The tangent cone $T(S, y)$ is always a closed cone.

Definition 3.1.5 (*Borwein's Proper Efficiency*)†

A point $\hat{x} \in X$ is said to be a properly efficient solution of the multiobjective optimization problem (P) if

$$T(Y + D, f(\hat{x})) \cap (-D) = \{0\}.$$

Proposition 3.1.5

If a point $\hat{x} \in X$ is a properly efficient solution of (P) by the definition of Borwein, then it is also an efficient solution of (P).

Proof If $\hat{x}$ is not efficient, there exists a nonzero vector $d \in D$ such that $d = f(\hat{x}) - y$ for some $y \in Y$. Let $d^k = (1 - 1/k)d \in D$ and $t_k = k$ for $k = 1, 2, \ldots$. Then

$$y + d^k = f(\hat{x}) - d + (1 - 1/k)d = f(\hat{x}) - (1/k)d \to f(\hat{x}) \qquad \text{as} \quad k \to \infty,$$

and

$$t_k(y + d^k - f(\hat{x})) = k(-d + (1 - 1/k)d) = -d \to -d \qquad \text{as} \quad k \to \infty.$$

Hence, $T(Y + D, f(\hat{x})) \cap (-D) \neq \{0\}$, and $\hat{x}$ is not properly efficient in the sense of Borwein.

For example, if

$$X = \{(x_1, x_2) : (x_1)^2 + (x_2)^2 \leqq 1\} \subset R^2,$$

$f_1(x) = x_1$, $f_2(x) = x_2$, and $D = R^2_+$. Then, $(-1, 0)$ and $(0, -1)$ are efficient solutions but not properly efficient solutions (in the sense of Borwein) (see Fig. 3.2).

Definition 3.1.6 (*Projecting Cone*)

Let $S \subset R^p$. The projecting cone of S, denoted by $P(S)$, is the set of all points h of the form $h = \alpha y$, where α is a nonnegative real number and $y \in S$. The projecting cone is also known as the cone generated by S.

Definition 3.1.7 (*Benson's Proper Efficiency*)‡

A point $\hat{x} \in X$ is said to be a properly efficient solution of the problem (P) if

$$\operatorname{cl} P(Y + D - f(\hat{x})) \cap (-D) = \{0\}.$$

† Borwein [B16].

‡ Benson [B7].

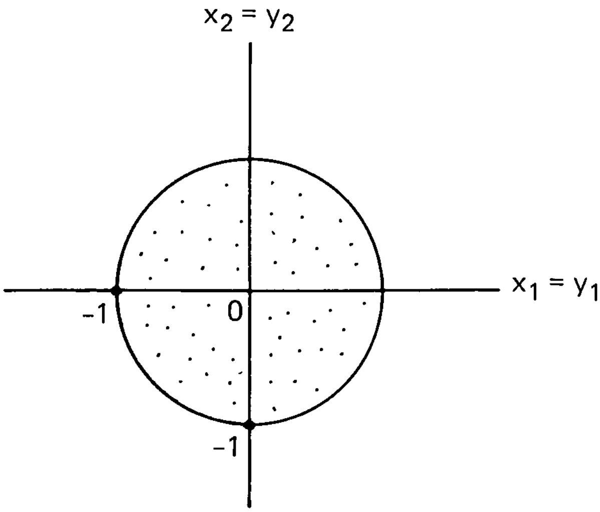

Fig. 3.2. Efficient but not properly efficient solutions.

Since $T(Y + D, f(\hat{x})) \subset \text{cl } P(Y + D - f(\hat{x}))$ Benson's proper efficiency strengthens Borwein's proper efficiency. The converse, however, does not always hold (see Example 3.1.1 later).

Lemma 3.1.1

Let S be a convex set and $y \in S$. Then

$$T(S, y) = \text{cl } P(S - y),$$

which is a closed convex cone.

Proof It is obvious that $\text{cl } P(S - y)$ is a closed convex cone. So it suffices to show that $\text{cl } P(S - y) \subset T(S, y)$, since the converse follows directly from the definitions. Since $T(S, y)$ is closed, we need only to prove that $P(S - y) \subset T(S, y)$. Let $h \in P(S - y)$. Then $h = \beta(y' - y)$ for some $\beta \geqq 0$ and $y' \in S$. Let

$$y^k = [1 - (1/k)]y + (1/k)y'$$

and $t_k = \beta k \geqq 0$. Then $t_k(y^k - y) = \beta(y' - y)$. Hence $y^k \in S$ from the convexity of S, and

$$y^k \to y \quad \text{and} \quad t_k(y^k - y) \to h \quad \text{as} \quad k \to \infty.$$

Thus $h \in T(S, y)$, and the proof is completed.

Theorem 3.1.1

If $\hat{x}$ is a properly efficient solution of the problem (P) in the sense of Benson, it is also a properly efficient solution of (P) in the sense of Borwein. If X is a convex set and if f is a D-convex function on X (see Definition 2.1.15), then the converse also holds; that is, Borwein's proper efficiency is equivalent to Benson's proper efficiency.

Proof If D is a convex cone and f is D-convex on the convex set X, the set $Y + D$ is a convex set (Proposition 2.1.21). Hence, $\operatorname{cl} P(Y + D - f(\hat{x})) = T(Y + D, f(\hat{x}))$ by Lemma 3.1.1. Therefore, the proper efficiency in the sense of Borwein is equivalent to that in the sense of Benson.

The following definition of proper efficiency by Henig does not require the domination cone D to be closed.

Definition 3.1.8 (*Henig's Proper Efficiency*)†

(i) A point $\hat{x} \in X$ is said to be a global properly efficient solution of (P) if

$$f(\hat{x}) \in \mathscr{E}(Y, D'),$$

for some convex cone D' with $D\backslash\{0\} \subset \operatorname{int} D'$.

(ii) A point $\hat{x} \in X$ is a local properly efficient solution of (P) if, for every $\varepsilon > 0$, there exists a convex cone D' with $D\backslash\{0\} \subset \operatorname{int} D'$, such that

$$f(\hat{x}) \in \mathscr{E}((Y + D) \cap (f(\hat{x}) + \varepsilon B), D'),$$

where B is the closed unit ball in R^p.

These definitions are essentially the same as Benson's and Borwein's, respectively.

Theorem 3.1.2

If D is closed and acute, then Henig's global (respectively, local) proper efficiency is equivalent to Benson's (respectively, Borwein's) proper efficiency.

Theorem 3.1.3

Any global properly efficient solution is also locally properly efficient. Conversely, if D is closed and acute, and if

$$Y^+ \cap (-D) = \{0\},$$

† Henig [H7].

then local proper efficiency is equivalent to global proper efficiency. Here Y^+ is an extension of the recession cone of Y and is defined by

$$Y^+ = \{y' : \text{there exist sequences } \{\alpha_k\} \subset R, \{y^k\} \subset Y \text{ such that } \alpha_k > 0, \alpha_k \to 0 \text{ and } \alpha_k y^k \to y'\}.$$

(The condition $Y^+ \cap (-D) = \{0\}$ will later be called the D-boundedness of Y. See Definition 3.2.4.)

The proofs of the above two theorems are rather long, and so they are omitted here. The reader may refer to Henig [H7].

When $D = R^p_+$, Geoffrion's definition of properly efficient solutions is well known.

Definition 3.1.9 (*Geoffrion's Proper Efficiency*)†

When $D = R^p_+$, a point $\hat{x}$ is said to be a properly efficient solution of (P) if it is efficient and if there is some real $M > 0$ such that for each i and each $x \in X$ satisfying $f_i(x) < f_i(\hat{x})$, there exists at least one j such that $f_j(\hat{x}) < f_j(x)$ and

$$(f_i(\hat{x}) - f_i(x))/(f_j(x) - f_j(\hat{x})) \leqq M.$$

Theorem 3.1.4

When $D = R^p_+$, Geoffrion's proper efficiency is equivalent to Benson's proper efficiency, and therefore is stronger than Borwein's proper efficiency.

Proof Geoffrion $\Rightarrow$ Benson: Suppose that $\hat{x}$ is efficient but not properly efficient in the sense of Benson, namely, that there exists a nonzero vector $d \in \operatorname{cl} P(Y + R^p_+ - f(\hat{x})) \cap (-R^p_+)$. Without loss of generality, we may assume that $d_1 < -1$, $d_i \leqq 0$ $(i = 2, \ldots, p)$. Let

$$t_k(f(x^k) + r^k - f(\hat{x})) \to d,$$

where $r^k \in R^p_+$, $t_k > 0$, and $x^k \in X$. By choosing a subsequence we can assume that

$$\tilde{I} = \{i : f_i(x^k) > f_i(\hat{x})\}$$

is constant for all k (and nonempty since $\hat{x}$ is efficient). Take a positive number M. Then there exists some number k_0 such that for $k \geqq k_0$

$$f_1(x^k) - f_1(\hat{x}) < -1/2t_k$$

and

$$f_i(x^k) - f_i(\hat{x}) \leqq 1/2Mt_k \quad (i = 2, \ldots, p).$$

†Geoffrion [G5].

Then for all $i \in \tilde{I}$, we have (for $k \geqq k_0$)

$$0 < f_i(x^k) - f_i(\hat{x}) \leqq 1/2Mt_k$$

and for $k \geqq k_0$,

$$\frac{f_1(\hat{x}) - f_1(x^k)}{f_i(x^k) - f_i(\hat{x})} > \frac{1/2t_k}{1/2Mt_k} = M.$$

Thus $\hat{x}$ is not properly efficient in the sense of Geoffrion.

Benson $\Rightarrow$ Geoffrion: Suppose that $\hat{x}$ is efficient but not properly efficient in the sense of Geoffrion. Let $\{M_k\}$ be an unbounded sequence of positive real numbers. Then, by reordering the objective functions, if necessary, we can assume that for each M_k there exists an $x^k \in X$ such that $f_1(x^k) < f_1(\hat{x})$ and

$$(f_1(\hat{x}) - f_1(x^k))/(f_i(x^k) - f_i(\hat{x})) > M_k$$

for all $i = 2, \ldots, p$ such that $f_i(x^k) > f_i(\hat{x})$. By choosing a subsequence of $\{M_k\}$, if necessary, we can assume that

$$\tilde{I} = \{i : f_i(x^k) > f_i(\hat{x})\}$$

is constant for all k (and nonempty since $\hat{x}$ is efficient). For each k let

$$t_k = (f_1(\hat{x}) - f_1(x^k))^{-1}.$$

Then t_k is positive for all k. Let

$$r_i^k = \begin{cases} 0 & \text{for} \quad i = 1 \quad \text{or} \quad i \in \tilde{I}, \\ f_i(\hat{x}) - f_i(x^k) & \text{for} \quad i \neq 1, \quad i \notin \tilde{I}. \end{cases}$$

Clearly $r^k \in R_+^p$, and

$$t_k(f_i(x^k) + r_i^k - f_i(\hat{x})) = \begin{cases} -1 & \text{for} \quad i = 1, \\ 0 & \text{for} \quad i \neq 1, \quad i \notin \tilde{I}, \end{cases}$$

$$0 < t_k(f_i(x^k) + r_i^k - f_i(\hat{x})) = t_k(f_i(x^k) - f_i(\hat{x})) < M_k^{-1} \quad \text{for} \quad i \in \tilde{I}.$$

Let

$$d_i = \lim_{k \to \infty} t_k[f_i(x^k) + r_i^k - f_i(x)].$$

Clearly $d_1 = -1$ and $d_i = 0$ when $i \neq 1$ and $i \notin \tilde{I}$. Since $\{M_k\}$ is an unbounded sequence of positive real numbers, we have

$$d_i = 0 \qquad \text{for} \quad i \in \tilde{I}.$$

Thus,

$$d = (-1, 0, \ldots, 0) \neq 0 \in \text{cl}\, P(f(X) + R_+^p - f(\hat{x})) \cap (-R_+^p).$$

Hence $\hat{x}$ is not a properly efficient solution of (P) in the sense of Benson, as was to be proved.

Paying attention to multiobjective programming problems, we may consider another type of proper efficiency that was defined earlier than the preceding ones.

A multiobjective programming problem is defined as

$$(\mathrm{P}') \qquad \begin{array}{ll} \text{minimize} & f(x) = (f_1(x), \ldots, f_p(x)) \\ \text{subject to} & x \in X = \{x : g(x) = (g_1(x), \ldots, g_m(x)) \leqq 0\}. \end{array}$$

That is, X is supposed to be specified by m inequality constraints. Here, all the functions f_i and g_j are assumed to be continuously differentiable.

Definition 3.1.10 *(Kuhn–Tucker's Proper Efficiency)*†

A point $\hat{x}$ is said to be a properly efficient solution of the problem (P′) if it is efficient and if there is no h such that

$$\langle \nabla f_i(\hat{x}), h \rangle \leqq 0 \qquad \text{for any} \quad i = 1, \ldots, p.$$

$$\langle \nabla f_i(\hat{x}), h \rangle < 0 \qquad \text{for some} \quad i,$$

and

$$\langle \nabla g_j(\hat{x}), h \rangle \leqq 0 \quad \text{for any} \quad j \in J(\hat{x}) = \{j : g_j(\hat{x}) = 0\}.$$

Theorem 3.1.5

Suppose that the functions f_i and g_j $(i = 1, \ldots, p, j = 1, \ldots, m)$ are convex. If $\hat{x}$ is a properly efficient solution of (P′) in the sense of Kuhn–Tucker, then it is also properly efficient in the sense of Geoffrion.

Proof This theorem can be established as a corollary of Theorems 3.5.1 and 3.5.2, which will be given later.

Definition 3.1.11 *(Kuhn–Tucker Constraint Qualification)*

Problem (P′) is said to satisfy the Kuhn–Tucker constraint qualification at $\hat{x} \in X$ if, for any $h \in R^n$ such that $\langle \nabla g_j(\hat{x}), h \rangle \leqq 0$ and for any $j \in J(\hat{x}) = \{j : g_j(\hat{x}) = 0\}$, there exist $\bar{t} > 0$, a vector-valued function θ on $[0, \bar{t}]$ differentiable at $t = 0$, and a real number $\alpha > 0$, such that

$$\theta(0) = \hat{x}, \qquad g(\theta(t)) \leqq 0 \qquad \text{for any} \quad t \in [0, \bar{t}], \qquad \dot{\theta}(0) = \alpha h.$$

† Kuhn and Tucker [K10].

Theorem 3.1.6

Suppose that (P′) satisfies the Kuhn–Tucker constraint qualification at $\hat{x}$. If $\hat{x}$ is a properly efficient solution of (P′) in the sense of Geoffrion, then it is also properly efficient in the sense of Kuhn–Tucker.

Proof Suppose that $\hat{x}$ is efficient but not properly efficient in the sense of Kuhn–Tucker. Then there exists an $h \in R^n$ such that

$$\begin{aligned} &\langle \nabla f_1(\hat{x}), h \rangle < 0, \\ &\langle \nabla f_i(\hat{x}), h \rangle \leqq 0 \qquad \text{for} \quad i = 2, \ldots, p, \\ &\langle \nabla g_j(\hat{x}), h \rangle \leqq 0 \qquad \text{for} \quad j \in J(\hat{x}). \end{aligned}$$

(Reorder the objective functions, if necessary.) From the Kuhn–Tucker constraint qualification, there exists a continuously differentiable arc $\theta(t)$ $(0 \leqq t \leqq \bar{t})$ such that $\theta(0) = \hat{x}$, $g(\theta(t)) \leqq 0$, and $\dot{\theta}(0) = \alpha h$. Consider a sequence of positive numbers $\{t_k\} \to 0$. By taking a subsequence, if necessary, we may assume that

$$\tilde{I} = \{i : f_i(\theta(t_k)) > f_i(\hat{x})\}$$

is constant. Since, for $i \in \tilde{I}$,

$$f_i(\theta(t_k)) - f_i(\hat{x}) = t_k \langle \nabla f_i(\hat{x}), \alpha h \rangle + o(t_k) > 0$$

and

$$\langle \nabla f_i(\hat{x}), h \rangle \leqq 0,$$

we have

$$\langle \nabla f_i(\hat{x}), \alpha h \rangle = 0, \qquad i \in \tilde{I}.$$

Then, since $\langle \nabla f_1(\hat{x}), h \rangle < 0$,

$$\frac{f_1(\hat{x}) - f_1(\theta(t_k))}{f_i(\theta(t_k)) - f_i(\hat{x})} = \frac{-\langle \nabla f_1(\hat{x}), \alpha h \rangle + o(t_k)/t_k}{\langle \nabla f_i(\hat{x}), \alpha h \rangle + o(t_k)/t_k}$$

diverges to $+\infty$ as $k \to \infty$. This implies that $\hat{x}$ is not properly efficient in the sense of Geoffrion. This completes the proof of the theorem.

Relationships among some different proper efficiency concepts are depicted in Fig. 3.3. Some examples are given in the following.

Example 3.1.1 [Borwein, but not Benson (=Geoffrion)—Fig. 3.4][†]

Let

$$\begin{aligned} X = {} & \{(x_1, x_2) \in R^2 : x_1 + x_2 \geqq 0\} \cup \{(x_1, x_2) \in R^2_+ : x_1 \geqq 1\} \\ & \cup \{(x_1, x_2) \in R^2 : x_2 \geqq 1\}, \end{aligned}$$

$$f_1(x_1, x_2) = x_1, \qquad f_2(x_1, x_2) = x_2,$$

$$D = R^2_+.$$

[†] Benson [B7].

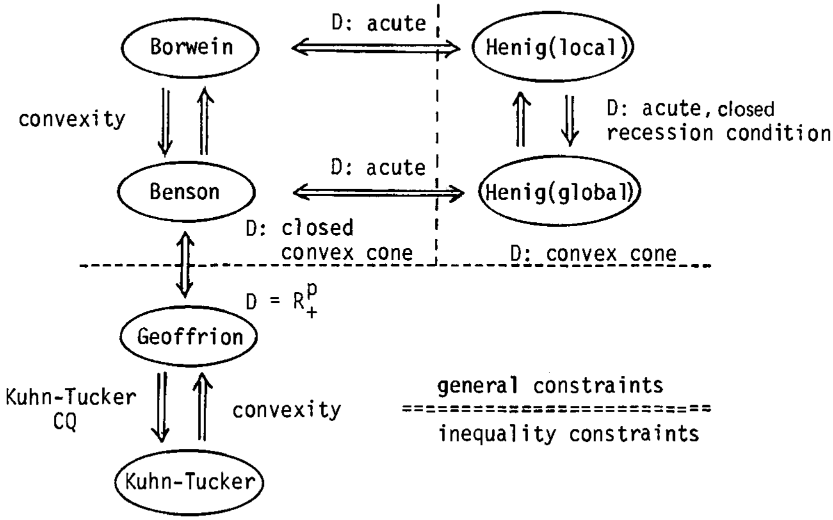

Fig. 3.3. Relationships among some definitions of proper efficiency.

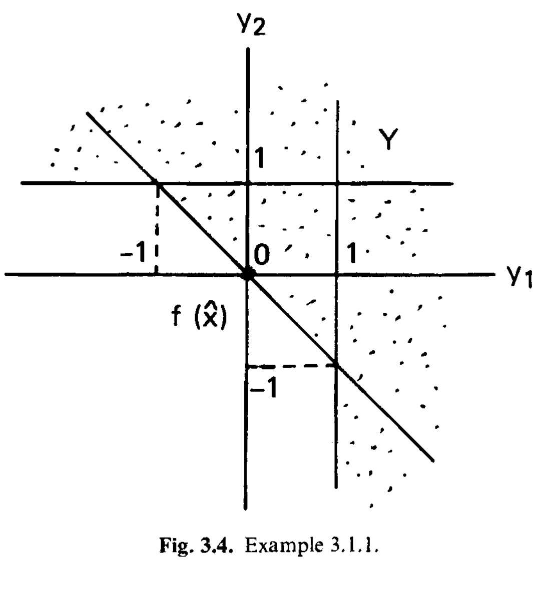

Fig. 3.4. Example 3.1.1.

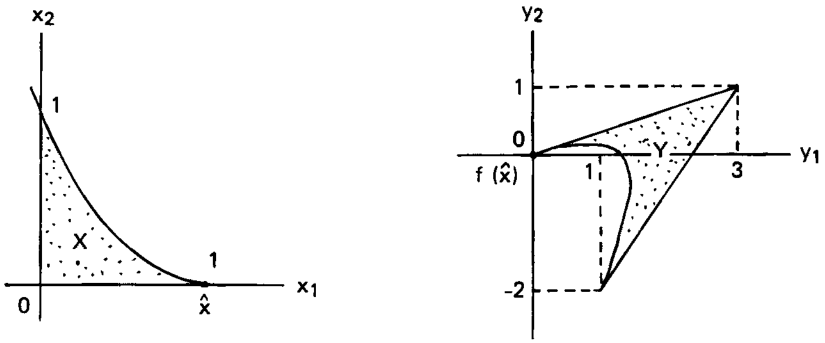

Fig. 3.5. Example 3.1.2.

Then $\hat{x} = (0, 0)$ is a properly efficient solution according to Borwein's definition. However, it is not properly efficient in the sense of Geoffrion or of Benson. Note that X is not convex and that Y is not R^2_+-bounded.

Example 3.1.2 [Geoffrion, but not Kuhn–Tucker—Fig. 3.5][†]

Let

$$X = \{(x_1, x_2) \in R^2 : -x_1 \leqq 0, -x_2 \leqq 0, (x_1 - 1)^3 + x_2 \leqq 0\},$$

$$f_1(x) = -3x_1 - 2x_2 + 3, \qquad f_2(x) = -x_1 - 3x_2 + 1,$$

$$D = R^2_+.$$

Then $\hat{x} = (1, 0)$ (at which the Kuhn–Tucker constraint qualification is not satisfied) is not a Kuhn–Tucker properly efficient solution, although it is properly efficient by Geoffrion's definition.

Example 3.1.3 [Kuhn–Tucker, but not Geoffrion—Fig. 3.6][†]

Let

$$X = \{x \in R : x \geqq 0\},$$

$$f_1(x) = -x^2 \quad \text{be not convex}, \qquad f_2(x) = x^3,$$

$$D = R^2_+.$$

Then $\hat{x} = 0$ is properly efficient in the sense of Kuhn–Tucker but not of Geoffrion.

Any properly efficient solution is, in general, an efficient solution, but not vice versa. (We will later prove that every efficient point is contained in the

[†] Tamura and Arai [T3].

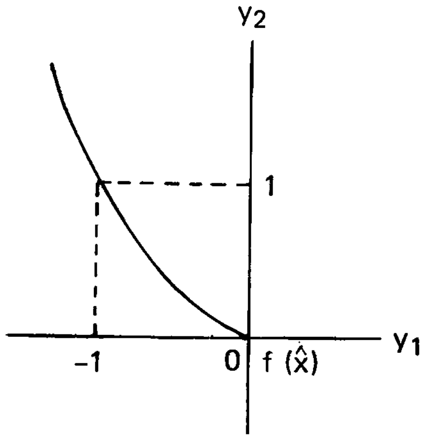

Fig. 3.6. Example 3.1.3.

closure of the set of all properly efficient points in the objective space under some additional conditions. See Theorems 3.2.11 and 3.4.6.) However, they coincide when Y is a polyhedral convex set.

Lemma 3.1.2

Let Y be a polyhedral convex set, i.e.,

$$Y = \{y : \langle b^i, y \rangle \leqq \beta_i,\ i = 1, \ldots, m\},$$

$\hat{y} \in Y$, and

$$I(\hat{y}) = \{i : \langle b^i, \hat{y} \rangle = \beta_i\}.$$

Then

$$T(Y, \hat{y}) = P(Y - \hat{y}) = \{h : \langle b^i, h \rangle \leqq 0 \text{ for } i \in I(\hat{y})\}.$$

Proof The proof of this lemma is easy and well known and so is left to the reader.

Theorem 3.1.7

If Y is a polyhedral convex set and D a pointed closed convex cone, then any efficient solution is properly efficient (in the sense of Benson or Borwein).

Proof Suppose that $\hat{x} \in X$ is not a properly efficient solution. Then we can prove that there exists a nonzero $h \in D$ such that $-h \in T(Y, f(\hat{x}))$, i.e.,

$$\langle b^i, -h \rangle \leqq 0 \qquad \text{for all} \quad i \in I(f(\hat{x})).$$

Then for sufficiently small $\alpha > 0$,

$$\langle b^i, f(\hat{x}) - \alpha h \rangle \leqq \beta_i \qquad \text{for all} \quad i = 1, \ldots, m.$$

Thus, $f(\hat{x}) - (f(\hat{x}) - \alpha h) \in D \backslash \{0\}$ and $f(\hat{x}) - \alpha h \in Y$. Therefore, $f(\hat{x}) \notin \mathscr{E}(Y, D)$.

Corollary 3.1.1

Any efficient solution of the linear multiobjective programming problem

$$\text{minimize} \qquad Cx \qquad \text{subject to} \qquad Ax \leqq b$$

is properly efficient for any pointed closed convex domination cone.

Proof This corollary is immediate from Proposition 2.1.15 and Theorem 3.1.7.

3.2 Existence and External Stability of Efficient Solutions

In this section, the existence of efficient solutions and the external stability of efficient sets are discussed. In an ordinary optimization problem

$$\text{minimize} \quad f(x) \qquad \text{subject to} \quad x \in X \subset R^n,$$

it is known well that the optimal solution $\hat{x}$ exists if the set X is compact and the objective function f is lower semicontinuous. We will extend this result to the case of multiobjective optimization in the first subsection. Some sufficient conditions will be obtained for the efficient set in the objective space to be nonempty. Moreover, a conclusion quite similar to the scalar case is valid for the existence of efficient solutions in the decision space.

A topic dealt with in the second subsection is the external stability of efficient sets. In a scalar optimization problem, every nonoptimal solution is clearly inferior to the optimal solution, if the latter exists. However, in a multiobjective optimization problem, this is not the case, even when the efficient set is not empty. Hence, we will investigate sufficient conditions for this property (i.e. the property that every nonefficient point is dominated by a certain efficient point) that will be called the external stability of the efficient set. Some authors used other terms such as nondominance boundedness (Yu [Y2]) and domination property (Benson [B8], Henig [H9], and Luc [L9]) instead of external stability. Since the external stability is closely related to the nonemptiness of efficient sets, they are discussed together in this section.

3.2.1 Existence of Efficient Solutions

First, the existence of efficient solutions requires acyclicity of the domination structure.

Definition 3.2.1 (*Acyclicity*)

A domination structure D is said to be acyclic if it has no cycle; in other words, if for all $n = 1, 2, \ldots$, it never occurs that

$$y^1 \in y^2 + D(y^2)\backslash\{0\},\ y^2 \in y^3 + D(y^3)\backslash\{0\}, \ldots, y^n \in y^1 + D(y^1)\backslash\{0\}$$

(i.e., $y^1 \prec y^2 \prec \cdots \prec y^n \prec y^1$).

Remark 3.2.1

If a domination structure D is acyclic, it is also asymmetric. Conversely, every transitive and asymmetric domination structure is acyclic.

The following theorem is the existence result with a general domination structure; namely, D is not assumed to be a constant cone.

Theorem 3.2.1†

If a domination structure D on Y is acyclic, the sets $D(y)\backslash\{0\}$ are open and Y is nonempty and compact, then

$$\mathscr{E}(Y, D) \neq \varnothing.$$

Proof If we suppose, to the contrary, that

$$\mathscr{E}(Y, D) = \varnothing.$$

then for any $y \in Y$ there exists $y' \in Y$ such that $y \in y' + D(y')\backslash\{0\}$. Thus

$$Y \subset \bigcup_{y \in Y} (y + D(y)\backslash\{0\}).$$

Thus, the family of the sets $\{y + D(y)\backslash\{0\}\}$ forms an open cover of Y. Since Y is compact, there is a finite subcover $\{y^i + D(y^i)\backslash\{0\}\}$ $(i = 1, \ldots, n)$. Then, for any $i \in \{1, \ldots, n\}$,

$$y^i \in y^j + D(y^j)\backslash\{0\} \qquad \text{for some} \quad j \in \{1, \ldots, n\}.$$

However, this contradicts the assumption that D is acyclic. Hence,

$$\mathscr{E}(Y, D) \neq \varnothing.$$

The following example shows that the condition of openness of $D(y)\backslash\{0\}$ cannot be eliminated in the above theorem.

† Hazen and Morin [H6].

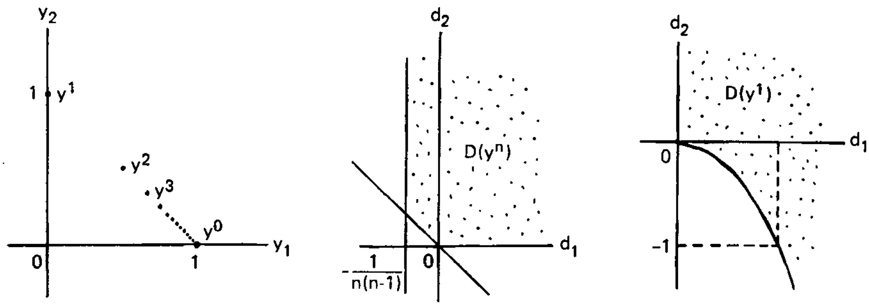

Fig. 3.7. $\mathscr{E}(Y, D) = \varnothing$. (Example 3.2.1.)

Example 3.2.1 (Fig. 3.7)

Let

$$Y = \{(1,0)\} \cup \bigcup_{n=1}^{\infty} \{(1-(1/n), (1/n))\},$$

$$y^0 = (1,0), \qquad y^n = (1-(1/n), (1/n)), \qquad n = 1, 2, \ldots;$$

$$D(y^n) = \{d : d_2 \geqq 0,\ d_1 \geqq -1/n(n-1),\ d_1 + d_2 \geqq 0\}, \qquad n = 2, 3, \ldots,$$

$$D(y^1) = \{d : d_1 \geqq 0,\ d_2 \geqq (-d_1)^2\},$$

$$D(y^0) = R^2_+.$$

Then

$$y^n \in y^{n+1} + D(y^{n+1}), \quad n = 1, 2, \ldots, \qquad \text{and} \qquad y^0 \in y^1 + D(y^1).$$

Therefore, $\mathscr{E}(Y, D) = \varnothing$, although Y is compact and D is acyclic.

When $D(y)$ is a constant convex cone D for all $y \in Y$, the elements of $\mathscr{E}(Y, D)$ are also called cone extreme points of Y (Yu [Y1]). They are characterized as

$$\hat{y} \in \mathscr{E}(Y, D) \qquad \text{if and only if} \quad (\hat{y} - D \cup \{0\}) \cap Y = \{\hat{y}\}.$$

In ordinary scalar minimization problems, $D = R^1_+$, and the existence of the minimal element is guaranteed under the condition that Y is bounded below and $Y + R^1_+$ is closed; that is, under a kind of *semi*compactness condition. A number of authors extended this condition to multiobjective optimization problems and obtained existence theorems for efficient points (e.g., Hartley [H5], Cesari and Suryanarayana [C1–C3], Birtan and Magnanti [B14], Corley [C15], Henig [H7, H8], Borwein [B16]). We first introduce the cone semicompactness condition by Corley [C15], which is a slight generalization of the definition by Wagner [W1].

Definition 3.2.2 (*Cone Semicompactness*)

Let D be a cone and Y be a set in R^p. Y is said to be D-semicompact if every open cover of Y of the form $\{(y^\gamma - \mathrm{cl}\, D)^c : y^\gamma \in Y, \gamma \in \Gamma\}$ has a finite subcover. Here Γ is some index set and the superscript c denotes the complement of a set.

Theorem 3.2.2

If D is an acute convex cone and Y is a nonempty D-semicompact set in R^p, then $\mathscr{E}(Y, D) \neq \varnothing$.

Proof In view of Proposition 3.1.1,

$$\mathscr{E}(Y, \mathrm{cl}\, D) \subset \mathscr{E}(Y, D).$$

Hence, it suffices to prove the case in which D is a pointed closed convex cone. In this case, D defines a partial order $\leqq_D$ on Y as

$$y^1 \leqq_D y^2 \qquad \text{if and only if} \quad y^2 - y^1 \in D.$$

An element in $\mathscr{E}(Y, D)$ is a minimal element of Y with respect to $\leqq_D$. Therefore, we will show that Y is inductively ordered and apply Zorn's lemma to establish the existence of minimal elements. Suppose to the contrary that Y is not inductively ordered. Then, there exists a totally ordered set $\bar{Y} = \{y^\gamma : \gamma \in \Gamma\}$ in Y which has no lower bound in Y. Thus,

$$\bigcap_{\gamma \in \Gamma} [(y^\gamma - D) \cap Y] = \varnothing,$$

otherwise any element of this intersection is a lower bound of $\bar{Y}$ in Y. It now follows that for any $y \in Y$, there exists $y^\gamma \in \bar{Y}$ such that $y \notin y^\gamma - D$. Since $y^\gamma - D$ is closed, the family $\{(y^\gamma - D)^c : \gamma \in \Gamma\}$ forms an open cover of Y. Moreover, $y^\gamma - D \subset y^{\gamma'} - D$ if and only if $y^\gamma \leqq_D y^{\gamma'}$, and so they are totally ordered by inclusion. Since Y is D-semicompact, the cover has a finite subcover, and hence there exists a single $y^{\bar{\gamma}} \in Y$ such that

$$Y \subset (y^{\bar{\gamma}} - D)^c.$$

However, this contradicts the fact $y^{\bar{\gamma}} \in Y$. Therefore, Y is inductively ordered by $\leqq_D$ and $\mathscr{E}(Y, D) \neq \varnothing$ by Zorn's lemma.

It may be troublesome to check the cone semicompactness condition. Hartley [H5] used the slightly stronger cone compactness condition.

Definition 3.2.3 (*Cone Compactness*)

Let D be a cone in R^p. A set $Y \subset R^p$ is said to be D-compact if, for any $y \in Y$, the set $(y - \mathrm{cl}\, D) \cap Y$ is compact.

Proposition 3.2.1

If Y is D-compact, then Y is D-semicompact.

Proof Let a family of sets

$$\{(y^\gamma - \mathrm{cl}\, D)^c : y^\gamma \in Y, \gamma \in \Gamma\}$$

be an open cover of Y. For an arbitrary $y^{\bar{\gamma}} \in Y$, the subfamily

$$\{(y^\gamma - \mathrm{cl}\, D)^c : y^\gamma \in Y, \gamma \in \Gamma, \gamma \neq \bar{\gamma}\}$$

forms an open cover of $(y^{\bar{\gamma}} - \mathrm{cl}\, D) \cap Y$. Since from the definition of D-compactness, $(y^{\bar{\gamma}} - \mathrm{cl}\, D) \cap Y$ is compact, this subfamily has a finite subcover, which together with $(y^{\bar{\gamma}} - \mathrm{cl}\, D)^c$ constitutes a finite subcover of Y. The proof is completed.

Remark 3.2.2

A compact set is D-compact and so D-semicompact. However, a D-compact set is not necessarily compact. For example, take $D = R^2_+$ and $Y = \{y \in R^2 : y_1 + y_2 \geqq 0\}$ (Hartley [H5]).

Theorem 3.2.3

Let D be an acute convex cone in R^p. If $Y \subset R^p$ is nonempty and D-compact, then $\mathscr{E}(Y, D) \neq \varnothing$.

Proof This result is immediate from Theorem 3.2.2 and Proposition 3.2.1. although the direct proof can be seen in Hartley [H5].

Some authors used conditions in terms of recession cones of feasible sets (Bitran and Magnanti [B14], Henig [H7]). Here we adopt an extended definition of the recession cone by Henig, since Y is not necessarily convex. As in Theorem 3.1.3, for $Y \supset R^p$, its extended recession cone Y^+ is defined by

$$Y^+ = \{y' : \text{there exist sequences } \{\alpha_k\} \subset R \text{ and } \{y^k\} \subset Y \text{ such that } \alpha_k > 0,\ \alpha_k \to 0, \text{ and } \alpha_k y^k \to y'\}.$$

Remark 3.2.3

It is clear that Y^+ is a closed cone and that $0^+ Y \subset Y^+$. On the other hand, $0^+ Y$ is not necessarily closed even if Y is convex (Rockafellar [R7], p. 63).

Lemma 3.2.1[†]

Let Y be a nonempty set. Then

(i) Y is bounded if and only if $Y^+ = \{0\}$,
(ii) if Y is closed and convex, $Y^+ = 0^+Y$.

Lemma 3.2.2[‡]

Let $\{Y_i\}_{i \in I}$ be an arbitrary collection of nonempty closed sets. Then

$$\left(\bigcap_{i \in I} Y_i\right)^+ \subset \bigcap_{i \in I} Y_i^+.$$

Lemma 3.2.3[§]

Let Y_1 and Y_2 be nonempty closed sets in R^p. If

$$Y_1^+ \cap (-Y_2^+) = \{0\},$$

then $Y_1 + Y_2$ is a closed set.

Proof Let $\{y_1^k + y_2^k\}$ be a sequence such that $y_1^k \in Y_1$, and $y_2^k \in Y_2$ and $y_1^k + y_2^k \to y$. If $\{y_1^k\}$ has no convergent subsequence, $\|y_1^k\| \to \infty$ as $k \to \infty$. Hence $(y_1^k + y_2^k)/\|y_1^k\| \to 0$. We may assume without loss of generality that $y_1^k/\|y_1^k\| \to y_1$ with $\|y_1\| = 1$. Clearly $y_1 \in Y_1^+$. Moreover $y_2^k/\|y_1^k\| \to -y_1$, which is in Y_2^+ from the definition. Thus $y_1 \in Y_1^+ \cap (-Y_2^+)$, which contradicts the hypothesis. Hence, suppose that $\{y_1^k\}$ converges to some $\bar{y}_1$. Then $\{y_2^k\}$ clearly converges to $y - \bar{y}_1$. Since Y_1 and Y_2 are closed, $\bar{y}_1 \in Y_1$, $y - \bar{y}_1 \in Y_2$, and so $y = \bar{y}_1 + (y - \bar{y}_1) \in Y_1 + Y_2$, as was to be proved.

Definition 3.2.4 (*Cone Closedness and Cone Boundedness*)

Let Y be a nonempty set in R^p, and let D be a cone in R^p. Then Y is said to be

(i) D-closed if $Y + \operatorname{cl} D$ is closed, and
(ii) D-bounded if $Y^+ \cap (-\operatorname{cl} D) = \{0\}$.

Remark 3.2.4

Let D be a pointed cone. If there exists a point $y^0 \in R^p$ such that $Y \subset y^0 + \operatorname{cl} D$, then it is clear that Y is D-bounded. This hypothesis was used as the definition of D-boundedness in Tanino and Sawaragi [T11].

[†] Henig [H7].
[‡] Henig [H7] cf. Proposition 2.1.9.
[§] cf. Proposition 2.1.11.

Lemma 3.2.4

Let D be a pointed, closed, convex cone and Y be a nonempty set in R^p. Then, $Y^+ \cap (-D) = \{0\}$ if and only if $(Y + D)^+ \cap (-D) = \{0\}$.

Proof Since $(Y + D)^+ \supset Y^+$, it is clear that $Y^+ \cap (-D) = \{0\}$ when $(Y + D)^+ \cap (-D) = \{0\}$. To prove the only if part, suppose that $(Y + D)^+ \cap (-D) \neq \{0\}$. Then there exist sequences $\{\alpha_k\} \subset R$ and $\{y^k + d^k\}$ such that $\alpha_k > 0$, $\alpha_k \to 0$, $y^k \in Y$, $d^k \in D$, $\alpha_k(y^k + d^k) \to -d \neq 0 \in -D$. There are two cases: (i) $\{\alpha_k d^k\}$ has a convergent subsequence, and (ii) $\{\alpha_k d^k\}$ has no convergent subsequence.

Case (i) By taking a subsequence, if necessary, we may assume without loss of generality that $\alpha_k d^k \to d'$. Since D is closed, $d' \in D$. Therefore, $\alpha_k y^k \to -d - d'$, which is a nonzero vector in $Y^+ \cap (-D)$ since D is a pointed convex cone. Hence, $Y^+ \cap (-D) \neq \{0\}$.

Case (ii) In this case, $\{\alpha_k d^k\}$ is unbounded, so $\{\alpha_k d^k\}^+ \neq \{0\}$ from Lemma 3.2.1(i). Namely, we may assume by taking a subsequence of $\{\alpha_k d^k\}$ that there exists another sequence $\{\beta_k\}$ such that

$$\beta_k > 0, \qquad \beta_k \to 0, \qquad \text{and} \qquad \beta_k(\alpha_k d^k) \to \bar{d} \neq 0.$$

Since D is closed, $\bar{d} \in D$. Then

$$(\beta_k \alpha_k) y^k \to -\bar{d}, \qquad \text{with} \quad \beta_k \alpha_k > 0, \quad \beta_k \alpha_k \to 0.$$

In fact,

$$\begin{aligned} \|(\beta_k \alpha_k) y^k + \bar{d}\| &= \|\beta_k(\alpha_k(y^k + d^k) + d) - (\beta_k \alpha_k d^k - \bar{d}) - \beta_k d\| \\ &\leqq \beta_k \|\alpha_k(y^k + d^k) + d\| + \|\beta_k \alpha_k d^k - \bar{d}\| + \beta_k \|d\| \\ &\to 0 \qquad \text{as} \quad k \to \infty. \end{aligned}$$

Thus, we have proved that $Y^+ \cap (-D) \neq \{0\}$.

Lemma 3.2.5

Let D be a convex cone and Y be a set in R^p. If $Y + \operatorname{cl} D$ is D-semicompact, then Y is also D-semicompact.

Proof Let $\{(y^\gamma - \operatorname{cl} D)^c : y^\gamma \in Y, \gamma \in \Gamma\}$ be an open cover of Y. For an arbitrary $y \in Y$, let $y \in (y^{\bar{\gamma}} - \operatorname{cl} D)^c$ with $y^{\bar{\gamma}} \in Y$. Since $\operatorname{cl} D$ is a convex cone.

$$(y^{\bar{\gamma}} - \operatorname{cl} D)^c \supset y + \operatorname{cl} D.$$

In fact, if $y + d \in y^{\bar{\gamma}} - \operatorname{cl} D$ for some $d \in \operatorname{cl} D$, $y \in y^{\bar{\gamma}} - \operatorname{cl} D$, which is a contradiction. Hence $\{(y^\gamma - \operatorname{cl} D)^c : y^\gamma \in Y, \gamma \in \Gamma\}$ is also an open cover of $Y + \operatorname{cl} D$. Since $Y + \operatorname{cl} D$ is D-semicompact, this cover has a finite subcover, which is, of course, a subcover of Y. Hence Y is D-semicompact.

Proposition 3.2.2

Let D be an acute convex cone in R^p and Y be a nonempty set in R^p. If Y is D-closed and D-bounded, then $Y + \mathrm{cl}\, D$ is D-compact, and Y is D-semicompact.

Proof Let $y \in Y + \mathrm{cl}\, D$. Since $y - \mathrm{cl}\, D$ and $Y + \mathrm{cl}\, D$ are both nonempty closed sets, $(y - \mathrm{cl}\, D) \cap (Y + \mathrm{cl}\, D)$ is closed. Therefore, from Lemmas 3.2.2 and 3.2.4

$$\begin{aligned} ((y - \mathrm{cl}\, D) \cap (Y + \mathrm{cl}\, D))^+ &\subset (y - \mathrm{cl}\, D)^+ \cap (Y + \mathrm{cl}\, D)^+ \\ &= (-\mathrm{cl}\, D) \cap (Y + \mathrm{cl}\, D)^+ \\ &= \{0\} \end{aligned}$$

since Y is D-bounded. Thus, in view of Lemma 3.2.1(i), $(y - \mathrm{cl}\, D) \cap (Y + \mathrm{cl}\, D)$ is bounded. Therefore, $Y + \mathrm{cl}\, D$ is D-compact, and so $Y + \mathrm{cl}\, D$ is D-semicompact from Proposition 3.2.1. Hence, from Lemma 3.2.5, Y is D-semicompact.

Theorem 3.2.4

Let D be an acute convex cone in R^p and Y be a nonempty, D-closed, D-bounded set in R^p. Then $\mathscr{E}(Y, D) \neq \varnothing$.

Proof This theorem can be proved easily by Theorem 3.2.2 and Proposition 3.2.2.

We now give some examples of D-semicompact sets with $D = R^2_+$.

Example 3.2.2

(i) (Fig. 3.8) Let

$$Y = \{(y_1, y_2) : (y_1)^2 + (y_2)^2 \leqq 1,\ y_1 < 0,\ y_2 < 0\}.$$

Then Y is R^2_+-compact, R^2_+- bounded, but not R^2_+-closed, and

$$\mathscr{E}(Y, R^2_+) = \{(y_1, y_2) : (y_1)^2 + (y_2)^2 = 1,\ y_1 < 0,\ y_2 < 0\}.$$

(ii) (Fig. 3.9) Let

$$Y = \{(y_1, y_2) : y_1 y_2 = 1,\ y_1 < 0\}.$$

Then Y is R^2_+-compact, R^2_+-closed, but not R^2_+-bounded. In fact $\{(-1, 0), (0, -1)\} \subset \mathrm{Y}^+ \cap (-\mathrm{R}^2_+)$. In this case, $\mathscr{E}(Y, R^2_+) = Y$.

(iii) (Fig. 3.10) Let

$$\begin{aligned} Y = &\{(y_1, y_2) : (y_1)^2 + (y_2)^2 < 1\} \\ &\cup \{(y_1, y_2) : (y_1)^2 + (y_2)^2 = 1,\ y_1 \leqq 0,\ y_2 \leqq 0\}. \end{aligned}$$

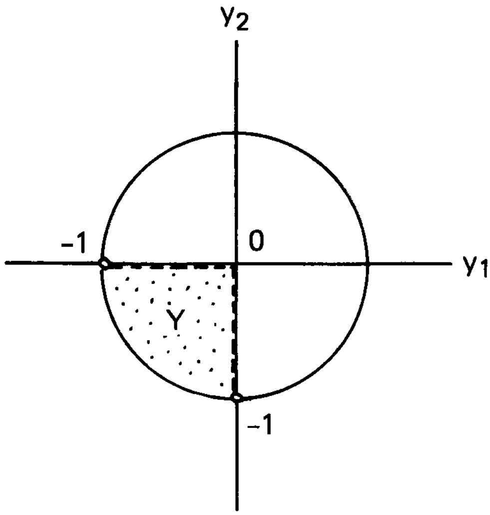

Fig. 3.8. R^2_+-compact, R^2_+-bounded but not R^2_+-closed set. (Example 3.2.2(i).)

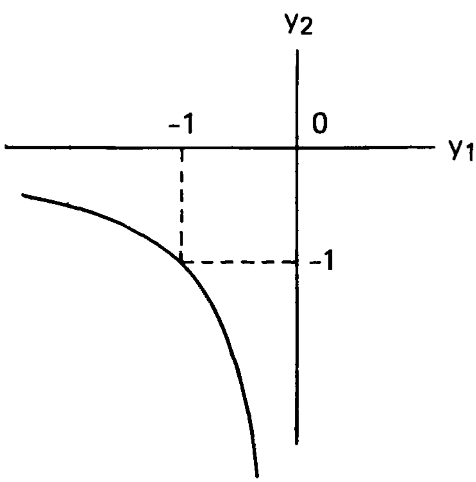

Fig. 3.9. R^2_+-compact, R^2_+-closed but not R^2_+-bounded set. (Example 3.2.2(ii).)

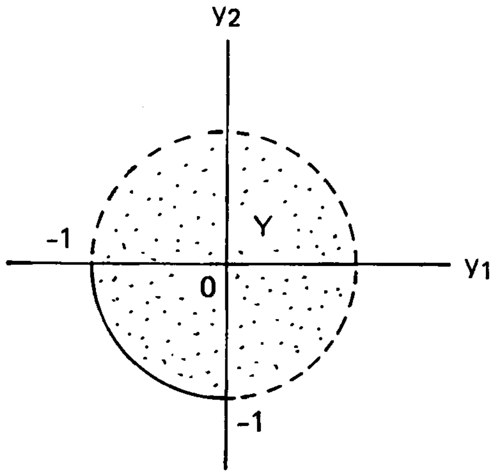

Fig. 3.10. R^2_+-closed, R^2_+-bounded but not R^2_+-compact set. (Example 3.2.2(iii).)

Then Y is R^2_+-closed and R^2_+-bounded, but not R^2_+-compact. In this case,

$$\mathscr{E}(Y, R^2_+) = \{(y_1, y_2) : (y_1)^2 + (y_2)^2 = 1,\ y_1 \leqq 0,\ y_2 \leqq 0\}.$$

As can be seen from these examples, the concepts introduced by the various authors are different, in general. However, they coincide in the convex case, as shown by the following proposition.

Proposition 3.2.3

Let D be a pointed closed convex cone and Y be a nonempty, closed convex set in R^p. Then the following are equivalent:

(i) $0^+Y \cap (-D) = \{0\}$;
(ii) Y is D-closed and D-bounded;
(iii) Y is D-compact;
(iv) $(0^+Y)^\circ \cap D^{s\circ} \neq \varnothing$;
(v) Y is D-semicompact.

Proof The equivalence of (i) and (iv) is immediate from Proposition 2.1.8. From Lemma 3.2.1(ii), condition (i) is equivalent to the D-boundedness of Y. In view of Proposition 2.1.11 or Lemma 3.2.3, $Y + D$ is closed if

$$0^+Y \cap (-0^+D) = 0^+Y \cap (-D) = \{0\},$$

i.e., if Y is D-bounded. Thus we have proved the equivalence of (i) and (ii). The proof of (ii) $\Rightarrow$ (iii) is essentially the same as the proof of Proposition

3.2.2. Next, we prove (iii) $\Rightarrow$ (i). Take any $y \in Y$. Then $(y - D) \cap Y$ is compact, and from Proposition 2.1.10,

$$0^+((y - D) \cap Y) = \{0\}.$$

From Proposition 2.1.9, this implies that

$$0^+ Y \cap (-D) = \{0\}.$$

Finally, (iii) $\Rightarrow$ (v) (Proposition 3.2.1) or (ii) $\Rightarrow$ (v) (Proposition 3.2.2), (v) $\Rightarrow$ $\mathscr{E}(Y, D) \neq \varnothing$, and we later prove that $\mathscr{E}(Y, D) \neq \varnothing \Rightarrow$ (i) in Theorem 3.2.5.

In the analysis of efficient sets, the D-compactness condition was used by Hartley [H5] and Naccache [N1], while Henig [H7] used a condition similar to D-boundedness and D-closedness. Borwein [B16] and Bitran and Magnanti [B14] used conditions (i) or (iv) in the above proposition for closed convex Y. Corley [C15] used the D-semicompactness condition. However, as can be seen from the above propositions, they are closely related to one another.

Now we can prove that if Y is closed convex, D-compactness is also necessary for $\mathscr{E}(Y, D)$ to be nonempty. (Sufficiency has been already established in Theorem 3.2.2 or 3.2.3.)

Theorem 3.2.5

Let D be a pointed closed convex cone and Y be a nonempty closed convex set. Then $\mathscr{E}(Y, D) \neq \varnothing$ if and only if any of the conditions in Proposition 3.2.3 is satisfied.

Proof We need only to prove that $0^+ Y \cap (-D) = \{0\}$ if $\mathscr{E}(Y, D) \neq \varnothing$. Suppose, to the contrary, that some nonzero $-d$ belongs to the set $0^+ Y \cap (-D)$. Then, for each $y \in Y$, $y - d \in Y$ and so $y \notin \mathscr{E}(Y, D)$. Hence, $\mathscr{E}(Y, D) = \varnothing$. This completes the proof of the theorem.

Remark 3.2.5

In Theorem 3.2.5, the condition that Y be closed and convex may be replaced by the one that Y be D-closed and D-convex.

When Y is a polyhedral convex set, the existence condition of efficient solutions (which are also properly efficient solutions by Theorem 3.1.7) can be obtained as follows.

Theorem 3.2.6

Let Y be a nonempty, polyhedral convex set and D be a pointed, closed convex cone in R^p. Then $\mathscr{E}(Y, D)$ is nonempty if and only if

(i) when Y is expressed as

$$Y = \{y : \langle b^i, y\rangle \leqq \beta_i,\ i = 1, \ldots, m\},$$

there is no $d \neq 0 \in D$ such that $\langle b^i, d\rangle \geqq 0$ or, equivalently, the finitely generated cone with the generators $\{b^1, \ldots, b^m\}$ and D^{so} have a common nonzero vector.

(ii) when Y is expressed as

$$Y = \left\{y : y = \sum_{i=1}^{m} \alpha_i a^i,\ \alpha_i \geqq 0\ (i = 1, \ldots, m),\ \sum_{i=1}^{k} \alpha_i = 1\right\},$$

$a^i \notin -D$ for any $i = k + 1, \ldots, m$.

Proof The results follow immediately from Theorem 3.2.5 and Propositions 2.1.8, 2.1.13, and 2.1.14.

We conclude this subsection with an existence theorem of efficient solutions in the decision space, which is a quite natural extension of the scalar case. For this purpose, we introduce an extended semicontinuity concept of functions according to Corley [C15].

Definition 3.2.5 (*Cone Semicontinuity*)

Let D be a cone in R^p. A function $f: R^n \to R^p$ is said to be D-semicontinuous if

$$f^{-1}(y - \mathrm{cl}\, D) = \{x \in R^n : y - f(x) \in \mathrm{cl}\, D\}$$

is closed for each $y \in R^p$.

Remark 3.2.6

If $p = 1$ and $D = R_+$ (resp. R_-), D-semicontinuity of f is nothing but the lower (resp. upper) semicontinuity. Moreover $(f_1, \ldots, f_p)$: $R^n \to R^p$ is R^p_+-semicontinuous when and only when each f_i $(i = 1, \ldots, p)$ is lower semicontinuous.

Lemma 3.2.6

Let X be a nonempty, compact set in R^n, D be a cone in R^p, and f be a D-continuous function from R^n into R^p. Then the set $Y = f(X)$ is D-semicompact.

Proof Let $\{(y^\gamma - \mathrm{cl}\, D)^c : y^\gamma \in Y, \gamma \in \Gamma\}$ be an open cover of Y. Since f is D-semicontinuous, $\{f^{-1}((y^\gamma - \mathrm{cl}\, D)^c) : y^\gamma \in Y, \gamma \in \Gamma\}$ forms an open cover of X. Since X is compact, it has a finite subcover whose image by f clearly constitutes a finite subcover of Y. Thus Y is D-semicompact.

Theorem 3.2.7

Let X be a nonempty compact set in R^n, D be a cone in R^p, and f be a D-semicontinuous function from R^n into R^p. Then there exists an efficient solution.

Proof Immediate from Lemma 3.2.6 and Theorem 3.2.2.

Corollary 3.2.1

Let $f = (f_1, \ldots, f_p)$ be a vector valued function from R^n into R^p. Let X be a nonempty compact set in R^n and each f_i $(i = 1, \ldots, p)$ be lower semicontinuous on X. Then the multiobjective optimization problem

$$\text{minimize} \quad f(x) = (f_1(x), \ldots, f_p(x)) \qquad \text{subject to} \quad x \in X$$

has a Pareto optimal solution.

Proof Immediate from Remark 3.2.6 and Theorem 3.2.7.

3.2.2 External Stability of Efficient Sets

In this subsection we introduce a new concept, *external stability* of the efficient set. In section 3.1 we defined the efficient set, which is the set of all the nondominated points in the objective space. Each point outside the efficient set is, therefore, dominated by some other point in the feasible set. However, is it dominated by a point in the efficient set? If this is the case, the efficient set is said to be externally stable.

Definition 3.2.6 (*External Stability*)

Let Y be a set of feasible points in R^p, S be a subset of Y, and D be a domination structure on Y. S is said to be externally stable if, for each $y \in Y \backslash S$, there exists some $\hat{y} \in S$ such that $y \in \hat{y} + D(\hat{y})$.

Remark 3.2.7

Since each $D(y)$ is assumed to contain the zero vector, the external stability condition can be rewritten as follows: For each $y \in Y$, there exists $\hat{y} \in S$ such that $y \in \hat{y} + D(\hat{y})$. Hence, if $D(y) \equiv D$ (constant) for all y, this can be also rewritten as $Y \subset S + D$.

External stability has been also studied in graph theory and in statistical decision theory, where it is often called completeness. In graph theory, another type of stability concept—internal stability—is known.

Definition 3.2.7 (*Internal Stability*)

A set $S \subset Y$ be said to be internally stable if $y \notin y' + D(y')\backslash\{0\}$ (i.e., $y \not\prec y'$) whenever $y, y' \in S$.

It is clear that the efficient set of Y is internally stable. Berge [B12] calls external and internal stability *absorbency* and *stability*, respectively. The reader should not confuse these stability concepts with stability with respect to parameter perturbation, which will be fully discussed in Chapter 4.

Definition 3.2.8 (*Kernel*)

Let Y be a set in R^p and D be a domination structure on Y. A subset of Y is called a kernel of Y with respect to D, denoted by $\mathscr{K}(Y, D)$, if it is both externally and internally stable.

Remark 3.2.8

If D is transitive and asymmetric, and a kernel $\mathscr{K}(Y, D)$ exists, then it is unique. In fact, if we assume the contrary, that there are two different kernels $\mathscr{K}(Y, D)$ and $\mathscr{K}'(Y, D)$, then we have $y \in \mathscr{K}(Y, D)$ but $y \notin \mathscr{K}'(Y, D)$ for some $y \in Y$. From $y \notin \mathscr{K}'(Y, D)$, there exists $y' \in \mathscr{K}'(Y, D)$ such that $y \in y' + D(y')\backslash\{0\}$. Since $\mathscr{K}(Y, D)$ is internally stable and $y \in \mathscr{K}(Y, D)$, *then* $y' \notin \mathscr{K}(Y, D)$ and so there exists $y'' \in \mathscr{K}(Y, D)$ such that $y' \in y'' + D(y'')\backslash\{0\}$. Since D is transitive and asymmetric, $y \in y'' + D(y'')\backslash\{0\}$, which leads to a contradiction to the internal stability of $\mathscr{K}(Y, D)$.

Proposition 3.2.4

Suppose that D is transitive and asymmetric. If $\mathscr{K}(Y, D)$ exists, then

$$\mathscr{K}(Y, D) = \mathscr{E}(Y, D).$$

Proof Let $\hat{y} \in \mathscr{K}(Y, D)\backslash\mathscr{E}(Y, D)$. Then these exists $y \in Y$ such that $\hat{y} \in y + D(y)\backslash\{0\}$. From the external stability of $\mathscr{K}(Y, D)$, there exists $y' \in \mathscr{K}(Y, D)$ such that $y \in y' + D(y')$. Since D is transitive and asymmetric, $\hat{y} \in y' + D(y')\backslash\{0\}$, which contradicts the internal stability of $\mathscr{K}(Y, D)$.

Conversely let $\hat{y} \in \mathscr{E}(Y, D)\backslash\mathscr{K}(Y, D)$. Then, from the external stability of $\mathscr{K}(Y, D)$, there exists $y \in \mathscr{K}(Y, D) \subset Y$ such that $\hat{y} \in y + D(y)\backslash\{0\}$. However, this is a contradiction.

Proposition 3.2.5

Let Y be a nonempty set and D be a transitive asymmetric domination structure on Y. Then $\mathscr{K}(Y, D)$ exists if and only if $\mathscr{E}(Y, D)$ is externally stable.

Proof In view of Proposition 3.2.4, $\mathscr{K}(Y, D) = \mathscr{E}(Y, D)$ if $\mathscr{K}(Y, D)$ exists. Hence $\mathscr{E}(Y, D)$ is externally stable. Conversely, if $\mathscr{E}(Y, D)$ is externally stable, then it becomes a kernel because it is internally stable.

As for some properties concerning the kernel, the reader may refer to White [W9, Chapter 2]. We consider in the remaining part of this section conditions for the external stability of the efficient set.

Theorem 3.2.8

Let Y be a nonempty compact set. Suppose that a domination structure D is transitive and upper semicontinuous (as a point-to-set map) on Y. Moreover, for each compact subset Y' of Y, $\mathscr{E}(Y', D)$ is assumed to be nonempty. Then $\mathscr{E}(Y, D)$ is externally stable and so is the kernel of Y.

Proof Let y be an arbitrary point in Y, and define a set

$$Y' = \{y' \in Y: y \in y' + D(y')\}.$$

Namely, the set Y' consists of y and all points in Y that dominate y. We must show that $Y' \cap \mathscr{E}(Y, D) \neq \emptyset$. It suffices for this to prove that (i) $\mathscr{E}(Y', D) \neq \emptyset$ and that (ii) $\mathscr{E}(Y', D) \subset \mathscr{E}(Y, D)$.

(*i*) We prove the compactness of Y', since it implies that $\mathscr{E}(Y', D) \neq \emptyset$ from the assumption. Since $Y' \subset Y$ and Y is bounded, Y' is also bounded. To prove the closedness of Y', let $\{y^k\} \subset Y'$ and $y^k \to y'$. Then

$$y \in y^k + D(y^k),$$

i.e.,

$$y - y^k \in D(y^k) \qquad \text{and} \qquad y - y^k \to y - y'.$$

Since D is upper semicontinuous, $y - y' \in D(y')$, whence $y' \in Y'$.

(ii) A vector $\hat{y} \in R^p$ is supposed not to be contained in $\mathscr{E}(Y, D)$. We suppose $\hat{y} \in Y'$, since otherwise it is clear that $\hat{y} \notin \mathscr{E}(Y', D)$. Then $\hat{y} \in Y$, and there exists $y'' \in Y$ such that $\hat{y} \in y'' + D(y'')\backslash\{0\}$. Since D is transitive and $\hat{y} \in Y'$, $y \in y'' + D(y'')$, which implies that $y'' \in Y'$. Hence, $\hat{y} \notin \mathscr{E}(Y', D)$, as was to be proved. This completes the proof of the theorem.

Theorem 3.2.9

Let $D(y) = D$ be a pointed closed convex cone and Y be a nonempty D-compact set. Then $\mathscr{E}(Y, D)$ is externally stable; that is, $Y \subset \mathscr{E}(Y, D) + D$.

Proof The proof is analogous to that of Theorem 3.2.8. In this case,

$$Y' = (y - D) \cap Y,$$

which is D-compact from the D-compactness of Y. Therefore, in view of Theorem 3.2.3, $\mathscr{E}(Y', D) \neq \varnothing$. On the other hand, since D is a pointed convex cone, we can immediately establish that $\mathscr{E}(Y', D) \subset \mathscr{E}(Y, D)$. Thus, $\mathscr{E}(Y, D)$ is externally stable.

Theorem 3.2.10

Let D be a pointed closed convex cone in R^p and Y be a nonempty D-bounded, D-closed set in R^p. Then $\mathscr{E}(Y, D)$ is externally stable, i.e., $Y \subset \mathscr{E}(Y, D) + D$.

Proof Take an arbitrary $y \in Y$ and put

$$Y' = (y - D) \cap (Y + D).$$

Clearly Y' is closed and nonempty since Y is D-closed. Since Y is D-bounded,

$$(Y')^+ \subset (y - D)^+ \cap (Y + D)^+ = (-D) \cap (Y + D)^+ = \{0\}.$$

Thus Y' is bounded and, hence, compact. Then, in view of Remark 3.2.2 and Theorem 3.2.3, $\mathscr{E}(Y', D) \neq \varnothing$. We can also prove that $\mathscr{E}(Y', D) \subset \mathscr{E}(Y, D)$ as before.

Theorem 3.2.10 enables us to obtain the following relationship between the efficient set and the properly efficient set.

Theorem 3.2.11†

Let D be a pointed closed convex cone in R^p and Y be a D-bounded, D-closed set in R^p. Then

$$\mathscr{E}(Y, D) \subset \operatorname{cl} \mathscr{P}(Y, D),$$

where $\mathscr{P}(Y, D)$ denotes the set of all properly efficient points of Y in the sense of Benson.

Proof If Y is empty, the result is trivial; we assume, therefore, that Y is nonempty. Let $\hat{y} \in \mathscr{E}(Y, D)$. In view of the lemma by Henig ([H7], Lemma 3.4), there exists a sequence $\{D_k\}$ of pointed, closed, convex cones such that $\bigcap_k D_k = D$ and

$$D_k \supset D_{k+1}, \quad D \subset \operatorname{int} D_k \cup \{0\}, \quad (Y + D)^+ \cap (-D_k) = \{0\} \quad \text{for every} \quad k.$$

† cf. Theorem 3.4.6.

Since $Y + D$ is closed and $(Y + D)^+ \cap (-D_k) = \{0\}$, then

$$(Y + D) + D_k = Y + D_k$$

is closed by Lemma 3.2.3. Moreover, $Y^+ \cap (-D_k) = \{0\}$. Hence, by Theorem 3.2.10, every $\mathscr{E}(Y, D_k)$ is externally stable. Thus, there exists a sequence $\{y^k\}$ such that

$$y^k \in \mathscr{E}(Y, D_k) \cap (\hat{y} - D_k).$$

Then y^k is a global properly efficient point of Y in the sense of Henig (Definition 3.1.8), which implies that $y^k \in \mathscr{P}(Y, D)$ from Theorem 3.1.2. Hence it suffices for the theorem to prove that a subsequence of $\{y^k\}$ converges to $\hat{y}$. Since $D_k \supset D_{k+1}$,

$$y^k \in (Y + D) \cap (\hat{y} - D_1) \qquad \text{for every} \quad k.$$

Both $Y + D$ and $\hat{y} - D_1$ are closed, and by Lemma 3.2.2,

$$\begin{aligned}((Y + D) \cap (\hat{y} - D_1))^+ &\subset (Y + D)^+ \cap (\hat{y} - D_1)^+ \\ &= (Y + D)^+ \cap (-D_1) = \{0\}.\end{aligned}$$

Hence, by Lemma 3.2.1(i), $(Y + D) \cap (\hat{y} - D_1)$ is compact. Therefore, we may assume without loss of generality that $\{y^k\}$ converges to some $\bar{y} \in (Y + D) \cap (\hat{y} - D_1)$. Since $\hat{y} - y^k \in D_k$ and $D_k \to D$, then $\hat{y} - \bar{y} \in D$. This implies that $\hat{y} = \bar{y}$, since $\hat{y} \in \mathscr{E}(Y, D) = \mathscr{E}(Y + D, D)$ (Proposition 3.1.2) and $\bar{y} \in Y + D$. Thus, we have proved that $y^k \to \hat{y}$, and the proof is completed.

In this theorem, the indispensability of the D-boundedness of Y can be understood by Example 3.1.1, where

$$Y = \{(y_1, y_2) : y_1 + y_2 \geqq 0\} \cup \{(y_1, y_2) : y_1 \geqq 1\} \cup \{(y_1, y_2) : y_2 \geqq 1\},$$

and $D = R^2_+$. Then Y is D-closed but not D-bounded. In this case,

$$\mathscr{E}(Y, D) = \{(y_1, y_2) : y_1 + y_2 = 0, -1 < y_1 < 1\},$$

while $\mathscr{P}(Y, D) = \varnothing$. The indispensability of the D-closedness of Y is illustrated in the following example.

Example 3.2.3 (Fig. 3.11)

Let $D = R^2_+$ and

$$Y = \{(y_1, y_2) : y_1 = (y_2)^2,\ y_2 < 0\} \cup \{(0, 1)\}.$$

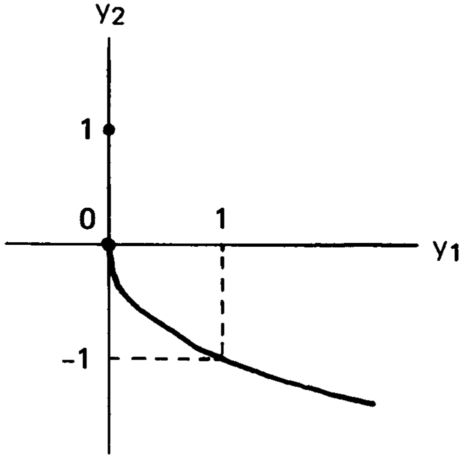

Fig. 3.11. $\mathscr{E}(Y, D) \not\subset \mathrm{cl}\,\mathscr{P}(Y, D)$. (Example 3.2.3.)

Then Y is D-bounded but not D-closed. In this case,

$$\mathscr{E}(Y, D) = \{(y_1, y_2) : y_1 = (y_2)^2,\ y_2 < 0\} \cup \{(0, 1)\},$$

while

$$\mathscr{P}(Y, D) = \{(y_1, y_2) : y_1 = (y_2)^2,\ y_2 < 0\},$$

$$\mathrm{cl}\,\mathscr{P}(Y, D) = \{(y_1, y_2) : y_1 = (y_2)^2,\ y_2 \leqq 0\}.$$

The converse of Theorem 3.2.11 (i.e., $\mathrm{cl}\,\mathscr{P}(Y, D) \subset \mathscr{E}(Y, D)$) does not always hold, as in the the following counterexample (see also Example 3.4.2).

Example 3.2.4 (Fig. 3.12)

Let

$$Y = \{(y_1, y_2) : y_2 = (y_1)^2,\ y_1 \leqq 0\} \cup \{(y_1, y_2) : y_1 = 0, -1 \leqq y_2 \leqq 0\},$$
$$D = R^2_+.$$

Then Y is D-closed, D-bounded, and

$$\mathscr{P}(Y, D) = \mathscr{E}(Y, D) = \{(y_1, y_2) : y_2 = (y_1)^2,\ y_1 < 0\} \cup \{(0, -1)\}.$$

Thus,

$$(0, 0) \in \mathrm{cl}\,\mathscr{P}(Y, D) \qquad \text{but} \qquad (0, 0) \notin \mathscr{E}(Y, D).$$

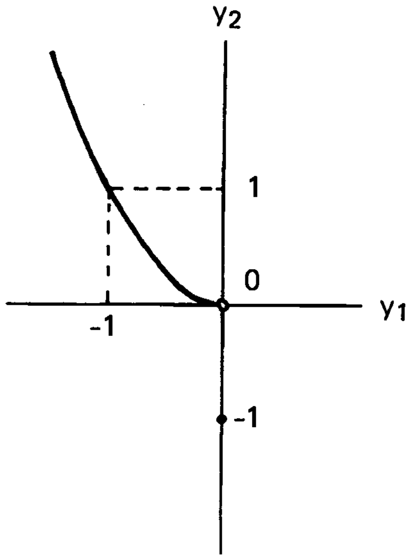

Fig. 3.12. cl $\mathscr{P}(Y, D) \not\subset \mathscr{E}(Y, D)$. (Example 3.2.4.)

Corollary 3.2.2 (cf. Theorem 3.4.6)

If D is a pointed closed convex cone in R^p and if Y is a closed convex or D-closed D-convex set in R^p, then

$$\mathscr{E}(Y, D) \subset \mathrm{cl}\, \mathscr{P}(Y, D).$$

Proof Since $\mathscr{E}(Y + D, D) = \mathscr{E}(Y, D)$ and $\mathscr{P}(Y + D, D) = \mathscr{P}(Y, D)$ when D is a pointed closed convex cone, it suffices to consider the first case (Y is closed and convex). If $\mathscr{E}(Y, D) = \varnothing$, then the result is obvious. On the other hand, if $\mathscr{E}(Y, D) \neq \varnothing$, then Y is D-bounded from Theorem 3.2.5, and we can apply Theorem 3.2.11 to obtain the result.

Theorem 3.2.12

Let D be a pointed closed convex cone, and let Y be a nonempty closed convex (or D-closed D-convex) set. Then the following are equivalent:

(i) Any of the conditions (i)–(v) in Proposition 3.2.3 is satisfied;
(ii) $\mathscr{E}(Y, D) \neq \varnothing$;
(iii) $\mathscr{P}(Y, D) \neq \varnothing$;
(iv) $\mathscr{E}(Y, D)$ is externally stable.

Proof (i) $\Leftrightarrow$ (ii) is shown in Theorem 3.2.5. (i) $\Rightarrow$ (iv) follows from Theorem 3.2.9 or Theorem 3.2.10. (ii) $\Rightarrow$ (iii) follows from Theorem 3.2.11. Finally, (iii) $\Rightarrow$ (ii) and (iv) $\Rightarrow$ (ii) are trivial.

Some examples of externally unstable sets will be given later in Chapter 4, in connection with the perturbation stability of efficient sets.

3.3 Connectedness of Efficient Sets

In this section we discuss the connectedness of efficient sets in multiobjective optimization problems. Throughout this section, the domination structure is given by a constant pointed closed convex cone D. The following results are mainly due to Naccache [N1].

Definition 3.3.1 (*Separatedness and Connectedness*)

A set $A \subset R^p$ is said to be separated if it can be written as

$$A = A_1 \cup A_2,$$

where

$$\operatorname{cl} A_1 \cap A_2 = A_1 \cap \operatorname{cl} A_2 = \varnothing$$

with two nonempty sets A_1 and A_2. Equivalently, A is separated if there exist two open sets O_1 and O_2 such that

$$A \subset O_1 \cup O_2, \qquad A \cap O_1 \neq \varnothing, \qquad A \cap O_2 \neq \varnothing, \qquad A \cap O_1 \cap O_2 = \varnothing.$$

If A is not separated, then A is said to be connected.

The following two lemmas are well known and useful in this section.

Lemma 3.3.1

If a set A is connected and

$$A \subset B \subset \operatorname{cl} A,$$

then the set B is also connected.

Lemma 3.3.2

If $\{A_i : i \in I\}$ is a family of connected sets such that $\bigcap_{i \in I} A_i \neq \varnothing$, then $\bigcup_{i \in I} A_i$ is connected.

Now let

$$S(\mu, Y) = \left\{ \hat{y} \in Y : \langle \mu, \hat{y} \rangle = \inf_{y \in Y} \langle \mu, y \rangle \right\}$$

and

$$\mathscr{D}(Y, D) = \bigcup_{\mu \in \operatorname{int} D^\circ} S(\mu, Y).$$

From the results given later in Section 3.4 (Theorems 3.4.3 and 3.4.6), we have the following lemma.

Lemma 3.3.3

$$\mathscr{D}(Y, D) \subset \mathscr{E}(Y, D)$$

and, moreover if Y is closed convex, then

$$\mathscr{E}(Y, D) \subset \operatorname{cl} \mathscr{D}(Y, D).$$

We first prove the connectedness of the set $\mathscr{D}(Y, D)$ in the case in which Y is a compact convex set.

Lemma 3.3.4

If Y is a compact convex set, then $\mathscr{D}(Y, D)$ is connected.

Proof Suppose, to the contrary, that $\mathscr{D}(Y, D)$ is not connected. Then there exist two open sets Y_1 and Y_2 such that

$$Y_i \cap \mathscr{D}(Y, D) \neq \varnothing, \qquad i = 1, 2, \qquad Y_1 \cap Y_2 \cap \mathscr{D}(Y, D) = \varnothing,$$
$$\mathscr{D}(Y, D) \subset Y_1 \cup Y_2.$$

Let

$$\mathscr{M}_i = \{\mu \in \operatorname{int} D^\circ : S(\mu, Y) \cap Y_i \neq \varnothing\} \qquad i = 1, 2.$$

Since $S(\mu, Y)$ is connected by convexity,

$$\mathscr{M}_i = \{\mu \in \operatorname{int} D^\circ : S(\mu, Y) \subset Y_i\} \qquad i = 1, 2.$$

Hence,

$$\mathscr{M}_1 \cap \mathscr{M}_2 = \varnothing.$$

Moreover, since $Y_i \cap \mathscr{D}(Y, D) \neq \varnothing$,

$$\mathscr{M}_i \cap \operatorname{int} D^\circ \neq \varnothing \qquad \text{for} \quad i = 1, 2,$$

and since $\mathscr{D}(Y, D) \subset Y_1 \cup Y_2$,

$$\operatorname{int} D^\circ \subset \mathscr{M}_1 \cup \mathscr{M}_2 \quad \text{(to say exactly } \operatorname{int} D^\circ = \mathscr{M}_1 \cup \mathscr{M}_2\text{)}.$$

Finally, each $\mathscr{M}_i$ is an open set (see Lemma 3.3.5 below). These all imply that int D° is not connected, which is a contradiction. Therefore $\mathscr{D}(Y, D)$ should be connected.

Lemma 3.3.5

If Y is compact convex, then the sets

$$\mathscr{M}_i = \{\mu \in \operatorname{int} D^\circ : S(\mu, Y) \subset Y_i\} \qquad i = 1, 2$$

in the proof of Lemma 3.3.4 are open.

Proof We will show that $\mathscr{M}_1$ is open. Suppose, to the contrary, that $\mathscr{M}_1$ is not open. Then there exist $\hat{\mu} \in \mathscr{M}_1$ and $\mu^k \in \operatorname{int} D^\circ \backslash \mathscr{M}_1 = \mathscr{M}_2$ $(k = 1, 2, \ldots)$ such that $\mu^k \to \hat{\mu}$. Take $y^k \in S(\mu^k, Y)$ $(k = 1, 2, \ldots)$. Since Y is compact we can assume without loss of generality that the sequence $\{y^k\}$ converges to an element $\hat{y} \in Y$, which is contained in $S(\hat{\mu}, Y)$. In fact, if $\hat{y} \notin S(\hat{\mu}, Y)$, there exists a $y' \in Y$ such that

$$\langle \hat{\mu}, y' \rangle < \langle \hat{\mu}, \hat{y} \rangle.$$

From the continuity of the inner product we see that for sufficiently large k,

$$\langle \mu^k, y' \rangle < \langle \mu^k, y^k \rangle,$$

which contradicts the fact that $y^k \in S(\mu^k, Y)$.

Since $y^k \in S(\mu^k, Y) \subset Y_2 \cap \mathscr{D}(Y, D)$ and $Y_1 \cap Y_2 \cap \mathscr{D}(Y, D) = \varnothing$, then $y^k \in Y_1^c$ (the complement of Y_1) for all $k = 1, 2, \ldots$. Since Y_1^c is a closed set, the limit $\hat{y}$ of $\{y^k\}$ is also in Y_1^c. Thus, we have $\hat{y} \notin Y_1$, which is contradictory to $\hat{\mu} \in \mathscr{M}_1$.

Theorem 3.3.1†

If D is a pointed, closed, convex cone and Y is a closed, convex, D-bounded set, then $\mathscr{E}(Y, D)$ is connected.

Proof Pick

$$d \in \operatorname{int} D$$

and define

$$y(\alpha) = \alpha d \qquad \alpha \in R.$$

We claim that, for any $y \in R^p$, there exists $\alpha > 0$ such that

$$y \in y(\alpha) - D.$$

† Naccache [N1].

If this is not true, then we can apply the standard separation theorem to the ray $\{y - \alpha d : \alpha > 0\}$ and the convex cone $-D$ to deduce the existence of $\mu \in R^p \backslash \{0\}$ such that

$$\langle \mu, y - \alpha d \rangle \geqq 0 \qquad \text{for all} \quad \alpha > 0$$

and

$$\langle \mu, d' \rangle \geqq 0 \qquad \text{for all} \quad d' \in D.$$

Since $d \in \operatorname{int} D$, $\langle \mu, d \rangle > 0$ from the latter inequality and so $\langle \mu, y - \alpha d \rangle < 0$ for sufficiently large α, which is a contradiction.

We can therefore choose $\hat{\alpha} > 0$ such that

$$(y(\hat{\alpha}) - D) \cap \mathscr{E}(Y, D) \neq \varnothing.$$

Let

$$E(\alpha) = \mathscr{E}((y(\alpha) - D) \cap Y, D).$$

The claim proven above shows that

$$\mathscr{E}(Y, D) = \bigcup_{\alpha \geqq \hat{\alpha}} E(\alpha).$$

Since $(y(\alpha) - D) \cap Y$ is compact convex from the assumptions (see Proposition 3.2.3), $E(\alpha)$ is connected from Lemmas 3.3.1, 3.3.3, and 3.3.4. Moreover, if $\alpha \geqq \hat{\alpha}$, then

$$E(\alpha) \supset E(\hat{\alpha}),$$

which implies that

$$\bigcap_{\alpha \geqq \hat{\alpha}} E(\alpha) = E(\hat{\alpha}) \neq \varnothing.$$

We have therefore expressed $\mathscr{E}(Y, D)$ as the union of a family of connected sets having a nonempty intersection and as such $\mathscr{E}(Y, D)$ must be connected. This completes the proof.

Remark 3.3.1

The condition that Y be closed and convex in Theorem 3.3.1 may be replaced by the assumption that Y is D-closed and D-convex according to Proposition 3.1.2.

Bitran and Magnanti [B14] obtained essentially the same result as Theorem 3.3.1. Warburton [W4] recently obtained connectedness results for the set of Pareto optimal and weak Pareto optimal solutions in the decision space (i.e., x space).

3.4 Characterization of Efficiency and Proper Efficiency

In this section we develop characterization theorems for efficient and properly efficient points. The domination structure is assumed to be specified by a convex cone D. The efficient set of Y with respect to D is denoted by $\mathscr{E}(Y, D)$ as before; that is,

$$\hat{y} \in \mathscr{E}(Y, D) \quad \text{if and only if } (Y - \hat{y}) \cap (-D) = \begin{cases} \{0\} \\ \varnothing \quad \text{when } D \text{ does not contain the origin.} \end{cases}$$

Moreover, when D is a closed convex cone, we use $\mathscr{P}(Y, D)$ to denote the set of all properly efficient points of Y in the sense of Benson. Namely,

$$\hat{y} \in \mathscr{P}(Y, D) \qquad \text{if and only if} \quad \operatorname{cl} P(Y + D - \hat{y}) \cap (-D) = \{0\}.$$

Three kinds of characterization are discussed in sequence. Some more results not discussed here can be seen in Gearhart [G4] and Jahn [J2].

3.4.1 Characterization by Scalarization

Most well-known results regarding the characterization of efficient or properly efficient points are via scalarization by vectors in the polar or strict polar cone of the domination cone. As in Section 3.3, let

$$S(\mu, Y) = \{\hat{y} \in Y : \langle \mu, \hat{y} \rangle = \inf\{\langle \mu, y \rangle : y \in Y\}\}.$$

Geometrically, $S(\mu, Y)$ is the set of supporting points of Y with the inner normal vector μ. It is obvious that $S(\alpha\mu, Y) = S(\mu, Y)$ for any $\alpha > 0$. Hence, we may often normalize μ as $\|\mu\| = 1$. Let

$$\mathscr{D}(Y, D) = \bigcup_{\mu \in D^{s\circ}} S(\mu, Y)$$

and

$$\bar{\mathscr{D}}(Y, D) = \bigcup_{\mu \in D^{\circ} \setminus \{0\}} S(\mu, Y).$$

Of course, $\mathscr{D}(Y, D) \subset \bar{\mathscr{D}}(Y, D)$.

Proposition 3.4.1

If Y is a closed set, then

$$\operatorname{cl} \mathscr{D}(Y, D) \subset \bar{\mathscr{D}}(Y, D).$$

Proof Let $y^k \in S(\mu^k, Y)$ with $\mu^k \in D^{so}$ and $y^k \to \hat{y}$. Since we can normalize μ^k as $\|\mu^k\| = 1$, we may assume that μ^k converges to some $\hat{\mu}$ with $\|\hat{\mu}\| = 1$ without loss of generality by taking a subsequence if necessary. Then $\hat{\mu} \in \operatorname{cl} D^{so} \subset D^{\circ}$ since $D^{so} \subset D^{\circ}$ and D° is closed. Moreover, $\hat{y} \in S(\hat{\mu}, Y)$. In fact, since $y^k \in S(\mu^k, Y)$,

$$\langle \mu^k, y^k - y \rangle \leqq 0 \qquad \text{for} \quad \forall k = 1, 2, \ldots, \qquad \forall y \in Y.$$

Taking the limit as $k \to \infty$, we have

$$\langle \hat{\mu}, \hat{y} - y \rangle \leqq 0 \qquad \text{for any} \quad y \in Y.$$

Hence $\hat{y} \in \bar{\mathscr{D}}(Y, D)$, as was to be proved.

We will now investigate the relationships between $\mathscr{P}(Y, D)$ (or $\mathscr{E}(Y, D)$) and $\mathscr{D}(Y, D)$ (or $\bar{\mathscr{D}}(Y, D)$).

Theorem 3.4.1

If D is a closed convex cone in R^p, then for an arbitrary Y in R^p,

$$\mathscr{D}(Y, D) \subset \mathscr{P}(Y, D).$$

Proof Let $\mu \in D^{so}$ and $\hat{y} \in S(\mu, Y)$. If we suppose that

$$-d \in \operatorname{cl} P(Y + D - \hat{y}) \qquad \text{and} \qquad d \in D,$$

then there exist sequences $\{y^k\} \subset Y$, $\{d^k\} \subset D$ and $\{\beta_k\} \subset R$ such that $\beta_k \geqq 0$ and

$$\beta_k(y^k + d^k - \hat{y}) \to -d \qquad \text{as} \quad k \to \infty.$$

Taking the inner product with μ, we have

$$\lim_{k \to \infty} \beta_k\{(\langle \mu, y^k \rangle - \langle \mu, \hat{y} \rangle) + \langle \mu, d^k \rangle\} = -\langle \mu, d \rangle.$$

Since $\hat{y} \in S(\mu, Y)$, $d^k \in D$, and $\mu \in D^{so}$, the left-hand side is nonnegative. So

$$\langle \mu, d \rangle \leqq 0,$$

and hence $d = 0$. Thus we have established that

$$\operatorname{cl} P(Y + D - \hat{y}) \cap (-D) = \{0\},$$

i.e., $\hat{y} \in \mathscr{P}(Y, D)$. This completes the proof.

Theorem 3.4.2

If D is a pointed, closed, convex cone and Y is a D-convex set in R^p, then

$$\mathscr{P}(Y, D) \subset \mathscr{D}(Y, D).$$

Proof Let $\hat{y} \in \mathscr{P}(Y, D)$, i.e.,

$$\mathrm{cl}\, P(Y + D - \hat{y}) \cap (-D) = \{0\}.$$

From Lemma 3.1.1, $\mathrm{cl}\, P(Y + D - \hat{y})$ is a closed convex cone. So we can apply Proposition 2.1.8 to obtain that

$$D^{so} \cap (\mathrm{cl}\, P(Y + D - \hat{y}))^{\circ} \neq \emptyset.$$

Thus, there exists a $\mu \in D^{so}$ such that

$$\mu \in (\mathrm{cl}\, P(Y + D - \hat{y}))^{\circ}.$$

Since $Y - \hat{y} \subset \mathrm{cl}\, P(Y + D - \hat{y})$,

$$\langle \mu, y - \hat{y} \rangle \geqq 0 \qquad \text{for any} \quad y \in Y.$$

Namely,

$$\langle \mu, y \rangle \geqq \langle \mu, \hat{y} \rangle \qquad \text{for all} \quad y \in Y.$$

Therefore, $\hat{y} \in \mathscr{D}(Y, D)$, as was to be proved.

Remark 3.4.1

In view of Theorems 3.4.1 and 3.4.2, when D is a pointed closed convex cone and Y is a D-convex set, every properly efficient point of Y can be completely characterized as a supporting point of Y with an inner normal vector in D^{so}. This geometric result is quite useful in the dicussion of multiobjective duality theory (Chapter 5).

Theorem 3.4.3

If D is a convex cone and Y is a set in R^p, then

$$\mathscr{D}(Y, D) \subset \mathscr{E}(Y, D).$$

Proof Suppose that $\hat{y} \notin \mathscr{E}(Y, D)$. If $\hat{y} \notin Y$, it is clear that $\hat{y} \notin \mathscr{D}(Y, D)$. So assume that $\hat{y} \in Y \backslash \mathscr{E}(Y, D)$, and there exists a $y' \in Y$ such that $\hat{y} - y' \in D \backslash \{0\}$. Then for any $\mu \in D^{so}$,

$$\langle \mu, \hat{y} - y' \rangle > 0.$$

Thus $\hat{y} \notin \mathscr{D}(Y, D)$, as was to be proved.

Theorem 3.4.4

If D is an acute convex cone and Y is a D-convex set in R^p, then

$$\mathscr{E}(Y, D) \subset \bar{\mathscr{D}}(Y, D).$$

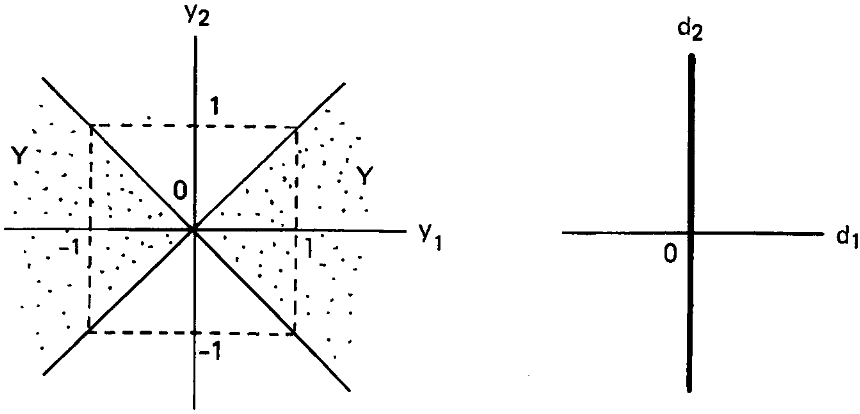

Fig. 3.13. $\mathscr{E}(Y, D) \not\subset \bar{\mathscr{D}}(Y, D)$. (Example 3.4.1.)

Proof Let $\hat{y} \in \mathscr{E}(Y, D)$. From Proposition 3.1.2, $\hat{y} \in \mathscr{E}(Y + D, D)$ and so $(Y + D - \hat{y}) \cap (-D) = \{0\}$ (or $\varnothing$ when D does not contain the origin). Applying the separation theorem (Theorem 2.1.1), we have the existence of a nonzero vector $\mu \in R^p$ such that

$$\langle \mu, y + d - \hat{y} \rangle \geqq 0 \qquad \text{for all} \quad y \in Y \qquad \text{and} \qquad d \in D,$$

$$\langle \mu, -d' \rangle \leqq 0 \qquad \text{for all} \quad d' \in D.$$

From the latter, $\mu \in D^\circ$. On the other hand, from the former,

$$\langle \mu, y \rangle \geqq \langle \mu, \hat{y} \rangle \qquad \text{for all} \quad y \in Y,$$

since we can take $d = 0$ (or as nearly equal to 0 as desired). Thus we have established that $\hat{y} \in S(\mu, Y) \subset \bar{\mathscr{D}}(Y, D)$. This completes the proof of the theorem.

In the above theorem, the acuteness condition of D cannot be eliminated. The following example was given by Yu [Y1].

Example 3.4.1[†] (Fig. 3.13)

Let

$$Y = \{(y_1, y_2) : -y_1 \leqq y_2 \leqq y_1\} \cup \{(y_1, y_2) : y_1 \leqq y_2 \leqq -y_1\},$$

$$D = \{(d_1, d_2) : d_1 = 0\}.$$

Then $D^\circ = \{(\mu_1, \mu_2) : \mu_2 = 0\}$, and so $\bar{\mathscr{D}}(Y, D) = \varnothing$. However,

$$\mathscr{E}(Y, D) = \{(0, 0)\}.$$

[†] Yu [Y1].

Corollary 3.4.1

Let D be an acute, open, convex cone and Y be a D-convex set in R^p. Then

$$\mathscr{E}(Y, D) = \bar{\mathscr{D}}(Y, D).$$

Proof In view of Proposition 2.1.5(iv), $D^{so} = D^\circ \backslash \{0\}$ when D is open. Hence $\mathscr{D}(Y, D) = \bar{\mathscr{D}}(Y, D)$, and the corollary directly follows from Theorems 3.4.3 and 3.4.4.

An element in $S(\mu, Y)$ is not necessarily efficient for $\mu \in D^\circ$, in contrast with Theorem 3.4.4. However, we can at least claim the following.

Theorem 3.4.5

Let D be an acute, convex cone and Y be a set in R^p. If $\mu \in D^\circ$ and $S(\mu, Y) = \{\hat{y}\}$, then $\hat{y} \in \mathscr{E}(Y, D)$. If Y is nonempty and compact, then for any $\mu \in D^\circ$, $S(\mu, Y) \cap \mathscr{E}(Y, D) \neq \varnothing$.

Proof If $\hat{y} \notin \mathscr{E}(Y, D)$, there exists $y' \neq \hat{y}$ in Y such that $\hat{y} - y' \in D$. Then $\langle \mu, \hat{y} - y' \rangle \geqq 0$, which implies that $S(\mu, Y) \neq \{\hat{y}\}$.

Next, assume that Y is nonempty and compact. Then, for any $\mu \in D^\circ$, $S(\mu, Y)$ is a nonempty, compact set, and hence there exists $\bar{y} \in \mathscr{E}(S(\mu, Y), D)$ from Theorem 3.2.2 and Remark 3.2.2. Then $\bar{y} \in \mathscr{E}(Y, D)$. In fact, if there were $y' \in Y$ such that $\bar{y} - y' \in D \backslash \{0\}$, then the inequality $\langle \mu, \bar{y} \rangle \geqq \langle \mu, y' \rangle$ holds. Therefore, $y' \in S(\mu, Y)$, which contradicts the supposition $\bar{y} \in \mathscr{E}(S(\mu, Y), D)$. Thus $\bar{y} \in S(\mu, Y) \cap \mathscr{E}(Y, D)$, and the proof of the theorem is completed.

The following theorem, which is a generalization of a theorem of Arrow *et al.* [A4], is due to Hartley [H5].

Theorem 3.4.6 (cf. Corollary 3.2.2)

If D is a pointed closed convex cone and if Y is a closed, convex set or if Y is a D-closed, D-convex set, then

$$\mathscr{E}(Y, D) \subset \operatorname{cl} \mathscr{D}(Y, D) = \operatorname{cl} \mathscr{P}(Y, D).$$

Proof Theorems 3.4.1 and 3.4.2 ensure that

$$\operatorname{cl} \mathscr{D}(Y, D) = \operatorname{cl} \mathscr{P}(Y, D)$$

in this case. Hence, the result follows immediately from Corollary 3.2.2. However, we will give a direct proof of the relationship. Hence we prove that $\mathscr{E}(Y, D) \subset \operatorname{cl} \mathscr{D}(Y, D)$. Since $\mathscr{E}(Y + D, D) = \mathscr{E}(Y, D)$ and $\mathscr{D}(Y, D) = \mathscr{D}(Y + D, D)$, it suffices to prove the first case (Y is a closed convex set). If Y is empty, the theorem is trivial, and so we assume that there is $\hat{y} \in \mathscr{E}(Y, D)$. We can take $\hat{y} = 0$ for simplicity without loss of generality.

First we prove the case in which Y is compact and convex. Choose $\bar{\mu} \in D^{so}$ and let $C(\varepsilon) = \varepsilon\bar{\mu} + D^{\circ}$ for any $\varepsilon > 0$. Then, for all ε sufficiently small, Y and $C(\varepsilon) \cap B$ are both nonempty, compact, and convex, where B is the closed unit ball. Hence, by the minimax theorem of Sion–Kakutani (Stoer and Witzgall [S15]), there exist $y(\varepsilon) \in Y$ and $\mu(\varepsilon) \in C(\varepsilon) \cap B$ such that

$$\langle \mu, y(\varepsilon)\rangle \leqq \langle \mu(\varepsilon), y(\varepsilon)\rangle \leqq \langle \mu(\varepsilon), y\rangle \qquad \text{for} \quad \forall y \in Y, \qquad \forall \mu \in C(\varepsilon) \cap B.$$

Since $0 \in Y$, we have

$$\langle \mu, y(\varepsilon)\rangle \leqq 0 \qquad \text{for all} \quad \mu \in C(\varepsilon) \cap B.$$

Since Y is compact, there exists a sequence of positive numbers $\{\varepsilon^k\}$ converging to 0 such that the sequence $\{y^k\} = \{y(\varepsilon^k)\}$ has a limit $\bar{y} \in Y$ as $k \to \infty$. For any $\mu \in D^{so} \cap B = (\operatorname{int} D^{\circ}) \cap B$, we can find an $\bar{\varepsilon} > 0$ such that $\mu \in C(\varepsilon) \cap B$ for all $\varepsilon \leqq \bar{\varepsilon}$, and thus $\langle \mu, y^k\rangle \leqq 0$ for all large enough k. Since D^{so} is a cone, this implies that

$$\langle \mu, \bar{y}\rangle \leqq 0 \qquad \text{for all} \quad \mu \in D^{so}$$

and thus

$$-\bar{y} \in (D^{so})^{\circ} = D.$$

Because $0 \in \mathscr{E}(Y, D)$, $\bar{y} = 0$ and defining

$$\mu^k = \mu(\varepsilon^k)/\|\mu(\varepsilon^k)\| \in D^{so} \cap \partial B,$$

we have

$$y^k \in S(\mu^k, Y) \subset \mathscr{D}(Y, D).$$

Now suppose Y is not necessarily compact (but closed and convex). Let $0 \in \mathscr{E}(Y, D)$. Since $Y \cap B$ is a nonempty, compact, convex set and $0 \in \mathscr{E}(Y \cap B, D)$, from the above result there exist sequences $\{\mu^k\} \subset D^{so}$ with $\|\mu^k\| = 1$ and $y^k \in S(\mu^k, Y \cap B)$ such that $y^k \to 0$. If k is sufficiently large, $y^k \in \operatorname{int} B$ and thus $y^k \in S(\mu^k, Y)$. In fact, if there were $y' \in Y$ such that

$$\langle \mu^k, y'\rangle < \langle \mu^k, y^k\rangle,$$

then $\alpha y' + (1 - \alpha)y^k \in Y \cap B$ for sufficiently small $\alpha > 0$ and

$$\langle \mu^k, \alpha y' + (1 - \alpha)y^k\rangle < \langle \mu^k, y^k\rangle,$$

which contradicts $y^k \in S(\mu^k, Y \cap B)$. This completes the proof of the theorem.

In this theorem, the converse inclusion $\mathscr{E}(Y, D) \supset \operatorname{cl} \mathscr{P}(Y, D)$ does not hold generally as can be seen in the following example.

Example 3.4.2[†]

Let

$$Y' = \{(y_1, y_2, y_3) : (y_1 - 1)^2 + (y_2 - 1)^2 = 1,\ y_1 \leqq 1,\ y_2 \leqq 1,\ y_3 = 1\},$$
$$Y = \mathrm{co}(Y' \cup \{(1,0,0)\}),$$
$$D = R^3_+.$$

Then Y is closed and convex, and $\hat{y} = (1,0,1) \notin \mathscr{E}(Y, D)$, since $\hat{y} = (1,0,0) + (0,0,1)$. However any $y \in Y'$ such that $y_1 < 1$, $y_2 < 1$ is contained in $\mathscr{P}(Y, D) = \mathscr{D}(Y, D)$. In fact, if we put

$$\bar{y} = (1 - \cos\theta,\ 1 - \sin\theta,\ 1) \qquad \text{for} \quad 0 < \theta < \tfrac{\pi}{2},$$

then

$$\mu = (1 - \alpha)(\cos\theta, \sin\theta, 0) + \alpha(0,0,1), \qquad 0 < \alpha < 1$$

is in $D^{s\circ}$ and, for any $y \in Y'$ with $y = (1 - \cos\theta', 1 - \sin\theta', 1)$, $0 \leqq \theta' \leqq \tfrac{\pi}{2}$,

$$\begin{aligned}\langle \mu, y - \bar{y} \rangle &= (1 - \alpha)\{(\cos\theta)(\cos\theta - \cos\theta') + (\sin\theta)(\sin\theta - \sin\theta')\} \\ &= (1 - \alpha)\{1 - \cos(\theta - \theta')\} \geqq 0.\end{aligned}$$

Moreover,

$$\begin{aligned}\langle \mu, (1,0,0) - \bar{y} \rangle &= (1 - \alpha)\cos^2\theta - (1 - \alpha)\sin\theta(1 - \sin\theta) - \alpha \\ &= (1 - \alpha)(1 - \sin\theta) - \alpha > 0\end{aligned}$$

for any α sufficiently small. Thus

$$\langle \mu, \bar{y} \rangle \leqq \langle \mu, y \rangle \qquad \text{for all} \quad y \in Y$$

and so $\bar{y} \in \mathscr{D}(Y, D)$. Letting $\theta \to 0$, we find $\hat{y} \in \mathrm{cl}\,\mathscr{D}(Y, D)$.

When Y is a polyhedral convex set, we obtain the following result.

Theorem 3.4.7

If Y is a polyhedral convex set and D is a pointed closed convex cone, then

$$\mathscr{P}(Y, D) = \mathscr{E}(Y, D) = \mathscr{D}(Y, D).$$

Proof The theorem is obvious from Theorems 3.1.7, 3.4.1 and 3.4.2.

[†] Arrow *et al.* [A4].

3.4.2 Characterization as Best Approximations to the Ideal Point

In this subsection we concentrate on compromise solutions to multiobjective optimization problems. The domination structure is assumed to be specified by the nonnegative orthant R^p_+. In other words, we pay attention to the Pareto optimal points. The idea of compromise solutions is based on a procedure that calculates the Pareto optimal points as approximations to an ideal or utopia point (Yu and Leitmann [Y3], Zeleny [Z3], Bowman [B20], Gearhart [G3], Salukvadze [S3]). We investigate the relationships between compromise solutions and efficient (or properly efficient) points partly according to Gearhart [G3].

We consider a multiobjective optimization problem

(P) minimize $\quad f(x) = (f_1(x), \ldots, f_p(x)) \quad$ subject to $\quad x \in X \subset R^n$

with the domination structure $D = R^p_+$. Therefore, our aim is to find the set of all Pareto optimal points (or solutions). Since D is fixed, the efficient set of $Y = f(X)$ is simply denoted by $\mathscr{E}(Y)$. We assume that the set $Y + R^p_+$ is closed (i.e., Y is R^p_+-closed) and that

$$\tilde{y}_i = \inf\{f_i(x) : x \in X\} > -\infty$$

for each $i = 1, \ldots, p$. The point $\tilde{y} = (\tilde{y}_1, \ldots, \tilde{y}_p)$ is often called the utopia point or the ideal point. However, we might take another ideal point $\bar{y}$ as $\bar{y} \leqq \tilde{y}$.

The l_∞-norm in R^p is defined by

$$\|y\|_\infty = \max\{|y_i| : 1 \leqq i \leqq p\}, \qquad y \in R^p.$$

Let A be a scalar index set bounded below and unbounded above (a simple example is $A = [1, \infty)$) and let a family of norms $\{\|\cdot\|_\alpha : \alpha \in A\}$ be given which has the following properties:

Property P1: $\|y\|_\infty \leqq M(\alpha)\|y\|_\alpha$ with $M(\alpha) > 0$, $\alpha \in A$, $y \in R^p$,

Property P2: $M(\alpha)\|y\|_\alpha \to \|y\|_\infty$ as $\alpha \to \infty$, $y \in \mathrm{R}^p$

Property P3: if $0 \leqq y \leq y'$, then $\|y\|_\alpha < \|y'\|_\alpha$ for each $\alpha \in A$.

Example 3.4.3

(i) The most popular example of such families is that of the l_α-norms, i.e.,

$$\|y\|_\alpha = \left[\sum_{i=1}^{p} |y_i|^\alpha\right]^{1/\alpha}, \qquad \alpha \in [1, \infty).$$

In this case $M(\alpha) = 1$.

(ii) Another example, which was used by Dinkelbach and Isermann [D6], is given by

$$\|y\|_\alpha = \|y\|_\infty + (1/\alpha)\left(\sum_{i=1}^{p} |y_i|\right), \qquad \alpha \in [1, \infty).$$

In this case $M(\alpha) = 1$.

(iii) The final example was proposed by Choo and Atkins [C10]. Let I_α be the $p \times p$ matrix whose (j, k) entry is given by

$$(I_\alpha)_{jk} = \begin{cases} 1 & \text{if } j = k \\ -1/(p-1)\alpha & \text{if } j \neq k \end{cases} \qquad \text{for} \quad \alpha \in (1, \infty).$$

Then I_α is nonsingular and

$$(I_\alpha^{-1})_{jk} = \begin{cases} \alpha((p-1)(\alpha-1)+1)/(\alpha-1)((p-1)\alpha+1) & \text{if} \quad j = k \\ \alpha/(\alpha-1)((p-1)\alpha+1) & \text{if} \quad j \neq k. \end{cases}$$

Here every component is positive and

$$(I_\alpha^{-1})_{jj} = [(p-1)(\alpha-1)+1](I_\alpha^{-1})_{jk} > (I_\alpha^{-1})_{jk} \qquad (j \neq k).$$

We define

$$\|y\|_\alpha = \|I_\alpha^{-1} y\|_\infty.$$

In this case

$$|y_i| = |(I_\alpha I_\alpha^{-1} y)_i| \leqq |(I_\alpha^{-1} y)_i| + \left(\sum_{\substack{j=1 \\ j \neq i}}^{p} |(I_\alpha^{-1} y)_j|\right) \Big/ (p-1)\alpha.$$

Hence

$$\|y\|_\infty \leqq (1 + 1/\alpha)\|I_\alpha^{-1} y\|_\infty = M(\alpha)\|y\|_\alpha,$$

where $M(\alpha) = 1 + 1/\alpha$.

Let $\mathring{S}^p$ be the set as follows:

$$\mathring{S}^p = \left\{\mu \in R^p : \sum_{i=1}^{p} \mu_i = 1,\ \mu_i > 0\ (i = 1, \ldots, p)\right\}.$$

For each $\mu \in \mathring{S}^p$ and $\alpha \in A$, the (μ, α)-norm of $y \in R^p$ is defined by

$$\|y\|_\alpha^\mu = \|\mu \odot y\|_\alpha,$$

and the (μ, ∞)-norm is defined by

$$\|y\|_\infty^\mu = \|\mu \odot y\|_\infty,$$

where $\mu \odot y = (\mu_1 y_1, \ldots, \mu_p y_p)$.

In this subsection $\bar{y} \in R^p$ will designate a fixed but arbitrary point such that $\bar{y} \leqq \tilde{y}$. This point will serve as the ideal point. Define

$$A(\mu, \alpha, Y) = \{\hat{y} \in Y : \|\hat{y} - \bar{y}\|_\alpha^\mu = \inf\{\|y - \bar{y}\|_\alpha^\mu : y \in Y\}\},$$

i.e., $A(\mu, \alpha, Y)$ is the set of best approximations to $\bar{y}$ out of Y in the (μ, α)-norm. Moreover, denote by $\mathscr{A}(Y)$ the set $\bigcup_{\mu \in \mathring{S}^p, \alpha \in A} A(\mu, \alpha, Y)$, i.e., $\mathscr{A}(Y)$ is the set of all best approximations to $\bar{y}$ out of Y in the (μ, α)-norm for some $\mu \in \mathring{S}^p$ and $\alpha \in A$.

Definition 3.4.1 (*Compromise Solution*)

A point $\bar{y}$ in $\mathscr{A}(Y)$ is called the compromise solution (in the objective space) to the multiobjective optimization problem (P).

We now investigate the relationships between $\mathscr{A}(Y)$ and $\mathscr{E}(Y)$ (or $\mathscr{P}(Y)$).

Theorem 3.4.8

If Y is R_+^p-closed and the family of norms $\{\|\cdot\|_\alpha : \alpha \in A\}$ satisfies Properties P1–P3, then

$$\mathscr{A}(Y) \subset \mathscr{P}(Y) \subset \mathscr{E}(Y) \subset \mathrm{cl}\,\mathscr{A}(Y).$$

Proof Since $\mathscr{P}(Y) \subset \mathscr{E}(Y)$, we will prove that $\mathscr{A}(Y) \subset \mathscr{P}(Y)$ and $\mathscr{E}(Y) \subset \mathrm{cl}\,\mathscr{A}(Y)$.

($\mathscr{A}(Y) \subset \mathscr{P}(Y)$): First, let $\hat{y} \in \mathscr{A}(Y)$. Then for some $\mu \in \mathring{S}^p$ and $\alpha \in A$,

$$\|\hat{y} - \bar{y}\|_\alpha^\mu \leqq \|y - \bar{y}\|_\alpha^\mu \qquad \text{for all} \quad y \in Y.$$

Suppose to the contrary that $\hat{y} \notin \mathscr{P}(Y)$. Then there exist sequences $\{\beta_k\} \subset R$, $\{y^k\} \subset Y$, $\{d^k\} \subset R_+^p$ such that $\beta_k > 0$ and

$$\beta_k(y^k + d^k - \hat{y}) \to -d \qquad \text{as} \quad k \to \infty \qquad \text{for} \quad \text{some nonzero } d \in R_+^p.$$

There are two cases for $\{\beta_k\}$: bounded and unbounded. If $\{\beta_k\}$ is bounded we may assume, without loss of generality (by taking a subsequence if necessary) that $\beta_k \to \beta_0$ for some $\beta_0 \geqq 0$. If $\beta_0 = 0$, noting that $y^k + d^k - \hat{y} \geqq \bar{y} - \hat{y}$, we have $-d \in (\bar{y} - \hat{y} + R_+^p)^+ = R_+^p$, which is a contradiction. If $\beta_0 > 0$, $y^k + d^k - \hat{y} \to -d/\beta_0 \neq 0 \in -R_+^p$ as $k \to \infty$. Since $Y + R_+^p$ is closed, $\hat{y} - d/\beta_0 \in Y + R_+^p$, i.e., there exists $y^0 \in Y$ such that $\hat{y} \geq y^0$. However, this implies that

$$\|\hat{y} - \bar{y}\|_\alpha^\mu > \|y^0 - \bar{y}\|_\alpha^\mu$$

by Property P3 and leads to a contradiction.

If $\{\beta_k\}$ is unbounded we may assume that, $\beta_k \to \infty$ as $k \to \infty$. Hence $y^k + d^k - \hat{y} \to 0$. If $\hat{y}_i = \bar{y}_i$ for some i, then $d_i = 0$. In fact,

$$\beta_k(y_i^k + d_i^k - \hat{y}_i) \geqq \beta_k(\bar{y}_i + d_i^k - \hat{y}_i) = \beta_k d_i^k \geqq 0$$

for this i and so $-d_i \geqq 0$. Thus, there exists a $\bar{\beta} > 0$ such that

$$0 \leqq \hat{y} - d/\beta - \bar{y} \leq \hat{y} - \bar{y} \qquad \text{for all} \quad \beta \geqq \bar{\beta}.$$

Then, from Property P3,

$$\|\hat{y} - d/\beta - \bar{y}\|_\alpha^\mu < \|\hat{y} - \bar{y}\|_\alpha^\mu \qquad \text{for all} \quad \beta \geqq \bar{\beta}.$$

Since $\beta_k \to \infty$, then $\beta_k \geqq \bar{\beta}$ for all k sufficiently large. Hence,

$$\|y^k + d^k - \bar{y}\|_\alpha^\mu < \|\hat{y} - \bar{y}\|_\alpha^\mu$$

for some k sufficiently large. In fact,

$$\begin{aligned}\|y^k + d^k - \bar{y}\|_\alpha^\mu &= \|y^k + d^k - \hat{y} + d/\beta_k + \hat{y} - d/\beta_k - \bar{y}\|_\alpha^\mu \\ &\leqq \|y^k + d^k - \hat{y}\|_\alpha^\mu + \|d\|_\alpha^\mu/\beta_k + \|\hat{y} - d/\beta_k - \bar{y}\|_\alpha^\mu\end{aligned}$$

and the first and second terms of the right-hand side converge to 0 as $k \to \infty$. Thus, we have

$$\|y^k - \bar{y}\|_\alpha^\mu < \|\hat{y} - \bar{y}\|_\alpha^\mu$$

for some k sufficiently large, since $y^k + d^k - \bar{y} \geqq y^k - \bar{y} \geqq 0$. This is also a contradiction and therefore $\hat{y} \in \mathscr{P}(Y)$.

($\mathscr{E}(Y) \subset \operatorname{cl} \mathscr{A}(Y)$): Let $\hat{y} \in \mathscr{E}(Y)$. We will prove that for each $\varepsilon > 0$, there exists $y^{(\varepsilon)} \in \mathscr{A}(Y)$ such that $\|y^{(\varepsilon)} - \hat{y}\|_\infty < \varepsilon$. First, there exists $y' > \hat{y}$ such that

$$\|y - \hat{y}\|_\infty < \varepsilon \qquad \text{for any} \quad y \in (y' - R_+^p) \cap Y.$$

In fact, if no such y' exists, then there exists a sequence $\{\hat{y}^k\} \subset R^p$ such that $\hat{y}^k > \hat{y}$, $\hat{y}^k \to \hat{y}$ and for each k, there is a $y^k \in (\hat{y}^k - R_+^p) \cap Y$ such that $\|y^k - \hat{y}\|_\infty \geqq \varepsilon$. Since $Y + R_+^p$ is closed and Y is R_+^p-bounded, we may assume without loss of generality that $y^k \to y^0 + d^0$ with $y^0 \in Y$ and $d^0 \geqq 0$. Then

$$\|y^0 + d^0 - \hat{y}\|_\infty \geqq \varepsilon.$$

However, $y^0 + d^0 \in (\hat{y} - R_+^p) \cap (Y + R_+^p) = \{\hat{y}\}$, since $\hat{y} \in \mathscr{E}(Y)$. Thus, we have a contradiction.

Now, since $\bar{y} \leqq \hat{y} < y'$, there exist some $\mu \in \mathring{S}^p$ and $\beta > 0$ such that

$$y' - \bar{y} = \beta(1/\mu_1, \ldots, 1/\mu_p).$$

Then

$$\mu_i(\hat{y}_i - \bar{y}_i) < \mu_i(y_i' - \bar{y}_i) = \beta, \qquad i = 1, \ldots, p.$$

Hence, $\|\hat{y} - \bar{y}\|_\infty^\mu < \beta$. Let $y(\alpha)$ be a best approximation to $\bar{y}$ out of Y in the (μ, α)-norm. Since $Y + R_+^p$ is closed, $y(\alpha)$ exists. Then, from Property P1,

$$\|y(\alpha) - \bar{y}\|_\infty^\mu \leqq M(\alpha)\|y(\alpha) - \bar{y}\|_\alpha^\mu \leqq M(\alpha)\|\hat{y} - \bar{y}\|_\alpha^\mu.$$

From Property P2, for all α sufficiently large,

$$\|y(\alpha) - \bar{y}\|_\infty^\mu \leqq \beta,$$

whence

$$y_i(\alpha) - \bar{y}_i \leqq \beta/\mu_i = y_i' - \bar{y}_i \qquad \text{for all} \quad i = 1, \ldots, p,$$

i.e., $y(\alpha) \leqq y'$. Thus $y(\alpha) \in (y' - R_+^p) \cap Y$, and so if we put $y^{(\varepsilon)} = y(\alpha)$ for sufficiently large α, there $\|y^{(\varepsilon)} - \hat{y}\|_\infty < \varepsilon$. This completes the proof of the theorem.

Remark 3.4.2

The proof of Theorem 3.4.8 suggests that, in order to generate the efficient set, the value of α should be taken large. However, as has been established as Theorem 3.4.6, it suffices to use the so-called l_1-norm [i.e. $\alpha = 1$ in the norm of Example 3.4.3(i)] only if Y is R_+^p-convex. Possibly, the magnitude of α that is required can be related to a measure of the nonconvexity of Y.

Remark 3.4.3

(i) In Theorem 3.4.8, the converse relationship $\mathscr{E}(Y) \supset \mathrm{cl}\,\mathscr{A}(Y)$ does not necessarily hold. In fact, if

$$Y = \{(y_1, y_2) : (y_1)^2 + (y_2 - 1)^2 \leqq 1\} \cup \{(y_1, y_2) : y_1 \geqq 0,\ y_2 \geqq -1\},$$

and $\bar{y} = (-1, -1)$, then $\hat{y} = (0, 0) \notin \mathscr{E}(Y)$, but $\hat{y} \in \mathrm{cl}\,\mathscr{E}(Y) \subset \mathrm{cl}\,\mathscr{A}(Y)$ (see Fig. 3.14).

(ii) On the other hand, the relationship $\mathscr{A}(Y) \supset \mathscr{P}(Y)$ holds when $\bar{y} < \tilde{y}$ and we use the family of norms (ii) or (iii) in Example 3.4.3. This fact will be shown later in Theorem 3.4.10.

In subsection 3.4.1 we studied the characterization of efficient sets by scalarization. If we use the l_1-norm in Example 3.4.1(i), then

$$S(\mu, Y) = A(\mu, 1, Y).$$

Namely, the procedure of best approximation by the l_1-norm is of particular interest, though it cannot generate all the efficient points without convexity

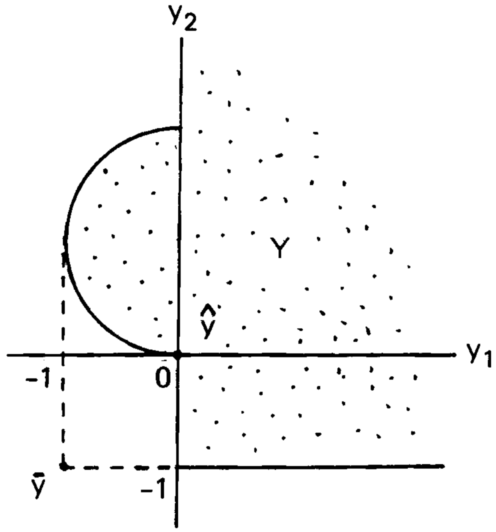

Fig. 3.14. $\mathscr{E}(Y) \not\supset \mathrm{cl}\,\mathscr{A}(Y)$. (Remark 3.4.3(i).)

assumption. We next consider the method of best approximations by the l_∞-norm (Tchebyshev norm). Let

$$A_\infty(\mu, Y) = \{\hat{y} \in Y : \|\hat{y} - \bar{y}\|_\infty^\mu = \inf\{\|y - \bar{y}\|_\infty^\mu : y \in Y\}\}$$

and

$$\mathscr{A}_\infty(Y) = \bigcup_{\mu \in \mathring{S}^p} A_\infty(\mu, Y).$$

Theorem 3.4.9

If Y is nonempty and R_+^p-closed, then

(i) $A_\infty(\mu, Y) \cap \mathscr{E}(Y) \neq \varnothing$ for each $\mu \in \mathring{S}^p$.

(ii) $\mathscr{E}(Y) \subset \mathrm{cl}(\mathscr{A}_\infty(Y) \cap \mathscr{E}(Y))$. Furthermore, if $\bar{y} < \tilde{y}$, then $\mathscr{E}(Y) \subset \mathscr{A}_\infty(Y)$.

Proof

(i) Since Y is R_+^p-closed, $A_\infty(\mu, y)$ is not empty. In fact, let $\{y^k\} \subset Y$ be a sequence such that

$$\|y^k - \bar{y}\|_\infty^\mu \to \inf\{\|y - \bar{y}\|_\infty^\mu : y \in Y\}.$$

Then, we may assume without loss of generality that $y^k \to y^0$. Since $Y + R^p_+$ is closed, $y^0 = y' + d$ for some $y' \in Y$ and $d \in R^p_+$. Then

$$0 \leqq y' - \bar{y} = y^0 - \bar{y} - d \leqq y^0 - \bar{y},$$

$$\|y' - \bar{y}\|^\mu_\infty \leqq \|y^0 - \bar{y}\|^\mu_\infty = \inf\{\|y - \bar{y}\|^\mu_\infty : y \in Y\}.$$

Hence, $y' \in A_\infty(\mu, Y)$. Now, let

$$\delta = \inf\left\{\sum_{i=1}^{p} y_i : y \in A_\infty(\mu, Y)\right\}$$

and let $\{\hat{y}^k\} \subset A_\infty(\mu, Y)$ be a sequence such that $\sum_{i=1}^p \hat{y}^k_i \to \delta$. Then we may assume that $\hat{y}^k$ converges to some $\hat{y}$, since $A_\infty(\mu, Y)$ is bounded. We can prove that $\hat{y} \in A_\infty(\mu, Y) + R^p_+$ as before. Then, it is clear that $\hat{y} \in \mathscr{E}(Y, D) \cap A_\infty(\mu, Y)$.

(ii) Let $\hat{y} \in \mathscr{E}(Y)$ and take an arbitrary $\varepsilon > 0$. Then, as in the proof of the latter half of Theorem 3.4.8, we can find $y' \in R^p$, $\mu \in \mathring{S}^p$, and $\beta > 0$ such that

$$\|y - \hat{y}\|_\infty < \varepsilon \qquad \text{for any} \quad y \in Y \qquad \text{such that} \quad y \leqq y',$$

$$y' - \bar{y} = \beta(1/\mu_1, \ldots, 1/\mu_p),$$

and

$$\|\hat{y} - \bar{y}\|^\mu_\infty < \beta.$$

From the first section of this theorem, there exists $y^\infty \in A_\infty(\mu, Y) \cap \mathscr{E}(Y)$. Then

$$\|y^\infty - \bar{y}\|^\mu_\infty \leqq \|\hat{y} - \bar{y}\|^\mu_\infty < \beta$$

and so $y^\infty \leqq y'$. Therefore $\|y^\infty - \hat{y}\|_\infty < \varepsilon$, which implies that

$$\hat{y} \in \operatorname{cl}(\mathscr{A}_\infty(Y) \cap \mathscr{E}(Y)).$$

Finally, if $\bar{y} < \tilde{y}$, a given $\hat{y} \in \mathscr{E}(Y)$ clearly belongs to the set $A_\infty(\mu, Y)$, where

$$0 < \hat{y} - \bar{y} = \beta(1/\mu_1, \ldots, 1/\mu_p) \qquad \text{with} \quad \beta > 0, \qquad \mu \in \mathring{S}^p.$$

In fact, if there exists $y \in Y$ such that

$$\mu_i(y_i - \bar{y}_i) < \beta = \mu_i(\hat{y}_i - \bar{y}_i) \qquad \text{for every} \quad i = 1, \ldots, p,$$

then $y < \hat{y}$, which is a contradiction, and the proof of the theorem is completed.

Remark 3.4.4

In Theorem 3.4.9(ii), the relationship $\mathscr{E}(Y) \subset \mathscr{A}_\infty(Y)$ (even $\mathscr{P}(Y) \subset \mathscr{A}_\infty(Y)$) may not hold when $\bar{y} \nless \tilde{y}$. In fact, let

$$Y = \{(y_1, y_2) : y_1 + y_2 \geqq 1, 0 \leqq y_1 \leqq 1\} \subset R^2,$$

and $\bar{y} = \tilde{y} = (0, 0)$. Then, $(0, 1) \in \mathscr{P}(Y)$, but $(0, 1) \notin \mathscr{A}_\infty(Y)$ (see Fig. 3.15).

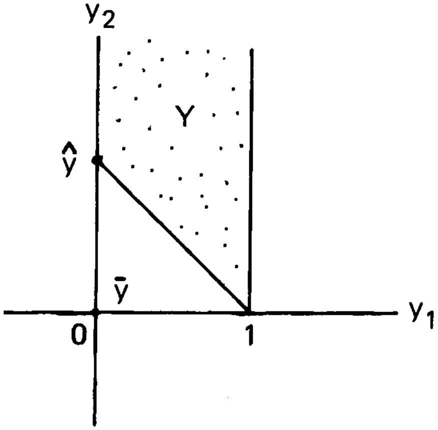

Fig. 3.15. $\mathscr{E}(Y) \not\subset \mathscr{A}_\infty(Y)$. (Remark 3.4.4.)

Remark 3.4.5

$A_\infty(\mu, Y) \subset \mathscr{E}(Y)$ does not hold in general; in other words, the best approximation by Tchebyshev norm is not necessarily efficient. The families of norms in Example 3.4.3(ii) and 3.4.3(iii) may overcome this defect.

Now we will show that properly efficient points are completely characterized as best approximations when $\bar{y} < \tilde{y}$.

Theorem 3.4.10

If Y is R^p_+-closed and $\bar{y} < \tilde{y}$, then

$$\mathscr{A}(Y) = \mathscr{P}(Y)$$

when we use the family of norms in Example 3.4.3(ii) or 3.4.3(iii).

Proof The inclusion $\mathscr{A}(Y) \subset \mathscr{P}(Y)$ has been already proved in Theorem 3.4.8; so we now prove that $\mathscr{P}(Y) \subset \mathscr{A}(Y)$. Let $\hat{y} \in \mathscr{P}(Y) \subset \mathscr{E}(Y)$. In view of the last part of the proof of Theorem 3.4.9(ii), $\hat{y}$ is a unique point in $A_\infty(\mu, Y)$

Let $\{\alpha_k\}$ be a sequence diverging to $+\infty$. If $\hat{y} \notin \mathscr{A}(Y)$, we can take a sequence $\{y^k\} \subset Y$ such that

$$\|\hat{y} - \bar{y}\|^{\mu}_{\alpha_k} > \|y^k - \bar{y}\|^{\mu}_{\alpha_k} \qquad \text{for all} \quad k = 1, 2, \ldots .$$

Since

$$M(\alpha_k)\|\hat{y} - \bar{y}\|^{\mu}_{\alpha_k} > M(\alpha_k)\|y^k - \bar{y}\|^{\mu}_{\alpha_k} \geqq \|y^k - \bar{y}\|^{\mu}_{\infty}$$

by Property P1 and the left-hand side converges to $\|\hat{y} - \bar{y}\|^{\mu}_{\infty}$ as $k \to \infty$ by Property P2, then $\{y^k\}$ is a bounded sequence. Therefore, we may assume without loss of generality that $y^k \to y$ for some y. Since Y is R^p_+-closed, $y = y' + d$ for $y' \in Y$ and $d \in R^p_+$. Then, by taking the limit of

$$M(\alpha_k)\|\hat{y} - \bar{y}\|^{\mu}_{\alpha_k} > \|y^k - \bar{y}\|^{\mu}_{\infty}$$

as $k \to \infty$, we have

$$\|\hat{y} - \bar{y}\|^{\mu}_{\infty} \geqq \|y' + d - \bar{y}\|^{\mu}_{\infty} \geqq \|y' - \bar{y}\|^{\mu}_{\infty}$$

by Property P3. Since $\hat{y}$ is an unique point in $A_{\infty}(\mu, Y)$, $y = y' = \hat{y}$; in other words, $y^k \to \hat{y}$. Let $d^k = \beta(y^k - \hat{y})/\|y^k - \hat{y}\|^{\mu}_{\infty}$. Then $\hat{y} + d^k \geqq \bar{y}$ and we can prove that

$$\|\hat{y} + d^k - \bar{y}\|^{\mu}_{\alpha_k} < \|\hat{y} - \bar{y}\|^{\mu}_{\alpha_k}.$$

In fact, when we use the family of norms (ii) of Example 3.4.3,

$$\begin{aligned}
\|\hat{y} + d^k - \bar{y}\|^{\mu}_{\alpha_k} &= \|\hat{y} - \bar{y} + d^k\|^{\mu}_{\infty} + \frac{1}{\alpha_k}\sum_i \mu_i(\hat{y}_i - \bar{y}_i + d^k_i) \\
&= \|\hat{y} - \bar{y}\|^{\mu}_{\alpha_k} + \beta\Big[\max_i\{\mu_i(y^k_i - \hat{y}_i)\} \\
&\qquad + \frac{1}{\alpha_k}\sum_i \mu_i(y^k_i - \hat{y}_i)\Big]\Big/\|y^k - \hat{y}\|^{\mu}_{\infty}.
\end{aligned}$$

Since

$$\begin{aligned}
\|\hat{y} - \bar{y}\|^{\mu}_{\alpha_k} > \|y^k - \bar{y}\|^{\mu}_{\alpha_k} &= \|y^k - \hat{y} + \hat{y} - \bar{y}\|^{\mu}_{\alpha_k} \\
&= \|\hat{y} - \bar{y}\|^{\mu}_{\alpha_k} + \Big[\max_i\{\mu_i(y^k_i - \hat{y}_i)\} \\
&\qquad + \frac{1}{\alpha_k}\sum_i \mu_i(y^k_i - \hat{y}_i)\Big],
\end{aligned}$$

the second term of the last equation is negative and therefore

$$\|\hat{y} + d^k - \bar{y}\|^{\mu}_{\alpha_k} < \|\hat{y} - \bar{y}\|^{\mu}_{\alpha_k}.$$

Similarly, when we use the family of norms (iii),

$$\|\hat{y} + d^k - \bar{y}\|^{\mu}_{\alpha_k} = \|I^{-1}_{\alpha_k}(\mu \odot (\hat{y} - \bar{y})) + I^{-1}_{\alpha_k}(\mu \odot d^k)\|_{\infty}.$$

Since

$$\begin{aligned}\|y^k - \bar{y}\|^{\mu}_{\alpha_k} &= \|I^{-1}_{\alpha_k}(\mu \odot (\hat{y} - \bar{y})) + I^{-1}_{\alpha_k}(\mu \odot (y^k - \hat{y}))\|_{\infty} \\ &< \|\hat{y} - \bar{y}\|^{\mu}_{\alpha_k} = \|I^{-1}_{\alpha_k}(\mu \odot (\hat{y} - \bar{y}))\|_{\infty}\end{aligned}$$

and

$$I^{-1}_{\alpha_k}(\mu \odot (\hat{y} - \bar{y})) = (\alpha_k\beta/(\alpha_k - 1), \ldots, \alpha_k\beta/(\alpha_k - 1)),$$

we have

$$I^{-1}_{\alpha_k}(\mu \odot (y^k - \hat{y})) < 0.$$

Hence

$$\|\hat{y} + d^k - \bar{y}\|^{\mu}_{\alpha_k} < \|\hat{y} - \bar{y}\|^{\mu}_{\alpha_k}.$$

Now, since $\|d^k\|^{\mu}_{\infty} = \beta$ for any k, we may assume that $\{d^k\}$ converges to some $\hat{d}$ with $\|\hat{d}\|^{\mu}_{\infty} = \beta$. Then, taking the limit of

$$\|\hat{y} + d^k - \bar{y}\|^{\mu}_{\infty} \leqq M(\alpha_k)\|\hat{y} + d^k - \bar{y}\|^{\mu}_{\alpha_k} < M(\alpha_k)\|\hat{y} - \bar{y}\|^{\mu}_{\alpha_k},$$

we have

$$\|\hat{y} + \hat{d} - \bar{y}\|^{\mu}_{\infty} \leqq \|\hat{y} - \bar{y}\|^{\mu}_{\infty}.$$

This implies that $\hat{d} \leqq 0$, since $\hat{y} + \hat{d} \geqq \bar{y}$ and $\mu \odot (\hat{y} - \bar{y}) = (\beta, \ldots, \beta)$. On the other hand, from the definition of $\{d^k\}$, $\hat{d} \in \operatorname{cl} P(Y + R^p_+ - \hat{y})$. However, this contradicts the assumption $\hat{y} \in \mathscr{P}(Y)$ and the proof of the theorem is complted.

3.4.3 Characterization as Solution of Constraint Problems

Another method of characterizing Pareto optimal solutions is to solve kth objective constraint problems associated with the original multiobjective optimization problem. The kth objective constraint problem is formulated by taking the kth objective function f_k as the objective function and letting all the other objective functions f_j ($j \neq k$) be inequality constraints (Benson and Morin [B10], Haimes *et al.* [H2], Chankong and Haimes [C5, C6]). That is, given the original multiobjective optimization problem

(P) minimize $f(x) = (f_1(x), \ldots, f_p(x)$ subject to $x \in X \subset R^n$, the kth objective constraint problem can be defined as a scalar optimization problem

$$(\mathrm{P}_k(\varepsilon)) \qquad \begin{array}{ll} \text{minimize} & f_k(x) \\ \text{subject to} & f_i(x) \leqq \varepsilon_i, \qquad i = 1, \ldots, p, \quad i \neq k, \quad x \in X, \end{array}$$

where $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_p) \in R^p$. A similar type of equality-constrained problems was studied by Lin [L6–L8].

The following theorems provide the relationships between the Pareto optimal solutions of the original multiobjective optimization problem (P) and the solutions of the kth objective constraint problems $(\mathrm{P}_k(\varepsilon))$.

Theorem 3.4.11

A point $\hat{x}$ is a Pareto optimal solution of (P) if and only if $\hat{x}$ solves $(\mathrm{P}_k(\hat{\varepsilon}))$ for every $k = 1, \ldots, p$, where $\hat{\varepsilon} = f(\hat{x})$.

Proof If $\hat{x}$ is not a Pareto optimal solution of (P), there exists $x \in X$ such that $f_i(x) \leqq f_i(\hat{x})$ with the strict inequality holding for at least one k. This implies that $\hat{x}$ does not solve $(P_k(\hat{\varepsilon}))$. Conversely if $\hat{x}$ does not solve $(\mathrm{P}_k(\hat{\varepsilon}))$ for some k, then there exists $x \in X$ such that $f_k(x) < f_k(\hat{x})$ and $f_i(x) \leqq f_i(\hat{x})$ $(i \neq k)$, implying that $\hat{x}$ is not a Pareto optimal solution of (P).

Theorem 3.4.12

If $\hat{x}$ is a unique solution of $(\mathrm{P}_k(\hat{\varepsilon}))$ for some $k \in \{1, \ldots, p\}$ with $\hat{\varepsilon} = f(\hat{x})$, then $\hat{x}$ is a Pareto optimal solution of (P).

Proof Since $\hat{x}$ uniquely minimizes $(\mathrm{P}_k(\hat{\varepsilon}))$, for all x satisfying $f_i(x) \leqq f_i(\hat{x})$ $(i \neq k), f_k(x) > f_k(\hat{x})$. Hence $\hat{x}$ is a Pareto optimal solution of (P).

As has been already stated, every properly efficient solution (in the sense of Geoffrion) is a Pareto optimal solution, but not vice versa in general. Benson and Morin showed that the converse holds when the constraint problems are stable as scalar optimization problems.

Definition 3.4.2 (*Perturbation Function*)

The perturbation function $w_k \colon R^{p-1} \to \overline{R} = [-\infty, +\infty]$ associated with $(\mathrm{P}_k(\varepsilon))$ is defined as

$$w_k(u) = \inf\{f_k(x) : f_i(x) - \varepsilon_i \leqq u_i,\ i = 1, \ldots, k-1, k+1, \ldots, p,\ x \in X\},$$

where $u = (u_1, \ldots, u_{k-1}, u_{k+1}, \ldots, u_p) \in R^{p-1}$.

Definition 3.4.3 (*Stability*)

($P_k(\varepsilon)$) is said to be stable when $w_k(0)$ is finite and when there exists a scalar $M > 0$ such that, for all $u \neq 0$,

$$[w_k(0) - w_k(u)]/\|u\| \leqq M.$$

Theorem 3.4.13

Assume that $f_1(x), \ldots, f_p(x)$ are convex functions on the nonempty convex set X and that $\hat{x}$ is a Pareto optimal (i.e. efficient) solution of the multiobjective optimization problem (P). Then $\hat{x}$ is a properly efficient solution of (P) if and only if ($P_k(\hat{\varepsilon})$) is stable for each $k = 1, \ldots, p$, where $\hat{\varepsilon} = f(\hat{x})$.

Proof Assume first that ($P_k(\hat{\varepsilon})$) is stable, and then define the dual problem of ($P_k(\hat{\varepsilon})$) by

$$(D_k(\hat{\varepsilon})) \qquad \varphi_k = \max_{\mu \geqq 0}\left\{\min_{x \in X}\left[f_k(x) + \sum_{i \neq k} \mu_i(f_i(x) - \hat{\varepsilon}_i)\right]\right\}$$

where μ is a $(p - 1)$-vector of dual variables. Since ($P_k(\hat{\varepsilon})$) is stable, adapting a result of Geoffrion [G6, Theorem 3], we find that ($D_k(\hat{\varepsilon})$) has an optimal solution $\hat{\mu}$ and $\hat{x}$ is an optimal solution for the problem

$$(P'_{k,\mu}) \qquad \text{miminize} \quad f_k(x) + \sum_{i \neq k} \hat{\mu}_i(f_i(x) - \hat{\varepsilon}_i) \qquad \text{subject to} \quad x \in X.$$

Since $-\sum \hat{\mu}_i \hat{\varepsilon}_i$ is constant for all $x \in X$, $\hat{x}$ is an optimal solution of the problem

$$(P_{k,\hat{\mu}}) \qquad \text{minimize} \quad f_k(x) + \sum_{i \neq k} \hat{\mu}_i f_i(x) \qquad \text{subject to} \quad x \in X.$$

By Lemma 3.4.1 below, $\hat{x}$ is a properly efficient solution of (P).

Conversely, suppose that $\hat{x}$ is a properly efficient solution of (P). Then, for each $k = 1, \ldots, p$ there exists (by Lemma 3.4.1) a $\hat{\mu} \in R_+^{p-1}$ such that $\hat{x}$ is an optimal solution of ($P_{k,\hat{\mu}}$). This implies that $\hat{x}$ is an optimal solution of ($P'_{k,\hat{\mu}}$), where ($P'_{k,\hat{\mu}}$) is given above. Thus $(\hat{x}, \hat{\mu})$ is feasible in ($D_k(\hat{\varepsilon})$), and so $\varphi_k \geqq f_k(\hat{x})$. Furthermore, since $\hat{x}$ is an optimal solution of ($P_{k,\hat{\mu}}$) with the optimal objective value $f_k(\hat{x})$, then $\varphi_k \leqq f_k(\hat{x})$ from the weak duality theorem for nonlinear programming. Thus, $\varphi_k = f_k(\hat{x})$, and $(\hat{x}, \hat{\mu})$ is an optimal solution to ($D_k(\hat{\varepsilon})$). By Geoffrion [G6], ($P_k(\hat{\varepsilon})$) is stable. This completes the proof of the theorem.

Lemma 3.4.1

Suppose that $f_1(x), \ldots, f_p(x)$ are convex functions on the convex set X. Then $\hat{x}$ is a properly efficient solution for (P) if and only if, for each k, there exists a $\mu = (\mu_1, \ldots, \mu_{k-1}, \mu_{k+1}, \ldots, \mu_p) \in R_+^{p-1}$ such that $\hat{x}$ is an optimal solution of the problem $(\mathrm{P}_{k,\mu})$.

Proof This result is obvious from Theorems 3.4.1 and 3.4.2.

Remark 3.4.2

Though stability may not be an easy property to demonstrate directly, it holds whenever some constraint qualification (e.g. Slater's constraint qualification) is satisfied. (See Geoffrion [G6].)

Details concerning the constraint problems can be found, for example, in Chankong and Haimes [C6]. The interested reader may refer to it.

3.5 Kuhn–Tucker Conditions for Multiobjective Programming

In this section we deal with the Kuhn–Tucker condition for a multiobjective programming problem

(P) minimize $f(x) = (f_1(x), \ldots, f_p(x))^{\mathrm{T}}$

subject to $x \in X = \{x \in R^n : g(x) = (g_1(x), \ldots, g_m(x))^{\mathrm{T}} \leqq 0\}$.

Here the domination structure is given by a convex cone D, in particular R_+^p. The necessary and/or sufficient conditions for efficiency (or Pareto optimality) in this problem have been studied by a number of researchers (Kuhn and Tucker [K10], Hurwicz [H16], DaCunha and Polak [D1], Neustadt [N14], Ritter [R4–R6], Smale [S10, S11], Wan [W3], Tamura and Miura [T4]). In this section we derive necessary/sufficient conditions for Pareto optimality under the assumption that every f_i $(i = 1, \ldots, p)$ and g_j $(j = 1, \ldots, m)$ is continuously differentiable. Results under the convexity assumption are closely related to the duality theory and so will be discussed fully in later chapters.

Theorem 3.5.1

A necessary condition for $\hat{x} \in X$ to be a properly efficient solution to (P) in the sense of Kuhn–Tucker is that there exist $\hat{\mu} \in R^p$ and $\hat{\lambda} \in R^m$ such that

(i) $\langle \hat{\mu}, \nabla f(\hat{x}) \rangle + \langle \hat{\lambda}, \nabla g(\hat{x}) \rangle = 0.$
(ii) $\langle \hat{\lambda}, g(\hat{x}) \rangle = 0,$
(iii) $\hat{\mu} > 0,\ \hat{\lambda} \geqq 0,$

where $\langle \hat{\mu}, \nabla f(\hat{x}) \rangle$ and $\langle \hat{\lambda}, \nabla g(\hat{x}) \rangle$ stand for $\sum_{i=1}^p \hat{\mu}_i \nabla f_i(\hat{x})$ and $\sum_{j=1}^m \hat{\lambda}_j \nabla g_j(\hat{x})$, respectively.

Proof From Kuhn–Tucker's definition of proper efficiency (Definition 3.1.10), there is no $h \in R^n$ such that $\langle \nabla f_i(\hat{x}), h \rangle \leqq 0$ $(i = 1, \ldots, p)$ with the strict inequalities for at least one i and that

$$\langle \nabla g_j(\hat{x}), h \rangle \leqq 0, \qquad j \in J(\hat{x}) = \{j : g_j(\hat{x}) = 0\}.$$

Then, from the Tucker's theorem of alternatives (Mangasarian [M2]), there exist $\hat{\mu}_i > 0$ $(i = 1, \ldots, p)$ and $\hat{\lambda}_j \geqq 0$ $(j \in J(\hat{x}))$ such that

$$\sum_{i=1}^{p} \hat{\mu}_i \nabla f_i(\hat{x}) + \sum_{j \in J(x)} \hat{\lambda}_j \nabla g_j(\hat{x}) = 0$$

Letting $\hat{\lambda}_j = 0$ for $j \notin J(\hat{x})$, we can immediately establish the theorem.

Corollary 3.5.1

If the Kuhn–Tucker constraint qualification is satisfied at $\hat{x} \in X$, a necessary condition for $\hat{x}$ to be a properly efficient solution to (P) in the sense of Geoffrion is the condition in Theorem 3.5.1.

Proof Immediate from Theorem 3.1.6 and 3.5.1.

Theorem 3.5.2

If every function f_i $(i = 1, \ldots, p)$ and g_j $(j = 1, \ldots, m)$ is convex, the condition in Theorem 3.5.1 is also sufficient for $\hat{x} \in X$ to be a properly efficient solution to (P).

Proof The condition in Theorem 3.5.1 implies that $f(\hat{x}) \in S(\hat{\mu}, Y)$ (see Section 3.4.1) and hence, from Theorem 3.4.1, $\hat{x}$ is a properly efficient solution of (P).

Theorem 3.5.3

Suppose that (P) satisfies the Kuhn–Tucker constraint qualification at $\hat{x} \in X$. Then, a necessary condition for $\hat{x}$ to be a weak Pareto optimal solution to (P) is that there exist $\hat{\mu} \in R^p$ and $\hat{\lambda} \in R^m$ such that

(i) $\langle \hat{\mu}, \nabla f(\hat{x}) \rangle + \langle \hat{\lambda}, \nabla g(\hat{x}) \rangle = 0,$
(ii) $\langle \hat{\lambda}, g(\hat{x}) \rangle = 0,$
(iii) $\hat{\mu} \geq 0, \hat{\lambda} \geqq 0.$

Proof Let $\hat{x}$ be a weak Pareto optimal solution of (P). We prove that there is no h such that

$$\langle \nabla f_i(\hat{x}), h \rangle < 0 \qquad i = 1, \ldots, p,$$
$$\langle \nabla g_j(\hat{x}), h \rangle \leqq 0 \qquad j \in J(\hat{x}) = \{j : g_j(\hat{x}) = 0\}.$$

Then the theorem follows from Motzkin's theorem of alternatives (Mangasarian [M2]). Suppose, to the contrary, that there exists an h satisfying the above inequalities. From the Kuhn–Tucker constraint qualification, there exists a continuously differentiable arc $\theta(t)$ $(0 \leqq t \leqq \bar{t})$ such that $\theta(0) = x$, $g(\theta(t)) \leqq 0$ and $\dot{\theta}(0) = \alpha h$ $(\alpha > 0)$ as

$$f_i(\theta(t)) = f_i(\hat{x}) + t\langle \nabla f_i(\hat{x}), \alpha h\rangle + o(t).$$

Since $\langle \nabla f_i(\hat{x}), h\rangle < 0$ for all $i = 1, \ldots, p$, then $f_i(\theta(t)) < f_i(\hat{x})$ for t sufficiently small. This contradicts the weak Pareto optimality of $\hat{x}$, as was to be proved.

Theorem 3.5.4

If every function f_i $(i = 1, \ldots, p)$ and g_j $(j = 1, \ldots, m)$ is convex, the condition in Theorem 3.5.3 is also sufficient for $\hat{x} \in X$ to be a weak Pareto optimal solution to (P).

Proof The condition in Theorem 3.5.3 implies that $f(\hat{x}) \in S(\hat{\mu}, Y)$, where $\hat{\mu} \in R^p_+ \setminus \{0\}$. Since, in view of Corollary 3.4.1, $S(\hat{\mu}, Y) \subset \mathscr{E}(Y, \operatorname{int} R^p_+)$ under the convexity assumption, $\hat{x}$ is a weak Pareto optimal solution of (P).

Remark 3.5.1

Tamura and Miura [T4] obtained efficiency conditions for the case where the domination cone D is a polyhedral convex cone. However, that case can be reduced to the case $D = R^p_+$ (cf. Lemma 2.3.4). In fact, if D is given as

$$D = \{d \in R^p : \langle b^i, d\rangle \geqq 0,\ i = 1, \ldots, l\},$$

then $f(x) - f(x') \in D$ for $x, x' \in X$ if and only if

$$\langle b^i, f(x) - f(x')\rangle \geqq 0, \qquad i = 1, \ldots, l,$$

i.e., if and only if

$$\langle b^i, f(x)\rangle - \langle b^i, f(x')\rangle \geqq 0 \qquad \text{for} \quad i = 1, \ldots, l.$$

Therefore, by putting

$$\bar{f}_i(x) = \langle b^i, f(x)\rangle \qquad \text{for} \quad i = 1, \ldots, l,$$

we can essentially identify the original problem (P) with a new l-objective optimization problem

$$\text{minimize} \quad \bar{f}(x) = (\bar{f}_1(x), \ldots, \bar{f}_p(x)) \qquad \text{subject to} \quad x \in X.$$

Remark 3.5.2

In the above discussion we assume the continuous differentiability of the functions. Recent development of nonsmooth analysis is enabling us to obtain a similar type of efficiency condition with nondifferentiable functions. (See, for example, Clarke [C11], Minami [M4].)

4 STABILITY

In this chapter we consider the stability of solution sets of multiobjective optimization problems with respect to perturbation of feasible solution sets, objective functions, and domination structures of the decision maker. Several sufficient conditions that ensure the semicontinuity of the solution set maps will be provided along with a series of examples.

4.1 Families of Multiobjective Optimization Problems

Stability analysis in nonlinear programming has been studied by several authors (Berge [B11], Dantzig *et al.* [D2], Evans and Gould [E2], Hogan [H11], Fiacco [F4]). The most fundamental results are concerned with the upper and lower semicontinuity† of the optimal value function (perturbation function)

$$w(u) = \inf\{f(x, u) : x \in X(u)\},$$

where u is a parameter vector in a space U, X a point-to-set map from U into R^n, and f a scalar-valued function on $R^n \times U$. The following results are well known.

Theorem 4.1.1‡

(i) If X is upper semicontinuous at $\hat{u}$ and uniformly compact near $\hat{u}$, and if f is lower semicontinuous on $X(\hat{u}) \times \hat{u}$, then w is lower semicontinuous at $\hat{u}$.

† We must discriminate between upper (or lower) semicontinuity of functions and that of point-to-set maps.

‡ Theorems 4.1.1 and 4.1.2 are from Hogan [H11].

(ii) If X is lower semicontinuous at $\hat{u}$ and f is upper semicontinuous on $X(\hat{u}) \times \hat{u}$, then w is upper semicontinuous at $\hat{u}$.

Some results are also obtained concerning the continuity of the point-to-set map

$$M(u) = \{x \in X(u) : f(x, u) = w(u)\},$$

which is a solution set map in the decision space.

Theorem 4.1.2

(i) If X is continuous at $\hat{u}$, and if f is continuous on $X(\hat{u}) \times \hat{u}$, then M is upper semicontinuous at $\hat{u}$.

(ii) If X is continuous at $\hat{u}$, f is continuous on $X(\hat{u}) \times \hat{u}$, M is nonempty and uniformly compact near $\hat{u}$, and $M(\hat{u})$ is single-valued, then M is continuous at $\hat{u}$.

In order to extend the above results to multiobjective optimization, we will define a multiobjective optimization problem having two kinds of parameters in this section. One parameter u, which varies over a set U, specifies the set of possible decisions (or feasible solutions) and the objective function as before. Another parameter v, which varies over a set V, specifies the domination structure of the decision maker in the objective space. Here, U and V may be arbitrary subsets of linear topological spaces in which the concept of convergence is defined in terms of sequences. However, in this book, we assume that U and V are subsets of the Euclidean spaces for simplicity.

Now, we can define a parametrized multiobjective optimization problem in the following way. First, the set of all feasible solutions and the vector-valued function, which changes depending on the parameter u are denoted by $X(u)$ and $f(x, u)$, respectively. These determine the set of feasible solutions in the objective space as

$$Y(u) = \{y \in R^p : y = f(x, u),\ x \in X(u)\}.$$

Second, the domination structure of the decision maker in the objective space is specified by the parameter v as a point-to-set map $D(v)$ from R^p into R^p. Thus, the aim of our multiobjective optimization problem is to find the set $N(u, v) := \mathscr{E}(Y(u), D(v))$ and the set

$$M(u, v) := \{x \in X(u) : f(x, u) \in N(u, v)\},$$

which are called the solution sets in the objective space and in the decision space, respectively.

In view of the preceding setting, we can consider the stability theory for this parametrized multiobjective optimization problem from the following two different points of view. The first one is concerned with the stability of the efficient sets $N(u, v)$ and $M(u, v)$ in the case in which the parameter u varies; that is, in which the set of feasible solutions changes. We use the following notations to investigate this case. When the parameter v is fixed at some nominal value $\hat{v} \in V$, we define a point-to-point set map N_1 from U into R^p by

$$N_1(u) = N(u, \hat{v}).$$

Analogously, a point-to-set map M_1 from U into R^n is defined by

$$M_1(u) = M(u, \hat{v}) = \{x \in X(u) : f(x, u) \in N_1(u)\}.$$

We will analyze the continuity of N_1 and M_1 in Sections 4.2 and 4.4, respectively.

The second point of view is concerned with the stability of the solution sets in the case in which the parameter v varies; that is, in which the domination structure of the decision maker changes. To investigate this case, we also use the following notations. When the parameter u is fixed at some nominal value $\hat{u} \in U$, we define a point-to-set map N_2 from V into R^p by

$$N_2(v) = N(\hat{u}, v)$$

and a point-to-set map M_2 from V into R^n by

$$M_2(v) = M(\hat{u}, v) = \{x \in X(\hat{u}) : f(x, \hat{u}) \in N_2(v)\}.$$

We will analyze the continuity of N_2 and M_2 in Sections 4.3 and 4.4, respectively.

In order to simplify the notation, the fixed set of feasible solutions $Y(\hat{u})$ and the fixed domination structure $D(\hat{v})$ will be denoted by Y and D, respectively, hereafter in this chapter.

The results in this chapter are mainly from Tanino and Sawaragi [T10]. Naccache [N2] developed stability results in connection with scalarization, and Jurkiewicz [J3] considered the stability of compromise solutions.

4.2 Stability for Perturbation of the Set of Feasible Solutions

This section is concerned with the stability from the first point of view described in the previous section. Namely, the stability of the solution set will be investigated when the set of feasible solutions specified by a parameter vector u is perturbed. It is a generalization of the study of the continuity of optimal value functions (or perturbation functions) in ordinary scalar optimization problems. In fact, if the objective function f is one-dimensional (scalar-valued), then the map N_1 is at most single-valued and so might be identified with the optimal value function

$$w(u) := \inf\{f(x, u) : x \in X(u)\}.$$

The continuity, especially lower semicontinuity, of this function is closely related to the duality in scalar optimization.

In this section the domination structure is fixed and so is considered to be a point-to-set map from

$$Y(U) := \bigcup_{u \in U} Y(u)$$

into R^p. Moreover, the point-to-set map int D, which is defined by

$$(\operatorname{int} D)(y) = \operatorname{int} D(y) \cup \{0\} \qquad \text{for every} \quad y,$$

is considered to provide another domination structure.

4.2.1 Upper Semicontinuity of the Map N_1

First, we consider sufficient conditions for the upper semicontinuity of the map N_1. The following lemma is useful in this chapter.

Lemma 4.2.1

Let Ω be a point-to-set map from a set Y into R^p. Suppose that Ω is lower semicontinuous at $\hat{y} \in Y$ and that $\Omega(y)$ is convex for every $y \in Y$ near $\hat{y}$. Let $\{y^k\} \subset Y$ and $\{z^k\} \subset R^p$ be sequences that converge to $\hat{y}$ and $\hat{z}$, respectively, and suppose that $\hat{z} \in \operatorname{int} \Omega(\hat{y})$. Then $z^k \in \Omega(y^k)$ except for a finite number of k's.

Proof We denote the closed unit ball in R^p by B. As $\hat{z} \in \operatorname{int} \Omega(\hat{y})$, there exists a positive real number δ such that $\hat{z} + \delta B \subset \Omega(\hat{y})$. Since $z^k \to \hat{z}$, there is a number m such that

$$z^k \in \hat{z} + \operatorname{int}(\delta/2)B \qquad \text{for all} \quad k \geqq m.$$

Let $k \geqq m$, and suppose that $z^k \notin \Omega(y^k)$. Here we may assume that $\Omega(y^k)$ is a convex set, since $\Omega(y)$ is convex for any y near $\hat{y}$. Then the separation theorem of convex sets (Theorem 2.1.1) guarantees the existence of a vector $\lambda \in R^p$ such that $\|\lambda\| = 1$ and

$$\Omega(y^k) \subset \{z \in R^p : \langle \lambda, z \rangle \leqq \langle \lambda, z^k \rangle\}.$$

Then the vector

$$\bar{z}^k = \hat{z} + (\delta/2 + \langle \lambda, z^k - \hat{z}\rangle)\lambda$$

satisfies

$$\bar{z}^k \in \hat{z} + \operatorname{int} \delta B \qquad \text{and} \qquad \bar{z}^k \notin \Omega(y^k) + \operatorname{int}(\delta/2)B.$$

In fact,

$$\begin{aligned}\|\bar{z}^k - \hat{z}\| &= |\delta/2 + \langle \lambda, z^k - \hat{z}\rangle| \, \|\lambda\| \leqq \delta/2 + |\langle \lambda, z^k - \hat{z}\rangle| \\ &\leqq \delta/2 + \|\lambda\| \, \|z^k - \hat{z}\| < \delta/2 + \delta/2 = \delta\end{aligned}$$

and

$$\langle \lambda, \bar{z}^k\rangle = \langle \lambda, \hat{z}\rangle + \delta/2 + \langle \lambda, z^k - \hat{z}\rangle = \delta/2 + \langle \lambda, z^k\rangle;$$

whence, for any $z \in \Omega(y^k)$,

$$\|\bar{z}^k - z\| \geqq |\langle \lambda, \bar{z}^k - z\rangle| = |(\langle \lambda, \bar{z}^k\rangle - \langle \lambda, z^k\rangle) + (\langle \lambda, z^k\rangle - \langle \lambda, z\rangle)| \geqq \delta/2.$$

Thus, if we assume that $z^k \notin \Omega(y^k)$ for an infinite number of k's, then we can take an infinite number of $\bar{z}^k$ satisfying the above properties. The set of those points $\bar{z}^k$ has a cluster point in the compact set $\hat{z} + \delta B$, which is denoted by $\bar{z}$. By taking a subsequence if necessary, we may assume without loss of generality that $\bar{z}^k$ exists for every $k \geqq m$ and $\bar{z}^k \to \bar{z}$. Since $\bar{z} \in \hat{z} + \delta B \subset \Omega(\hat{y})$ and Ω is lower semicontinuous at $\hat{y}$, there exist a number m' and a sequence $\{\tilde{z}^k\}$ such that $\tilde{z}^k \in \Omega(y^k)$ for $k \geqq m'$ and $\tilde{z}^k \to \bar{z}$. Then $\bar{z}^k - \tilde{z}^k \to 0$. This, however, contradicts the fact that

$$\bar{z}^k \notin \Omega(y^k) + \operatorname{int}(\delta/2)B.$$

Therefore, the number of k's for which $z^k \notin \Omega(y^k)$ is finite. This completes the proof.

The following theorem provides sufficient conditions under which the map N_1 is upper semicontinuous.

Theorem 4.2.1

The map N_1 is upper semicontinuous at $\hat{u} \in U$ if the following conditions are satisfied:

(i) the map Y is continuous at $\hat{u}$;
(ii) the map D is lower semicontinuous on $Y(\hat{u})$;
(iii) $D(y)$ is a convex set for every $y \in S \cap Y(u)$, with $u \in T$, where S and T are some neighborhoods of $Y(\hat{u})$ and $\hat{u}$, respectively;
(iv) the relationship $N_1(\hat{u}) = \mathscr{E}(Y(\hat{u}), \operatorname{int} D)$ holds.

Proof Let

$$\{u^k\} \subset U, \qquad u^k \to \hat{u}, \qquad y^k \in N_1(u^k), \qquad \text{and} \qquad y^k \to \hat{y}.$$

We must show that $\hat{y} \in N_1(\hat{u})$. First, note that $\hat{y} \in Y(\hat{u})$, since the map Y is upper semicontinuous at $\hat{u}$. Hence, if we assume (to the contrary) that $\hat{y}$ does not belong to the set

$$N_1(\hat{u}) = \mathscr{E}(Y(\hat{u}), \operatorname{int} D) \qquad \text{(condition (iv))},$$

there exist an element $\bar{y} \in Y(\hat{u})$ and a nonzero vector $d \in \operatorname{int} D(\bar{y})$ such that

$$\hat{y} = \bar{y} + d.$$

Since Y is lower semicontinuous at $\hat{u}$, there exist a number m and a sequence $\{\bar{y}^k\} \subset R^p$ such that $\bar{y}^k \in Y(u^k)$ for $k \geqq m$ and $\bar{y}^k \to \bar{y}$. This and the convergence $y^k \to \hat{y}$ imply that

$$y^k - \bar{y}^k \to \hat{y} - \bar{y} = d.$$

Hence we can apply Lemma 4.2.1 to obtain

$$y^k - \bar{y}^k \in D(\bar{y}^k),$$

except for a finite number of k's. Thus,

$$y^k \in \bar{y}^k + D(\bar{y}^k) \qquad \text{and} \qquad y^k \neq \bar{y}^k \qquad \text{for some} \quad k,$$

which contradicts the fact that $y^k \in N_1(u^k)$. Therefore, $\hat{y} \in N_1(\hat{u})$; that is, N_1 is upper semicontinuous at $\hat{u}$. This completes the proof of the theorem.

We will give some examples to illustrate that each condition in the above theorem is essential for assuring the upper semicontinuity of the point-to-set map N_1 at $\hat{u}$. Namely, we will illustrate that the lack of only one of the conditions might violate the upper semicontinuity of N_1. For each example, the condition that does not hold is indicated in the parenthesis. All the other conditions are satisfied.

Example 4.2.1 (Upper Semicontinuity of Y at $\hat{u}$—Fig. 4.1)

Let

$$U = \{u \in R : u < 0\}, \qquad \hat{u} = -1,$$

$$Y(u) = \{(y_1, y_2) : y_2 \geqq uy_1,\ y_1 \leqq 1,\ y_2 \leqq 1\} \subset R^2, \qquad \text{for } u < 0,\ u \neq -1,$$

$$Y(-1) = \{(y_1, y_2) : y_2 \geqq -y_1 + 1,\ y_1 \leqq 1,\ y_2 \leqq 1\},$$

$$D(y) = R^2_+, \qquad \text{for all} \quad y \in Y(U).$$

Then,

$$N_1(u) = \{(y_1, y_2) : y_2 = uy_1,\ 1/u \leqq y_1 \leqq 1\} \qquad \text{for} \quad u < 0, \quad u \neq -1,$$

$$N_1(-1) = \{(y_1, y_2) : y_2 = -y_1 + 1,\ 0 \leqq y_1 \leqq 1\}.$$

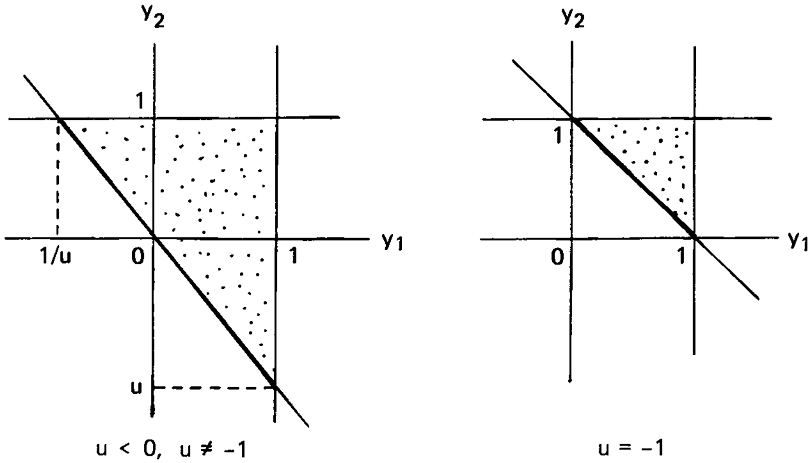

Fig. 4.1. Example 4.2.1.

Example 4.2.2 (Lower Semicontinuity of Y at $\hat{u}$—Fig. 4.2)

Let

$$U = \{u \in R : u < 0], \qquad \hat{u} = -1,$$

$$Y(u) = \{(y_1, y_2) : y_2 \geqq uy_1,\ y_1 \leqq 1,\ y_2 \leqq 1\} \subset R^2 \qquad \text{for} \quad u < 0, \quad u \neq -1$$

$$Y(-1) = \{(y_1, y_2) : y_2 \geqq -y_1 - 1,\ y_1 \lneqq 1,\ y_2 \leqq 1\},$$

$$D(y) = R^2_+ \qquad \text{for all} \quad y \in Y(U).$$

Then,

$$N_1(u) = \{(y_1, y_2) : y_2 = uy_1,\ 1/u \leqq y_1 \leqq 1\} \qquad \text{for} \quad u < 0, \quad u \neq -1,$$

$$N_1(-1) = \{(y_1, y_2) : y_2 = -y_1 - 1,\ -2 \leqq y_1 \leqq 1\}.$$

Example 4.2.3 (Lower Semicontinuity of D on $Y(\hat{u})$—Fig. 4.3)

Let

$$U = \{u \in R : 0 < u \leqq 1\}, \qquad \hat{u} = 1,$$

$$Y(u) = \{(y_1, y_2) : (y_1)^2 + (y_2)^2 \leqq u^2\} \subset R^2,$$

$$D(y) = \begin{cases} R^2_+ & \text{if} \quad (y_1)^2 + (y_2)^2 < 1, \\ \{(d_1, d_2) : d_2 \geqq 0,\ d_1 + d_2 \geqq 0\} & \text{if} \quad (y_1)^2 + (y_2)^2 = 1. \end{cases}$$

Then,

$$N_1(u) = \{(y_1, y_2) : (y_1)^2 + (y_2)^2 = u^2,\ y_1 \leqq 0,\ y_2 \leqq 0\} \qquad \text{for} \quad 0 < u < 1,$$

$$N_1(1) = \{(y_1, y_2) : (y_1)^2 + (y_2)^2 = 1,\ y_1 \leqq 0,\ y_1 \geqq y_2\}.$$

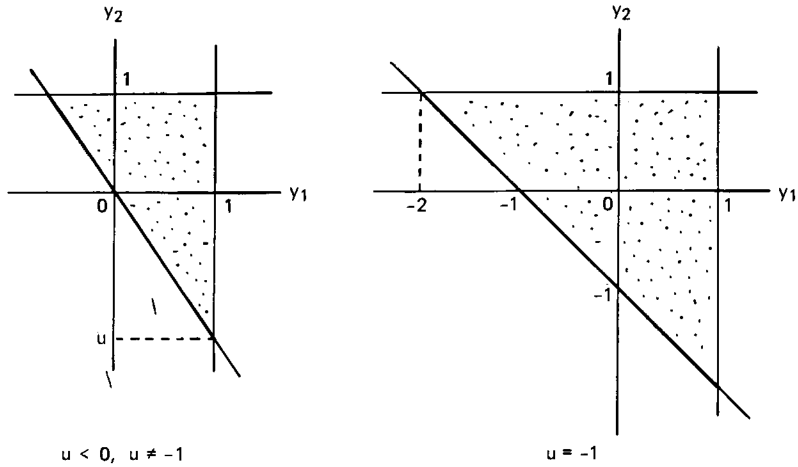

Fig. 4.2. Example 4.2.2.

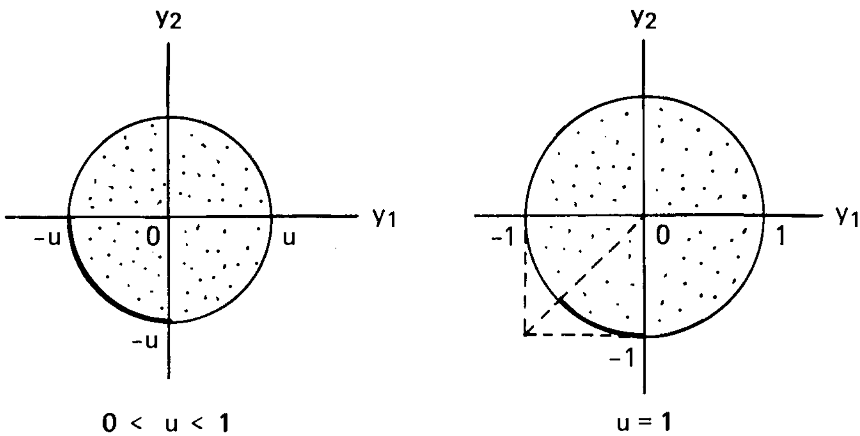

Fig. 4.3. Example 4.2.3.

Example 4.2.4 (Convexity of the Sets $D(y)$—Fig. 4.4)

Let

$$U = \{u \in R : 0 < u \leqq 1\}, \qquad \hat{u} = 1,$$

$$Y(u) = \{(y_1, y_2) : y_2 = uy_1,\ 0 \leqq y_1 \leqq 1\} \subset R^2,$$

$$D(y) = R^2_+ \cap \{(d_1, d_2) : (d_1 - d_2)((2u - 1)d_1 - d_2) \geqq 0\}$$

if $y_2 = uy_1$, $y \neq 0$; (in particular $D(y) = R^2_+$ if $y_2 = y_1$, $y \neq 0$),

$$D(0) = \{(0, 0)\}.$$

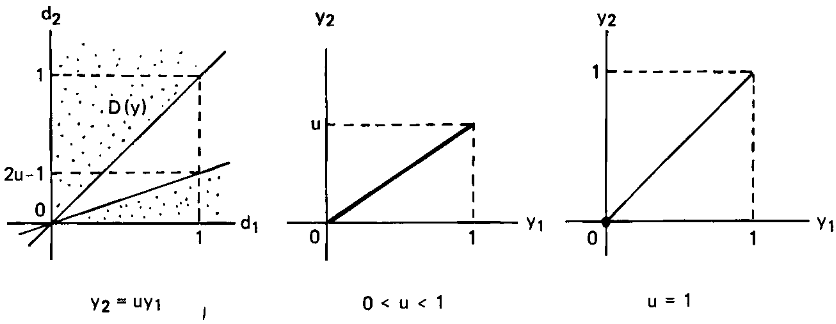

Fig. 4.4. Example 4.2.4.

Then,

$$N_1(u) = \{(y_1, y_2) : y_2 = uy_1,\ 0 \leqq y_1 \leqq 1\} = Y(u) \qquad \text{for} \quad 0 < u < 1,$$
$$N_1(1) = \{(0, 0)\}.$$

Example 4.2.5 (The Relationship $N_1(\hat{u}) = \mathscr{E}(Y(\hat{u}), \operatorname{int} D)$—Fig. 4.5)

Let

$$U = \{u \in R : u \leqq 0\}, \qquad \hat{u} = 0,$$
$$Y(u) = \{(y_1, y_2) : y_2 \leqq y_1,\ y_2 \geqq uy_1\} \subset R^2,$$
$$D(y) = R_+^2 \qquad \text{for all} \quad y \in Y(U).$$

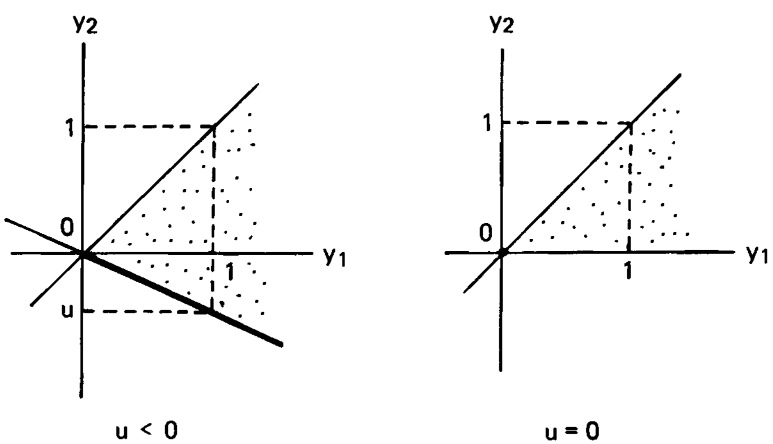

Fig. 4.5. Example 4.2.5.

Then,

$$N_1(u) = \{(y_1, y_2) : y_1 \geqq 0, y_2 = uy_1\} \qquad \text{for} \quad u < 0,$$
$$N_1(0) = \{(0,0)\} \neq \mathscr{E}(Y(0), \operatorname{int} D) = \{(y_1, y_2) : y_1 \geqq 0, y_2 = 0\}.$$

4.2.2 Lower Semicontinuity of the Map N_1

Next, we consider sufficient conditions for the lower semicontinuity of the map N_1.

Theorem 4.2.2

The map N_1 is lower semicontinuous at $\hat{u} \in U$ if the following conditions are satisfied:

(i) the map Y is continuous at $\hat{u}$;
(ii) Y is uniformly compact near $\hat{u}$;
(iii) $N_1(u) = \mathscr{E}(Y(u), D)$ is externally stable for every $u \in U$ near $\hat{u}$;
(iv) the map D is upper semicontinuous on $Y(\hat{u})$.

Proof Let

$$\{u^k\} \subset U, \qquad u^k \to \hat{u}, \qquad \text{and} \qquad \hat{y} \in N_1(\hat{u}).$$

Since Y is lower semicontinuous at $\hat{u}$, there exist a number m' and a sequence $\{y^k\}$ such that $y^k \in Y(u^k)$ for $k \geqq m'$ and $y^k \to \hat{y}$. From condition (iii), there is a number m'' such that $N_1(u^k)$ is externally stable for all $k \geqq m''$. Let

$$m = \max(m', m'').$$

Then, for $k \geqq m$, there is $\bar{y}^k \in N_1(u^k)$ such that

$$y^k \in \bar{y}^k + D(\bar{y}^k).$$

For $1 \leqq k < m$, let $\bar{y}^k$ be an arbitrary point in R^p. Since Y is uniformly compact near $\hat{u}$, the sequence $\{\bar{y}^k\}$ has a cluster point, which is denoted by $\bar{y}$, and is contained in $Y(\hat{u})$ since Y is upper semicontinuous at $\hat{u}$. In other words, the sequence $\{y^k - \bar{y}^k\}$ has a cluster point $\hat{y} - \bar{y}$. Since D is upper semicontinuous on $Y(\hat{u})$,

$$\hat{y} - \bar{y} \in D(\bar{y});$$

that is,

$$\hat{y} \in \bar{y} + D(\bar{y}).$$

If we notice that $\hat{y} \in N_1(\hat{u})$, $\bar{y}$ must coincide with $\hat{y}$. Namely, $\hat{y}$ is a unique cluster point for the bounded sequence $\{\bar{y}^k\}$. Thus, we have proved that the sequence $\{\bar{y}^k\}$ satisfies $\bar{y}^k \in N_1(u^k)$ for $k \geqq m$ and converges to $\hat{y}$. Hence N_1 is upper semicontinuous at $\hat{u}$, and the proof of the theorem is completed.

Some examples illustrate the necessity of each condition in the above theorem. The condition in the parenthesis does not hold, but all the other conditions in the theorem are satisfied.

Example 4.2.6 (Upper Semicontinuity of the Map Y at $\hat{u}$)

See Example 4.2.1.

Example 4.2.7 (Lower Semicontinuity of the Map Y at $\hat{u}$)

See Example 4.2.2.

Example 4.2.8 (Uniform Compactness of the Map Y near $\hat{u}$—Fig. 4.6)

Let

$$U = \{u \in R : u \geqq 0\}, \qquad \hat{u} = 0,$$
$$Y(u) = \{(y_1, y_2) : (y_1)^2 + (y_2)^2 = 1/u\} \cup \{(0,0)\} \subset R^2, \qquad \text{for} \quad u > 0,$$
$$Y(0) = \{(0,0)\},$$
$$D(y) = R^2_+ \qquad \text{for all} \quad y \in R^2.$$

Then,

$$N_1(u) = \{(y_1, y_2) : (y_1)^2 + (y_2)^2 = 1/u,\ y_1 \leqq 0,\ y_2 \leqq 0\} \qquad \text{for} \quad u > 0,$$
$$N_1(0) = \{(0,0)\}.$$

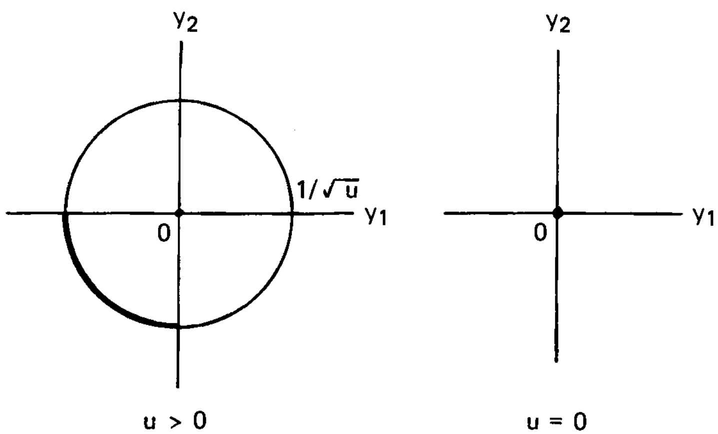

Fig. 4.6. Example 4.2.8.

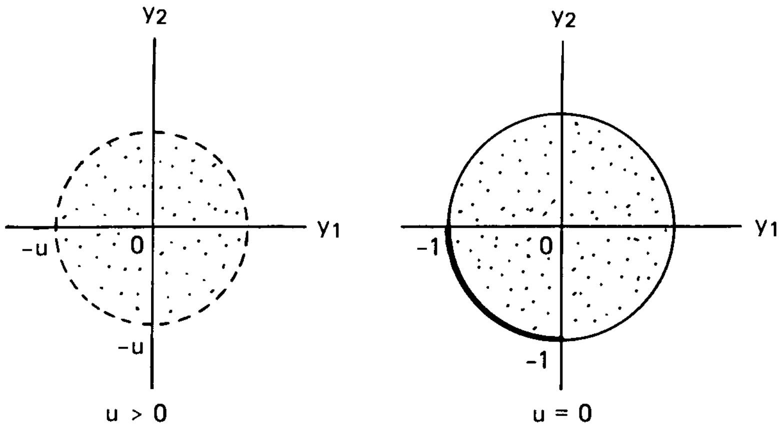

Fig. 4.7. Example 4.2.9.

Example 4.2.9 (External Stability of $N_1(u)$, Compactness of the Set $Y(u)$—Fig. 4.7)

Let

$$
\begin{aligned}
U &= \{u \in R : u > 0\}, \qquad \hat{u} = 1, \\
Y(u) &= \{(y_1, y_2) : (y_1)^2 + (y_2)^2 < u^2\} \subset R^2 \qquad \text{for} \quad u > 0, \quad u \neq 1, \\
Y(1) &= \{(y_1, y_2) : (y_1)^2 + (y_2)^2 \leqq 1\}, \\
D(y) &= R^2_+ \qquad \text{for all} \quad y \in R^2.
\end{aligned}
$$

Then

$$
\begin{aligned}
N_1(u) &= \varnothing \qquad \text{for} \quad u > 0, \quad u \neq 1, \\
N_1(1) &= \{(y_1, y_2) : (y_1)^2 + (y_2)^2 = 1, y_1 \leqq 0, y_2 \leqq 0\}.
\end{aligned}
$$

Example 4.2.10 (External Stability of $N_1(u)$, Upper Semicontinuity of D on $Y(u)$—Fig. 4.8)

Let

$$
\begin{aligned}
U &= \{u = 1/n : n \text{ is a positive integer}\} \cup \{0\}, \qquad \hat{u} = 0, \\
Y(u) &= \{(y_1, y_2) : 0 \leqq y_1 \leqq 1, y_2 = u\} \subset R^2,
\end{aligned}
$$

$$
D(y) = \begin{cases} \{(d_1, d_2) : 0 \leqq d_1 \leqq k/n - y_1\} & \text{for} \quad (k-1)/n < y_1 \leqq k/n, \\ & \qquad k = 1, \dots, n, \quad y_2 = 1/n, \\ \{(d_1, d_2) : d_1 = 0\} & \text{for} \quad y_1 = 0 \quad \text{or} \quad y_2 = 0. \end{cases}
$$

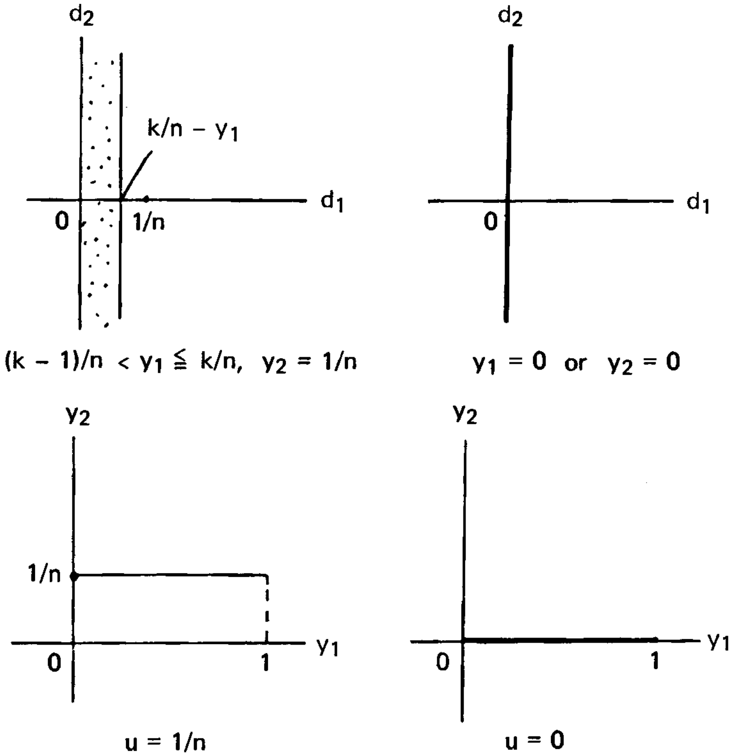

Fig. 4.8. Example 4.2.10.

Then,

$$N_1(u) = \{(0, u)\} \qquad \text{for} \quad u = 1/n,$$

$$N_1(0) = \{(y_1, y_2) : 0 \leqq y_1 \leqq 1,\ y_2 = 0\}.$$

Example 4.2.11 (External Stability of $N_1(u)$, Asymmetry of D—Fig 4.9)

Let

$$U = \{u \in R : u \geqq 0\}, \qquad \hat{u} = 0,$$

$$Y(u) = \{(y_1, y_2) : 0 \leqq y_1 \leqq 1,\ 0 \leqq y_2 \leqq uy_1\} \subset R^2;$$

$$\text{in particular,} \qquad Y(0) = \{(y_1, y_2) : 0 \leqq y_1 \leqq 1,\ y_2 = 0\},$$

$$D(y) = \{(d_1, d_2) : d_1 = 0\} \qquad \text{for all} \quad y \in Y(U).$$

Then,

$$N_1(u) = \{(0, 0)\} \qquad \text{for} \quad u > 0,$$

$$N_1(0) = \{(y_1, y_2) : 0 \leqq y_1 \leqq 1,\ y_2 = 0\}.$$

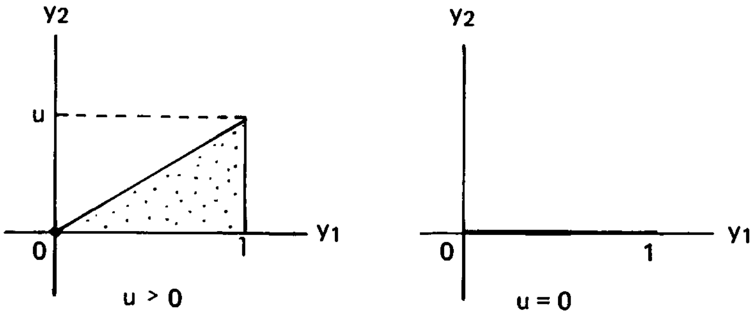

Fig. 4.9. Example 4.2.11.

Example 4.2.12 (External Stability of $N_1(u)$, Transitivity of D—Fig. 4.10)

Let

$$U = \{u \in R : 0 \leqq u \leqq \tfrac{1}{2}\}, \qquad \hat{u} = 0,$$

$$Y(u) = \{(y_1, y_2) : u \leqq y_1 \leqq 1, y_2 = 0\}$$
$$\cup \{(y_1, y_2) : u \leqq y_1 \leqq 1, y_1 = y_2\}$$

$$D(y) = \begin{cases} \{(d_1, d_2) : d_1 = 0, d_2 \geqq 0\} & \text{if } y_2 = 0, y_1 \neq 0, \\ \{(d_1, d_2) : d_2 = 2d_1 \leqq 0\} & \text{if } y_1 = y_2 \neq 0, \\ \{(d_1, d_2) : d_1 \leqq 0, d_2 \geqq 2d_1\} & \text{if } y = (0, 0). \end{cases}$$

Then,

$$N_1(u) = \{(y_1, y_2) : \tfrac{1}{2} < y_1 \leqq 1, y_2 = 0\} \quad \text{for} \quad 0 < u \leqq \tfrac{1}{2},$$
$$N_1(0) = \{(0, 0)\} \cup \{(y_1, y_2) : \tfrac{1}{2} < y_1 \leqq 1, y_2 = 0\}.$$

Example 4.2.13 (Upper Semicontinuity of the Map D on $Y(\hat{u})$—Fig. 4.11)

Let

$$U = \{u \in R : u \geqq 0\}, \qquad \hat{u} = 0,$$

$$Y(u) = \{(y_1, y_2) : 0 \leqq y_1 \leqq 1, uy_1 \leqq y_2 \leqq 1\} \subset R^2,$$

$$D(y) = \begin{cases} \{(d_1, d_2) : d_1 \geqq 0, d_2 \geqq ud_1\} & \text{if } y_2 = uy_1, y \neq 0 \text{ for } u > 0, \\ \{(d_1, d_2) : d_1 = 0, d_2 \geqq 0\} & \text{if } y_1 = 0, y_2 \neq 0, \\ \{(d_1, d_2) : d_1 \geqq 0, d_2 > 0\} & \text{if } y_2 = 0. \end{cases}$$

Then,

$$N_1(u) = \{(0, 0)\} \quad \text{for} \quad u > 0,$$
$$N_1(0) = \{(y_1, y_2) : 0 \leqq y_1 \leqq 1, y_2 = 0\}.$$

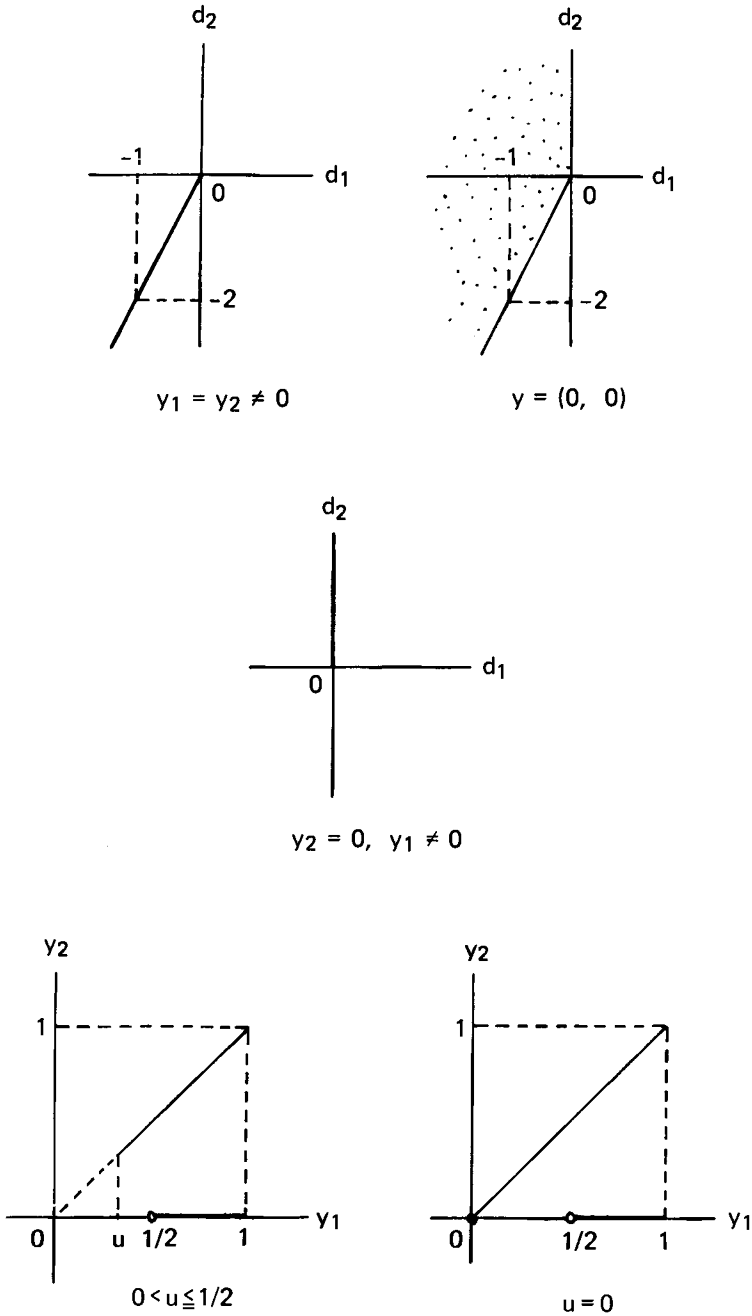

Fig. 4.10. Example 4.2.12.

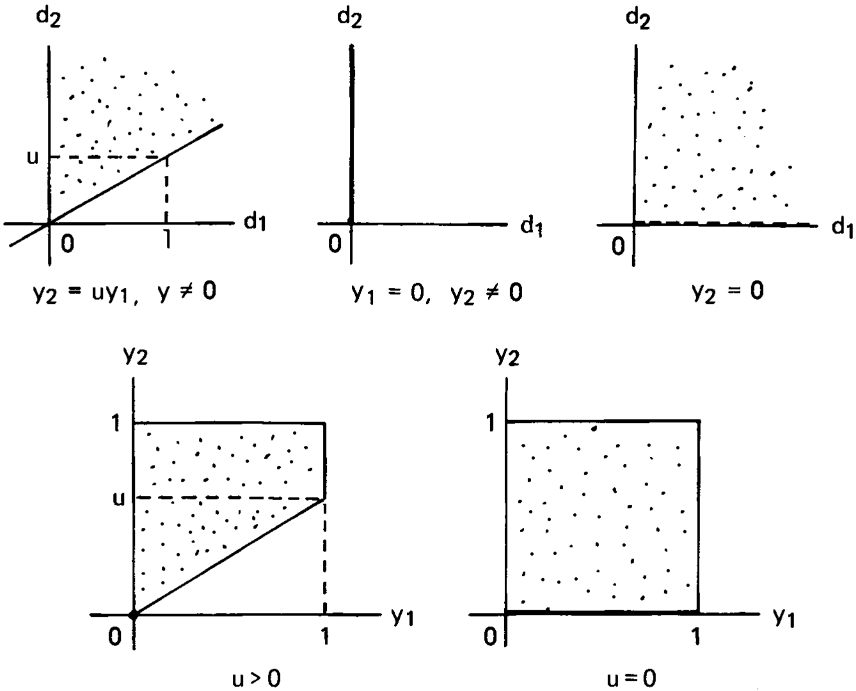

Fig. 4.11. Example 4.2.13.

4.3 Stability for Perturbation of the Domination Structure

This section deals with the stability of the solution set from the second point of view described in Section 4.1. In other words, we will investigate the continuity of the point-to-set map N_2, which is defined on the space V by parameter vectors v that specify the domination structure of the decision maker.

This point of view is peculiar to multiobjective optimization problems. The aim of the decision maker in ordinary optimization problems is to maximize (or minimize) scalar objective functions and so is quite clear; hence, there is no room for such consideration. On the other hand, in multiobjective optimization problems, the domination structure of the decision maker is a very important and essential factor. Hence, it is quite natural to consider that it is perturbed according to changes of environmental factors or of feelings of the decision maker. Thus, the study of the stability of the solution set for perturbations of the domination structure seems to be necessary and interesting.

Throughout this section, the set of feasible solutions in the objective space is fixed as Y and the domination structure D is considered to be a point-to-set map from $Y \times V$ into R^p. As has been already noted in Section 4.1, the point-to-set map N_2 from V into R^p is defined by

$$N_2(v) = \mathscr{E}(Y, D(\cdot, v)).$$

Some sufficient conditions for the continuity of this point-to-set map N_2 will be investigated in this section.

4.3.1 Upper Semicontinuity of the Map N_2

First, sufficient conditions under which the map N_2 is upper semicontinuous are obtained as follows.

Theorem 4.3.1

The map N_2 is upper semicontinuous at $\hat{v} \in V$ if the following conditions are satisfied:

(i) Y is a closed set;
(ii) the map $D(y, \cdot)$ is lower semicontinuous at $\hat{v}$ for each fixed $y \in Y$;
(iii) $D(y, v)$ is a convex set for all $y \in Y$ and every $v \in V$ near $\hat{v}$;
(iv) the relationship $N_2(\hat{v}) = \mathscr{E}(Y, \operatorname{int} D(\cdot, \hat{v}))$ holds.

Proof Let

$$\{v^k\} \subset V, \qquad v^k \to \hat{v}, \qquad y^k \in N_2(v^k), \qquad \text{and} \qquad y^k \to \hat{y}.$$

We must show that $\hat{y} \in N_2(\hat{v})$. Since Y is a closed set, then $\hat{y} \in Y$. Hence, if we assume to the contrary that $\hat{y}$ does not belong to the set

$$N_2(\hat{v}) = \mathscr{E}(Y, \operatorname{int} D(\cdot, \hat{v})) \qquad \text{(condition (iv))},$$

then there exist a $y \in Y$ and a nonzero $d \in \operatorname{int} D(y, \hat{v})$ such that $\hat{y} = y + d$. Let

$$y^k - y = d^k, \qquad k = 1, 2, \ldots.$$

Then, in view of Lemma 4.2.1,

$$d^k \in D(y, v^k)$$

except for a finite number of k's, since

$$d^k \to d \in \operatorname{int} D(y, \hat{v})$$

and the point-to-set map $D(y, \cdot)$ is lower semicontinuous at $\hat{v}$. This implies that

$$y^k \in y + D(y, v^k) \qquad \text{and} \qquad y^k \neq y \qquad \text{for some} \quad k,$$

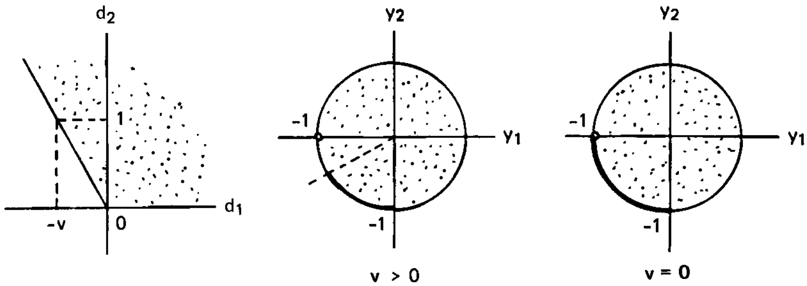

Fig. 4.12. Example 4.3.1.

which contradicts the assumption $y^k \in N_2(v^k)$. Therefore $\hat{y}$ must be contained in $N_2(\hat{v})$. Namely, N_2 is upper semicontinuous at $\hat{v}$ and the proof is completed.

Some examples are given in the following to illustrate the necessity of each condition in the theorem.

Example 4.3.1 (Closedness of the Set Y—Fig. 4.12)

Let

$$V = \{v \in R : v \geqq 0\}, \qquad \hat{v} = 0,$$

$$Y = \{(y_1, y_2) : (y_1)^2 + (y_2)^2 \leqq 1\} \backslash \{(-1, 0)\} \subset R^2,$$

$$D(y, v) = \{(d_1, d_2) : d_2 \geqq 0,\ d_1 + vd_2 \geqq 0\} \qquad \text{for all} \quad y \in Y.$$

Then,

$$N_2(v) = \{(y_1, y_2) : (y_1)^2 + (y_2)^2 = 1,\ y_1 \leqq 0,\ y_2 \leqq vy_1\} \qquad \text{for} \quad v > 0,$$

$$N_2(0) = \{(y_1, y_2) : (y_1)^2 + (y_2)^2 = 1,\ y_1 \leqq 0, y_2 < 0\}.$$

Example 4.3.2 (Lower Semicontinuity of the Map $D(y, \cdot)$ at $\hat{v}$—Fig. 4.13)

Let

$$V = \{v \in R : v > 0\}, \qquad \hat{v} = 1,$$

$$Y = \{(y_1, y_2) : (y_1)^2 + (y_2)^2 \leqq 1\} \subset R^2,$$

$$D(y, v) = R^2_+ \qquad \text{for all} \quad y \in Y \qquad \text{if} \quad v > 0, \quad v \neq 1,$$

$$D(y, 1) = \{(d_1, d_2) : d_2 \geqq 0,\ d_1 + d_2 \geqq 0\} \qquad \text{for all} \quad y.$$

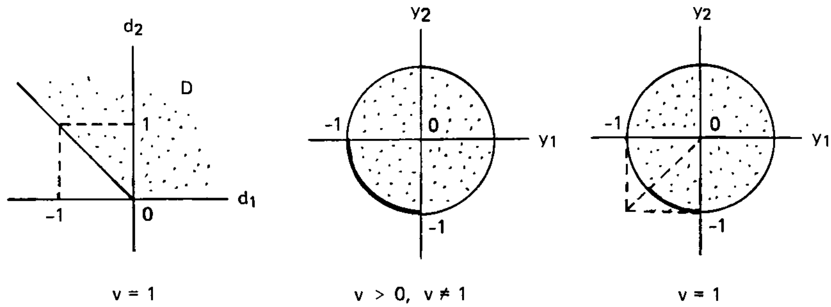

Fig. 4.13. Example 4.3.2.

Then,

$$N_2(v) = \{(y_1, y_2) : (y_1)^2 + (y_2)^2 = 1, y_1 \leqq 0, y_2 \leqq 0\}$$
$$\text{for} \quad v > 0, \quad v \neq 1,$$
$$N_2(1) = \{(y_1, y_2) : (y_1)^2 + (y_2)^2 = 1, y_1 \leqq 0, y_2 \leqq y_1\}.$$

Example 4.3.3 (Convexity of the Set $D(y, v)$—Fig. 4.14)

Let

$$V = \{v \in R : 0 < v \leqq 1\}, \qquad \hat{v} = 1,$$
$$Y = \{(y_1, y_2) : 0 \leqq y_1 \leqq 1, y_1 = y_2\} \subset R^2,$$
$$D(y, v) = \{(d_1, d_2) : d_1 \geqq 0, d_2 \geqq 0,$$
$$(d_1 - (2v - 1)d_2)((2v - 1)d_1 - d_2) \geqq 0\} \qquad \text{for all} \quad y \in Y.$$

In particular,

$$D(y, 1) = R^2_+ \qquad \text{for all} \quad y \in Y.$$

Then,

$$N_2(v) = \{(y_1, y_2) : 0 \leqq y_1 \leqq 1, y_1 = y_2\} \qquad \text{for} \quad 0 < v < 1,$$
$$N_2(1) = \{(0, 0)\}.$$

Example 4.3.4 (The Relationship $N_2(\hat{v}) = \mathscr{E}(Y, \operatorname{int} D(\cdot, \hat{v}))$—Fig. 4.15)

Let

$$V = \{v \in R : v \geqq 0\}, \qquad v = 0,$$
$$Y = \{(y_1, y_2) : -y_1 + y_2 \leqq 0, y_2 \geqq 0\} \subset R^2,$$
$$D(y, v) = \{(d_1, d_2) : -vd_1 + d_2 \geqq 0, d_1 + vd_2 \geqq 0\} \qquad \text{for all} \quad y \in Y,$$

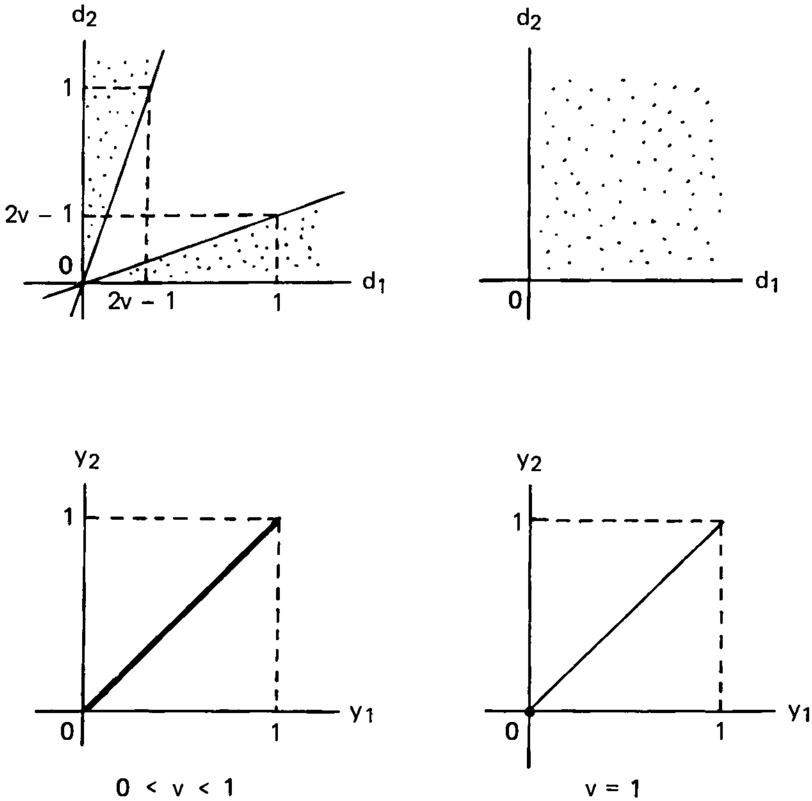

Fig. 4.14. Example 4.3.3.

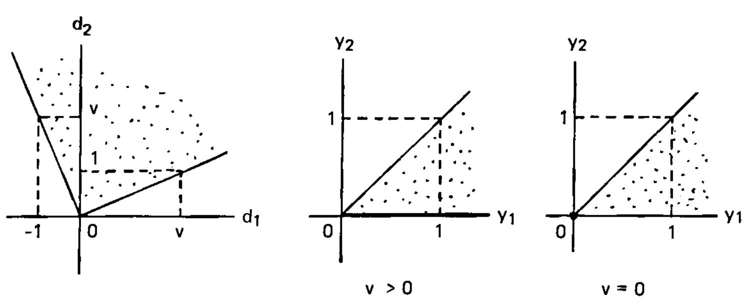

Fig. 4.15. Example 4.3.4.

In particular,

$$D(y, 0) = R^2_+ \qquad \text{for all} \quad y \in Y.$$

Then,

$$N_2(v) = \{(y_1, y_2) : y_1 \geqq 0, y_2 = 0\} \qquad \text{for} \quad v > 0,$$

$$\begin{aligned} N_2(0) &= \{(0, 0)\} \neq \mathscr{E}(Y, \operatorname{int} D(\cdot, 0)) \\ &= \{(y_1, y_2) : y_1 \geqq 0, y_2 = 0\}. \end{aligned}$$

4.3.2 Lower Semicontinuity of the Map N_2

The fourth theorem in this chapter provides sufficient conditions for the lower semicontinuity of the map N_2.

Theorem 4.3.2

The map N_2 is lower semicontinuous at $\hat{v} \in V$ if the following conditions are satisfied:

(i) Y is a compact set;
(ii) $N_2(v) = \mathscr{E}(Y, D(\cdot, v))$ is externally stable for every v near $\hat{v}$;
(iii) D is upper semicontinuous on $Y \times \hat{v}$.

Proof If Y is empty, the theorem is obviously true. So we assume that Y is nonempty. Let

$$\{v^k\} \subset V, \qquad v^k \to \hat{v}, \qquad \text{and} \qquad \hat{y} \in N_2(\hat{v}).$$

From condition (ii), there exist a number m and $y^k \in N_2(v^k)$ for $k \geqq m$ such that

$$\hat{y} \in y^k + D(y^k, v^k).$$

Let y^k be an arbitrary point in Y for $1 \leqq k < m$. Since Y is a compact set, the sequence $\{y^k\}$ has a cluster point in Y, which is denoted by $\bar{y}$. In other words, the sequence $\{\hat{y} - y^k\}$ has a cluster point $\hat{y} - \bar{y}$. Since

$$\hat{y} - y^k \in D(y^k, v^k) \qquad \text{for} \quad k \geqq m$$

and D is upper semicontinuous on $Y \times \hat{v}$,

$$\hat{y} - \bar{y} \in D(\bar{y}, \hat{v}).$$

By combining this with the fact $\hat{y} \in N_2(\hat{v})$, we have $\hat{y} = \bar{y}$. Namely, $\hat{y}$ is a unique cluster point for the bounded sequence $\{y^k\}$. Thus, we have proved that the sequence $\{y^k\}$ satisfies

$$y^k \in N_2(v^k) \qquad \text{for} \quad k \geqq m \qquad \text{and} \qquad y^k \to \hat{y}.$$

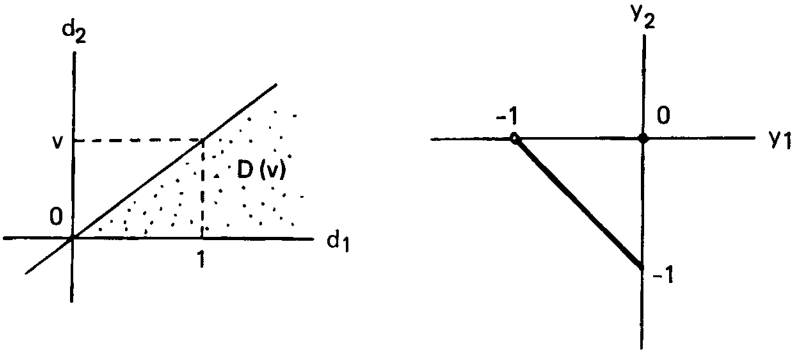

Fig. 4.16. Example 4.3.5.

Therefore, the map N_2 is lower semicontinuous at $\hat{v}$ and the proof of the theorem is completed.

Some examples are given in the following.

Example 4.3.5 (Closedness of the Set Y—Fig. 4.16)

Let

$$V = \{v \in R : v \geqq 0\}, \qquad \hat{v} = 0,$$

$$Y = \{(y_1, y_2) : y_1 + y_2 = -1,\ -1 < y_1 \leqq 0\} \cup \{(0,0)\} \subset R^2,$$

$$D(y, v) = \{(d_1, d_2) : 0 \leqq d_2 \leqq vd_1\} \qquad \text{for all} \quad y \in Y.$$

Then,

$$N_2(v) = \{(y_1, y_2) : y_1 + y_2 = -1,\ -1 < y_1 \leqq 0\} \qquad \text{for} \quad v > 0,$$

$$N_2(0) = \{(y_1, y_2) : y_1 + y_2 = -1,\ -1 < y_1 \leqq 0\} \cup \{(0,0)\}.$$

Example 4.3.6 (Boundedness of the Set Y—Fig. 4.17)

In Example 4.3.5, we replace Y by a new set defined by

$$Y = \{(y_1, y_2) : y_1 < 0,\ y_1 y_2 = 1\} \cup \{(0,0)\} \subset R^2.$$

Then,

$$N_2(v) = \{(y_1, y_2) : y_1 < 0,\ y_1 y_2 = 1\} \qquad \text{for} \quad v > 0,$$

$$N_2(0) = \{(y_1, y_2) : y_1 < 0,\ y_1 y_2 = 1\} \cup \{(0,0)\}.$$

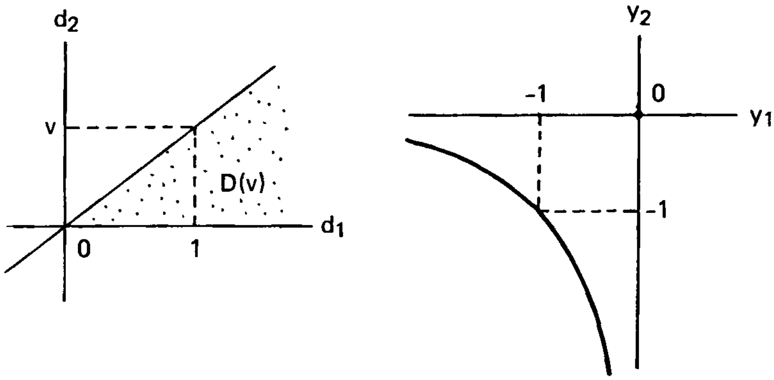

Fig. 4.17. Example 4.3.6.

Example 4.3.7 (External Stability of $N_2(v)$, Upper Semicontinuity of $D(\cdot, v)$ on Y for v near $\hat{v}$—Fig. 4.18)

Let

$$V = \{v = 1/n : n \text{ is a positive integer}\} \cup \{0\}, \qquad \hat{v} = 0,$$

$$Y = \{y : 0 \leqq y \leqq 1\} \subset R,$$

$$D(y, v) = \begin{cases} \{d : 0 \leqq d \leqq k/n - y\} & \text{for} \quad (k-1)/n < y \leqq k/n, \\ & \qquad\qquad k = 1, \dots, n, \\ \{0\} & \text{for} \quad y = 0 \qquad \text{if} \quad v = 1/n. \end{cases}$$

Then,

$$N_2(v) = \{0\}, \qquad \text{for} \quad v = 1/n,$$

$$N_2(0) = \{y : 0 \leqq y \leqq 1\}.$$

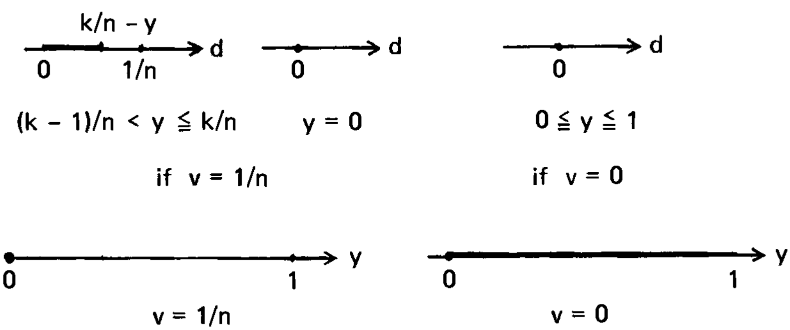

Fig. 4.18. Example 4.3.7.

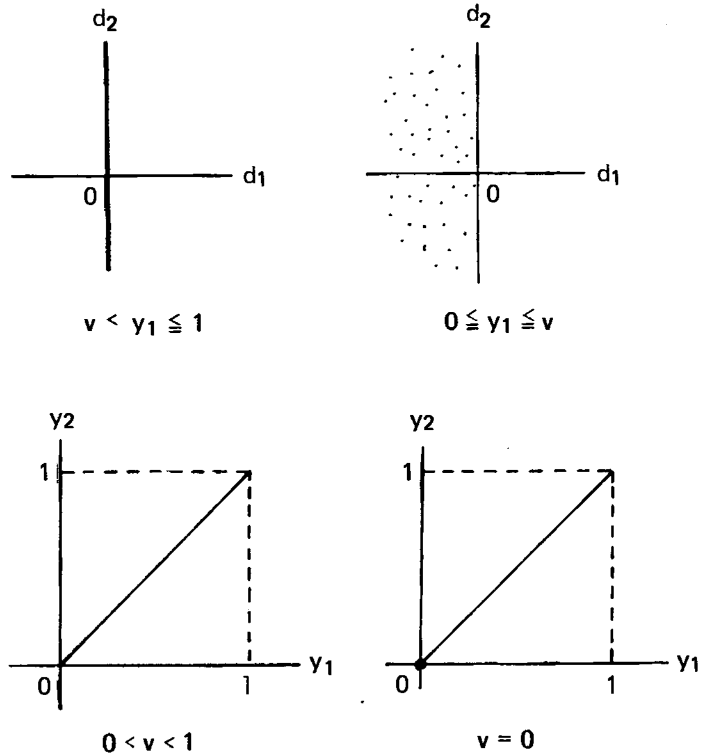

Fig. 4.19. Example 4.3.8.

Example 4.3.8 (External Stability of $N_2(v)$, Asymmetry of $D(\cdot, v)$—Fig. 4.19)

Let

$$V = \{v \in R : 0 \leqq v < 1\}, \qquad \hat{v} = 0,$$

$$Y = \{(y_1, y_2) : 0 \leqq y_1 \leqq 1, y_2 = 0\}$$
$$\cup \{(y_1, y_2) : 0 \leqq y_1 \leqq 1, y_1 = y_2\},$$

$$D(y, v) = \begin{cases} \{(d_1, d_2) : d_1 = 0\} & \text{if} \quad v < y_1 \leqq 1, \\ \{(d_1, d_2) : d_1 \leqq 0\} & \text{if} \quad 0 \leqq y_1 \leqq v. \end{cases}$$

Then,

$$N_2(v) = \varnothing \qquad \text{for} \quad 0 < v < 1,$$
$$N_2(0) = \{(0, 0)\}.$$

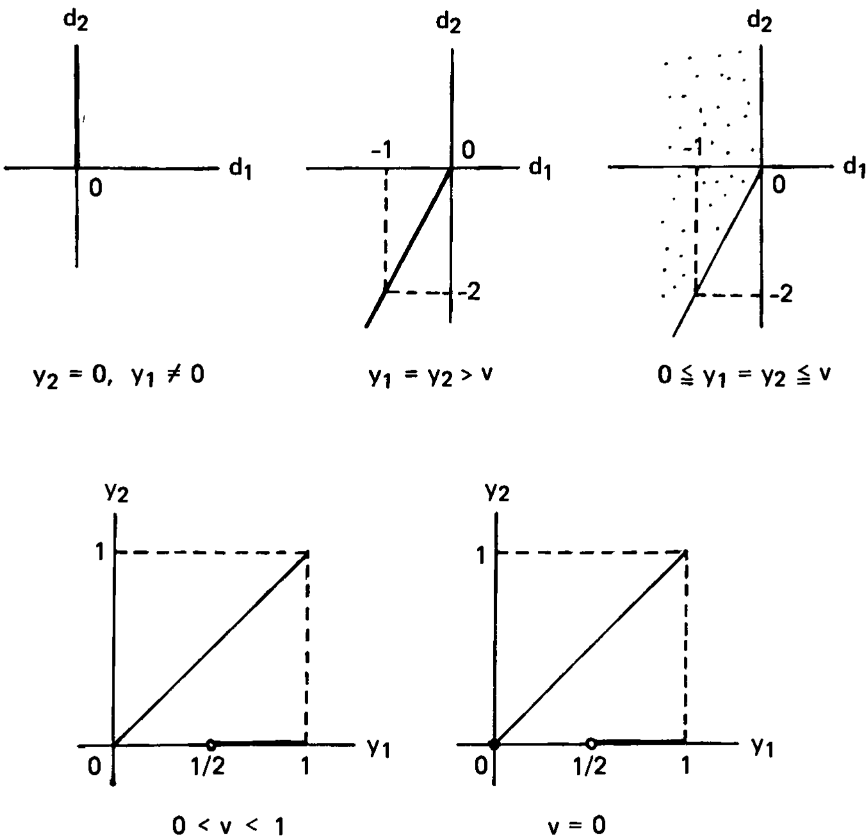

Fig. 4.20. Example 4.3.9.

Example 4.3.9 (External Stability of $N_2(v)$, Transitivity of $D(\cdot, v)$—Fig. 4.20)

In Example 4.3.8 we replace $D(y, v)$ by

$$D(y, v) = \begin{cases} \{(d_1, d_2) : d_1 = 0,\ d_2 \geqq 0\} & \text{if} \quad y_2 = 0, \quad y_1 \neq 0, \\ \{(d_1, d_2) : d_1 \leqq 0,\ d_2 = 2d_1\} & \text{if} \quad y_1 = y_2 > v, \\ \{(d_1, d_2) : d_1 \leqq 0,\ d_2 \geqq 2d_1\} & \text{if} \quad 0 \leqq y_1 = y_2 \leqq v. \end{cases}$$

Then,

$$N_2(v) = \{(y_1, y_2) : \tfrac{1}{2} < y_1 \leqq 1,\ y_2 = 0\},$$
$$N_2(0) = \{(0, 0)\} \cup \{(y_1, y_2) : \tfrac{1}{2} < y_1 \leqq 1,\ y_2 = 0\}.$$

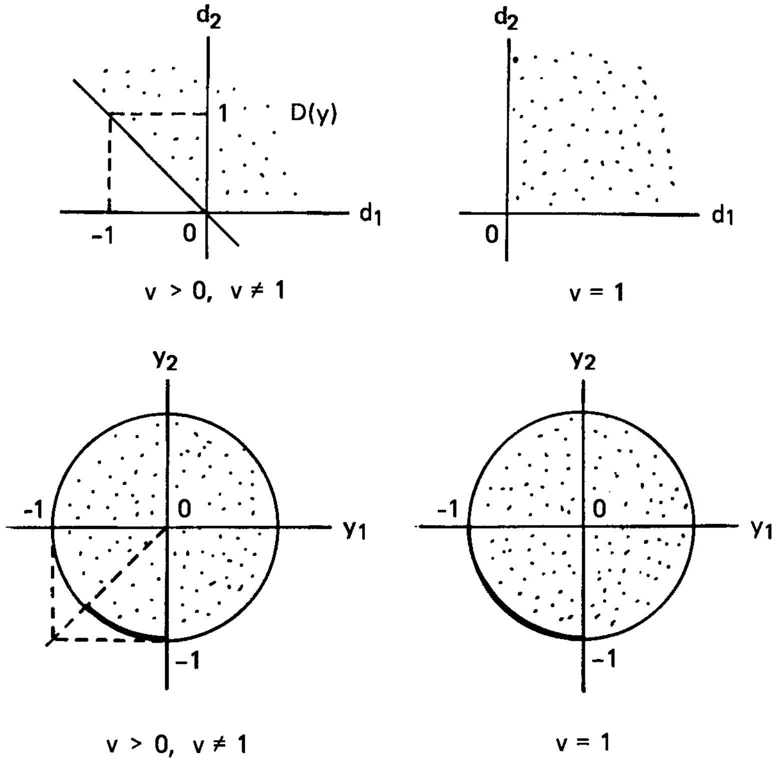

Fig. 4.21. Example 4.3.10.

Example 4.3.10 (Upper Semicontinuity of the Map $D(y, \cdot)$ at $\hat{v}$—Fig. 4.21)
Let

$$V = \{v \in R : v > 0\}, \qquad \hat{v} = 1,$$
$$Y = \{(y_1, y_2) : (y_1)^2 + (y_2)^2 \leqq 1\} \subset R^2,$$
$$D(y, v) = \{(d_1, d_2) : d_2 \geqq 0,\ d_1 + d_2 \geqq 0\}$$
$$\text{for all} \quad y \in Y \qquad \text{if} \quad v > 0, \quad v \neq 1,$$
$$D(y, 1) = R^2_+ \qquad \text{for all} \quad y \in Y.$$

Then,

$$N_2(v) = \{(y_1, y_2) : (y_1)^2 + (y_2)^2 = 1,\ y_1 \leqq 0,\ y_1 - y_2 \geqq 0\}$$
$$\text{for} \quad v > 0, \quad v \neq 1,$$
$$N_2(1) = \{(y_1, y_2) : (y_1)^2 + (y_2)^2 = 1,\ y_1 \leqq 0,\ y_2 \leqq 0\}.$$

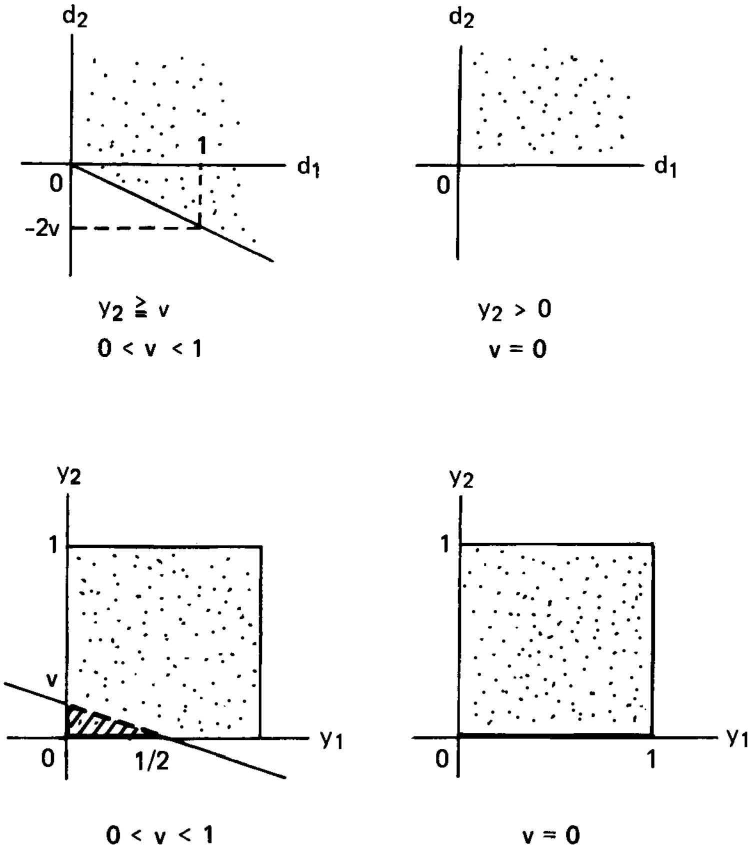

Fig. 4.22. Example 4.3.11.

Example 4.3.11 (Upper Semicontinuity of the Map $D(\cdot, \hat{v})$ on Y—Fig. 4.22)

Let

$$V = \{v \in R : 0 \leqq v \leqq 1\}, \qquad \hat{v} = 0,$$

$$Y = \{(y_1, y_2) : 0 \leqq y_1 \leqq 1,\ 0 \leqq y_2 \leqq 1\} \subset R^2,$$

$$D(y, v) = \begin{cases} \{(d_1, d_2) : d_1 \geqq 0,\ 2vd_1 + d_2 \geqq 0\} & \text{if} \quad y_2 \geqq v,\ v > 0, \\ \{(0, 0)\} & \text{if} \quad y_2 < v,\ v > 0, \end{cases}$$

$$D(y, 0) = \begin{cases} \{(d_1, d_2) : d_1 \geqq 0,\ d_2 \geqq 0\} & \text{if} \quad y_2 > 0, \\ \{(0, 0)\} & \text{if} \quad y_2 = 0. \end{cases}$$

Then,

$$N_2(v) = \{(y_1, y_2) : y_1 \geqq 0, 0 \leqq y_2 < -2vy_1 + v\} \cup \{(0, v)\}$$
$$\text{for} \quad 0 < v \leqq 1,$$
$$N_2(0) = \{(y_1, y_2) : 0 \leqq y_1 \leqq 1, y_2 = 0\}.$$

4.4 Stability in the Decision Space

In the preceding sections, we have considered the stability of the efficient set in the objective space in terms of the continuity of the two point-to-set maps N_1 and N_2. In this section, we will consider the stability of the efficient set in the decision space by investigating the continuity of the two point-to-set maps M_1 and M_2, which have been already defined in Section 4.1.

First, as in Section 4.1, let X be a point-to-set map from a set U into R^n, let f be a function from $R^n \times U$ into R^p, and let

$$Y(u) = \{y \in R^p : y = f(x, u), x \in X(u)\}.$$

Then, the lemma is basic for the results in this section.

Lemma 4.4.1

(i) If the map X is upper semicontinuous at $\hat{u} \in U$ and uniformly compact near $\hat{u}$, and if the function f is continuous on $X(\hat{u}) \times \hat{u}$, then the map Y is upper semicontinuous at $\hat{u}$.

(ii) If the map X is lower semicontinuous at $\hat{u}$ and f is continuous on $X(\hat{u}) \times \hat{u}$, then Y is lower semicontinuous at $\hat{u}$.

Proof

(i) Let

$$\{u^k\} \subset U, \qquad u^k \to \hat{u}, \qquad y^k \in Y(u^k), \qquad \text{and} \qquad y^k \to \hat{y}.$$

From the definition of $Y(u)$, there exists a sequence $\{x^k\}$ such that

$$x^k \in X(u^k) \qquad \text{and} \qquad y^k = f(x^k, u^k).$$

Since X is uniformly compact near $\hat{u}$, we may assume without loss of generality (by taking a subsequence if necessary) that x^k converges to some point $\hat{x}$. Then $\hat{x} \in X(\hat{u})$, since X is upper semicontinuous at $\hat{u}$. Since f is continuous at $(\hat{x}, \hat{u})$, then $f(x^k, u^k) \to f(\hat{x}, \hat{u})$. This implies that

$$\hat{y} = f(\hat{x}, \hat{u}) \in Y(\hat{u}).$$

Hence, Y is upper semicontinuous at $\hat{u}$.

(ii) Let

$$\{u^k\} \subset U, \qquad u^k \to \hat{u}, \qquad \text{and} \qquad \hat{y} \in Y(\hat{u}).$$

From the definition of $Y(\hat{u})$, there exists $\hat{x} \in X(\hat{u})$ such that $\hat{y} = f(\hat{x}, \hat{u})$. Since X is lower semicontinuous at $\hat{u}$, there exist a number m and a sequence $\{x^k\}$ such that

$$x^k \in X(u^k) \qquad \text{for} \quad k \geqq m \qquad \text{and} \qquad x^k \to \hat{x}.$$

If we let $y^k = f(x^k, u^k)$, then

$$y^k \in Y(u^k) \qquad \text{for} \quad k \geqq m \qquad \text{and} \qquad f(x^k, u^k) \to f(\hat{x}, \hat{u}),$$

in view of the continuity of f. Thus we have obtained that $y^k \to \hat{y}$, whence Y is lower semicontinuous at $\hat{u}$. This completes the proof of the lemma.

Next, in addition to X and f, let N be a point-to-set map from U into R^p such that $N(u) \subset Y(u)$, and let

$$M(u) = \{x \in X(u) : f(x, u) \in N(u)\}.$$

Then, we get the following lemma, which is also fundamental in this section. This lemma provides the relationships between the continuity of point-to-set maps in the decision and objective spaces.

Lemma 4.4.2

(i) If X is upper semicontinuous at $\hat{u}$, if N is upper semicontinuous at $\hat{u}$, and if f is continuous at $X(\hat{u}) \times \hat{u}$, then the map M is upper semicontinuous at $\hat{u}$.

(ii) If X is upper semicontinuous at $\hat{u}$ and uniformly compact near $\hat{u}$, if N is lower semicontinuous at $\hat{u}$, and if f is continuous and one-to-one on $X(\hat{u}) \times \hat{u}$, then M is lower semicontinuous at $\hat{u}$.

Proof

(i) Let

$$\{u^k\} \subset U, \qquad u^k \to \hat{u}, \qquad x^k \in M(u^k), \qquad \text{and} \qquad x^k \to \hat{x}.$$

Since X is upper semicontinuous at $\hat{u}$, $\hat{x} \in X(\hat{u})$. From the definition of $M(u)$, $f(x^k, u^k) \in N(u^k)$. From the continuity of f,

$$f(x^k, u^k) \to f(\hat{x}, \hat{u}).$$

Hence, $f(\hat{x}, \hat{u}) \in N(\hat{u})$, since the map N is upper semicontinuous at $\hat{u}$. Therefore, $\hat{x} \in M(\hat{u})$, and so the map M is upper semicontinuous at $\hat{u}$.

(ii) Let

$$\{u^k\} \subset U, \qquad u^k \to \hat{u}, \qquad \text{and} \qquad \hat{x} \in M(\hat{u}).$$

From the definition of $M(u)$, $f(\hat{x}, \hat{u}) \in N(\hat{u})$. Since N is lower semicontinuous at $\hat{u}$ and $N(u) \subset Y(u)$, there exist a number m and a sequence $\{x^k\}$ such that

$$x^k \in M(u^k) \quad \text{for} \quad k \geqq m \quad \text{and} \quad f(x^k, u^k) \to f(\hat{x}, \hat{u}).$$

Since X is uniformly compact near $\hat{u}$, the sequence $\{x^k\}$ has a cluster point $\bar{x}$. Since X is upper semicontinuous at $\hat{u}$, $\bar{x} \in X(\hat{u})$. Then, since f is continuous,

$$f(\bar{x}, \hat{u}) = f(\hat{x}, \hat{u}),$$

and so we have $\bar{x} = \hat{x}$, since $f(\cdot, \hat{u})$ is one-to-one on $X(\hat{u})$; that is, $\hat{x}$ is a unique cluster point of the bounded sequence $\{x^k\}$. This implies that $x^k \to \hat{x}$, and therefore M is lower semicontinuous at $\hat{u}$. This completes the proof of the lemma.

We can combine these lemmas and the theorems obtained in Sections 4.2 and 4.3 to yield a series of theorems as follows.

First, with regard to the continuity of the point-to-set map M_1, the following theorems are obtained. (Notice the fact that Y is uniformly compact near u if X is so, and f is continuous on cl $\bigcup_{u \in T} X(u) \times T$ along with some compact neighborhood T in U of $\hat{u}$.)

Theorem 4.4.1

The map M_1 is upper semicontinuous at $\hat{u} \in U$ under the following conditions:

(i) the map X is continuous at $\hat{u}$;
(ii) X is uniformly compact near $\hat{u}$;
(iii) f is continuous on $X(\hat{u}) \times \hat{u}$;
(iv) the map D is lower semicontinuous on $Y(\hat{u})$;
(v) $D(y)$ is a convex set for every $y \in S \cap Y(u)$ with $u \in T$, where S and T are some neighborhoods of $Y(\hat{u})$ and $\hat{u}$, respectively; and
(vi) $N_1(\hat{u}) = \mathscr{E}(Y(\hat{u}), \text{int } D)$.

Theorem 4.4.2

The map M_1 is lower semicontinuous at $\hat{u} \in U$ under the following conditions:

(i) the map X is continuous at $\hat{u}$;
(ii) X is uniformly compact near $\hat{u}$;
(iii) f is continuous on cl $\bigcup_{u \in T} X(u) \times T$ with a compact neighborhood T in U of $\hat{u}$ and one-to-one on $X(\hat{u}) \times \hat{u}$;
(iv) $N_1(u)$ is externally stable for every $u \in U$ near $\hat{u}$; and
(v) the map D is upper semicontinuous on $Y(\hat{u})$.

Next, in order to consider the continuity of the point-to-set map M_2, we assume for simplicity that the set of feasible solutions is X and that the vector-valued objective function from X into R^p is $f(x)$. The following theorems are readily obtained from Theorems 4.3.1 and 4.3.2, respectively.

Theorem 4.4.3

The map M_2 is upper semicontinuous at $\hat{v} \in V$ under the following conditions:

(i) $Y = f(X)$ is a closed set;
(ii) f is continuous on X;
(iii) the map $D(y, \cdot)$ is lower semicontinuous at $\hat{v}$ for each fixed $y \in Y$;
(iv) $D(y, v)$ is a convex set for all $y \in Y$ and every $v \in V$ near $\hat{v}$;
(v) $N_2(\hat{v}) = \mathscr{E}(Y, \operatorname{int} D(\cdot, \hat{v}))$.

Theorem 4.4.4

The map M_2 is lower semicontinuous at $\hat{v} \in V$ under the following conditions:

(i) X is a compact set;
(ii) f is continuous and one-to-one on X;
(iii) $N_2(v)$ is externally stable for every $v \in V$ near $\hat{v}$;
(iv) the map D is upper semicontinuous on $Y \times \hat{v}$.

Remark 4.4.1

Sufficient conditions for the point-to-set map $X(u)$ to be upper (resp. lower) semicontinuous at some point are given in Proposition 2.2.1 (resp. 2.2.2) when $X(u)$ is determined by inequalities such as

$$X(u) = \{x \in X : g(x, u) \leqq 0\}.$$

4.5 Stability of Properly Efficient Solutions

This section is devoted to the stability of the set of properly efficient points (in the objective space) for a multiobjective optimization problem

$$\text{(P)} \qquad \text{minimize} \quad f(x) \qquad \text{subject to} \quad x \in X$$

with respect to two types of perturbation: perturbation of the feasible set and perturbation of the domination structure. Every domination structure is assumed to be provided by a closed convex cone. As before, let U and V be parameter spaces that specify the feasible set and the domination cone, respectively. We define two point-to-set maps $Q_1: U \to R^p$ and $Q_2: V \to R^p$ by

$$Q_1(u) = \mathscr{P}(Y(u), D),$$

where D is a fixed, closed, convex cone in R^p, and by

$$Q_2(v) = \mathscr{P}(Y, D(v)),$$

where Y is a fixed set in R^p (cf. $N_1(u)$ and $N_2(v)$). In order for properly efficient sets to be well-defined, it is assumed throughout this section that D or $D(v)$ is a pointed closed convex cone for every $v \in V$ (If D is not pointed, $\mathscr{P}(Y, D) = \varnothing$). We shall obtain sufficient conditions for the point-to-set maps Q_1 or Q_2 to be upper or lower semicontinuous.

Theorem 4.5.1

The map Q_1 is upper semicontinuous at $\hat{u} \in U$ if the following conditions are satisfied:

(i) the map Y is continuous at $\hat{u}$;
(ii) $Q_1(\hat{u}) = \mathscr{E}(Y(\hat{u}), \operatorname{int} D)$.

Proof Let

$$\{u^k\} \subset U, \qquad u^k \to \hat{u}, \qquad y^k \in Q_1(u^k), \qquad \text{and} \qquad y^k \to \hat{y}.$$

Since Y is upper semicontinuous at $\hat{u}$, $\hat{y} \in Y(\hat{u})$. Hence, if $\hat{y}$ is not contained in

$$Q_1(\hat{u}) = \mathscr{E}(Y(\hat{u}), \operatorname{int} D),$$

there exists $\bar{y} \in Y(\hat{u})$ such that $\hat{y} - \bar{y} \in \operatorname{int} D$. Since Y is lower semicontinuous at $\hat{u}$, there exist a number m and a sequence $\{\bar{y}^k\}$ such that $\bar{y}^k \in Y(u^k)$ for $k \geqq m$ and $\bar{y}^k \to \bar{y}$. Then, since $y^k - \bar{y}^k \to \hat{y} - \bar{y} \in \operatorname{int} D$, $y^k - \bar{y}^k \in \operatorname{int} D$ for sufficiently large k. Hence, $y^k \notin \mathscr{E}(Y(u^k), \operatorname{int} D) \supset \mathscr{E}(Y(u^k), D) \supset Q_1(u^k)$, which contradicts the assumption. Therefore $\hat{y} \in Q_1(u)$. This completes the proof.

Indispensability of conditions (i) and (ii) in Theorem 4.5.1 can be understood by Examples 4.2.1, 4.2.2, and 4.2.5. Though the second condition is indispensable, it may be too restrictive. For example, let

$$Y(u) = \{(y_1, y_2) : (y_1)^2 + (y_2)^2 \leqq 1\} \qquad \text{for all} \quad u \in U = R,$$
$$D = R^2_+.$$

Then,

$$Q_1(u) = \{(y_1, y_2) : (y_1)^2 + (y_2)^2 \leqq 1, y_1 < 0, y_2 < 0\} \qquad \text{for all} \quad u,$$

and so Q_1 is not upper semicontinuous at any $u \in U$, despite the fact that this example is quite natural.

Theorem 4.5.2

The map Q_1 is lower semicontinuous at $\hat{u} \in U$ if the following conditions are satisfied:

(i) the map Y is continuous at $\hat{u}$;
(ii) Y is uniformly compact near $\hat{u}$;
(iii) $Y(u)$ is D-closed for every $u \in U$ near $\hat{u}$.

Proof Let

$$\{u^k\} \subset U, \qquad u^k \to \hat{u}, \qquad \text{and} \qquad \hat{y} \in Q_1(\hat{u}).$$

Since Y is lower semicontinuous at $\hat{u}$, there exist a number m and a sequence $\{y^k\}$ such that $y^k \in Y(u^k)$ for $k \geqq m$ and $y^k \to \hat{y}$. Since $Y(u^k)$ is bounded (condition (ii)) and D-closed (condition (iii)) for all k sufficiently large,

$$y^k \in Y(u^k) \subset \mathscr{E}(Y(u^k)) + D \subset \mathrm{cl}\, Q_1(u^k) + D$$

by Theorems 3.2.10 and 3.2.11. Thus, we may assume that there exist sequences $\{\bar{y}^k\}$ and $\{\hat{y}^k\}$ such that $\bar{y}^k \in \mathrm{cl}\, Q_1(u^k)$, $\hat{y}^k \in Q_1(u^k)$, $y^k - \bar{y}^k \in D$, and $\|\bar{y}^k - \hat{y}^k\| < 1/k$, for $k \geqq m$. Since Y is uniformly compact near $\hat{u}$, $\{\hat{y}^k\}$ may be assumed to converge to some $\bar{y}$. Since Y is upper semicontinuous at $\hat{u}$, $\bar{y} \in Y(\hat{u})$. Since

$$\|\bar{y}^k - \bar{y}\| \leqq \|\bar{y}^k - \hat{y}^k\| + \|\hat{y}^k - \bar{y}\| < 1/k + \|\hat{y}^k - \bar{y}\|,$$

$\bar{y}^k \to \bar{y}$ as $k \to \infty$. Hence $y^k - \bar{y}^k \to \hat{y} - \bar{y} \in D$, as D is closed. Since $\hat{y} \in Q_1(\hat{u})$, $\hat{y} = \bar{y}$. Thus, we have established that $\hat{y}^k \to \hat{y}$ and $\hat{y}^k \in Q_1(u^k)$ for $k \geqq m$. This completes the proof.

In view of Examples 4.2.6–4.2.9, each condition of the theorem cannot be dispenced with.

Theorem 4.5.3

The map Q_2 is upper semicontinuous at $\hat{v} \in V$ if the following conditions are satisfied:

(i) Y is a closed set;
(ii) the map $\mathrm{int}\, D$ is lower semicontinuous at $\hat{v}$;
(iii) $Q_2(\hat{v}) = \mathscr{E}(Y, \mathrm{int}\, D(\hat{v}))$.

Proof Let

$$\{v^k\} \subset V, \qquad v^k \to \hat{v}, \qquad y^k \in Q_2(v^k), \qquad \text{and} \qquad y^k \to \hat{y}.$$

Since Y is a closed set, $\hat{y} \in Y$. If we suppose that $\hat{y} \notin Q_2(\hat{v}) = \mathscr{E}(Y, \operatorname{int} D(\hat{v}))$, there exists $y \in Y$ such that

$$\hat{y} - y \in \operatorname{int} D(v).$$

Since $y^k - y \to \hat{y} - y \neq 0$, in view of Lemma 4.2.1, $y^k - y \in D(v^k)\backslash\{0\}$ for all k sufficiently large. Then, for those k, $y^k \notin Q_2(v^k)$, which leads to a contradiction. Hence $\hat{y} \in Q_2(\hat{v})$, and the proof is complete.

Indispensability of each condition can be understood by Examples 4.2.1, 4.3.2, and 4.3.4.

Theorem 4.5.4

The map Q_2 is lower semicontinuous at $\hat{v} \in V$ if the following condition is satisfied:

(i) the map D is upper semicontinuous at $\hat{v}$.

Proof Let

$$\{v^k\} \subset V, \qquad v^k \to \hat{v}, \qquad \text{and} \qquad \hat{y} \in Q_2(\hat{v}).$$

Then, in view of Theorem 3.1.2, there exists a convex cone D' such that

$$\hat{y} \in \mathscr{E}(Y, D') \qquad \text{and} \qquad D(\hat{v})\backslash\{0\} \subset \operatorname{int} D'.$$

We will prove that $D(v^k)\backslash\{0\} \subset \operatorname{int} D'$ for all k sufficiently large. If we suppose the contrary, then, by considering a subsequence if necessary, we may take a sequence $\{d^k\}$ such that

$$d^k \in D(v^k), \qquad \|d^k\| = 1 \qquad \text{and} \qquad d^k \notin \operatorname{int} D' \qquad \text{for all} \quad k = 1, 2, \ldots.$$

We may assume without loss of generality that $\{d^k\}$ converges to some $\hat{d}$ with $\|\hat{d}\| = 1$. Then, $\hat{d} \in D(\hat{v})$, since D is upper semicontinuous at $\hat{v}$. On the other hand, $\hat{d} \notin \operatorname{int} D'$ since $d^k \notin \operatorname{int} D'$ for any k. This contradicts the fact that $D(\hat{v})\backslash\{0\} \subset \operatorname{int} D'$. Therefore, $D(v^k)\backslash\{0\} \subset \operatorname{int} D'$ for all k sufficiently large. Hence $\hat{y} \in Q_2(v^k) = \mathscr{P}(Y, D(v^k))$ for those k by Theorem 3.1.2, and the proof of the theorem is completed.

Examples 4.3.5, 4.3.6, and 4.3.10 illustrate that the condition (i) is essential to the theorem.

Remark 4.5.1

Comparing Theorems 4.5.1–4.5.3 with Theorems 4.2.1, 4.2.2, and 4.3.1, we can conclude that the conditions stated in the theorems in this subsection are also sufficient for the upper and lower semicontinuity of the maps N_1

and N_2, respectively, if Q_i is replaced by N_i $(i = 1, 2)$. However, the condition in Theorem 4.5.4 is not sufficient for the lower semicontinuity of N_2 as can be seen by Examples 4.3.5 and 4.3.6.

Remark 4.5.2

In a linear multiobjective optimization problem, proper efficiency is equivalent to efficiency (Theorem 3.1.7). Therefore, the results in this subsection are valid for the stability of the efficient set in linear problems.

5 LAGRANGE DUALITY

This chapter will be devoted to Lagrange duality for efficient solutions in multiobjective optimization where the domination structure is supposed to be a pointed closed convex cone. As was seen in the previous chapters, efficient solutions correspond to minimal (maximal) solutions in ordinary mathematical programming. Therefore, for convenience in this chapter we define *D-minimizing* as finding efficient solutions with respect to a cone ordering $\leq_D$; in particular, we define *minimizing* for cases with the cone ordering with R^p_+. Similarly we use the notation $\mathrm{Min}_D\, Y$ for representing $\mathscr{E}(Y, \leq_D)$; in particular, we define $\mathrm{Min}\, Y$ as $\mathscr{E}(Y, \leq_{R^p_+})$. D-maximization and Max_D are used in a similar fashion for the cone ordering $\leq_{-D}$.

The first section gives several results on duality in linear multiobjective programming, which are mainly due to Gale *et al.* [G2], Isermann [I5–I7], and Nakayama [N6]. The next section develops duality in nonlinear multiobjective optimization in parallel with ordinary convex programming, which was originally developed by Tanino and Sawaragi [T7, T9]. Some of modified versions are due to Nakayama [N5].

The third section considers duality from a geometric viewpoint. A geometric meaning of the vector-valued Lagrangian function will be given. Other duality theorems will be derived from geometric consideration. The main parts of this section are due to Nakayama [N5, N7]. Earlier results along a similar line can be found in Nieuwenhuis [N15] and Jahn [J1].

5.1 Linear Cases

Possibly the first work on duality for multiobjective optimization was given by Gale *et al.* [G2] for linear cases. They considered the following matrix optimization problem.

Let D, Q, M, and N be pointed convex polyhedral cones in R^p, R^m, R^n, and R^r, respectively. This means, in particular, that int $D^\circ \neq \emptyset$. In what follows, we shall suppose that int $N \neq \emptyset$. Furthermore, we shall identify the set of all $m \times n$ matrices with $R^{m\times n}$. This notation is also valid for matrices with other dimensions. Define

$$\bar{\mathscr{K}}_+ := \{K \in R^{p\times r} : KN \subset D\},$$

$$\mathscr{K}_+ := \{K \in R^{p\times r} : K(\operatorname{int} N) \subset D\backslash\{0\}\}.$$

Then, the order $\geqq$ for $p \times r$ matrices is introduced as follows:

$$K^1 \geqq K^2 \quad \text{if and only if} \quad K^1 - K^2 \in \bar{\mathscr{K}}_+,$$

$$K^1 \geq K^2 \quad \text{if and only if} \quad K^1 - K^2 \in \mathscr{K}_+.$$

The following problem is a simple extension of that given by Gale *et al.* [G2]:

($\mathrm{P_{GKT}}$) Minimize K subject to

$$Cx \leqq_D Ky, \qquad Ax \geqq_Q By, \qquad x \geqq_M 0, \qquad y >_N 0.$$

Here, $A \in R^{m\times n}$, $B \in R^{m\times r}$, $C \in R^{p\times n}$, $K \in R^{p\times r}$, $x \in R^n$, and $y \in R^r$.

The dual problem associated with problem ($\mathrm{P_{GKT}}$) is then given by

($\mathrm{D_{GKT}}$) Maximize K subject to

$$B^{\mathrm{T}}\lambda \geqq_{N^\circ} K^{\mathrm{T}}\mu, \qquad A^{\mathrm{T}}\lambda \leqq_{M^\circ} C^{\mathrm{T}}\mu, \qquad \lambda \geqq_{Q^\circ} 0, \qquad \mu >_{D^\circ} 0.$$

Remark 5.1.1

The problems ($\mathrm{P_{GKT}}$) and ($\mathrm{D_{GKT}}$) represent a class of matrix optimization that includes vector optimization as a special case. In fact, in the case where B and K are vectors and y is a positive scalar, problem ($\mathrm{P_{GKT}}$) reduces to the usual formulation of vector optimization problems; that is,

($\mathrm{P'_{GKT}}$) Minimize k subject to

$$Cx \leqq_D k, \qquad Ax \geqq_Q b, \qquad x \geqq_M 0.$$

The dual problem associated with the problem ($\mathrm{P'_{KGT}}$) then becomes

($\mathrm{D'_{GKT}}$) Maximize k subject to

$$b^T\lambda \geq k^T\mu, \qquad A^{\mathrm{T}}\lambda \leqq_{M^\circ} C^{\mathrm{T}}\mu, \qquad \lambda \geqq_{Q^\circ} 0, \qquad \mu >_{D^\circ} 0.$$

Before proceeding to duality for the stated problem, we will prove the theorem of alternative in a generalized form.

Lemma 5.1.1

For a matrix $A \in R^{m \times n}$ and a convex cone $S \subset R^n$, set

$$AS := \{Ax : x \in S\}.$$

Then

$$(AS)^\circ = \{\lambda \in R^m : A^T\lambda \in S^\circ\}.$$

Proof Easy.

Lemma 5.1.2 (*Farkas Lemma*)

In order that $\langle b, \lambda \rangle \leqq 0$ for any $\lambda \in Q^\circ$ such that $-A^T\lambda \in M^\circ$, it is necessary and sufficient that $Ax \geqq_Q b$ for some $x \geqq_M 0$.

Proof The given λ condition is equivalent to

$$-b \in ((-AM)^\circ \cap Q^\circ)^\circ.$$

Furthermore, since M and Q are convex polyhedral cones, from Proposition 2.1.6

$$((-AM)^\circ \cap Q^\circ)^\circ = -AM + Q.$$

Finally, the given λ condition is equivalent to

$$(b + Q) \cap (AM) \neq \emptyset,$$

which is also equivalent to the given x condition.

The following two lemmas are extensions of those by Gale *et al.* [G2].

Lemma 5.1.3

In order that $B^T\lambda \in N^\circ \backslash \{0\}$ for no $\lambda \in Q^\circ$ such that $-A^T\lambda \in M^\circ$, it is necessary and sufficient that $By \leqq_Q Ax$ for some $x \geqq_M 0$ and $y >_N 0$.

Proof The given λ condition $\Leftrightarrow B^T\lambda \notin N^\circ \backslash \{0\}$ for any $\lambda \in Q^\circ \cap (-AM)^\circ$

$\Leftrightarrow \langle y, B^T\lambda \rangle \not> 0$ for some $y \in \operatorname{int} N$ and any $\lambda \in Q^\circ \cap (-AM)^\circ$

$\Leftrightarrow \langle b, \lambda \rangle \not> 0$ for any $\lambda \in Q^\circ \cap (-AM)^\circ$ (Here, $b = By$)

$\Leftrightarrow Ax \geqq_Q b$ for some $x \geqq_M 0$ (from Lemma 5.1.2)

$\Leftrightarrow Ax \geqq_Q By$ for some $x \geqq_M 0$ and $y >_N 0$.

Lemma 5.1.4

$\hat{K}$ is an efficient solution to problem ($\mathrm{P_{GKT}}$) if and only if

(i) $C\bar{x} \leqq_D \hat{K}\bar{y}$

for some $\bar{x} \in M$ and $\bar{y} \in \operatorname{int} N$ such that $A\bar{x} \geqq_Q B\bar{y}$, and

(ii) $Cx \leq_D \hat{K}y$

for no $x \in M$ and $y \in N$ such that $Ax \geqq_Q By$.

Similarly, $\hat{K}$ is an efficient solution to the problem ($\mathrm{D_{GKT}}$), if and only if

(ii′) $B^{\mathrm{T}}\bar{\lambda} \geqq_{N^\circ} \hat{K}^{\mathrm{T}}\bar{\mu}$

for some $\bar{\lambda} \in Q^\circ$ and $\bar{\mu} \in \operatorname{int} D^\circ$ such that $A^{\mathrm{T}}\bar{\lambda} \leqq_{M^\circ} C^{\mathrm{T}}\bar{\mu}$, and

(i′) $B^{\mathrm{T}}\lambda \geq_{N^\circ} \hat{K}^{\mathrm{T}}\mu$

for no $\lambda \in Q^\circ$ and $\mu \in D^\circ$ such that $A^{\mathrm{T}}\lambda \leqq_{M^\circ} C^{\mathrm{T}}\mu$.

Proof We shall prove here the first part of the lemma, since the second one follows readily in a similar fashion.

if Suppose that $\hat{K}$ satisfying condition (i) is not a solution to problem ($\mathrm{P_{GKT}}$). Then there exists a matrix K' such that

$$K' \leq \hat{K}$$

and

$$C\tilde{x} \leqq_D K'\tilde{y}$$

for some $\tilde{x} \geqq_M 0$ and $\tilde{y} >_N 0$ such that $A\tilde{x} \geqq_Q B\tilde{y}$. Hence, we have $C\tilde{x} \leqq_D K'\tilde{y} \leq_D \hat{K}\tilde{y}$, which contradicts condition (ii).

only if Suppose, to the contrary, that $\hat{K}$ does not satisfy condition (ii). Then there exist some $x' \in M$ and $y' \in N$ such that

$$Cx' \leq_D \hat{K}y' \qquad \text{and} \qquad Ax' \geqq_Q By'.$$

By taking condition (i) into account, it follows that

$$C(\bar{x} + x') \leq_D \hat{K}(\bar{y} + y')$$

for $\bar{x} + x' \in M$ and $\bar{y} + y' \in \operatorname{int} N$ such that $A(\bar{x} + x') \geqq_Q B(\bar{y} + y')$. Take a vector $d' \in D\backslash\{0\}$ in such a way that $d' \leq_D \hat{K}(\bar{y} + y') - C(\bar{x} + x')$. Choose a matrix $\Delta K \in \mathscr{K}_+$ such that $\Delta K(\bar{y} + y') = d'$. In fact, for some vector e of N°

with $\langle e, \bar{y} + y' \rangle = 1$, a possible ΔK is given by $\Delta K := (d'_1 e, d'_2 e, \ldots, d'_p e)^{\mathrm{T}}$, where d'_i is the ith component of d'. Then

$$(\hat{K} - \Delta K)(\bar{y} + y') \geq_D C(\bar{x} + x'),$$

and

$$K' := \hat{K} - \Delta K \leq \hat{K},$$

which implies that $\hat{K}$ cannot be a solution to the problem ($\mathrm{P_{GKT}}$).

A duality between the problems ($\mathrm{P_{GKT}}$) and ($\mathrm{D_{GKT}}$) can be stated as follows.

Theorem 5.1.1 *(Gale–Kuhn–Tucker Duality)*

(i) A matrix $\hat{K}$ is an efficient solution to problem ($\mathrm{P_{GKT}}$) if and only if it is a solution to problem ($\mathrm{D_{GKT}}$).

(ii) If $\hat{K}$ is an efficient solution to problem ($\mathrm{P_{GKT}}$) for some $\hat{x} \geqq_M 0$ and $\hat{y} >_N 0$, then we have $\hat{K}\hat{y} = C\hat{x}$.

(iii) If $\hat{K}$ is a feasible solution to problem ($\mathrm{P_{GKT}}$) for some $\hat{x} \geqq_M 0$ and $\hat{y} >_N 0$ and is also a feasible solution to problem ($\mathrm{D_{GKT}}$) for some $\hat{\lambda} \in Q^\circ$ and $\hat{\mu} \in \operatorname{int} D^\circ$, then $\hat{K}$ is an efficient solution to both problems ($\mathrm{P_{GKT}}$) and ($\mathrm{D_{GKT}}$).

Proof (i) It is easily shown that, for any two convex cones S and T

$$(S \times T)^\circ = S^\circ \times T^\circ.$$

Now with a help of Lemma 5.1.3 we have

$$\text{(i) of Lemma 5.1.4} \Leftrightarrow \begin{bmatrix} A \\ -C \end{bmatrix} \bar{x} \geqq_{Q \times D} \begin{bmatrix} B \\ -\hat{K} \end{bmatrix} \bar{y} \qquad \text{for some} \quad \bar{x} \in M$$

$$\text{and } \bar{y} \in \operatorname{int} N$$

$$\Leftrightarrow (B^{\mathrm{T}}, -\hat{K}^{\mathrm{T}}) \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \geq_{N^\circ} 0 \qquad \text{for no} \quad (\lambda, \mu) \in Q^\circ \times D^\circ$$

$$\text{such that} \quad (A^{\mathrm{T}}, -C^{\mathrm{T}}) \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \leqq_{M^\circ} 0$$

$$\Leftrightarrow \text{(i}'\text{) of Lemma 5.1.4.}$$

Similarly,

$$\text{(ii$'$) of Lemma 5.1.4} \Leftrightarrow \begin{bmatrix} -A^{\mathrm{T}} \\ B^{\mathrm{T}} \end{bmatrix} \bar{\lambda} \geqq_{M^\circ \times N^\circ} \begin{bmatrix} -C^{\mathrm{T}} \\ \hat{K}^{\mathrm{T}} \end{bmatrix} \bar{\mu} \qquad \text{for some} \quad \bar{\lambda} \in Q^\circ$$

$$\text{and } \bar{\mu} \in \operatorname{int} D^\circ$$

$$\Leftrightarrow (-C, \hat{K}) \begin{bmatrix} x \\ y \end{bmatrix} \geq_D 0 \qquad \text{for no} \quad (x, y) \in M \times N$$

$$\text{such that} \quad (-A, B) \begin{bmatrix} x \\ y \end{bmatrix} \leqq_Q 0$$

$$\Leftrightarrow \text{(ii) of Lemma 5.1.4.}$$

On the other hand, statements (ii) and (iii) follow immediately from Lemma 5.1.4. This completes the proof.

Another formulation for vector optimization with more reciprocity was suggested by Kornbluth [K8]:

(P_K)

D-minimize $\quad Cx \quad$ subject to $\quad Ax \geqq_Q By, \quad x \geqq_M 0, \quad y >_N 0.$

(D_K)

D-maximize $\quad B^{\mathrm{T}}\lambda \quad$ subject to $\quad A^{\mathrm{T}}\lambda \leqq_{M^\circ} C^{\mathrm{T}}\mu, \quad \lambda \geqq_{Q^\circ} 0, \quad \mu >_{D^\circ} 0.$

Theorem 5.1.2 (Kornbluth Duality)

There exists an efficient solution $\hat{x}$ to problem (P_K) for some $y = \hat{y}$ if and only if there exists an efficient solution $\hat{\lambda}$ to problem (D_K) for some $\mu = \hat{\mu}$.

Proof See Kornbluth [K8].

The following relationship between problems (P_{GKT}) and (P_K) (resp. (D_{GKT}) and (D_K)) was revealed by Rödder [R9].

Theorem 5.1.3

(i) If $(\hat{K}, \hat{x}, \hat{y})$ solves problem (P_{GKT}), then $\hat{x}$ is an efficient solution to problem (P_K) for $y = \hat{y}$.

(ii) If $\hat{x}$ is an efficient solution to problem (P_K) for $y = \hat{y}$, then there exists a matrix $\hat{K}$ such that $(\hat{K}, \hat{x}, \hat{y})$ solves problem (P_{GKT}).

(iii) Statements analogous to (i) and (ii) hold for problems (D_{GKT}) and (D_K).

Proof (i) is obvious. Since (iii) is dual to (i) and (ii), we shall prove (ii) here. Suppose that $\hat{x}$ is an efficient solution to problem ($\mathrm{P_K}$). It is readily shown that there exist $\hat{\mu} \in D^\circ$ such that

$$\hat{\mu}^{\mathrm{T}} C\hat{x} \leqq \hat{\mu}^{\mathrm{T}} Cx$$

for all $x \geqq_M 0$ such that

$$Ax \geqq_Q B\hat{y}.$$

Considering the dual problem associated with this scalarized linear programming problem, it follows that there exists $\hat{\lambda} \in Q^\circ$ such that

$$\hat{y}^{\mathrm{T}} B^{\mathrm{T}} \hat{\lambda} \geqq \hat{y}^{\mathrm{T}} B^{\mathrm{T}} \lambda$$

for any $\lambda \in Q^\circ$ such that

$$A^{\mathrm{T}}\lambda \leqq_{M^\circ} C^{\mathrm{T}}\hat{\mu}.$$

According to the well-known duality theorem in linear programming, we have

$$\hat{y}^{\mathrm{T}} B^{\mathrm{T}} \hat{\lambda} = \hat{\mu}^{\mathrm{T}} C\hat{x}.$$

This condition implies that

$$K\hat{y} = C\hat{x} \qquad \text{and} \qquad \hat{\mu}^{\mathrm{T}} K = \hat{\lambda}^{\mathrm{T}} B$$

have a common solution $\hat{K}$ (see, for example, Penrose [P1]). Hence, it follows immediately from (iii) of Theorem 5.1.1 that $\hat{K}$ is an efficient solution to problem ($\mathrm{P_{GKT}}$). This completes the proof.

On the other hand, Isermann [I6] has given a more attractive formulation without auxiliary parameters such as y and μ. In the following, we shall consider it in an extended form. Let $\mathscr{L}_0$ be class of $p \times m$ matrices such that there exists $\mu \in \operatorname{int} D^\circ$ holding $\Lambda^{\mathrm{T}}\mu \in Q^\circ$ and $A^{\mathrm{T}}\Lambda^{\mathrm{T}}\mu \leqq_{M^\circ} C^{\mathrm{T}}\mu$. The primal and dual problems are then defined as

($\mathrm{P_I}$)

$$D\text{-minimize} \quad \{Cx : x \in X\} \quad \text{where} \quad X := \{x \in M : Ax \geqq_Q b\}.$$

($\mathrm{D_I}$) $$D\text{-maximize} \quad \{\Lambda b : \Lambda \in \mathscr{L}_0\}.$$

The following duality properties hold for these problems. Here, we denote by $\mathrm{Min}_D(\mathrm{P_I})$ and $\mathrm{Max}_D(\mathrm{D_I})$) the set of D-minimal (D-maximal) points in the objective space of problem ($\mathrm{P_I}$) (problem ($\mathrm{D_I}$)).

Theorem 5.1.4†

(i) $\Lambda b \not\geq_D Cx$ for all $(\Lambda, x) \in \mathscr{L}_0 \times X$.

(ii) Suppose that $\bar{\Lambda} \in \mathscr{L}_0$ and $\bar{x} \in X$ satisfy $\bar{\Lambda}b = C\bar{x}$. Then $\bar{\Lambda}$ is a D-maximal solution to dual problem $(\mathrm{D_I})$ and $\bar{x}$ is a D-minimal solution to primal problem $(\mathrm{P_I})$.

(iii) $\mathrm{Min}_D(\mathrm{P_I}) = \mathrm{Max}_D(\mathrm{D_I})$.

Proof

(i) Suppose, to the contrary, that there exist some $\bar{x} \in X$ and $\bar{\Lambda} \in \mathscr{L}_0$ such that

$$\bar{\Lambda}b \geq_D C\bar{x}. \tag{5.1.2}$$

Note here by definition that there exists $\bar{\mu} \in \operatorname{int} D^\circ$ such that

$$(\bar{\Lambda}A)^{\mathrm{T}}\bar{\mu} \leqq_{M^\circ} C^{\mathrm{T}}\bar{\mu}, \qquad \text{and} \qquad \bar{\Lambda}^{\mathrm{T}}\bar{\mu} \in Q^\circ.$$

Therefore, since $\bar{x} \in M$, we have

$$\langle \bar{\mu}, \bar{\Lambda}A\bar{x} \rangle \leqq \langle \bar{\mu}, C\bar{x} \rangle. \tag{5.1.3}$$

Furthermore, from Eq. (5.1.2)

$$\langle \bar{\mu}, \bar{\Lambda}b \rangle > \langle \bar{\mu}, C\bar{x} \rangle. \tag{5.1.4}$$

It then follows from Eqs. (5.1.3) and (5.1.4) that

$$\langle \bar{\mu}, \bar{\Lambda}A\bar{x} \rangle < \langle \bar{\mu}, \bar{\Lambda}b \rangle. \tag{5.1.5}$$

On the other hand, since $\bar{\Lambda}^{\mathrm{T}}\mu \in Q^\circ$ and $A\bar{x} \geqq_Q b$, we have $\langle \bar{\mu}, \bar{\Lambda}A\bar{x} \rangle \geqq \langle \bar{\mu}, \bar{\Lambda}b \rangle$, which contradicts Eq. (5.1.5).

(ii) Suppose, to the contrary, that $\bar{\Lambda}b \notin \mathrm{Max}_D(\mathrm{D_I})$. Then there exists $\hat{\Lambda} \in \mathscr{L}_0$ such that

$$\hat{\Lambda}b \geq_D \bar{\Lambda}b = C\bar{x},$$

which contradicts (i). Therefore, $\bar{\Lambda}$ is a D-maximal solution to the dual problem. In a similar fashion, we can conclude that $\bar{x}$ is a D-minimal solution to the primal problem.

(iii) First we shall prove $\mathrm{Min}_D(\mathrm{P_I}) \subset \mathrm{Max}_D(\mathrm{D_I})$. Let $\hat{x}$ be a D-minimal solution to the primal problem $(\mathrm{P_I})$. Then it is readily seen from Theorem 3.4.7 that there exists some $\hat{\mu} \in \operatorname{int} D^\circ$ such that

$$\hat{\mu}^{\mathrm{T}}C\hat{x} = \operatorname*{Min}_{x \in X} \hat{\mu}^{\mathrm{T}}Cx.$$

† Isermann [16]

It is sufficient to prove the statement in the case where $\hat{x}$ is a basic solution. Transform the original inequality constraints $Ax \geqq_Q b$ into

$$Ax - y = b, \qquad \text{and} \qquad y \geqq_Q 0.$$

Let B denote the submatrix of $[A, -I]$ that consists of m columns corresponding to the basic variables. Then from the initial simplex tableu

$$\begin{bmatrix} A & -I & b \\ \hat{\mu}^T C & 0 & 0 \end{bmatrix}$$

we obtain the final tableu

$$\begin{bmatrix} B^{-1}A & -B^{-1} & B^{-1}b \\ \hat{\mu}^T(C - C_B B^{-1}A) & \hat{\mu}^T C_B B^{-1} & -\hat{\mu}^T C_B B^{-1} b \end{bmatrix}$$

by using the simplex method. According to the well-known properties of linear programming, we have

$$(C - C_B B^{-1}A)^T \hat{\mu} \geqq_{M^\circ} 0$$
$$(C_B B^{-1})^T \hat{\mu} >_{Q^\circ} 0.$$

Setting here $\hat{\Lambda} = C_B B^{-1}$, these relations can be rewritten as

$$C^T \hat{\mu} \geqq_{M^\circ} A^T \hat{\Lambda}^T \hat{\mu},$$
$$\hat{\Lambda}^T \hat{\mu} \in Q^\circ,$$

which shows that $\hat{\Lambda} \in \mathscr{L}_0$.

On the other hand,

$$\hat{\Lambda} b = C_B B^{-1} b = C_B x_B = C(x_B, 0) = C\hat{x}.$$

In view of the result (ii), the last relation implies that $\hat{\Lambda}$ is a D-maximal solution to dual problem $(\mathrm{D_I})$. Hence we have $\mathrm{Min}_D(\mathrm{P_I}) \subset \mathrm{Max}_D(\mathrm{D_I})$.

Next, we shall prove the reverse inclusion. Suppose that $\hat{\Lambda}$ is a D-maximal solution to the dual problem $(\mathrm{D_I})$. Then it is clear that for every $\mu \in \operatorname{int} D^\circ$ there exists no Λ with $\Lambda^T \mu \in Q^\circ$ such that

$$A^T \Lambda^T \mu \leqq_{M^\circ} C^T \mu, \qquad \text{and} \qquad \mu^T \Lambda b > \mu^T \hat{\Lambda} b.$$

Setting here $\lambda = \Lambda^T \mu$, it follows that for no $\lambda \in Q^\circ$ and no $\mu \in \operatorname{int} D^\circ$

$$A^T \lambda \leqq_{M^\circ} C^T \mu \qquad \text{and} \qquad \lambda^T b > \mu^T \hat{\Lambda} b. \tag{5.1.6}$$

More strongly, we can conclude that there exist no $\lambda \in Q^\circ$ and no $\mu \in D^\circ$ satisfying Eqs. (5.1.6). In fact, suppose to the contrary, that there exist some $\lambda' \in Q^\circ$ and $\mu' \in D^\circ$ such that

$$A^T \lambda' \leqq_{M^\circ} C^T \mu' \qquad \text{and} \qquad \lambda'^T b > \mu'^T \hat{\Lambda} b.$$

On the other hand, since $\hat{\Lambda}$ is a solution to the dual problem, there exist $\hat{\lambda} = \hat{\Lambda}^T\hat{\mu} \in Q^\circ$ and $\hat{\mu} \in \operatorname{int} D^\circ$ such that

$$A^T\hat{\lambda} \leqq_{M^\circ} C^T\hat{\mu} \qquad \text{and} \qquad \hat{\lambda}^T b = \hat{\mu}^T\hat{\Lambda}^T b.$$

Therefore, we have

$$A^T(\hat{\lambda} + \lambda') \leqq_{M^\circ} C^T(\hat{\mu} + u')$$
$$(\hat{\lambda} + \lambda')^T b > (\hat{\mu} + \mu')^T\hat{\Lambda} b.$$

This implies the existence of solutions $\hat{\mu} + \mu' \in \operatorname{int} D^\circ$, $\hat{\lambda} + \lambda' \in Q^\circ$ to Eqs. (5.1.6), which is a contradiction.

Rewriting Eqs. (5.1.6), we may see that there exist no $\lambda \in Q^0$ and no $\mu \in D^0$ for which

$$\begin{bmatrix} -A & b \\ C & -\bar{\Lambda} b \end{bmatrix}^T \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \in M^\circ \times R^1_+ \backslash \{0\}$$

is satisfied. Thus from Lemma 5.1.5 below, there exists a $(\hat{x}\ \alpha) \in \operatorname{int}(M^0 \times R^1_+)^0 = \operatorname{int}(M \times R^1_+)$ satisfying

$$\begin{bmatrix} -A & b \\ C & -\Lambda b \end{bmatrix} \begin{bmatrix} \hat{x} \\ \alpha \end{bmatrix} \in (-Q) \times (-D)$$

Since $\alpha > 0$, we finally have

$$A\hat{x} \geqq_Q b \qquad \text{and} \qquad C\hat{x} \leqq_D \hat{\Lambda} b. \tag{5.1.7}$$

By using result (i), the last relation reduces to

$$C\hat{x} = \hat{\Lambda} b. \tag{5.1.8}$$

From the result (ii), the relations (5.1.7) and (5.1.8) imply that $\hat{x}$ is a D-minimal solution to the primal problem. This completes the proof.

The following lemma, which was used in the proof of Theorem 5.1.4, is a simple extension of that due to Gale [G1] and also a special case of Lemma 5.1.3 with $M = \{0\}$.

Lemma 5.1.5

Either

(I) $\qquad B^T\lambda \in N^\circ \backslash \{0\} \qquad$ for some $\quad \lambda \in Q^\circ$

or

(II) $\qquad By \in -Q \qquad$ for some $\quad y \in \operatorname{int} N$

holds, but never both.

5.2 Duality in Nonlinear Multiobjective Optimization

This section will be concerned with duality in nonlinear multiobjective optimization problem formulated as follows:

(P) $\qquad D\text{-minimize} \qquad \{f(x) : x \in X\}$

where

$$X := \{x \in X' : g(x) \leqq_Q 0,\ X' \subset R^n\}.$$

Since efficient solutions correspond to maximal solutions in ordinary mathematical programming, it is hard to derive duality for efficient solutions in the absence of the compactness condition. Throughout this section, we impose the following assumptions:

(i) X' is a nonempty compact convex set,
(ii) D and Q are pointed closed convex cones with nonempty interiors of R^p and R^m, respectively,
(iii) f is continuous and D-convex,
(iv) g is continuous and Q-convex.

Under the assumptions, it can be readily shown that for every $u \in R^m$, both sets

$$X(u) := \{x \in X' : g(x) \leqq_Q u\}$$

and

$$Y(u) := f[X(u)] := \{y \in R^p : y = f(x),\ x \in X',\ g(x) \leqq_Q u\} \tag{5.2.1}$$

are compact, $X(u)$ is convex, and $Y(u)$ is D-convex. Although these conditions might seem too strong, something like these conditions would be more or less inevitable as long as we consider efficient solutions. Several authors have been recently attempting to relax the compactness condition by using the notion of inf instead of min, which will be restated later.

5.2.1 Perturbation (or Primal) Map and Lagrange Multiplier Theorem

Let us consider the primal problem (P) by embedding it in a family of perturbed problems with $Y(u)$ given by (5.2.1):

(P_u) $\qquad D\text{-minimize} \qquad Y(u).$

Clearly primal problem (P) is identical to problem ($\mathrm{P_u}$) with $u = 0$. Now define the set Γ as

$$\Gamma := \{u \in R^m : X(u) \neq \varnothing\}.$$

It is known that the set Γ is convex (see, for example, Luenberger [L14]).

Definition 5.2.1 (*Perturbation* (*or Primal*) *Map*)

The point-to-set map $\bar{w}\colon \Gamma \to R^p$ defined by

$$W(u) := \mathrm{Min}_D \mathrm{Y}(\mathrm{u})$$

is called a perturbation (or primal) map.

Observe that the perturbation map corresponds to the perturbation (or primal) function

$$w(u) = \min\{f(x) : x \in X', g(x) \leqq u\}$$

in ordinary mathematical programming. Obviously the original problem (P) can be regarded as determining the set $W(0)$ and $f^{-1}[W(0)] \cap X$. In the following we shall investigate the properties of W.

Proposition 5.2.1

For every $u \in \Gamma$,

$$W(u) + D = Y(u) + D.$$

Proof Note first that

$$W(u) = \mathrm{Min}_D\, Y(u) \subset Y(u).$$

On the other hand, since $Y(u)$ is D-compact, Theorem 3.2.9 yields

$$Y(u) \subset W(u) + D,$$

from which

$$Y(u) + D \subset W(u) + D + D = W(u) + D,$$

because D is a convex cone. This completes the proof.

Proposition 5.2.2

For each $u \in \Gamma$, $W(u)$ is a D-convex set in R^p.

Proof Immediate from Proposition 5.2.1 and the D-convexity of $Y(u)$.

Proposition 5.2.3

The map W is D-monotone on Γ, namely,

$$W(u^1) \subset W(u^2) + D$$

for any $u^1, u^2 \in \Gamma$ such that $u^1 \leqq_Q u^2$.

Proof Obviously, $Y(u^1) \subset Y(u^2)$ whenever $u^1 \leqq_Q u^2$. Hence

$$W(u^1) \subset Y(u^1) \subset Y(u^2) \subset W(u^2) + D.$$

Proposition 5.2.4

W is a D-convex point-to-set map on Γ.

Proof In view of Proposition 5.2.1, it suffices to show that

$$\alpha Y(u^1) + (1 - \alpha)Y(u^2) \subset Y(\alpha u^1 + (1 - \alpha)u^2) + D$$

for any $u^1, u^2 \in \Gamma$ and any $\alpha \in [0, 1]$. If we suppose that

$$y \in \alpha Y(u^1) + (1 - \alpha)Y(u^2),$$

then there exist $x^1, x^2 \in X'$ such that

$$g(x^1) \leqq_Q u^1, \qquad g(x^2) \leqq_Q u^2, \qquad \text{and} \qquad y = \alpha f(x^1) + (1 - \alpha)f(x^2).$$

Since X' is a convex set, $\alpha x^1 + (1 - \alpha)x^2 \in X'$. Furthermore, from the Q-convexity of g,

$$g(\alpha x^1 + (1 - \alpha)x^2) \leqq_Q \alpha g(x^1) + (1 - \alpha)g(x^2) \leqq_Q \alpha u^1 + (1 - \alpha)u^2,$$

which implies $\alpha x^1 + (1 - \alpha)x^2 \in X(\alpha u^1 + (1 - \alpha)u^2)$ and, thus,

$$f(\alpha x^1 + (1 - \alpha)x^2) \in Y(\alpha u^1 + (1 - \alpha)u^2).$$

On the other hand, from the D-convexity of f

$$\alpha f(x^1) + (1 - \alpha)f(x^2) \in f(\alpha x^1 + (1 - \alpha)x^2) + D.$$

Finally we have

$$y \in Y(\alpha u^1 + (1 - \alpha)u^2) + D.$$

This completes the proof of the proposition.

Remark 5.2.1

Propositions 5.2.3 and 5.2.4 correspond to the fact that the primal function w in ordinary mathematical programming is monotonically non-increasing and convex.

It is well known that, in scalar convex optimization, the convexity of w ensures that by adding an appropriate linear functional $\langle \lambda, u \rangle$ to $w(u)$, the resulting combination $w(u) + \langle \lambda, u \rangle$ is minimized at $u = 0$ (see, for example, Luenberger [L14]). Geometrically speaking, this appropriate linear functional $\langle \lambda, u \rangle$ leads to a supporting hyperplane of epi w at $u = 0$. In an analogous manner, the D-convexity of the point-to-set map W ensures that, if an appropriate linear vector-valued functional Λu is added to $W(u)$, there will exist no point of $W(u) + \Lambda u$ that dominates a given point of $W(0)$. This linear vector-valued functional leads geometrically to a supporting conical variety (i.e., a translation of cone) of $W(0)$ at the given point. Such a geometric meaning will be discussed in more detail in the next subsection. We now have a Lagrange multiplier theorem for multiobjective optimization.

Theorem 5.2.1

If $\hat{x}$ is a properly efficient solution to problem (P), and if Slater's constraint qualification holds (i.e., there exists $x' \in X'$ such that $g(x') <_Q 0$), then there exists a $p \times m$ matrix $\hat{\Lambda}$ such that $\hat{\Lambda} Q \subset D$ and

$$f(\hat{x}) \in \operatorname{Min}_D\{f(x) + \hat{\Lambda} g(x) : x \in X'\},$$
$$\hat{\Lambda} g(\hat{x}) = 0.$$

Proof Let $X = \{x \in R^n : g(x) \leqq_Q 0\} \cap X'$. Since $\hat{x}$ is a properly efficient solution of $f(X)$ with respect to $\leq_D$, there exists a vector $\hat{\mu} \in \operatorname{int} D^\circ$ such that

$$\langle \hat{\mu}, f(\hat{x}) \rangle \leqq \langle \hat{\mu}, f(x) \rangle \qquad \text{for any} \quad x \in X.$$

Note here that $\langle \hat{\mu}, f(x) \rangle$ is a convex function on X'. In fact, due to the D-convexity of f, since

$$\alpha f(x^1) + (1-\alpha) f(x^2) - f(\alpha x^1 + (1-\alpha)x^2) \in D$$

for any $x^1, x^2 \in X'$ and any $\alpha \in [0, 1]$, we have

$$\alpha \langle \hat{\mu}, f(x^1) \rangle + (1-\alpha) \langle \hat{\mu}, f(x^2) \rangle - \langle \hat{\mu}, f(\alpha x^1 + (1-\alpha)x^2) \rangle \geqq 0.$$

Therefore, the well-known Lagrange multiplier theorem in scalar convex optimization leads to the existence of a vector $\hat{\lambda} \in Q^\circ$ such that

$$\langle \hat{\mu}, f(\hat{x}) \rangle + \langle \hat{\lambda}, g(\hat{x}) \rangle \leqq \langle \hat{\mu}, f(x) \rangle + \langle \hat{\lambda}, g(x) \rangle \tag{5.2.2}$$

for any $x \in X'$ and

$$\langle \hat{\lambda}, g(\hat{x}) \rangle = 0.$$

Now, for such $\hat{\mu}$ and $\hat{\lambda}$, take a $p \times m$ matrix $\hat{\Lambda}$ with $\hat{\Lambda}^{\mathrm{T}}\hat{\mu} = \hat{\lambda}$ in such a way that

$$\hat{\Lambda} = (\hat{\lambda}_1 e, \hat{\lambda}_2 e, \ldots, \hat{\lambda}_m e),$$

where e is a vector of D with $\langle \hat{\mu}, e \rangle = 1$. Then clearly $\hat{\Lambda}Q \subset D$ and $\hat{\Lambda}g(\hat{x}) = 0$. If we suppose that for this $\hat{\Lambda}$

$$f(\hat{x}) \notin \mathrm{Min}_D\{f(x) + \hat{\Lambda}g(x) : x \in X'\},$$

there exists $\bar{x} \in X'$ such that

$$f(\hat{x}) - f(\bar{x}) - \hat{\Lambda}g(\bar{x}) \in D\backslash\{0\}.$$

Hence,

$$\begin{aligned}\langle \hat{\mu}, f(\hat{x}) \rangle &> \langle \hat{\mu}, f(\bar{x}) \rangle + \langle \hat{\mu}, \hat{\Lambda}g(\bar{x}) \rangle \\ &= \langle \hat{\mu}, f(\bar{x}) \rangle + \langle \hat{\lambda}, g(\bar{x}) \rangle,\end{aligned}$$

which contradicts the relation (5.2.2). Therefore,

$$f(\hat{x}) \in \mathrm{Min}_D\{f(x) + \hat{\Lambda}g(x) : x \in X'\}.$$

Remark 5.2.2

In the case where $D \ni (1, \ldots, 1)^{\mathrm{T}}$, by normalizing $\hat{\mu}$ in such a particular way that $\sum \hat{\mu}_i = 1$, we can adopt $(1, 1, \ldots, 1)^{\mathrm{T}}$ as the vector e in the proof of Theorem 5.2.1. We then have

$$\Lambda g(x) = (\langle \lambda, g(x) \rangle, \ldots, \langle \lambda, g(x) \rangle)^{\mathrm{T}},$$

which was used in Tanino and Sawaragi [T9]. All results in this subsection are also valid for such a particular form of Λg.

5.2.2 Vector-Valued Lagrangian Function and Its Saddle Point

We have suggested a geometric expression of the Lagrange multiplier theorem for multiobjective optimization in the previous section. In this subsection, we will state an equivalent algebraic form of the theorem. Hereafter, we shall denote by $\mathscr{L}$ a family of all $p \times m$ matrices Λ such that

$$\Lambda Q \subset D.$$

Such matrices are said to be positive in some literature (Ritter [R6], Corley [C16]). Note that for given $\mu \in D^{\circ}\backslash\{0\}$ and $\lambda \in Q^{\circ}$, there exist $\Lambda \in \mathscr{L}$ such that

$$\Lambda^{\mathrm{T}}\mu = \lambda.$$

In fact, as in the proof of Theorem 5.2.1, for some vector e of D with $\langle \mu, e \rangle = 1$,

$$\Lambda = (\lambda_1 e, \lambda_2 e, \ldots, \lambda_m e)$$

is a desired one. This fact will be often used later.

Definition 5.2.2 (*Vector-Valued Lagrangian Function*)

A vector-valued Lagrangian function for problem (P) is defined on $X' \times \mathscr{L}$ by

$$L(x, \Lambda) = f(x) + \Lambda g(x).$$

Definition 5.2.3 (*Saddle Point for Vector-Valued Lagrangian Function*)

A pair $(\hat{x}, \hat{\Lambda}) \in X' \times \mathscr{L}$ is said to be a saddle point for the vector-valued Lagrangian function $L(x, \Lambda)$, if the following holds:

$$L(\hat{x}, \hat{\Lambda}) \in \mathrm{Min}_D\{L(x, \hat{\Lambda}) : x \in X'\} \cap \mathrm{Max}_D\{L(\hat{x}, \Lambda) : \Lambda \in \mathscr{L}\}$$

Theorem 5.2.2

The following three conditions are necessary and sufficient for a pair $(\hat{x}, \hat{\Lambda}) \in X' \times \mathscr{L}$ to be a saddle point for the vector-valued Lagrangian function $L(x, \Lambda)$:

(i) $L(\hat{x}, \hat{\Lambda}) \in \mathrm{Min}_D\{L(x, \hat{\Lambda}) : x \in X'\}$,
(ii) $g(\hat{x}) \leqq_Q 0$,
(iii) $\hat{\Lambda} g(\hat{x}) = 0$.

Proof

necessity Condition (i) is the same as a part of the definition of saddle point for $L(x, \Lambda)$. As for condition (ii), note that $L(\hat{x}, \hat{\Lambda}) \in \mathrm{Max}_D\{f(\hat{x}) + \Lambda g(\hat{x}) : \Lambda \in \mathscr{L}\}$ implies that

$$f(\hat{x}) + \hat{\Lambda} g(\hat{x}) \not\leq_D f(\hat{x}) + \Lambda g(\hat{x}) \qquad \text{for any} \quad \Lambda \in \mathscr{L}, \tag{5.2.3}$$

from which we have

$$\langle \hat{\mu}, \Lambda g(\hat{x}) - \hat{\Lambda} g(\hat{x}) \rangle \leqq 0 \tag{5.2.4}$$

for some $\hat{\mu} \in D^{\circ} \backslash \{0\}$ and for any $\Lambda \in \mathscr{L}$. Suppose that $g(\hat{x}) \nleqq_Q 0$. Then there is $\hat{\lambda} \in Q^{\circ}$ such that $\langle \hat{\lambda}, g(\hat{x}) \rangle > 0$. Making $\|\hat{\lambda}\|$ sufficiently large and taking $\Lambda \in \mathscr{L}$ such that $\hat{\mu}^{\mathrm{T}} \Lambda = \hat{\lambda}^{\mathrm{T}}$, we obtain the relation

$$\langle \hat{\mu}, \Lambda g(\hat{x}) \rangle - \langle \hat{\mu}, \hat{\Lambda} g(\hat{x}) \rangle > 0,$$

which contradicts relation (5.2.4). Thus, $g(\hat{x}) \leqq_Q 0$.

Using this result, $\hat{\Lambda}g(\hat{x}) \leqq_D 0$ for $\hat{\Lambda} \in \mathscr{L}$. On the other hand, substituting $\Lambda = 0$ into Eq. (5.2.3) yields $\hat{\Lambda}g(\hat{x}) \nleq_D 0$. Finally, we have $\hat{\Lambda}g(\hat{x}) = 0$.

sufficiency Since $\Lambda g(\hat{x}) \in -D$ for any $\Lambda \in \mathscr{L}$ as long as $g(\hat{x}) \leqq_Q 0$, it follows that

$$\mathrm{Max}_D\{\Lambda g(\hat{x}) : \Lambda \in \mathscr{L}\} = \{0\}.$$

Thus, from $\hat{\Lambda}g(\hat{x}) = 0$ we have $L(\hat{x}, \hat{\Lambda}) \in \mathrm{Max}_D\{f(\hat{x}) + \Lambda g(\hat{x}) : \Lambda \in \mathscr{L}\}$. This result and condition (i) imply that the pair $(\hat{x}, \hat{\Lambda})$ is a saddle point for $L(x, \Lambda)$.

Corollary 5.2.1

Suppose that $\hat{x}$ is a properly efficient solution to problem (P) and that Slater's constraint qualification is satisfied. Then, there exists a $p \times m$ matrix $\hat{\Lambda} \in \mathscr{L}$ such that $(\hat{x}, \hat{\Lambda})$ is a saddle point for the vector-valued Lagrangian function $L(x, \Lambda)$.

Proof Immediate from Theorems 5.2.1 and 5.2.2.

Thus, we have verified that a properly efficient solution to the problem (P) together with a matrix give a saddle point for the vector-valued Lagrangian function under the convexity assumptions and the appropriate regularity conditions. Conversely, the saddle point provides a sufficient condition for optimality for the problem (P).

Theorem 5.2.3

If $(\hat{x}, \hat{\Lambda}) \in X' \times \mathscr{L}$ is a saddle point for the vector-valued Lagrangian function $L(x, \Lambda)$, then $\hat{x}$ is an efficient solution to problem (P).

Proof Suppose that $\hat{x}$ is not a solution to problem (P); namely, that there exists $\bar{x} \in X$ such that $f(\bar{x}) \leq_D f(\hat{x})$. Since $g(\bar{x}) \leqq_Q 0$ and $\hat{\Lambda} \in \mathscr{L}$ yield $\hat{\Lambda}g(\bar{x}) \in -D$, we finally have $f(\bar{x}) + \hat{\Lambda}g(\bar{x}) \leq_D f(\hat{x})$, which contradicts $L(\hat{x}, \hat{\Lambda}) \in \mathrm{Min}_D\{L(x, \hat{\Lambda}) : x \in X'\}$.

5.2.3 Dual Map and Duality Theory

Recall that the dual function in ordinary convex optimization is defined by

$$\phi(\lambda) = \inf\{f(x) + \langle \lambda, g(x) \rangle : x \in X'\}.$$

To begin with, we define a point-to-set map corresponding to the dual function.

Definition 5.2.4 (*Dual map*)

Define for any $\Lambda \in \mathscr{L}$,

$$\Omega(\Lambda) = \{L(x, \lambda) : x \in X'\} = \{f(x) + \Lambda g(x) : x \in X'\}$$

and

$$\Phi(\Lambda) = \mathrm{Min}_D \Omega(\Lambda).$$

The point-to-set map $\Phi: \mathscr{L} \to R^p$ is called a dual map.

Under the terminology, we can define a dual problem associated with the primal problem (P) as follows (Tanino and Sawaragi [T9]):

$$(\mathrm{D_{TS}}) \qquad D\text{-maximize} \bigcup_{\Lambda \in \mathscr{L}} \Phi(\Lambda).$$

In the following, we summarize several properties of dual maps.

Proposition 5.2.5

For each $\Lambda \in \mathscr{L}$, $\Phi(\Lambda)$ is a D-convex set in R^p.

Proof Since the maps f and g are D-convex and Q-convex, respectively, the map $L(\cdot, \Lambda)$ is D-convex over X' for each fixed $\Lambda \in \mathscr{L}$. Hence $\Omega(\Lambda)$ is a compact D-convex set in R^p, for X' is a compact convex set. Therefore,

$$\Phi(\Lambda) + D = \Omega(\Lambda) + D$$

by Theorem 3.2.9, and thus $\Phi(\Lambda)$ is also D-convex.

Proposition 5.2.6

Φ is a D-concave point-to-set map on Γ. Namely, for any Λ^1, $\Lambda^2 \in \mathscr{L}$ and any $\alpha \in [0, 1]$

$$\Phi(\alpha\Lambda^1 + (1-\alpha)\Lambda^2) \subset \alpha\Phi(\Lambda^1) + (1-\alpha)\Phi(\Lambda^2) + D.$$

Proof Note that Theorem 3.2.9 yields

$$\mathrm{Min}_D A \subset A \subset B \subset \mathrm{Min}_D B + D$$

for any two sets A and B such that $A \subset B$, and B is compact. From this and the relation

$$\{\alpha(f(x) + \Lambda^1 g(x)) + (1-\alpha)(f(x) + \Lambda^2 g(x)) : x \in X'\}$$
$$\subset \alpha\{f(x) + \Lambda^1 g(x) : x \in X'\} + (1-\alpha)\{f(x) + \Lambda^2 g(x) : x \in X'\},$$

we have

$$\begin{aligned}
&\Phi(\alpha\Lambda^1 + (1-\alpha)\Lambda^2) \\
&\quad = \mathrm{Min}_D\{f(x) + (\alpha\Lambda^1 + (1-\alpha)\Lambda^2)g(x) : x \in X'\} \\
&\quad = \mathrm{Min}_D\{\alpha(f(x) + \Lambda^1 g(x)) + (1-\alpha)(f(x) + \Lambda^2 g(x)) : x \in X'\} \\
&\quad \subset \mathrm{Min}_D[\alpha\{f(x) + \Lambda^1 g(x) : x \in X'\} \\
&\qquad + (1-\alpha)\{f(x) + \Lambda^2 g(x) : x \in X'\}] + D.
\end{aligned}$$

Hence, in view of Proposition 3.1.3,

$$\begin{aligned}
\Phi(\alpha\Lambda^1 + (1-\alpha)\Lambda^2) &\subset \mathrm{Min}_D\, \alpha\{f(x) + \Lambda^1 g(x) : x \in X'\} \\
&\quad + \mathrm{Min}_D(1-\alpha)\{f(x) + \Lambda^2 g(x) : x \in X'\} + D \\
&= \alpha\, \mathrm{Min}_D\{f(x) + \Lambda^1 g(x) : x \in X'\} \\
&\quad + (1-\alpha)\, \mathrm{Min}_D\{f(x) + \Lambda^2 g(x) : x \in X'\} + D \\
&= \alpha\Phi(\Lambda^1) + (1-\alpha)\Phi(\Lambda^2) + D.
\end{aligned}$$

Remark 5.2.3

Proposition 5.2.6 is an extension of the fact that the dual function $\phi(\lambda)$ is concave.

We can now establish the following relationship between the dual map $\Phi(\Lambda)$ and the primal map $W(u)$, which is an extension of the following relation between the dual function $\phi(\lambda)$ and the primal function $w(u)$:

$$\phi(\lambda) = \inf\{w(u) + \langle \lambda, u \rangle : u \in \Gamma\}.$$

Proposition 5.2.7

The following relation holds:

$$\Phi(\Lambda) = \mathrm{Min}_D \bigcup_{u \in \Gamma} (W(u) + \Lambda u).$$

Proof Let

$$y^1 = f(x^1) + \Lambda g(x^1)$$

for any $x^1 \in X'$. Then, letting $u^1 = g(x^1)$,

$$y^1 = f(x^1) + \Lambda u^1.$$

Note here that

$$f(x^1) \in W(u^1) + D,$$

because $f(x^1) \in Y(u^1)$. Hence

$$y^1 \in W(u^1) + \Lambda u^1 + D \subset \bigcup_{u \in \Gamma} (W(u) + \Lambda u) + D,$$

from which

$$\Omega(\Lambda) \subset \bigcup_{u \in \Gamma} (W(u) + \Lambda u) + D,$$

$$\Omega(\Lambda) + D \subset \bigcup_{u \in \Gamma} (W(u) + \Lambda u) + D.$$

Conversely, suppose that

$$y^1 \in W(u^1) + \Lambda u^1$$

for some $u^1 \in \Gamma$, namely that

$$y^1 - \Lambda u^1 \in \mathrm{Min}_D \mathrm{Y}(\mathrm{u}^1).$$

Thus,

$$y^1 - \Lambda u^1 = f(x^1)$$

for some $x^1 \in X'$ such that

$$g(x^1) \leqq_Q u^1.$$

Then, for $\Lambda \in \mathscr{L}$,

$$y^1 = f(x^1) + \Lambda u^1 \geqq_D f(x^1) + \Lambda g(x^1),$$

and hence

$$y^1 \in L(x^1, \Lambda) + D \subset \Omega(\Lambda) + D.$$

Therefore,

$$\bigcup_{u \in \Gamma} (W(u) + \Lambda u) \subset \Omega(\Lambda) + D,$$

and thus

$$\bigcup_{u \in \Gamma} (W(u) + \Lambda u) + D \subset \Omega(\Lambda) + D.$$

Finally, we have

$$\Omega(\Lambda) + D = \bigcup_{u \in \Gamma} (W(u) + \Lambda u) + D,$$

and hence

$$\mathrm{Min}_D(\Omega(\Lambda) + D) = \mathrm{Min}_D\left(\bigcup_{u \in \Gamma} (W(u) + \Lambda U) + D\right),$$

which establishes the proposition, because in general $\mathrm{Min}_D A = \mathrm{Min}_D(A + D)$ whenever D is a pointed convex cone (cf. Proposition 3.1.2).

Remark 5.2.4

Recall that Theorem 5.2.1 implies that, given a properly efficient solution $\hat{x}$, then $f(\hat{x}) \in \Phi(\hat{\Lambda})$ for some $\hat{\Lambda} \in \mathscr{L}$. Hence, we may see from Proposition 5.2.7 that

$$f(\hat{x}) \in \mathrm{Min}_D \bigcup_{u \in \Gamma} (W(u) + \hat{\Lambda} u)$$

for some $\hat{\Lambda} \in \mathscr{L}$. This is an extension of a property of the primal function w in ordinary convex optimization stated in Subsection 5.2.1.

The following theorems represent some properties of efficient solutions to primal problem (P) in connection with the dual map Φ and so might be considered as duality theorems for multiobjective optimization.

Theorem 5.2.4 (*Weak Duality*)

For any $x \in X$ and $y \in \Phi(\Lambda)$

$$y \not\geq_D f(x).$$

Proof For any y with the property

$$y \not\geq_D f(x) + \Lambda g(x) \qquad \text{for all} \quad x \in X',$$

the result of the theorem follows immediately from Lemma 2.3.3, because for any $x \in X$ and $\Lambda \in \mathscr{L}$,

$$\Lambda g(x) \leqq_D 0.$$

Theorem 5.2.5

(i) Suppose that $\hat{x} \in X$, $\hat{\Lambda} \in \mathscr{L}$ and $f(\hat{x}) \in \Phi(\hat{\Lambda})$. Then $\hat{y} = f(\hat{x})$ is an efficient point to the primal problem (P) and also to the dual problem ($\mathrm{D_{TS}}$).

(ii) Suppose that $\hat{x}$ is a properly efficient solution to problem (P) and that Slater's constraint qualification is satisfied. Then

$$f(\hat{x}) \in \mathrm{Max}_D \bigcup_{\Lambda \in \mathscr{L}} \Phi(\Lambda).$$

Proof

(i) If $f(\hat{x})$ is not an efficient point to problem (P), then there exists $\bar{x} \in X$ such that $f(\bar{x}) \leq_D f(\hat{x})$. Since $g(\bar{x}) \leqq_Q 0$ and $\hat{\Lambda} \in \mathscr{L}$ yield $\hat{\Lambda} g(\bar{x}) \in -D$, we finally have $f(\bar{x}) + \hat{\Lambda} g(\bar{x}) \leq_D f(\hat{x})$, which contradicts $f(\hat{x}) \in \mathrm{Min}_D\{f(x) + \hat{\Lambda} g(x) : x \in X'\}$. Hence $f(\hat{x})$ is an efficient point to problem (P). Furthermore, suppose that $f(\hat{x}) \leq_D \bar{y}$ for some $\bar{y} \in \bigcup_{\Lambda \in \mathscr{L}} \Phi(\Lambda)$. Let $\bar{\Lambda} \in \mathscr{L}$ be such that $\bar{y} \in \Phi(\bar{\Lambda})$. Then, since $\bar{\Lambda} g(\hat{x}) \leqq_D 0$, we have

$f(\hat{x}) + \bar{\Lambda}g(\hat{x}) \leq_D \bar{y}$, which contradicts $\bar{y} \in \text{Min}_D\{f(x) + \bar{\Lambda}g(x) : x \in X'\}$. Therefore, $f(\hat{x})$ is also a solution to dual problem (D).

(ii) Theorem 5.2.1 ensures that there exists a $p \times m$ matrix $\hat{\Lambda} \in \mathscr{L}$ such that $f(\hat{x}) \in \Phi(\hat{\Lambda})$, which leads to the conclusion via result (i).

5.3 Geometric Consideration of Duality

5.3.1 Geometric Meaning of Vector-Valued Lagrangian Functions

In Theorem 5.2.1, we established that for a properly efficient solution $\hat{x}$ to problem (P) there exists a $p \times m$ matrix $\hat{\Lambda}$ such that $\Lambda Q \subset D$ and

$$f(x) + \hat{\Lambda}g(x) \not\leq_D f(\hat{x}) \qquad \text{for all} \quad x \in X'$$

under an appropriate regularity condition such as Slater's constraint qualification. The theorem in ordinary scalar convex optimization that corresponds to this theorem asserts the existence of a supporting hyperplane for epi w at $(0, f(\hat{x}))$. On the other hand, in multiobjective optimization, for the D-epigraph of $W(u)$ defined by

$$D\text{-epi}\, W := \{(u, y) : u \in \Gamma,\ y \in R^p,\ y \in W(u) + D\},$$

Theorem 5.2.1 geometrically implies the existence of a supporting conical variety (i.e., a translation of the cone) supporting D-epi W at $(0, f(\hat{x}))$. We shall discuss this in more detail. All assumptions on f, g, D, and Q of the previous section are inherited in this section.

As is readily shown, D-epi W is a closed convex set in $R^m \times R^p$ under our assumptions. Hereafter we shall assume that the pointed closed convex cone D is polyhedral. Then, as is well known, there exists a matrix M_1 such that

$$D = \{y \in R^p : M_1 y \geqq 0\}.$$

Since D is pointed, M_1 has full rank p. We now have the following lemma.

Lemma 5.3.1

For a given $\Lambda \in \mathscr{L}$, the following two conditions are equivalent to each other:

(i) $f(x) + \Lambda g(x) \not\leq_D f(\hat{x}) + \Lambda g(\hat{x}) \qquad \text{for all} \quad x \in X'$, (5.3.1)

(ii) $M_1 f(x) + M_1 \Lambda g(x) \not\leq M_1 f(\hat{x}) + M_1 \Lambda g(\hat{x})$ for all $x \in X'$.

Proof Immediate from Lemma 2.3.4.

Let $M_2 = M_1 \Lambda$ and define a cone K in $R^m \times R^p$ by

$$K := \{(u, y) : M_1 y + M_2 u \leqq 0,\ u \in R^m,\ y \in R^p\}. \tag{5.3.2}$$

For this cone K, the following properties hold. Let the y-intercept of K in $R^m \times R^p$ be denoted by Y_K, namely

$$Y_K := \{y : (0, y) \in K, 0 \in R^m, y \in R^p\};$$

then from the definition of M_1, we have $Y_K = -D$. Furthermore, let $l(K)$ denote the lineality space of K, i.e., $-K \cap K$. Then, since $l(K)$ accords to $\mathrm{Ker}(M_1, M_2)$, it is m-dimensional. For

$$\begin{aligned} l(K) &= \{(u, y) \in R^m \times R^p : M_1 y + M_2 u = 0\} \\ &= \{(u, y) \in R^m \times R^p : y + \Lambda u = 0\}, \end{aligned}$$

because the $s \times p$ matrix M_1 has the maximal rank p. In addition, since the row vectors of M_1 are generators of D° and $\Lambda Q \subset D$, every row vector of M_2 is in Q°.

We now can establish with the help of Lemma 5.3.1 a geometric interpretation of Theorem 5.2.1. We say the cone K supports D-epi W at $(\hat{u}, \hat{y})$, if $K \cap ((D\text{-epi } W) - (\hat{u}, \hat{y})) \subset l(K)$. Clearly, K supports D-epi W at $(\hat{u}, \hat{y})$ if and only if

$$M_1(y - \hat{y}) + M_2(u - \hat{u}) \nleq 0 \qquad \text{for all} \quad (u, y) \in D\text{-epi } W,$$

or equivalently

$$y + \Lambda u \nleq_D \hat{y} + \Lambda \hat{u} \qquad \text{for all} \quad (u, y) \in D\text{-epi } W.$$

Therefore, we can conclude that Theorem 5.2.1 asserts the existence of supporting conical varieties (a translation of K) for D-epi W at $(g(\hat{x}), f(\hat{x}))$ (see Fig. 5.1).

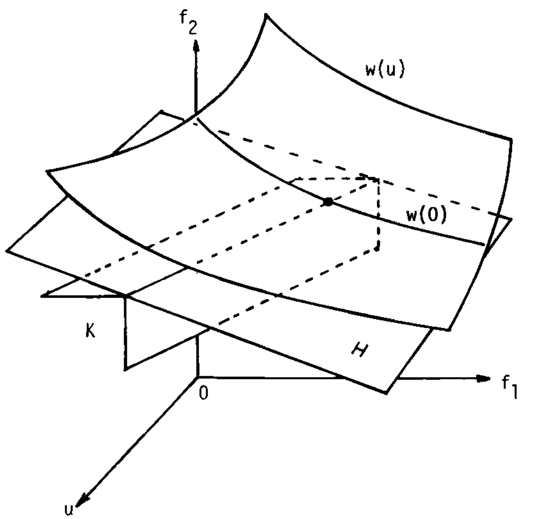

Fig. 5.1. Geometric interpretation of vector-valued Lagrangian.

Remark 5.3.1

Observe that in ordinary scalar optimization, the supporting conical varieties becomes half-spaces, because the lineality is identical to the dimension of g. This accords with the well-known result.

We now investigate a relationship between supporting hyperplanes and supporting conical varieties. Let $H(\lambda, \mu : \gamma)$ be a hyperplane in $R^m \times R^p$ with the normal (λ, μ) such that

$$H(\lambda, \mu : \gamma) := \{(u, y) : \langle \mu, y \rangle + \langle \lambda, u \rangle - \gamma = 0, u \in R^m, y \in R^p\}.$$

Define

$$h(\lambda, \mu : \gamma) := \langle \mu, y \rangle + \langle \lambda, u \rangle - \gamma.$$

Then, associated with the hyperplane $H(\lambda, \mu : \gamma)$, several kinds of half-spaces are defined as follows:

$$H^+(\lambda, \mu : \gamma) := \{(u, y) \in R^m \times R^p : h(\lambda, \mu : \gamma) \geqq 0\},$$

$$\mathring{H}^+(\lambda, \mu : \gamma) := \{(u, y) \in R^m \times R^p : h(\lambda, \mu : \gamma) > 0\}.$$

H^- and $\mathring{H}^-$ are similarly defined by replacing the inequality $>$ (resp., $=$) with $<$ (resp., $=$). In particular, let $H(\lambda, \mu)$ denote the supporting hyperplane for D-epi W with the inner normal (λ, μ); that is,

$$H(\lambda, \mu) := H(\lambda, \mu : \hat{\gamma}) \qquad \text{where} \quad \hat{\gamma} = \sup\{\gamma : H^+(\lambda, \mu : \gamma) \supset D\text{-epi}\, W\}. \tag{5.3.3}$$

Lemma 5.3.2

The lineality space of the cone K given by (5.3.2) with $M_2 = M_1\Lambda$ is included in the hyperplane $H(\lambda, \mu : 0)$ if and only if the matrix Λ satisfies $\Lambda^{\mathrm{T}}\mu = \lambda$.

Proof Since the $s \times p$ matrix M_1 has the maximal rank p, $M_1(y + \Lambda u) = 0$ is equivalent to $y + \Lambda u = 0$. Hence, for $M_2 = M_1\Lambda$

$$\begin{aligned}
\langle \mu, y \rangle + \langle \lambda, u \rangle = 0 \qquad & \text{for any} \quad (u, y) \text{ such that } M_1 y + M_2 u = 0 \\
& \Leftrightarrow \langle \mu, y \rangle + \langle \lambda, u \rangle = 0 \qquad \text{for any} \quad (u, y) \\
& \quad\ \text{such that } y + \Lambda u = 0 \\
& \Leftrightarrow \text{there exists a } p\text{-dimensional vector } \xi \\
& \quad\ \text{such that } (\mu^{\mathrm{T}}, \lambda^{\mathrm{T}}) = \xi^{\mathrm{T}}(I, \Lambda),
\end{aligned}$$

where I denotes the $p \times p$ identity matrix, and in the last equivalence the well-known Minkowski–Farkas lemma is used. This completes the proof.

Lemma 5.3.3

For any supporting hyperplane $H(\lambda, \mu)$ for D-epi W, we have $\lambda \in Q^\circ$ and $\mu \in D^\circ$.

Proof Let $(\hat{u}, \hat{y}) \in D$-epi W be the supporting point of D-epi W by $H(\lambda, \mu)$. Then, $\mu \in D^{\circ}$ follows because $(\hat{u}, \hat{y} + d) \in W(\hat{u}) + D \subset D$-epi W for any $d \in D$. Similarly, $\lambda \in Q^{\circ}$ follows from the monotonicity of $W(u)$; that is $(\hat{u} + q, \hat{y}) \in D$-epi W for any $q \in Q$.

The following lemma clarifies a relation between supporting conical varieties and supporting hyperplanes.

Lemma 5.3.4

Let $H(\lambda, \mu)$ be a supporting hyperplane for D-epi W with a supporting point $(\hat{u}, \hat{y})$. If $\mu \in \operatorname{int} D^{\circ}$, then any conical variety of the cone K given by relation (5.3.2) whose lineality variety passing through $(\hat{u}, \hat{y})$ is included in $H(\lambda, \mu)$ supports D-epi W at $(\hat{u}, \hat{y})$.

Conversely, if some conical variety of K supports D-epi W at $(\hat{u}, \hat{y})$, then there exists a hyperplane $H(\lambda, \mu : \gamma)$ with $\mu \neq 0$ supporting D-epi W at $(\hat{u}, \hat{y})$ that contains the lineality variety of the supporting conical variety.

Proof Let $\mu \in \operatorname{int} D^{\circ}$. If, to the contrary, a conical variety of K given by relation (5.3.2), whose lineality variety passing through $(\hat{u}, \hat{y})$ is included in $H(\lambda, \mu)$, does not support D-epi W at $(\hat{u}, \hat{y})$, then there exists a point $(\bar{u}, \bar{y}) \in D$-epi W other than $(\hat{u}, \hat{y})$ such that

$$M_1(\bar{y} - \hat{y}) + M_2(\bar{u} - \hat{u}) \leq 0,$$

where $M_2 = M_1 \Lambda$ and $\Lambda^{\mathrm{T}} \mu = \lambda$. Hence from the definition of D we have

$$\bar{y} - \hat{y} + \Lambda(\bar{u} - \hat{u}) \leq_D 0,$$

which implies

$$\langle \mu, \bar{y} - \hat{y} + \Lambda(\bar{u} - \hat{u}) \rangle < 0,$$

as long as $\mu \in \operatorname{int} D^{\circ}$. This contradicts the fact that the hyperplane $H(\lambda, \mu)$ supports D-epi W at $(\hat{u}, \hat{y})$.

Conversely, suppose that a conical variety of K supports D-epi W at $(\hat{u}, \hat{y})$; that is,

$$M_1(y - \hat{y}) + M_2(u - \hat{u}) \nleq 0 \qquad \text{for all} \quad (u, y) \in D\text{-epi } W.$$

Clearly the set $\{M_1 y + M_2 u : (u, y) \in D\text{-epi } W\}$ is convex. Hence, there exists an s-dimensional vector $\alpha \geq 0$ such that

$$\langle \alpha, M_1(y - \hat{y}) \rangle + \langle \alpha, M_2(u - \hat{u}) \rangle \geqq 0 \qquad \text{for all} \quad (u, y) \in D\text{-epi } W,$$

from which we have a supporting hyperplane with $\mu = M_1^{\mathrm{T}} \alpha$, $\lambda = M_2^{\mathrm{T}} \alpha$ that satisfies

$$\langle \mu, y - \hat{y} \rangle + \langle \lambda, u - \hat{u} \rangle \geqq 0 \qquad \text{for all} \quad (u, y) \in D\text{-epi } W.$$

Since D° is pointed due to int $D \neq \varnothing$ and the row vectors of M_1 are generators of D°, the relation $\mu = M_1^{\mathrm{T}}\alpha$ and $\lambda = M_2^{\mathrm{T}}\alpha$ for $\alpha \geq 0$ leads to the fact that $\mu \neq 0$ and

$$\langle \mu, y - \hat{y} \rangle + \langle \lambda, u - \hat{u} \rangle = 0$$
$$\text{for any} \quad (u, y) \text{ such that } M_1(y - \hat{y}) + M_2(u - \hat{u}) = 0,$$

which implies the lineality variety passing through $(\hat{u}, \hat{y})$ of the supporting conical variety under consideration is included by the hyperplane $H(\lambda, \mu)$ supporting D-epi W at $(\hat{u}, \hat{y})$.

We now summarize several properties for the support property of conical varieties, saddle points of the vector-valued Lagrangian function, duality, and unconstrained D-minimization of the vector-valued Lagrangian function as the following mutual equivalence theorem.

Theorem 5.3.1

For $(\hat{x}, \hat{\Lambda}) \in X' \times \mathscr{L}$ the following four conditions are equivalent to one another:

(i) Let M_1 be an $s \times p$ matrix with full rank whose row vectors generate D°, and let M_2 be an $s \times m$ matrix such that $M_1\hat{\Lambda} = M_2$ for $\hat{\Lambda} \in \mathscr{L}$. Then $\hat{x}$ solves primal problem (P), and conical varieties of the cone given by

$$K = \{(u, y): M_1 y + M_2 u \leqq 0,\ y \in R^p,\ u \in R^m\}$$

support D-epi W at $(0, f(\hat{x}))$.

(ii) $L(\hat{x}, \hat{\Lambda}) \in \mathrm{Min}_D\{L(x, \hat{\Lambda}) : x \in X'\} \cap \mathrm{Max}_D\{L(\hat{x}, \Lambda) : \Lambda \in \mathscr{L}\}$.

(iii) $\hat{x} \in X$, and $f(\hat{x}) \in \Phi(\hat{\Lambda})$.

(iv) $L(\hat{x}, \hat{\Lambda}) \in \mathrm{Min}_D\{L(x, \hat{\Lambda}) : x \in X'\}$, $g(\hat{x}) \leqq_Q 0$, and $\hat{\Lambda} g(\hat{x}) = 0$.

Remark 5.3.2

Note that, as in Theorem 5.2.5, condition (iii) means that $f(\hat{x})$ is an efficient point to the primal problem (P) and also to the dual problem (D).

Proof of Theorem 5.3.1 We have already seen (ii) $\Leftrightarrow$ (iv) (Theorem 5.2.2) and (ii) or (iv) $\Rightarrow$ (i) (Theorem 5.2.3 and Lemma 5.3.1), and, hence, we prove the remaining parts.

(iv) $\Rightarrow$ (iii) Since the equivalence of (ii) and (iv) yields $\hat{\Lambda} g(\hat{x}) = 0$, we have

$$f(\hat{x}) = f(\hat{x}) + \hat{\Lambda} g(\hat{x}) = L(\hat{x}, \hat{\Lambda}) \in \mathrm{Min}_D\{f(x) + \hat{\Lambda} g(\hat{x}) : x \in X'\} = \Phi(\hat{\Lambda}).$$

$\hat{x} \in X$ is immediate.

(iii) ⇒ (iv) Note that for any $y \in \Phi(\hat{\Lambda})$ and any $x \in X'$,

$$y \not\geq_D f(x) + \hat{\Lambda}g(x). \tag{5.3.4}$$

Since $f(\hat{x}) \in \Phi(\hat{\Lambda})$, putting $x = \hat{x}$ into relation (5.3.4) with $y = f(\hat{x})$ yields $\hat{\Lambda}g(\hat{x}) \not\leq_D 0$. On the other hand, $\hat{\Lambda}g(\hat{x}) \leqq_D 0$ for $\hat{\Lambda} \in \mathscr{L}$, because $g(\hat{x}) \leqq_Q 0$. Therefore, $\hat{\Lambda}g(\hat{x}) = 0$ and $L(\hat{x}, \hat{\Lambda}) = f(\hat{x}) \in \Phi(\hat{\Lambda}) := \mathrm{Min}_D\{L(x, \hat{\Lambda}) : x \in X'\}$.

(i) ⇒ (iv) Under our assumption of M_1, M_2, and $\hat{\Lambda}$, if conical varieties of the cone K support D-epi W at $(0, f(\hat{x}))$, then for any $(u, y) \in D$-epi W

$$M_1(y - f(\hat{x})) + M_2(u - 0) \not\leq 0,$$

or equivalently due to Lemma 5.3.1,

$$y - f(\hat{x}) + \hat{\Lambda}u \not\leq_D 0.$$

Since $(g(\hat{x}), f(\hat{x})) \in D$-epi W, we have $\hat{\Lambda}g(\hat{x}) \not\leq_D 0$. By a similar argument in in the proof of (iii) ⇒ (iv), we can conclude $\hat{\Lambda}g(\hat{x}) = 0$. Therefore, since $(g(x), f(x)) \in D$-epi W for any $x \in X'$, we have

$$f(\hat{x}) + \hat{\Lambda}g(\hat{x}) \in \mathrm{Min}_D\{f(x) + \hat{\Lambda}g(x) : x \in X'\}.$$

5.3.2 Another Geometric Approach to Duality

Now we shall take another approach to duality in multiobjective optimization. As in the previous section, the convexity assumption on f and g will be also imposed here, but X' is not necessarily compact.

Define

$$G := \{(u, y) \in R^m \times R^p : y \geqq_D f(x),\ u \geqq_Q g(x), \text{ for some } x \in X'\},$$
$$Y_G := \{y : (0, y) \in G,\ 0 \in R^m,\ y \in R^p\}.$$

We restate the primal problem as follows:

(P) $\qquad D$-minimize $\quad \{f(x) : x \in X\}$,

where

$$X := \{x \in X' : g(x) \leqq_Q 0,\ X' \subset R^n\}.$$

Associated with this primal problem, we shall consider the following two kinds of dual problems:

($\mathrm{D_N}$) $\qquad D$-maximize $\quad \bigcup_{\Lambda \in \mathscr{L}} Y_{S(\Lambda)}$, where

$$Y_{S(\Lambda)} := \{y \in R^p : f(x) + \Lambda g(x) \not\leq_D y, \text{ for all } x \in X'\}$$

and

$$(\mathrm{D_J}) \qquad D\text{-maximize} \quad \bigcup_{\substack{\mu \in \operatorname{int} D^\circ \\ \lambda \in Q^\circ}} Y_{H^-(\lambda,\mu)},$$

where

$$Y_{H^-(\lambda,\mu)} := \{y \in R^p : \langle \mu, f(x)\rangle + \langle \lambda, g(x)\rangle \geqq \langle \mu, y\rangle \text{ for all } x \in X'\}.$$

The following results assert weak duality theorems for these problems.

Theorem 5.3.2

(i) For any $y \in \bigcup_{\Lambda \in \mathscr{L}} Y_{S(\Lambda)}$ and for any $x \in X$,

$$y \not\geq_D f(x).$$

(ii) For any $y \in \bigcup_{\mu \in \operatorname{int} D^\circ,\, \lambda \in D^\circ} Y_{H^-(\lambda,\mu)}$ and for any $x \in X$,

$$y \not\geq_D f(x).$$

Proof Part (i) follows in a similar fashion to Theorem 5.2.4. To show part (ii), suppose that for some $\mu \in \operatorname{int} D^\circ$, some $\lambda \in Q^\circ$, and all $x \in X'$

$$\langle \mu, f(x)\rangle + \langle \lambda, g(x)\rangle \geqq \langle \mu, y\rangle.$$

Thus, for any $x \in X$

$$\langle \mu, f(x)\rangle \geqq \langle \mu, y\rangle.$$

This implies that for any $x \in X$,

$$f(x) \not\leq_D y.$$

This completes the proof of the theorem.

The following lemma is essential for deriving strong duality.

Lemma 5.3.5

For a given (λ, μ), let

$$H^+(\lambda, \mu : \hat{\gamma}) := \{(\bar{u}, \bar{y}) \in R^m \times R^p : h(\lambda, \mu : \hat{\gamma}) := \langle \mu, \bar{y}\rangle + \langle \lambda, \bar{u}\rangle - \hat{\gamma} \geqq 0\},$$

where

$$\hat{\gamma} = \inf\{\langle \mu, y\rangle + \langle \lambda, u\rangle : (u, y) \in G\}.$$

Suppose that G is a nonempty closed convex subset in $R^m \times R^p$ and that there exists at least a halfspace $H^+(\lambda^0, \mu^0 : \hat{\gamma}^0)$ with $\mu^0 \in \operatorname{int} D^0$, $\lambda^0 \in Q^0$ and $\hat{\gamma}^0 > -\infty$, which contains G. Then

$$G = \bigcap_{\substack{\mu \in \operatorname{int} D^\circ \\ \lambda \in Q^\circ}} H^+(\lambda, \mu : \hat{\gamma}).$$

Proof Since G is a closed convex set, it is well known that G is also represented by the intersection of the closed halfspaces including it. Therefore, recalling that normals (λ, μ) of the supporting hyperplanes for G satisfy $\lambda \in Q^\circ$ and $\mu \in D^\circ$ (Lemma 5.3.3), we have

$$G = \bigcap_{\mu \in D^\circ,\, \lambda \in Q^\circ} H^+(\lambda, \mu : \hat{\gamma}).$$

Hence, we shall prove

$$\bigcap_{\mu \in D^\circ,\, \lambda \in Q^\circ} H^+(\lambda, \mu : \hat{\gamma}) = \bigcap_{\substack{\mu \in \operatorname{int} D^\circ \\ \lambda \in Q^\circ}} H^+(\lambda, \mu : \hat{\gamma}).$$

It is obvious that

$$\bigcap_{\mu \in D^\circ,\, \lambda \in Q^\circ} H^+(\lambda, \mu : \hat{\gamma}) \subset \bigcap_{\substack{\mu \in \operatorname{int} D^\circ \\ \lambda \in Q^\circ}} H^+(\lambda, \mu : \hat{\gamma}).$$

To prove the reverse inclusion, it suffices to show that for a hyperplane $H(\lambda^1, \mu^1 : \hat{\gamma}^1)$ with $\mu^1 \in \partial D^\circ$, $\lambda^1 \in Q^\circ$ such that $G \subset H^+(\lambda^1, \mu^1 : \hat{\gamma}^1)$, and for any $(\tilde{u}, \tilde{y}) \in \mathring{H}^-(\lambda^1, \mu^1 : \hat{\gamma}^1)$, there exists a hyperplane $H(\lambda, \mu : \gamma)$ with $\mu \in \operatorname{int} D^\circ$ and $\lambda \in Q^\circ$ such that $G \subset H^+(\lambda, \mu : \gamma)$ and $(\tilde{u}, \tilde{y}) \in \mathring{H}^-(\lambda, \mu : \gamma)$.

To this end, for affine functions h_1 and h_0 associated respectively with $H(\lambda^1, \mu^1 : \hat{\gamma}^1)$ and $H(\lambda^0, \mu^0 : \hat{\gamma}^0)$, let

$$\begin{aligned} h(u, y) &= \theta h_1 + h_0 \\ &= \langle \theta\mu^1 + \mu^0, y\rangle + \langle \theta\lambda^1 + \lambda^0, u\rangle - (\theta\hat{\gamma}^1 + \hat{\gamma}^0). \end{aligned}$$

Clearly, for any $\theta > 0$ we have $h(u, y) \geqq 0$ for all $(u, y) \in G$, and $h(\tilde{u}, \tilde{y}) < 0$ for sufficiently large $\theta > 0$. Note that since $\mu^1 \in \partial D^\circ$ and $\mu^0 \in \operatorname{int} D^\circ$, $\theta\mu^1 + \mu^0 \in \operatorname{int} D^\circ$ for any $\theta > 0$. Therefore, the hyperplane $H(\theta\lambda^1 + \lambda^0, \theta\mu^1 + \mu^0 : \theta\hat{\gamma}^1 + \hat{\gamma}^0)$ with a sufficiently large $\theta > 0$ is the desired one. This completes the proof.

Proposition 5.3.1

Suppose that G is closed and that there is at least a properly efficient solution to the primal problem. Then, under the condition of Slater's constraint qualification,

$$(Y_G)^c \subset \bigcup_{\substack{\mu \in \operatorname{int} D^\circ \\ \lambda \in Q^\circ}} Y_{H^-(\lambda, \mu)} \subset \bigcup_{\Lambda \in \mathscr{L}} Y_{S(\Lambda)} \subset \operatorname{cl}(Y_G)^c.$$

Proof Let $\tilde{y} = f(\tilde{x})$ be a properly efficient point in the objective space to the primal problem. Then there exists $\tilde{\mu} \in \operatorname{int} D^\circ$ such that

$$\langle \tilde{\mu}, f(x) \rangle \geqq \langle \tilde{\mu}, \tilde{y} \rangle \qquad \text{for all} \quad x \in X.$$

It is easy to show that $\langle \tilde{\mu}, f(x) \rangle$ is convex with respect to $x \in X'$. Hence, the well-known Lagrange multiplier rule for scalar convex programms under the Slater's constraint qualification yields the existence of $\tilde{\lambda} \in Q^\circ$ such that

$$\langle \tilde{\mu}, f(x) \rangle + \langle \tilde{\lambda}, g(x) \rangle \geqq \langle \tilde{\mu}, \tilde{y} \rangle \qquad \text{for all} \quad x \in X'. \tag{5.3.5}$$

Since there exists $x \in X'$ such that $f(x) \leqq_D y$ and $g(x) \leqq_Q u$ for each $(u, y) \in G$, it follows from Eq. (5.3.5) that

$$\langle \tilde{\mu}, y \rangle + \langle \tilde{\lambda}, u \rangle \geqq \langle \tilde{\mu}, \tilde{y} \rangle \qquad \text{for all} \quad (u, y) \in G.$$

This implies that there exists a halfspace defined by a hyperplane with normal $(\tilde{\lambda}, \tilde{\mu})$ such that $\tilde{\mu} \in \operatorname{int} D^\circ$ and $\tilde{\lambda} \in Q^\circ$ that contains G. Since G is closed, we can now apply Lemma 5.3.5. Note here that $(Y_G)^c = Y_{G^c}$. Hence, it follows from Lemma 5.3.5 that

$$\begin{aligned}(Y_G)^c &= \bigcup_{\substack{\mu \in \operatorname{int} D^\circ \\ \lambda \in Q^\circ}} Y_{\mathring{H}^-(\lambda, \mu : \hat{y})} \\ &\subset \bigcup_{\substack{\mu \in \operatorname{int} D^\circ \\ \lambda \in Q^\circ}} Y_{H^-(\lambda, \mu)}.\end{aligned}$$

Next suppose that $\bar{y} \in Y_{H^-(\bar{\lambda}, \bar{\mu})}$ for some $\bar{\lambda} \in Q^\circ$ and some $\bar{\mu} \in \operatorname{int} D^\circ$; namely,

$$\langle \bar{\mu}, f(x) \rangle + \langle \bar{\lambda}, g(x) \rangle \geqq \langle \bar{\mu}, \bar{y} \rangle \qquad \text{for all} \quad x \in X'.$$

In other words, for a matrix $\bar{\Lambda} \in \mathscr{L}$ such that $\bar{\Lambda}^T \bar{\mu} = \bar{\lambda}$,

$$\langle \bar{\mu}, f(x) + \bar{\Lambda} g(x) \rangle \geqq \langle \bar{\mu}, \bar{y} \rangle.$$

Therefore,

$$f(x) + \bar{\Lambda} g(x) \nleq_D \bar{y} \qquad \text{for all} \quad x \in X'.$$

Thus,

$$\bigcup_{\substack{\mu \in \operatorname{int} D^\circ \\ \lambda \in Q^\circ}} Y_{H^-(\lambda, \mu)} \subset \bigcup_{\Lambda \in \mathscr{L}} Y_{S(\Lambda)}$$

We now show the last inclusion in the proposition. To this end, suppose that $\bar{y} \in Y_{S(\bar{\Lambda})}$ for some $\bar{\Lambda} \in \mathscr{L}$. Then, it follows immediately from the weak duality (Theorem 5.3.2) that,

$$\bar{y} \ngeq_D y \qquad \text{for all} \quad y \in Y_G.$$

This implies

$$\bar{y} \notin \operatorname{int} Y_G .$$

For, supposing to the contrary that $\bar{y} \in \operatorname{int} Y_G$, there exists $y' \in \operatorname{int} Y_G$ such that $\bar{y} \geq_D y'$, which leads to a contradiction. Hence,

$$\bar{y} \in (\operatorname{int} Y_G)^c = \operatorname{cl}(Y_G)^c.$$

This completes the proof.

For simplicity of notation, let $\mathrm{Min}_D(\mathrm{P})$, $\mathrm{Max}_D(\mathrm{D_N})$, and $\mathrm{Max}_D(\mathrm{D_J})$ denote the set of all efficient points in the objective space of the problems (P), $(\mathrm{D_N})$, and $(\mathrm{D_J})$, respectively. The following lemma will be used in deriving strong duality.

Lemma 5.3.6

The following holds:

$$\mathrm{Min}_D(P) = \mathrm{Min}_D\, Y_G .$$

Proof Note that

$$Y_G = \{f(x) : x \in X\} + D.$$

Since D is a pointed convex cone, according to Proposition 3.1.2 we have

$$\mathrm{Min}_D\, Y_G = \mathrm{Min}_D[\{f(x) : x \in X\} + D] := \mathrm{Min}_D(P).$$

We can now establish a strong duality. The first part of the following theorem is due to Nakayama [N5], and the second part is due to Jahn [J1].

Theorem 5.3.3

Assume that G is closed, that there exists at least a D-minimal solution to the primal problem, and that these solutions are all proper. Then, under the condition of Slater's constraint qualification, the following holds:

(i) $\mathrm{Min}_D(\mathrm{P}) = \mathrm{Max}_D(\mathrm{D_N})$
(ii) $\mathrm{Min}_D(\mathrm{P}) = \mathrm{Max}_D(\mathrm{D_J})$.

Proof It can be shown in a similar fashion to Theorem 5.2.5 that each efficient point of the primal problem is also an efficient point of the dual problems under consideration; this is left as an exercise. We prove that each efficient point of the dual problem is also an efficient point of the primal problem.

To show part (i), suppose that $\bar{y} \in \mathrm{Max}_D(\mathrm{D_N})$. Then, according to the weak duality (Theorem 5.3.2),

$$f(x) \not\leq_D \bar{y} \qquad \text{for all} \quad x \in X,$$

and, hence,

$$y \not\leq_D \bar{y} \qquad \text{for all} \quad y \in Y_G. \tag{5.3.6}$$

On the other hand, clearly $\bar{y} \in \mathrm{Max}_D(\mathrm{D_N})$ implies $\bar{y} \notin \operatorname{int} \bigcup_{\Lambda \in \mathscr{L}} Y_{S(\Lambda)}$. This fact and Proposition 5.3.1 yield $\bar{y} \in \partial\mathrm{cl}(Y_G)^c = \partial Y_G$; from this and relation (5.3.6) we have $\bar{y} \in \mathrm{Min}_D(Y_G)$. Finally $\bar{y} \in \mathrm{Min}_D(\mathrm{P})$ follows from Lemma 5.3.6. This completes the proof.

The proof of (ii) is quite similar, and thus it is omitted here.

5.3.3 Normality Condition

In the stated duality, we assumed that the convex set G is closed and that Slater's constraint qualification is satisfied, which seem relatively restrictive. Instead of these conditions, in his original work Jahn [J1] assumed that Y_G is closed and some normality condition. We shall discuss this condition in more detail.

For a scalar primal optimization problem

$$(\mathrm{P}') \qquad \inf\{f(x) : g(x) \leqq_Q 0,\ x \in X' \subset R^n\},$$

we define

$$G' := \{(u, y) : y \in R^1,\ u \in R^m,\ y \geqq f(x), \text{ and } u \geqq_Q g(x) \text{ for some } x \in X'\},$$

$$Y_{G'} := \{y : (0, y) \in G',\ 0 \in R^m,\ y \in R^1\},$$

$$\Gamma' := \{u : g(x) \leqq_Q u \text{ for some } x \in X'\},$$

$$w(u) := \inf\{f(x) : g(x) \leqq_Q u,\ u \in \Gamma'\},$$

and the dual problem

$$(\mathrm{D}') \qquad \sup\{\phi(\lambda) : \lambda \in Q^\circ\} \text{ where } \phi(\lambda) := \inf\{f(x) + \lambda^{\mathrm{T}} g(x) : x \in X'\}.$$

In this event, note that, in general, (e.g., Ponstein [P2])

$$G' \subset \operatorname{epi} w \subset \operatorname{cl} G'.$$

Then problem (P′) is said to be *normal*, if

$$\mathrm{cl}(Y_{G'}) = Y_{\mathrm{cl}\,G'}.$$

It is well known (e.g., Ponstein [P2], Rockafellar [R7]) that for convex problems, if (P′) has a finite optimal value of f and if the normality condition holds, then the optimal value of primal problem (P′) equals to that of dual problem (D′) (strong duality). The normality condition plays a role to exclude the so-called duality gap here. Notice that Slater's constraint qualification yields the normality.

In Jahn's formulation for multiobjective optimization, the stated normality condition can be extended as follows. For a given μ, let

$$G(\mu) := \{(u, \alpha) : \alpha = \langle \mu, y \rangle, (u, y) \in G\},$$

where

$$G := \{(u, y) \in R^m \times R^p : y \geqq_D f(x),\ u \geqq_Q g(x) \text{ for some } x \in X'\}$$

Furthermore, as in the previous subsection, let $A_{G(\mu)}$ denote the α-intercept of $G(\mu)$ and let Y_G denote the y-intercept of G, namely,

$$A_{G(\mu)} := \{\alpha : (0, \alpha) \in G(\mu),\ 0 \in R^m,\ \alpha \in R^1\},$$

$$Y_G := \{y : (0, y) \in G,\ 0 \in R^m,\ y \in R^p\}.$$

The convexity of G is assumed throughout this subsection.

Definition 5.3.3 (*J-Normal*)

The primal problem (P) is said to be J-normal, if for every $\mu \in \operatorname{int} D^\circ$

$$\operatorname{cl}(A_{G(\mu)}) = A_{\operatorname{cl} G(\mu)}.$$

Definition 5.3.4 (*J-Stable*)

The primal problem (P) is said to be J-stable, if it is J-normal and for an arbitrary $\mu \in \operatorname{int} D^\circ$ the problem

$$\sup_{\lambda \in Q^\circ} \inf_{x \in X'} \langle \mu, f(x) \rangle + \langle \lambda, g(x) \rangle$$

has at least one solution.

On the other hand, Nieuwenhuis [N15] suggested extending the normality condition in a more natural way to multiobjective optimization.

Definition 5.3.5 (*N-Normal*)

The primal problem (P) is said to be N-normal, if

$$\operatorname{cl} Y_G = Y_{\operatorname{cl} G}.$$

Lemma 5.3.7

Suppose that $\operatorname{int} Q \neq \varnothing$. Then Slater's constraint qualification ($\exists x^0, g(x^0) \in \operatorname{int} Q$) yields J-stability and N-normality.

Proof J-stability follows immediately from a property in ordinary convex programming with Slater's constraint qualification. To show N-normality, note first that for some x^0 such that $g(x^0) <_Q 0$ and any $d \in \operatorname{int} D$

$$(0, f(x^0) + d) \in \operatorname{int} G,$$

which is an immediate result of the monotonicity of $Y(u)$ defined by relation (5.2.1), (i.e., $Y(u^1) + D \supset Y(u^2)$ for $u^1 > u^2$). We arrive at the desired result via Rockafellar [R7, Corollary 6.5.1]. On the other hand, a direct proof is as follows. Since always $\operatorname{cl} Y_G \subset Y_{\operatorname{cl} G}$, we shall prove the reverse inclusion. Let $\bar{y} \in Y_{\operatorname{cl} G}$. Furthermore, for a fixed $d \in \operatorname{int} D$, let $y' = f(x^0) + d$. Then, since $(0, y') \in \operatorname{int} G$ and $(0, \bar{y}) \in \operatorname{cl} G$, we have for any $\alpha \in [0, 1)$

$$\alpha(0, \bar{y}) + (1 - \alpha)(0, y') \in G,$$

which implies

$$\alpha\bar{y} + (1 - \alpha)y' \in Y_G.$$

Letting $\alpha \to 1$, we finally have $\bar{y} \in \operatorname{cl} Y_G$. This completes the proof.

Theorem 5.3.4†

Suppose that Y_G is a nonempty convex subset in R^p. Then the following hold:

(i) If $\inf\{\langle\mu, y\rangle : y \in Y_G\}$ is finite for some $\mu \in \operatorname{int} D^\circ$, then J-normality implies N-normality.

(ii) If $0^+(\operatorname{cl} Y_G) = D$, then N-normality implies J-normality.

Proof

(i) Since $\operatorname{cl} Y_G \subset Y_{\operatorname{cl} G}$ is always true, we shall show the reverse inclusion. Define hyperplanes and associated halfspaces in the objective spaces as follows:

$$h'(\mu : \rho) := \langle\mu, y\rangle - \rho,$$

$$H'(\mu : \rho) := \{y \in R^p : h'(\mu : \rho) = 0\},$$

$$H'^+(\mu : \rho) := \{y \in R^p : h'(\mu : \rho) \geqq 0\},$$

$$\mathring{H}'^+(\mu : \rho) := \{y \in R^p : h'(\mu : \rho) > 0\},$$

and

$$\hat{\rho} := \inf\{\langle\mu, y\rangle : y \in Y_G).$$

† Borwein and Nieuwenhuis [B19].

Here, H'^-, and $\mathring{H}'^-$ are defined by reversing the inequality of H'^+ and $\mathring{H}'^+$, respectively. Notice that under the condition, by the same reason of Lemma 5.3.5

$$\mathrm{cl}\, Y_G = \bigcap_{\mu \in \mathrm{int} D^\circ} H'^+(\mu : \rho).$$

Hence, if $\bar{y} \notin \mathrm{cl}\, Y_G$, there exists $\bar{\mu} \in \mathrm{int}\, D^\circ$ such that

$$\langle \bar{\mu}, y \rangle > \langle \bar{\mu}, \bar{y} \rangle \qquad \text{for any} \quad y \in \mathrm{cl}\, Y_G.$$

Thus, $\langle \bar{\mu}, \bar{y} \rangle \notin \mathrm{cl}\, A_{G(\bar{\mu})}$. Furthermore, by the J-normality $\langle \bar{\mu}, \bar{y} \rangle \notin A_{\mathrm{cl}G(\mu)}$, which implies also $\bar{y} \notin Y_{\mathrm{cl}G}$, thereby establishing the desired result.

(ii) We shall prove $\mathrm{cl}\, A_{G(\mu)} \supset A_{\mathrm{cl}\, G(\mu)}$ for all $\mu \in \mathrm{int}\, D^\circ$. To this end, take $\bar{\mu} \in \mathrm{int}\, D^\circ$ and $\bar{\alpha} \notin \mathrm{cl}\, A_{G(\bar{\mu})}$. Then clearly under our condition, we have for some $\varepsilon > 0$ and for any $y \in Y_G$ (and hence for any $y \in \mathrm{cl}\, Y_G$)

$$\langle \bar{\mu}, y \rangle \geqq \bar{\alpha} + \varepsilon.$$

Suppose, to the contrary, that $\bar{\alpha} \in A_{\mathrm{cl}\, G(\bar{\mu})}$. Then there exists a sequence $\{(u^k, \alpha_k)\} \in G(\bar{\mu})$ converging to $(0, \bar{\alpha})$. Furthermore, $(u^k, \alpha_k) \in G(\bar{\mu})$ implies the existence of an element $y^k \in R^p$ such that $\langle \bar{\mu}, y^k \rangle = \alpha_k$ and $(u^k, y^k) \in G$. If $y^k \to y'$, then $(0, y') \in \mathrm{cl}\, G$; namely, $y' \in Y_{\mathrm{cl}G}$, and, hence, by N-normality $y' \in \mathrm{cl}\, Y_G$. This yields $\bar{\alpha} \geqq \bar{\alpha} + \varepsilon$, which is a contradiction establishing the theorem.

Therefore, in order to show $y^k \to y'$, we prove the boundedness of the sequence $\{y^k\}$. Suppose to the contrary that $\|y^k\| \to +\infty$. Then for an appropriate subsequence

$$\lim_{k \to \infty} \left(\frac{u^k}{\|y^k\|}, \frac{y^k}{\|y^k\|} \right) = (0, \hat{y})$$

is a direction of recession of cl G. Therefore, $\hat{y}$ is a direction of recession of $Y_{\mathrm{cl}G}$, and, hence, of cl Y_G due to N-normality. Finally, it follows from the assumption that $\hat{y} \in D$. As $\hat{y} \neq 0$ and $\bar{\mu} \in \mathrm{int}\, D^\circ$, this implies $\langle \bar{\mu}, \hat{y} \rangle > 0$. But

$$\left\langle \bar{\mu}, \frac{y^k}{\|y^k\|} \right\rangle = \frac{\alpha_k}{\|y^k\|} \to 0,$$

which leads to a desirable contradiction.

We can now relax some of conditions of Proposition 5.3.1.

Corollary 5.3.1

Suppose that there is at least a properly D-minimal solution to the primal problem. Then, under the condition of J-stability,

$$(\mathrm{cl}\, Y_G)^c \subset \bigcup_{\substack{\mu \in \mathrm{int}\, D^\circ \\ \lambda \in Q^\circ}} Y_{H^-(\lambda,\mu)} \subset \bigcup_{\Lambda \in \mathscr{L}} Y_{S(\Lambda)} \subset \mathrm{cl}(\mathrm{cl}\, Y_G)^c.$$

Proof The result can be obtained in almost the same way as Proposition 5.3.1. It should be noted that N-normality is assured by the existence of a (proper) D-minimal solution to the primal problem and by J-normality (Theorem 5.3.4). Notice that, in deriving the relation (5.3.5), Slater's constraint qualification can be replaced by J-stability. By considering $\mathrm{cl}\, G$ instead of G itself and using the N-normality condition, the first inclusion follows immediately. The proof of the second inclusion is the same as that of Proposition 5.3.1. The third inclusion can be obtained similarly by replacing G by $\mathrm{cl}\, G$ and by using N-normality.

Corollary 5.3.2

Suppose that Y_G is closed, $\mathrm{Min}_D(\mathrm{P}) \neq \varnothing$, and the D-minimal solutions are all proper. Then, under the condition of J-stability,

$$\mathrm{Min}_D(P) = \mathrm{Max}_D(D_N) = \mathrm{Max}_D(D_J).$$

Proof Immediate from Theorem 5.3.3 and Corollary 5.3.1.

5.3.4 Duality for Weak Efficiency

We shall consider duality for weak efficiency in parallel with the previous section. Convexity (D-convexity or Q-convexity) assumptions of f, g, and X' are also assumed here, but X' is not necessarily compact nor closed. Here, G, Y_G, and $Y_{H^-(\lambda,\mu)}$ are the same as in the previous section. Let

$$Y_{S'(\Lambda)} := \{y \in R^p : f(x) + \Lambda g(x) \not<_D y,\ \forall x \in X'\}.$$

Proposition 5.3.2

Suppose that Y_G is a nonempty D-bounded set in R^p. Then under the condition of N-normality

$$\mathrm{cl}(\mathrm{cl}\, Y_G)^c = \mathrm{cl}\Bigg(\bigcup_{\substack{\mu \in D^\circ\setminus\{0\} \\ \lambda \in Q^\circ}} Y_{H^-(\lambda,\mu)}\Bigg) = \mathrm{cl}\Bigg(\bigcup_{\Lambda \in \mathscr{L}} Y_{S'(\Lambda)}\Bigg).$$

Proof Let $\mathring{H}^-(\lambda, \mu)$ denote an open halfspace with normal $(\lambda, \mu) \in Q^\circ \times (D^\circ \backslash \{0\})$ that does not have a common point with cl G; namely, for $\hat{\gamma} = \inf\{\langle \mu, y\rangle + \langle \lambda, u\rangle : (u, y) \in G\}$,

$$\mathring{H}^-(\lambda, \mu) := \{(\bar{u}, \bar{y}) : \hat{\gamma} > \langle \mu, \bar{y}\rangle + \langle \lambda, \bar{u}\rangle\}$$

By definition, it is easy to show that

$$(\text{cl } G)^c = \bigcup_{\substack{\mu \in D^\circ \backslash \{0\} \\ \lambda \in Q^\circ}} \mathring{H}^-(\lambda, \mu).$$

Therefore,

$$Y_{(\text{cl} G)^c} = \bigcup_{\substack{\mu \in D^\circ \backslash \{0\} \\ \lambda \in Q^\circ}} Y_{\mathring{H}^-(\lambda, \mu)}.$$

By using $Y_{(\text{cl} G)^c} = (Y_{\text{cl} G})^c$, we finally have (due to N-normality)

$$\text{cl}(\text{cl } Y_G)^c = \text{cl}\left(\bigcup_{\substack{\mu \in D^\circ \backslash \{0\} \\ \lambda \in Q^\circ}} Y^\circ_{H^-(\lambda, \mu)}\right) = \text{cl}\left(\bigcup_{\substack{\mu \in D^\circ \backslash \{0\} \\ \lambda \in Q^\circ}} Y_{H^-(\lambda, \mu)}\right).$$

We now turn to the proof of

$$\text{cl}\left(\bigcup_{\substack{\mu \in D^\circ \backslash \{0\} \\ \lambda \in Q^\circ}} Y_{H^-(\lambda, \mu)}\right) = \text{cl}\left(\bigcup_{\Lambda \in \mathscr{L}} Y_{S'(\Lambda)}\right).$$

For simplicity of notation, define

$$B_1 := \bigcup_{\substack{\mu \in D^\circ \backslash \{0\} \\ \lambda \in Q^\circ}} Y_{H^-(\lambda, \mu)}$$

$$B_2 := \bigcup_{\Lambda \in \mathscr{L}} Y_{S'(\Lambda)}.$$

Suppose that $y' \in \text{cl } B_1$. Then there exists a sequence $\{y^k\} \subset B_1$ converging to y'. Here $y^k \in B_1$ implies that there exist $\mu^k \in D^\circ \backslash \{0\}$ and $\lambda^k \in Q^\circ$ such that for any $x \in X'$

$$\langle \mu^k, f(x)\rangle + \langle \lambda^k, g(x)\rangle \geqq \langle \mu^k, y^k\rangle.$$

Choose $\Lambda^k \in \mathscr{L}$ such that $\Lambda^{kT}\mu^k = \lambda^k$. Then, since $\mu^k \in D^\circ \backslash \{0\}$ yields $\langle \mu^k, d\rangle > 0$ for an arbitrary $d \in \text{int } D$, we have

$$\langle \mu^k, f(x) + \Lambda^k g(x)\rangle > \langle \mu^k, y^k - d\rangle,$$

from which

$$(y^k - \operatorname{int} D) \cap \{(f(x) + \Lambda^k g(x)) : x \in X'\} = \varnothing.$$

This implies

$$y^k \in Y_{S'(\Lambda^k)}.$$

Hence,

$$y' \in \operatorname{cl} B_2,$$

which proves $\operatorname{cl} A \subset \operatorname{cl} B_2$.

Next suppose that $y' \in \operatorname{cl} B_2$. Then there exists a sequence $\{y^k\} \subset B_2$ converging to y'. Here $y^k \in B_2$ implies that there exists $\Lambda^k \in \mathscr{L}$ such that the set $\{f(x) + \Lambda^k g(x) : x \in X'\}$ is properly separated from the set $y^k - \operatorname{int} D$ by some hyperplane. In other words, there exist $\mu^k \in D^\circ \backslash \{0\}$ such that for an arbitrary $d \in \operatorname{int} D$ and for any $x \in X'$

$$\langle \mu^k, f(x) + \Lambda^k g(x) \rangle > \langle \mu^k, y^k - d \rangle.$$

Let $\lambda^k = \Lambda^{kT}\mu^k$. Then since $d \in \operatorname{int} D^\circ$ is arbitrary, we have for any $x \in X'$

$$\langle \mu^k, f(x) \rangle + \langle \lambda_k, g(x) \rangle \geqq \langle \mu^k, y^k \rangle,$$

which implies $y^k \in B_1$, and, hence, $y' \in \operatorname{cl} B_1$. This completes the proof.

Let us denote the set of weak efficient solutions of Y with respect to $<_D$ (resp. $<_{-D}$) by w-$\operatorname{Min}_D Y$ (resp. w-$\operatorname{Max}_D Y$). We then have the following duality theorem.

Lemma 5.3.8

For an arbitrary set Y in R^p, suppose that w-$\operatorname{Min}_D(Y + D) \neq \varnothing$ and w-$\operatorname{Max}_D \operatorname{cl}(Y - D) \neq \varnothing$. Then the following hold

(i) $\partial(Y + D) =$ w-$\operatorname{Min}_D \operatorname{cl}(Y + D)$,

(ii) $\partial(Y - D) =$ w-$\operatorname{Max}_D \operatorname{cl}(Y - D)$,

(iii) w-$\operatorname{Min}_D \operatorname{cl}(Y + D) =$ w-$\operatorname{Max}_D \operatorname{cl}(\operatorname{cl}(Y + D))^c$.

Proof

(i) It is easy to show that w-$\operatorname{Min}_D \operatorname{cl}(Y + D) \subset \partial \operatorname{cl}(Y + D)$. Hence, we shall prove the reverse inclusion. Suppose that $\bar{y} \in \partial(Y + D)$. Furthermore, to the contrary, suppose that $\bar{y} \notin$ w-$\mathrm{Min}_D \operatorname{cl}(Y + D)$; that is, there exists a point $y' \in \operatorname{cl}(Y + D)$ such that $\bar{y} \in y' + \operatorname{int} D$. Since $y' + \operatorname{int} D$ is an open set included by $\operatorname{cl}(Y + D)$, $\bar{y}$ can never be a boundary point of $Y + D$, which leads to a contradiction.

(ii) the result follows immediately since w-Min_{-D} is identical to w-Max_D.
(iii) Note first that

$$(\operatorname{cl}(Y + D))^c = (\operatorname{cl}(Y + D))^c - D. \tag{5.3.8}$$

In fact, for an arbitrary subset B of R^p, $B = B + D$ implies $B^c = B^c - D$. For, we have $B^c - D \subset B^c$ as follows; if $y' \in B^c - D$, there exist $\bar{y} \in B^c$ and $\bar{d} \in D$ such that $y' = \bar{y} - \bar{d}$. Suppose, to the contrary, that $y' \in B$, then we have $\bar{y} = y' + \bar{d} \in B + D = B$, which leads to a contradiction. Hence, for a proof of (5.3.8), it suffices to show $\operatorname{cl}(Y + D) = \operatorname{cl}(Y + D) + D$. To this end, suppose that $\hat{y} \in \operatorname{cl}(Y + D) + D$. Then there exist $\bar{y} \in \operatorname{cl}(Y + D)$ and $\bar{d} \in D$ such that $\hat{y} = \bar{y} + \bar{d}$. Furthermore, $\bar{y} \in \operatorname{cl}(Y + D)$ implies that there exist sequences $\{y^k\}$ in Y and $\{d^k\}$ in D such that $y^k + d^k$ converges to $\bar{y}$. Now observe that $y^k + d^k + \bar{d} \in Y + D$, because D is a convex cone. Finally, since $\hat{y}$ is a limit point of $y^k + d^k + \bar{d}$, we have $\hat{y} \in \operatorname{cl}(Y + D)$, from which Eq. (5.3.8) follows.

Now turn to the proof of part (iii). According to parts (i) and (ii), and Eq. (5.3.8),

$$\begin{aligned}
\text{w-Min}_D \operatorname{cl}(Y + D) &= \partial \operatorname{cl}(Y + D) \\
&= \partial \operatorname{cl}(\operatorname{cl}(Y + D))^c \\
&= \partial \operatorname{cl}[(\operatorname{cl}(Y + D))^c - D] \\
&= \text{w-Max}_D \operatorname{cl}[(\operatorname{cl}(Y + D))^c - D] \\
&= \text{w-Max}_D \operatorname{cl}(\operatorname{cl}(Y + D))^c.
\end{aligned}$$

Theorem 5.3.3

Under the same assumptions as in Proposition 5.3.2, the following holds:

$$\text{w-Min}_D \operatorname{cl} Y_G = \text{w-Max}_D \operatorname{cl} \bigcup_{\Lambda \in \mathscr{L}} Y_{S'(\Lambda)} = \text{w-Max}_D \operatorname{cl}\left(\bigcup_{\substack{\mu \in D^\circ \setminus \{0\} \\ \lambda \in Q^\circ}} Y_{H^-(\lambda, \mu)} \right).$$

Proof According to Lemma 5.3.8, w-$\mathrm{Min}_D \operatorname{cl} Y_G =$ w-$\mathrm{Max}_D \operatorname{cl}(Y_G)^c$; from this and from Proposition 5.3.2, the theorem follows immediately.

Remark 5.3.3

In the stated duality in this subsection, there appeared no perturbation map and, hence, no dual map. In order to develop duality theory in connection with these notions, we may expect to modify the result of the previous sections by using the notion of inf (or sup) defined as $\mathrm{Inf}_D\, Y := \mathrm{Min}_D \operatorname{cl} Y$ instead of $\mathrm{Min}_D\, Y$ itself. As for w-Inf_D, see Kawasaki [K3] and Nieuwenhuis [N15], which will be introduced briefly in the following chapter. As yet, unfortunately, there has been no development of inf playing an effective role in duality for strong efficiency.

6 CONJUGATE DUALITY

This chapter is devoted to the duality theory via conjugacy in multiobjective optimization. The theory has not yet been brought to completion, but is now under development. There remain several unsolved problems. Hence, we present three different kinds of results owing to different authors, namely, Tanino and Sawaragi [T11], Kawasaki [K2, K3] and Brumelle [B21]. They are based, respectively, on efficiency, weak efficiency, and strong optimality.

6.1 Conjugate Duality Based on Efficiency

This section is devoted to the conjugate duality based on efficiency, which was developed by Tanino and Sawaragi [T11]. Some well-known concepts in ordinary conjugate duality, such as conjugate functions, subgradients and stability, will be extended to the case of multiobjective optimization. Since we consider only the case in which the domination cone is R^p_+ or $-R^p_+$, the sets $\mathscr{E}(Y, R^p_+)$ and $\mathscr{E}(Y, -R^p_+)$ for $Y \subset R^p$ will be simply denoted by Min Y and Max Y, respectively. Of course when $p = 1$, Min Y and Max Y coincide with the sets $\{\min Y\}$ and $\{\max Y\}$, respectively. In this chapter, R^p_+ is often abbreviated as epi F instead of R^p_+-epi F.

6.1.1 Conjugate Maps

In this subsection, we introduce the concept of conjugate maps for vector-valued functions and point-to-set maps. When f is an extended real-valued function on R^n (i.e., when f: $R^n \to \overline{R}$) its conjugate function f^* is defined by

$$f^*(x^*) = \sup\{\langle x^*, x\rangle - f(x) : x \in R^n\} \qquad \text{for} \quad x^* \in R^n.$$

(See Definition 2.1.20.) Unfortunately, the concept of *supremum* of a set Y in R^p or $\bar{R}^p$ in the sense of efficiency has not yet been established, though some definitions have been proposed. (See Gros [G11], Nieuwenhuis [N15], Kawasaki [K2], and Ponstein [P2, P3]. Note that the definition by Zowe [Z9] or Brumelle [B21] is entirely different because it is given as a least upper bound with respect to the order relation $\leqq$ in R^p.) Hence, we use Max and Min as an extension of usual sup and inf, respectively.

In order to consider the vector-valued case, we must define a paired space of R^n with respect to R^p. A most natural paired space is the set of all $p \times n$ matrices, which is denoted by $R^{p \times n}$. However its dimension $p \times n$ is often too large. Another idea is to take $(R^n)^* = R^n$ as a paired space as in the scalar case and to consider a p-dimensional vector $\langle \mathbf{t}, \mathbf{x} \rangle := (\langle t, x \rangle, \ldots, \langle t, x \rangle)^{\mathrm{T}}$ for $t \in R^n$. Since $\langle \mathbf{t}, \mathbf{x} \rangle = (t, \ldots, t)^{\mathrm{T}} x$, the vector variable is a special case of the matrix variable. Hence, we mainly discuss the case of the matrix variable, though most results in this section are valid in both cases.

Definition 6.1.1 (*Conjugate Map and Biconjugate Map*)

Let F be a point-to-set map from R^n into R^p. The point-to-set map $F^* : R^{p \times n} \to R^p$ defined by

$$F^*(T) = \operatorname{Max} \bigcup_{x \in R^n} [Tx - F(x)] \qquad \text{for} \quad T \in R^{p \times n}$$

is called the conjugate map of F. In particular, the conjugate map F^{**} for the conjugate map F^* is called the biconjugate map of F; i.e.,

$$F^{**}(x) = \operatorname{Max} \bigcup_{T \in R^{p \times n}} [Tx - F^*(T)].$$

When f is a vector-valued function from R^n to $R^p \cup \{+\infty\}$, let $\operatorname{dom} f = \{x \in R^n : f(x) \neq +\infty\}$ and define the conjugate map f^* of f by

$$f^*(T) = \operatorname{Max}\{Tx - f(x) : x \in \operatorname{dom} f\}.$$

Here $+\infty$ is the imaginary point whose every component is $+\infty$. We identify the function f as the point-to-set map that is equal to $\{f(x)\}$ for $x \in \operatorname{dom} f$ and is empty otherwise. The biconjugate map f^{**} can be defined as the conjugate map of f^*.

Proposition 6.1.1

Let F be a point-to-set map from R^n into R^p and $\bar{x} \in R^n$. If we define another point-to-set map G by $G(x) = F(x + \bar{x})$, then

(i) $G^*(T) = F^*(T) - T\bar{x}$,
(ii) $G^{**}(x) = F^{**}(x + \bar{x})$.

Proof (i)

$$\begin{aligned} G^*(T) &= \operatorname{Max} \bigcup_x [Tx - G(x)] \\ &= \operatorname{Max} \bigcup_x [Tx - F(x + \bar{x})] \\ &= \operatorname{Max} \bigcup_{x'} [T(x' - \bar{x}) - F(x')] \\ &= \operatorname{Max} \bigcup_{x'} [Tx' - F(x')] - T\bar{x} \\ &= F^*(T) - T\bar{x}. \end{aligned}$$

(ii)

$$\begin{aligned} G^{**}(x) &= \operatorname{Max} \bigcup_T [Tx - G^*(T)] \\ &= \operatorname{Max} \bigcup_T [T(x + \bar{x}) - F^*(T)] \\ &= F^{**}(x + \bar{x}). \end{aligned}$$

Proposition 6.1.2

Let F be a point-to-set map from R^n into R^p and $\bar{y} \in R^p$. Then

$$(F + \bar{y})^*(T) = F^*(T) - \bar{y},$$

$$(F + \bar{y})^{**}(x) = F^{**}(x) + \bar{y}.$$

Proof Left to the reader.

Proposition 6.1.3†

Let F be a point-to-set map from R^n into R^p. If $y \in F(\hat{x})$ and $y' \in F^*(T)$, then

$$y + y' \not\leq T\hat{x} \qquad (\text{i.e., } T\hat{x} - (y + y') \notin R^p_+ \setminus \{0\}).$$

Proof Since $y \in F(\hat{x})$, $T\hat{x} - y \in \bigcup_x [Tx - F(x)]$. Hence, if $y' \leq T\hat{x} - y$, it contradicts the assumption $y' \in F^*(T) = \operatorname{Max} \bigcup_x [Tx - F(x)]$.

Corollary 6.1.1

Let F be a point-to-set map from R^n into R^p. If $y \in F(0)$ and $y' \in -F^*(T)$, then $y \not\leq y'$. Moreover, if $y \in F(x)$ and $y' \in F^{**}(x)$, then $y \not\leq y'$.

† An extension of Fenchel's inequality; cf. Proposition 2.1.27.

Proof The former part is immediate from Proposition 6.1.3. Hence, for $y \in F(0)$ and $y' \in F^{**}(0)$, $y \not\leq y'$, and the latter part follows from Proposition 6.1.1.

Lemma 6.1.1

Let F_1 and F_2 be point-to-set maps from R^n into R^p. Then

$$\operatorname{Max} \bigcup_x [F_1(x) + F_2(x)] \subset \operatorname{Max} \bigcup_x [F_1(x) + \operatorname{Max} F_2(x)].$$

If $\operatorname{Max} F_2(x)$ is externally stable (i.e., $F_2(x) \subset \operatorname{Max} F_2(x) - R^p_+$) for every $x \in R^n$, then the converse inclusion also holds.

Proof Let $\hat{y} \in \operatorname{Max} \bigcup_x [F_1(x) + F_2(x)]$. Then there exists $\hat{x} \in R^n$ such that $\hat{y} = y^1 + y^2$ for some $y^1 \in F_1(\hat{x})$ and $y^2 \in F_2(\hat{x})$. If we suppose that $y^2 \notin \operatorname{Max} F_2(\hat{x})$, there exists $\bar{y}^2 \in F_2(\hat{x})$ such that $y^2 \leq \bar{y}^2$. Then $\hat{y} = y^1 + y^2 \leq y^1 + \bar{y}^2$, which is a contradiction. Therefore, $y^2 \in \operatorname{Max} F_2(\hat{x})$. Since $\bigcup_x [F_1(x) + F_2(x)] \supset \bigcup_x [F_1(x) + \operatorname{Max} F_2(x)]$, then

$$\hat{y} \in \operatorname{Max} \bigcup_x [F_1(x) + \operatorname{Max} F_2(x)].$$

Next, suppose that $\operatorname{Max} F_2(x)$ is externally stable for every x, then

$$F_2(x) - R^p_+ = \operatorname{Max} F_2(x) - R^p_+ \qquad \text{for every} \quad x.$$

Thus

$$F_1(x) + F_2(x) - R^p_+ = F_1(x) + \operatorname{Max} F_2(x) - R^p_+ \qquad \text{for any} \quad x,$$

$$\bigcup_x [F_1(x) + F_2(x)] - R^p_+ = \bigcup_x [F_1(x) + \operatorname{Max} F_2(x)] - R^p_+.$$

Taking the Max of both sides, we have

$$\operatorname{Max} \bigcup_x [F_1(x) + F_2(x)] = \operatorname{Max} \bigcup_x [F_1(x) + \operatorname{Max} F_2(x)]$$

due to Proposition 3.1.2. This completes the proof of the lemma.

Corollary 6.1.2

Let F be a point-to-set map from R^n into R^p. Then

$$F^*(T) \subset \operatorname{Max} \bigcup_x [Tx - \operatorname{Min} F(x)].$$

Moreover, if $\operatorname{Min} F(x)$ is externally stable for every x, the equality holds.

Proof Take $F_1(x) = \{Tx\}$ and $F_2(x) = -F(x)$ in Lemma 6.1.1.

Corollary 6.1.3

Let F be a point-to-set map from R^n into R^p. If $\operatorname{Max} F(x)$ is externally stable for every x, then

$$\operatorname{Max} \bigcup_x F(x) = \operatorname{Max} \bigcup_x \operatorname{Max} F(x).$$

Proof Take $F_1(x) = \{0\}$ and $F_2(x) = F(x)$ in Lemma 6.1.1.

Remark 6.1.1

In the above corollaries or the lemma, we cannot dispense with the external stability condition of $F(x)$. For example, let $F: R \to R^2$ with

$$F(x) = \begin{cases} \begin{bmatrix} 0 \\ 0 \end{bmatrix} & \text{if } \ x = 0, \\ \{y \in R^2 : (y_1)^2 + (y_2)^2 < 1\} & \text{if } \ x \neq 0. \end{cases}$$

Then, for $T = (0, 0)^{\mathrm{T}}$,

$$F^*(T) = \varnothing,$$

while

$$\operatorname{Max} \bigcup_x [Tx - \operatorname{Min} F(x)] = \{0\}.$$

Similarly $\operatorname{Max} \bigcup_x F(x) = \varnothing$, but $\operatorname{Max} \bigcup_x \operatorname{Max} F(x) = \{0\}$. This drawback is probably due to the fact that we are using Max instead of Sup.

6.1.2 Subgradients of Vector-Valued Functions and Point-To-Set Maps

In this subsection, we shall introduce the concept of subgradients of vector-valued functions and point-to-set maps. A subgradient x^* of a scalar-valued function f at x is defined in Definition 2.1.23. The definition can be formally extended to a nonconvex function, though it essentially requires convexity (at least locally). A subgradient has the geometrical interpretation as the normal vector to a supporting hyperplane of epi f at $(x, f(x))$ (Remark 2.1.3). If we notice the separability of epi f and its supporting hyperplane, we may extend the definition of subgradients to the vector-valued case, which is also an intuitively direct extension of Definition 2.1.23.

Definition 6.1.2 *(Subgradient and Subdifferential)*†

(i) Let f be a function from R^n to $R^p \cup \{+\infty\}$. A $p \times n$ matrix T is said to be a subgradient of f at $\hat{x} \in \operatorname{dom} f$ if

$$f(x) \nleq f(\hat{x}) + T(x - \hat{x}) \qquad \text{for any} \quad x \in R^n;$$

i.e., if

$$f(\hat{x}) - T\hat{x} \in \operatorname{Min}\{f(x) - Tx \in R^p : x \in R^n\} = \operatorname{Min}\{f(x) - Tx : x \in \operatorname{dom} f\}.$$

The set of all subgradients of f at $\hat{x}$ is called the subdifferential of f at $\hat{x}$ and is denoted by $\partial f(\hat{x})$. If $\partial f(\hat{x})$ is not empty, then f is said to be subdifferentiable at $\hat{x}$.

(ii) Let F be a point-to-set map from R^n into R^p and $\hat{y} \in F(\hat{x})$. A $p \times n$ matrix T is said to be a subgradient of F at $(\hat{x}; \hat{y})$ if

$$\hat{y} - T\hat{x} \in \operatorname{Min} \bigcup_{x \in R^n} [F(x) - Tx].$$

The set of all subgradients of F at $(\hat{x}; \hat{y})$ is called the subdifferential of F at $(\hat{x}; \hat{y})$ and is denoted by $\partial F(\hat{x}; \hat{y})$. If $\partial F(\hat{x}; \hat{y})$ is not empty for every $\hat{y} \in F(\hat{x})$, then F is said to be subdifferentiable at $\hat{x}$.

Remark 6.1.2

As has been discussed before, T is a subgradient of F at $(\hat{x}; \hat{y})$ if and only if the polyhedral convex set

$$\begin{aligned}&\{(x, y) : y \leqq \hat{y} + T(x - \hat{x})\} \\ &\quad = \{(x, y) : \langle (t^i, -e^i), (x, y)\rangle \geqq \langle (t^i, -e^i), (\hat{x}, \hat{y})\rangle,\ i = 1, \ldots, p\}\end{aligned}$$

is a *supporting* polyhedral convex set of epi F at $(\hat{x}, \hat{y})$, where t^i is the ith row vector of T and e^i is the ith unit vector. Thus, F need not be convex in order for F to be subdifferentiable. For example, let F be a constant point-to-set map on R:

$$F(x) = \{y \in R^2 : (y_1)^2 + (y_2)^2 = 1,\ y_1 \geqq 0,\ y_2 \geqq 0\}.$$

Then the 0 matrix is a subgradient of F at $(x; y)$ for any x and $y \in F(x)$. On the other hand, epi F is not a convex set.

Remark 6.1.3

Unlike the scalar case, the subdifferential is not necessarily a closed convex set even when f is a finite convex vector-valued function (cf. Corollary 2.1.2).

(i) Let $f(x) = (f_1(x), f_2(x))^{\mathrm{T}}$ with

$$f_1(x) = \begin{cases} 0 & \text{for} \quad x \geqq 0, \\ -x & \text{for} \quad x < 0, \end{cases} \qquad \text{and} \qquad f_2(x) = -x.$$

† Cf. Definition 2.1.23.

Then $T = (t_1, 0)^{\mathrm{T}} \in \partial f(0)$ for any t_1 such that $t_1 < 0$. However, $(0, 0)^{\mathrm{T}} \notin \partial f(0)$, since $0 \notin \operatorname{Min}\{f(x) : x \in R^n\}$. Thus $\partial f(0)$ is not closed.

(ii) Let

$$f(x) = \begin{bmatrix} 0 \\ 0 \\ x \end{bmatrix} \qquad \text{for} \quad x \in R.$$

Then we can easily check that

$$\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \in \partial f(0), \qquad \text{but} \quad \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \notin \partial f(0).$$

Thus $\partial f(0)$ is not convex.

The following proposition provides a characterization of minimal solutions by the subgradient 0, that is, the stationary condition in an extended sense.

Proposition 6.1.4

(i) Let f be a function from R^n to $R^p \cup \{+\infty\}$. Then

$$f(\hat{x}) \in \operatorname{Min}\{f(x) : x \in \operatorname{dom} f\} \qquad \text{if and only if} \quad 0 \in \partial f(\hat{x}).$$

(ii) Let F be a point-to-set map from R^n into R^p. For $\hat{y} \in F(\hat{x})$,

$$\hat{y} \in \operatorname{Min} \bigcup_x F(x) \qquad \text{if and only if} \quad 0 \in \partial F(\hat{x}; \hat{y}).$$

Proof Immediate from the definition of subgradients.

The following propositions show the relationships between conjugate or biconjugate maps and subgradients.

Proposition 6.1.5

(i) Let f be a function from R^n to $R^p \cup \{+\infty\}$. Then

$$T \in \partial f(x) \qquad \text{if and only if} \quad Tx - f(x) \in f^*(T).$$

(ii) Let F be a point-to-set map from R^n to R^p. Then, for $y \in F(x)$,

$$T \in \partial F(x; y) \quad \text{if and only if} \quad Tx - y \in F^*(T).$$

Proof This proposition is obvious from the definitions of conjugate maps and subgradients.

Proposition 6.1.6

(i) Let f be a function from R^n to $R^p \cup \{+\infty\}$. Then, $f(\hat{x}) \in f^{**}(\hat{x})$ if and only if $\partial f(\hat{x}) \neq \varnothing$; i.e. if and only if f is subdifferentiable at $\hat{x}$.

(ii) Let F be a point-to-set map from R^n into R^p. For $\hat{y} \in F(\hat{x})$, $\partial F(\hat{x}; \hat{y}) \neq \emptyset$ if and only if $\hat{y} \in F^{**}(\hat{x})$. Thus F is subdifferentiable at $\hat{x}$ if and only if $F(\hat{x}) \subset F^{**}(\hat{x})$.

Proof We prove (ii) only because the proof of (i) is similar. From Proposition 6.1.5, $\partial F(\hat{x}; \hat{y}) \neq \emptyset$ if and only if there exists $T \in R^{p \times n}$ such that

$$T\hat{x} - \hat{y} \in F^*(T).$$

Hence, if $\hat{y} \in F^{**}(\hat{x})$, then it is clear that $\partial F(\hat{x}; \hat{y}) \neq \emptyset$. Conversely, suppose that $\partial F(\hat{x}; \hat{y}) \neq \emptyset$, namely that $\hat{y} \in T\hat{x} - F^*(T)$ for some T. From Proposition 6.1.3, we have $\hat{y} \not\leq y$ for any $y \in T'\hat{x} - F^*(T')$ with any T'. Therefore $\hat{y} \in F^{**}(\hat{x})$, as was to be proved.

As is pointed out in Remark 6.1.2, convexity of F is not always necessary for its subdifferentiability. However, it is often sufficient as is shown in the next proposition.

Proposition 6.1.7

Let F be a convex point-to-set map from R^n into R^p satisfying $\operatorname{Min} F(\hat{x}) = F(\hat{x})$ for some $\hat{x} \in R^n$. If $F(x) \neq \emptyset$ in some neighborhood of $\hat{x}$ and each $F^*(T)$ can be characterized completely by scalarization, then F is subdifferentiable at $\hat{x}$.

Proof We may assume without the loss of generality that $\hat{x} = 0$. Let $\hat{y} \in F(0) = \operatorname{Min} F(0)$. Then $(0, \hat{y})$ is clearly a boundary point of epi F, which is a convex set from the assumption. Hence, by Theorem 2.1.4, there exists $(\lambda, \mu) \neq 0 \in R^n \times R^p$ such that

$$\langle \mu, \hat{y} \rangle \leqq \langle \lambda, x \rangle + \langle \mu, y \rangle \qquad \text{for all} \quad (x, y) \in \operatorname{epi} F.$$

Since we can take every component of y as large as desired, then $\mu \geqq 0$. If we assume that $\mu = 0$, then

$$\langle \lambda, x \rangle \geqq 0 \qquad \text{for all} \quad (x, y) \in \operatorname{epi} F.$$

Since $F(x) \neq \emptyset$ in some neighborhood of 0, this implies that $\lambda = 0$ and so leads to a contradiction. Hence $\mu \neq 0$ and we can choose T as $-T^{\mathrm{T}}\mu = \lambda$. Then

$$\langle \mu, \hat{y} \rangle \leqq \langle \mu, y - Tx \rangle \qquad \text{for all} \quad (x, y) \in \operatorname{epi} F,$$

and so

$$\langle \mu, \hat{y} \rangle \leqq \langle \mu, y - Tx \rangle \qquad \text{for all} \quad x \in R^n \qquad \text{and} \qquad y \in F(x).$$

Since $F^*(T)$ is assumed to be characterized completely by scalarization, the above inequality implies that

$$\hat{y} \in -F^*(T).$$

Hence, $T \in \partial F(0; \hat{y})$ from Proposition 6.1.5, and the proof is completed.

Remark 6.1.4

In view of the preceding proof, we may assume that $\mu \in S^p = \{\mu' \in R^p : \mu' \geqq 0, \sum_{i=1}^p \mu_i' = 1\}$, and therefore may take $T^{\mathrm{T}} = (-\lambda, \ldots, -\lambda)$.

We have the following relationship between $\partial f(x)$ and $\partial f_i(x)$ $(i = 1, \ldots, p)$ for convex vector-valued function.

Proposition 6.1.8

Let f be a proper convex vector-valued function from R^n to $R^p \cup \{+\infty\}$ such that $\mathrm{ri}(\mathrm{dom} f) \neq \varnothing$. Then

$$\partial f(\hat{x}) \subset \bigcup_{\mu \in S^p} \left\{ T \in R^{p \times n} : T^{\mathrm{T}}\mu \in \sum_i \mu_i \, \partial f_i(\hat{x}) \right\}, \ \forall \hat{x},$$

where $S^p = \{\mu \in R^p : \mu \geqq 0, \sum_{i=1}^p \mu_i = 1\}$. Conversely, if $T^{\mathrm{T}}\mu \in \sum_i \mu_i \, \partial f_i(\hat{x})$ for some $\mu \in S^p$ and if $-f^*(T)$ can be characterized completely by scalarization; i.e.,

$$-f^*(T) = \left\{ f(x') - Tx' : \langle \mu, f(x') - Tx' \rangle \right.$$

$$\left. = \min_{x \in \mathrm{dom} f} \langle \mu, f(x) - Tx \rangle \quad \text{for some} \quad \mu \in S^p \right\},$$

then $T \in \partial f(\hat{x})$.

Proof If $T \in \partial f(\hat{x})$, then from Proposition 6.1.5,

$$f(\hat{x}) - T\hat{x} \in -f^*(T) = \mathrm{Min}\{f(x) - Tx : x \in \mathrm{dom} f\}.$$

Since $f(x) - Tx$ is a convex function, there exists $\mu \in S^p$ such that

$$\langle \mu, f(\hat{x}) \rangle - \langle T^{\mathrm{T}}\mu, \hat{x} \rangle = \min\{\langle \mu, f(x) \rangle - \langle T^{\mathrm{T}}\mu, x \rangle : x \in \mathrm{dom} f\}.$$

(See Theorem 3.4.4.) Hence,

$$0 \in \sum_{i=1}^{p} \mu_i \, \partial f_i(\hat{x}) - T^{\mathrm{T}}\mu.$$

Conversely, if $T^{\mathrm{T}}\mu \in \sum_i \mu_i \, \partial f_i(\hat{x})$ and if $f^*(T)$ can be characterized completely by scalarization, then we can trace the above procedure inversely.

Proposition 6.1.9

Let f be a vector-valued function from R^n to $R^p \cup \{+\infty\}$. If $t^i \in \partial f_i(\hat{x})$ for every $i = 1, \ldots, p$, then $T = (t^1, \ldots, t^p)^{\mathrm{T}} \in \partial f(\hat{x})$.

Proof This result is obvious.

When f is convex, we can say more about the relationships between $f^{**}(x)$ and $f(x)$.

Proposition 6.1.10

Let f be a proper convex function from R^n to $R^p \cup \{+\infty\}$. If $f(\hat{x})$ is finite, each f_i $(i = 1, \ldots, p)$ is subdifferentiable at $\hat{x}$, and the set $\{f(x) - Tx \in R^p : x \in R^n\}$ is R^p_+-closed for every $T \in R^{p \times n}$, then

$$f^{**}(\hat{x}) = \{f(\hat{x})\}.$$

Proof We may assume without the loss of generality that $\hat{x} = 0$. (See Proposition 6.1.1.) Since each f_i is subdifferentiable at 0, there exists $\hat{t}^i \in \partial f_i(0)$ $(i = 1, : p)$. In view of Proposition 6.1.9, $\hat{T} = (\hat{t}^1, \ldots, \hat{t}^p)^{\mathrm{T}} \in \partial f(0)$, and hence $f(0) \in f^{**}(0)$ by Proposition 6.1.6. Hence, we prove that $y \notin f^{**}(0)$ if $y \neq f(0)$. From Corollary 6.1.1, $y \in f(0) + R^p_+ \backslash \{0\}$ is impossible for $y \in f^{**}(0)$. Therefore, it suffices to prove that, for any $y \in R^p$ such that $y_i < f_i(0)$ for some i and $y_j > f_j(0)$ for other j's, there exist $T \in R^{p \times n}$ and $\hat{y} \in R^p$ satisfying

$$\hat{y} \in -f^*(T) \qquad \text{and} \qquad y \leq \hat{y}.$$

Hereafter, for simplicity, we assume without loss of generality that $f(0) = 0$. (See Proposition 6.1.2.) From the definition of $\hat{T}$,

$$Y(\hat{T}) := \{f(x) - \hat{T}x \in R^p : x \in R^n\} \subset R^p_+.$$

We may also assume without the loss of generality that $y_1 < 0$, $y_2 > 0, \ldots, y_p > 0$. Take a positive number ε as

$$\varepsilon < \min\{-y_1/4, y_2/2, \ldots, y_p/2\}.$$

If $f(x) = +\infty$ for all $x \neq 0$, then $-f^*(T) = \{0\}$ for all T and so $f^{**}(0) = \{0\}$. Hence, assume that $f(x) \neq +\infty$ for some $x \neq 0$. Then, since $f(0) = 0$ and f is convex, there exists some $\bar{x} \neq 0$ such that

$$0 \leqq f_i(\bar{x}) - \langle \hat{t}^i, \bar{x} \rangle < \varepsilon, \qquad i = 1, \ldots, p.$$

Define a vector $t \in R^p$ by

$$t_i = \begin{cases} y_i/2 & \text{for} \quad i = 1 \\ y_i & \text{for} \quad i = 2, \ldots, p, \end{cases}$$

let $\Delta T = (\alpha_1 t, \ldots, \alpha_n t)$ with $\sum_{j=1}^n \alpha_j \bar{x}_j = -\frac{3}{2}$ and let $\bar{y} = f(\bar{x}) - (\hat{T} + \Delta T)\bar{x}$. Then

$$\bar{y}_i = f_i(\bar{x}) - \langle \hat{t}^i, \bar{x} \rangle - \sum_{j=1}^{n} \alpha_j \bar{x}_j t_i,$$

and hence

$$\tfrac{3}{4} y_1 \leqq \bar{y}_1 < \tfrac{1}{2} y_1; \qquad \tfrac{3}{2} y_i \leqq \bar{y}_i < 2 y_i, \qquad i = 2, \ldots, p.$$

$\bar{y} \in Y(\hat{T} + \Delta T)$ and $Y(\hat{T} + \Delta T)$ is R^p_+-closed from the assumption. Moreover,

$$Y(\hat{T} + \Delta T) \subset Y(\hat{T}) + \{\Delta T x : x \in R^n\} \subset R^p_+ + \{\beta t : \beta \in R\}.$$

and $\pm t \notin R^p_+$. Therefore

$$0^+(Y(\hat{T} + \Delta T) + R^p_+) \cap (-R^p_+) = \{0\};$$

that is, $Y(\hat{T} + \Delta T)$ is R^p_+-bounded. Thus, from Theorem 3.2.10, there exists $\hat{y} \in \operatorname{Min}(Y(\hat{T} + \Delta T)) = -f^*(\hat{T} + \Delta T)$ such that $\hat{y} \leqq \bar{y}$. Since $Y(\hat{T} + \Delta T) \subset R^p_+ + \{\beta t : \beta \in R\}$, $\hat{y} \geqq \beta t$ for some $\beta \in$ R. Since

$$y_1/2 > \bar{y}_1 \geqq \hat{y}_1 \geqq \beta t_1 = \beta y_1/2,$$

we have $\beta > 1$. On the other hand,

$$2y_i > \bar{y}_i \geqq \hat{y}_i \geqq \beta t_i = \beta y_i, \qquad \text{for} \quad i = 2, \ldots, p,$$

and hence $\beta < 2$. Then, from $\beta > 1$, $\hat{y}_i > y_i$ for $i = 2, \ldots, p$, and from $\beta < 2$, $\hat{y}_1 > y_1$. Thus, we have proved that

$$y < \hat{y} \in -f^*(\hat{T} + \Delta T),$$

and hence $y \notin f^{**}(0)$, which completes the proof.

Remark 6.1.5

Proposition 6.1.10 is not necessarily valid for the vector dual variable. For example, let

$$f(x) = \begin{cases} \begin{bmatrix} x \\ x^2 \end{bmatrix} & \text{if} \quad 0 \leqq x \leqq 1, \\ +\infty & \text{otherwise.} \end{cases}$$

Then

$$\begin{bmatrix} \frac{1}{8} \\ -\frac{1}{16} \end{bmatrix} \in -f^*\left(\frac{1}{2}\right) \cap \operatorname{Max} \bigcup_{\lambda} [-f^*(\lambda)].$$

Proposition 6.1.11

Let f be a closed convex function from R^n to $R^p \cup \{+\infty\}$. If f is subdifferentiable at each $x \in \operatorname{dom} f$ and if $f(\hat{x}) = +\infty$, then $f^{**}(\hat{x}) = \emptyset$.

Proof If $f(x) \equiv +\infty$, it is clear that $f^{**}(\hat{x}) = \emptyset$. Hence, assume that $\operatorname{dom} f \neq \emptyset$. We may consider the case $\hat{x} = 0$. Since $0 \notin \operatorname{dom} f$ and $\operatorname{dom} f$ is a closed convex set, by the minimum norm theorem for convex sets (Luenberger [L14, Theorem 3.10]), there exists $\bar{x} \neq 0 \in \operatorname{dom} f$ such that

$$\langle x - \bar{x}, \bar{x} \rangle \geqq 0 \qquad \text{for any} \quad x \in \operatorname{dom} f.$$

Since f is assumed to be subdifferentiable at $\bar{x}$, there exists $T \in \partial f(\bar{x})$. Let

$$\Delta T = -\alpha(\bar{x}, \ldots, \bar{x})^{\mathrm{T}},$$

with $\alpha > 0$. Then $T + \Delta T \in \partial f(\bar{x})$. In fact, if we suppose to the contrary that $T + \Delta T \notin \partial f(\bar{x})$, there exists $x' \in \operatorname{dom} f$ such that

$$f(x') - (T + \Delta T)x' \leq f(\bar{x}) - (T + \Delta T)\bar{x};$$

i.e.,

$$f(x') - Tx' \leq f(\bar{x}) - T\bar{x} + \Delta T(x' - \bar{x}).$$

Here

$$(\Delta T(x' - \bar{x}))_i = -\alpha\langle \bar{x}, x' - \bar{x} \rangle \leqq 0,$$

whence

$$f(x') - Tx' \leq f(\bar{x}) - T\bar{x},$$

which contradicts $T \in \partial f(\bar{x})$. Therefore $T + \Delta T \in \partial f(\bar{x})$; that is,

$$f(\bar{x}) - (T + \Delta T)\bar{x} \in -f^*(T + \Delta T).$$

Moreover

$$f(\bar{x}) - (T + \Delta T)\bar{x} = f(\bar{x}) - T\bar{x} + \alpha \begin{bmatrix} \langle \bar{x}, \bar{x} \rangle \\ \langle \bar{x}, \bar{x} \rangle \end{bmatrix} \to +\infty \qquad \text{as} \quad \alpha \to +\infty.$$

Hence,

$$f^{**}(0) = \operatorname{Max} \bigcup_T [-f^*(T)] = \emptyset,$$

as was to be proved.

6.1.3 Multiobjective Optimization Problems and Duality

In this subsection we consider a multiobjective optimization problem

$$\text{(P)} \qquad \underset{x}{\text{minimize}} \quad f(x),$$

where f is an extended real vector-valued function from R^n to $R^p \cup \{+\infty\}$ and $+\infty$ is the imaginary point whose every component is $+\infty$. In other words, (P) is the problem to find $\hat{x} \in R^n$ such that

$$f(\hat{x}) \in \operatorname{Min}\{f(x) \in R^p : x \in R^n\}.$$

We denote the set $\operatorname{Min}\{f(x) \in R^p : x \in R^n\}$ simply by Min(P) and every $\hat{x}$ satisfying the above relationship is called a solution of Problem (P). If some component of $f(x)$ is $+\infty$, then x is regarded to be infeasible. Hence, we may take no thought of it and so may assume that $f(x) = +\infty$.

We analyze Problem (P) by embedding it in a family of perturbed problems. The space of perturbation is assumed to be R^m for simplicity. Thus we consider a function $\varphi: R^n \times R^m \to R^p \cup \{+\infty\}$ such that

$$\varphi(x, 0) = f(x) \qquad \text{for any} \quad x \in R^n,$$

and a family of perturbed problems

$$(\mathrm{P_u}) \qquad \underset{x}{\text{minimize}} \quad \varphi(x, u)$$

or

$$(\mathrm{P_u}) \quad \text{find} \quad \hat{x} \in R^n \qquad \text{such that}$$
$$\varphi(\hat{x}, u) \in \operatorname{Min}\{\varphi(x, u) \in R^p : x \in R^n\}.$$

Let φ^* be the conjugate map of φ; i.e.,

$$\varphi^*(T, \Lambda) = \operatorname{Max}\{Tx + \Lambda u - \varphi(x, u) \in R^p : x \in R^n,\ u \in R^m\}.$$

We consider the following problem as the dual problem of the problem (P) with respect to the given perturbations:

$$\text{(D)} \quad \text{find} \quad \hat{\Lambda} \in R^{p \times m} \qquad \text{such that}$$
$$-\varphi^*(0, \hat{\Lambda}) \cap \operatorname{Max} \bigcup_{\Lambda} -\varphi^*(0, \Lambda) \neq \emptyset.$$

This problem may be written formally as

$$\underset{\Lambda}{\text{maximize}} \quad -\varphi^*(0, \Lambda).$$

However, it is not an ordinary multiobjective optimization problem, since $-\varphi^*(0, \Lambda)$ is not a function but rather a point-to-set map. The set

Max $\bigcup_\Lambda -\varphi^*(0, \Lambda)$ will be simply denoted by Max(D) and every $\hat{\Lambda}$ satisfying the above relationship will be called a solution of the problem (D).

The first result is the so-called weak duality theorem.

Proposition 6.1.12

For any $x \in R^n$ and $\Lambda \in R^{p\times m}$,

$$\varphi(x, 0) \notin -\varphi^*(0, \Lambda) - R^p_+\backslash\{0\}.$$

Proof Let $y = \varphi(x, 0)$ and $y' \in \varphi^*(0, \Lambda)$. Then, from Proposition 6.1.3,

$$y + y' \not\leq 0x + \Lambda 0 = 0,$$

as was to be proved.

Corollary 6.1.4

For any $y \in$ Min(P) and $y' \in$ Max(D), $y \not\leq y'$.

Proof Immediate from Proposition 6.1.12.

We now define the perturbation map for (P), which is an extension of the perturbation function or the optimal value function in ordinary scalar optimization. Let W be a point-to-set map from R^m into R^p defined by

$$W(u) = \text{Min}\{\varphi(x, u) \in R^p : x \in R^n\}.$$

Clearly,

$$\text{Min(P)} = W(0).$$

Lemma 6.1.2

If the function φ is convex on $R^n \times R^m$ and $W(u)$ is externally stable (i.e., $\{\varphi(x, u) \in R^p : x \in R^n\} \subset W(u) + R^p_+$) for each $u \in R^m$, then the perturbation map W is convex.

Proof If the assumptions of the lemma hold,

$$\begin{aligned}
\text{epi } W &= \{(u, y) \in R^m \times R^p : y \in W(u) + R^p_+\} \\
&= \{(u, y) \in R^m \times R^p : y \in \{\varphi(x, u) \in R^p : x \in R^n\} + R^p_+\} \\
&= \{(u, y) \in R^m \times R^p : y \geqq \varphi(x, u),\ x \in R^n\}.
\end{aligned}$$

The last set is the image of epi φ under the projection $(x, u, y) \to (u, y)$. Since epi φ is convex and convexity is preserved under projection, epi W is convex. This completes the proof of the lemma.

We can consider the conjugate map W^* of W, which is directly connected with φ^* as shown in the following lemma.

Lemma 6.1.3

The following relationship holds for every $\Lambda \in R^{p \times m}$:

$$W^*(\Lambda) \supset \varphi^*(0, \Lambda)$$

with the equality holding when every $W(u)$ is externally stable.

Proof

$$\begin{aligned} W^*(\Lambda) &= \operatorname{Max} \bigcup_u [\Lambda u - W(u)] \\ &= \operatorname{Max} \bigcup_u [0x + \Lambda u - \operatorname{Min}\{\varphi(x, u) \in R^p : x \in R^n\}]. \end{aligned}$$

Hence, in view of Corollary 6.1.2,

$$\varphi^*(0, \Lambda) \subset W^*(\Lambda)$$

with the equality holding when every $W(u)$ is externally stable. This completes the proof.

Throughout this chapter we assume that $W(u)$ is externally stable for each $u \in R^m$. Thus we can rewrite the dual problem (D) as follows:

(D) find $\hat{\Lambda} \in R^{p \times m}$ such that

$$-W^*(\hat{\Lambda}) \cap \operatorname{Max} \bigcup_{\Lambda} [-W^*(\Lambda)] \neq \varnothing.$$

Lemma 6.1.4

$$\operatorname{Max}(D) = \operatorname{Max} \bigcup_{\Lambda} [-W^*(\Lambda)] = W^{**}(0).$$

Proof Immediate from Lemma 6.1.3 and the definition of W^{**}.

Thus,

$$\operatorname{Min}(P) = W(0) \qquad \text{and} \qquad \operatorname{Max}(D) = W^{**}(0).$$

Therefore, the discussion on the duality [i.e., the relationship between Min(P) and Max(D)] can be replaced by the discussion on the relationship between $W(0)$ and $W^{**}(0)$. Proposition 6.1.6 justifies considering the following class of multiobjective optimization problems.

Definition 6.1.3 (*Stable Problem*)

The multiobjective optimization problem (P) is said to be stable if the perturbation map W is subdifferentiable at 0.

Proposition 6.1.13

If the function φ is convex, $\{\varphi(x, u) \in R^p : x \in R^n\} \neq \emptyset$ for every u in some neighborhood of 0, and $W^*(\Lambda)$ can be characterized completely by scalarization for each Λ, then problem (P) is stable. Here we should note that we may take a vector subgradient λ as pointed out in Remark 6.1.4.

Proof Consider the point-to-set map Y from R^m to R^p defined by

$$Y(u) = \{\varphi(x, u) \in R^p : x \in R^n\}.$$

From the assumptions, Y is a convex point-to-set map and $Y(u) \neq \emptyset$ in some neighborhood of 0. Noting that $W(u) \subset Y(u)$ for every u, we can prove the proposition in a manner quite similar to the proof of Proposition 6.1.7.

In view of Proposition 6.1.6(ii), the problem (P) is stable if and only if

$$\mathrm{Min(P)} = W(0) \subset W^{**}(0) = \mathrm{Max(D)}.$$

Thus we have the following theorem, which is a generalization of the strong duality theorem in scalar optimization.

Theorem 6.1.1

(i) The problem (P) is stable if and only if, for each solution $\hat{x}$ of (P), there exists a solution $\hat{\Lambda}$ of the dual problem (D) such that

$$\varphi(\hat{x}, 0) \in -\varphi^*(0, \hat{\Lambda}), \qquad \text{i.e.,} \qquad f(\hat{x}) \in -W^*(\hat{\Lambda}),$$

or equivalently

$$(0, \hat{\Lambda}) \in \partial\varphi(\hat{x}, 0), \qquad \text{i.e.,} \qquad \hat{\Lambda} \in \partial W(0; f(\hat{x})).$$

(ii) Conversely, if $\hat{x} \in R^n$ and $\hat{\Lambda} \in R^{p \times m}$ satisfy the above relationship, then $\hat{x}$ is a solution of (P) and $\hat{\Lambda}$ is a solution of (D).

Proof These results are obvious from the previous discussions.

6.1.4 Lagrangians and Saddle Points

In this subsection we define the Lagrangian and its saddle points for Problem (P) and investigate their properties.

Definition 6.1.4 (*Lagrangian*)

The point-to-set map $L: R^n \times R^{p \times m} \to R^p$, defined by

$$-L(x, \Lambda) = \operatorname{Max}\{\Lambda u - \varphi(x, u) \in R^p : u \in R^m\},$$

i.e.,

$$L(x, \Lambda) = \operatorname{Min}\{\varphi(x, u) - \Lambda u \in R^p : u \in R^m\}$$

is called the Lagrangian of the multiobjective optimization problem (P) relative to the given perturbations.

We can write

$$L(x, \Lambda) = -\varphi_x^*(\Lambda),$$

where φ_x denotes the function $u \to \varphi(x, u)$ for a fixed $x \in R^n$, and φ_x^* denotes the conjugate map of φ_x.

Definition 6.1.5 (*Saddle Point*)

A point $(\hat{x}, \hat{\Lambda}) \in R^n \times R^{p \times m}$ is called a saddle point of L if

$$L(\hat{x}, \hat{\Lambda}) \cap \left[\operatorname{Max} \bigcup_{\Lambda} L(\hat{x}, \Lambda)\right] \cap \left[\operatorname{Min} \bigcup_{x} L(x, \hat{\Lambda})\right] \neq \emptyset.$$

It will be useful to express problems (P) and (D) in terms of the map L.

Proposition 6.1.14

For each $\Lambda \in R^{p \times m}$, either $-L(x, \Lambda)$ is externally stable, i.e.,

$$\{\Lambda u - \varphi(x, u) \in R^p : u \in R^m\} \subset -L(x, \Lambda) - R_+^p$$

for every $x \in R^n$, or $\{\Lambda u - \varphi(x, u) \in R^p : u \in R^m\}$ is $(-R_+^p)$-convex $(-R_+^p)$-unbounded or empty for every x. Then

$$\varphi^*(0, \Lambda) = \operatorname{Max} \bigcup_{x} [-L(x, \Lambda)],$$

and, hence, the dual problem (D) can be written formally as

$$\text{(D)} \qquad \operatorname{Max} \bigcup_{\Lambda} \operatorname{Min} \bigcup_{x} L(x, \Lambda).$$

Proof In the former case, in view of Corollary 6.1.3,

$$\begin{aligned}
\operatorname{Max} \bigcup_{x} [-L(x, \Lambda)] &= \operatorname{Max} \bigcup_{x} \operatorname{Max}\{\Lambda u - \varphi(x, u) \in R^p : u \in R^m\} \\
&= \operatorname{Max} \bigcup_{x} \{\Lambda u - \varphi(x, u) \in R^p : u \in R^m\} \\
&= \operatorname{Max}\{0x + \Lambda u - \varphi(x, u) \in R^p : x \in R^n, u \in R^m\} \\
&= \varphi^*(0, \Lambda).
\end{aligned}$$

On the other hand, in the latter case, both $\bigcup_x[-L(x, \Lambda)]$ and $\varphi^*(0, \Lambda)$ are empty. Hence,

$$\operatorname{Max}(D) = \operatorname{Max} \bigcup_{\Lambda}[-\varphi^*(0, \Lambda)] = \operatorname{Max} \bigcup_{\Lambda}\left(-\operatorname{Max} \bigcup_{x}[-L(x, \Lambda)]\right)$$

$$= \operatorname{Max} \bigcup_{\Lambda}\left[\operatorname{Min} \bigcup_{x} L(x, \Lambda)\right].$$

Since $L(x, \Lambda) = -\varphi_x^*(\Lambda)$,

$$\operatorname{Max} \bigcup_{\Lambda} L(x, \Lambda) = \operatorname{Max} \bigcup_{\Lambda}[-\varphi_x^*(\Lambda)] = \varphi_x^{**}(0).$$

Hence, we can directly apply Propositions 6.1.10 and 6.1.11 to obtain the following result.

Proposition 6.1.15

Suppose that $\varphi(x, \cdot)$ is convex for each fixed $x \in R^n$, and that

(i) when $f(x) = \varphi(x, 0)$ is finite, each $\varphi_i(x, \cdot)$ is subdifferentiable at 0, and the set $\{\varphi(x, u) - \Lambda u \in R^p : u \in R^m\}$ is R^p_+-closed for each $\Lambda \in R^{p \times m}$,

(ii) when $f(x) = \varphi(x, 0) = +\infty$, φ_x is closed and subdifferentiable at each $u \in \operatorname{dom} \varphi_x$.

Then in case (i), $\operatorname{Max} \bigcup_{\Lambda} L(x, \Lambda) = \{\varphi_x(0)\} = \{f(x)\}$, and in case (ii), $\operatorname{Max} \bigcup_{\Lambda} L(x, \Lambda) = \varnothing$. Thus the primal problem (P) can be written formally as

$$\text{(P)} \qquad \operatorname{Min} \bigcup_{x} \operatorname{Max} \bigcup_{\Lambda} L(x, \Lambda).$$

These expressions of the problems (P) and (D) suggest that the saddle point of the Lagrangian is closely connected with a pair of solutions to the problems (P) and (D).

Theorem 6.1.2

Suppose that the assumptions both in Propositions 6.1.14 and 6.1.15 are satisfied. Then the following statements are equivalent to each other:

(i) $(\hat{x}, \hat{\Lambda})$ is a saddle point of L,

(ii) $\hat{x}$ is a solution of the problem (P), $\hat{\Lambda}$ is a solution of the problem (D), and

$$\varphi(\hat{x}, 0) \in -\varphi^*(0, \hat{\Lambda}).$$

Proof This theorem is obvious from Propositions 6.1.12, 6.1.14, and 6.1.15.

Theorem 6.1.3

Suppose that the hypothesis of Theorem 6.1.2 holds and that the problem (P) is stable. Then $\hat{x} \in R^n$ is a solution of (P) if and only if there exists $\hat{\Lambda} \in R^{p \times m}$ such that $(\hat{x}, \hat{\Lambda})$ is a saddle point of the Lagrangian L.

Results similar to the above theorems for convex problems can be obtained with the vector dual variable (cf. Remark 6.1.4). The details can be seen in Tanino and Sawaragi [T11].

6.1.5 Duality in Multiobjective Programming

In this subsection we apply the results obtained in the preceding subsections to multiobjective programming. The results will be essentially the same as those obtained in Chapter 5.

We consider a multiobjective programming problem

$$\text{(P)} \qquad \text{minimize} \quad f(x) \qquad \text{subject to} \quad g(x) \leqq 0$$

as a primal problem, where f and g are vector-valued functions on R^n of dimension p and m, respectively. Of course we assume the existence of feasible solutions, i.e., let $X = \{x \in R^n : g(x) \leqq 0\} \neq \emptyset$. If we define a vector-valued function $\tilde{f}: R^n \to R^p \cup \{+\infty\}$ by

$$\tilde{f}(x) = f(x) + \begin{bmatrix} \delta(x \mid X) \\ \vdots \\ \delta(x \mid X) \end{bmatrix},$$

i.e., by

$$\tilde{f}(x) = \begin{cases} f(x) & \text{if} \quad g(x) \leqq 0, \\ +\infty & \text{otherwise,} \end{cases}$$

the problem (P) can be rewritten as

$$\text{(P)} \qquad \text{minimize} \quad \tilde{f}(x).$$

A family of perturbed problems is provided by a function

$$\varphi(x, u) = \begin{cases} f(x) & \text{if} \quad g(x) \leqq u, \\ +\infty & \text{otherwise,} \end{cases}$$

with the m-dimensional parameter vector u. Then the perturbation map W is

$$\begin{aligned} W(u) &= \mathrm{Min}\{\varphi(x, u) \in R^p : x \in R^n\} \\ &= \mathrm{Min}\{f(x) : x \in R^n,\ g(x) \leqq u\}. \end{aligned}$$

In this subsection we use the vector dual variable to define the dual problem and the Lagrangian because we can obtain simpler expressions of them. For $\lambda \in R^m$,

$$-\varphi^*(0, \lambda) = \mathrm{Min}\{\varphi(x, u) - \langle \boldsymbol{\lambda}, \mathbf{u} \rangle \in R^p : x \in R^n,\ u \in R^m\}$$
$$= \mathrm{Min}\{f(x) - \langle \boldsymbol{\lambda}, \mathbf{u} \rangle : x \in R^n,\ u \in R^m,\ g(x) \leqq u\}.$$

Setting $u = g(x) + v$, we have

$$-\varphi^*(0, \lambda) = \mathrm{Min}\{f(x) - \langle \boldsymbol{\lambda}, \mathbf{g}(\mathbf{x}) \rangle - \langle \boldsymbol{\lambda}, \mathbf{v} \rangle : x \in R^n,\ v \in R^m,\ v \geqq 0\}.$$

Hence, unless $\lambda \leqq 0$, $-\varphi^*(0, \lambda) = \varnothing$. Moreover, for $\lambda \leqq 0$,

$$-\varphi^*(0, \lambda) = \mathrm{Min}\{f(x) - \langle \boldsymbol{\lambda}, \mathbf{g}(\mathbf{x}) \rangle : x \in R^n\}.$$

Thus, the dual problem with the vector dual variable can be written as

$$\text{(D)} \qquad \mathrm{Max} \bigcup_{\lambda \leqq 0} \mathrm{Min}\{f(x) - \langle \boldsymbol{\lambda}, \mathbf{g}(\mathbf{x}) \rangle : x \in R^n\}.$$

Theorem 6.1.1 provides the following duality theorem for multiobjective programming problems.

Theorem 6.1.4

(i) The problem (P) is stable if and only if, for each solution $\hat{x}$ of (P), there exists $\hat{\lambda} \in R^m$ with $\hat{\lambda} \leqq 0$ such that $\hat{\lambda}$ is a solution of the dual problem (D),

$$f(\hat{x}) \in \mathrm{Min}\{f(x) - \langle \hat{\boldsymbol{\lambda}}, \mathbf{g}(\mathbf{x}) \rangle : x \in R^n\},$$

and $\langle \hat{\lambda}, g(\hat{x}) \rangle = 0$.

(ii) Conversely, if $\hat{x} \in X$ and $\hat{\lambda} \in R^m$ with $\hat{\lambda} \leqq 0$ satisfy the above conditions, then $\hat{x}$ and $\hat{\lambda}$ are solutions of (P) and (D), respectively.

Proof We prove that $\langle \hat{\lambda}, g(\hat{x}) \rangle = 0$. Since $g(\hat{x}) \leqq 0$ and $\hat{\lambda} \leqq 0$, $\langle \hat{\lambda}, g(\hat{x}) \rangle \geqq 0$. If $\langle \hat{\lambda}, g(\hat{x}) \rangle > 0$,

$$f(\hat{x}) > f(\hat{x}) - \langle \hat{\boldsymbol{\lambda}}, \mathbf{g}(\hat{\mathbf{x}}) \rangle,$$

which contradicts the efficiency of $f(\hat{x})$. Hence $\langle \hat{\lambda}, g(\hat{x}) \rangle = 0$. The remaining proof is obvious.

In this case the stability of the problem (P) is essentially ensured by convexity and the Slater's constraint qualification.

Proposition 6.1.16

In each function f_i $(i = 1, \ldots, p)$ and g_j $(j = 1, \ldots, m)$ is convex, if $W^*(\lambda)$ can be characterized completely by scalarization for each λ (for example, if every f_i is strictly convex), and if the Slater's constraint qualification is satisfied (i.e., there exists $\bar{x} \in X$ such that $g(\bar{x}) < 0$), then the problem (P) is stable.

Proof We can directly apply Proposition 6.1.13.

Next, we compute the Lagrangian with the vector dual variable:

$$L(x, \lambda) = \operatorname{Min}\{\varphi(x, u) - \langle \boldsymbol{\lambda}, \mathbf{u} \rangle \in R^p : u \in R^m\}$$
$$= \operatorname{Min}\{f(x) - \langle \boldsymbol{\lambda}, \mathbf{u} \rangle : u \in R^m,\ g(x) \leqq u\}.$$

It is clear that $L(x, \lambda) = \varnothing$ for all x unless $\lambda \leqq 0$. On the other hand, when $\lambda \leqq 0$,

$$L(x, \lambda) = \{f(x) - \langle \boldsymbol{\lambda}, \mathbf{g}(\mathbf{x}) \rangle\} \qquad \text{for all} \quad x.$$

Thus,

$$-\varphi^*(0, \lambda) = \operatorname{Min} \bigcup_x L(x, \lambda).$$

In this case, a point $(\hat{x}, \hat{\lambda})$ is a saddle point of L if and only if $\hat{\lambda} \leqq 0$ and

$$f(\hat{x}) - \langle \hat{\boldsymbol{\lambda}}, \mathbf{g}(\hat{\mathbf{x}}) \rangle \in \left[\operatorname{Min} \bigcup_x L(x, \hat{\lambda})\right] \cap \left[\operatorname{Max} \bigcup_\lambda L(\hat{x}, \lambda)\right].$$

The relationship

$$f(\hat{x}) - \langle \hat{\boldsymbol{\lambda}}, \mathbf{g}(\hat{\mathbf{x}}) \rangle \in \operatorname{Max} \bigcup_\lambda L(x, \lambda)$$

is equivalent to the inequality

$$\langle \hat{\lambda}, g(\hat{x}) \rangle \leqq \langle \lambda, g(\hat{x}) \rangle \qquad \text{for all} \quad \lambda \in R^m, \quad \lambda \leqq 0.$$

Thus we have the following theorem.

Theorem 6.1.5

Suppose that the problem (P) is stable. Then $\hat{x} \in R^n$ is a solution of (P) if and only if there exists $\hat{\lambda} \in R^m$ such that $(\hat{x}, \hat{\lambda})$ is a saddle point of L; i.e.,

(i) $f(\hat{x}) - \langle \hat{\boldsymbol{\lambda}}, \mathbf{g}(\hat{\mathbf{x}}) \rangle \in \operatorname{Min}\{f(x) - \langle \hat{\boldsymbol{\lambda}}, \mathbf{g}(\mathbf{x}) \rangle : x \in R^n\}$,
(ii) $\hat{\lambda} \leqq 0$,
(iii) $\langle \hat{\lambda}, g(\hat{x}) \rangle \leqq \langle \lambda, g(\hat{x}) \rangle$ for all $\lambda \in R^m$, $\lambda \leqq 0$.

6.1.6 Infimum and Supremum Based on Efficiency

The concept of infimum and supremum for sets in R has been extended to that for sets in R^p in some ways. They can be categorized in three types:

(i) extension based on lower and upper bound (Brumelle [B21]),
(ii) extension based on efficiency (Gros [G11] and Ponstein [P4]),

(iii) extension based on weak efficiency (Nieuwenhuis [N15] and Kawasaki [K2]).

The infimum of a set Y in R is defined as the greatest lower bound, namely, $\hat{y} = \inf Y$ if and only if

(1) $\hat{y} \leqq y$ for any $y \in Y$; and
(2) if $y' \leqq y$ for any $y \in Y$, then $y' \leqq \hat{y}$.

This definition can be extended directly to a set in R^p. It is well known in mathematics and used by Zowe [Z9, Z10] and Brumelle [B21] (see Section 6.3.). Moreover, since $\leqq = \ngeq = \not> $ in R, we may replace $\leqq$ in the above conditions with $\ngeq$ or $\not>$ to define the infimum of a set Y in R^p. The definition given by Nieuwenhuis [N15] corresponds to the case in which $\not>$ is used. Furthermore, the infimum of $Y \subset R$ has the property that

$$\inf Y = \min(\mathrm{cl}(Y + R)).$$

Hence, we can use this formula in order to define the infimum of $Y \subset R^p$. The definitions given by Gros [G11] and Kawasaki [K2] correspond to the cases in which minimum and weak minimum, respectively, are used. (See this subsection and Section 6.2.)

We will discuss (i) and (iii) in Section 6.3 and Section 6.2, respectively. In this subsection, we touch on (ii), namely the definitions of infimum and supremum in R^p by Gros [G11] and Ponstein [P4].

Gros defined infimum and supremum of a set Y in R^p by

$$\mathrm{Inf}\, Y = \mathrm{Min}(\mathrm{cl}(Y + R^p_+ \setminus \{0\})) = \mathrm{Min}(\mathrm{cl}(Y + R^p_+)),$$

$$\mathrm{Sup}\, Y = \mathrm{Max}(\mathrm{cl}(Y - R^p_+ \setminus \{0\})) = \mathrm{Max}(\mathrm{cl}(Y - R^p_+)).$$

When Y is R^p_+ (resp. $-R^p_+$)-closed, $\mathrm{Inf}\, Y = \mathrm{Min}\, Y$ (resp. $\mathrm{Sup}\, Y = \mathrm{Max}\, Y$). This definition is intuitive. However, if we adopt it to define the conjugate map, we cannot claim that $f(0) \in f^{**}(0)$ even in very simple cases. (See Example 6.1.2 later.) Gros himself defined the conjugate with the help of scalarization, and so the conjugate obtained is quite similar to that given by Kawasaki (Definition 6.2.3).

Ponstein defined infimum for a set Y in $\bar{R}^p$. He extended the convergence and closedness in R^p to those in $\bar{R}^p$ as follows:

(i) A sequence $\{y^k\} \subset R^p$ converges to some $y \in \bar{R}^p$ if

$$\lim_{k \to \infty} y_i^k = y_i, \qquad i = 1, \dots, p.$$

This definition is extended to sequences $\{y^k\} \subset \bar{R}^p$ such that, for each $i = 1, \ldots, p$, either y_i^k is finite for all k, or $y_i^k = +\infty$ for all k, or $y_i^k = -\infty$ for all k.

(ii) A set $Y \subset \bar{R}^p$ is closed if $y \in Y$ whenever $y^k \in Y$ and $\lim_{k\to\infty} y^k = y$.

(iii) The closure of Y is the intersection of all closed sets containing Y and is also denoted by cl Y.

Ponstein defined Inf Y as

$$\text{Inf } Y = \begin{cases} \{+\infty\} & \text{if} \quad Y = \emptyset \\ \text{Min(cl } Y) & \text{if} \quad Y \neq \emptyset, \end{cases}$$

where Min implies efficiency with respect to the order $\leqq$ naturally induced in $\bar{R}^p$. He proved external stability of Inf Y and interchangeability of two Inf's and also dualized multiobjective optimization problems.

Example 6.1.1

Let

$$Y = \{y = (y_1, y_2) : y_1 \leqq 0, y_2 > 0\} \cup \{0\} \subset R^2 \subset \bar{R}^2.$$

Then

(i) Min $Y = \{0\}$,

(ii) Inf $Y = \emptyset$ by Gros's definition,

(iii) Inf $Y = \{(-\infty, 0)\}$ by Ponstein's definition.

Unfortunately both definitions of infimum by Gros and Ponstein might not be adequate for defining the conjugate as is shown in the following example.

Example 6.1.2

Let

$$f(x) = \begin{bmatrix} x \\ -x \end{bmatrix} \quad \text{for} \quad x \in R.$$

Then,

$$\text{for} \quad T = \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \in R^{2\times 1}, \qquad -f^*(T) = \text{Inf}\left\{\begin{bmatrix} (1-t_1)x \\ -(1+t_2)x \end{bmatrix} : x \in R\right\}.$$

We can easily check that

$$\bigcup_T (-f^*(T)) =$$

$$\begin{cases} (R^2 \backslash (R^2_+ \cup R^2_-)) \cup \{0\} & \text{by Gros's definition} \\ (R^2 \backslash (R^2_+ \cup R^2_-)) \cup \left\{\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -\infty \end{bmatrix}, \begin{bmatrix} -\infty \\ 0 \end{bmatrix}, \begin{bmatrix} -\infty \\ -\infty \end{bmatrix}, \begin{bmatrix} -\infty \\ +\infty \end{bmatrix}, \begin{bmatrix} +\infty \\ -\infty \end{bmatrix}\right\} & \\ & \text{by Ponstein's definition.} \end{cases}$$

Hence,

$$f^{**}(0) = \begin{cases} \varnothing & \text{(Gros)} \\ \left\{ \begin{bmatrix} 0 \\ +\infty \end{bmatrix}, \begin{bmatrix} +\infty \\ 0 \end{bmatrix} \right\} & \text{(Ponstein),} \end{cases}$$

which does not contain

$$f(0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

6.2 Conjugate Duality Based on Weak Efficiency

As has been shown in the previous section, the duality result is not always reflexive if we use the vector dual variable (i.e., $W(0) \subsetneq W^{**}(0)$, generally). Recently Kawasaki [K2, K3] gave some new notions and some fundamental tools and results needed for developing the reflexive duality theory of multiobjective optimization problems. His approach is based on weak efficiency instead of usual efficiency. Weak efficiency is easier to deal with than efficiency, since it is directly related to scalarization. (See Corollary 3.4.1.) Moreover, his primal problems are also defined by vector-valued point-to-set maps (called relations in his papers).

The above two facts are undesirable from the practical point of view in multiobjective optimization. However, his results can enrich the conjugate duality theory and so are well worth paying attention to. Hence, in this section, we overview Kawasaki's conjugate duality theory. In order to avoid making this section longer than needed, we omit all the proofs of theorems, propositions, and lemmas. If the reader is interested, please see the original papers. We should also note that some notations and terms in this section are different from the original ones. Particularly, we often add the modifier *weak* not to confuse the concepts in Sections 6.1 and 6.2.

6.2.1 Weak Supremum and Infimum

In this subsection, the definition and some fundamental properties of weak supremum and infimum will be given. The p-dimensional point consisting of a common element $\alpha \in R$ is simply denoted by $\boldsymbol{\alpha}$. Let

$$R^p_\infty = R^p \cup \{\pm\infty\}.$$

Of, course, $-\infty < y < +\infty$ for all $y \in R^p$. Let

$$K = \operatorname{int} R^p \cup \{0\}.$$

For any $y \in R^p$, $y \pm \infty$ is defined by

$$y \pm \infty = \pm\infty.$$

The sum $\infty + (-\infty)$ is left undefined, since we can avoid it.

Definition 6.2.1 *(Weak Maximal (Minimal) Set)*

Let Y be a subset of R^p_∞. The set of efficient points of Y with respect to the ordering $>$ (resp. $<$) is denoted by w-Max Y (resp. w-Min Y), which will be called the weak maximal (resp. minimal) set of Y. Namely,

$$\text{w-Max } Y \text{ (resp. w-Min } Y) = \{\hat{y} \in Y : \text{there is no } y \in Y \text{ such that } \hat{y} < y \text{ (resp. } y < \hat{y})\}.$$

Definition 6.2.2 *(Weak Supremum (Infimum) Set)*

Let Y be a subset of R^p_∞. We shall define $\overline{Y-K}$ as

$$\overline{Y-K} = \begin{cases} R^p_\infty & \text{if } R^p \subset Y-K \text{ or } +\infty \in Y-K \\ \emptyset & \text{if } Y-K = \emptyset \\ \mathrm{cl}[(Y-K) \cap R^p] \cup \{-\infty\} & \text{otherwise.} \end{cases}$$

Furthermore, we shall define w-Sup Y as

$$\text{w-Sup } Y = \text{w-Max } \overline{Y-K},$$

which is called the weak supremum set of Y and its element is called the weak supremum point of Y. By replacing K with $-K$, we can define $\overline{Y+K}$ by

$$\overline{Y+K} = \begin{cases} R^p_\infty & \text{if } R^p \subset Y+K \text{ or } +\infty \in Y+K \\ \emptyset & \text{if } Y+K = \emptyset \\ \mathrm{cl}[(Y+K) \cap R^p] \cup \{+\infty\} & \text{otherwise.} \end{cases}$$

Moreover,

$$\text{w-Inf } Y = \text{w-Min } \overline{Y+K},$$

which is called the weak infimum set of Y and its element is called the weak infimum point of Y.

Of course, when $p = 1$, w-Inf Y coincides with the set $\{\inf Y\}$. (See Fig. 6.1 to compare Min Y, w-Min Y, and w-Inf Y.) Kawasaki's definition is essentially the same as Nieuwenhuis's definition except that the points $\pm\infty$ are added.

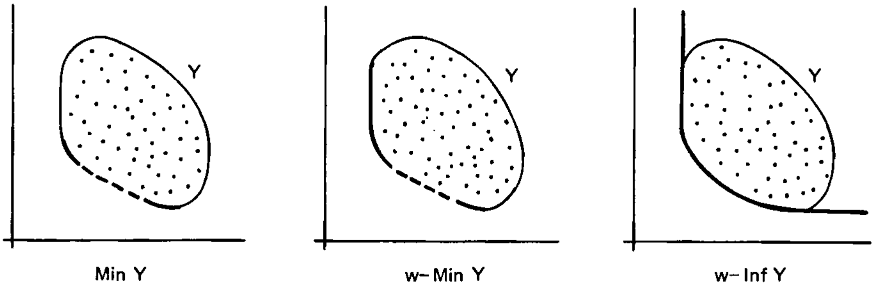

Fig. 6.1.

We have

$$\text{w-inf}\, Y = -\text{w-Sup}(-Y).$$

Let Y and Y' be subsets of R^p_∞. We write $Y \leqslant Y'$ if there are no points $y \in Y$ and $y' \in Y'$ such that $y' < y$. We write $Y \underset{\rightarrow}{\leqslant} Y'$ if $Y \leqslant Y'$ and for any $y \in Y$ there exists some $y' \in Y'$ such that $y < y'$ or $y = y'$. We write $Y \underset{\leftarrow}{\leqslant} Y'$ if $-Y' \underset{\rightarrow}{\leqslant} -Y$. Moreover, we write $Y \underset{\leftrightarrow}{\leqslant} Y'$ if both $Y \underset{\rightarrow}{\leqslant} Y'$ and $Y \underset{\leftarrow}{\leqslant} Y'$ hold.

The following propositions essentially show that w-Sup Y is externally stable not only for the set $\overline{Y - K}$ but also for its complement.

Proposition 6.2.1

Let Y be a subset of R^p. For each $y \in \overline{Y - K}$, there exists a point $y' \in$ w-Sup Y such that $y < y'$ or $y = y'$. In particular, for each $y \in \overline{Y - K} \setminus \{-\infty\}$, there exists a number $t \in [0, +\infty]$ such that $y + \mathbf{t} \in$ w-Sup Y.

Proposition 6.2.2

Let Y be a nonempty subset of R^p_∞. Then, for each $y \notin \overline{Y - K}$, there exists a point $y' \in$ w-Sup Y such that $y' < y$. In particular, if $y \neq \infty$, then there exists a number $t \in [-\infty, 0)$ such that $y + \mathbf{t} \in$ w-Sup Y.

Corollary 6.2.1

Let Y be a nonempty subset of R^p_∞. If w-Sup $Y \neq \pm\{\infty\}$, then

$$(\text{w-Sup}\, Y + \text{int}\, R^p_+)^c = \overline{Y - K} \qquad \text{in} \quad R^p_\infty.$$

Corollary 6.2.2

Let $Y_1, Y_2 \subset R^p_\infty$. If $Y_1 - K \subset Y_2 - K$, then w-Sup $Y_1 \underset{\rightarrow}{\leqslant}$ w-Sup Y_2. Moreover, if Y_1 is not empty, then w-Sup $Y_1 \underset{\leftrightarrow}{\leqslant}$ w-Sup Y_2.

Lemma 6.2.1

Let Y be a subset of R^p_∞. Then we have

$$\text{w-Inf}(\text{w-Sup } Y) = \text{w-Sup } Y,$$

$$\text{w-Sup}(\text{w-Inf } Y) = \text{w-Inf } Y.$$

6.2.2 Weak Conjugate Maps and Γ^p-Regularization

This subsection is devoted to weak conjugate maps and Γ^p-regularization of vector-valued point-to-set maps. The definition of a new type of conjugate maps is based on the weak supremum instead of the usual efficiency.

Definition 6.2.3 (*Weak Conjugate Map and Weak Biconjugate Map*)†

Let F be a point-to-set map from R^n into R^p_∞. The weak conjugate map w-F^*: $R^n \to R^p_\infty$ of F is defined by

$$\text{w-}F^*(x^*) = \text{w-Sup} \bigcup_x [\langle \mathbf{x}^*, \mathbf{x} \rangle - F(x)].$$

Furthermore, the weak biconjugate map w-F^{**}: $R^n \to R^p_\infty$ is defined by

$$\text{w-}F^{**}(x) = \text{w-Sup} \bigcup_{x^*} [\langle \mathbf{x}^*, \mathbf{x} \rangle - \text{w-}F^*(x^*)].$$

Proposition 2.1.26 shows that the conjugate function f^* of a scalar-valued function f is a closed convex function, which is the pointwise supremum of the collection of all affine functions h such that $h \leqq f$ (Proposition 2.1.25). The set of closed convex functions f is often denoted by Γ. This class is extended to vector-valued point-to-set maps.

Definition 6.2.4 (*Class* Γ^p)

The set of point-to-set maps G: $R^n \to R^p_\infty$ that which can be written as

$$G(x) = \text{w-Sup} \bigcup_v (\langle \mathbf{x}_v^*, \mathbf{x} \rangle + a_v)$$

is denoted by $\Gamma^p(R^n)$, or simply by Γ^p, where $x_v^* \in R^n$, $a_v \in R^p_\infty$ for all v and the set of indices v is not empty.

For two point-to-set maps F and G from R^n into R^p_∞, we write $F \preccurlyeq G$ if $F(x) \preccurlyeq G(x)$ for all $x \in R^n$.

† Cf. Definition 6.1.1.

Definition 6.2.5 (Γ^p-*Regularization*)

Let F be a point-to-set map from R^n into R^p_∞. The point-to-set map $\tilde{F}$ defined by

$$\tilde{F}(x) = \text{w-Sup}\left[\bigcup_{\substack{\langle x^*, \cdot \rangle + a \leqslant F \\ x \in R^n,\, a \in R^p_\infty}} (\langle x^*, x \rangle + a) \right]$$

is called the Γ^p-regularization of F.

The following theorem shows the relationship between the Γ^p-regularization and the weak conjugate map.

Theorem 6.2.1

Let F be a point-to-set map from R^n into R^p_∞, and $\tilde{F}$ be the Γ^p-regularization of F. Then

(i) $\tilde{F} \in \Gamma^p$,
(ii) $\tilde{F} \underset{\leftarrow}{\leqslant} F$,
(iii) if $G \in \Gamma^p$ and $G \leqslant F$, then $G \underset{\leftrightarrow}{\leqslant} \tilde{F}$.
(iv) $\tilde{F} = F$ holds if and only if $F \in \Gamma^p$,
(v) $\tilde{F} = \text{w-}F^{**}$, if $F(x)$ is not identically empty.

Next, we shall give a characterization of Γ^p.

Definition 6.2.6 (*Scalarizable Point-To-Set Map*)

Let F be a point-to-set map from R^n into R^p_∞. We call F scalarizable if F satisfies

$$\text{w-Sup}\, F(x) = F(x) \neq \varnothing \qquad \text{for all} \quad x \in R^n.$$

Remark 6.2.1

By Lemma 6.2.1, F is scalarizable if and only if

$$\text{w-Inf}\, F(x) = F(x) \neq \varnothing \qquad \text{for all} \quad x \in R^n.$$

Definition 6.2.7 (*Scalarized Function*)

Let F be a scalarizable point-to-set map from R^n into R^p_∞, and let a be an arbitrary point of R^p. Let Δ be the diagonal set of R^p, and let $\Delta_\infty = \Delta \cup \{\pm\infty\}$. Namely

$$\Delta = \{\boldsymbol{\alpha} \in R^p : \alpha \in R\} \qquad \text{and} \qquad \Delta_\infty = \{\boldsymbol{\alpha} : \alpha \in \bar{R}\}.$$

Then, by Propositions 6.2.1 and 6.2.2, the set $(a + \Delta_\infty) \cap F(x)$ is a singleton, and its element can be expressed uniquely as $a + F_a(x)$, where $F_a(x) \in \bar{R}$. Thus we can define the function $F_a: R^n \to \overline{R}$, which is called the scalarized function of F through a.

We can characterize Γ^p in terms of scalarizable point-to-set maps and scalarized functions.

Theorem 6.2.2

Let F be a point-to-set map from R^n into R^p_∞. Then F is contained in Γ^p if and only if F is scalarizable and all the scalarized functions of F are contained in Γ^1.

Proposition 6.2.3

For any scalarizable point-to-set map F,

$$(F_a)^* = (\text{w-}F^*)_{-a} \qquad \text{for all} \quad a \in R^p,$$

$$(F_a)^{**} = (\text{w-}F^{**})_a \qquad \text{for all} \quad a \in R^p.$$

6.2.3 Weak Subdifferentials of Point-To-Set Maps

In this subsection we shall define weak subgradients of vector-valued point-to-set maps.

Definition 6.2.8 (*Weak Subgradient and Weak Subdifferential*)†

Let F be a point-to-set map from R^n into R^p_∞. F is said to be weakly subdifferentiable at $(x; y) \in R^n \times R^p_\infty$ if $y \in F(x) \cap R^p$ and there exists some $x^* \in R^n$ such that

$$\langle \mathbf{x}^*, \mathbf{x}' - \mathbf{x} \rangle + y \not< F(x') \qquad \text{for all} \quad x' \in R^n.$$

We call x^* a weak subgradient of F at $(x; y)$. The set of all weak subgradients of F at $(x; y)$ is called the weak subdifferential of F at $(x; y)$ and is denoted by w-$\partial F(x; y)$. Let

$$\text{w-}\partial F(x) = \bigcup_{y \in F(x)} \text{w-}\partial F(x; y).$$

Lemma 6.2.2

Let F be a point-to-set map from R^n into R^p_∞. Then

(i) w-$\partial F(x; y) \subset$ w-∂(w-F^{**})$(x; y)$ for all $x \in R^n$ and $y \in R^p$,

(ii) If w-$F^{**}(x) \subset F(x)$, then w-$\partial F(x; y) =$ w-∂(w-F^{**})$(x; y)$ for all $y \in$ w-$F^{**}(x)$,

† Cf. Definition 6.1.2.

(iii) y is contained in $F(x) \cap \text{w-Min}[\bigcup_{x'} F(x')] \cap R^p$ if and only if $0 \in \text{w-}\partial F(x; y)$,

(iv) $x^* \in \text{w-}\partial F(x; y)$ if and only if $\langle x^*, x\rangle \in y + F^*(x^*)$ and $y \in F(x) \cap R^p$,

(v) if $x^* \in \text{w-}\partial F(x; y)$, then $x \in \text{w-}\partial(\text{w-}F^*)(x^*; \langle \mathbf{x}^*, \mathbf{x}\rangle - y)$. In particular, if $F \in \Gamma^p$, then the above two conditions are equivalent to each other.

Corollary 6.2.3

Let F be a point-to-set map from R^n into R^p_∞. Then

(i) $\text{w-}\partial F(x) \subset \text{w-}\partial(\text{w-}F^{**})(x)$ for all $x \in R^n$,
(ii) if $\text{w-}F^{**}(x) \subset F(x)$, then $\text{w-}\partial F(x) = \text{w-}\partial F^{**}(x)$,
(iii) $F(x) \cap \text{w-Min}[\bigcup_{x'} F(x')] \cap R^p \neq \emptyset$ if and only if $0 \in \text{w-}\partial F(x)$,
(iv) $x^* \in \text{w-}\partial F(x)$ if and only if $\langle \mathbf{x}^*, \mathbf{x}\rangle \in F(x) + \text{w-}F^*(x^*)$,
(v) if $x^* \in \text{w-}\partial F(x)$, then $x \in \text{w-}\partial(\text{w-}F^*)(x^*)$. In particular, if $F \in \Gamma^p$, then the above two conditions are equivalent to each other.

We can obtain the relationship between weak subdifferentials of scalarizable point-to-set maps and subdifferentials of scalarized functions.

Proposition 6.2.4

Let F be a scalarizable point-to-set map from R^n into R^p_∞. For any $x \in R^n$ and $y \in F(x) \cap R^p$, we have

$$\text{w-}\partial F(x; y) = \partial F_y(x).$$

6.2.4 Duality

In this subsection, by using the results obtained in the previous subsections, we provide duality results similar to those in Section 6.1. First, let Φ be a point-to-set map from $R^n \times R^m$ into R^p_∞. We assume that it is not identically empty on $R^n \times \{0\}$.

The primal problem can be written as

(P) find $\hat{x} \in R^n$ such that

$$\Phi(\hat{x}, 0) \cap \text{w-Min}\left[\bigcup_x \Phi(x, 0)\right] \cap R^p \neq \emptyset.$$

This problem can be equivalently rewritten as

(P) find $\hat{x} \in R^n$ such that

$$\Phi(\hat{x}, 0) \cap \text{w-Inf}\left[\bigcup_x \Phi(x, 0)\right] \cap R^p \neq \emptyset.$$

This problem can be imbedded into a family of perturbed problems

$(\mathrm{P_u})$ find $\hat{x} \in R^n$ such that

$$\Phi(\hat{x}, u) \cap \text{w-Inf}\left[\bigcup_x \Phi(x, u)\right] \cap R^p \neq \varnothing.$$

The dual problem and the family of perturbed dual problems are given as

(D) find $\hat{\lambda} \in R^m$ such that

$$-\text{w-}\Phi^*(0, \hat{\lambda}) \cap \text{w-Sup}\left[\bigcup_\lambda -\text{w-}\Phi^*(0, \lambda)\right] \cap R^p \neq \varnothing,$$

(D_μ) find $\hat{\lambda} \in R^m$ such that

$$-\text{w-}\Phi^*(\mu, \hat{\lambda}) \cap \text{w-Sup}\left[\bigcup_\lambda -\text{w-}\Phi^*(\mu, \lambda)\right] \cap R^p \neq \varnothing.$$

Define point-to-set maps $W: R^m \to R^p$ (which is slightly different from W in Section 6.1 though the same symbol is used) and $V: R^n \to R^p$ by

$$W(u) = \text{w-Inf}\left[\bigcup_x \Phi(x, u)\right] \quad \text{and} \quad V(\mu) = -\text{w-Sup}\left[\bigcup_\lambda -\text{w-}\Phi^*(\mu, \lambda)\right].$$

We denote $W(0)$ and $-V(0)$ by w-Inf P and w-Sup D, respectively. The following lemma is analogous to Lemma 6.1.3.

Lemma 6.2.3

(i) $\text{w-}W^*(\lambda) = \text{w-}\Phi^*(0, \lambda)$ for all $\lambda \in R^m$,

(ii) $\text{w-}V^*(x) = \text{w-}\Phi^{**}(x, 0)$ for all $x \in R^n$. In particular, if $\Phi \in \Gamma^p(R^n \times R^m)$, then $\text{w-}V^*(x) = \Phi(x, 0)$ for all $x \in R^n$.

Lemma 6.2.4†

(i) $\text{w-Sup(D)} = \text{w-}W^{**}(0)$,

(ii) if $\Phi \in \Gamma^p(R^n \times R^m)$, then $\text{w-Inf(P)} = -\text{w-}V^{**}(0)$.

The weak duality theorem can be strengthened to the following.

Proposition 6.2.5‡

$$\text{w-Sup(D)} \underset{\leftrightarrow}{\leqslant} \text{w-Inf(P)}.$$

Convexity of Φ and W are related via their scalarized functions.

† Cf. Lemma 6.1.4.

‡ Cf. Corollary 6.1.4.

Lemma 6.2.5[†]

If Φ is scalarizable and every scalarized function of Φ is convex, then every scalarized function of W is also convex.

Stability of the primal (resp. dual) problem can be defined in terms of scalarized functions of W (resp. V). We also introduce a new concept normality extended from the case of scalar optimization.

Definition 6.2.9 *(Normality)*[‡]

If every scalarized function of W (resp. V) is lower semicontinuous and finite at 0, then the primal (resp. dual) problem (P) (resp. (D)) is said to be normal.

Definition 6.2.10 *(Weak Stability)*[§]

If every scalarized function of W (resp. V) is weakly subdifferentiable at 0, then the primal (resp. dual) problem (P) (resp. (D)) is said to be weakly stable.

Proposition 6.2.6

Suppose that $\Phi \in \Gamma^p$, that w-Inf(P) is a nonempty subset of R^p, and that for each $a \in R^p$ there exists $x_a \in R^n$ such that $\Phi_a(x_a, \cdot)$ is finite and continuous at 0. Then (P) is weakly stable.

Proposition 6.2.7[‖]

If $\Phi \in \Gamma^p$, then the following three conditions are equivalent to each other:

(i) (P) is normal,
(ii) (D) is normal,
(iii) w-Inf(P) = w-Sup(D) and this set is a nonempty subset of R^p.

Proposition 6.2.8

(i) For any $\lambda \in R^m$ and $y \in R^p_\infty$, $y \in -\text{w-}\Phi^*(0, \lambda) \cap \text{w-Sup(D)} \cap R^p$ if and only if $\lambda \in \text{w-}\partial(\text{w-}W^{**})(0; y)$.

(ii) Suppose that $\Phi \in \Gamma^p$. For any $x \in R^n$ and $y \in R^p_\infty$, $y \in \Phi(x, 0) \cap \text{w-Inf(P)} \cap R^p$ if and only if $x \in \text{w-}\partial(\text{w-}V^{**})(0; -y)$.

[†] Cf. Lemma 6.1.2.
[‡] Cf. Ekeland and Temam [E1, Definition III.2.1].
[§] Cf. Definition 6.1.3.
[‖] Cf. Ekeland and Temam [E1, Proposition III.2.1].

Proposition 6.2.9

If $\Phi \in \Gamma^p$, the following three conditions are equivalent to each other:

(i) (P) is weakly stable,
(ii) $W(0)$ is a nonempty subset of R^p, and W is weakly subdifferentiable at $(0; y)$ for any $y \in W(0)$.
(iii) w-Inf(P) = w-Sup(D) = w-Max(D), and this is a nonempty subset of R^p, where w-Max(D) = w-Max$[\bigcup_\lambda -\text{w-}\Phi^*(0, \lambda) - K]$.

The main strong duality theorem is given as follows.

Theorem 6.2.3[†]

Suppose that $\Phi \in \Gamma^p$ and that the primal problem (P) is weakly stable. Then the dual problem (D) has at least one solution. If $\hat{x} \in R^n$ is a solution of (P), then there exists a solution $\hat{\lambda}$ of (D) such that

$$\Phi(\hat{x}, 0) \cap -\text{w-}\Phi^*(0, \hat{\lambda}) \neq \emptyset$$

or equivalently

$$(0, \hat{\lambda}) \in \text{w-}\partial\Phi(\hat{x}, 0).$$

Conversely, if $\hat{x} \in R^n$ and $\hat{\lambda} \in R^m$ satisfy the above relationship, then $\hat{x}$ is a solution of (P), $\hat{\lambda}$ is a solution of (D) and

$$\Phi(\hat{x}, 0) \cap -\text{w-}\Phi^*(0, \hat{\lambda}) \subset \text{w-Inf(P)} \cap \text{w-Sup(D)} \cap R^p.$$

Remark 6.2.2

The above duality theorem satisfies a reflexivity. In fact, we can exchange the roles of (P) and (D) in Theorem 6.2.3.

As a matter of fact, as is shown in the following proposition, the primal problem (P) can be reduced to ordinary scalar optimization problem when Φ is scalarizable.

Proposition 6.2.10

If Φ is scalarizable, then $\hat{x} \in R^n$ is a solution of (P) if and only if there exists $a \in R^p$ such that $\hat{x}$ is a solution of the following scalar optimization problem $(\mathrm{P_a})$:

$(\mathrm{P_a})$ minimize $\Phi_a(x, 0)$.

[†] Cf. Theorem 6.1.1.

6.2.5 Weak Lagrangians and Weak Saddle Points

In this subsection, we shall introduce a weak Lagrangian of the primal problem (P) relative to the given perturbations Φ and clarify the relationship between the pairs of solutions of (P) and (D) and weak saddle points of the weak Lagrangian.

Definition 6.2.11 (*Weak Lagrangian*)[†]

The point-to-set map w-L: $R^n \times R^m \to R^p_\infty$, defined by

$$\text{w-}L(x, \lambda) = \text{w-Inf}\left[\bigcup_u (\Phi(x, u) - \langle \boldsymbol{\lambda}, \mathbf{u} \rangle)\right] \qquad \text{for} \quad x \in R^n, \quad \lambda \in R^m$$

is called the weak Lagrangian of the primal problem (P) relative to the given perturbations.

Definition 6.2.12 (*Weak Saddle Point*)[‡]

A point $(\hat{x}, \hat{\lambda}) \in R^n \times R^m$ is called a weak saddle point of the weak Lagrangian w-L, if there exists $\hat{y} \in \text{w-}L(\hat{x}. \hat{\lambda}) \cap R^p$ such that

$$\text{w-}L(\hat{x}, \lambda) \leqslant \hat{y} \leqslant \text{w-}L(x, \hat{\lambda}) \qquad \text{for all} \quad x \in R^n \qquad \text{and} \qquad \lambda \in R^m.$$

Let

$$\text{w-Inf w-Sup w-}L = \text{w-Inf} \bigcup_x \left[\text{w-Sup} \bigcup_\lambda \text{w-}L(x, \lambda)\right],$$

$$\text{w-Sup w-Inf w-}L = \text{w-Sup} \bigcup_\lambda \left[\text{w-Inf} \bigcup_x \text{w-}L(x, \lambda)\right].$$

Proposition 6.2.11

$$\text{w-Sup w-Inf w-}L \underset{\leftrightarrow}{\leqslant} \text{w-Inf w-Sup w-}L.$$

Proposition 6.2.12

The following two conditions are equivalent to each other:

(i) $(\hat{x}, \hat{\lambda})$ is a weak saddle point of w-L and $\hat{y} \in L(\hat{x}, \hat{\lambda})$ satisfies w-$L(\hat{x}, \lambda) \leqslant \hat{y} \leqslant$ w-$L(x, \hat{\lambda})$ for all $x \in R^n$ and $\lambda \in R^m$,

(ii) $\hat{y} \in \text{w-Inf}[\bigcup_x \text{w-}L(x, \hat{\lambda})] \cap \text{w-Sup}[\bigcup_\lambda \text{w-}L(\hat{x}, \lambda)] \cap R^p$.

[†] Cf. Definition 6.1.4.

[‡] Cf. Definition 6.1.5.

Proposition 6.2.13

$(\hat{x}, \hat{\lambda}) \in R^n \times R^m$ is a weak saddle point of w-L if and only if there exists $a \in R^p$ such that $(\hat{x}, \hat{\lambda})$ is a saddle point of the scalarized function $(\text{w-}L)_a$.

The primal problem and the dual problem can be represented by the weak Lagrangian as follows.

Proposition 6.2.14†

(i) $-\text{w-}\Phi^*(0, \lambda) = \text{w-Inf} \bigcup_x \text{w-}L(x, \lambda)$ for all $\lambda \in R^m$ and hence w-Sup(D) = w-Sup w-Inf w-L,

(ii) if $\Phi \in \Gamma^p$, then $\Phi(x, 0) = \text{w-Sup} \bigcup_\lambda \text{w-}L(x, \lambda)$ for all $x \in R^n$ and hence w-Inf(P) = w-Inf w-Sup w-L.

Theorem 6.2.4‡

If $\Phi \in \Gamma^p$, then the following conditions are equivalent to each other:

(i) $(\hat{x}, \hat{\lambda}) \in R^n \times R^m$ is a weak saddle point of the weak Lagrangian,

(ii) $\hat{x}$ and $\hat{\lambda}$ are solutions of (P) and (D), respectively, and $\Phi(\hat{x}, 0) \cap -\text{w-}\Phi^*(0, \hat{\lambda}) \neq \varnothing$,

(iii) $\Phi(\hat{x}, 0) \cap -\text{w-}\Phi^*(0, \hat{\lambda}) \cap R^p \neq \varnothing$.

Theorem 6.2.5§

Suppose that $\Phi \in \Gamma^p$ and that (P) is weakly stable. Then the following conditions are equivalent to each other:

(i) $\hat{x}$ is a solution of (P),

(ii) there exists $\hat{\lambda} \in R^m$ such that $(\hat{x}, \hat{\lambda})$ is a weak saddle point of the weak Lagrangian. In this case $\hat{\lambda}$ is a solution of (D).

6.3 Conjugate Duality Based on Strong Supremum and Infimum

Brumelle [B21] took an approach different from Tanino and Sawaragi's or Kawasaki's to derive duality for convex multiobjective optimization problems. His definitions of conjugates and subgradients are all based on the concept of strong supremum and infimum, which are the least upper bound and the greatest lower bound, respectively, for a partially ordered set. Since

† Cf. Proposition 6.1.14 and 6.1.15.

‡ Cf. Theorem 6.1.2.

§ Cf. Theorem 6.1.3.

strong infimum essentially implies componentwise infimum, we can directly extend the well-known results concerning conjugate functions and subgradients of scalar-valued convex functions. (See Section 2.1.)

Brumelle also concentrated on properly efficient solutions, which can be completely characterized as solutions of scalarized problems with negative (or positive) weighting vectors. (See Theorems 3.4.1 and 3.4.2.) Since every efficient solution of a linear multiobjective optimization problem is properly efficient (Corollary 3.1.1), his results are particularly meaningful in the linear case.

In this section we overview the conjugate duality according to Brumelle [B21]. As in Section 6.2, the proofs of theorems, propositions, and lemmas are omitted.

6.3.1 Strong Supremum and Infimum, Conjugate Functions, and Subgradients

We provide some fundamental concepts—strong supremum, conjugates and subgradients—in this subsection. They were also used by Zowe [Z9, Z10]. However Zowe and Craven [C17] considered problems of finding strongly supremal points (i.e., strongly optimal solutions), which are different from the multiobjective optimization problems in this book. It might be better to call the concepts defined in this section without the modifier *strong*, since they are direct extensions from the concepts in the scalar case. However, in this book, we attach it in order to distinguish them from the concepts defined earlier.

First we define strong supremum and maximum of a set in the space $R^p_\infty = R^p \cup \{\pm\infty\}$.

Definition 6.3.1 *(Strong Supremum (Infimum) and Maximum (Minimum))*

Let Y be a subset of R^p_∞. A point $\hat{y}$ is said to be a strong supremum of Y and denoted by $\hat{y} = \mathbf{sup}\, Y$ if and only if $\hat{y}$ is an upper bound for Y in R^p_∞ (i.e., $\hat{y} \geqq y$ for any $y \in Y$) and $\hat{y} \leqq \bar{y}$ whenever $\bar{y}$ is any other upper bound of Y. If Y is empty, then $\mathbf{sup}\, Y = -\infty$. If, in addition, $\hat{y} \in Y$ then we call $\hat{y}$ a strong maximum of Y and denote it by $\mathbf{max}\, Y$. Strong infimum ($\mathbf{inf}\, Y$) and strong minimum ($\mathbf{min}\, Y$) can be defined analogously.

Note that $\mathbf{sup}\, Y$ is a unique element of R^p_∞ for any $Y \subset R^p_\infty$ and that if $\mathbf{sup}\, Y \in R^p$, then $(\mathbf{sup}\, Y)_i = \sup\{y_i : y \in Y\}$. Namely, $\mathbf{sup}\, Y$ is essentially the componentwise supremum of Y.

Definition 6.3.2 (*Strong Conjugate and Strong Biconjugate*)†

Let f be a convex function from R^n to R^p_∞. The strong conjugate function $\mathbf{f}^*$ of f is the function from $R^{p\times n}$ to R^p_∞ defined by

$$\mathbf{f}^*(T) = \mathbf{sup}\{Tx - f(x) : x \in R^n\}.$$

Furthermore, the strong biconjugate function $\mathbf{f}^{**} : R^n \to R^p_\infty$ is defined by

$$\mathbf{f}^{**}(x) = \mathbf{sup}\{Tx - \mathbf{f}^*(T) : T \in R^{p\times n}\}.$$

Note that if $\mathbf{f}^*(T) \in R^p$, then $\mathbf{f}^*(T) = (f_1^*(t^1), \ldots, f_p^*(t^p))^{\mathrm{T}}$, where $T = (t^1, \ldots, t^p)^{\mathrm{T}}$ with each $t^i \in R^n$.

Definition 6.3.3 (*Strong Closure*)‡

Let f be a vector-valued convex function from R^n to R^p_∞. The strong closure of f, which is denoted by $\mathbf{cl}\, f$, is defined by $(\mathbf{cl}\, f)(x) = \mathbf{inf}\{y : (x, y) \in \mathrm{cl}(\mathrm{epi}\, f)\}$ if f never takes the value $-\infty$ and $(\mathbf{cl}\, f)(x) = -\infty$ if $f(\bar{x}) = -\infty$ for some $\bar{x}$. The function f is said to be strongly closed if $f = \mathbf{cl}\, f$.

Proposition 6.3.1§

If f is a convex function from R^n to R^p, then $\mathbf{f}^*$ is convex and $\mathbf{f}^{**} = \mathbf{cl}\, f$.

Example 6.3.1

As a simple example, consider

$$f(x) = (x, -x)^{\mathrm{T}} \qquad \text{for} \quad x \in R.$$

Then

$$\mathbf{f}^*(T) = \begin{cases} 0 & \text{if} \quad T = (1, -1) \\ +\infty & \text{otherwise.} \end{cases}$$

Thus, $\mathbf{f}^*(T)$ takes the infinite value too often.

Next, we define strong subgradients of convex vector-valued functions.

Definition 6.3.4 (*Strong Subgradient*)‖

Let f be a convex vector-valued function from R^n to R^p_∞. A matrix $T \in R^{p\times n}$ is said to be a strong subgradient of f at $\hat{x}$ if

$$f(x) \geqq f(\hat{x}) + T(x - \hat{x}) \qquad \text{for any} \quad x \in R^n.$$

† Cf. Definition 2.1.20, 6.1.1, and 6.2.3.
‡ Cf. Definition 2.1.19.
§ Cf. Proposition 2.1.26.
‖ Cf. Definition 2.1.23, 6.1.2, and 6.2.8.

The set of strong subgradients of f at $\hat{x}$ is called the strong subdifferential of f at $\hat{x}$ and is denoted by $\partial f(\hat{x})$. If $\partial f(\hat{x})$ is not empty, f is said to be strongly subdifferentiable at $\hat{x}$. $\partial \mathbf{f}^*(T)$ can be defined analogously.

Proposition 6.3.2

Let f be a convex function from R^n to R^p_∞. Then

$$\partial f(x) = \bigtimes_{i=1}^{p} \partial f_i(x) = \{T = (t^1, \ldots, t^p)^T : t^i \in \partial f_i(x),\ i = 1, \ldots, p\}.$$

Proposition 6.3.3†

Let f be a proper convex function from R^n to R^p_∞. Then

$$f(x) + \mathbf{f}^*(T) \geqq Tx$$

for all $x \in R^n$ and $T \in R^{p \times n}$, with equality holding if and only if $T \in \partial f(x)$. If f is closed, then equality is obtained if and only if $x \in \partial \mathbf{f}^*(T)$.

6.3.2 Proper Efficiency and Zerolike Matrix

In this subsection we consider the characterization of properly efficient solutions of a convex vector-valued function in terms of zerolike matrices. Let f be a proper convex function from R^n to R^p_∞. We should recall that $\hat{x}$ is a properly efficient solution of a multiobjective optimization problem

$$\text{minimize} \quad f(x)$$

if and only if $\hat{x}$ minimizes the scalar-valued function $\langle \mu, f(x) \rangle$ for some $\mu \in \operatorname{int} R^p_+$ (Theorems 3.4.1 and 3.4.2). In this case we simply say that $\hat{x}$ is properly efficient for f.

Definition 6.3.5 (*Zerolike Matrix*)

A matrix $T \in R^{p \times n}$ is said to be zerolike (written $T \sim 0$ or $0 \sim T$) if there exists a vector $\mu \in \operatorname{int} R^p_+$ such that the n vector $T^T\mu = 0$.

Proposition 6.3.4

Let f be a proper convex function from R^n to R^p_∞. Then $\hat{x}$ is properly efficient for f if and only if there exists a zerolike $T \in \partial f(\hat{x})$, in other words, if and only if there exists a zerolike $T \in R^{p \times n}$ such that

$$f(\hat{x}) \in \mathbf{min}\{f(x) - Tx : x \in R^n\}.$$

We must note that, when $T \in \partial f(\hat{x})$, T_{ij} can be interpreted as the rate of change of the ith objective per unit change in the jth activity at $\hat{x}$.

† Cf. Proposition 2.1.27 and 6.1.3.

6.3.3 Duality

Now we can derive duality in convex multiobjective optimization based on the preceding concepts. Though the original paper by Brumelle [B21] generalized the duality theory of convex bifunctions by Rockafellar [R8], we rewrite it in a form analogous to Sections 6.1 and 6.2.

Let φ be a function from $R^n \times R^m$ to R^p_∞. The function φ is assumed to be proper and convex throughout this section. The primal problem is

(P) minimize $\varphi(x, 0)$.

We would like to find $\hat{x}$ that is properly efficient for $\varphi(\cdot, 0)$ (simply for φ). In view of Proposition 6.3.4, for each $T \in R^{p \times n}$, we define the convex function φ_T: $R^n \times R^m \to R^p_\infty$ as

$$\varphi_T(x, u) = \varphi(x, u) - Tx.$$

The notation $(u, T) \sim 0$ means $u = 0$ and $T \sim 0$. Dually, for each u, let

$$\boldsymbol{\varphi}^*_u(T, \Lambda) = \boldsymbol{\varphi}^*(T, \Lambda) - \Lambda u.$$

We consider two kinds of perturbed problems that are regarded as strong optimization problems

($\mathrm{P_T}$) minimize $\varphi_T(x, 0)$,

($\mathrm{D_T}$) maximize $-\boldsymbol{\varphi}^*_0(T, \Lambda) = -\boldsymbol{\varphi}^*(T, \Lambda)$ (T is fixed).

Definition 6.3.6 (*Admissibility*)

If $\hat{x}$ is a minimizing solution of ($\mathrm{P_T}$), that is, if

$$\varphi_T(\hat{x}, 0) = \mathbf{min}\{\varphi_T(x, 0) : x \in R^n\},$$

for some zerolike T, then $\hat{x}$ is said to be admissible for φ. Dually, if $\hat{\Lambda}$ is a maximizing solution of ($\mathrm{D_T}$), that is, if

$$-\boldsymbol{\varphi}^*(T, \hat{\Lambda}) = \mathbf{max}\{-\boldsymbol{\varphi}^*(T, \Lambda) : \Lambda \in R^{p \times m}\},$$

for some zerolike T, then $\hat{\Lambda}$ is said to be admissible for $\boldsymbol{\varphi}^*$.

Proposition 6.3.5

The following three conditions are equivalent:

(i) $\hat{x}$ is properly efficient for φ,
(ii) $\hat{x}$ is admissible for φ,
(iii) there are some $\mu \in \operatorname{int} R^p_+$ and some $\hat{T} \in \mathscr{T}(\mu)$ such that

$$\langle \mu, \varphi_{\hat{T}}(\hat{x}, 0) \rangle \leqq \langle \mu, \varphi_T(x, 0) \rangle \qquad \text{for all} \quad (x, T) \in R^n \times \mathscr{T}(\mu),$$

where

$$\mathscr{T}(\mu) = \{T \in R^{p \times n} : T^{\mathrm{T}}\mu = 0\}.$$

In view of this proposition we might say that $\hat{\Lambda}$ is properly efficient for φ^* if there are $\mu \in \operatorname{int} R^p_+$ and $\hat{T} \in \mathscr{T}(\mu)$ such that

$$\langle \mu, -\varphi^*(\hat{T}, \hat{\Lambda})\rangle \geqq \langle \mu, -\varphi^*(T, \Lambda)\rangle \qquad \text{for all} \quad (T, \Lambda) \in \mathscr{T}(\mu) \times R^{p\times m}.$$

Note that if $\hat{\Lambda}$ is properly efficient for φ^*, then $\hat{\Lambda}$ is properly efficient for $-\varphi^*(T, \cdot)$ for some $T \sim 0$, and hence $\hat{\Lambda}$ is admissible for φ^*; however, the converse statements do not hold.

In view of Proposition 6.3.5, the set of properly efficient solutions of (P) is equal to the set of admissible points, which is the union of the solution sets of $(\mathrm{P_T})$ with $T \sim 0$. Hence, we may define the dual problem (D) of (P) as the problem of finding the set of properly efficient points of φ^*. Namely

(D) find $\hat{\Lambda} \in R^{p\times m}$ such that
there exist $\mu \in \operatorname{int} R^p$ and $\hat{T} \in \mathscr{T}(\mu)$ such that
$\langle \mu, -\varphi^*(\hat{T}, \hat{\Lambda})\rangle \geqq \langle \mu, -\varphi^*(T, \Lambda)\rangle$ for all
$(T, \Lambda) \in \mathscr{T}(\mu) \times R^{p\times m}$.

Note that (D) is different from the problem

$$\text{maximize} \quad -\varphi^*(0, \Lambda) \qquad \text{subject to} \quad \Lambda \in R^{p\times m}.$$

However, in this case, the set of properly efficient points is included in the set

$$\bigcup_{T\sim 0} \{\hat{\Lambda} \in R^{p\times m} : -\varphi^*(T, \hat{\Lambda}) = \max\{-\varphi^*(T, \Lambda) : \Lambda \in R^{p\times m}\}\}$$

but is not equal to it. A drawback of Brumelle's formulation is that the unperturbed dual problem involves the dual perturbation T. This causes excessive sensitivity of the problem (See Ponstein [P3].)

Brumelle defined another efficiency concept for φ^*, motivated by Isermann's results for linear problems.

Definition 6.3.7 *(Isermann Efficiency)*

If there exists some zerolike T such that

$$-\varphi^*(\hat{T}, \hat{\Lambda}) \leq -\varphi^*(T, \Lambda) \qquad \text{for no} \quad T \sim 0 \in R^{p\times n} \qquad \text{and no} \qquad \Lambda \in R^{p\times m},$$

then $\hat{\Lambda}$ is said to be Isermann efficient for φ^*.

We define the perturbation functions w^T on R^m and v_u on $R^{p\times n}$ as

$$w_T(u) = \inf\{\varphi_T(x, u) : x \in R^n\}$$

and

$$v_u(T) = -\sup\{-\varphi^*_u(T, \Lambda) : \Lambda \in R^{p\times m}\},$$

respectively. The functions w_T and v_u are convex with $\operatorname{dom} w_T = \{u : \varphi_T(\cdot, u) \not\equiv +\infty\}$ and $\operatorname{dom} v_u = \{T : \boldsymbol{\varphi}^*_u(T, \cdot) \not\equiv +\infty\}$. The following is a weak duality theorem.

Theorem 6.3.1

For each $u \in R^m$, $x \in R^n$, $\Lambda \in R^{p \times m}$, and $T \in R^{p \times n}$,

$$\varphi_T(x, u) \geq -\boldsymbol{\varphi}^*_{\mathbf{u}}(T, \Lambda).$$

Consequently, for each $u \in R^m$ and $T \in R^{p \times n}$,

$$w_T(u) \geqq -v_u(T),$$

and $\varphi_T(x, 0) \leq -\boldsymbol{\varphi}^*(T, \Lambda)$ does not hold for any $x \in R^n$, $\Lambda \in R^{p \times m}$ and $T \sim 0 \in R^{p \times n}$.

Corollary 6.3.1

If $\varphi_T(x, 0) = -\boldsymbol{\varphi}^*(\hat{T}, \hat{\Lambda})$, $\hat{T}^{\mathrm{T}}\mu = 0$, and $\mu \in \operatorname{int} R^p_+$, then $\hat{x}$ and $\hat{\Lambda}$ have the following properties:

(i) $\hat{x}$ minimizes $\langle \mu, \varphi_{\hat{T}}(x, 0) \rangle$ for $x \in R^n$, and $(\hat{T}, \hat{\Lambda})$ maximized $\langle \mu, -\boldsymbol{\varphi}^*(T, \Lambda) \rangle$ for $(T, \Lambda) \in \mathscr{T}(\mu) \times R^{p \times m}$. This $\hat{x}$ and $\hat{\Lambda}$ are properly efficient for φ and $\boldsymbol{\varphi}^*$, respectively.

(ii) $\hat{x}$ is efficient for φ; i.e.,

$$\varphi(\hat{x}, 0) \geq \varphi(x, 0) \qquad \text{for no} \quad x \in R^n,$$

(iii) if $\hat{T}\hat{x} = 0$, then $\hat{\Lambda}$ is Isermann efficient for $\boldsymbol{\varphi}^*$.

The result corresponding to Lemma 6.1.3 and Lemma 6.2.3 is the following.

Lemma 6.3.1

(i) For each $T \in R^{p \times n}$, $\mathbf{w}^*_{\mathbf{T}}(\Lambda) = \boldsymbol{\varphi}^*(T, \Lambda)$ for all Λ,
(ii) if φ is closed, then for each $u \in R^m$, $\mathbf{v}^*_{\mathbf{u}}(x) = \varphi(x, u)$ for all x.

Definition 6.3.8 (*Strong Normality*)†

If w_T is closed at u, problem (P) is said to be strongly normal at (u, T). Similarly, if v_u is closed at T, problem (D) is said to be strongly normal at (u, T). If (P) (resp. (D)) is strongly normal at (u, T) for each $(u, T) \sim 0$, then (P) (resp. (D)) is said to be strongly normal.

† Cf. Definition 6.2.9.

Theorem 6.3.2†

If φ is closed, then the following are equivalent:

(i) (P) is strongly normal at (u, T),
(ii) (D) is strongly normal at (u, T),
(iii) $w_T(u) = -v_u(T)$.

Lemma 6.3.2

If φ is closed, then any one of the following conditions is sufficient for guaranteeing that (P) (or equivalently (D)) is strongly normal at $(0, T)$:

(i) $0 \in \operatorname{ri}(\{u : \varphi(\cdot, u) \not\equiv +\infty\})$.
(ii) $T \in \operatorname{ri}(\{T' : \boldsymbol{\varphi}^*(T', \cdot) \not\equiv +\infty\})$.
(iii) φ is polyhedral (i.e., $\operatorname{epi} \varphi$ is a polyhedral convex set) and $\varphi(\cdot, 0) \not\equiv +\infty$.
(iv) $\boldsymbol{\varphi}^*$ is polyhedral and $\boldsymbol{\varphi}^*(T, \cdot) \not\equiv +\infty$.

We can now define the strong Lagrangian for φ_T and its saddle point.

Definition 6.3.9 *(Strong Lagrangian and Strong Saddle Point)*‡

The strong Lagrangian $\mathbf{L}$: $R^n \times R^{p \times m} \to R^p_\infty$ of the primal problem (P) is defined by

$$\mathbf{L}(x, \Lambda) = \mathbf{inf}\{\varphi(x, u) - \Lambda u : u \in R^m\}.$$

Moreover, the strong Lagrangian $\mathbf{L}_T$ of the problem $(\mathrm{P_T})$ is defined by

$$\mathbf{L}_T(x, \Lambda) = \mathbf{inf}\{\varphi_T(x, u) - \Lambda u : u \in R^m\}.$$

Clearly $\mathbf{L}_T(x, \Lambda) = \mathbf{L}(x, \Lambda) - Tx$. A point $(\hat{x}, \hat{\Lambda})$ is said to be a strong saddle point of $\mathbf{L}_T$ if

$$\mathbf{L}_T(\hat{x}, \Lambda) \leqq \mathbf{L}_T(\hat{x}, \hat{\Lambda}) \leqq \mathbf{L}_T(x, \hat{\Lambda}) \qquad \text{for all} \quad x \in R^n, \quad \Lambda \in R^{p \times m}.$$

Theorem 6.3.3§

Suppose that φ is closed and let $x \in R^n$, $\Lambda \in R^{p \times m}$ and $0 \sim T \in R^{p \times n}$. The following conditions are equivalent:

(i) Problem (P) is normal at $(0, T)$ and $\hat{x}$ and $\hat{\Lambda}$ are properly efficient solutions of (P) and (D), respectively, with

$$\varphi_T(\hat{x}, 0) = \mathbf{min}\{\varphi_T(x, 0) : x \in R^n\}$$

† Cf. Proposition 6.2.7.
‡ Cf. Definition 6.1.4, 6.1.5, 6.2.11, and 6.2.12.
§ Cf. Theorem 6.1.2 and 6.2.4.

and

$$-\boldsymbol{\varphi}^*(T, \hat{\Lambda}) = \mathbf{max}\{-\boldsymbol{\varphi}^*(T, \Lambda) : \Lambda \in R^{p \times m}\}.$$

(ii) $(\hat{x}, \hat{\Lambda})$ is a strong saddle point of the strong Lagrangian $\mathbf{L}_T$.
(iii) $\varphi_T(\hat{x}, 0) \leqq -\boldsymbol{\varphi}^*(T, \hat{\Lambda})$ (in which case equality must actually hold).

Theorem 6.3.4

Suppose that φ is closed, that dom $\varphi(\cdot, 0) \neq \varnothing$, and that $T \in R^{p \times n}$ is zero like. If φ is polyhedral or $0 \in \text{ri}\ \{u : \varphi(x, u) \not\equiv +\infty\}$, then the following conditions are equivalent:

(i) $\hat{x}$ is a properly efficient solution of (P) and strongly minimizes $\varphi_T(\cdot, 0)$,
(ii) $(\hat{x}, \hat{\Lambda})$ is a strong saddle point of the strong Lagrangian $\mathbf{L}_T$ for some $\hat{\Lambda} \in R^{p \times m}$,
(iii) $(T, 0) \in \partial \mathbf{L}(\hat{x}, \hat{\Lambda})$ for some $\hat{\Lambda} \in R^{p \times m}$.

Theorem 6.3.5

Suppose that φ is polyhedral and that $0 \notin \text{aff}(\text{dom}\ \varphi(\cdot, 0))$, where aff$(S)$ denotes the affine hull of a set S. If $\hat{x}$ is an efficient solution of the problem (P), then there exists $(\hat{T}, \hat{\Lambda})$ such that $-\boldsymbol{\varphi}^*(\hat{T}, \hat{\Lambda}) = \varphi(\hat{x}, 0)$ and $\hat{\Lambda}$ is Isermann efficient.

7 METHODOLOGY

This chapter deals with methodology for practical implementation of multiobjective decision making. Above all, utility (value) analysis and interactive programming methods play key roles in this area. Our aim in this chapter is not merely to introduce these methods, but to discuss critical points. For details and practical implementation of utility (value) analysis, Keeney and Raiffa [K6] is excellent. Regarding interactive programming methods, Chankong and Haimes [C6] and Hwang and Masud [H17] include good surveys. Here, we shall emphasize controversial points of the typical methods from a viewpoint of practical implementation and give a direction toward future research.

7.1 Utility and Value Theory

Following the same notations as in the preceding chapters (which may be a little different from those of other literatures), let X denote the set of alternatives, and let each of its elements be evaluated by several criteria or objective functions $f_i: X \to R^1$ $(i = 1, \ldots, p)$. In the utility and value theory, such criteria for evaluating alternatives are often called *attributes*. In this section, we do not distinguish these terminologies so long as there is no fear of confusion. Preference of decision makers are supposed to be expressed on the criteria space Y (in which each y_i reflects the criterion f_i), rather than on the set of alternatives itself. This section will be concerned with existence, representation, and uniqueness of a function $u: Y \to R^1$ representing the decision maker's preference. Such a function representing the decision maker's preference is in general called a preference-preserving function or a *preference function*. In particular, following Keeney and Raiffa [K6], a preference function in decision making under uncertainty is called a *utility function*, which is used to represent a general preference function in some other literatures. On the other hand, a preference function in decision making under certainty is called a *value function*.

7.1.1 Preference Functions

This subsection will be concerned primarily with the existence problem of preference functions.

Definition 7.1.1 (*Homomorphism*)

Let $(A, \succ_1)$ and $(B, \succ_2)$ be some ordered sets. An onto mapping $u: A \to B$ is called a homomorphism from $(A, \succ_1)$ to $(B, \succ_2)$ if the following holds for any $a^1, a^2 \in A$:

$$a^1 \succ_1 a^2 \Rightarrow u(a^1) \succ_2 u(a^2).$$

Then $(A, \succ_1)$ and $(B, \succ_2)$ are called homomorphic by the mapping u.

Definition 7.1.2 (*Isomorphism*)

Let $(A, \succ_1)$ and $(B, \succ_2)$ be some ordered sets. An onto and one-to-one mapping $u: A \to B$ is called an isomorphism from $(A, \succ_1)$ to $(B, \succ_2)$, if the following holds for any $a^1, a^2 \in A$:

$$a^1 \succ_1 a^2 \Leftrightarrow u(a^1) \succ_2 u(a^2).$$

Then $(A, \succ_1)$ and $(B, \succ_2)$ are called isomorphic by the mapping u.

Remark 7.1.1

A mapping u is an isomorphism if and only if both u and u^{-1} are homomorphisms.

In order to make the treatment of the decision maker's preference easier, let us consider the set of real numbers with the usual inequality relation $(R^1, >)$ isomorphic or homomorphic to $(Y, \succ)$. A preference function is defined as an isomorphic (or homomorphic) function $u: (Y, \succ) \to (R^1, >)$.

In examining the existence of such a preference function, it is meaningful to know the structure of the set of real numbers. Among several characteristics of the set of real numbers, the structure of order and addition is essential and also plays an important role to our aim. Let Re denote the set of real numbers with the natural order $>$. By this notation we discriminate such a set of primitive real numbers from R^1 with some topological properties. The first noticeable point is that the natural order is a total order (Definition 2.3.1). Further, let Ra denote the set of rational numbers. It is well known that the rational numbers are *countable* and *dense* in Ra itself, i.e., for any $\alpha, \beta \in \mathrm{Ra}$

$$\alpha > \beta \Rightarrow \alpha > \gamma > \beta \qquad \text{for some} \quad \gamma \in \mathrm{Ra}.$$

This encourages us to score each element of a totally ordered and countable set $Y = \{y^1, y^2, \ldots, y^k, \ldots\}$ by the function $u: Y \to \mathrm{Ra}$ such that for any positive integer n,

(i) $u(y^1) = 0$,
(ii) $y^{n+1} \succ y^k,\ 1 \leqq \forall k \leqq n \Rightarrow u(y^{n+1}) = n$,
$y^{n+1} \prec y^k,\ 1 \leqq \forall k \leqq n \Rightarrow u(y^{n+1}) = -n$.

Otherwise,

$$u(y^{n+1}) = (u(y^i) + u(y^j))/2,$$

since there are integers i and j $(1 \leqq i, j \leqq n)$ such that

$$y^i \succ y^{n+1} \succ y^j$$

$$y^k \succsim y^i \qquad \text{or} \qquad y^j \succsim y^k, \qquad 1 \leqq \forall k \leqq n.$$

This observation yields Theorem 7.1 below.

For an uncountable set Y, we make use of the denseness of Ra in Re; that is, for any $\alpha, \beta \in \mathrm{Re}$

$$\alpha > \beta \Rightarrow \alpha > \gamma > \beta \qquad \text{for some} \quad \gamma \in \mathrm{Ra}.$$

Of course, it is required that Y have a subset that is countable and order-dense in Y. (See Theorem 7.2 below.) It should be noted here that the denseness of Ra does not imply its connectedness or continuity. In fact, there are infinitely many gaps† in Ra. Similarly, there might be a gap in an uncountable set Y with a subset countable and order-dense in Y. In defining a preference function $Y \to R^1$, special attention must be paid to such gaps. (For a proof of Theorem 7.2 below, see, for example, Krantz *et al.* [K9] and Fishburn [F7].)

Now recall that the set of real numbers is completed by making up these gaps. To this end, irrational numbers are introduced, which produces the *continuity property* of Re.

In addition to the order structure, an operation on Re is defined called *addition*, $+$, by assigning $a + b$ for any $a, b \in$ Re (which is again in Re) with the following properties:

(1) $(a + b) + c = a + (b + c)$,
(2) $a + b = b + a$,
(3) given $a, b \in$ Re, there exists $c \in$ Re such that $a + c = b$,
(4) $a < b \Rightarrow a + c < b + c$, for any $c \in$ Re,

where property (4) connects the order and the addition on Re and is called the monotonicity of addition. An additive structure stands for a set together with the operation of addition defined on the set. The additive structure on the ordered set (Re, $<$) with the properties (1)–(4) provides a basis for *measurability* (*cardinality*) of real numbers. The addition on Re along with the order $>$ leads to the well-known *Archimedean property*:

For given $a > 0$ and $b > 0$, there exists an integer n such that $na > b$.

In order to get a measurable preference function for $(Y, \succ)$, some notions corresponding to the addition and the Archimedean property will be necessary (cf. Theorem 7.1.4 and Remark 7.1.4 below).

In the following, we shall list several existence theorems for preference functions without proof. (For more details see, for exampl, Krantz *et al.* [K9], Fishburn [F7], and Keeney and Raiffa [K6].)

† For a totally ordered set $(A, \succ)$, the term *a cut of A* implies to partition A in such a way that $A = A' \cup A''$, where A' and A'' are disjoint subsets of A and for any $a' \in A'$ and $a'' \in A''$, we have $a'' \succ a'$. For such a cut, the following three cases are possible: (i) The lower part A' has a greatest element and the upper part A'' has a least element; (ii) the lower part A' has no greatest element and the upper part A'' has no least element; (iii) either A' has a greatest element and A'' has no least element or A' has no greatest element and A'' has a least element.

For any cut of discrete sets, case (i) is necessarily possible. On the other hand, for any cut of dense sets either case (ii) or (iii) is possible, whereas case (i) is never possible. In case (ii), we say that the cut has a gap. For the set of rational numbers Ra, infinitely many cuts with gaps are possible. For example, consider $\text{Ra} = \{\alpha \in \text{Ra} \mid \alpha^2 \leqq 2\} \cup \{\alpha \in \text{Ra} \mid \alpha^2 > 2\}$. Irrational numbers are introduced to make up for such gaps in Ra so that only case (iii) arises for any cut in the set of real numbers, which is defined as the set consisting of both Ra and the set of irrational numbers (Dedekind cut).

7.1.1.1 Value Functions

Theorem 7.1.1

If $(Y, \succ)$ is a totally ordered set and if Y is countable, then there exists a value function u: $Y \to R^1$ such that for every $y^1, y^2 \in Y$

$$y^1 \succ y^2 \Leftrightarrow u(y^1) > u(y^2).$$

Such a value function is unique up to a monotonically increasing function.†

Definition 7.1.3 (*Order Dense*)

Let $\succ$ be an order relation. Then a subset Z of Y is said to be $\succ$-order dense in Y, if for any $y^1, y^2 \in Y \backslash Z$ there exists $z \in Z$ such that

$$y^1 \succ y^2 \Rightarrow y^1 \succ z, \qquad z \succ y^2.$$

Theorem 7.1.2

If $(Y, \succ)$ is a totally ordered set and if Y has a subset Z countable and $\succ$-order dense in Y, then there exists a value function u: $Y \to R^1$ such that for every $y^1, y^2 \in Y$

$$y^1 \succ y^2 \Leftrightarrow u(y^1) > u(y^2).$$

Such a value function is unique up to monotonically increasing functions.

Remark 7.1.2

In cases of $(Y, \succ)$ being an weakly ordered set, we define the order $\succ'$ on the set of equivalent classes $Y/\sim$ as

$$a \succ' b, \qquad a, b \in Y/\sim \Leftrightarrow x \succ y, \qquad x \in a, \qquad y \in b.$$

† In the following, it is important to keep in mind what meaning the numerical values in (Re, $>$) isomorphic or homomorphic to $(Y, \succ)$ have. If these numbers make sense only for their orders, they are said to be ordinal. Clearly, since their order relations are unchangeable by any monotonically increasing transformation, the ordinal scale is rather redefined as the one that is unique up to a monotonically increasing transformation. Numerical values that make sense for their differences are said to be measurable or cardinal. There are two kinds of measurable scales: The first is the interval scale unique up to a positive affine transformation, which implies that both the order among the numerical values and the ratio $(u(y^1) - u(y^2))/(v(y^1) - v(y^2))$ are invariant under a positive affine transformation v of u; that is, $v = \alpha + \beta u$, for $\beta > 0$. This relationship also says that the interval scale has arbitrariness of unit and origin, which can be seen in an example of the measure of temperature (Celsius and Fahrenheit). The second scale is the ratio scale unique up to a positive similar transformation, which implies that both the order among the numerical values and the ratio $u(y)/v(y)$ are invariant under a positive similar transformation v of u; that is, $v = \beta u$, for $\beta > 0$. The ratio scale leaves arbitrariness of unit as in examples of length and weight. In general, two preference functions u and v with the relationship $u = \alpha + \beta v$ for $\beta > 0$, are said to be *equivalent* as cardinal measure.

Then, since $\succ'$ is clearly a total order, Theorem 7.1.1 and 7.1.2 can easily be modified for a weakly ordered set $(Y, \succ)$. In Theorem 7.1.1, the result holds for a weakly ordered set $(Y, \succ)$ by replacing *Y is countable* with *the set of equivalent classes $Y/\sim$ is countable*. Theorem 7.1.2 holds for a weakly ordered set $(Y, \succ)$ by replacing *$(Y, \succ)$ has a subset countable and $\succ$-order dense in Y* with *the set of equivalent classes $Y/\sim$ has a subset countable and $\succ'$-order dense in $Y/\sim'$*.

In order to reduce the selection of the maximal element of $(Y, \succ)$ to the maximization of the preference-representing function $u(\cdot)$, the continuity property of $u(\cdot)$ encourages the existence of a maximal element under some appropriate conditions and applications of various optimization techniques. A mild, sufficient condition for continuity of value functions was derived by Debreu [D3].

Theorem 7.1.3

Suppose that the preference $\succ$ is defined on a subset Y of p-dimensional Euclidean space R^p with the natural topology τ. Let τ' be the induced topology of τ to Y. Then, if Y is connected with respect to the topology τ' and if for any $y' \in Y$

$$Y^{y'} := \{y \in Y : y \succ y'\} \in \tau', \qquad \text{and} \qquad Y_{y'} := \{y \in Y : y \prec y'\} \in \tau',$$

then there exists a continuous value function u such that for any $y^1, y^2 \in Y$

$$y^1 \succ y^2 \Leftrightarrow u(y^1) > u(y^2).$$

Such a value function is unique up to a monotonically increasing function.

7.1.1.2 Measurable Value Functions

For a long time, there have been many controversies with regard to measurable (cardinal) value functions. (See, for example, Fishburn [F7].) One of the reasons for rejecting them in traditional economics is that value functions with ordinal measure provide a sufficiently effective tool for many economic analyses. However, measurable value functions would be expected to show great power as a tool for decision making rather than for mere analysis using ordinal value functions.

Definition 7.1.4 (*Preference–Difference Order*)

Define $\succ^*$ as a binary relation on $Y \times Y$ reflecting the order over differences of preference $\succ$ on Y; i.e., for any $y^1, y^2, y^3, y^4 \in Y$, $y^1 y^2 \succ^* y^3 y^4$ implies that the strength of preference for y^1 over y^2 is greater than the strength of preference for y^3 over y^4.

Theorem 7.1.4[†]

Let $\succ^*$ denote a preference difference order as in Definition 7.1.4. Suppose the following conditions hold:

(C1) $\succ^*$ is a weak order.
(C2) $y^1y^2 \succ^* y^3y^4 \Rightarrow y^4y^3 \succ^* y^2y^1$.
(C3) $y^1y^2 \succ^* \hat{y}^1\hat{y}^2$ and $y^2y^3 \succ^* \hat{y}^2\hat{y}^3 \Rightarrow y^1y^3 \succ^* \hat{y}^1\hat{y}^3$.
(C4) $y^1y^2 \succ^* y^3y^4 \succ^* y^1y^1 \Rightarrow y^1y^5 \sim^* y^3y^4 \sim^* y^6y^2$ for some $y^5, y^6 \in Y$.
(C5) If $\{y^1, y^2, \ldots, y^k, \ldots\}$ is a bounded standard sequence[‡] such that

$$y^{k+1}y^k \sim^* y^2y^1 \qquad \text{for every} \quad k, \quad \text{not} \quad y^2y^1 \sim^* y^1y^1,$$

and

$$\hat{y}\bar{y} \succ^* y^ky^1 \succ^* \bar{y}\hat{y} \qquad \text{for some} \quad \hat{y}, \bar{y} \in Y \quad \text{and for all} \quad k,$$

then the sequence is finite.

Then there exists a function $u\colon Y \to R^1$ such that for any y^1, y^2, y^3, and y^4 in Y,

$$y^1y^2 \succ^* y^3y^4 \Leftrightarrow u(y^1) - u(y^2) > u(y^3) - u(y^4).$$

The function $u(\cdot)$ is unique up to a positive affine transformation.

Remark 7.1.3

Since the value function $u(\cdot)$ in Theorem 7.1.4 is unique up to a positive affine transformation, it is a cardinal measure. Hence it is called a *measurable value function*.

Remark 7.1.4

For a precise proof see, for example, Krantz *et al.* [K9]. Fishburn [F7] gave another proof by using the fact that the conditions (C1)–(C5) yield an additive structure for $(Y \times Y, \succ^*)$, whence according to Theorem 7.1.15 below there exists an additive value function representing the preference $\succ^*$ over two attributes $Y \times Y$ such that

$$y^1y^2 \succ^* y^3y^4 \Leftrightarrow u_1(y^1) + u_2(y^2) > u_1(y^3) + u_2(y^4).$$

A measurable value function $u(\cdot)$ can be obtained by setting $u(y) = u_1(y) - u_2(y)$. In referring to Theorem 7.1.15, note that condition (C3) corresponds to the Thomsen condition, condition (C4) to the solvability condition, and condition (C5) to the Archimedean condition.

[†] Krantz *et al.* [K9].

[‡] In this case, a standard sequence $\{y^1, y^2, \ldots, y^k, \ldots\}$ is defined as the one such that $y^3y^1 = 2y^2y^1, \ldots, y^ky^1 = (k-1)y^2y^1$. In addition, the sequence is said to be bounded if there exists a $\tilde{y} \in Y$ such that $\tilde{y} \succ y^k, \forall k$.

7.1.1.3 von Neumann–Morgenstern Utility Functions[†]

In decision making under risk, an alternative does not necessarily produce exactly one consequence, rather it possibly causes several consequences with some probability distribution. Therefore, risky alternatives are identified with probability measures over Y. Note that any convex combination of probability measures is also a probability measure. For general spaces other than linear spaces, roughly speaking such operations as the convex combination are usually called mixture operations. A set with a mixture operation is called a mixture set.

Definition 7.1.5 (*Mixture Set*)[‡]

A set S is said to be a mixture set if S is closed under the operation on S assigning $\alpha a + (1 - \alpha)b$ to each $a, b \in S$, and to each $\alpha \in [0, 1]$; i.e., $\alpha a + (1 - \alpha)b$ is again in S and if the operation satisfies for any $a, b \in S$, and for any $\alpha, \beta \in [0, 1]$,

(i) $1a + 0b = a$,
(ii) $\alpha a + (1 - \alpha)b = (1 - \alpha)b + \alpha a$.
(iii) $\alpha[\beta a + (1 - \beta)b] + (1 - \alpha)b = \alpha\beta a + (1 - \alpha\beta)b$.

The following theorems are due to Fishburn [F7] which weakened the conditions of Herstein and Milnor [H10].

Theorem 7.1.5

Let S be a mixture set. Then there exists an affine utility function $u: S \to R^1$ representing the preference $\succ$ over S; namely for any $a, b \in S$

$$\begin{aligned} a \succ b &\Leftrightarrow u(a) > u(b) \\ u(\alpha a + (1 - \alpha)b) &= \alpha u(a) + (1 - \alpha)u(b), \end{aligned} \tag{7.1.1}$$

if and only if the following hold:

(C1) $\succ$ is a weak order,
(C2) $a \succ b \Rightarrow \alpha a + (1 - \alpha)c \succ \alpha b + (1 - \alpha)c$ for any $c \in S$ and any $\alpha \in (0, 1)$,
(C3) $a \succ b \succ c \Rightarrow \alpha a + (1 - \alpha)c \succ b \succ \beta a + (1 - \beta)c$ for some $\alpha, \beta \in (0, 1)$.

This utility function is unique up to a positive linear transformation.

[†] von Neumann and Morgenstern [N13] and Luce and Suppes [L13].
[‡] Herstein and Milnor [H10].

Remark 7.1.5

Condition C2 is called the *independence condition*, whereas C3 is called the *continuity condition*.

Now we turn to our decision problem. Let $Y = R^p$ and denote the set of probability measures on Y as $\tilde{Y}$. Here, $\tilde{Y}$ is considered as the set of risky prospects including Y as the set of sure prospects. Note that $\tilde{Y}$ is a mixture set. The well-known expected utility principle follows from this observation through Theorem 7.1.5. For example, consider a case with discrete random variables on Y. Suppose that there exists a utility function $u(\cdot)$ on the set $\tilde{Y}$ of simple probability measures. Then, we consider the utility function restricted on Y by setting

$$u(y) = u(p) \qquad \text{for} \quad p \in \tilde{Y} \qquad \text{such that} \quad p(y) = 1.$$

As is well known, there exists some finite number r for a simple probability measure such that

$$\{y : p(y) > 0\} = \{y^1, y^2, \ldots, y^r\}.$$

Therefore, using the relationship (7.1.1) for $p^i \in \tilde{Y}$ such that $p^i(y^i) = 1$ successively,

$$u(p) = u\left(\sum_{i=1}^{r} p(y^i)p^i\right) = \sum_{i=1}^{r} p(y^i)u(y^i),$$

which represents the expected utility of the risky prospect p. This discussion can be readily extended to cases with general probability measures.

Definition 7.1.6 (*Expected Utility*)

Corresponding to a risky outcome p, $p(y)$ denotes a probability distribution for discrete $y \in Y$, and a probability density function for continuous $y \in Y$. The expected utility of p is defined by

$$E(p, u) = \begin{cases} \sum_{y \in Y} p(y)u(y) & \text{for discrete} \quad y, \\ \int_{y \in Y} p(y)u(y)\, dy & \text{for continuous} \quad y. \end{cases}$$

Remark 7.1.6

Since, in general, a probability density function does not necessarily exist, sometimes a cumulative probability distribution $P(y)$ is used for defining the expected utility with the help of the Lebesgue–Stieltjes integral:

$$E(u, P) = \int_{y \in Y} u(y)\, dP(y).$$

Note that if a cumulative distribution $P(y)$ is absolutely continuous, then a corresponding density function exists.

Theorem 7.1.6

Suppose that Y is a subset of R^p and $\tilde{Y}$ represents the set of risky prospects represented by probability measures over Y. Then for a given $(\tilde{Y}, \succ)$ there exists a utility function $u(\cdot)\colon Y \to R^1$ such that for any $p, q \in \tilde{Y}$

$$p \succ q \Leftrightarrow E(p, u) > E(q, u)$$

if and only if the following hold:

(C1) $\succ$ is a weak order,
(C2) $p \succ q \Rightarrow \alpha p + (1 - \alpha)r \succ \alpha q + (1 - \alpha)r$ for any $r \in \tilde{Y}$ and any $\alpha \in (0, 1)$,
(C3) $p \succ q \succ r \Rightarrow \alpha p + (1 - \alpha)r \succ q \succ \beta p + (1 - \beta)r$ for some $\alpha, \beta \in (0, 1)$.

This utility function is unique up to a positive affine transformation.

Remark 7.1.7

In practice, we often use the term *lottery* or *prospect* instead of probability measure. Letting Y be a subset of R^1 throughout this Chapter, we denote by $[y^1, y^2, \ldots, y^k : \alpha^1, \alpha^2, \ldots, \alpha^k]$ a lottery with the gain $y^i \in Y$ with the probability α^i $(i = 1; \ldots, k)$. Let the preference $\succ$ be a weak order. For $y^* \succsim y \succsim y^0$ $(y^* \succ y^0)$, a von Neumann–Morgenstern utility function can be identified by finding a probability α such that the sure prospect y is indifferent to the lottery $[y^*, y^0 : \alpha, 1 - \alpha]$. Letting y^* be the best level, y^0 the worst, $u(y^*) = 1$, and $u(y^0) = 0$, we can assign a number of $[0, 1]$ to the utility $u(y)$ for any y. When it is difficult to find such a probability α, we find a sure prospect y indifferent to the fifty–fifty lottery $[y', y'' : 0.5, 0.5]$. Then we have $u(y) = (u(y') + u(y''))/2$. Starting with a pair of y' and y'' with known utilities $u(y')$ and $u(y'')$ (for example, y^* and y^0), we can get a utility $u(y)$ for any y (*midpoint method*).

Remark 7.1.8

Let Y be a subset of R^1, and let $\tilde{Y}$ be the set of lotteries over Y. Then a sure prospect $\hat{y}$ is called a *certainty equivalent* of a lottery $p \in \tilde{Y}$, if $\hat{y}$ is indifferent to p. The von Neumann–Morgenstern utility functions give information on the attitude of the decision maker toward risk. The decision maker's preference is said to be *risk averse* if, for any nondegenerated lottery p, he/she prefers the expected (monetary) value $\bar{y}$ $(=\sum yp(y))$ to the lottery itself; that is,

$$u(\bar{y}) > u(\hat{y}).$$

On the other hand, a preference that prefers lotteries to their expectations is said to be *risk prone* and *risk neutral* if lotteries are indifferent to their expectations. It can be readily shown that risk averse (resp. prone) utility functions monotonically increasing over Y are concave (resp. convex). (See, for example, Keeney and Raiffa [K6].) The difference $\bar{y} - \hat{y}$ is called a *risk premium*, which shows how much the decision maker accepts the decrease of gain than the expectation in order to avoid risk. Risk premiums reflect the degree of risk aversion. Since the form of von Neumann–Morgenstern utility functions itself also reflects the degree of risk aversion, a measure of risk aversion can be defined in terms of utility functions (Arrow [A3], Pratt [P5]) in cases with a monotonically increasing utility function u as

$$r(y) := -u''(y)/u'(y) = -(d/dy)[\log u'(y)],$$

which is called a *local risk aversion* at y. It is easily shown that

$$r(y) \equiv c > 0 \Leftrightarrow u(y) \sim -\exp[-cy],$$
$$r(y) \equiv 0 \Leftrightarrow u(y) \sim y,$$
$$r(y) \equiv c < 0 \Leftrightarrow u(y) \sim \exp[-cy],$$

where $\sim$ represents the equivalence of utility functions. (See, for example, Keeney and Raiffa [K6].) These relationships are often used in determining utility functions.

7.1.2 Decomposed Representation for Multiattribute Preference Functions

The section will be concerned with some decomposed representations for multiattribute preference functions. Our main concerns are how much information is needed for the decomposed representation and how the whole preference function is decomposed. Keeney and Raiffa [K6] gave an excellent development to this theme. In the following, we take a geometric approach that is slightly different from theirs.

7.1.2.1 Decomposability

First, we investigate conditions under which the overall preference function can be decomposed in terms of its conditional preference functions. For simplicity, for a vector $y = (y_1, y_2, \ldots, y_n)$, we use abbreviated notations such as

$$\begin{aligned} y_{i^-} &= (y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n) \\ y_{ij^-} &= (y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_{j-1}, y_{j+1}, \ldots, y_n), \\ (y_i, y_{i^-}) &= (y_i, y_j, y_{ij^-}) = (y_1, y_2, \ldots, y_n). \end{aligned}$$

Similarly,

$$\begin{aligned} Y_{i^-} &= Y_1 \times \cdots \times Y_{i-1} \times Y_{i+1} \times \cdots \times Y_n, \\ Y_{ij^-} &= Y_1 \times \cdots \times Y_{i-1} \times Y_{i+1} \times \cdots \times Y_{j-1} \times Y_{j+1} \times \cdots \times Y_n. \end{aligned}$$

Definition 7.1.7 (*Conditional Preference*)

For a given preference $\succ$ over multiple attributes $Y_1 \times Y_2 \times \cdots \times Y_n$, the *conditional preference* $\succ_{\hat{y}_{i^-}}$ over Y_i at $\hat{y}_{i^-}$ implies the induced preference of $\succ$ on the subset $Y_{\hat{y}_{i^-}} := \{y = (y_1, \ldots, y_n) \mid y_i \in Y_i, y_j = \hat{y}_j \text{ for } j \in \{1, \ldots, n\} \backslash \{i\}\}$.

Similarly, for a given overall preference function $u(y_1, y_2, \ldots, y_n)$ over $Y_1 \times Y_2 \times \cdots \times Y_n$, the *conditional preference function* $u_i(\cdot\,; \hat{y}_{i^-})$ over Y_i at a fixed $\hat{y}_{i^-}$ is defined by $u_i(y_i; \hat{y}_{i^-}) = u(y_i, \hat{y}_{i^-})$.

Definition 7.1.8 (*Preferential Independence*)

The attribute Y_i is said to be *preferentially independent* of Y_{i^-}, if the conditional preference $\succ_{y_{i^-}}$ over Y_i is not affected by the level y_{i^-} of Y_{i^-}, namely, if

$$y_i^1 \succ_{\hat{y}_{i^-}} y_i^2 \quad \text{for some} \quad \hat{y}_{i^-} \in Y_{i^-} \Rightarrow y_i^1 \succ_{y_{i^-}} y_i^2 \quad \text{for all} \quad y_{i^-} \in Y_{i^-},$$

or equivalently, if

$$(y_i^1, y'_{i^-}) \succ (y_i^2, y'_{i^-}) \quad \text{for some} \quad y'_{i^-} \in Y_{i^-} \Rightarrow (y_i^1, y_{i^-}) \succ (y_i^2, y_{i^-}) \quad \text{for all} \quad y_{i^-} \in Y_{i^-}.$$

Theorem 7.1.7

Suppose that there exists an order preserving preference function $u(y_1, \ldots, y_n)$ for the preference structure $(\Pi_{i=1}^n Y_i, \succ)$, namely

$$(y_1, \ldots, y_n) \succ (y'_1, \ldots, y'_n) \Leftrightarrow u(y_1, \ldots, y_n) > u(y'_1, \ldots, y'_n).$$

Then the preference function $u(\cdot)$ is decomposable in terms of the conditional preference function over each attribute; namely, there exists a conditional preference function $u_i(y_i)$ for each Y_i at some point $b_{i^-} \in Y_{i^-}$ $(i = 1, \ldots, n)$ and a strictly increasing† one-to-one function $\Phi\colon R^n \to R$ such that

$$u(y_1, \ldots, y_n) = \Phi(u_1(y_1), \ldots, u_n(y_n)) \tag{7.1.2}$$

if and only if each attribute Y_i is preferentially independent of Y_{i^-}. Moreover, such a function Φ is unique up to an increasing transformation.

Proof

only if Suppose that there exists a conditional preference function $u_i(y_i)$ for each attribute Y_i at some $b_{i^-} \in Y_{i^-}$ and a one-to-one increasing function Φ with the properties of Eq. (7.1.2). Then

$$\begin{aligned}
y_i \succ_{b_{i^-}} y_i' &\Rightarrow u_i(y_i) > u_i(y_i') \\
&\Rightarrow (\alpha^1, \ldots, \alpha^{i-1}, u_i(y_i), \alpha^{i+1}, \ldots, \alpha^n) \geq (\alpha^1, \ldots, \alpha^{i-1}, u_i(y_i'), \alpha^{i+1}, \ldots, \alpha^n) \\
&\qquad\qquad \text{for any} \quad \alpha^k \in R^1 \ (k \neq i) \\
&\Rightarrow (u_1(y_1), \ldots, u_{i-1}(y_{i-1}), u_i(y_i), u_{i+1}(y_{i+1}), \ldots, u_n(y_n)) \\
&\qquad \geq (u_i(y_i), \ldots, u_{i-1}(y_{i-1}), u_i(y_i'), u_{i+1}(y_{i+1}), \ldots, u_n(y_n)) \\
&\qquad\qquad \text{for every} \quad y_{i^-} \in Y_{i^-} \\
&\Rightarrow \Phi(u_1(y_1), \ldots, u_{i-1}(y_{i-1}), u_i(y_i), u_{i+1}(y_{i+1}), \ldots, u_n(y_n)) \\
&\qquad > \Phi(u_1(y_1), \ldots, u_{i-1}(y_{i-1}), u_i(y_i'), u_{i+1}(y_{i+1}), \ldots, u_n(y_n)) \\
&\qquad\qquad \text{for every} \quad y_{i^-} \in Y_{i^-} \\
&\Rightarrow (y_i, y_{i^-}) \succ (y_i', y_{i^-}) \qquad \text{for every} \quad y_{i^-} \in Y_{i^-} \\
&\Rightarrow y_i \succ_{y_{i^-}} y_i' \qquad \text{for every} \quad y_{i^-} \in Y_{i^-}.
\end{aligned}$$

This relationship shows that $\succ_{b_{i^-}}$ is not affected by the level b_{i^-} of Y_{i^-}, which implies that Y_i is preferentially independent of Y_{i^-}.

if Let $u_i(y_i)$ denote the conditional preference function over Y_i at $b_{i^-} \in Y_{i^-}$; that is, $u_i(y_i) = u(b_i, \ldots, b_{i-1}, y_i, b_{i+1}, \ldots, b_n)$. Then we can show the fact that a function $\Phi\colon R^n \to R$ with the form of Eq. (7.1.2) is well defined as follows. Since Y_i is preferentially independent of Y_{i^-}, then $\sim_{b_i}$ is not affected by b_i, whereby we simply denote $\sim_{b_i}$ by $\sim_i$. Then we have

$$\begin{aligned}
a_i \sim_i a_i' \quad (i = 1, \ldots, n) \Leftrightarrow (a_1, a_2, \ldots, a_n) &\sim (a_1', a_2, \ldots, a_n) \\
&\sim (a_1', a_2', a_3, \ldots, a_n) \qquad (7.1.3) \\
&\sim (a_1', a_2', \ldots, a_n'),
\end{aligned}$$

† A function $\Phi\colon R^n \to R$ is said to be strictly increasing if $(\alpha_1, \ldots, \alpha_n) \geq (\alpha_1', \ldots, \alpha_n') \Rightarrow \Phi(\alpha_1, \ldots, \alpha_n) > \Phi(\alpha_1', \ldots, \alpha_n')$.

and hence

$$u_i(a_i) = u_i(a_i') \quad (i = 1, \ldots, n) \Leftrightarrow u(a) = u(a'). \tag{7.1.4}$$

The relation $\Rightarrow$ in Eq. (7.1.4) guarantees the existence of a function Φ with Eq. (7.1.2) and $\Leftarrow$ in Eq. (7.1.4) yields the one-to-one property of Φ. The increasing property and the uniqueness of Φ up to an increasing function are obvious.

7.1.2.2 Decomposed Representation

Theorem 7.1.7 tells nothing about the decomposed representation of multiattribute preference function. Note that in the theorem only order relation on the preference is used, and no information on the preference intensity seems necessary there. We may infer that some additional cardinal information on the preference should be required for getting some representation theorems. This is true, as will be seen below. Especially for two attributes cases, the condition of preference independence is too mild to bring about any information on representation. Surprisingly, however, for cases with more than two attributes, some slight modifications of preference independence produce some information on the preference intensity and, hence, some representation theorem. (Regarding this point see, for example, Krantz *et al.* [K9].) In the following we shall concentrate our consideration on cases with two attributes Y and Z for simplicity and better understanding. For cases with more than two attributes, see Krantz *et al.* [K9], Fishburn [F7], and Keeney and Raiffa [K6].

Cardinal Independence. By reference to Fig. 7.1, we may notice that some particular relationships are needed among conditional preference functions in order to get a decomposed representation of the whole preference function using the limited information of these conditional preference functions. Now recall that preference functions with a cardinal

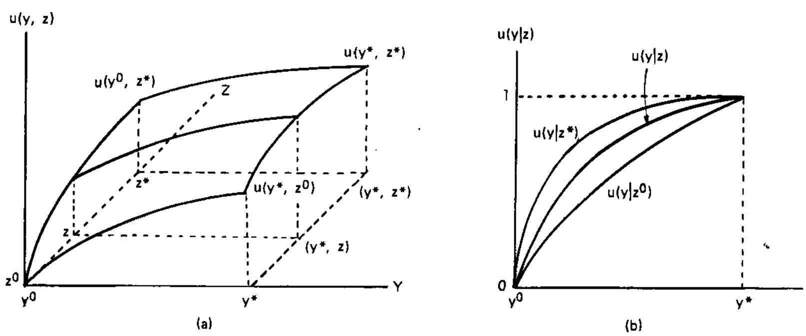

Fig. 7.1. (a) Utility surface and (b) conditional utility curves.

measure are unique up to a positive affine transformation. In other words, a preference function u is equivalent to another preference function v as cardinal measure if

$$u = \alpha + \beta v \qquad \text{for some} \quad \alpha \quad \text{and} \quad \beta > 0. \tag{7.1.5}$$

Consider a particular case in which all conditional preference functions over the attribute Y are equivalent to each other independent of the level of Z, namely, all conditional preference functions $u(\cdot, z)$ over Y become the same if their origins and units are set in common. More formally, let $u(y \mid z)$ be the normalized conditional preference function over Y at z defined by

$$u(y \mid z) = \frac{u(y, z) - u(y^0, z)}{u(y^*, z) - u(y^0, z)} \tag{7.1.6}$$

for given y^0 and $y^* \in Y$ (y^0 and y^* may be set respectively as the worst level and the best level of Y, if any), for which we have $u(y^0 \mid z) = 0$ and $u(y^* \mid z) = 1$ for all $z \in Z$. Then the special case under consideration requires that all these normalized conditional preference functions become identical with each other regardless of the level z. Under such a circumstance, therefore, since these normalized conditional preference functions are not affected by the level $z \in Z$, they can be delegated by $u(y \mid z^0)$. We define $u_1(y) = u(y \mid z^0)$. Then letting $u(y^0, z^0) = 0$, we have

$$k_1 u_1(y) = u(y, z^0), \qquad k_1 = u(y^*, z^0). \tag{7.1.7}$$

Henceforth, for convenience in this book we refer to this special independence condition on the conditional preference functions as the cardinal indpendence.

Definition 7.1.9 (*Cardinal Independence*)

The attribute Y is said to be cardinally independent of the attribute Z, if all conditional preference functions over Y are equivalent to each other as cardinal measure independently of the level of Z.

Theorem 7.1.8

The attributes Y and Z are cardinally independent of each other, if and only if

$$u(y, z) = k_1 u_1(y) + k_2 u_2(z) + k_{12} u_1(y) u_2(z) \tag{7.1.8}$$

Here, for (y^0, z^0) and (y^*, z^*) such that $(y^*, z^0) \succ (y^0, z^0)$ and $(y^0, z^*) \succ (y^0, z^0)$, we normalize each preference function as

(1) $u(y^0, z^0) = 0,\ u(y^*, z^*) = 1,$

(2) $u_1(y^0) = 0,\ u_1(y^*) = 1,$

(3) $u_2(z^0) = 0,\ u_2(z^*) = 1.$

Then we have

(i) $k_1 = u(y^*, z^0)$,
(ii) $k_2 = u(y^0, z^*)$,
(iii) $k_{12} = 1 - k_1 - k_2$.

Proof It has been seen that if the attribute Y is cardinally independent of Z, then Eq. (7.1.7) follows through the normalization Eq. (7.1.6) for each conditional preference function over Y at every $z \in Z$ and $u(y^0, z^0) = 0$. Similarly, if Z is cardinally independent of Y, then for $u_2(z) := u(z \mid y^0)$ with

$$u(z \mid y) = (u(y, z) - u(y, z^0))/(u(y, z^*) - u(y, z^0)) \tag{7.1.9}$$

and $u(y^0, z^0) = 0$, we have

$$k_2 u_2(z) = u(y^0, z), \qquad k_2 = u(y^0, z^*).$$

Note here that from Eq. (7.1.6)

$$u(y, z) = u(y^0, z) + [u(y^*, z) - u(y^0, z)]u_1(y), \tag{7.1.10}$$

and similarly from (7.1.9)

$$u(y, z) = u(y, z^0) + [u(y, z^*) - u(y, z^0)]u_2(z). \tag{7.1.11}$$

Hence, substituting Eq. (7.1.11) with $y = y^*$ into (7.1.10),

$$\begin{aligned} u(y, z) &= k_2 u_2(z) + [k_1 + (1 - k_1)u_2(z) - k_2 u_2(z)]u_1(y) \\ &= k_1 u_1(y) + k_2 u_2(z) + (1 - k_1 - k_2)u_1(y)u_2(z). \end{aligned}$$

The reverse is obvious.

As a particular case, with $k_1 + k_2 = 1$, we have an additive decomposition

$$u(y, z) = k_1 u_1(y) + k_2 u_2(z). \tag{7.1.12}$$

Let us consider the geometric meaning for this case in more detail. Recall that the cardinal independence condition admits the arbitrariness of both the origin and the unit for each conditional preference function. If the preference function $u(y, z)$ has an additively decomposed representation [Eq. (7.1.12)], then it is easily seen that each conditional preference function over Y has a common unit and leaves only a difference of origin. Conversely, if each conditional preference function over Y has a common unit and leaves a difference of origin, then for all $y \in Y$ and $z \in Z$

$$u(y, z) = \alpha(z) + u(y, z^0). \tag{7.1.13}$$

Substituting $u(y^0, z^0) = 0$ into Eq. (7.1.13), we have $\alpha(z) = u(y^0, z)$, whence

$$u(y, z) = u(y^0, z) + u(y, z^0). \tag{7.1.14}$$

Equation (7.1.14) yields Eq. (7.1.12). It should be noted here that if each conditional preference function over Y has only the arbitrariness of origin, then the one over Z automatically has the same property.

The above discussion leads to the following theorem.

Definition 7.1.10 (*Additive Independence*)

The attributes Y and Z are said to be additive independent, if each conditional preference function over either Y or Z has a common unit and leaves only the arbitrariness of origin.

Remark 7.1.9

As stated above, the condition of additive independence is symmetric, whereas the condition of cardinal independence is not so. Namely, even if Y is cardinally independent of Z, Z is not necessarily cardinally independent of Y.

Theorem 7.1.9

The attributes Y and Z are additive independent if and only if

$$u(y, z) = k_1 u_1(y) + k_2 u_2(z).$$

Here, for (y^0, z^0) and (y^*, z^*) such that $(y^*, z^0) \succ (y^0, z^0)$ and $(y^0, z^*) \succ (y^0, z^0)$, we normalize each preference function as

(1) $u(y^0, z^0) = 0,\ u(y^*, z^*) = 1,$
(2) $u_1(y^0) = 0,\ u_1(y^*) = 1,$
(3) $u_2(z^0) = 0,\ u_2(z^*) = 1.$

Then we have

(i) $k_1 = u(y^*, z^0),$
(ii) $k_2 = u(y^0, z^*),$
(iii) $k_1 + k_2 = 1.$

Cardinal Independence in Decision Making under Risk. Now we shall consider practical implications of the stated independence conditions. In the following, we shall use the notation $\hat{Y}$ for representing the set of all probability measures over Y, which also stands for the set of risky outcomes over Y. Recall that the expected preference principle by von Neumann and Morgenstern asserts that in decision making under risk there exists a utility function such that

$$p \succ q \Leftrightarrow \int u(y)\, dP(y) > \int u(y)\, dQ(y).$$

where $P(y)$ and $Q(y)$ represent the probability distribution functions corresponding to the risky outcomes p and q, respectively. Moreover, ar important thing is that such a utility function is unique up to a positive affine transformation. Note that, according to von Neumann–Morgenstern utility theory, utility functions that are equivalent to each other as cardinal measure reflect the same preference order over risky outcomes. Therefore, if the conditional preference over risky outcomes $(\tilde{Y}, z)$ does not depend on the level z, then the conditional utility functions over Y are equivalent as cardinal measure independently of z. As a result, as long as we consider utility functions in the sense of von Neumann–Morgenstern, the condition of cardinal independence can be restated by a preference independence over risky outcomes.

Definition 7.1.11 (*Utility Independence*)

The attribute Y is utility independent of another attribute Z, if the preference over risky prospects $(\tilde{Y}, z)$ does not depend on z; namely, if

$$(\tilde{y}^1, z^0) \succ (\tilde{y}^2, z^0) \qquad \text{for some} \quad z^0 \Rightarrow (\tilde{y}^1, z) \succ (\tilde{y}^2, z) \qquad \text{for all} \quad z \in Z,$$

where $\tilde{y}^1$ and $\tilde{y}^2$ are probability measures on Y.

Theorem 7.1.10

In decision making under risk, the overall utility function $u(y, z)$ is decomposable in the bilinear form in Theorem 7.1.8, if and only if Y and Z are utility independent of each other.

Remark 7.1.10

Fishburn and Keeney [F12] suggested for risky choices a condition called the generalized utility independence without restriction of the positiveness of β in Eq. (7.1.5). Under the condition, we can get a more general multiplicative representation. However, it will be seen later that this is included in the representation form under the more general condition of interpolation independence (Theorem 7.1.16).

As to the additive independence in decision making under risk, Eq. (7.1.5) yields the following expected utility for some risky alternative with the probability distribution $p(y, z)$

$$\begin{aligned}\int_{y^0}^{y^*}\int_{z^0}^{z^*} p(y, z)u(y, z)\, dy\, dz &= \int_{y^0}^{y^*}\int_{z^0}^{z^*} p(y, z)[\alpha(z) + k_1 u_1(y)]\, dy\, dz \\ &= \int_{z^0}^{z^*}\left[\int_{y^0}^{y^*} p(y, z)\, dy\right]\alpha(z)\, dz \\ &\quad + \int_{y^0}^{y^*}\left[\int_{z^0}^{z^*} p(y, z)\, dz\right]k_1 u_1(y)\, dy. \qquad (7.1.15)\end{aligned}$$

Note here that the terms $\int_{y^0}^{y^*} p(y, z)\,dy$ and $\int_{z^0}^{z^*} p(y, z)\,dz$ denote the marginal distribution of $p(y, z)$ on Z and Y, respectively. In addition, it should be noted that in Eq. (7.1.15) only the marginal distribution causes the difference of expected utility for risky alternatives. Now we have the following implication of the additive independence in risky decisions.

Definition 7.1.12 (*Additive Utility Independence*)

In decision making under risk with two attributes Y and Z, Y and Z are said to be additive utility independent, if the preference over $Y \times Z$ is decided by only the marginal distributions of the risky alternatives.

Theorem 7.1.11

In decision making under risk, the overall utility function $u(y, z)$ is decomposable in the additive form of Theorem 7.1.9, if and only if the attributes Y and Z are additive utility independent.

Cardinal Independence in Decision Making under Certainty. We have seen through the above discussion that some cardinal information on the preference is indispensable for additive or bilinear decomposition. How can such cardinal information under certain circumstances be elicited from the decision maker? As a simple example, in particular, we assume the existence of measurable value function defined in the preceding section.

Let Y^* be a nonempty subset of $Y \times Y$ and let $\succ^*$ be a binary relation on Y^* reflecting a weak order of preference difference for $\succ$ on Y. We have seen in Theorem 7.1.4 that some appropriate conditions ensure a measurable value function $u(\cdot)$ such that if $y^1 \succ y^2$ and $y^3 \succ y^4$, then

$$y^1y^2 \succ^* y^3y^4 \Leftrightarrow u(y^1) - u(y^2) > u(y^3) - u(y^4). \tag{7.1.16}$$

In addition, this measurable value function is unique up to a positive affine transformation; namely, two measurable value functions u and v are equivalent (i.e., reflect the same order $\succ^*$), if and only if there exist α and positive β such that

$$u = \alpha + \beta v.$$

Therefore, it can be readily seen that the attribute Y is cardinally independent in the preference order $\succ^*$ of the attribute Z, if and only if the conditional order of $\succ^*$ on Y at z is not affected by z; namely,

$$\begin{aligned} &(y^1, z^0)(y^2, z^0) \succ^* (y^3, z^0)(y^4, z^0) \qquad \text{for some} \quad z^0 \in Z \\ &\quad \Rightarrow (y^1, z)(y^2, z) \succ^* (y^3, z)(y^4, z) \qquad \text{for every} \quad z \in Z. \end{aligned} \tag{7.1.17}$$

In other words, this condition implies that the *order* of the preference difference over Y is not affected by the level of Z.

Definition 7.1.13 (*Weak Difference Independence*)†

The attribute Y is said to be weak difference independent of the attribute Z, if, for any y^1, y^2, y^3, and y^4 over Y, the relation (7.1.17) holds.

As a corollary of Theorem 7.1.8, we have the following theorem.

Theorem 7.1.12‡

In decision making under certainty with two attributes Y and Z, suppose that there exists a measurable value function $u(y, z)$. Then, $u(y, z)$ is decomposable in the bilinear form of Theorem 7.1.8, if and only if the attributes Y and Z are preferentially independent and weak difference independent of each other.

Recall that, in Theorem 7.1.8, the additive decomposition is possible if and only if $k_1 + k_2 = 1$, where $k_1 = u(y^*, z^0)$, $k_2 = u(y^0, z^*)$, $u(y^*, z^*) = 1$, $u(y^0, z^0) = 0$, $(y^*, z^0) \succ (y^0, z^0)$, and $(y^0, z^*) \succ (y^0, z^0)$. Since y^*, y^0, z^*, and z^0 are introduced merely in order to define the unit and the origin, they can be arbitrarily chosen. Hence, letting some y^1, y^2, z^1, and z^2 act for y^*, y^0, z^*, and z^0, respectively,

$$k_1 + k_2 = 1 \Leftrightarrow u(y^1, z^1) - u(y^2, z^1) = u(y^1, z^2) - u(y^2, z^2),$$

which leads to the following theorem.

Theorem 7.1.13

The measurable value function $u(y, z)$ is decomposable in the additive form of Theorem 7.1.9, if and only if, in addition to the assumptions of Theorem 7.1.12, the following holds:

$$(y^1, z^1)(y^2, z^1) \sim^* (y^1, z^2)(y^2, z^2) \tag{7.1.18}$$

for arbitrarily chosen $y^1, y^2 \in Y$ and $z^1, z^2 \in Z$, such that

$$(y^1, z^1) \succ (y^2, z^1) \tag{7.1.19}$$

and

$$(y^0, z^1) \nsim (y^0, z^2) \qquad \text{for some} \quad y^0 \in Y. \tag{7.1.20}$$

† Dyer and Sarin [D10].

‡ Theorems 7.1.12 and 7.1.13 from Dyer and Sarin [D10].

Remark 7.1.11

As a result of Theorem 7.1.13, it should be noted that if, under the assumption of weak difference independence, Eq. (7.1.18) holds for *some* $y^1, y^2 \in Y$ and *some* $z^1, z^2 \in Z$ satisfying Eqs. (7.1.19) and (7.1.20), then Eq. (7.1.18) holds for *all* $y^1, y^2 \in Y$ and *all* $z^1, z^2 \in Z$ satisfying Eqs. (7.1.19) and (7.1.20). In fact, the additive decomposition implies each conditional measurable function has a common unit, and hence the preference difference between two arbitrarily given points over $Y \times z$ is identical independently of z.

Without the assumption of weak difference independence, the additive decomposition can be obtained as follows. Recall that the additive independence implies that all conditional value functions over each attribute have a common unit and only their origins vary depending on levels of another attribute. This implies, in other words, that the difference of the conditional preference function $u(y \mid z)$ between the points (y^1, z) and (y^2, z) is not affected by z, whereby according to Eq. (7.1.16), the difference of the intensity of the conditional preference over Y does not depend on the level of Z.

Definition 7.1.14 (*Difference Independence*)[†]

The attribute Y is said to be difference independent of Z, if for all $y^1, y^2 \in Y$ such that $(y^1, z^0) \succsim (y^2, z^0)$ for some $z^0 \in Z$

$$(y^1, z)(y^2, z) \sim^* (y^1, z^0)(y^2, z^0) \qquad \text{for all} \quad z \in Z.$$

Theorem 7.1.14

Assume that a measurable value function $u(y, z)$ exists for the given preference over $Y \times Z$. Then $u(y, z)$ is decomposable in the additive form if and only if the attributes Y and Z are preferentially independent of each other and Y is difference independent of Z.

Conditional measurable value functions may be calibrated by some methods such as the direct rating method or the direct midpoint method (Fishburn [F6]). Once some conditional measurable value functions are assessed, the weak difference independence condition can be easily checked by comparing their normalized function forms.

However, in cases that the assessment of conditional measurable value function is not easy, or trade-off between one attribute and another can be relatively easily made, the trade-off can be used for eliciting cardinal information on the preference for the additive decomposition.

[†] Dyer and Sarin [D10].

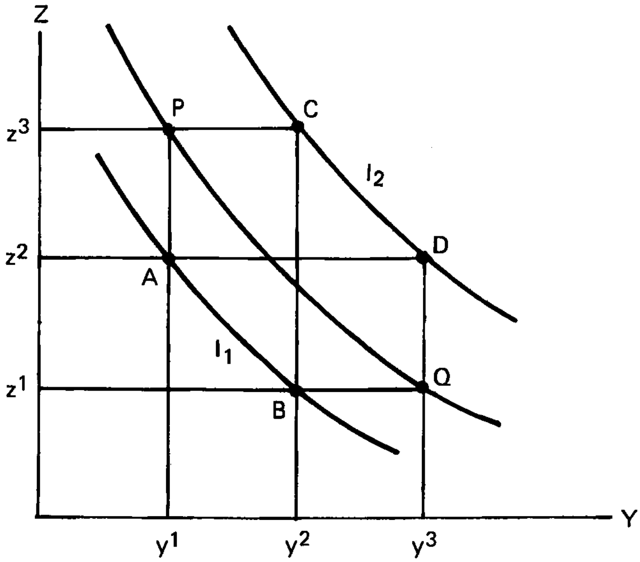

Fig. 7.2. Thomsen condition.

The additive independence condition can be restated in connection with the trade-off between Y and Z as follows. In Fig. 7.2, since the points A and B are on an indifference curve I_1, the difference of the conditional preference over Y between y^1 and y^2 at $z = z^1$ equals that of the conditional preference over Z between z^1 and z^2 at $y = y^1$. These differences of conditional preferences may be represented by Δu_1^1 and Δu_2^1 in terms of measurable value function. A similar argument can be made for the points C and D on the indifference curve I_2. Now the additive independence condition asserts that the difference Δu_1^1 of the conditional preference over Y between y^1 and y^2 is not affected by the level of Z. The same is true for other differences of conditional preferences. Therefore, the difference of conditional preference between the points P and Q can be represented by $\Delta u_1^1 + \Delta u_1^2$ for the attribute Y and $\Delta u_2^1 + \Delta u_2^2$ for the attribute Z. In view of that $\Delta u_1^1 = \Delta u_2^1$ and $\Delta u_1^2 = \Delta u_2^2$, it finally follows that the point P should be indifferent to the point Q. This observation leads to the well-known Thomsen condition.

Definition 7.1.15 (*Thomsen Condition*)†

A weakly ordered set $(Y \times Z, \succ)$ is said to satisfy the Thomsen condition, if for any $y^1, y^2, y^3 \in Y$ and $z^1, z^2, z^3 \in Z$,

$$\left.\begin{array}{l}(y^1, z^2) \sim (y^2, z^1) \\ (y^2, z^3) \sim (y^3, z^2)\end{array}\right\} \Rightarrow (y^1, z^3) \sim (y^3, z^1).$$

† Kranz *et al.* [K9].

Remark 7.1.12

The conditions of preference independence and Thomsen give sufficient power for the additive decomposability of value functions. However, some additional conditions are required for ensuring the existence of such an additively decomposed value function.

Definition 7.1.16 (*Solvability Condition*)

A weakly ordered set $(Y \times Z, \succ)$ is said to satisfy the solvability condition if there exists some $\hat{y} \in Y$ such that

$$(\hat{y}, z_*) \sim (y^1, z^*),$$

for any $y_*, y^*, y^1 \in Y$ and $z_*, z^* \in Z$ with $(y_*, z_*) \precsim (y^1, z^*) \precsim (y^*, z_*)$.

Definition 7.1.17 (*Archimedean Condition*)

A weakly ordered set $(Y \times Z, \succ)$ is said to satisfy the Archimedean condition, if the following holds. If, for any $\bar{z}, \hat{z} \in Z$, the sequence $\{y^1, y^2, \ldots, y^k, \ldots\}$ such that $(y^k, \bar{z}) \sim (y^{k+1}, \hat{z})$ $(\forall k)$ is bounded, then it is finite.

Remark 7.1.13

Recall that the set of real numbers satisfies the continuity property and has a subset countable and dense in it in connection with the structure of order and addition. The condition of solvability and that of Archimedean ensure similar properties for $(Y \times Z, \succ)$. (For details, see Krantz *et al.* [K9].)

Theorem 7.1.15†

Suppose the following conditions hold for $(Y \times Z, \succ)$:

(C1) $\succ$ is a weak order;
(C2) the attributes Y and Z are mutually preferentially independent;
(C3) the Thomsen condition holds;
(C4) the solvability condition holds;
(C5) the Archimedean condition holds.

Then there exists a value function $u: Y \times Z \to R^1$ which is decomposable in an additive form.

† Kranz *et al.* [K9].

Proof Since the precise proof is complicated, we shall sketch an outline of the proofs. Given y, z, y_*, and z_*, let $\zeta(y)$, $\eta(z)$ be solutions such that

$$(y_*, \zeta(y)) \sim (y, z_*), \qquad (\eta(z), z_*) \sim (y_*, z),$$

and let $y^1 \circ y^2 \in Y$ be a solution to

$$((y^1 \circ y^2), z_*) \sim (y^1, \zeta(y^2)).$$

A key idea to the proof is that for any bounded subset of $(Y \times Z, \succ)$, $(Y, \succ_Y, 0)$ is of additive structure; i.e., there exists a function u_1 unique up to a positive affine transformation such that

$$y^1 \succ_Y y^2 \Leftrightarrow u_1(y^1) > u_1(y^2)$$
$$u_1(y^1 \circ y^2) = u_1(y^1) + u_1(y^2).$$

Then, the preference independence condition yields

$$(y^1, z^1) \succ (y^2, z^2) \Leftrightarrow y^1 \circ \eta(z^1) \succ_Y y^2 \circ \eta(z^2),$$

and hence letting $u_2(z) = u_1(\eta(z))$

$$(y^1, z^1) \succ (y^2, z^2) \Leftrightarrow u_1(y^1) + u_2(z^1) > u_1(y^2) + u_2(z^2).$$

The theorem is established by extending u_1 and u_2 to the whole space $(Y \times Z, \succ)$.

For the additive decomposition of value functions, several methods have been developed for testing the Thomsen condition and the assessment of the value function simultaneously. We shall introduce briefly a method called the lock step method based on the above proof (Keeney and Raiffa [K6]) (Fig. 7.3).

(1) Define y^0 and z^0 as the least preferred level of Y and Z, respectively. Set $u(y^0, z^0) = u_1(y^0) = u_2(z^0) = 0$.

(2) Take y^1 in such a way that $(y^1, z^0) \succ (y^0, z^0)$ and set $u_1(y^1) = 1$.

(3) Find z^1 such that $(y^0, z^1) \sim (y^1, z^0)$ (due to the solvability condition) and set $u_2(z^1) = 1$.

(4) Find y^2 and z^2 such that $(y^2, z^0) \sim (y^1, z^1) \sim (y^0, z^2)$ and set $u_1(y^2) = u_2(z^2) = 2$.

(5) Check whether or not $(y^1, z^2) \sim (y^2, z^1)$, which is a part of the Thomsen condition. If the condition holds, go to the next step. Otherwise, we can not have the additive decomposition.

(6) Find y^3 and z^3 such that $(y^3, z^0) \sim (y^2, z^1) \sim (y^1, z^2) \sim (y^0, z^3)$ and set $u_1(y^3) = u_2(z^3) = 3$.

(7) Check whether or not $(y^3, z^1) \sim (y^2, z^2) \sim (y^1, z^3)$.

(8) Continue the same process for some appropriate number of sample points.

(9) Decide the function forms of $u_1(y)$ and $u_2(z)$ by some appropriate curve fitting method and set $u(y, z) = u_1(y) + u_2(z)$.

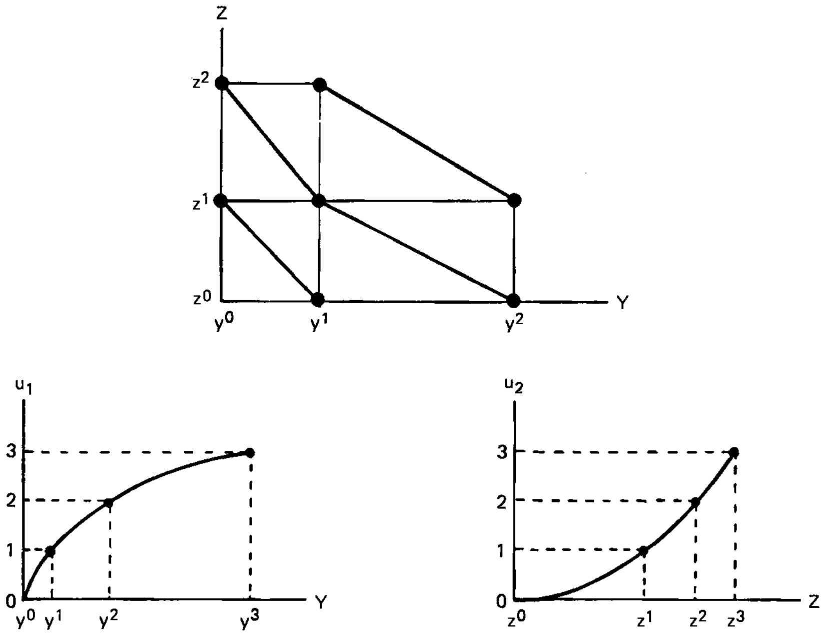

Fig. 7.3. Lock step method.

Interpolation Independence and Convex Dependence. Recall that the conditions of cardinal independence and additive independence impose a severe restriction upon interrelationships among conditional preference functions over each attribute. Now we shall relax the interrelationships. Bell [B4] suggested the condition of interpolation independence, which implies that each conditional preference function over an attribute is expressed by a certain convex combination of two other conditional preference functions over the attribute. A more general formulation, called the convex dependence, was developed by Tamura and Nakamura [T5].

Definition 7.1.18 (*Interpolation Independence*)

Suppose that there exist some values y^0 and y^* of Y and z^0 and z^* of Z such that for all y in Y and z in Z the decision maker strictly prefers (y^*, z) to (y^0, z) and (y, z^*) to (y, z^0). Then Y is said to be interpolation independent of (or 1-convex dependent on) Z, if each normalized conditional preference function $u(y\,|\,z)$ over Y can be written as a convex combination of $u(y\,|\,z^0)$ and $u(y\,|\,z^*)$; that is, if for some $\alpha(z)$

$$u(y\,|\,z) = \alpha(z)u(y\,|\,z^*) + (1 - \alpha(z))u(y\,|\,z^0) \quad \text{for any} \quad y \in Y \quad \text{and} \quad z \in Z. \tag{7.1.21}$$

Remark 7.1.14

As is readily seen, the notion of interpolation independence includes several other independence conditions as special cases. For example, if $u(y \mid z^0) = u(y \mid z^*)$ and $u(z \mid y^0) = u(z \mid y^*)$, the condition that Y is interpolation independent of Z reduces to the condition that Y is cardinally independent of Z.

Note that, conversely, if Y is cardinally independent of Z, then not only is Y interpolation independent of Z, but Z is also interpolation independent of Y.

Theorem 7.1.16†

The attributes Y and Z are interpolation independent of each other if and only if

$$\begin{aligned} u(y, z) = {} & k_1 u(y \mid z^0) + k_2 u(z \mid y^0) - k u(y \mid z^0) u(z \mid y^0) \\ & + (k - k_1) u(y \mid z^0) u(z \mid y^*) + (k - k_2) u(y \mid z^*) u(z \mid y^0) \\ & + (1 - k) u(y \mid z^*) u(z \mid y^*), \end{aligned} \tag{7.1.22}$$

where $k_1 = u(y^*, z^0)$, $k_2 = u(y^0, z^*)$, and k is an independent constant.

Proof

if Suppose that Eq. (7.1.22) holds. Then Eq. (7.1.22) yields

$$u(y^0, z) = k_2 u(z \mid y^0) \qquad \text{and} \qquad u(y^*, z) = k_1 + (1 - k_1) u(z \mid y^*),$$

because $u(y^0 \mid z) = 0$ and $u(y^* \mid z) = 1$ for any $z \in Z$. Hence,

$$\begin{aligned} u(y \mid z) &= (u(y, z) - u(y^0, z))/(u(y^*, z) - u(y^0, z)) \\ &= \frac{\begin{array}{c} \{k_1 - k u(z \mid y^0) + (k - k_1) u(z \mid y^*)\} u(y \mid z^0) \\ + \{(k - k_2) u(z \mid y^0) + (1 - k) u(z \mid y^*)\} u(y \mid z^*) \end{array}}{k_1 + (1 - k_1) u(z \mid y^*) - k_2 u(z \mid y^0)}, \end{aligned}$$

which is of the form of Eq. (7.1.21) showing that Y is interpolation independent of Z. Similarly, it can be shown that Z is interpolation independent of Y.

only if Suppose that Y and Z are mutually interpolation independent. For a special case where $u(y \mid z^0) = u(y \mid z^*)$ and $u(z \mid y^0) = u(z \mid y^*)$, the condition of Y and Z being mutually interpolation independent implies that Y and Z are mutually cardinally independent, which leads to Eq. (7.1.22) with $k = k_2$. Therefore, for the remainder of the proof, we assume that $u(y \mid z^0)$ is not identical to $u(y \mid z^*)$ and that $u(z \mid y^0)$ is not identical to $u(z \mid y^*)$.

† Bell [B4].

From the definition of the normalized conditional preference function over Y with $u(y^*, z^0) = k_1$ and $u(y^0, z^*) = k_2$ [cf. Eq. (7.1.6)], we have

$$u(y, z) = k_2 u(z \mid y^0) + [k_1 - k_2 u(z \mid y^0) + (1 - k_1)u(z \mid y^*)]u(y \mid z).$$

from which the condition that Y be interpolation independent of Z produces

$$u(y, z) = k_2 u(z \mid y^0) + G_2(z)[\alpha_2(z)u(y \mid z^*) + (1 - \alpha_2(z))u(y \mid z^0)], \quad (7.1.23)$$

where

$$G_2(z) = k_1 - k_2 u(z \mid y^0) + (1 - k_1)u(z \mid y^*).$$

Similarly, from the condition that Z be interpolation independent of Y,

$$u(y, z) = k_1 u(y \mid z^0) + G_1(y)[\alpha_1(y)u(z \mid y^*) + (1 - \alpha_1(y))u(z \mid y^0)], \quad (7.1.24)$$

where

$$G_1(y) = k_2 - k_1 u(y \mid z^0) + (1 - k_2)u(y \mid z^*).$$

Fixing z at some $\hat{z} \neq z^0$ in Eqs. (7.1.23) and (7.1.24), we may see that for some constants c_1, c_2, and c_3

$$G_1(y)\alpha_1(y) = c_1 + c_2 u(y \mid z^0) + c_3 u(y \mid z^*).$$

Here $c_1 = 0$ and $c_2 + c_3 = 1 - k_1$ because $G_1(y^0) = k_2$, $G_1(y^*) = 1 - k_1$, $\alpha_1(y^0) = 0$, and $\alpha_1(y^*) = 1$. Finally, using a constant c instead of c_2,

$$G_1(y)\alpha_1(y) = cu(y \mid z^0) + (1 - k_1 - c)u(y \mid z^*).$$

Similarly, for some constant d,

$$G_2(z)\alpha_2(z) = du(z \mid y^0) + (1 - k_2 - d)u(z \mid y^*).$$

Substituting them into Eqs. (7.1.23) and (7.1.24),

$$(k_1 - k_2 + c - d)(u(y \mid z^0) - u(y \mid z^*))(u(z \mid y^0) - u(z \mid y^*)) = 0,$$

whence

$$k_1 + c = k_2 + d.$$

Substituting $G_2(z)\alpha_2(z)$ with $k = k_2 + d$ into Eq. (7.1.23), we finally have

$$\begin{aligned} u(y, z) &= k_2 u(z \mid y^0) + [k_1 - k_2 u(z \mid y^0) + (1 - k_1)u(z \mid y^*)]u(y \mid z^0) \\ &\quad + (u(y \mid z^*) - u(y \mid z^0))[(k - k_2)u(z \mid y^0) + (1 - k)u(z \mid y^*)] \\ &= k_1 u(y \mid z^0) + k_2 u(z \mid y^0) - ku(y \mid z^0)u(z \mid y^0) \\ &\quad + (k - k_1)y(y \mid z^0)u(z \mid y^*) + (k - k_2)u(y \mid z^*)u(z \mid y^0) \\ &\quad + (1 - k)u(y \mid z^*)u(z \mid y^*). \end{aligned}$$

Remark 7.1.15

The decomposed representation of Theorem 7.1.16 clearly includes other representations stated above as special cases. It reduces to the one under bilateral independence (Fishburn [F9]) if $(k_1 + k_2 - 1)k = k_1k_2$, and to the one under the generalized utility independence if $k = 1$. Further, if $u(y \mid z^0) = u(y \mid z^*)$ and $u(z \mid y^0) = u(z \mid y^*)$, it reduces to the one under the cardinal independence, and in addition, if $k_1 + k_2 = 1$, the additive decomposition results.

7.1.2.3 Assessment of Multiattribute Preference Functions

Although we introduced some methods for assessing some of particular multiattribute utility functions, we shall summarize here an effective method based on conditional preference functions using a graphic display terminal of computer. Consider cases with two attributes Y and Z for simplicity.

(1) Set the best level y^* and z^* and the worst level y^0 and z^0 for attributes Y and Z, respectively.

(2) Assess some representative conditional preference functions over each attribute in the normalized form, i.e., in such a way that $u(y^0 \mid z) = 0$ and $u(y^* \mid z) = 1$ for any $z \in Z$ and similarly for $u(z \mid y)$.

(3) Depict graphs of the normalized conditional preference functions over each attribute on the graphic display.

(4) If all normalized conditional preference functions over Y are perceived identical, then it may be considered that the attribute Y is cardinally independent of Z. In addition, if Z is also cardinally independent of Y, then we will have in general the bilinear utility function. In this case, assess the scale constants k_1 and k_2 by asking the trade-off between Y and Z; i.e., in risky choices, ask for the decision maker probabilities α and β such that

$$(y^*, z^0) \sim [(y^*, z^*), (y^0, z^0) : \alpha, 1 - \alpha]$$

$$(y^0, z^*) \sim [(y^*, z^*), (y^0, z^0) : \beta, 1 - \beta],$$

from which $k_1 = \alpha$ and $k_2 = \beta$. If $k_1 + k_2 = 1$, then the utility function reduces to the additive form, otherwise it reduces to the bilinear form, in which the constant k is decided by solving $1 + k = (1 + kk_1)(1 + kk_2)$.

(5) If only Y is cardinally independent of Z, and if Z is not of Y, then Z is interpolation independent of Y (see Remark 7.1.14), and hence both Y and Z are interpolation independent of each other. Go to step (7). If neither Y nor Z is cardinally independent of the other, go to step (6) for the check of interpolation independence.

(6) Set

$$u_1(y:\alpha) = \alpha u(y\,|\,z^*) + (1-\alpha)u(y\,|\,z^0).$$

For a sample level $\hat{z}$, take $\hat{y}$ appropriately (for example, $\hat{y} = (y^0 + y^*)/2$). Find a solution $\hat{\alpha}$ to the following equation

$$u(\hat{y}, \hat{z}) = \alpha u(\hat{y}\,|\,z^*) + (1-\alpha)u(\hat{y}\,|\,z^0).$$

Compare the utility curves of $u(y\,|\,\hat{z})$ and $u_1(y:\hat{\alpha})$. If we observe that these two curves are identical for several levels of Z, then Y may be considered to be interpolation independent of Z. Check for the conditional preference functions over Z in a similar way. Go to step (7) if both Y and Z are interpolation independent of each other, otherwise some other decomposed representation would be possible (for example, nth convex dependence decomposition, Tamura and Nakamura [T5]).

(7) Assess the scale constants k_1 and k_2 in the similar way to step (4). Decide k directly from the utility at some representative point.

Due to recent remarkable developments of micro computers, the assessment method stated above is expected to be useful for practical problems in an interactive way with ease. For cases with more than two attributes, the consistency check of trade-off constants may be a problem. A suggestion using Saaty's method for cardinal ranking (Saaty [S1]) is seen in Sawaragi *et al.* [S5].

7.1.3 Discussions

In the preceding subsections, we have introduced some of the representative theorems regarding existence, uniqueness, and decomposed representation for multiattribute preference functions. These theorems are of a normative theory based on some assumptions. In applying them in practical problems, therefore, it is meaningful to check the properness of these assumptions from some viewpoint of behavioral science and to know the limits of the theorems.

7.1.3.1 Allais Paradox

One of the most famous criticisms for von Neumann–Morgenstern expected utility theory was claimed by Allais (see, for example, Allais and Hagen [A1]). We shall explain it using the example given by Kahneman and Tversky [K1]:

Consider a pair of lotteries L_1 with a gain of 3000 for sure and $L_2 = [4000, 0:0.80, 0.20]$, and another pair of $L_3 = [3000, 0:0.25, 0.75]$

and $L_4 = [4000, 0: 0.20, 0.80]$. According to an experiment by Kahneman and Tversky, many respondents preferred L_1 for the first pair and L_4 for the second pair. Note that their preferences are not consistent with the von Neumann–Morgenstern expected utility theory. In fact, for the preference over the first pair, a von Neumann–Morgenstern utility function $u(\cdot)$ with $u(0) = 0$ produces

$$\begin{aligned} u(3000) &= 1.0u(3000) + 0.0u(0) > 0.8u(4000) + 0.2u(0) \\ &= 0.8u(4000). \end{aligned} \tag{7.1.25}$$

On the other hand, for the preference over the second pair,

$$\begin{aligned} 0.2u(4000) &= 0.2u(4000) + 0.8u(0) > 0.25u(3000) + 0.75u(0) \\ &= 0.25u(3000) \end{aligned} \tag{7.1.26}$$

We have from Eq. (7.1.25) $u(3000)/u(4000) > 0.8$, whereas from Eq. (7.1.26) $u(3000)/u(4000) < 0.8$, which lead to a contradiction.

Now let us consider a reason why such a contradiction for the von Neumann–Morgenstern utility function arise. To this end, observe that the lottery L_3 can be regarded as a compound lottery $[L_1, 0: 0.25, 0.75]$ and also L_4 as $[L_2, 0: 0.25, 0.75]$. The independence condition C2 of Theorem 7.1.6 implies that if we prefer L_1 to L_2, then we must prefer L_3 to L_4, which is violated by the preference stated above. In many other reports also (see, for example, Allais and Hagen [A1] and Machina [M1]), such a violation of the independence condition has been observed.

Kahneman and Tversky considered that the stated preference incompatible with the independence condition of von Neuman–Morgenstern expected utility theory is caused from the *certainty effect*, which implies that people overweight outcomes that are considered certain, relative to outcomes which are merely probable.

Another effect violating the independence condition is observed by the following example. Consider a pair of lotteries $L_5 = [3000, 0: 0.90, 0.10]$ and $L_6 = [6000, 0: 0.45, 0.55]$, and $L_7 = [3000, 0: 0.02, 0.98]$ and $L_8 = [6000, 0: 0.01, 0.99]$. Most people would prefer L_5 to L_6 for the first pair, because although both L_5 and L_6 have the same expected monetary value, the probability of winning for L_5 is much larger than that for L_6. For the second, on the other hand, probabilities of winning for the lotteries are both minuscule. In this situation where winning is not almost probable, most people would prefer L_8 with a large gain to L_7. This effect suggested by Kahneman and Tversky is called the *get-rich-quick effect* in the literature (Sakai [S2]). Regarding L_7 as a compound lottery $[L_5, 0: 1/450, 449/450]$ and L_8 as $[L_6, 0: 1/450, 449/450]$, we may see that the preceding example violates the independence condition of Theorem 7.1.6.

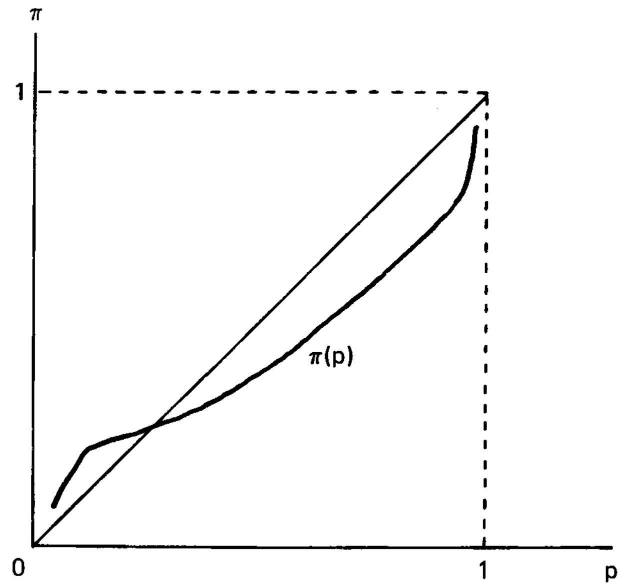

Fig. 7.4. Weighting function for probability.

Observing further effects that violate the expected utility theory, say, the *reflection effect*† and the *isolation effect*‡ Kahneman and Tversky [K1] suggested a weighted value function model (prospect theory) for a simple probability measure p over Y

$$u(p) = \sum_{y \in Y} \pi(p(y))u(y),$$

where $\pi(p)$ is a weighting function for probability with a property of Fig. 7.4. It has been reported that such a modification of the expected utility overcomes several drawbacks in the von Neumann–Morgenstern expected utility theory. For more details, see the original work by Kahneman and Tversy [K1].

7.1.3.2 Group Decisions and Equity

In group decisions, the well-known Arrow's general impossibility theorem ([A2], [F8]) asserts that it is impossible in a general sense (under Arrow's five

† The preference tendency is called the *reflection effect* if we observe the preference over a pair of risky alternatives reverses when all the outcome values are reversed in sign. For example, the risk aversion on the positive domain leads to the risk prone in the negative domain.

‡ A risky alternative can be reconsidered as a compound lottery in various ways; sometimes a preference changes over ways of decomposition. This phenomenon is called the *isolation effect*.

conditions) to aggregate individuals' preference orders into a social one.† Numerous researches have tried to overcome the difficulty pointed by Arrow, but it is considered to be essentially impossible to get out of the stated dilemma. The most controversial subject is directed towards the independence of irrelevant alternatives which is imposed in order to avoid the measurability of preferences. It has been known, however, that even if each individual value (utility) can be evaluated quantitatively, there does not exist any social choice function without interpersonal comparison of values (utilities) (Sen [S7]).

A group Bayesian approach tries to find a group utility function by regarding the group (society) as a man having rationality consistent with the expected utility theory (Raiffa [R1]). Harsanyi [H4] showed that if both the individuals and the group satisfy the conditions of Theorem 7.1.6, and if the group is indifferent between two risky alternatives whenever all individuals are indifferent, then the group utility function should be a linear combination of individual utilities. Note here that such a linearity of the group utility function is mainly due to the independence condition of Theorem 7.1.6. In fact, the proof by Harsanyi proceeds as follows. Under the assumption, the group utility function is a function of the expected utilities of individuals. The group utility function based upon Theorem 7.1.6 is a linear function over the set of risky prospects of individual expected utilities. We shall show this in more detail as follows. Suppose that, for a risky alternative P, each individual i has an expected utility u_i. Hence, we identify P as a prospect $(u_1, \ldots, u_n)$. Similarly, identify O as $(0, \ldots, 0)$. Let Q be an uncertain prospect $[P, O : \alpha, 1 - \alpha]$. Then clearly, $Q = \alpha P + (1 - \alpha)O$. According to Theorem 7.1.5, $u(Q) = \alpha u(P) + (1 - \alpha)u(O)$, whereas $u(Q) = u(\alpha u_1, \ldots, \alpha u_n)$, $u(O) = u(0, \ldots, 0) = 0$ and $u(P) = u(u_1, \ldots, u_n)$, which leads to the first order

† Define a social choice function F by a functional relationship between the tuple of individual preference $(P^1, \ldots, P^n)$ and the social preference P; namely, $P = F(P^1, \ldots, P^n)$. Arrow's general impossibility theorem says that under the assumption that there are at least two persons in the group and three alternatives, there is no social choice function satisfying the following five conditions:

(C1) The social preference P is a weak order.

(C2) The domain of F must include all possible combinations of individual orderings (*unrestricted domain*).

(C3) For any pair x, x' in X, if xP^ix' for all i, then xPx' (*Pareto principle*).

(C4) Let P and P' be the social binary relations determined by F corresponding, respectively, to two sets of individual preferences, $(P^1, \ldots, P^n)$ and $(P^{1\prime}, \ldots, P^{n\prime})$. If for all pairs of alternatives x, x' in a subset S of X, xPx' if and only if xP^ix' for all i, then the greatest set of S with respect to P is identical to the one with respect to P' (*independence of irrelevant alternatives*).

(C5) There is no individual i such that xP^ix' implies that xPx' for every pair of elements x, x' in X (*nondictatorship*).

homogenuity of $u(u_1, \ldots, u_n)$ for $0 \leqq \alpha \leqq 1$. In addition, similar arguments for $O = \alpha R + (1 - \alpha)P$ with $R = (ku_1, \ldots, ku_n)$, where $k = 1 - 1/\alpha$ and $P = \alpha S + (1 - \alpha)O$ with $S = (ku_1, \ldots, ku_n)$, where $k = 1/\alpha$ yield the first order homogenuity of $u(u_1, \ldots, u_n)$ for $k < 0$ and $k > 1$, respectively. The additivity of $u(u_1, \ldots, u_n)$ follows immediately by considering a prospect $S_i = (0, \ldots, 1, \ldots, 0)$ $(i = 1, \ldots, n)$ and $T = [S_1, \ldots, S_n : 1/n, \ldots, 1/n]$.

We have seen that the independence condition of Theorem 7.1.6 plays a key role in deriving the linear group utility function via its linearity over uncertain prospect. However, Diamond [D5] criticized this point. Let us consider a problem with two individuals and two specific alternatives. Let the alternative A be $(u_1 = 1,\ u_2 = 0)$ for sure and B $(u_1 = 0,\ u_2 = 1)$ for sure. Unless we discriminate two individuals, the group should be indifferent between A and B. Now mix these two alternatives with the probability 0.5; that is, $C = [A, A : 0.5, 0.5]$ and $D = [A, B : 0.5, 0.5]$. According to the independence condition (in fact, according to a weaker condition called the *sure thing principle*†), the group should be indifferent between C and D as long as A is indifferent to B. However, D seems more preferable to C in the sense that in D both individuals have a possibility getting the prize, whereas in C the second person has no possibility getting the prize. Therefore, Diamond claimed the group does not accept the expected utility theory.

In addition, Keeney [K4] pointed out that as long as the group utility function is viewed as a function of individual expected utilities, it can not necessarily take equity into account. For example, consider a problem with two individuals and two alternatives: the alternative P results in either $(u_1 = 1,\ u_2 = 0)$ with probability one-half or $(u_1 = 0,\ u_2 = 1)$ with probability one-half. The alternative Q yields either $(u_1 = 1,\ u_2 = 1)$ or $(u_1 = 0,\ u_2 = 0)$, each with probability one-half. For each alternative, both individual expected utilities are 0.5, whence the linear group utility function does not distinguish Q from P. However, from the viewpoint of equity, Q seems more preferable to P. This phenomenon may be considered a natural result from the fact that the group utility function as a function of individual expected utilities is caused from an individualism in the sense that the group is indifferent between two alternatives whenever all individuals are indifferent between them.

As is readily seen in the above discussion, it is hard to construct a group utility function consistent with the expected utility theory and equity. Sen [S7] proposed that the social welfare function should depend not only on the mean value of individual utilities but also on some measure of inequality (dispersion) in these individual utilities. Based on a similar idea, Nakayama

† For any $p, q \in \tilde{Y}$, $p \sim q \Rightarrow \alpha p + (1 - \alpha)r \sim \alpha q + (1 - \alpha)r$ for any $r \in \tilde{Y}$ and $\alpha \in (0, 1)$.

et al. [N10] suggested the extended contributive rule (ECR) method by combining the utilitarian index (linear combination of individual utilities) and a certain egalitarian index (variance of individual utilities), whose aim is to help mutual understanding of the group members toward a consensus rather than to get a social welfare index. Some kinds of utility functions that incorporate equity are discussed in Keeney [K5], Bodily [B5] and others. As to group decisions in terms of parametric domination structures, see Tanino *et al.* [T12, T13], and from a viewpoint of fuzzy set theory, see Tanino [T8].

7.1.3.3 Measurability of von Neumann–Morgenstern Utility Function

It is readily seen that the measurability of von Neumann–Morgerstern utility functions strongly depends on the independence condition of Theorem 7.1.5. From a viewpoint of behavioral science, however, we have observed that the independence condition may be violated in many practical cases, where a sure prospect cannot necessarily be considered an uncertain prospect with the probability one. Therefore, a utility function obtained by some measurement methods based on the certainty equivalent does not necessarily reflect the preference correctly. More strongly, several researchers such as Arrow [A3], Baumol [B2], Luce and Raiffa [L12], and Dyer and Sarin [D9] argued that von Neumann–Morgenstern utility functions do not measure the intensity of preference, though they evaluate the decision maker's attitude toward risk. Recently, some trials of evaluating the preference intensity and the risk attitude separately have been reported (Dyer and Sarin [D10] and Sarin [S4]). In the previous section, we discussed a direct measurement of preference intensity and decomposed representation of measurable multiattribute value functions.

7.1.3.4 Utility Functions Nonlinear in the Probability

We may see that the expected utility theory by von Neumann–Morgenstern is in a sense a linear model for the preference in risky choices. Of course, the real world is generally nonlinear. It depends on the nonlinearity whether or not the expected utility theory fits to the reality as a first-order approximation. It would still remain valid for many practical problems keeping the stated limits in mind. Finally, we would like to mention that some devices for expected utility theory without the independence condition have been developed recently (Fishburn [F10, F11] and Machina [M1]), which are nonlinear in probability while keeping its measurability. For a review of recent developments of utility and value theory, see Farquhar [F3].

7.2 Stochastic Dominance

We have observed in the previous sections that the expected utility theory plays a major role in risky decision making problems, while at the same time, there are some difficulties in assessing utility functions. The theory of stochastic dominance has been developed for ranking risky alternatives without a complete knowledge on the utility function.

In the following, we suppose that our risky decisions follow the expected utility theory by von Neumann–Morgenstern, but we know only partial information on the utility function of the decision maker. For simplicity, we identify risky alternatives as cumulative probability distributions over a positive scalar attribute Y. Hence, for a risky alternative F, the expected utility can be represented by

$$E(u, F) = \int_0^\infty u(y)\, dF(y).$$

Definition 7.2.1

Consider utility functions that are bounded and appropriately smooth on $[0, \infty)$. Let u', u'', and u''' denote the first, second, and third derivatives of the utility function u, respectively. Then the classes of utility functions $\mathscr{U}_1$, $\mathscr{U}_2$, and $\mathscr{U}_3$ are defined as

(i) $\mathscr{U}_1 := \{u \mid u \in C^1,\ u'(y) > 0,\ \forall y \in [0, \infty)\}$.
(ii) $\mathscr{U}_2 := \{u \mid u \in C^2,\ u \in \mathscr{U}_1,\ u''(y) < 0,\ \forall y \in [0, \infty)\}$.
(iii) $\mathscr{U}_3 := \{u \mid u \in C^3,\ u \in \mathscr{U}_2,\ u'''(y) > 0,\ \forall y \in [0, \infty)\}$.

Remark 7.2.1

As is readily recognized, $\mathscr{U}_1$ is the class of utility functions for which the decision maker's preference is strictly increasing over outcomes. Furthermore, $\mathscr{U}_2$ is the class of utility functions which are of $\mathscr{U}_1$ and risk averse. Finally, it is readily seen that decreasing risk averse utility functions belong to the class $\mathscr{U}_3$.

Definition 7.2.2 (*Stochastic Domination*)

For $i = 1, 2, 3,$

$$F >_i G \quad \text{if and only if} \quad E(u, F) > E(u, G), \qquad \forall u \in \mathscr{U}_i.$$

We refer to $>_1$ as the first-degree stochastic dominance (FSD) to $>_2$ as second-degree stochastic dominance (SSD) and $>_3$ as third-degree stochastic dominance (TSD).

These domination relationships induce some kinds of partial orderings over risky prospects. Therefore, although stochastic dominance may not, in general be expected, to lead to the best alternative by itself, it may be helpful to narrow down the alternative set.

Definition 7.2.3

Let $\mathscr{P}$ be the set of all right-continuous distribution functions with $F(0) = 0$. For each $F \in \mathscr{P}$, we define F^n as

$$F^1(y) := F(y),$$
$$F^{n+1} := \int_0^y F^n(y)\,dy, \qquad y \in [0, \infty).$$

Remark 7.2.2

The right continuity of F and G of $\mathscr{P}$ implies that $F^2 \neq G^2$ when $F \neq G$. Therefore, $F = G$ follows whenever $F^n = G^n$ for some $n \geqq 2$.

Denote by $G(y) \geq F(y)$, $\forall y \in [0, \infty)$ that $G(y) \geqq F(y)$, $\forall y \in [0, \infty)$ and $G(y) > F(y)$ for at least some interval of $[0, \infty)$. Then, the first two parts of the following theorem are due to Hadar and Russell [H1], and the last part is due to Whitmore [W10]. (See also Fishburn and Vickson [F13].)

Theorem 7.2.1

The following hold:

(i) $F >_1 G \Leftrightarrow G^1(y) \geq F^1(y)$ for all $y \in [0, \infty)$,
(ii) $F >_2 G \Leftrightarrow G^2(y) \geq F^2(y)$ for all $y \in [0, \infty)$,
(iii) $F >_3 G \Leftrightarrow \mu_F \geqq \mu_G$ and $G^3(y) \geq F^3(y)$ for all $y \in [0, \infty)$,

where μ_F denotes the mean value of F given by

$$\mu_F := \int_0^\infty y\,dF(y).$$

Proof Let $D^i = F^i - G^i$ $(i = 1, 2, 3)$. Then since $u(y)$ is bounded over $[0, \infty)$,

$$\begin{aligned} E(u, F) - E(u, G) &= \int_0^\infty u(y)\,dD^1(y) \\ &= [u(y)D^1(y)]_0^\infty - \int_0^\infty u'(y)\,dD^2(y) \\ &= -\int_0^\infty u'(y)\,dD^2(y). \end{aligned}$$

from which the part $\Leftarrow$ for $i = 1$ is clear.

The part $\Rightarrow$ for $i = 1$ is established by leading a contradiction for a utility function

$$u(y) = \begin{cases} \alpha \exp[\alpha^{-1}(y - y_0)], & \text{for} \quad y \leqq y_0, \\ \alpha + \{1 - \exp[-\alpha(y - y_0)]\}/\alpha, & \text{for} \quad y \geqq y_0, \end{cases}$$

if there exists $y_0 \in [0, \infty)$ such that $F(y) \geqq G(y)$ for $y \leqq y_0$, and $F(y) \leqq G(y)$ for $y_0 \leqq y$.

Similarly, setting $u^{(i)} = d^i u/dy^i$, the part $\Leftarrow$ for $i = 2$ and 3 follows from

$$\int_0^\infty u^{(i-1)}(y) dD^i(y) = [u^{(i-1)}(y)D^i(y)]_0^\infty - \int_0^\infty u^{(i)}(y)\, dD^{i+1}(y)$$

$$= -\int_0^\infty u^{(i)}(y)\, dD^{i+1}(y),$$

because for distributions F and G with finite mean values, $D^i(y)$ $(i = 1, 2)$ is finite, and $\lim_{y\to\infty} u^{(i)}(y) = 0$ $(i = 1, 2)$ is due to the monotonical inceasingness and boundedness from above. In addition, $\Rightarrow$ is derived in a similar fashion by a contradiction using a utility function for $>_2$

$$u(y) = -(1/\alpha^2)\exp[-\alpha(y - y_0)],$$

and for $>_3$

$$u(y) = \begin{cases} -\frac{1}{2}(y - y_0)^2, & y \leqq y_0 \\ 0, & y > y_0. \end{cases}$$

The necessity of $\mu_F \geqq \mu_G$ for $>_3$ is obtained by a contradiction using a utility function

$$u(y) = ky - 1 + \exp[-ky],$$

which yields $E(u, F) = -1 + k\mu_F + o(k)$ where $o(k)/k \to 0$ as $k \downarrow 0$. For more details, see Fishburn and Vickson [F13].

Remark 7.2.3

Theorem 7.2.1 implies that if the cumulative distribution F is pointwise smaller than G, then F is preferred to G in terms of the expected utilities as long as the decision maker's preference is monotonically increasing over outcomes (Fig. 7.5). However, in the case of F and G intersecting at a point, $>_1$ gives no information on the domination between F and G. In such a case, if F^2 is pointwise smaller than G^2, then F is preferred to G in terms of expected utilities as long as the decision maker's preference is monotonically

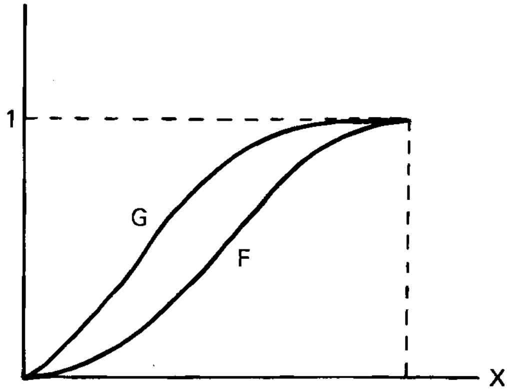

Fig. 7.5. $F >_1 G$.

increasing over outcomes and risk averse (Fig. 7.6). Finally, in the case of F^2 and G^2 intersecting at a point, if the decision maker's preference is decreasing risk averse, we have a domination between F and G by checking each mean value together with F^3 and G^3.

Remark 7.2.4

As is well known, EV-dominance, for which alternatives with larger expected value and smaller variance are preferred, cannot be necessarily adequate for decision problems, if we take the preference of the decision maker into account. (See, for example, Keeney and Raiffa [K6].) Moreover, if $\mu_F = \mu_G$, then $F >_2 G$ ($F >_3 G$) implies $\sigma_F < \sigma_G$ (resp. $\sigma_F \leqq \sigma_G$). Clearly, $\mu_F = \mu_G$ and $F >_1 G$ are inconsistent. On the other hand, if $\mu_F > \mu_G$, then $F >_i G$ ($i = 1, 2, 3$) is neither necessary nor sufficient for $\sigma_F \leqq \sigma_G$. In view of

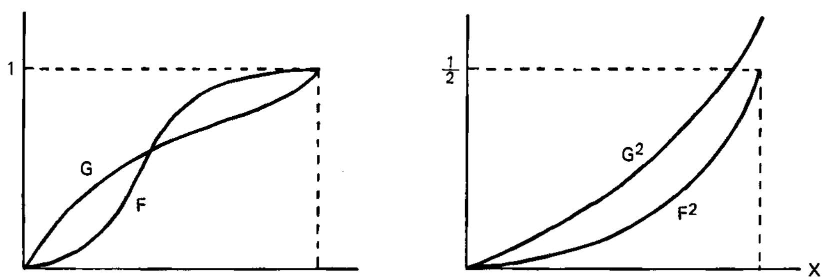

Fig. 7.6. $F >_2 G$.

these facts, stochastic dominance seems to be sufficiently adequate for reducing the alternative set in the light of decision maker's preference.

7.2.1 Multivariate Stochastic Dominance

Now we consider stochastic dominance for decision problems with multiple attributes, in particular, with two attributes $Y \times Z$ for simplicity.

Definition 7.2.4 (*Multivariate Risk Aversion*)†

If for any y_1, y_2, z_1, and z_2 with $y_1 < y_2$ and $z_1 < z_2$ a decision maker prefers the risky alternative which gives an even chance for (y_1, z_2) or (y_2, z_1) to the one which gives an even chance for (y_1, z_1) or (y_2, z_2), then his preference is said to be multivariate risk averse.

Lemma 7.2.1†

Let the utility function of the decision maker be smooth enough. Then the decision maker's preference is multivariate risk averse if and only if

$$\partial^2 u(y, z)/\partial y\, \partial z < 0 \qquad \text{for all} \quad y \quad \text{and} \quad z.$$

Proof The lemma follows immediately from

$$u(y_1, z_2) + u(y_2, z_1) - u(y_1, z_1) - u(y_2, z_2) = -\int_{y_1}^{y_2}\int_{z_1}^{z_2} u_{yz}(y, z)\, dz\, dy.$$

The following theorem is originally due to Levy and Paroush [L4].

Theorem 7.2.2

Let $F(y, z)$ and $G(y, z)$ be two cumulative probability distribution functions absolutely continuous on R_+^2 with $F_1(y)$, $F_2(z)$ and $G_1(y)$, $G_2(z)$ as marginals. In addition, suppose that the utility functions under consideration $u(y, z)$ are smooth enough, and $\lim_{y\to\infty} u(y, z)$ and $\lim_{z\to\infty} u(y, z)$ exist for any $(y, z) \in R_+^2$. Then if the decision maker's preference is increasing over outcomes,

$$F >_{\text{MRA}} G \qquad \text{if and only if} \quad F(y, z) \le G(y, z) \qquad \text{for all} \quad (y, z) \in R_+^2.$$

Here, the notation $>_{\text{MRA}}$ represents a stochastic dominance for the class of utility functions that are multivariate risk averse.

† Definition 7.2.4 and Lemma 7.2.1 are from Richard [R2].

Proof

if Letting $D(y, z) = F(y, z) - G(y, z)$, $D_{yz} = \partial^2 D/\partial y\, \partial z$, $u_y = \partial u/\partial y$, etc.,

$$\begin{aligned} E(u, F) - E(u, G) &= \int_0^\infty \int_0^\infty u(y, z) D_{yz}(y, z)\, dy\, dz \\ &= \int_0^\infty [u(y, z) D_z(y, z)]_{y=0}^{y=\infty}\, dz \\ &\quad - \int_0^\infty \int_0^\infty D_z(y, z) u_y(y, z)\, dy\, dz. \end{aligned}$$

Set $d(y, z) = f(y, z) - g(y, z)$, where $f(y, z)$ and $g(y, z)$ are the density functions associated with $F(y, z)$ and $G(y, z)$, respectively. Then

$$d_2(z) := \int_0^\infty d(y, z)\, dy = \lim_{y\to\infty} \int_0^y d(t_1, z)\, dt_1 = \lim_{y\to\infty} D_z(y, z)$$

and $D_2(z) = \int_0^z d_2(t_2)\, dt_2 = F_2(z) - G_2(z)$. Therefore,

$$\begin{aligned} \text{first term} &= \int_0^\infty \left\{ \lim_{y\to\infty} u(y, z) d_2(z) \right\} dz \\ &= \lim_{y\to\infty} \int_0^\infty u(y, z) d_2(z)\, dz \\ &= -\lim_{y\to\infty} \int_0^\infty D_2(z) u_z(y, z)\, dz. \end{aligned}$$

Similarly,

$$\begin{aligned} d_1(y) &= \lim_{z\to\infty} D_y(y, z) \text{ and } D_1(y) = \int_0^y d_1(t_1)\, dt_1 = \int_0^y \lim_{z\to\infty} D_y(t_1, z)\, dt_1 \\ &= \lim_{z\to\infty} \int_0^y D_y(t_1, z)\, dt_1 = \lim_{z\to\infty} D(y, z). \end{aligned}$$

Hence,

$$\begin{aligned} \text{second term} &= -\int_0^\infty \int_0^\infty D_z(y, z) u_y(y, z)\, dy\, dz \\ &= -\int_0^\infty [D(y, z) u_y(y, z)]_{z=0}^{z=\infty}\, dy \\ &\quad + \int_0^\infty \int_0^\infty D(y, z) u_{yz}(y, z)\, dy\, dz \\ &= -\lim_{z\to\infty} \int_0^\infty D_1(y) u_y(y, z)\, dy \\ &\quad + \int_0^\infty \int_0^\infty D(y, z) u_{yz}(y, z)\, dy\, dz. \end{aligned}$$

Since $u_y(y, z) > 0$, $u_z(y, z) > 0$ and $u_{yz}(y, z) < 0$ for all $(y, z) \in R^2_+$, $F >_{\mathrm{MRA}} G$ holds as long as $D(y, z) < 0$ for any $(y, z) \in R^2_+$.

only if Provided that $F(y, z) \geq G(y, z)$ for some $(\bar{y}, \bar{z})$, then a contradiction results from using a utility function

$$u(y, z) = \begin{cases} y + z & \text{for all } (y, z) \text{ such that either } y < \bar{y} \text{ or } z < \bar{z} \\ y + z + \alpha yz & \text{for all } (y, z) \text{ such that both } \bar{y} \leqq y < \bar{y} + h \\ & \qquad\qquad\qquad \text{and } \bar{z} \leqq z < \bar{z} + h, \\ y + z + K & \text{otherwise,} \end{cases}$$

where h is a sufficiently small positive scalar, and $K = (\bar{y} + h) + (\bar{z} + h) + \alpha(\bar{y} + h)(\bar{z} + h)$.

Corollary 7.2.1†

If the utility function $u(y, z)$ is of an additive form:

$$u(y, z) = k_1 u_1(y) + k_2 u_2(z),$$

then for the dominance $>_{\mathrm{MFSD}}$ for the class of increasing utility functions

$$F >_{\mathrm{MFSD}} G \qquad \text{if and only if} \quad \begin{cases} F_1(y) \leq G_1(y) & \forall y \in (0, \infty) \\ F_2(z) \leq G_2(z) & \forall z \in (0, \infty). \end{cases}$$

Remark 7.2.5

If F and G are probability independent, the preference between F and G depends only on their marginal distributions. Then the utility function becomes of an additive form (ref. Theorem 7.1.1). Therefore, if F and G are probability independent, then $>_{\mathrm{MFSD}}$ can be judged based only on their marginals.

7.2.2 A Method-Via-Value Function

In practical assessment of multiattribute utility functions, decision makers tend to be confused with judging their preference based on lotteries. For these cases, we can make use of multiattribute value functions. Measurable value functions are assessed on the basis of information on the intensity of preference without using any lotteries. Therefore, value functions themselves do not represent decision maker's attitude toward risk and hence their direct applications are limited to deterministic (nonrisky) problems. However, by

† Levy and Paroush [L4].

making use of value functions as inputs to utility functions, several merits appear in risky situations (Keeney and Raiffa [K6], and Bodily [B15]). Under the circumstance, when we have only partial knowledge about a multiattribute value function $v(y_1, \ldots, y_p)$ and a utility function over the value $u(v)$, stochastic dominance is introduced in a similar way as in the single attribute cases.

Definition 7.2.5

Letting $v(0, 0) = 0$ and $u(0) = 0$,

$$\mathscr{U}_1^* := \{u = u(v) \mid u' > 0,\ \partial v/\partial y > 0,\ \partial v/\partial z > 0,\ \forall (y, z) \in R_+^2\}$$

$$\mathscr{U}_2^* := \{u = u(v) \mid u \in \mathscr{U}_1^*,\ u'' < 0,\ \partial^2 v/\partial y^2 < 0,\ \partial^2 v/\partial z^2 < 0,\ \forall (y, z) \in R_+^2\}.$$

Remark 7.2.6

It is well known that the condition of $\partial^2 v/\partial y^2 < 0$ and $\partial^2 v/\partial z^2 < 0$ means decreasing marginal value. Therefore, when the value v is increasing over outcomes and its marginal value of each attribute is decreasing, then risk aversion utility functions over v forms the class $\mathscr{U}_2^*$.

Define cumulative distributions over $Y = Y \times Z$ as

$$\bar{F}(v) := \mathrm{Prob}[v(y, z) \leqq v] = \int_0^\infty F_{2|1}(v_2(v : y) \mid y)\, dF_1,$$

where $v_2(v : y)$ is the solution of z to $v(y, z) = v$. Then the expected utility can be written as

$$\begin{aligned} E(u(y, z), F(y, z)) &= \int_0^\infty \int_0^\infty u(y, z)\, d^2F(y, z) \\ &= \int_0^{v_\infty} u(v)\, d\bar{F}(v). \end{aligned}$$

Definition 7.2.6

Define a multivariate stochastic dominance $>_i^*$ for $i = 1, 2$ as

$$F(y, z) >_i^* G(y, z) \Leftrightarrow E(u(y, z), F(y, z)) > E(u(y, z), G(y, z)), \qquad \forall u \in \mathscr{U}_i^*.$$

Note here that $F(y, z) >_i^* G(y, z)$ is equivalent to $\bar{F}(v) >_i \bar{G}(v)$. Then we have the following theorem, in which the first part is due to Huang *et al.* [H13], and the second is due to Takeguchi and Akashi [T1]. For the proof, see their original papers.

Theorem 7.2.3

(i) Suppose that

(a) $G_1(y) > F_1(y)$ for all $y \in (0, \infty)$,
(b) $\partial G_{2|1}(z \mid y)/\partial y < 0$ for all $(y, z) \in R_+^2$,
(c) $G_{2|1}(z \mid y) > F_{2|1}(z \mid y)$ for all $(y, z) \in R_+^2$.

Then we have

$$F(y, z) >_1^* G(y, z).$$

(ii) Suppose in place of (c) above

(c′) $\int_0^z G_{2|1}(t \mid y)\, dt > \int_0^z F_{2|1}(t \mid y)\, dt$ for all $(y, z) \in R_+^2$.

Then under the conditions (a), (b) and (c′),

$$F(y, z) >_2^* G(y, z).$$

7.2.3 Discussion

In order for the stochastic dominance theory to be valid for practical problems, it is necessary to know the probability distributions of alternatives, which might seem too restrictive. However, recall that the expected utility theory requires information on both the probability distribution and the utility function. Since the examination of multivariate stochastic dominance is rather difficult in practice, it would be expected to be especially effective for single attribute cases with many alternatives. In fact, the stochastic dominance theory has been extensively developed in portfolio selections. Another application to equal distribution of public risk would be possible: As is well known, risk measurement in von Neumann–Morgenstern utility functions is also applicable to inequality measurement. (See, for example, Atkinson [A6], and Bodily [B15].) A discussion on this topic is seen in Nakayama *et al.* [N12]. As to the statistic dominance for problems with incomplete information on the state of nature, see Fishburn [F5], and regarding the statistic–stochastic dominance, see Takeguchi and Akashi [T2].

7.3 Multiobjective Programming Methods

Utility/value analysis tries to represent the decision maker's preference as a utility/value function. To this aim, as was seen in the previous sections, some restrictive assumptions are imposed on the decision maker's preference structure. However, if our aim is mainly to find a solution to the given problem, it is not always necessary to grasp the global preference structure of the decision maker. Interactive programming methods have been developed for obtaining a solution by utilizing only partial information about the decision maker's preference. We now discuss several kinds of interactive programming methods. However, our aim is not merely to introduce those methods but to discuss their practical implications. For surveys of more methods see, for example, Hwang and Masud [H17], Chankong and Haimes [C6].

7.3.1 Goal Programming

The term *goal programming* was originally proposed by Charnes and Cooper [C7]. Its purpose is to find a solution as close as possible to a given goal or target that is, in some cases, unattainable (but not necessarily so). Let $\bar{f} = (\bar{f}_1, \bar{f}_2, \ldots, \bar{f}_p)$ denote such a goal. Then the general formulation of goal programming is given by

$$\text{(GP)} \qquad \text{Minimize} \quad d(f(x), \bar{f}) \quad \text{subject to} \quad x \in X, \tag{7.3.1}$$

where $d(\cdot)$ is a distance function associated with a chosen metric.

Lemma 7.3.1

Let y^+ and y^- be vectors of R^p. Then consider the following problem:

$$\text{(GP}') \qquad \begin{aligned} \text{Minimize} \quad & g(y^+, y^-) \\ \text{subject to} \quad & f(x) + y^+ - y^- = \bar{f} \\ & y^+, y^- \geqq 0 \\ & x \in X. \end{aligned}$$

Suppose that the function g is monotonically increasing with respect to elements of y^+ and y^- and strictly monotonically increasing with respect to at least either y_i^+ or y_i^- for each i $(1 \leqq i \leqq p)$. Then, the solution $\hat{y}^+$ and $\hat{y}^-$ to the preceding problem satisfy

$$\hat{y}_i^+ \hat{y}_i^- = 0, \qquad i = 1, 2, \ldots, p. \tag{7.3.2}$$

Proof Suppose, to the contrary, that the solution $\hat{y}^+$, $\hat{y}^-$, and $\hat{x}$ to the problem under consideration do not satisfy the relation (7.3.2), then for some

i $(1 \leqq i \leqq p)$ both $\hat{y}_i^+$ and $\hat{y}_i^-$ are positive. Hence there exists a positive number δ such that

$$\tilde{y}_i^+ := \hat{y}_i^+ - \delta > 0$$
$$\tilde{y}_i^- := \hat{y}_i^- - \delta > 0.$$

Letting $\tilde{y}^+$ denote the vector such that

$$\tilde{y}_j^+ = \begin{cases} \hat{y}_j^+ & j \neq j, \\ \tilde{y}_i^+ & j = i \end{cases}$$

and $\tilde{y}^-$ be defined in a similar way, we have

$$f(\hat{x}) + \tilde{y}^+ - \tilde{y}^- = \bar{f}.$$

It follows, then, from our assumption that

$$g(\tilde{y}^+, \tilde{y}^-) < g(\hat{y}^+, \hat{y}^-),$$

which contradicts the fact that $\hat{y}^+$ and $\hat{y}^-$ are the minimal solution to the given problem.

Remark 7.3.1

According to Lemma 7.3.1, for solutions $\hat{y}^+$, $\hat{y}^-$, and $\hat{x}$ to the minimization problem in Lemma 7.3.1, we have

$$\hat{y}_i^+ = \bar{f}_i - f_i(\hat{x}), \qquad \text{if } f_i(\hat{x}) \leqq \bar{f}_i,$$
$$\hat{y}_i^- = f_i(\hat{x}) - \bar{f}_i, \qquad \text{if } f_i(\hat{x}) \geqq \bar{f}_i.$$

Therefore, if less of the level of the ith criterion is preferred to more, then $\hat{y}_i^+$ and $\hat{y}_i^-$ represent the overattainment and the underattainment of the ith criterion, respectively.

Remark 7.3.2

The following particular forms of $g(\cdot)$ are usually used in many practical cases:

(i) $g(y^+, y^-) = \sum_{i=1}^{p} (y_i^+ + y_i^-)$

(ii) $g(y^+, y^-) = \sum_{i=1}^{p} y_i^+$

(iii) $g(y^+, y^-) = \sum_{i=1}^{p} y_i^-$

In case (i), the function g becomes identical to $\|\bar{f} - f(x)\|_1$. This is often used when the most desirable solution $\hat{x}$ should satisfy $f(\hat{x}) = \bar{f}$. On the other hand, when less of each objective function is preferred to more, case (iii) may occur. Note that case (ii) does not make sense under such a circumstance. In case (iii), if there exists a feasible solution $x \in X$ such that $f(x) \leqq \bar{f}$, the optimal value of the problem (GP′) becomes 0. Therefore, the problem (GP′) with the objective function g with case (iii) is equivalent to

$$\text{Find} \qquad x \in X \qquad \text{such that} \quad f(x) \leqq \bar{f}.$$

Also in case (iii), if there exists no feasible solution $x \in X$ such that $f(x) \leqq \bar{f}$, then the problem (GP′) reduces to $\mathrm{Min}\|f(x) - \bar{f}\|_1$ over X. In such a circumstance, it is better to consider $\bar{f}$ as the aspiration level rather than the goal. For cases in which more of each objective function is preferred to less, a similar discussion is possible for case (ii). Throughout the preceding consideration, it should be noted that the goal programming can treat not only optimization but also satisficing.

Remark 7.3.3

If the goal $\bar{f}$ is unattainable, the distance between $f(x)$ and $\bar{f}$ can be considered to represent a measure of regret resulted from unattainability of $f(x)$ to $\bar{f}$. As a more general distance function, Charnes and Cooper suggested using Minkowski's metric

$$d(f(x), \bar{f}) = \sum_{i=1}^{p} (\mu_i|\bar{f}_i - f_i(x)|^{\alpha})^{1/\alpha}, \qquad \mu \geq 0.$$

In particular,

$$\text{for } \alpha = 1 \qquad d(f(x), \bar{f}) = \sum_{i=1}^{p} \mu_i|f_i(x) - \bar{f}_i| \qquad (l_1\text{-norm})$$

$$\text{for } \alpha = 2 \qquad d(f(x), \bar{f}) = \sum_{i=1}^{p} (\mu_i|f_i(x) - \bar{f}_i|^2)^{1/2} \qquad (l_2\text{-norm})$$

$$\text{for } \alpha = \infty \qquad d(f(x), \bar{f}) = \operatorname*{Max}_{1 \leqq i \leqq p} \mu_i|f_i(x) - \bar{f}_i| \qquad (l_\infty\text{-norm or Tchebyshev norm}).$$

Several years later, the solution that minimizes such a regret was resuggested as a compromise solution (Zeleny [Z3]). Although there is no essential difference between them, the terminology of goal programming has been used mainly for problems with the formulation stated in Remark 7.3.2, whereas that of the compromise solution is used for the formulation with more general distance functions. For more details and extensions, refer to

Charnes and Cooper [C8], Hwang and Masud [H17], Ignizio [I1, I2], and Lee [L1].

7.3.3.1.1 Discussion. There have been many applications of goal programming to practical problems (see, for example, Charnes and Cooper [C8], Hwang and Masud [H17], Ignizio [I1], and Lee [L1]). However, in order that decision makers accept the solution obtained by the method, it should reflect the decision makers' value judgment. Although goal programming is recognized to be very easy to handle, it has the following two major drawbacks:

(i) How do we choose the distance function?
(ii) How do we decide the weighting vector $(\mu_1, \ldots, \mu_p)$?

We may notice that the distance function is not necessarily required to approximate the decision maker's preferences globally, but it should be so at least locally near the solution. In many applications so far, the l_1-distance (or either y^+ or y^- in some cases) has been used as the distance function. As a result, in order to get better approximation of the preference of decision makers, several devices have been developed such as solving the goal programming iteratively (see, for example, Ignizio [I1]).

Problem (ii) is more serious. In earlier days, no rule was suggested for deciding the weight, and, hence, systems analysists adopted the weight based on their experience. In 1972, Dyer [D7] suggested a method for deciding the weight under interaction with decision makers. Consider a case in which less of each objective function is preferred to more. In view of Remark 7.3.2, a formulation of goal programming may be given by

$$\text{(IGP)} \qquad \begin{aligned} &\text{Minimize} && \mu^{\mathrm{T}} y^- \\ &\text{subject to} && f(x) + y^+ - y^- = \bar{f} \\ &&& y^+, y^- \geqq 0 \end{aligned}$$

Dyer's idea originates from an observation that the distance function of problem (IGP) gives a piecewise linear approximation of an additive separable value function $V(f(x))$. Therefore, he suggested to set

$$\mu_i = (\partial V/\partial f_i)/(\partial V/\partial f_j).$$

In the interactive goal programming by Dyer, decision makers are requested to answer these weights as the marginal rate of substitutions with respect to their preferences. With this interaction, the goal programming is solved iteratively until the iteration converges. This interaction, however, suffers from the difficulty that the marginal rate of substitution seems to be beyond man's ability. This will be restated in more detail in the next subsection.

7.3.2 Interactive Optimization Methods

Although a number of interactive optimization methods have been developed for solving multiobjective programming problems, we discuss two of them here, typical in the historical and philosophical sense; namely, the interactive Frank–Wolfe method (Geoffrion *et al.* [G7]) and the surrogate worth trade-off method (Haimes *et al.* [H2]). For references to other methods, see, for example, Chankong and Haimes [C5], Goicoeche *et al.* [G8] and Hwang and Masud [H17].

7.3.2.1 Interactive Frank–Wolfe Method

The Frank–Wolfe method, orginally suggested for solving quadratic programming problems, can be applied to general convex programming problems

$$\text{(CP)} \qquad \text{Maximize} \quad V(x) \quad \text{subject to} \quad x \in X.$$

It has been known that the method converges to a solution with a good initial convergence rate under the condition of V being a real-valued differentiable pseudoconcave function of the n-vector $x = (x_1, x_2, \dots, x_n)$ of real variables and the set $X \subset R^n$ being convex and compact (Zangwill [Z2]). Its algorithm is given by

Step 0. Choose $x^1 \in X$ and set $k = 1$.

Step 1. Determine the search direction $d^k = y^k - x^k$, where y^k is a solution of

$$\text{Maximize } \nabla_x f(x^k) y \qquad \text{over} \quad y \in X.$$

Step 2. Determine the step length α^k that solves

$$\text{Maximize} \qquad f(x^k + \alpha d^k) \qquad \text{over} \quad \alpha \in [0, 1].$$

If $f(x^k + \alpha^k d^k) \leqq f(x^k)$, then stop. Otherwise, let $x^{k+1} = x^k + \alpha^k d^k$, and go to Step 1 replacing k with $k + 1$.

In multiobjective programming, we presume for the total objective function the decision maker's value function $V(f_1, \dots, f_p)$ over the objective space, which is not known explicitly. Note that

$$\nabla_x V(f(x^k)) = \sum_{i=1}^{p} \frac{\partial V f(x^k)}{\partial f_i} \nabla_x f_i(x^k). \tag{7.3.3}$$

The term $\partial V / \partial f_i$ is the so-called marginal value with respect to f_i, whose evaluation has been supposed to be impossible in traditional economics. However, as in traditional economics, it also suffices for us to use the marginal rate of substitution (MRS) instead of the marginal value for our present purpose. Let $A = (\hat{f}_1, \ldots, \hat{f}_p)$ be indifferent to $B = (\hat{f}_1, \ldots, \hat{f}_i + \Delta f_i, \ldots, \hat{f}_j - \Delta f_j, \ldots, \hat{f}_p)$. Namely, in order to keep the indifference relation between A and B, the improvement of f_j by Δf_j is required for compensating for the sacrifice of f_i by Δf_i. Then the marginal rate of substitution m_{ij} of f_i for f_j at $\hat{f}$ is defined as the limit of the rate of substitution $\Delta f_j / \Delta f_i$ as B tending to A. That is,

$$m_{ij} = \lim_{B \to A} \left(\frac{\Delta f_j}{\Delta f_i} \right).$$

As is readily seen, there is a relation between the marginal rate of substitution and the marginal value

$$m_{ij} = (\partial V / \partial f_i) / (\partial V / \partial f_j).$$

Hence, the relation (7.3.3) can be rewritten as

$$\nabla_x V(f(x^k)) = \frac{\partial V(f^k)}{\partial f_s} \sum_{i=1}^{p} m_{is}^k \nabla_x f_i(x^k).$$

Note here that the term $\partial V(f^k)/\partial f_s$ does not have any effect on the solution of the maximization problem (CP), so long as s is selected in such a way that the term becomes negative at each iteration. Therefore, in using the Frank–Wolfe method for our multiobjective problem in an interactive way, the decision maker is requested to answer his MRS at Step 1 and the most preferred point along the search direction at Step 2.

7.3.2.1.1 Discussion. We have the following questions in the interactive Frank–Wolfe method:

(i) Can we guarantee the differentiability and quasiconcavity of $V(f_1, \ldots, f_p)$ so that the method may converge?
(ii) How is the MRS of the decision maker assessed?
(iii) Can the decision maker make the line search easily?

As to question (i), Wehrung [W6] proved that the convergence property of the method under some appropriate conditions on the decision maker's preference rather than the value function $V(f)$ is

(1) complete,
(2) transitive,
(3) continuous in the sense that both sets

$$S^+(\tilde{f}) := \{\hat{f} \in f(X) \mid \hat{f} \succsim \tilde{f}\}$$

and

$$S^-(\tilde{f}) := \{\hat{f} \in f(X) \,|\, \hat{f} \precsim \tilde{f}\}$$

are closed for any $\tilde{f}$,

(4) convex in the sense that for any f^1 and f^2

$$f^1 \succ f^2 \Rightarrow \alpha f^1 + (1-\alpha) f^2 \succsim f^2 \qquad \text{for all} \quad \alpha \in [0, 1),$$

(5) there exists a unique MRS at each point $(f_1, \ldots, f_p)$.

It should be noted that the convexity of preference may be interpreted to mean that the preference is discriminating enough to exclude thick indifference curves (Debreu [D3]). The conditions of the convexity of preference and the unique existence of MRS will be seen later to be very strong in many practical situations. Observe also that the check of these conditions is laborious, even if we accept them.

Question (ii) is serious. We may easily become aware of difficulties in estimating the MRS of decision makers. Although some devices for getting the MRS have been made (Dyer [D7]), it has been observed that many decision makers cannot easily answer their MRS (Wallenius [W2] and Cohon and Marks [C14]. In fact, in getting an MRS (m_{ij}), decision makers are requested to specify the amount Δf_i by which they agree to sacrifice f_i for improving f_j by a given small amount Δf_j *in order to keep the indifference relation*. It has been observed in many cases that such a judgment troubles decision makers very much (see, for example, Wallenius [W2]). Furthermore, recall that the MRS is given as a limit of a ratio of finite differences. Strictly speaking, such a limit operation, implied by the terminology of *marginal*, seems beyond human ability. As is widely perceived, human beings cannot recognize a small change within some extent. This threshold of human recognition is called the just noticeable difference (JND). Therefore, we can not make the amount Δf_j arbitrarily small. This observation shows the fact that the MRS obtained through the interaction with decision makers is just an approximation (possibly rough) to the correct one. An effect of this error in evaluation of the MRS to the convergence property of the method was analyzed by Dyer [D8], who concluded the method is stable to such an error. However, we may see in Fig. 7.7 that some of the underlying assumptions in Dyer's conclusion (in particular, the exact line search) are too strong in actual cases. If the JND of the attribute X is $x_B - x_A$, in the evaluation of the MRS we cannot access further from the point B to the point A, which implies that the MRS represents the direction d. When the difference between A and C is within the JND along the direction d, the decision maker answers that A is most preferred along the direction. Namely, the optimization process terminates at the point A, which differs from the true optimum (if any) and possibly even from the satisfactory solution.

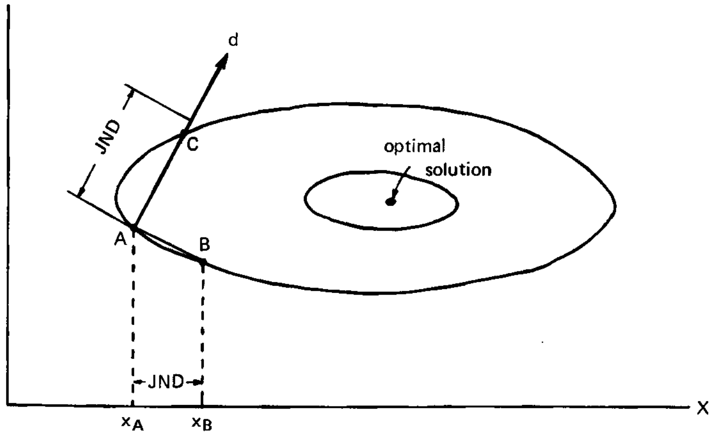

Fig. 7.7.

Note here that, in the presence of the JND, indifference curves in a usual sense cannot so much as exist (and hence neither can the MRS), because the JND causes an intransitive indifference. For more details on the JND and intransitive indifference, see Chipman [C9], Luce [L11], and Tversky [T15].

Question (iii) is not as severe as question (ii). Recall that we cannot expect the exact line search even in traditional mathematical programming with explicitly known objective functions; hence, it suffices for convergence to find an approximate point $\bar{f}$ more preferable to f^k. Even so, it is not very easy for decision makers to answer a more preferable point along the direction, because it requires them to rank vectors, that is, to evaluate multiple criteria simultaneously. In particular, this situation becomes severe as we come close to the solution. Furthermore, poor line searches cause poor rates of convergence. Wehrung [W7] suggested a method that determines a subset of a predetermined finite number of equal subintervals of $[f^k, y^k]$ based on the Fibonacci search method, the golden section method, the bisection method, and other methods. On the other hand, Oppenheimer [O1] suggested the iterative utilization of a proxy function with a known form as a local approximation of $V(f(x))$. As a proxy function, he suggested the use of the exponential sum

$$P_1(f) = -\sum_{i=1}^{p} a_i \exp(-w_i f_i)$$

or the power sum

$$P_2(f) = -\sum_{i=1}^{p} a_i f_i^{\alpha_i}.$$

The parameters a_i, w_i, and α_i are decided for P_1, for example, by solving the equation

$$\begin{aligned} m_{ij} &= (\partial P(f)/\partial f_i)(\partial P(f)/\partial f_j) \\ &= (w_i a_i \exp(-w_i f_i))/(w_j a_j \exp(-w_j f_j)). \end{aligned}$$

Using Oppenheimer's method, the number of assessments of the decision maker's preference is expected to decrease. However, in general, the convergence (and convergence rate) would be sensitive to the form of proxy function and the assessed MRS.

7.3.2.2 *The Surrogate Worth Trade-Off Method*

If the less of each objective function f_i is preferred to the more, it is natural to consider the decision maker's preference to be monotonous; namely,

$$f^1 < f^2 \Rightarrow f^1 \succ f^2.$$

Then, according to Theorem 3.1.1, the most preferable solution with respect to the decision maker's preference should belong to the set of efficient solutions. In the Geoffrion–Dyer–Feinberg method, the obtained solution may be proved theoretically to be an efficient solution under some appropriate conditions such as the convexity of the preference, the criteria f_i, and the set X. In practice, however, this cannot necessarily be expected due to several kinds of errors including human factors. Haimes *et al.* [H2] suggested a method, called the surrogate worth trade-off (SWT) method, in which the search is limited on the efficient set without using decision makers' MRS explicitly. Before stating the method, note the following property of the preferentially optimal solution.

Assumption 7.3.1

The set of efficient solutions in the objective space defines a smooth surface E in R^p.

Definition 7.3.1 (*Efficiency Trade-Off Ratio*)

Let $(t_1, \ldots, t_{p-1}, 1)$ denote the direction ratio of the outer normal of E at an efficient point $\hat{f}$. Then t_j gives information on the trade-off between f_p and f_j at the point $\hat{f}$ for keeping the efficiency. Therefore, t_j is called the efficiency trade-off ratio between f_p and f_j.

Lemma 7.3.2

Let $f^0 = f(x^0)$ be a preferentially optimal solution, and let $m_i = m_{ip}$ $(i = 1, \ldots, p-1)$ for the MRS of the decision maker. We then have

$$f^0 \in E, \qquad m_i(f^0) = t_i(f^0) \quad (i = 1, \ldots, p-1)$$

Furthermore, under the convexity of the preference, each objective function, and the set X, the preceding condition is also sufficient for the preferential optimality of f^0.

Proof With the help of a value function representing the decision maker's preference, the lemma follows from the well-known Kuhn–Tucker theorem (Haimes *et al.* [H2]). Also see Nakayama [N3] for a proof in terms of the preference relation.

The algorithm of the SWT method is summarized as follows:

Step 1. Generate a representative subset of efficient solutions. To this end, the auxiliary optimization stated in Section 3.4.3 is recommended for appropriately chosen vectors $(\varepsilon_1, \ldots, \varepsilon_{p-1})$, where the reference objective function is taken as f_p without loss of generality.

Step 2. The decision maker is asked to score each efficient solution obtained. Let λ_j be the Lagrange multiplier associated with the constraint $f_j(x) \leqq \varepsilon_j$ in the auxiliary optimization at step 1. Under the hypothesis of the smoothness of the efficient surface E, λ_j gives information on the efficiency trade-off between f_p and f_j at the solution $f(\hat{x})$. We ask the decision maker how much he would like to improve f_p by λ_j units per one-unit degradation of f_j at $f(\hat{x})$ while other criteria being fixed at $f_s(\hat{x})$ $(s \neq j, p)$ and would he please score the degree of desire with its sign. Here the plus sign indicates the case in which the amount of improvement of f_p, λ_j produced by the unit sacrifice of f_j is too short (i.e., $m_j > \lambda_j$), while the minus sign is associated with the opposite case $(m_j < \lambda_j)$. If he is indifferent about that trade, the score is to be zero $(m_j = \lambda_j)$. Figure 7.8 justifies this manner of scoring. These scores are recorded as the surrogate worth w_{pj}. The same procedure is made for all representative efficient solutions obtained at step 1.

Step 3. An optimal solution $\hat{x}$ can be obtained as the one satisfying

$$w_{pj}(f(\hat{x})) = 0 \qquad \text{for all} \quad j \neq p. \tag{7.3.3}$$

If there is no such a solution among the representative efficient solutions, then we decide the function forms of $w_{pj}(f)$ by some appropriate methods such as multiple regression analysis. After that, we can get a most preferable solution by solving Eq. (7.3.3).

7.3.2.2.1 Discussion. We may see easily several difficulties in the SWT method; they include the smoothness of the efficient surface, generating the

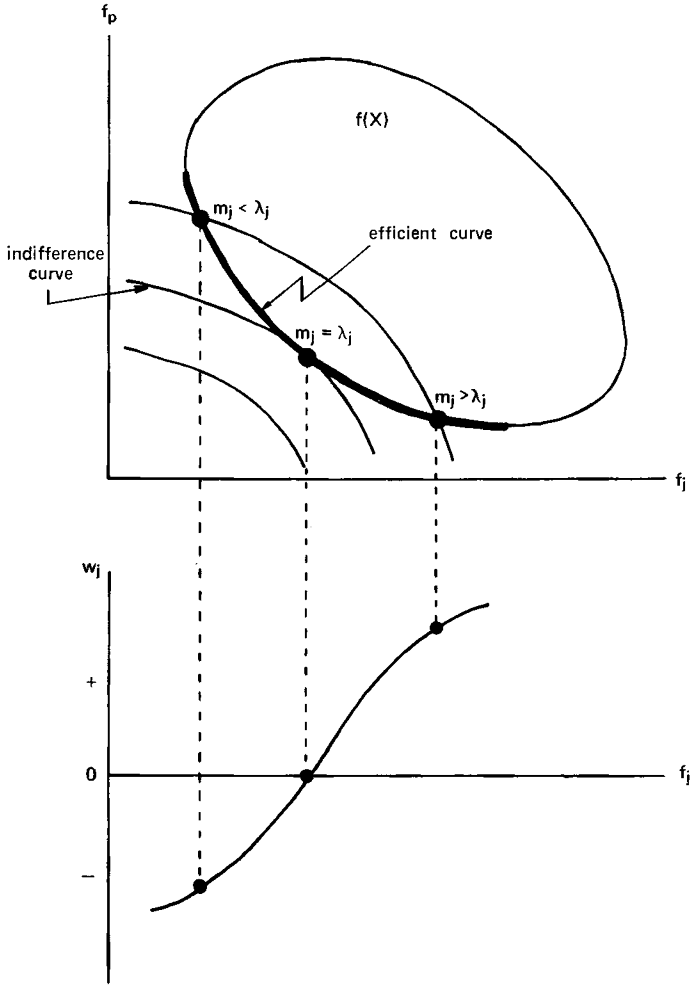

Fig. 7.8. SWT method.

representative set of efficient solutions, scoring the surrogate worth, and multiple regression and solving a system of nonlinear equations $w_{pj}(x) = 0$. We shall discuss these points below.

The smoothness of the efficient surface is guaranteed only for limited cases, for example, in which each objective function is smooth and strictly convex. If the efficient surface is not smooth, we could have a condition for the preferential optimality of f^0

$$-(m_1, \ldots, m_{p-1}, 1) \in (T(f(X), f^0))^{\circ}$$

(see, for example, Nakayama [N3]). In this case, the Lagrange multiplier $(\lambda_1, \ldots, \lambda_{p-1}, 1)$ obtained by the auxiliary optimization at step 1 is merely a vector of $((T(f(X), f^0))^\circ$, and hence $m_j = \lambda_j$ $(j = 1, \ldots, p-1)$ cannot be expected, in general.

We need to solve an auxiliary nonlinear optimization problem in order to get an efficient solution. In many practical problems, this step requires much computational effort, and hence the least number of representative efficient solutions is more desirable. However, if the number is too few, then the most preferable solution might not be in the given representative efficient set. The case, therefore, charges us with the extra effort of multiple regression and solving the nonlinear equations $w_{pj} = 0$.

In the original version of the SWT method, the surrogate worth is needed only to score an ordinal measure. However, since such an ordinal score does not make sense for comparison of the differences among them, the subsequent multiple regression and the nonlinear equation $w_{pj}(f(x)) = 0$ are very difficult to solve. In Chankong and Haimes [C6], the surrogate worth is supposed to be scored by a cardinal measure. However, the consistency of the surrogate worth as a cardinal measure becomes a new problem.

In cases with only a few objective functions, a solution satisfying $w_{pj}(f(x)) = 0$ $(j = 1, \ldots, p-1)$ may be relatively easily obtained in the given representative efficient set. In these lucky cases, we do not need to invoke the multiple regression analysis to get the function form of $w_{pj}(f)$ and solve the system of nonlinear equations $w_{pj}(f) = 0$. In general, however (particularly in cases with a large number of objective functions) such a lucky case is hopeless, and, therefore, the stated extra efforts are needed. The difficulty for these operations has been seen in the preceding discussion for the ordinal surogate worth.

In order to overcome some of these difficulties, several modifications have been developed. Note that in the SWT method since the search is restricted on the efficient surface, the independent variable is the $p-1$-dimensional vector $(f_1, \ldots, f_{p-1})$. Let the efficient surface be represented by the equation $E(f_1, \ldots, f_p) = 0$. If we have $f_p = \omega(f_1, \ldots, f_{p-1})$ by solving the equation, then for the value function restricted on the efficient surface E,

$$\tilde{V}(f_1, \ldots, f_{p-1}) := V(f_1, \ldots, f_{p-1}, \omega(f_1, \ldots, f_{p-1}))$$
$$\tilde{V}_j = V_j + V_p \omega_j \qquad (j = 1, \ldots, p-1), \tag{7.3.4}$$

where $\tilde{V}_j = \partial \tilde{V}/\partial f_j$, $V_j = \partial V/\partial f_j$, and $\omega_j = \partial \omega/\partial f_j$. On the other hand, on the efficient surface E,

$$E_j + E_p \omega_j = 0 \qquad (j = 1, \ldots, p-1)$$

where $E_j = \partial E/\partial f_j$. Therefore, making use of the fact that

$$m_i = V_j/V_p \qquad \text{and} \qquad t_j = E_j/E_p,$$

we have, under the monotonicity of the preference,

$$\tilde{V}_j \gtreqqless 0 \Leftrightarrow m_j \gtreqqless t_j \qquad (j = 1, \ldots, p-1). \tag{7.3.5}$$

In order to decrease the number of representative efficient solutions, Chankong and Haimes [C4] suggested an interactive SWT method on the basis of relation (7.3.5). Instead of the pregiven set of representative efficient solutions, they improve the search point (which is an efficient solution) iteratively using the information of the surrogate worth at the point; namely,

$$w_{pj}(f(x)) \underset{(<)}{>} 0 \Rightarrow \varepsilon_j^{k+1} = \varepsilon_j^k \underset{(-)}{+} \delta f_j$$

for sufficiently small δf_j $(j = 1, \ldots, p-1)$. Observe that this method corresponds to the simple ascent method for the reduced value function $\tilde{V}$. If we use the MRS of the decision maker explicitly, the steepest ascent direction of the reduced value function may be obtained through relation (7.3.4). It should be noted that, in general, the simple ascent method bears a poorer rate of convergence than the steepest acent method. Furthermore, the method of selecting the magnitude of δf_j becomes a problem. If the value of δf_j is too small, we will get a poor rate of convergence. On the other hand, in this method we do not need to invoke the multiple regression analysis nor solve the equations $w_{pj}(f(x)) = 0$.

According to our experiences, we have observed that the graphical information is more understandable for decision makers than just numerical information. In particular, the efficient curve of f_p versus f_j provides the global information on the trade-off for efficiency, and therefore the decision maker can answer his preferred solution on the displayed curve much more easily than just on the basis of the efficiency trade-off ratio, which merely provides the local information on the trade-off for efficiency. Recall that the SWT method searches for a most preferable solution in the space $(f_1, \ldots, f_{p-1})$ by solving auxilliary optimization problems to get efficient solutions. We may then think of using the relaxation method (coordinatewise optimization method) in that space. Nakayama *et al.* [N11] suggested the graphical interactive relaxation method based on this consideration, in which the decision maker is requested only to answer his preferred solution on the shown efficient curve between a paired objective function. The pair of objective functions is changed cyclically while other objective functions are fixed at the level of the answered preferred solution so far (see Fig. 7.9). In the method using graphic displays in an interactive way, the burden of the decision maker is much less than the other methods stated above. However, it usually needs much computational effort to get each efficient curve. Moreover, its convergence property is limited to the case with smooth efficient surfaces.

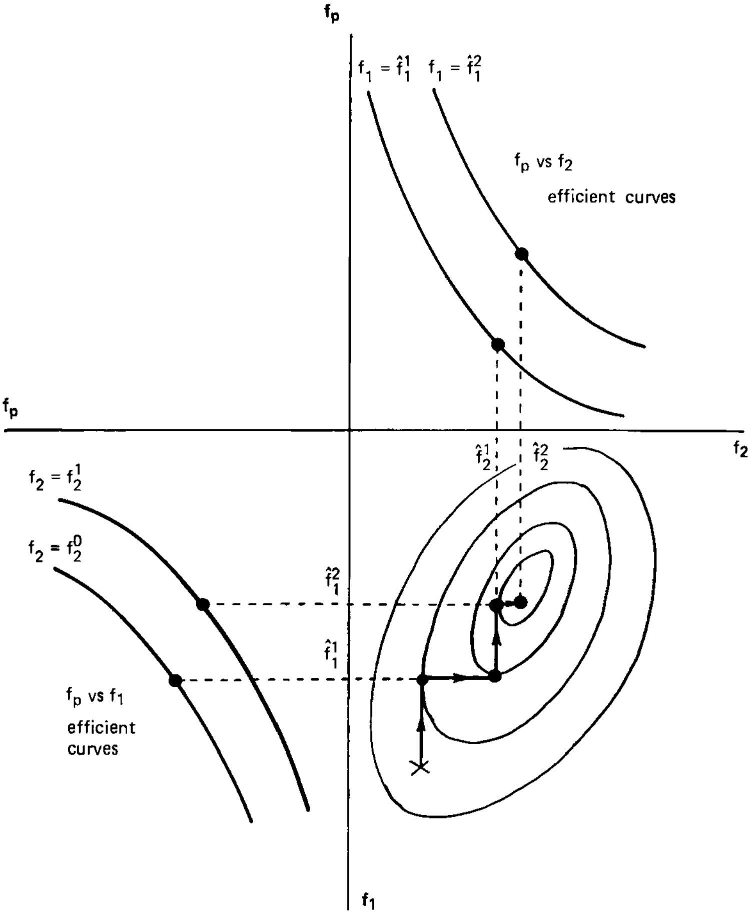

Fig. 7.9. IRM.

7.3.3 Interactive Satisficing Methods

As long as we take an approach of optimization, we cannot help requiring decision makers to make such a high degree of judgment as marginal rate of substitution or ordering vectors. In order to decrease the burden of decision makers, interactive programming methods that take the approach of satisficing as a decision rule have been developed. We have seen that goal programming may be used, in which the solution is just a satisfactory one if the aspiration level is feasible, or it is as close as possible to the aspiration levels if it is not feasible.

Note here that satisficing was originally suggested due to the observation of the limit of human cognitive ability and available information through decision making (Simon [S9]). In our problem formulation, however, we are supposed to be able to evaluate the values of criteria functions by some means, although we also observe the limit of human cognitive ability. For these problems, mere satisficing is not necesarily satisfactory to the decision maker as a decision rule.

For example, let us consider a case in which a decision maker ordered two designers to design some industrial product. Almost at the same time and almost at the same expense, these two designers completed their designs, which were both satisfactory to the decision maker. However, one of the designs is superior to the other in terms of all criteria. In this case, there seems no doubt that the decision maker should adopt the superior one. This example allows us to use a solution $\tilde{x}$ which is satisfactory and in addition gurantees that there is no other feasible solution that is superior to $\tilde{x}$ in terms of all criteria.

Definition 7.3.2 (*Satisfactory Pareto Solution*)

A solution that is satisfactory and (weakly) Pareto efficient is called a satisfactory (resp. weak) Pareto solution. In other words, when $\bar{f}$ is an aspiration level of the decision maker, an alternative $\tilde{x}$ is said to be a satisfactory Pareto solution if

$$f(x) \not\leq f(\tilde{x}) \qquad \text{for all} \quad x \in X$$

and

$$f(\tilde{x}) \leqq \bar{f}.$$

A satisfactory weak Pareto solution can be defined analogously by replacing $\not\leq$ with $\not<$ in the first inequality.

In general, satisfactory (weak) Pareto solutions constitute a subset of X. However, we can narrow down the set of satisfactory Pareto solutions and obtain a solution close to the preferentially optimal one by tightening the aspiration level, if necessary. On the other hand, even if there does not exist any satisfactory Pareto solution for the initial aspiration level, we can attain one by relaxing the aspiration level. Originally, the aspiration level is fuzzy and flexible. It seems that decision makers change their aspiration levels according to situations.

From this point of view, the interactive satisficing methods provide a tool that supports decision makers in finding an appropriate satisfactory (weak) Pareto solution by some adjustment of the aspiration level. Usually the aspiration level is changed in such a way that

$$\bar{f}^{k+1} = T \circ P(\bar{f}^k),$$

where $\bar{f}^k$ represents the aspiration level at the kth iteration. The operator P selects the Pareto solution nearest in some sense to the given aspiration level $\bar{f}^k$. The operator T is the trade-off operator, which changes the kth aspiration level $\bar{f}^k$ if the decision maker does not compromise with the shown solution $P(\bar{f}^k)$. Of course, since $P(\bar{f}^k)$ is a Pareto solution, there exists no feasible solution that makes all criteria better than $P(\bar{f}^k)$, and the decision maker must trade-off among criteria if he wants to improve some of criteria. Based on this trade-off, a new aspiration level is decided as $T \circ P(\bar{f}^k)$. The process is continued until the decision maker obtains an agreeable solution.

7.3.3.1 STEM

STEM (Step Method) by Benayoun *et al.* [B6] was originally developed for solving linear multiobjective programming problem, and its extension to general nonlinear cases is straightforward (Shimizu [S8]). The following discussion will be made for nonlinear multiobjective problems.

Let f^* be an ideal point, and let us try to find a solution as close as possible to this ideal point. Although several distance function may be available as in goal programming, l_∞-norm is usually recommended:

$$\min_{x \in X} \max_{1 \leqq i \leqq p} \mu_i |f_i(x) - f_i^*| \qquad \text{subject to}$$

$$f_i(x) - f_i^* \leqq \Delta f_i, \qquad (i = 1, \ldots, p),$$

where Δf_i is an admissible deviation from the ideal level f_i^*.

Equivalently,

$$\min_{x, \xi} \xi \qquad \text{subject to}$$

$$\mu_i(f_i(x) - f_i^*) \leqq \xi, \qquad (i = 1, \ldots, p)$$

$$f_i(x) - f_i^* \leqq \Delta f_i \qquad (i = 1, \ldots, p)$$

$$x \in X.$$

Note here that the aspiration level is given by $\bar{f}_i = f_i^* + \Delta f_i$. The solution to these problems is clearly a satisfactory weak Pareto solution if the aspiration level is feasible; namely, $\bar{f} \in f(X)$. However, taking into account the total balance among the levels of objective functions at the obtained solution, decision makers are sometimes dissatisfied with some of them. More importantly, the aspiration level might happen to be infeasible in some cases. Under such a circumstance, some of aspiration levels should be relaxed in order to get a satisfactory solution. In other words, the decision maker is now faced with a trade-off problem.

On the basis of the above consideration, STEM and its modifications provide a tool for getting a reasonable solution by changing the aspiration level in an interactive way.

Step 0. Set $k = 1$, and determine the ideal point f^* and the weight μ_i by some appropriate way (see Remark 7.3.4 later). Find a solution x^k to the following problem:

$$\operatorname*{Min}_{x,\xi} \xi \qquad \text{subject to} \qquad \mu_i(f_i(x) - f_i^*) \leqq \xi, \quad (i = 1, \dots, p) \qquad x \in X.$$

Step 1. Seeing $f(x^k)$ obtained above, the decision maker classifies the objective functions into two classes: the ones with satisfactory levels (its index set is denoted by S) and the ones with unsatisfactory levels (its index set is denoted by U). If there is no unsatisfactory f_i, then the iteration terminates. Otherwise the decision maker relaxes the level of the satisfactory objective functions, whose amount is set Δf_i.

Step 2. Solve

$$\operatorname*{Min}_{x,\xi} \xi \qquad \text{subject to}$$

$$\mu_i(f_i(x) - f_i^*) \leqq \xi, \qquad i \in U,$$

$$f_i(x) - f_i^* \leqq \Delta f_i, \qquad i \in S,$$

$$x \in X.$$

Set the solution to be x^{k+1} and return to step 1.

Remark 7.3.4

In using the l_α-type distance function, it is very important to make a common scale for the objective functions. For example, if the positive and negative values of the objective functions are mixed, or if their numerical orders are extremely different from each other, then in the Min–Max problem some of objective functions are sensitive and others not. Note that the origin is set up by the ideal point f^*, and the weight μ_i corresponds to the scale factor.

In STEM the ideal point f^* is set by

$$f_i^* = \operatorname*{Min}_{x \in X} f_i(x) := f_i^*(x_i^*),$$

and the weight μ_i is given by normalizing

$$w_i' = \frac{f_i^{\max} - f_i^*}{\hat{f}_i} \cdot \left(\sqrt{\sum_{j=1}^{n} c_{ij}^2}\right)^{-1},$$

$$
\begin{array}{c|cccc|}
 & f_1 & f_2 & & f_p \\
\hline
f_1 & f_1^*(x_1^*) & f_2(x_1^*) & \cdots\cdots & f_p(x_1^*) \\
f_2 & f_1(x_2^*) & f_2^*(x_2^*) & \cdots\cdots & f_p(x_2^*) \\
 & \vdots & \vdots & & \vdots \\
f_p & f_1(x_p^*) & f_2(x_p^*) & \cdots\cdots & f_p^*(x_p^*) \\
\hline
\end{array}
$$

Fig. 7.10. Payoff matrix.

where c_{ij} is the coefficient of the ith objective function $f_i(x) = \sum c_{ij}x_j$, $f_i^{\max}$ is the maximum component of the ith row vector in the so-called payoff matrix given by Fig. 7.10 and $\hat{f}_i = f_i^{\max}$ if $f_i^{\max} > 0$, otherwise $\hat{f}_i = |f_i^*|$.

7.3.3.1.1 Discussion. Intuitively speaking, the stated method relaxes the objective functions that attain their aspiration levels at an early stage and distributes the benefit caused by their sacrifice to other objective functions equally; the process is continued until all objective functions become satisfactory. However, it is not clarified just how much sacrifice of some objective functions gives how much improvement to other objective functions. As a result, it is up to the decision maker's guess with the help of the payoff matrix. Additionally, in many practical cases, the decision maker desires to improve some of the objective functions much more strongly than he may agree to sacrifice others. Therefore, it seems more natural to ask the decision makers how much they want to improve unsatisfactory objectives as well as how much they agree to sacrifice satisfactory ones. From this point of view, Wierzbicki [W13, W14] and Lewandowski and Grauer [L5] suggested a method called the reference point method or dynamic interactive decision analysis support systems (DIDASS) (see also Grauer *et al.* [G9]). Independently from their works, Nakayama [N4, N8] also developed a similar method called the satisficing trade-off method, which will be discussed in more detail later.

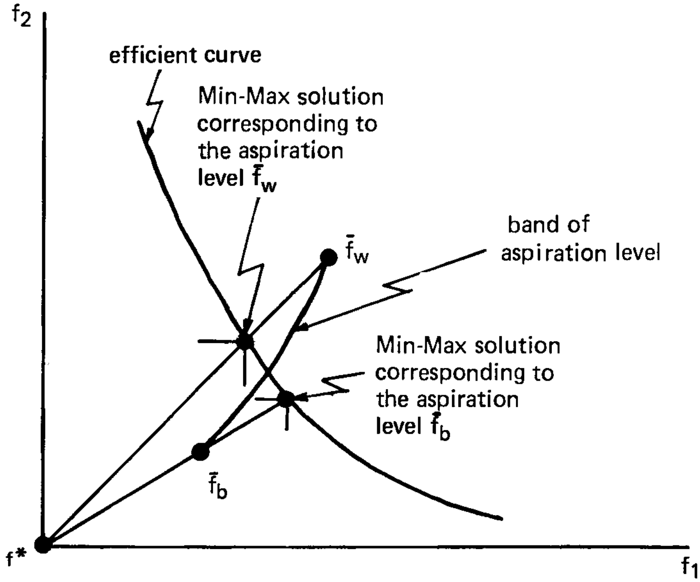

Fig. 7.11. Satisficing trade-off method.

7.3.3.2 Satisficing Trade-Off Method

First, we note several important properties for the Min–Max problem (Fig. 7.11)

$$(\mathrm{P}_\infty) \qquad \operatorname*{Min}_{x\in X} \operatorname*{Max}_{1\leq i\leq p} \mu_i|f_i(x) - f_i^*|.$$

Lemma 7.3.3

For $f_i^* < \mathrm{Min}\{f_i(x) : x \in X\}$, each Pareto solution can be obtained by solving the Min–Max problem (P_∞) with an appropriate positive weight μ_i $(i = 1, \ldots, p)$.

Proof Immediate from Theorem 3.4.9.

Lemma 7.3.4

Suppose that for any $x \in X$

$$f_i^* < f_i(x), \qquad (i = 1, \ldots, p). \tag{7.3.6}$$

If we set for a given aspiration level $\bar{f}$ with $\bar{f} > f^*$

$$\mu_i = 1/(\bar{f}_i - f_i^*), \qquad (i = 1, \ldots, p), \tag{7.3.7}$$

then the solution $\tilde{x}$ to the Min–Max problem (P_∞) is a satisfactory weak Pareto solution in case of $\bar{f}$ being feasible, while it is assured to be a weak Pareto solution even in case of $\bar{f}$ being infeasible.

Proof It is well known that the solution to the Min–Max problem (P_∞) is a weak Pareto solution. Hence, we shall show that if the aspiration level $\bar{f}$ is feasible, then the solution $\tilde{x}$ to the Min–Max problem becomes a satisfactory solution; namely, $f(\tilde{x}) \leqq \bar{f}$.

Let

$$\rho(y) := \operatorname*{Max}_{1 \leqq i \leqq p} \mu_i |y_i - f_i^*|$$

and define the level set at $\bar{f}$ by

$$L := \{y = (y_1, \ldots, y_p) \mid \rho(y) \leqq \rho(\bar{f})\}.$$

Furthermore, setting

$$S := \{y = (y_1, \ldots, y_p) \mid y_i \leqq \bar{f}_i,\ i = 1, \ldots, p\},$$

we have

$$L \subset S. \tag{7.3.8}$$

In fact, the way of setting the weight μ_i yields $\mu_i |\bar{f}_i - f_i^*| = 1$ $(i = 1, \ldots, p)$, and hence for any $y \in L$ and $i = 1, \ldots, p$,

$$\mu_i |y_i - f_i^*| \leqq \mu_i(\bar{f}_i - f_i^*),$$

which yields $y_i \leqq \bar{f}_i$ $(i = 1, \ldots, p)$.

Now note that if $\bar{f}$ is feasible, then there exists an $\hat{x} \in X$ such that $f_i(\hat{x}) \leqq \bar{f}_i$ for all $i = 1, \ldots, p$. Then

$$\begin{aligned} \mu_i |f_i(\hat{x}) - f_i^*| &= \mu_i(f_i(\hat{x}) - f_i^*) \\ &\leqq \mu_i(\bar{f}_i - f_i^*), \end{aligned}$$

thereby

$$\{f(x) \mid x \in X\} \cap L \neq \varnothing. \tag{7.3.9}$$

Relations (7.3.8) and (7.3.9) imply that the solution to $\mathrm{Min}_{x \in X}\, \rho(f(x))$ belongs to the set S.

Lemma 7.3.5

Let $(\tilde{x}, \tilde{\xi})$ be a solution to

$$(\mathrm{P}'_\infty)\quad \operatorname*{Min}_{x, \xi}\ \xi \quad \text{subject to} \quad \mu_i(f_i(x) - f_i^*) \leqq \xi, \qquad (i = 1, \ldots, p), \quad x \in X,$$

which is equivalent to the Min–Max problem (P_∞), and let $\tilde{\lambda} = (\tilde{\lambda}_1, \ldots, \tilde{\lambda}_p)$ be the optimal Lagrange multipliers. If $\tilde{x}$ is interior to the set X and each f_i has appropriate smoothness, then we have

$$\sum_{i=1}^{p} \tilde{\lambda}_i = 1, \qquad \tilde{\lambda}_i \geqq 0, \qquad (i = 1, \ldots, p) \tag{7.3.10}$$

$$\sum_{i=1}^{p} \tilde{\lambda}_i \mu_i \nabla f_i(\tilde{x}) = 0. \tag{7.3.11}$$

Moreover, if the problem (P'_∞) is convex, namely, if each f_i is convex and the set X is convex, then for any $x \in X$

$$\sum_{i=1}^{p} \tilde{\lambda}_i \mu_i (f_i(x) - f_i(\tilde{x})) \geqq 0. \tag{7.3.12}$$

Proof Relations (7.3.10) and (7.3.11) immediately follow from the well-known Kuhn–Tucker theorem for the Lagrangean

$$L(x, \xi : \lambda) = \xi + \sum_{i=1}^{p} \lambda_i(\mu_i(f_i(x) - f_i^*)) - \xi).$$

In addition, the well-known theory of convex programming provides that for the solution $(\tilde{x}, \tilde{\xi})$ and $\tilde{\lambda}$ and for all $x \in X$ and ξ

$$L(x, \xi : \tilde{\lambda}) \geqq L(\tilde{x}, \tilde{\xi} : \tilde{\lambda}),$$

from which relation (7.3.12) follows by virtue of $\sum_{i=1}^{p} \tilde{\lambda}_i = 1$.

The algorithm of the satisficing trade-off method is summarized as follows:

Step 1 (setting the ideal point). The ideal point $f^* = (f_1^*, \ldots, f_p^*)$ is set, where f_i^* is small enough [for example, $f_i^* = \mathrm{Min}\{f_i(x) \mid x \in X\} - \varepsilon\,(\varepsilon > 0)$]. This value is fixed throughout the following process.

Step 2 (setting the aspiration level). The aspiration level $\bar{f}_i^k$ of each objective function f_i at the kth iteration is asked of the decision maker. Here $\bar{f}_i^k$ should be set in such a way that $\bar{f}_i^k > f_i^*$ $(i = 1, \ldots, p)$. Set $k = 1$.

Step 3 (weighting and finding a Pareto solution by the Min–Max method). Set

$$\mu_i^k = 1/(\bar{f}_i^k - f_i^*), \tag{7.3.13}$$

and solve the Min–Max problem

$$(\mathrm{P}_\infty^k) \qquad \operatorname*{Min}_{x \in X} \operatorname*{Max}_{1 \leqq i \leqq p} \mu_i^k |f_i(x) - f_i^*|$$

or equivalently

$$(\mathrm{P}_\infty'^k) \quad \operatorname*{Min}_{x, \xi} \xi \quad \text{subject to} \quad \mu_i^k(f_i(x) - f_i^*) \leqq \xi, \quad (i = 1, \ldots, p), \quad x \in X.$$

Let x^k be a solution to these problems.

Step 4 (trade-off). Based on the value of $f(x^k)$, the decision maker classifies the criteria into three groups, namely, the class of criteria that he wants to improve more, the class of criteria that he may agree to relax, and the class of criteria that he accepts as they are. The index set of each class is represented by I_{I}^k, I_{R}^k, I_{A}^k, respectively. If $I_{\mathrm{I}}^k = \varnothing$, then stop the procedure. Otherwise, the decision maker is asked his new acceptable level of criteria $\tilde{f}_i^k$ for the class of I_{I}^k and I_{R}^k. For $i \in I_{\mathrm{A}}^k$, set $\tilde{f}_i^k = f_i(x^k)$.

Step 5 (feasibility check). Let λ_i $(i = 1, \ldots, p)$ be the optimal Lagrange multipliers to the problem (P'^k_∞). If for a small nonnegative ε

$$\sum_{i=1}^{p} \lambda_i \mu_i^k(\tilde{f}_i^k - f_i(x^k)) \geqq -\varepsilon, \tag{7.3.14}$$

set the new aspiration level $\bar{f}_i^{k+1}$ as $\tilde{f}_i^k$ and return to the step 3; otherwise, $\tilde{f}_i^k$ might be infeasible in the sense of linear approximation as will be explained later. By taking the degree of difficulty for solving the Min–Max problem into account, we choose either to trade off again or to return to the step 3 by setting $\bar{f}_i^{k+1} = \tilde{f}_i^k$. In the case of trading off again, the acceptable level of criteria for I_{I}^k and/or I_{R}^k should be reset lower than before, and then return to the beginning of the step 5.

The outstanding features of the satisficing trade-off method include the following:

(1) We do not need to pay much attention to setting the ideal point f^*. It suffices to set f^* sufficiently small enough to cover all or almost of all Pareto solutions as candidates for a decision solution in the following process. In case in which $\mathrm{Min}\{f_i(x) \mid x \in X\}$ is finite, for example, set $f_i^* = \mathrm{Min}\{f_i(x) \mid x \in X\} - \varepsilon$ $(\varepsilon > 0)$. Otherwise, set f_i^* to be sufficiently small.

(2) The weights μ_i $(i = 1, \ldots, p)$ are automatically set by the ideal point f^* and the aspiration level $\bar{f}$. For the weight in Eq. (7.3.13), the value of $\mu_i(f_i^* - f_i(x))$ can be considered to represent the normalized degree of nonattainability of $f_i(x)$ to the ideal point f_i^*. This enables us to ignore the need to pay extra attention to the difference among the dimension and the numerical order of criteria.

(3) By solving the Min–Max problem under the preceding conditions we can get a satisfactory weak Pareto solution in the case in which the aspiration level $\bar{f}$ is feasible, and a weak Pareto solution even in the case in which $\bar{f}$ is unfeasible. Interpreting this intuitively, in case in which $\bar{f}$ is feasible the obtained satisfactory weak Pareto solution is the one that improves equally in some sense each criterion as much as possible, and in case in which $\bar{f}$ is infeasible we get a weak Pareto solution nearest to the ideal point that shares an equal amount of normalized sacrifice for each criterion. This practical

meaning encourages the decision maker to accept easily the obtained solution.

(4) At the stage of trade-off, the new aspiration level is set in such a way that

$$\bar{f}_i^{k+1} < f_i(x^k) \qquad \text{for any} \quad i \in I_{\mathrm{I}}^k,$$

$$\bar{f}_i^{k+1} > f_i(x^k) \qquad \text{for any} \quad i \in I_{\mathrm{R}}^k,$$

In this event, if $\bar{f}^{k+1}$ is feasible, then in view of Lemma 7.3.4 we have a satisfactory weak Pareto solution by solving the Min–Max problem. In setting the new aspiration level, therefore, it is desirable to pay attention so that $\bar{f}$ becomes feasible.

(5) In order that the new aspiration level $\bar{f}^{k+1}$ may be feasible, the criteria f_i $(i \in I_{\mathrm{R}}^k)$ should be relaxed enough to compensate for the improvement of $f_i(x^k)$ $(i \in I_{\mathrm{I}}^k)$. To make this trade-off successful without solving a new Min–Max problem, we had better make use of sensitivity analysis based on Lemma 7.3.5. Since we already know x^k is a Pareto solution, the feasibility of x^{k+1} can be checked by relation (7.3.14). Here ε is introduced to make relation (7.3.14) available for nonconvex cases in which relation (7.3.12) does not necessarily hold. Moreover, observe in view of Eq. (7.3.11) that

$$\alpha_i = \lambda_i \mu_i \qquad (i = 1, \ldots, p) \tag{7.3.15}$$

represents the efficient trade-off, which reflects the mutual effect of change of each criterion restricted to the Pareto surface. Based on this information of efficient trade-off, the decision maker can easily judge how much the criterion f_i for $i \in I_{\mathrm{I}}^k$ should be improved and how much the criteria f_i for $i \in I_{\mathrm{R}}^k$ should be relaxed. In particular, since a little relaxation of f_j with j such that $\lambda_j = 0$ cannot compensate the improvement of f_i $(i \in I_{\mathrm{I}}^k)$ under the condition that $\bar{f}^{k+1}$ should be on the Pareto surface, we must relax at least one f_j for j such that $\lambda_j \neq 0$ or give a sufficiently large relax for f_j such that $\lambda_j = 0$.

Remark 7.3.4

The stated feasibility check is just for the purpose of decreasing the number of Min–Max problems that must be solved and is not necessarily performed strictly. In some cases, several trials for trade-off cannot succeed in holding relation (7.3.14) and the decision maker would tend to be tired of trading off again and again. In this circumstance, we had better go to solving Min–Max problem immediately even if relation (7.3.14) does not hold. However, in cases that require a lot of effort to solve Min–Max problems, a few retrials of trade-off could not be helped.

Remark 7.3.5

In some cases, decision makers want to know the global feature of the feasible set in trading-off. To this end, it is convenient to show the so-called payoff matrix by minimizing each objective function independently, which was introduced in STEM. Of course, since the payoff matrix does not change during the interactive process, systems analysists had better prepare it in advance of the decision process.

Remark 7.3.6

It may be possible to change the ideal point f^* while the aspiration level $\bar{f}$ is fixed during the decision process. This is considered to be a dual approach of the satisficing trade-off method. However, if the initial aspiration level happens to be a Pareto solution, then we cannot get other Pareto solution by solving the Min–Max problem as long as the aspiration level is fixed. Hence, when the decision maker wants another solution, he ,must change the aspiration level. Taking such a case into account, the way of changing the aspiration level seems better.

7.3.3.2.1 Discussion. As we have seen earlier, the solution obtained by the satisficing trade-off method as well as STEM is not necessarily a Pareto solution but is guaranteed only to be a weak Pareto solution. For weak Pareto solutions, there may exist another solution that improves some criteria while other criteria are left unchanged, and hence weak Pareto solutions seem to be inadequate as a decision making solution. We can get a Pareto solution from a weak Pareto solution as follows:

$$\text{(AP)} \qquad \text{Maximize} \quad \sum_{i=1}^{p} \varepsilon_i \quad \text{subject to}$$

$$f_i(x) + \varepsilon_i = f_i(\hat{x}), \qquad (i = 1, \ldots, p)$$

$$\varepsilon_i \geqq 0, \qquad (i = 1, \ldots, p)$$

$$x \in X,$$

where $\hat{x}$ is a (weak) Pareto solution. If all ε_i for the solution to (AP) are zero, then $\hat{x}$ itself is a Pareto solution. If there are some $\varepsilon_i \neq 0$, then the solution $\tilde{x}$ to problem (AP) is a Pareto solution.

However, in some practical problems it is very expensive to perform such an auxilliary optimization. In order to avoid such an additional scalar optimization problem, some researchers suggested the use of some kinds of augmented norms (Dinkelbach and Isermann [D6]; Steuer and Choo [S14]). One typical example is given by

$$S_a = \max_{1 \leqq i \leqq p} \mu_i|f_i^* - f_i(x)| + \frac{1}{\alpha} \sum_{i=1}^{p} \mu_i|f_i^* - f_i(x)|. \tag{7.3.16}$$

Recall that, in general, in interactive satisficing methods the aspiration level is changed by $\bar{f}^{k+1} = T \circ P(\bar{f}^k)$, where P is an operator that finds the Pareto solution *nearest in some sense* to the aspiration level $\bar{f}^k$. This can be done by using some scalarization function. The sense of nearest strongly depends on what kind of the scalarization function we use. We impose the following requirements on the scalarization functions:

1. They can cover all Pareto solutions.
2. Solutions to the auxilliary scalar optimization should be Pareto solutions.
3. Solutions to the auxiliary scalar optimization should be satisfactory, if the aspiration level is feasible.

Since our decision-making solution is a satisfactory Pareto solution, the scalarization function should hold the property that its maximization or minimization with appropriately chosen parameters can produce any Pareto solution in order to be applicable for any cases no matter what the problem may be and no matter what the preference of the decision maker may be. Furthermore, the solution to the auxilliary scalar optimization for the scalarization function should be nothing but a Pareto solution, because our aim is to find a satisfactory Pareto solution. Finally, if the aspiration level is feasible, then there exists a satisfactory Pareto solution. Therefore, it would not make sense to show the decision maker a nonsatisfactory solution when the aspiration level is feasible. These observations make the preceding three requirements reasonable.

Unfortunately, it should be noted, however, that there is no scalaraization function satisfying all these requirements. It is well known that l_1-norm cannot necessarily yield any Pareto solution if the problem is nonconvex. Even if we use an l_α-norm ($1 < \alpha < \infty$), we cannot always get all Pareto solutions depending on the degree of nonconvexity of the problem. In Fig. 7.12, we can see that the only norm (scalarization function) that can yields any Pareto solution in any problem is the weighted Tchebyshev norm. However, the weighted Tchebyshev norm violates requirement 2. Namely, the weighted Tchebyshev norm produces not only Pareto solutions but also weak Pareto solutions. For any $0 < \alpha < \infty$, the solution obtained by minimizing S_a is guaranteed to be a Pareto solution. However, it can be easily seen that the augmented norm S_a violates requirements 1 and 3 (Fig.

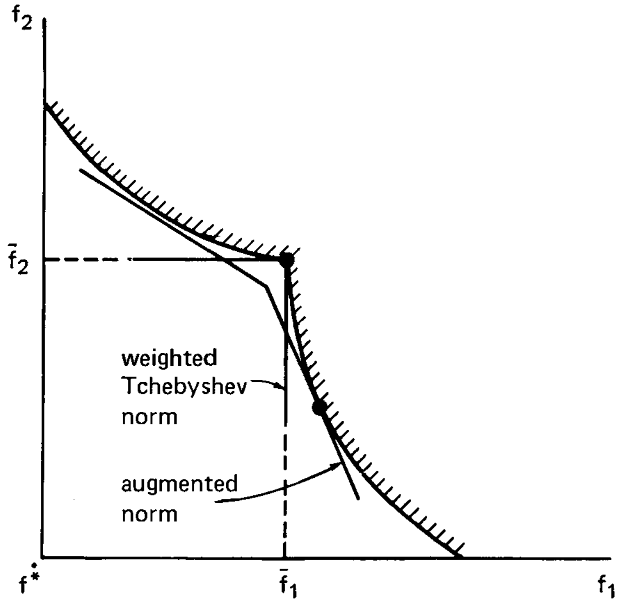

Fig. 7.12.

7.12). Recall here that the weighted Tchebyshev norm satisfies requirement 3 (Lemma 7.3.4).

7.3.4 Interactive Programming Methods as Man–Machine Interfaces

It should be noted that interactive programming methods have a feature as a man–machine interface. Therefore, it goes without saying that it is very important in developing these interactive methods to make the best use of the strong points of men and machines. Machines (computers) are strong at routine, iterative computation and can treat large scale and complex computation with high speed. On the other hand, men are good at global (but, possibly, rough) judgment, pattern recognition, flair, and learning.

With these points in mind, we impose the following properties on desirable interactive multiobjective programming methods:

1. (easy) The way of trading-off is easy. In other words, decision makers can easily grasp the total balance among the objectives.
2. (simple) The judgment and operation required of decision makers is as simple as possible.
3. (understandable) The information shown to decision makers is as intuitive and understandable as possible.

4. (quick response) The treatment by computers is as quick as possible.
5. (rapid convergence) The convergence to the final solution is rapid.
6. (explanatory) Decision makers can easily accept the obtained solution. In other words, they can understand why it is so and what it came from.
7. (learning effect) Decision makers can learn through the interaction process.

Interactive programming methods seem promising, in particular for design problems. However, in applying optimization techniques, we often encounter some difficulties. For example, in structural design problems such as bridges, function forms of some of criteria cannot be obtained explicitly, and their values are usually obtained by complex structural analysis. Similarly, values of criteria in the design of camera lenses are obtained by simulation of ray trace; moreover, the number of criteria is sometimes over 100.

From such a practical viewpoint, many existing interactive optimization methods require too many auxiliary optimizations during the whole interaction process. Moreover, some of them require too high a degree of judgment of decision makers such as the marginal rate of substitution, which seems to be beyond men's ability.

For the purpose of getting a solution, interactive satisficing methods seem advantageous, because they are usually simple and easy to carry out. However, the fact that we can obtain a solution with less information on decision makers' preference is in some cases a demerit as well as a merit. For, in many decision problems, there are multiple persons concerned with those problems, and, hence, it is important to get better mutual understanding and communication. To this end, methods providing the preference information of these people seem rather advantageous. This point is outstanding in particular for problems of public sectors. Since the real world is complex, it is not so easy to conclude which method is the best. Many practical applications could encourage us to sophisticate methodology in future.

This page intentionally left blank

REFERENCES

A1. Allais, M., and Hagen, O. (eds.), "Expected Utility Hypotheses and the Allais Paradox," Reidel, Dordrecht, Holland, 1979.

A2. Arrow, K. J., "Social Choice and Individual Values," Wiley, New York, 1957.

A3. Arrow, K. J., 'Essays in the Theory of Risk-Bearing," Markham, Chicago, 1971.

A4. Arrow, K. J., Barankin, E. W., and Blackwell, D., Admissible points of convex sets, *in* "Contributions to the Theory of Games," H. W. Kuhn and A. W. Tucker, (eds.), Vol. 2, Princeton Univ. Press, Princeton, New Jersey, 1953.

A5. Arrow, K. J., Hurwicz, L., and Uzawa, H., "Studies in Linear and Non-Linear Programming," Stanford Univ. Press, Stanford, California, 1958.

A6. Atkinson, A. B., On the measurement of inequality, *J. Econom. Theory* **2**, 244–263 (1970).

A7. Aubin, J. P., "Mathematical Methods of Game and Economic Theory," North-Holland Publ., Amsterdam, 1979.

B1. Barbu, V., and Precupanu, T., "Convexity and Optimization in Banach Spaces," Sijthoff and Noordhoff, Alphen aan de Rijn, 1978.

B2. Baumol, W. H., The cardinal utility which is ordinal, *Econom. J.* **68**, 665–672 (1958).

B3. Bazaraa, M. S., and Shetty, C. M., "Foundations of Optimization," Springer-Verlag, Berlin and New York, 1976.

B4. Bell, D. E., Multiattribute utility functions: decompositions using interpolation, *Management Sci.* **25**, 744–753 (1979).

B5. Bell, D. E., Keeney, R. L., and Raiffa, H. (eds.), "Conflicting Objectives in Decisions," Wiley, New York, 1976.

B6. Benayoun, R., de Mongolfier, J., Tergny, J., and Larichev, O., Linear programming with multiple objective functions: STEP Method (STEM), *Math. Programming* **1**, 366–375 (1971).

B7. Benson, H. P., An improved definition of proper efficiency for vector minimization with respect to cones, *J. Math. Anal. Appl.* **71**, 232–241 (1979).

B8. Benson, H. P., On a domination property for vector maximization with respect to cones, *J. Optim. Theory Appl.* **39**, 125–132 (1983). Errata Corrige, *J. Optim. Theory* **43**, 477–479 (1984).

B9. Benson, H. P., Efficiency and proper efficiency in vector maximization with respect to cones, *J. Math. Anal. Appl.* **93**, 273–289 (1983).

B10. Benson, H. P., and Morin, T. L., The vector maximization problem: proper efficiency and stability, *SIAM J. Appl. Math.* **32**, 64–72 (1977).

B11. Berge, C., "Topological Spaces," MacMillan, New York, 1963.

B12. Berge, C., "Graphs and Hypergraphs," North-Holland Publ., Amsterdam, 1973.

B13. Bitran, G. R., Duality for nonlinear multiple-criteria optimization problems, *J. Optim. Theory Appl.* **35**, 367–401 (1981).

B14. Bitran, G. R., and Magnanti, T. L., The structure of admissible points with respect to cone dominence, *J. Optim. Theory Appl.* **29**, 573–614 (1979).

B15. Bodily, S., Stability, equality, balance and multivariate risk, *in* "Organizations: Multiple Agents with Multiple Criteria," J. N. Morse, (ed.), Springer-Verlag, Berlin and New York, 1980.

B16. Borwein, J. M., Proper efficient points for maximizations with respect to cones, *SIAM J. Control Optim.* **15**, 57–63 (1977).

B17. Borwein, J. M., The geometry of Pareto efficiency over cones, *Math. Operationsforsch. und Statist. Ser. Optimization* **11**, 235–248 (1980).

B18. Borwein, J. M., On the existence of Pareto efficient points, *Math. Oper. Res.* **8**, 64–73 (1983).

B19. Borwein, J. M., and Nieuwenhuis, J. W., Two kinds of normality in vector optimization, *Math. Programming* **28**, 185–191 (1984).

B20. Bowman, V. J., Jr., On the relationship of the Tchebycheff norm and efficient frontiers of multiple-criteria objectives, *in* "Multiple Criteria Decision Making," H. Thriez and S. Zionts, (eds.), Springer-Verlag, Berlin and New York, 1976.

B21. Brumelle, S., Duality for multiple objective convex programs, *Math. Oper. Res.* **6**, 159–172 (1981).

C1. Cesari, L., and Suryanarayana, M. B., Existence theorems for Pareto optimization in Banach spaces, *Bull. Amer. Math. Soc.* **82**, 306–308 (1976).

C2. Cesari, L., and Suryanayarana, M. B., An existence theorem for Pareto problems, *Nonlinear Anal.* **2**, 225–233 (1978).

C3. Cesari, L., and Suryanayarana, M. B., Existence theorems for Pareto optimization: multivalued and Banach space valued functionals, *Trans. Amer. Math. Soc.* **244**, 37–65 (1978).

C4. Chankong, V., and Haimes, Y. Y., "The Interactive Surrogate Worth Trade-Off (ISWT) for Multiobjective Decision-Making," Springer-Verlag, Berlin and New York, 1978.

C5. Chankong, V., and Haimes, Y. Y., On the characterization of noninferior solutions of the vector optimization problem, *Automatica—J. IFAC* **18**, 697–707 (1982).

C6. Chankong, V., and Haimes, Y. Y., "Multiobjective Decision Making: Theory and Methodology," North-Holland Publ., Amsterdam, 1983.

C7. Charnes, A., and Cooper, W., "Management Models and Industrial Applications of Linear Programming, Vol. I," Wiley, New York, 1961.

C8. Charnes, A., and Cooper, W. W., Goal programming and multiple objective optimization, *J. Oper. Res. Soc.* **1**, 39–54 (1977).

C9. Chipman, J. S., Consumption theory without transitive indifference, *in* "Preferences, Utility, and Demand," J. S. Chipman, L. Hurwicz, M. J. Richter, and H. F. Sonnenschein, (eds.), Harcourt Brace Jovanovich, New York, 1971.

C10. Choo, E. U., and Atkins, D. R., Proper efficiency in nonconvex multicriteria programming, *Math. Oper. Res.* **8**, 467–470 (1983).

C11. Clarke, F. H., "Optimization and Nonsmooth Analysis," Wiley, New York, 1983.

C12. Cochrane, J. L., and Zeleny, M., (eds.), "Multiple Criteria Decision Making," Univ. of South Carolina Press, Columbia, South Carolina, 1973.

C13. Cohon, J. L., "Multiobjective Programming and Planning," Academic Press, New York, 1978.

C14. Cohon, J. L., and Marks, D. H., A review and evaluation of multiobjective programming techniques, *Water Resources Res.* **11**, 208–220 (1975).

C15. Corley, H. W., An existence result for maximizations with respect to cones, *J. Optim. Theory Appl.* **31**, 277–281 (1980).

C16. Corley, H. W., Duality for maximizations with respect to cones, *J. Math. Analy. Appl.* **84**, 560–568 (1981).

C17. Craven, B. D., Strong vector minimization and duality, *Z. Angew. Math. Mech.* **60**, 1–5 (1980).

D1. Da Cunha, N. O., and Polak, E., Constrained minimization under vector-valued criteria in finite-dimensional spaces, *J. Math. Anal. Appl.* **19**, 103–124 (1967).

D2. Dantzig, G. B., Folkman, J., and Shapiro, N., On the continuity of minimum set of continuous functions, *J. Math. Anal. Appl.* **17**, 519–548 (1967).

D3. Debreu, G., "Theory of Value," Wiley, New York, 1959.

D4. Delahare, J. P., and Denel, J., The continuities of the point-to-set maps, definitions and equivalences, *Math. Programming Study.* 10, 8–12 (1979).

D5. Diamond, P. A., Cardinal welfare, individualistic ethics, and interpersonal comparison of utility: comment, *J. Polit. Econom.* **75**, 765–766 (1967).

D6. Dinkelbach, W., and Isermann, H., On decision making under multiple criteria and under incomplete information, *in* "Multiple Criteria Decision Making," J. L. Cochrane and M. Zeleny, (eds.), Univ. of South Carolina Press, Columbia, South Carolina, 1973.

D7. Dyer, J. S., Interactive goal programming, *Management Sci.* **19**, 62–70, (1972).

D8. Dyer, J. S., The effect of errors in the estimation of the gradient on the Frank–Wolfe algorithm, with implications for interactive programming, *Oper. Res.* **22**, 160–174 (1974).

D9. Dyer, J. S., and Sarin, R. K., Group preference aggregation rules based on strength of preference, *Management Sci.* **25**, 822–831 (1979).

D10. Dyer, J. S., and Sarin, R. K., Measurable multiattribute value functions, *Oper. Res.* **27**, 810–822 (1979).

E1. Ekeland, I., and Temam, R., "Convex Analysis and Variational Problems", North-Holland Publ., Amsterdam, 1976.

E2. Evans, J. P., and Gould, F. J., Stability in nonlinear programming, *Oper. Res.* **18**, 107–118 (1970).

F1. Fan, K., On systems of linear inequalities, *in* "Linear Inequalities and Related Systems," H. W. Kuhn and A. W. Tucker, (eds.), Princeton Univ. Press, Princeton, New Jersey, 1956.

F2. Fandel, G., and Gal, T., (eds.), "Multiple Criteria Decision Making Theory and Applications, Proceedings, Bonn, 1979," Springer-Verlag, Berlin and New York, 1979.

F3. Farquhar, P. H., Research directions in multiattribute utility analysis, *in* "Essays and Surveys: Multiple Criteria Decision Making," P. Hansen, (ed.), Springer Publ., New York, 1983.

F4. Fiacco, A. V., "Introduction to Sensitivity and Stability Analysis in Nonlinear Programming," Academic Press, New York, 1983.

F5. Fishburn, P. C., "Decision and Value Theory," Wiley, New York, 1964.

F6. Fishburn, P. C., Methods of estimating additive utilities, *Management Sci.* **13**, 435–453 (1967).

F7. Fishburn, P. C., "Utility Theory for Decision Making," Wiley, New York, 1970.

F8. Fishburn, P. C., "The Theory of Social Choice," Princeton Univ. Press, Princeton, New Jersey, 1973.

F9. Fishburn, P. C., Bernoullian utilities for multiple-factor situations, *in* "Multiple Criteria Decision Making," J. L. Cochrane and M. Zeleny, (eds.), Univ. of South Carolina Press, Columbia, South Carolina, 1973.

F10. Fishburn, P. C., Nontransitive measurable utility, *J. Mathematical Psychology* **26**, 31–67 (1982).

F11. Fishburn, P. C., Transitive measurable utility, *J. Econom. Theory* **31**, 293–317 (1983).

F12. Fishburn, P. C., and Keeney, R. L., Generalized utility independence and some implications, *Oper. Res.* **23**, 928–940 (1975).

F13. Fishburn, P. C., and Vickson, R. G., Theoretical foundations of stochastic dominance, *in* "Stochastic Dominance," G. A. Whitmore and M. C. Findlay, (eds.), Lexington Books, Lexington, Massachusetts, 1978.

F14. French, S., Hartley, R., Thomas, L. C., and White, D. J., "Multi-Objective Decision Making," Academic Press, London, 1983.

G1. Gale, D., "The Theory of Linear Economic Models," McGraw-Hill, New York, 1960.

G2. Gale, D., Kuhn, H. W., and Tucker, A. W., Linear programming and the theory of games, *in* "Activity Analysis of Production and Allocation," T. C. Koopmans, (ed.), Wiley, New York, 1951.

G3. Gearhart, W. B., Compromise solutions and estimation of the noninferior set, *J. Optim. Theory Appl.* **28**, 29–47 (1979).

G4. Gearhart, W. B., Characterization of properly efficient solutions by generalized scalarization methods, *J. Optim. Theory Appl.* **41**, 491–502 (1983).

G5. Geoffrion, A. M., Proper efficiency and the theory of vector maximization, *J. Math. Anal. Appl.* **22**, 618–630 (1968).

G6. Geoffrion, A. M., Duality in nonlinear programming: a simplified applications oriented development, *SIAM Rev.* **13**, 1–37 (1974).

G7. Geoffrion, A. M., Dyer, J. S., and Feinberg, A., An interactive approach for multicriterion optimization, with an application to the operation of an academic department, *Management Sci.* **19**, 357–368 (1972).

G8. Goicoeche, A. G., Hansen, D. R., and Duckstein, L., "Multiobjective Decision Analysis with Engineering and Business Applications," Wiley, New York, 1982.

G9. Grauer, M., Lewandowski, A., and Wierzbicki, A. P., DIDASS-theory, implementation and experiences, *in* [G10].

G10. Grauer, M., and Wierzbicki, A. P., (eds.), "Interactive Decision Analysis; Proceedings of an International Workshop on Interactive Decision Analysis and Interpretative Computer Intelligence, IIASA, Laxenburg, Austria 1983." Springer-Verlag, Berlin and New York, 1984.

G11. Gros, C., Generalization of Fenchel's duality theorem for convex vector optimization, *European J. Oper. Res.* **2**, 368–376 (1978).

G12. di Guglielmo, F., Nonconvex duality in multiobjective optimization, *Math. Oper. Res.* **2**, 285–291 (1977).

H1. Hadar, J., and Russell, W. R., Stochastic dominance and diversification, *J. Econom. Theory* **3**, 288–305 (1971).

H2. Haimes, Y. Y., Hall, W. A., and Freedman, H. B., "Multiobjective Optimization in Water Resources Systems, The Surrogate Worth Trade-off Method," Elsevier Scientific, New York, 1975.

H3. Hansen, P., "Essays and Surveys on Multiple Criteria Decision Making," Springer-Verlag, Berlin and New York, 1983.

H4. Harsarnyi, J. C., Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility, *J.* Polit. Economy **63**, 309–321 (1955).

H5. Hartley, R., On cone-efficiency, cone-convexity, and cone-compactness, *SIAM J. Appl. Math.* **34**, 211–222 (1978).

H6. Hazen, G. B., and Morin, T. L., Optimality conditions in nonconical multiple-objective programming, *J. Optim. Theory Appl.* **40**, 25–60 (1983).

H7. Henig, M. I., Proper efficiency with respect to cones, *J. Optim. Theory Appl.* **36**, 387–407 (1982).

H8. Henig, M. I., Existence and characterization of efficient decisions with respect to cones, *Math. Programming* **23**, 111–116 (1982).

H9. Henig, M. I., The domination property in multicriterion optimization, The Israel Institute of Business Research, Working Paper No. 781/83 (1983).

H10. Herstein, I. N., and Milnor, J., An axiomatic approach to measurable utility, *Econometrica* **21**, 291–297 (1953).

H11. Hogan, W. W., Point-to-set maps in mathematical programming, *SIAM Rev.* **15**, 591–603 (1973).

H12. Holmes, R. B., "A Course on Optimization and Best Approximation," Springer-Verlag, Berlin and New York, 1972.

H13. Huang, C. C., Kira, D., and Vertinsky, I., Stochastic dominance rules for multiattribute utility functions, *Rev. Econom. Stud.* XLV, 611–615 (1978).

H14. Huard, P., (ed.), "Point-to-Set Maps and Mathematical Programming," *Math.* Programming Stud. **10** (1979).

H15. Huard, P., Background to point-to-set maps in mathematical programming, *Math. Programming Stud.* **10**, 1–7 (1979).

H16. Hurwicz, L., Programming in linear spaces, *in* "Studies in Linear and Nonlinear Programming," K. J. Arrow, L. Hurwicz, and H. Uzawa, (eds.), Stanford Univ. Press, Stanford, California, 1958.

H17. Hwang, C. L., and Masud, A. S. M., "Multiple Objective Decision Making Methods and Applications," Springer-Verlag, Berlin and New York, 1979.

H18. Hwang, C. L., and Yoon, K., "Multiple Attribute Decision Making: Methods and Applications, State-of-the-Art Survey," Springer-Verlag, Berlin and New York, 1981.

I1. Ignizio, J. P., "Goal Programming and Extensions," Lexington Books, Lexington, Massachusetts, 1976.

I2. Ignizio, J. P., Generalized goal programming, *Comp. Oper. Res.* **10**, 277–291 (1983).

I3. Ijiri, Y., "Management Goals and Accounting for Control," North-Holland Publ., Amsterdam, 1965.

I4. Ioffe, A. D., and Tihomirov, V. M., "Theory of External Problems," North-Holland Publ., Amsterdam, 1979.

I5. Isermann, H., The relevance of duality in multiple-objective linear programming, *in* "Multiple Criteria Decision Making," M. K. Starr and M. Zeleny, (eds.), North-Holland Publ., Amsterdam, 1977.

I6. Isermann, H., On some relations between a dual pair of multiple objective linear programs, *Z. Oper. Res.* **22**, 33–41 (1978).

I7. Isermann, H., Duality in multiple objective linear programming, *in* "Multiple Criteria Problem Solving," S. Zionts, (ed.), Springer-Verlag, Berlin and New York, 1978.

J1. Jahn, J., Duality in vector optimization, *Math. Programming* **25**, 343–353 (1983).

J2. Jahn, J., Scalarization in vector optimization, *Math. Programming* **29**, 203–218 (1984).

J3. Jurkiewicz, E., Stability of compromise solution in multicriteria decision-making problems, *J. Optim. Theory Appl.* **40**, 77–83 (1983).

K1. Kahneman, D., and Tversky, A., Prospect theory: an analysis of decision under risk, *Econometrica* **47**, 263–291 (1979).

K2. Kawasaki, H., Conjugate relations and weak subdifferentials, *Math. Oper. Res.* **6**, 593–607 (1981).

K3. Kawasaki, H., A duality theorem in multiobjective nonlinear programming, *Math. Oper. Res.* 7, 95–110 (1982).

K4. Keeney, R. L., A group preference axiomatization with cardinal utility, *Management Sci.* **23**, 140–145 (1976).

K5. Keeney, R. L., Equity and public risk, *Oper. Res.* **28**, 527–534 (1980).

K6. Keeney, R. L., and Raiffa, H., "Decision with Multiple Objectives: Preferences and Value Tradeoffs," Wiley, New York, 1975.

K7. Klinger, A., Improper solutions of the vector maximum problems, *Oper. Res.* **15**, 570–572 (1967).

K8. Kornbluth, J. S. H., Duality, indifference and sensitivity analysis in multiple objective linear programming, *Oper. Res. Quart.* **25**, 599–614 (1974).

K9. Krantz, D., Luce, D., Suppes, P., and Tversky, A., "Foundations of Measurement," Vol. 1, Academic Press, New York, 1971.

K10. Kuhn, H. W., and Tucker, A. W., Nonlinear programming, *in* "Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability," Berkeley, California, pp. 481–492 (1951).

L1. Lee, S. M., "Goal Programming for Decision Analysis," Auerbah Publishers, Philadelphia, 1972.

L2. Leitmann, G., (ed.), "Multicriteria Decision Making and Differential Games," Plenum, NewYork, 1976.

L3. Leitmann, G., and Marzollo, A., (eds.), "Multicriteria Decision Making," Springer-Verlag, Berlin and New York, 1975.

L4. Levy, H., and Paroush, J., Toward multivariate efficiency criteria, *J. Econom. Theory* **7**, 129–142 (1974).

L5. Lewandowski, A., and Grauer, M. The reference point optimization approach-methods of efficient implementation, *IIASA* WP-82-26 (1982).

L6. Lin, J. G., Multiple-objective problems: Pareto-optimal solutions by the method of proper equality constraints, *IEEE Trans. Automatic Control* **AC-21**, 641–650 (1976).

L7. Lin, J. G., Proper equality constraints and maximization of index vectors, *J. Optim. Theory Appl.* **20**, 215–244 (1976).

L8. Lin, J. G., Multiple-objective optimization by a multiplier method of proper equality constraints – Part I: Theory, *IEEE Trans. Automatic Control* **AC-24**, 567–573 (1979).

L9. Luc, D. T., On the domination property in vector optimization, *J. Optim. Theory Appl.* **43**, 327–330 (1984).

L10. Luc, D. T., Duality theory in multiobjective programming, *J. Optim. Theory Appl.* **43**, 557–582 (1984).

L11. Luce, R. D., Semiorders and a theory of utility discrimination, *Econometrica* **24**, 178–191 (1956).

L12. Luce, D., and Raiffa, H., "Games and Decisions; Introduction and Critical Survey," Wiley, London, 1957.

L13. Luce, R. D., and Suppes, P., Preference, utility and subjective probability, *in* "Handbook of Mathematical Psychology III," R. D. Luce, R. R. Bush, and E. Galanter, (eds.), Wiley, New York, 1965.

L14. Luenberger, D. G., "Optimization by Vector Space Methods," Wiley, New York, 1969.

M1. Machina, M. J., Expected utility analysis without the independence axiom, *Econometrica* **50**, 277–323 (1982).

M2. Mangasarian, O. L., "Nonlinear Programming," McGraw-Hill, New York, 1969.

M3. Mazzoleni, P., Duality and reciprocity for vector programming, *European J. Oper. Res.* **10**, 42–50 (1980).

M4. Minami, M., Weak Pareto-optimal necessary conditions in a nondifferentiable multiobjective program on a Banach space, *J. Optim. Theory Appl.* **41**, 451–461 (1983).

M5. Morse, J. N., (ed.), "Organizations: Multiple Agents with Multiple Criteria, Proceedings, University of Delaware, Newark, 1980." Springer-Verlag, Berlin and New York, 1981.

N1. Naccache, P. H., Connectedness of the set of nondominated outcomes in multicriteria optimization, *J. Optim. Theory Appl.* **25**, 459–467 (1978).

N2. Naccache, P. H., Stability in multicriteria optimization, *J. Math. Anal. Appl.* **68**, 441–453 (1979).

N3. Nakayama, H., Subjective programming in multi-criterion decision making, *in* F2, 252–265 (1980).

N4. Nakayama, H., Proposal of satisficing trade-off method for multiobjective programming, *Trans. Soc. Inst. Control Eng.* **20**, 29–35 (1984), (in Japanese).

N5. Nakayama, H., Geometric consideration of duality in vector optimization, *J. Optim. Theory Appl.* **44**, 625–655 (1984).

N6. Nakayama, H., Duality in linear vector optimization, *IIASA Working Paper* **WP 84-86** (1984).

N7. Nakayama, H., Lagrange duality and its geometric interpretation, *in* "Proceedings of the seminar of Mathematics of Multiobjective Optimization," CISM (to be published).

N8. Nakayama, H., and Furukawa, K., Satisficing trade-off method applied to structural design problems in civil engineering, *Large Scale Systems* (to be published).

N9. Nakayama, H., Takeguchi, T., and Sano, M., Intersective graphics for portfolio selection, *in* H3, 280–289 (1983).

N10. Nakayama, H., Tanino, T., Matsumoto, K., Matsuo, H., Inoue, K., and Sawaragi, Y., Methodology for group decision support with an application to assessment of residential environment, *IEEE Trans. Systems Man Cybernet.* **SMC-9**, 477–485 (1979).

N11. Nakayama, H., Tanino, T., and Sawaragi, Y., An interactive optimization method in multicriteria decision making, *IEEE Trans. Systems Man Cybernet.* **SMC-10**, 163–169 (1980).

N12. Nakayama, H., Tanino, T., and Sawaragi, Y., Stochastic dominance for decision problems with multiple attributes and/or multiple decision makers, *in* "Proceedings of the Eighth Triennial World Congress of IFAC," pp. 1397–1402 (1982).

N13. von Neumann, J., and Morgenstern, O., "Theory of Games and Economic Behavior," Princeton Univ. Press, Princeton, New Jersey, 1947.

N14. Neustadt, L. W., "Optimization," Princeton Univ. Press, Princeton, New Jersey, 1976.

N15. Nieuwenhuis, J. W., Supremal points and generalized duality, *Math. Operationsforsch. Statist. Ser. Optimization* **11**, 41–59 (1980).

N16. Nieuwenhuis, J. W., Some Minimax Theorems in Vector-Valued Functions, *J. Optim. Theory Appl.* **40**, 463–475 (1983).

N17. Nieuwenhuis, J. W., About Isermann duality, *J. Optim. Theory Appl.* **41**, 481–490 (1983).

N18. Nijkamp, P., and van Delft, A., "Multi-criteria Analysis and Regional Decision Making," Martinus Hijhoff Social Sciences Division, Leiden, 1977.

O1. Oppenheimer, K. R., A proxy approach to multiattribute decision making, *Management Sci.* **24**, 675–689 (1978).

P1. Penrose, R., A generalized inverse for matrices, *Proc. Cambridge Philos. Soc.* **51**, 406–413 (1955).

P2. Ponstein, J., "Approaches to the Theory of Optimization," Cambridge Univ. Press, London and New York, 1980.

P3. Ponstein, J., On the dualization of multiobjective optimization problems, University of Groningen, Econometric Institute, Report 87 (OR-8201), 1982.

P4. Ponstein, J., Infima of sets, proper efficiency and multiobjective optimization, University of Groningen, Econometric Institute, Report 88 (OR-8202), 1982.

P5. Pratt, J. W., Risk aversion in the small and in the large, *Econometrica* **32**, 122–136 (1964).

R1. Raiffa, H., "Decision Analysis: Introductory Lectures on Choice Under Uncertainty," Addison-Wesley, Reading, Massachusetts, 1968.

R2. Richard, S. F., Multivariate risk aversion, *Management Sci.* **22**, 12–21 (1975).

R3. Rietveld, P., "Multiple Objective Decision Method and Regional Planning," North-Holland Publ., Amsterdam, 1980.

R4. Ritter, K., Optimization theory in linear spaces I, *Math. Ann.* **182**, 189–206 (1969).

R5. Ritter, K., Optimization theory in linear spaces II, *Math. Ann.* **183**, 169–180 (1969).

R6. Ritter, K., Optimization theory in linear spaces III, *Math. Ann.* **184**, 133–154 (1970).

R7. Rockafellar, R. T., "Convex Analysis," Princeton Univ. Press, Princeton, New Jersey, 1970.

R8. Rockafellar, R. T., Conjugate Duality and Optimization, *in* "CBMS Lecture Notes Series No. 16," SIAM, Philadelphia, 1974.

R9. Rödder, W., A generalized saddlepoint theory, *European J. Oper. Res.* **1**, 55–59 (1977).

R10. Rosinger, E. E., Duality and alternative in multiobjective optimization, *Proc. Amer. Math. Soc.* **64**, 307–312 (1977).

R11. Ruys, P. H. M., and Weddepohl, H. N., Economic theory and duality, *in* "Convex Analysis and Mathematical Economics," J. Kriens, (ed.), Springer-Verlag, Berlin and New York, 1979.

S1. Saaty, T. L., A scaling method for priorities in hierarchical structures, *J. Mathematical Psychology* **15**, 234–281 (1977).

S2. Sakai, Y., "Economics under Uncertainty," Yuhikaku, (in Japanese), 1982.

S3. Salukvadze, M. E., "Vector-Valued Optimization Problems in Control Theory," Academic Press, New York, 1979.

S4. Sarin, R. K., Strength of preference and risky choice, *Oper. Res.* **30**, 982–997 (1982).

S5. Sawaragi, Y., Nakayama, H., and Tanino, T., Modeling and management control of environmental systems, *in* "Optimization and Control of Dynamic Operational Research Models," S. G. Tzafestas, (ed.), North-Holland Publ., Amsterdam, 1982.

S6. Schönfeld, P., Some duality theorem for the non-linear vector maximum problem, *Unternehmensforsch.* **14**, 51–63 (1970).

S7. Sen, A., "Collective Choice and Social Welfare," Holden-Day, San Francisco, California, 1970.

S8. Shimizu, K., "Theory of Multiobjective and Conflict," Kyoritsu Syuppan, (in Japanese), 1982.

S9. Simon, H., "Models of Man," Wiley, New York, 1957.

S10. Smale, S., Global analysis and economics I: Pareto optimum and a generalization of Morse theory, *in* "Dynamical Systems," M. Peixoto, (ed.), Academic Press, New York, 1973.

S11. Smale, S., Global analysis and economics III: Pareto optima and price equilibria, *J. Math. Econom.* **1**, 213–221 (1974).

S12. Sposito, V. A., and David, H. T., A note on Farkas lemma over cone domains, *SIAM J. Appl. Math.* **22**, 356–358 (1972).

S13. Starr, M. K., and Zeleney, M., (eds.), "Multiple Criteria Decision Making," North-Holland Publ., Amsterdam, 1977.

S14. Steuer, R. E., and Choo, E., An interactive weighted Tchebysheff procedure for multiple objective programming, *Math. Programming* **26**, 326–344 (1983).

S15. Stoer, J., and Witzgall, C., "Convexity and Optimization in Finite Dimensions I," Springer-Verlag, Berlin and New York, 1970.

T1. Takeguchi, T., and Akashi, H., Multi-attribute decision making by stochastic dominance, *Trans. Soc. Instr. Control Eng.* **19**, 13–19 (1983), (in Japanese).

T2. Takeguchi, T., and Akashi, H., Analysis of decisions under risk with incomplete knowledge, *IEEE Systems Man Cybernet.* (to be published).

T3. Tamura, K., and Arai, S., On proper and improper efficient solutions of optimal problems with multicriteria, *J. Optim. Theory Appl.* **38**, 191–205 (1982).

T4. Tamura, K., and Miura, S., Necessary and sufficient conditions for local and global nondominated solutions in decision problems with multiobjectives, *J. Optim. Theory Appl.* **28**, 501–523 (1979).

T5. Tamura, H., and Nakamura, N., Decompositions of multiattribute utility functions based on convex dependence, *Oper. Res.* **31**, 488–506 (1983).

T6. Tang, Y., Conditions for constrained efficient solutions of multi-objective problems in Banach spaces, *J. Math. Anal. Appl.* **96**, 505–519 (1983).

T7. Tanino, T., Saddle points and duality in multi-objective programming, *Internat. J. Systems Sci.* **13**, 323–335 (1982).

T8. Tanino, T., Fuzzy preference orderings in group decision making, *Internat. J. Fuzzy Sets Systems* **12**, 117–131 (1984).

T9. Tanino, T., and Sawaragi, Y., Duality theory in multiobjective programming, *J. Optim. Theory Appl.* **27**, 509–529 (1979).

T10. Tanino, T., and Sawaragi, Y., Stability of nondominated solutions in multicriteria decision-making, *J. Optim. Theory Appl.* **30**, 229–253 (1980).

T11. Tanino, T., and Sawaragi, Y., Conjugate maps and duality in multiobjective optimization, *J. Optim. Theory Appl.* **31**, 473–499 (1980).

T12. Tanino, T., Nakayama, H., and Sawaragi, Y., On methodology for group decision support, *in* "Organizations: Multiple Agents with Multiple Criteria," J. N. Morse, (ed.), Springer-Verlag, Berlin and New York, 1981.

T13. Tanino, T., Nakayama, H., and Sawaragi, Y., Parametric preference orderings in group decision making, *in* "Proceedings of Eighth Triennial World Congress of IFAC," pp. 1749–1754 (1982).

T14. Thiriez, H., and Zionts, S., (eds.), "Multiple Criteria Decision Making," Jouy-en-Josas, France, Springer-Verlag, Berlin and New York, 1976.

T15. Tversky, A., Intransitivity of preferences," *Psychological Rev.* **76**, 31–48 (1969).

W1. Wagner, D. H., Semicompactness with respect to a Euclidean cone, *Canad. J. Math.* **29**, 29–36 (1977).

W2. Wallenius, J., Comparative evaluation of some interactive approaches to multicriterion optimization, *Management Sci.* **21**, 1387–1396 (1973).

W3. Wan, Y. H., On local Pareto optima, *J. Math. Econom.* **2**, 35–42 (1975).

W4. Warburton, A. R., Quasiconcave vector maximization: connectedness of the sets of Pareto-optimal and weak Pareto-optimal alternatives, *J. Optim. Theory Appl.* **40**, 537–557 (1983).

W5. Warfield, J. N., "Societal Systems – Planning, Policy and Complexity," Wiley, New York, 1976.

W6. Wehrung, D. A., Interactive identification and optimization using a binary preference relation, *Oper. Res.* **26**, 322–332 (1978).

W7. Wehrung, D. A., Multidimensional line search using a binary preference relation, *Oper. Res.* **27**, 356–363 (1979).

W8. Wendell, R. E., and Lee, D. N., Efficiency in multiple objective optimization problems, *Math. Programming* **12**, 406–414 (1977).

W9. White, D. J., "Optimality and Efficiency," Wiley, New York, 1982.

W10. Whitmore, G. A., Third-degree stochastic dominance, *Amer. Econom. Rev.* **60**, 457–459 (1970).

W11. Whitmore, G. A., and Findlay, M. C., "Stochastic Dominance," Lexington Books, Lexington, Massachusetts, 1978.

W12. Wierzbicki, A. P., Basic properties of scalazing functionals for multiobjective optimization, *Math. Oper. Statist. Ser. Optim.* **8**, 55–60 (1977).

W13. Wierzbicki, A., The use of reference objectives in multiobjective optimization, *in* "Multiple Criteria Decision Making Theory and Application," G. Fandel, and T. Gal, (eds.), Springer-Verlag, Berlin and New York, 1980.

W14. Wierzbicki, A., A mathematical basis for satisficing decision making, *in* "Organizations: Multiple Agents with Multiple Criteria," J. Morse, (ed.), Springer Publ., New York, 1981.

W15. Wilhelm, J., "Objectives and Multi-objective Decision Making under Uncertainty," Springer-Verlag, Berlin and New York, 1975.

Y1. Yu, P. L., Cone convexity, cone extreme points, and nondominated solutions in decision problems with multiple objectives, *J. Optim. Theory Appl.* **14**, 319–377 (1974).

Y2. Yu, P. L., Second order game problem decision dynamics in gaming phenomena, *J. Optim. Theory Appl.* **27**, 144–166 (1979).

Y3. Yu, P. L., and Leitmann, G., Compromise solutions, domination structures, and Salukvadze's solution, *J. Optim. Theory Appl.* **13**, 362–378 (1974).

Y4. Yu, P. L., and Zeleny, M., The set of all undominated solutions in linear cases and the multicriteria simplex method, *J. Math. Anal. Appl.* **49**, 430–468 (1974).

Z1. Zadeh, L. A., Optimality and nonscalar-valued performance criteria, *IEEE Trans. Automat. Control* **AC-8**, 59–60 (1963).

Z2. Zangwill, W. I., "Nonlinear Programming: A Unified Approach," Prentice-Hall, Englewood Cliffs, New Jersey, 1969.

Z3. Zeleny, M., Compromise programming, *in* "Multiple Criteria Decision Making," J. L. Cochrane and M. Zeleny, (eds.), Univ. of South Carolina Press, Columbia, South Carolina, 1973.

Z4. Zeleny, M., "Linear Multi-objective Programming," Springer-Verlag, Berlin and New York, 1974.

Z5. Zeleny, M., (ed.), "Multiple Criteria Decision Making: Kyoto," Springer-Verlag, Berlin and New York, 1976.

Z6. Zeleny, M., "Multiple Criteria Decision Making," McGraw-Hill, New York, 1982.

Z7. Zionts, S., (ed.), "Multiple Criteria Problem Solving: Proceedings, Buffalo, NY 1977," Springer-Verlag, Berlin and New York, 1978.

Z8. Zionts, S., and Wallenius, J., An interactive programming method for solving the multiple criteria problem, *Management Sci.* **22**, 652–663 (1976).

Z9. Zowe, J., A duality theory for a convex programming in order-complete vector lattices, *J. Math. Anal. Appl.* **50**, 273–287 (1975).

Z10. Zowe, J., The saddle-point theorem of Kuhn and Tucker in ordered vector spaces, *J. Math. Anal. Appl.* **57**, 41–55 (1977).

This page intentionally left blank

INDEX

E

F

G

H

I

J

K

L

M