IIASA COLLABORATIVE PROCEEDINGS SERIES

CP–82–S12

MULTIOBJECTIVE AND STOCHASTIC OPTIMIZATION

IIASA COLLABORATIVE PROCEEDINGS SERIES

CP–81–S1 LARGE-SCALE LINEAR PROGRAMMING
Proceedings of an IIASA Workshop
G.B. Dantzig, M.A.H. Dempster, and M.J. Kallio, *Editors*

CP–81–S2 THE SHINKANSEN PROGRAM: TRANSPORTATION, RAILWAY, ENVIRONMENTAL, REGIONAL, AND NATIONAL DEVELOPMENT ISSUES
A. Straszak, *Editor*

CP–82–S1 HUMAN SETTLEMENT SYSTEMS: SPATIAL PATTERNS AND TRENDS
Selected Papers from an IIASA Conference
T. Kawashima and P. Korcelli, *Editors*

CP–82–S2 RISK: A SEMINAR SERIES
H. Kunreuther, *Editor*

CP–82–S3 THE OPERATION OF MULTIPLE RESERVOIR SYSTEMS
Proceedings of an International Workshop, Jodlowy Dwor, Poland
Z. Kaczmarek and J. Kindler, *Editors*

CP–82–S4 NONPOINT NITRATE POLLUTION OF MUNICIPAL WATER SUPPLY SOURCES: ISSUES OF ANALYSIS AND CONTROL
Proceedings of an IIASA Task Force Meeting
K.-H. Zwirnmann, *Editor*

CP–82–S5 MODELING AGRICULTURAL–ENVIRONMENTAL PROCESSES IN CROP PRODUCTION
Proceedings of an IIASA Task Force Meeting
G. Golubev and I. Shvytov, *Editors*

CP–82–S6 LIQUEFIED ENERGY GASES FACILITY SITING: INTERNATIONAL COMPARISONS
H. Kunreuther, J. Linnerooth, and R. Starnes, *Editors*

CP–82–S7 ENVIRONMENTAL ASPECTS IN GLOBAL MODELING
Proceedings of the 7th IIASA Symposium on Global Modeling
G. Bruckmann, *Editor*

CP–82–S8 PROGRESS IN NONDIFFERENTIABLE OPTIMIZATION
E.A. Nurminski, *Editor*

CP–82–S9 INNOVATION POLICY AND COMPANY STRATEGY
H. Maier and J. Robinson, *Editors*

CP–82–S10 THE KINKI INTEGRATED REGIONAL DEVELOPMENT PROGRAM
Y. Sawaragi and A. Straszak, *Editors*

CP–82–S11 EUROPEAN AND UNITED STATES CASE STUDIES IN APPLICATION OF THE CREAMS MODEL
V. Svetlosanov and W.G. Knisel, *Editors*

CP–82–S12 MULTIOBJECTIVE AND STOCHASTIC OPTIMIZATION
Proceedings of an IIASA Task Force Meeting
M. Grauer, A. Lewandowski, and A.P. Wierzbicki, *Editors*

MULTIOBJECTIVE AND STOCHASTIC OPTIMIZATION

Proceedings of an IIASA Task Force Meeting
November 30–December 4, 1981

M. Grauer, A. Lewandowski,
and A.P. Wierzbicki
Editors

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
Laxenburg, Austria
1982

International Standard Book Number 3-7045-0057-7

Collaborative papers in this *Special* series sometimes report work done at the International Institute for Applied Systems Analysis and sometimes work done elsewhere. They are reviewed at IIASA, but receive only limited external review, and are issued after limited editorial attention. The views or opinions they express do not necessarily represent those of the Institute, its National Member Organizations, or other organizations supporting the work.

PREFACE

The International Institute for Applied Systems Analysis (IIASA) has long been involved in the study of problems with conflicting objectives -- indeed, this research was initiated by the first director of the Institute, Howard Raiffa, whose own particular interests lie in this area. Problems with conflicting objectives arise in fields as disparate as economics and engineering, and are central to many applications of decision analysis, planning theory, and conflict management. Methods for handling these problems are therefore very important; the development of quantitative approaches in this area has been the main aim of the IIASA study.

The early stages of this research were discussed at an IIASA Workshop in 1975 and summarized in a book "Conflicting Objectives in Decisions", edited by D.E. Bell, R.L. Keeney and H. Raiffa, and published by Wiley in 1977. A Task Force Meeting with the general title "Multiobjective and Stochastic Optimization" was held at IIASA at the end of 1981 to review more recent work in the field -- this volume contains the Proceedings of the Task Force Meeting.

These Proceedings report the work of scientists from many different countries and describe a variety of approaches to one basic problem. It is hoped that this book provides a guide to the various schools of thought in multiobjective analysis and will be useful to practitioners working in this field.

FOREWORD

During the week 30 November-4 December 1981, the System and Decision Sciences group at IIASA organized a Task Force Meeting on Multiobjective and Stochastic Optimization. The participants came from all over the world, but had one thing in common -- an active interest in multiobjective and stochastic optimization methodology, algorithms and software.

The field of multiobjective analysis and optimization under conditions of uncertainty is currently expanding very fast. For this reason, it was decided to publish the Proceedings in a lecture note format (without editing) so that a complete record of the papers presented at the meeting would be available relatively rapidly. In some cases, the papers were revised by their authors following the meeting; however, many contributions have not undergone revision and are reproduced here in their original form.

The book is divided into four main sections, the first of which contains five papers dealing with the theoretical aspects of multiobjective and stochastic optimization. The seven papers included in Section II are concerned with those aspects of multiobjective analysis which have a direct relationship to decision

making (some papers in Sections III and IV are also linked to decision making or decision support, although not so directly). Section III contains four papers dealing with uncertainties and multiobjective analysis. The first and last papers in this section also present solution techniques which are illustrated by means of examples. The final section contains papers which concentrate on solution techniques and indicate how they can be applied to practical problems; the software presented in this section can be regarded as a step toward computerized decision support systems. Naturally, some of the papers in other sections also touch on applications of multiobjective and stochastic optimization: examples are drawn from a wide range of activities, including regional planning, environmental control, wage negotiation and energy planning.

The Editors wish to take this opportunity to thank all of the contributors for participating in the Task Force Meeting and for permitting IIASA to publish their work in these Proceedings. They would also like to thank Gabi Adam for her help in arranging the meeting, Helen Gasking for supervising the publication of Proceedings, and Edith Gruber for her assistance in the organization and coordination of the meeting.

Manfred Grauer
Andrzej Lewandowski
Andrzej Wierzbicki

CONTENTS

I

THEORETICAL ASPECTS OF MULTIOBJECTIVE AND STOCHASTIC OPTIMIZATION

MULTIOBJECTIVE TRAJECTORY OPTIMIZATION AND MODEL SEMIREGULARIZATION

A.P. Wierzbicki

International Institute for Applied Systems Analysis, Laxenburg, Austria and
Technical University of Warsaw, Poland

1. MOTIVATION

Dynamic optimization problems are usually formulated in terms of minimization (or maximization) of a given objective functional, also called performance functional. Even if the performance of a dynamic system is specified in terms of closeness to a given trajectory, a performance functional corresponding to a distance from this trajectory is still being used. However, not all practical problems can be usefully formulated as optimization problems with given performance functionals.

Very often, particularly in economic applications, the purpose of optimization is not to propose 'the optimal solution', but rather to generate reasonable alternatives in response to users' requirements while eliminating clearly inferior alternatives. It is not likely that a user would specify his requirements in form of a performance functional. More likely, he would specify his aspirations in form of a reasonable or desirable trajectory of the dynamic system being investigated. Since the desirable trajectory reflects his judgment and experience, it might not be attainable for a particular model of the dynamic system being studied. However, if the desirable trajectory happens to be attainable, the user can often specify also what trajectories should be considered as naturally better than the desirable ones.

As an example, consider a dynamic economic model that specifies, for various monetary and fiscal policies, the resulting economic growth and inflation rates. An economist, while working with this model, is perfectly able to specify reasonable growth and inflation rates trajectories although these trajectories may not be attainable for the model. If they are attainable however, he would not be satisfied by them, particularly if he knew that he could obtain either higher growth rate or lower inflation rate or both. Thus, we cannot use the classical device of minimization of a performance functional corresponding to the distance from the desired trajectory; this device works well only when the desired trajectory is naturally better than the attainable ones. Another classical device is the formulation of a social welfare functional and its maximization; but the information needed for formulating the social welfare functional is much larger than the information contained in a desirable trajectory. Moreover, a social welfare functional implies 'the optimal solution' without allowing for the possibility of checking various alternatives by changing the desired trajectory.

Therefore, a concept of multiobjective trajectory optimization based on reference trajectories has been recently introduced (Wierzbicki 1979) and practically applied to some issues in economic modeling (Kallio et al. 1980). This concept, while being strongly related to some basic concepts in satisficing decision making (Wierzbicki 1980), deserves a separate study. The purpose of this paper is to present, in more detail, the theory, some computational approaches and applicational aspects of multiobjective trajectory optimization.

2. BASIC THEORY IN A NORMED SPACE

All the theory in this section could be introduced in referring to a more detailed dynamic model, for example, the classical control model described by an ordinary differential state equation and an output equation. However, the precise form of a dynamic model does not matter, and the theory is also applicable for models described by difference-differential equations (with delays), by partial differential equations, integral equations, etc.

To obtain a possible compact presentation of basic ideas, let us start with an abstract formulation in normed spaces. Let $u \in E_u$ be a *control trajectory*, shortly called *control*; E_u is a Banach space, say, the space of essentially bounded functions $L^\infty([t_0;t_1],R^m)$, or the space of square integrable functions $L^2([t_0,t_1],R^m)$, etc. Additionally, control constraints $u \in V \subset E_u$ might be given. Let $x \in E_x$ be a *state trajectory*, shortly *state*, defined by a mapping $X:E_u \to E_x$, $x = X(u)$. Conditions, under which the mapping X corresponds to a model of a *dynamic system* and can be expressed as a resolving operation for a *state equation* are given, for example, in Kalman et al. 1969, and will not be discussed here. A proper choice of a Banach space E_x might be the Sobolev space of absolutely continuous functions with essentially bounded derivatives $W^\infty([t_0;t_1],R^n)$ or with square integrable derivatives $W^2([t_0;t_1],R^n)$ -- see, e.g., Wierzbicki, 1977b. However, these properties are needed only for a more detailed development of the form of the dynamic model, and, at this stage of abstraction, E_u and E_x could be just any linear topological spaces.

More important are the assumptions concerning *output trajectory*, shortly *output* $y \in E_y$, defined as a result of a mapping $Y:E_x \times E_u \to E_y$, $y = Y(x,u)$. A properly chosen Banach space E_y should have the same character as the space E_u; thus, $E_y = L^\infty([t_0;t_1],R^p)$ or $E_y = L^2([t_0;t_1],R^p)$. Since the notion of an output is relative to the purpose of the model, we might consider only those output variables that are relevant for the purpose of multiobjective trajectory optimization, the number of those variables being p. Thus, a notion of a partial preordering (partial ordering of equivalence classes) is assumed to be given in the output space E_y. Although more general assumptions are possible, it is convenient to suppose that this partial preordering is transitive and, therefore, can be defined by specifying a positive cone $D \subset E_y$; the cone D is assumed to be closed, convex and proper, i.e. $D \neq E_y$. The partial preordering relation takes then the form

$$(1) \qquad y_1, y_2 \in E_y \quad , \quad y_1 \preccurlyeq y_2 \iff y_2 - y_1 \in D$$

with the corresponding equivalence relation

$$(2) \qquad y_1, y_2 \in E_y \quad , \quad y_1 \sim y_2 \Longleftrightarrow y_2 - y_1 \in D \cap -D$$

and the strong partial preordering relation

$$(3) \qquad y_1, y_2 \in E_y \quad , \quad y_1 \prec y_2 \Longleftrightarrow y_2 - y_1 \in \tilde{D} \stackrel{df}{=\!=} D \setminus (D \cap -D)$$

as well as the strict partial preordering relation

$$(4) \qquad y_1, y_2 \in E_y \quad , \quad y_1 \ll y_2 \Longleftrightarrow y_2 - y_1 \in \mathring{D}$$

where $\mathring{D}$ is the interior of the cone D. In some spaces, naturally defined positive cones might have empty interiors; however, we can define then the quasi-strict partial preordering through replacing $\mathring{D}$ in (4) by $\mathring{D}^q$, the quasi-interior of D

$$(5) \qquad \mathring{D}^q = \{y \in D : \langle y^*, y \rangle > 0 \ , \forall y^* \in \tilde{D}^* \stackrel{df}{=\!=} D^* \setminus (D^* \cap -D^*)\}$$

where

$$(6) \qquad D^* = \{y^* \in E_y^* : \langle y^*, y \rangle \geq 0, \forall y \in D\}$$

is the dual cone to D, E_y^* being the dual space to E_y and $\langle \cdot, \cdot \rangle$ denoting the duality relation between E_y^* and E_y (the general form of a linear continuous functional from $E_{y^*}^*$ over E_y).

For example, if $E_y = L^2([t_0;t_1], R^p)$, then a positive cone can be naturally defined by

$$(7) \quad D = L_+^2([t_0;t_1], R^p) =$$

$$\{y \in L^2([t_0;t_1], R^p) : y^i(t) \geq 0, \text{ a.e. for } t \in [t_0;t_1], \forall i = 1, \ldots, p\} \ .$$

The equivalence classes (2) are then composed of functions that are equal to each other almost everywhere on $[t_0;t_1]$, which coincides with classical definitions of equivalence classes in L^2. The strong partial preordering (3) relates functions which have components $y_1^i(t) \leq y_2^i(t)$ a.e. on $[t_0;t_1]$, $\forall i = 1, \ldots, p$, such that the inequality $y_1^i(t) < y_2^i(t)$ holds for at least one i and at least on a subset of $[t_0;t_1]$ of nonzero measure. Since the cone (7) has

empty interior, there are no $y_1, y_2 \in E_y$ that are strictly related. However, $D^* = D$ in this case (L^2 is a Hilbert space and its dual can be made identical with it). Moreover, D has a nonempty quasi-interior:

$$(8)\quad \mathring{D}^q = \{y \in L^2([t_0;t_1],R^p) : y^i(t) > 0 \text{ a.e. for } t \in [t_0;t_1], \forall i = 1,\ldots,p\}$$

and the quasi-strict partial preordering relates functions with components $y_1^i(t) < y_2^i(t)$ a.e. on $[t_0;t_1]$, $\forall i = 1,\ldots,p$. For other examples of positive cones see Wierzbicki and Kurcyusz, 1977.

The set of admissible controls V and the mappings X,Y define together the *set of attainable outputs*

$$(9)\quad Y_V = Y(X(V),V) \subset E_y \quad .$$

Usually, we cannot describe the full set Y_V analytically because the mappings X,Y are too complicated; however, it is assumed that we can generate elements of this set, at least numerically, by solving the dynamic model for a given $u \in V$. On the other hand, suppose we are interested only in D-*maximal elements* $\hat{y} \in \hat{Y}_V$

$$(10)\quad \hat{Y}_V = \{\hat{y} \in Y_V : Y_V \cap (\hat{y} + \tilde{D}) = \phi\}$$

which are natural generalizations of Pareto-maximal outputs for the case of trajectory optimization. If the cone $\mathring{D}$ is nonempty, it is sometimes convenient to consider also *weak* D-*maximal elements* $\hat{y} \in \hat{Y}_V^w$

$$(11)\quad \hat{Y}_V^w = \{\hat{y} \in Y_V : Y_V \cap (\hat{y} + \mathring{D}) = \phi\}$$

or *quasi-weak* D-*maximal elements* $\hat{y} \in \hat{Y}_V^{wq}$ obtained as in (11) while replacing $\mathring{D}$ by $\mathring{D}^q$. Clearly, $\hat{Y}_V \subset \hat{Y}_V^{wq} \subset \hat{Y}_V^w \subset Y_V$. Sometimes it is also convenient to consider a smaller set $\hat{Y}_V^\varepsilon \subset \hat{Y}_V$ of D_ε-*maximal* elements of Y_V:

$$(12)\quad \hat{Y}_V^\varepsilon = \{\hat{y} \in Y_V : Y_V \cap (\hat{y} + \tilde{D}_\varepsilon) = \phi\}$$

where D_ε is defined as a conical ε-neighborhood of D:

$$(13) \qquad D_\varepsilon = \{y \in E_y : \mathrm{dist}(y,D) < \varepsilon \|y\|\} \; ; \; \tilde{D}_\varepsilon = D_\varepsilon \setminus (D_\varepsilon \cap -D_\varepsilon) \quad .$$

Since dist(y,D) is a continuous functional of y, the cone D_ε is an open cone, that is, an open set augmented with the point 0 or the set $D_\varepsilon \cap -D_\varepsilon$. Thus, $\tilde{D}_\varepsilon$ is an open set, and D_ε-maximality is equivalent to weak D_ε-maximality.

For example, if $D = L^2_+([t_0;t_1],R^p)$ as in (7), then, using an argument via projections on cones in Hilbert spaces as in Wierzbicki and Kurcyusz (1977) it can be shown that:

$$(14) \quad D_\varepsilon = \{y \in L^2([t_0;t_1],R^p) : \|y_-\| < \varepsilon \|y\|\} \; ; \; y^i_-(t) = \min(0, y^i(t))$$

and D_ε has an interior: at any point $y \in D$ we can center a ball with radius $\delta < \varepsilon$, contained in D_ε.

A classical method of generating D-maximal elements of Y_V is that of maximizing a (quasi-) strictly positive linear functional $y^* \in \mathring{D}^{*q}$ over $y \in Y_V$:

$$(15) \qquad \hat{y} \in \mathrm{Arg} \max_{y \in Y_V} \langle y^*, y \rangle \quad , \qquad y^* \in \mathring{D}^{*q} \Rightarrow \hat{y} \in \hat{Y}_V \quad .$$

However, it is very difficult to express the experience and judgment of a user of the model in terms of a linear functional (called also weighting functional) $y^* \in \mathring{D}^{*q}$; in the case of dynamic trajectory optimization, it often becomes practically impossible. On the other hand, it is quite practical to express the experience and judgment in terms of a desirable output trajectory $\bar{y} \in E_y$, which should not be constrained to Y_V nor otherwise, called *reference trajectory* (also aspiration level trajectory, reference point).

Many authors -- see Wierzbicki (1979) for a review -- have considered the use of the norm $\|\bar{y} - y\|$ for generating D-maximal elements of Y_V. The most general results were obtained by Rolewicz (1975) for any Banach space E_y:

$$(16) \qquad \hat{y} \in \operatorname{Arg} \max_{y \in Y_V} \|\bar{y} - y\| \quad , \qquad \bar{y} \in Y_{VD} \Rightarrow \hat{y} \in \hat{Y}_V$$

if $\bar{y}$ is D-dominating Y_V:

$$(17) \qquad Y_{VD} = \{\bar{y} \in E_y : \bar{y} - y \in D \text{ for all } y \in Y_V\} = \{\bar{y} \in E_y : Y_V \subset \bar{y} - D\}$$

and if the following condition is satisfied:

$$(18) \qquad D \cap (y - D) \subset B(0, \|y\|) \cup \{y\} \quad , \qquad \forall y \in E_y$$

where $B(0,\rho)$ denotes the open ball in the space E_y with radius ρ and center at 0. If E_y is Hilbert, then the condition (18) is satisfied iff

$$(19) \qquad D \subseteq D^* \quad .$$

However, the conditions (18) or (19), limiting the choice of the norm and the positive cone, are not very restrictive for applications; really restrictive is the requirement that $\bar{y}$ should be D-dominating all attainable outputs. To overcome this limitation, the notion of an *achievement scalarizing functional* has been introduced -- see, e.g., Wierzbicki (1980). An achievement scalarizing functional is a nonlinear continuous functional $s: E_y \to R^1$, with argument $y - \bar{y}$, where $y \in Y_V$ is an attainable output trajectory and $\bar{y} \in E_y$ is an arbitrary (not constrained to Y_V nor to Y_{VD}) desirable reference trajectory. An achievement scalarizing functional should, moreover, satisfy two axiomatic requirements:

(i) it should be *(quasi-) strictly order preserving*

$$(20) \qquad y_2 - y_1 \in \overset{\circ}{D} \text{ (or } y_2 - y_1 \in \overset{\circ}{D}^q) \Rightarrow s(y_1 - \bar{y}) < s(y_2 - \bar{y})$$

or, if possible, *strongly order preserving*

$$(21) \qquad y_2 - y_1 \in \tilde{D} \Rightarrow s(y_1 - \bar{y}) < s(y_2 - \bar{y})$$

(ii) it should be *order representing*

$$(22) \qquad S_0 \stackrel{df}{=} \{y \in E_y : s(y - \bar{y}) \geq 0\} = \bar{y} + D \quad ;$$

$$s(y - \bar{y}) = 0 \quad \text{for all } y - \bar{y} \in D \setminus \mathring{D} \ (\text{or } y - \bar{y} \in D \setminus \mathring{D}^q)$$

or, at least, *order approximating* for some small $\varepsilon > 0$;

$$(23) \qquad \bar{y} + D \subset S_0 \stackrel{df}{=} \{y \in E_y : s(y - \bar{y}) \geq 0\} = \bar{y} + D_{\varepsilon 0} \subseteq \bar{y} + D_\varepsilon ; \ s(0) = 0$$

where the cone $D_{\varepsilon 0}$ is not necessarily of the form (13) and is a closed cone. However, in order to preserve similarity with $\tilde{D}_\varepsilon$, $\tilde{D}_{\varepsilon 0}$ is defined by $\tilde{S}_0 \stackrel{df}{=} \{y \in E_y : s(y - \bar{y}) > 0\} = \bar{q} + \tilde{D}_{\varepsilon 0}$. Therefore, $\tilde{D}_{\varepsilon 0}$ is an open set, and $D_{\varepsilon 0}$-maximality is equivalent to weak $D_{\varepsilon 0}$-maximality. The set $\hat{Y}_V^{\varepsilon 0} = \{\hat{y} \in Y_V : Y_V \cap (\hat{y} + \tilde{D}_{\varepsilon 0}) = \phi\}$ is understood in the above sense.

Thus, we can distinguish *strict achievement scalarizing functionals*, which satisfy the requirements (20) and (22), and *strong achievement scalarizing functionals*, which satisfy the requirements (21) and (23); the requirements (21) and (22) cannot be satisfied together. It is known that, if s is strongly order preserving, then, for any $\bar{y} \in E_y$:

$$(24) \qquad \hat{y} \in \operatorname{Arg} \max_{y \in Y_V} s(y - \bar{y}) \Rightarrow \hat{y} \in \hat{Y}_V$$

and if s is only (quasi-) strictly order preserving, then:

$$(25) \qquad \hat{y} \in \operatorname{Arg} \max_{y \in Y_V} s(y - \bar{y}) \Rightarrow \hat{y} \in \hat{Y}_V^w \ (\text{or } \hat{y} \in \hat{Y}_V^{wq}) \quad .$$

On the other hand, as shown in Wierzbicki (1980), if s is a strict achievement scalarizing functional, then

$$(26) \qquad \hat{y} \in \hat{Y}_V^w \ (\text{or } \hat{y} \in \hat{Y}^{wq}) \Rightarrow \hat{y} \in \operatorname{Arg} \max_{y \in Y_V} s(y - \hat{y}) \ , \ \max_{y \in Y_V} s(y - \hat{y}) = 0$$

and, if s is a strong achievment scalarizing functional, then

$$(27) \qquad \hat{y} \in \hat{Y}_V^{\varepsilon 0} \Rightarrow \hat{y} \in \operatorname{Arg} \max_{y \in Y_V} s(y - \hat{y}) \ , \ \max_{y \in Y_V} s(y - \hat{y}) = 0 \quad .$$

The conditions (26), (27) constitute not only necessary conditions for D-maximality even for nonconvex sets Y_V (corresponding to the separation of the sets Y_V and $\hat{y} + \mathring{D}^q$ or $\hat{y} + \tilde{D}_{\varepsilon 0}$ by the nonlinear functional s), but are also rather practical means for checking whether a given desirable $\bar{y}$ is attainable with surplus, attainable without surplus and D-maximal, or not attainable. In fact, for a strong achievement scalarizing functional s

$$(28) \qquad \bar{y} \in (Y_V - D_{\varepsilon 0}) \setminus \hat{Y}_V^{\varepsilon 0} \Rightarrow \max_{y \in Y_V} s(y - \bar{y}) > 0$$

$$\bar{y} \in \hat{Y}_V^{\varepsilon 0} \subset Y_V - D_{\varepsilon 0} \Rightarrow \max_{y \in Y_V} s(y - \bar{y}) = 0$$

$$\bar{y} \notin Y_V - D_{\varepsilon 0} \Rightarrow \max_{y \in Y_V} s(y - \bar{y}) < 0$$

where $(Y_V - D_{\varepsilon 0}) \setminus \hat{Y}_V^{\varepsilon 0}$ is the set of all output trajectories $D_{\varepsilon 0}$-dominated by an attainable trajectory, $Y_V - D_{\varepsilon 0} = \{\bar{y} \in E_y : \bar{y} = y - d, y \in Y_V, d \in D_{\varepsilon 0}\}$. The proof of relations (28) follows directly from the definition of $\tilde{D}_{\varepsilon 0}$ by $\tilde{S}_0 = \{y \in E_y : s(y - \bar{y}) > 0\} = \bar{q} + \tilde{D}_{\varepsilon 0}$. Similar conclusions hold for strict achievement scalarizing functionals.

Another important conclusion (see Wierzbicki 1980) from the conditions (26), (27) is *the controllability of modeling results by the user*: if, say, a strong achievement scalarizing functional is applied, then the user can obtain any $D_{\varepsilon 0}$-maximal output trajectory $\hat{y}$ as a result of maximization of $s(y - \bar{y})$ by suitably changing the reference trajectory $\bar{y}$, no matter what are other detailed properties of the functionals. Therefore, detailed properties of the functional s can be chosen in order to facilitate either computational optimization procedures, or the interaction between the user and the optimization model, or as a compromise between these two goals.

Various forms of achievement scalarizing functionals have been discussed in Wierzbicki (1980) in the case when $E_y = R^p$, together with some special forms when E_y is a Hilbert space. Here we consider in some more detail the construction of achievement scalarizing functionals in normed spaces.

A general construction of a strict achievement scalarizing functional in the case of $\mathring{D} \neq \phi$ can be obtained as follows. Suppose a value functional $v: D \to R^1$ is given (that is, any strictly order preserving, nonnegative functional v defined for $y \in D$ -- similarly as in Debreu (1959)) and is equal zero for all $y \in D \setminus \mathring{D}$. Then:

$$(29) \qquad s(y-\bar{y}) = \begin{cases} v(y-\bar{y}), & \text{if } y-\bar{y} \in D \\ -\rho \operatorname{dist}(y-\bar{y}, D), & \text{if } y-\bar{y} \notin D \quad ; \quad \rho > 0 \end{cases}$$

is a strict achievement scalarizing functional. It is clearly order representing. If $y-\bar{y} \in D$, it is strictly order preserving. If $y_2 - y_1 \in \mathring{D}$, $y_2 - \bar{y} \in D$ and $y_1 - \bar{y} \notin D$, then $s(y_2 - \bar{y}) - s(y_1 - \bar{y}) > 0$ by the definition (29). If $y_2 - y_1 \in \mathring{D}$, $y_2 - \bar{y} \notin D$ and, thus, $y_1 - \bar{y} \notin D$, then denote $y_2 - y_1 = \tilde{y} \in \mathring{D}$ and observe that

$$\operatorname{dist}(y_2-\bar{y}, D) = \min_{d \in D} \|y_1+\tilde{y}-\bar{y}-d\| = \min_{\tilde{d} \in D-\tilde{y}} \|y_1-\bar{y}-\tilde{d}\| = \operatorname{dist}(y_1-\bar{y}, D-\tilde{y}).$$

On the other hand, since $\tilde{y} \in \mathring{D}$ and D is a convex cone, hence $D \subset \mathring{D} - \tilde{y}$. Any interior point of $D - \tilde{y}$ has a larger distance from the exterior point $y_1 - \bar{y}$ than $\operatorname{dist}(y_1 - \bar{y}, D-\tilde{y})$; hence $\operatorname{dist}(y_2-\bar{y}, D) < \operatorname{dist}(y_1 - \bar{y}, D)$ and $s(y_2 - y) - s(y_1 - \bar{y}) > 0$ in all cases of $y_2 - y_1 \in \mathring{D}$, the functional (29) is strictly order preserving.

However, the functional (29) has several drawbacks. First, even if it would be possible to extend it for cases when $\mathring{D} = \phi$ and $\mathring{D}^q \neq \phi$, such an extension is not essential: in applications, weak or quasi-weak D-maximal elements of Y_V are not interesting, and much more important are D_ε-maximal elements. Moreover, the choice of a value functional with desired properties might be difficult in infinite-dimensional spaces, since the simplest value functional -- a positive linear functional -- cannot be continuously modified to zero for $y - \bar{y} \in D \setminus \mathring{D}$. Therefore, we shall relax the requirement of order representation to that of order approximation, while trying to obtain in return strong order preservation.

Choose any strongly positive linear functional $y^* \in \mathring{D}^{*q}$, of unit norm, $\|y^*\| = 1$. Then:

$$(30) \qquad s(y-\bar{y}) = \langle y^*, y-\bar{y}\rangle - \rho\, \mathrm{dist}(y-\bar{y}, D) \quad , \qquad \rho > 1$$

is a strong achievement scalarizing functional, with $\varepsilon > \frac{1}{\rho}$. In fact, $\langle y^*, y-\bar{y}\rangle$ is strongly order preserving, due to the definition $\overset{\circ}{D}{}^{*q} = \{y^* \in E_y^*: \langle y^*, y\rangle > 0 \ \forall y \in \tilde{D}\}$. The functional $-\mathrm{dist}(y-\bar{y}, D)$ is order preserving (neither strongly nor strictly), by an argument similar to the analysis of the functional (29). However, the sum of a strongly order preserving and an order preserving functional is, clearly, strongly order preserving. Moreover, by the definition of the norm in the dual space, $\langle y^*, y-\bar{y}\rangle \leq \|y-\bar{y}\|$ if $\|y^*\| = 1$. If, additionally, $y \in S_0 = \{y \in E_y: s(y-\bar{y}) \geq 0\}$, then $\rho\, \mathrm{dist}(y-\bar{y}, D) \leq \langle y^*, y-\bar{y}\rangle \leq \|y-\bar{y}\|$; hence $S_0 \subset \bar{y} + D_\varepsilon$ for $\varepsilon > \frac{1}{\rho}$. Clearly, $\bar{y} + D \subset S_0$ and $s(0) = 0$; thus the functional (30) is order approximating.

The functional (30) has also some drawbacks in applications. First, the choice of y^* is arbitrary; however, it does not much influence the applicability of the functional (30), particularly if $\rho >> 1$, since $\bar{y}$ is very often chosen as not attainable. Thus, any reasonable y^* -- for example, corresponding to equal weights for all components of output trajectories and all instants of time -- might be chosen; according to the controllability conclusion, this does not restrict the possibility of influencing the resulting D_ε-maximal output trajectories $\hat{y}$ by changing the reference trajectories $\bar{y}$. Second, the functional (30) is nondifferentiable. Although recent development of nondifferentiable optimization algorithms is remarkable, not all of these algorithms are directly applicable for dynamic optimization. Therefore, it might be useful to consider also achievement scalarizing functionals that are differentiable.

Observe that achievement scalarizing functionals are constructed by using a strictly or strongly order preserving functional of value functional type and supplementing it by a term expressing a distance from $y-\bar{y}$ to the cone D. While the first part can be chosen to be differentiable, it is the second part that introduces nondifferentiability. To facilitate computation and differentiation of functionals related to the distance, suppose E_y is a Hilbert space. Then, due to the Moreau theorem (1962; see Wierzbicki and Kurcyusz, 1977), the following holds:

$$\text{dist}(y-\bar{y},D) = \|(y-\bar{y})^{-D^*}\| = \|(\bar{y}-y)^{D^*}\| \tag{31}$$

when $(\cdot)^{-D^*}$ or $(\cdot)^{D^*}$ denotes the operation of projection on the cone $-D^*$ or D^*. Moreover, $\|(\bar{y}-y)^{D^*}\|^2$ is differentiable in y and its derivative is precisely $-(\bar{y}-y)^{D^*}$. Thus, if E_y is Hilbert, a differentiable modification of (30) is as follows:

$$s(y-\bar{y}) = \langle y^*, y-\bar{y}\rangle - \frac{1}{2}\rho\|(\bar{y}-y)^{D^*}\|^2 \quad , \qquad \rho > 0 \quad . \tag{32}$$

This functional is strongly order preserving, by the same argument as in the analysis of (30), and its maximal points are D-maximal for any $\rho > 0$. However, the functional (32) is not order approximating and, if $\bar{y} = \hat{y}$ is D_ε-maximal, then the maximal points of (32) will generally not coincide with $\hat{y}$ for any $\rho > 0$. On the other hand, if ρ is sufficiently large, the maximal points of (32) usually approximate quite closely the maximal points of (30), and the requirement of order approximation does not play a decisive role. Thus, the functional (32) for sufficiently large ρ might have useful applications.

If E_y is Hilbert, then there is also a technically differentiable form of a strong achievement scalarizing functional, satisfying both (21) and (23):

$$s(y-\bar{y}) = \frac{1}{2}\|y-\bar{y}\|^2 - \frac{1}{2}\rho\|(\bar{y}-y)^{D^*}\|^2 \quad ; \quad \rho > 1, \quad D \subseteq D^* \tag{33}$$

with $\varepsilon > \rho^{-\frac{1}{2}}$, see Wierzbicki (1977a). In (33), the role of a value functional for $y-\bar{y} \in D$ plays the (square) norm; hence the condition $D \subseteq D^*$, equivalent to the Rolewicz condition (18), is necessary for the strong order preservation property. If $y-\bar{y} \notin D$, the (square) norm is modified by the (square) distance term; if $\rho > 1$, this modification is sufficiently strong to imply strong order preservation. The property of order approximation results immediately from the form of (33).

Consider, however, a functional similar to (33):

$$s(y-\bar{y}) = \|y-\bar{y}\| - \rho\|(\bar{y}-y)^{D^*}\| \quad ; \quad \rho > 1, \quad D \subseteq D^* \quad . \tag{34}$$

It is also a strong achievement scalarizing functional. It is clearly order approximating with $\varepsilon > \rho^{-1}$. Moreover, due to the Moreau theorem, $s(y-\bar{y}) = (a^2(y) + b^2(y))^{\frac{1}{2}} - \rho b(y)$, where $a(y) = \|(y-\bar{y})^D\|$, $b(y) = \|(y-\bar{y})^{-D^*}\|$. The operation of projection on cones, $(\cdot)^D$ or $(\cdot)^{-D^*}$, has the property (see Wierzbicki and Kurcyusz, 1977) that $\|(y-\bar{y}+\tilde{y})^{-D^*}\| \leq \|(y-\bar{y})^{-D^*}\|$ for all $\tilde{y} \in D$ and $\|(y-\bar{y}+\tilde{y})^D\| \geq \|(y-\bar{y})^D\|$ for all $\tilde{y} \in D^*$, hence also for $\tilde{y} \in D$ if $D \subseteq D^*$. Thus, if $y_2 - y_1 \in D$, then $a(y_2) \geq a(y_1)$ and $b(y_2) \leq b(y_1)$. Since $a(y_2) = a(y_1)$ and $b(y_2) = b(y_1)$ imply together $y_2 = y_1$, hence, if $y_2 - y_1 \in \tilde{D}$, we can have either $a(y_2) > a(y_1)$ and $b(y_2) \leq b(y_1)$ or $a(y_2) \geq a(y_1)$ and $b(y_2) < b(y_1)$. Now, consider the function $\psi(a,b) = (a^2 + b^2)^{\frac{1}{2}} - \rho b$. This function is clearly strictly increasing in respect to a. Since $\frac{\partial \psi}{\partial b}(a,b) = b(a^2 + b^2)^{-\frac{1}{2}} - \rho < 0$ for $\rho > 1$, the function ψ is strictly decreasing in respect to b. Therefore, if $y_2 - y_1 \in \tilde{D}$, then $s(y_2 - \bar{y}_1) - s(y_1 - \bar{y}) > 0$, and the functional (34) is strongly order preserving.

On the other hand, after a suitable choice of (different) values of ρ in (33) and (34), the level set $S_0 = \{y \in E_y : s(y-\bar{y}) \geq 0\}$ can be made identical for these two functionals, and this level set has necessarily a corner point at $y = \bar{y}$. Thus, the differentiability of (33) has only technical character, and an essential nondifferentiability in terms of corner points of level sets is necessarily related to strong and strict achievement scalarizing functionals. Therefore, for computational purposes, it is useful to introduce another class of *approximate scalarizing functionals*. The approximate scalarizing functionals are supposed to have strong order preservation property (21), which implies that their maximal points are D-maximal. However, the requirement of order approximation (23) is further related by substituting D_ε, the conical ε-neighborhood of D, by another form of an ε-neighborhood:

$$D_{\varepsilon\gamma} = \{y \in E_y : \mathrm{dist}(y,D) < \varepsilon\gamma(\|y\|)\} \tag{35}$$

where $\gamma(\cdot)$ is any given strictly increasing function. For example, it is easy to check that (32) is an approximate scalarizing functional, with $\gamma(\|y\|) = \|y\|^{\frac{1}{2}}$. Approximate scalarizing functionals are not strictly applicable for checking D-maximality of a given

$\hat{y}$ via condition (27), nor attainability of a given $\bar{y}$ via conditions (28), since a maximum point of an approximate scalarizing functional might be different from a given D-maximal $\hat{y} = \bar{y}$. However, the set $D_{\varepsilon\gamma}$ approximates the cone D sufficiently closely for small ε, and the difference between its maximal point and a given D-maximal $\hat{y} = \bar{y}$ can be made very small. Thus, for practical purposes, approximate scalarizing functionals have all the advantages of strong scalarizing functionals.

To illustrate further the distinction between strong and approximate scalarizing functionals, consider still another variant of such functionals. Suppose we have, originally, a single-objective optimization problem with a performance functional:

$$y^0 = Y^0(X(u),u) \in R^1 \quad . \tag{36}$$

Suppose that, after maximizing this functional and observing, for example, that there are many controls u and states x that result in nearly the same value of y^0 (a frequent case of practical nonuniqueness of solutions), we decided to supplement this performance functional with other objectives, stated in terms of a desirable shape of output trajectories:

$$y^r = Y^r(X(u),u) \in E_y^r \tag{37}$$

where E_y^r is a normed space, with a positive cone D^r. After defining $y = (y^0, y^r)$, $E_y = R^1 \times E_y^r$ and $D = R_+^1 \times D^r$ we bring the problem back to the previous formulation, and any of the scalarizing functionals defined above can be used. However, this specific case suggests also a specific form of a strong scalarizing functional:

$$s(y-\bar{y}) = y^0 - \bar{y}^0 - \rho\,\mathrm{dist}(y^r - \bar{y}, D^r) \quad ; \quad \rho > 0 \quad . \tag{38}$$

It is easy to check that this functional is order approximating with $\varepsilon > 1/\rho$. Moreover, it is strongly order preserving in a modified sense, with $\tilde{D} = (R_+^1 \times D^r) \setminus (\{0\} \times (D^r \cap -D^r))$ replaced by

$\tilde{\tilde{D}} = (R^1_+ \setminus \{0\}) \times (D^r \setminus (D^r \cap -D^r)) = \tilde{R}^1_+ \times \tilde{D}^r$. This modified sense of strong partial preordering results in modified D-maximal points that might be weakly D^r-maximal, in the second component y^r, but are always strongly maximal in the first component y^0. In fact, if $y_2 - y_1 \in \tilde{\tilde{D}}$, then $y^0_2 > y^0_1$ and $y^r_2 - y^r_1 \in \tilde{D}^r$. Since the functional $-\text{dist}(y^r - \bar{y}^r, D^r)$ is (neither strictly nor strongly) order preserving, the first term in (38) guarantees that $s(y_2 - \bar{y}) > s(y_1 - \bar{y})$ for $y_2 - y_1 \in \tilde{\tilde{D}}$.

Suppose E^r_y is Hilbert and consider the following approximate scalarizing functional

$$(39) \qquad s(y-\bar{y}) = y^0 - \bar{y}^0 - \tfrac{1}{2}\rho \| (\bar{y}^r - y^r)_{D^{r*}} \|^2 \quad ; \quad \rho > 0$$

By a similar argument, this functional is strongly order preserving with $\tilde{\tilde{D}}$ replacing $\tilde{D}$. It is not order approximating, only γ-order approximating with $D_{\varepsilon\gamma}$ defined as in (35) and $\gamma(\|y\|) = \|y\|^{\frac{1}{2}}$.

Observe that the functionals (38), (39) correspond to one of the classical, widely used approaches to multiobjective optimization. In this approach, we choose one of the objectives -- say, y^0 -- to be maximized and represent other objectives -- say, y^r -- by parametrically changing constraints, $y^r - \bar{y}^r \in D^r$. The functionals (38), (39) represent, respectively, an exact and an exterior quadratic penalty functional for such a formulation. However, it is not widely known that, when using such penalty functionals, one does not have to increase ρ to infinity or otherwise iterate (e.g., introduce shifts) on penalty functionals. Since these functionals are (modified) strongly order preserving, each maximal point of them is (modified) D-maximal, no matter what $\rho > 0$ has been chosen and what are the actual violations $(\bar{y} - y^r)_{D^{r*}}$ of the constraints $y^r - \bar{y}^r \in D^r$, treated here as a type of soft constraints. This feature of the scalarizing functionals (38), (39) is particularly useful for dynamic optimization with trajectory constraints (taking a form, for example, of state constraints), since the iterations on penalty functions might be particularly cumbersome in such a case. While using functions (38), (39) for multiobjective trajectory optimization, it is

sufficient to choose a reasonable value of $\rho > 0$ and to maximize (38) or (39) once in order to obtain a (modified) D-maximal alternative solution corresponding to a desirable shape $\bar{y}^r$ of output trajectory y^r.

Via penalty functions, functionals (38), (39) -- and, in fact, all other achievement scalarizing functionals -- are related to two other basic notions in mathematical optimization and modelling: those of generalized Lagrangian functionals and of regularization of solutions of ill-posed problems.

3. RELATIONS TO GENERALIZED LAGRANGIAN FUNCTIONALS

Consider the classical form of a mathematical programming problem with generalized inequalities:

$$(40) \qquad \underset{u \in U_0}{\text{minimize}}\ f^0(u) \quad ; \quad U_0 = \{u \in E_u : g(u) \in -D \subset E_g\}$$

where $f^0 : E_u \to R'$, $g : E_u \to E_g$, D is a positive cone in E_g. Suppose E_x is a Banach space and E_g is a Hilbert space. Under various forms of regularity conditions -- see, e.g., Kurcyusz (1974) -- the necessary conditions for $\hat{u}$ being an optimal solution to this problem can be expressed via the well-known normal Lagrangian functional

$$(41) \qquad L(\eta, u) = f^0(u) + \langle \eta, g(u) \rangle$$

and take the known form

$$(42) \qquad L_u(\hat{\eta}, \hat{u}) = f_u^0(\hat{u}) + g_u^*(\hat{u})\hat{\eta} = 0$$

where $g_u^*(\hat{u})$ is the adjoint operator to $g_u(\hat{u})$, and

$$(43) \qquad g(\hat{u}) \in -D \quad ; \quad \langle \hat{\eta}, g(\hat{u}) \rangle = 0 \quad ; \quad \hat{\eta} \in D^*$$

where $\hat{\eta} \in E_g^*$ is a normal Lagrange multiplier related to the solution $\hat{u}$. The triple condition (43) might be referred to as Kuhn-Tucher complementarity triple, widely known. However, it is not widely known that complementarity triple (43) is, in fact,

equivalent to a single nonlinear equation for $\hat{\eta}$ (although this result has been, in fact, used in R^n by Rockafellar (1974), in a Hilbert space by Wierzbicki and Kurcyusz (1977) and independently proven in R^n by Mangasarian (1976)).

To show this in the case when E_g is a Hilbert space, we use the Moreau (1962) theorem: for any closed convex cone $D \subset E_g$ and any $p \in E_g$, $p_1 = (p)^{-D}$ and $p_2 = (p)^{D^*}$ are the projections of p on the cones -D, D*, respectively, if and only if

$$(44) \qquad p_1 + p_2 = p, \quad p_1 \in -D, \quad \langle p_2, p_1 \rangle = 0, \quad p_2 \in D^* \quad .$$

Thus, denote $g(\hat{u}) + \hat{\eta} = p$; it is easy to check then that (43) holds if and only if

$$(45) \qquad (g(\hat{u}) + \hat{\eta})^{D^*} = \hat{\eta}$$

or, equivalently, iff $(g(\hat{u}) + \hat{\eta})^{-D} = g(\hat{u})$ (one of these equations suffices and the other is redundant because of the definition $g(\hat{u}) + \hat{\eta} = p$.)

This basic fact has various consequences. For example, the sensitivity analysis of solutions of (40) might be based on appropriate implicit function theorems instead of analyzing the sensitivity of a system of inequalities, which is now the typical approach to this question -- see, e.g., Robinson (1976). Another important conclusion from equation (45) is that there are modified Lagrangian functionals that should possess an *unconstrained saddle point* in η, u at $\hat{\eta}, \hat{u}$. In fact, these are augmented Lagrangian functionals as introduced by Hesteness (1969) for problems with equality constraints in R^n, by Rockafellar (1974) for problems with inequality constraints in R^n, by Wierzbicki and Kurcyusz (1977) for problems with inequality constraints in a Hilbert space, and studied by many other authors. For problem (40), the augmented Lagrangian functional takes the form

$$(46) \qquad \Lambda(\eta, u, \rho) = f^0(u) + \tfrac{1}{2}\rho \| (g(u) + \frac{\eta}{\rho})^{D^*} \|^2 - \tfrac{1}{2}\rho \| \frac{\eta}{\rho} \|^2 \quad , \ \rho > 0$$

and the first-order necessary conditions (42), (43) $\Longrightarrow$ (45) take the form

$$(47) \qquad \Lambda_u(\eta,u,\rho) = f_u^0(\hat{u}) + g_u^{*}(\hat{u})(\rho g(\hat{u}) + \hat{\eta})^{D^*} = 0$$

$$(48) \qquad \Lambda_\eta(\hat{\eta},\hat{u},\rho) = \frac{1}{\rho}\,((\rho,g(\hat{u}) + \hat{\eta})^{D^*} - \hat{\eta}) = 0$$

Other necessary and sufficient conditions for optimality of $\hat{u}$ in terms of saddle-points of (46) are given in Wierzbicki and Kurcyusz (1977).

Consider now the following specification of problem (40), taking into account (36), (37)

$$(49) \qquad f^0(u) = -Y^0(X(u),u) \quad ; \quad g(u) = \bar{y}^r - Y^r(X(u),u) \in -D^r$$

where u might be additionally constrained explicitly by $u \in V$. Consider the augmented Lagrangian functional (46) with $\eta = 0$:

$$(50) \qquad \Lambda(0,u,\rho) = -Y^0(X(u),u) + \tfrac{1}{2}\rho\| (\bar{y}^r - Y^r(X(u),u))^{D^{r*}} \|^2 =$$

$$= -s(Y(X(u),u)-\bar{y}) - \bar{y}^0$$

with $s(y-\bar{y})$ defined as in (39). The order-preservation properties of the approximate scalarizing functional (39) can be now interpreted as follows. Even if we fix $\eta = 0$ and admit violations of the constraint $\bar{y}^r - Y^r(X(u),u) \in -D^r$, and even under additional constraints $u \in V$, any minimal point of the augmented Lagrangian functional (50) is a D-maximal point of the set $Y_V = Y(X(V),V) = Y^0(X(V),V) \times Y^r(X(V),V)$ in the sense of the strong partial preordering induced by the cone $\tilde{D} = \tilde{R}_+^1 \times \tilde{D}^r$. Moreover, since:

$$(51) \qquad \Lambda(\eta,u,\rho) = -s(Y(X(u),u) - \bar{y} - \frac{\eta}{\rho}) - \bar{y}^0 - \tfrac{1}{2}\rho\| \frac{\eta}{\rho} \|^2$$

and the above conclusion holds independently of $\bar{y}$, hence it also holds for any fixed η. Thus, the conclusion can be considered as another generalization of Everett's theorem (196) and the reference trajectory $\bar{y}$ is, in a sense, related to the generalized Lagrange multiplier η.

However, the last analogy should not be taken too mechanistically. For example, the properties (28) of a strong scalarizing functional can be rewritten as

$$(52) \qquad \min_{y \in Y_V} \max_{u \in V} s(Y(X(u),u) - \bar{y}) = 0$$

and the min-max points $(\hat{\bar{y}},\hat{u})$ correspond to D-maximal points of the set $Y_V = Y(X(V),V)$. On the other hand, (52) is not a saddle-point property, since $s(y-\bar{y})$ is not convex in $\bar{y}$, and it is easy to show examples such that $\max_{u \in V} \min_{\bar{y} \in Y_V} s(Y(X(u),u)-\bar{y}) < 0$. In order to obtain saddle-point properties, convexifying terms in η would have to be added to $s(Y(X(u),u) - \bar{y})$, as it was done in (51).

4. MULTIOBJECTIVE TRAJECTORY OPTIMIZATION AS SEMIREGULARIZATION OF MODEL SOLUTIONS

The monography of Tikkonov and Arsenin (1977) summarizes an extensive research on one of the basic problems of mathematical modeling -- that of regularization of solutions of ill-posed problems. Many results of this research relate to the usefulness of using distance functionals when solving problems with non-unique solutions or quasi-solutions (generalized solutions). The nonuniqueness of solutions of a mathematical model implies usually that the solutions would change discontinuously with small changes of parameters of the model. For example, if a dynamic linear programming model has practically nonunique solutions, that is, if there is one optimal basic solution but many other basic solutions result in almost the same value of the objective function, then a small change of parameters of the model results in large changes of the solution -- see Avenhaus (1980). The regularization of solutions of such a type of models consists then in choosing from experience a *reference solution* and considering the solution of the model that is closest to the reference solution in a chosen sense of distance; as proven by Tikkanov and Arsenin, this results not only in the selection of a solution, but also in continuous dependence of the selected solution on parameters of the model.

The regularization method can be illustrated as follows. Suppose a mathematical programming problem consists in minimizing the functional

$$f^0(u) = -Y^0(X(u),u) \tag{53}$$

for $u \in V$. Suppose the solutions of this problem are (possibly only practically) nonunique. Let a reference trajectory $\bar{y}^r$ be given in a normed space E_y^r of the outputs of the model, $y^r = Y^r(X(u),u)$. By a *normal solution* of the problem of minimizing $f^0(u)$ for $u \in V$ we define such a solution of this problem that minimizes, additionally, $\|\bar{y}^r - Y^r(X(u),u)\|$. This normalization is, clearly, relative to the output space E_y^r. However, it is easy to see that if, say, $f^0(u)$ and V are convex, Y and X are linear, and the unit ball in E_y^r is strongly convex, then the normal solution is unique relative to the output space -- that is, it determines uniquely the output trajectory y^r. Moreover, this output trajectory depends continuously on the reference trajectory $\bar{y}^r$. A stable computational method of determining the normal solution approximately consists in minimizing the functional:

$$\Phi(\bar{y}^r,u,\rho) = -Y^0(X(u),u) + \tfrac{1}{2}\rho\|\bar{y}^r - Y^r(X(u),u)\|^2 \tag{54}$$

for $\rho \to 0$. Again, under appropriate assumptions, it can be shown that output trajectories corresponding to minimal points of (54) converge to the output trajectory corresponding to the normal solution as $\rho \to 0$.

However, observe that (54) can be obtained from (50) if E_y^r is Hilbert and $D^r = \{0\}$, $D^{r*} = E_y^r$. Thus, the multiobjective trajectory optimization is strongly related to model regularization. Actually, the former can be considered as a generalization of the latter. In fact, define *semi-normal* solutions of the problem of minimizing $f^0(u)$ for $u \in V$ as such that minimize, additionally, $\mathrm{dist}(Y^r(X(u),u), \bar{y}^r + D^r)$, where D^r is a positive cone in the space of output trajectories E_y^r. Now, even if f(u) and V were convex and Y and X linear, the output trajectory y^r corresponding to a semi-normal solution need not be unique -- since there might be many points in a convex set that are equidistant

to a convex cone. However, the semi-normal solutions have good practical interpretation; the corresponding output trajectories are either close to or better than the desired reference trajectory $\bar{y}^r$, depending on its attainability. Moreover, when minimizing the functional (50), instead of (54), we obtain D-maximal points of the set $Y_V = Y(X(V),V) = Y^0(X(V),V) \times Y^r(X(V),V)$ for each $\rho > 0$. The same applies, clearly, to the functional (54), if we assume $D^r = \{0\}$, which gives another interpretation of regularization techniques. Thus, multiobjective trajectory optimization is a type of *model semiregularization technique:* for the selection of a solution of the model, a reference output trajectory is used together with a notion of a partial preordering of the output space.

5. COMPUTATIONAL ISSUES AND APPLICATIONS: A DIFFERENTIABLE TIME-CONTINUOUS CASE

If an achievement scalarizing functional is differentiable, then any method of dynamic optimization can be applied as a tool for obtaining an attainable, D-maximal trajectory $\hat{y}$ in response to a desirable trajectory $\bar{y}$. An efficient class of dynamic optimization techniques applicable in this case are gradient trajectory techniques, or *control space gradient techniques,* based on a reduction of the gradient of the minimized functional to control space. A general method for such a gradient reduction, independent on the particular type of the state equation, is described, for example, in Wierzbicki (1977b). Here we present only the simplest and well-known case of gradient reduction for problems with ordinary differential state equations.

As an example, consider the approximate scalarizing functional (39) and suppose y^0 is described by

$$y^0 = \int_{t_0}^{t_1} F^0(x(t),u(t),t)\,dt + F^1(x(t_1)) \quad . \tag{55}$$

Moreover, assume the mapping X be given by solutions of the state equation

$$x = X(u) \longrightarrow \dot{x}(t) = F(x(t),u(t),t) \quad ; \quad x(t_0) = x_0 \in R^n \tag{56}$$

and the mapping Y^r -- by the output equation

$$(57) \qquad y^r = Y^r(X(u),u) \longrightarrow y^r(t) = G^r(x(t),u(t),t) \in R^p \quad .$$

Take $E_y^r = L^2([t_0;t_1],R^p)$ and $D^r = L_+^2([t_0;t_1],R^p)$; then

$$(58) \quad s(Y(X(u),u)-\bar{y}) = \int_{t_0}^{t_1} G^0(x(t),u(t),\bar{y}^r(t),t)dt + F^1(x(t_1)) - \bar{y}^0$$

where

$$(59) \quad G^0(x(t),u(t),\bar{y}^r(t),t) = F^0(x(t),u(t),t) - \tfrac{1}{2}\rho \sum_{i=1}^{p} (\bar{y}^{ri}(t) - G^{ri}(x(t),u(t),t))_+^2$$

and $(\tilde{y}^i)_+ = \max(0,\tilde{y}^i)$ for $\tilde{y}^i \in R^1$. By choosing $D^r = L_+^2([t_0;t_1],R^p)$ we assumed that all outputs improve as the corresponding values $y^{ri}(t)$ increase for (almost) all $t \in [t_0;t_1]$. Now, a reference output trajectory $\bar{y}^r(t) = (\bar{y}^{r1}(t),\ldots,\bar{y}^{ri}(t),\ldots,\bar{y}^{rp}(t))$ for $t \in [t_0;t_1]$ is assumed to be given by the model user. In fact, if p is not too large -- say, 3 or 4 -- the user can easily draw the number p of curves representing output trajectories desired by him. Moreover, experiments show that he is also able to evaluate easily the corresponding responses of the optimization model, $\hat{y}^0$ and $\hat{y}^r(t) = (\hat{y}^{r1}(t),\ldots,\hat{y}^{ri}(t),\ldots,\hat{y}^{rp}(t))$ for $t \in [t_0;t_1]$ and, if he does not like them, to change the reference trajectory in order to obtain new responses. Observe that the reference value $\bar{y}^0$ plays, in this case, a technical role and can be omitted. Thus, an interactive multiobjective dynamic optimization procedure can be organized, provided we could supply an efficient technique of maximizing the functional (57) subject to the state equation (56) and, possibly, other constraints. To simplify the presentation, suppose other constraints are already expressed as penalty terms in the functions F^0 or F^1.

Denote $S(u) = s(Y(X(u),u) - \bar{y})$. Then $S_u(u)$, the gradient of the functional (58) reduced to the control space, can be computed in the following way. The Hamiltonian function for the problem of maximizing (58) subject to (56) has the form

$$(60) \qquad H(\Psi(t),x(t),u(t),\bar{y}^r(t),t) = G^0(x(t),u(t),\bar{y}^r(t),t) + \Psi(t)F(x(t),u(t),t)$$

where $\Psi(t)F(x(t),u(t),t)$ is a short denotation for scalar product in R^n and $\Psi(t)$ is the costate (the adjoint variable for the state). To compute $S_u(u)(t)$ for $t \in [t_0;t_1]$, given $u(t)$ for $t \in [t_0;t_1]$, we first determine $x(t) = X(u)(t)$ by solving (56), written equivalently as

$$(61) \qquad \dot{x}(t) = H_\Psi(\Psi(t),x(t),u(t),\bar{y}^r(t),t) \quad ; \quad x(t_0) = x_0 \quad .$$

Then the costate $\Psi(t)$ is determined for $t \in [t_0;t_1]$ by solving, in the reverse direction of time, the adjoint equation

$$(62) \qquad \dot{\Psi}(t) = -H_x(\Psi(t),x(t),u(t),\bar{y}^r(t),t) \quad ; \quad \Psi(t_1) = F_x^1(x(t_1))$$

and the reduced gradient in the control space is determined by

$$(63) \qquad S_u(u)(t) = H_u(\Psi(t),x(t),u(t),\bar{y}^r(t),t) \quad .$$

Typical conjugate directions algorithms of nonlinear programming can be adapted for making use of this reduced gradient. However, Fortuna (1974) has shown that, for dynamic optimization, conjugate directions perform much better if a modified reduced gradient is being used:

$$(64) \qquad \tilde{S}_u(u)(t) = -H_{uu}^{-1}(\Psi(t),x(t),u(t),\bar{y}^r(t),t)H_u(\Psi(t),x(t),u(t),\bar{y}^r(t),t) \quad .$$

This modification removes possible ill-conditioning of the algebraic part of the Hessian operator $S_{uu}(u)$, leaving only possible ill-conditioning of the compact part of this operator -- and the compact part has, in the limit, negligible influence on the convergence of conjugate direction algorithms in a Hilbert space. This abstract reasoning has been also confirmed by extensive computational tests.

Now, each continuous-time dynamic optimization problem, when solved on a digital computer, is ultimately discretized over time. While a discussion of results of recent world-wide extensive research on approximations of time-continuous optimization problems is beyond the scope of the paper, it is worthwhile to note some comments on this issue.

A conscientious approach to discretization of a time-continuous problem should start with the question whether time-continuity is really an essential aspect of the analyzed model. In many cases, time-continuity is assumed only for analytical convenience, and the actual model can be better built, parameter-fitted and validated in its time-discrete version. In such cases of a priori discretization, it is certainly better to abandon time-continuity at the very beginning and to develop the time-discrete versions, say, of the equations (55) ... (64). Some qualitative properties and conclusions from the time-continuous analysis might be still applied to time-discrete models; for example, the Fortuna modification of the reduced gradient, although motivated strictly for the time-continuous case only, gives good results also in the time-discrete case.

In rather special cases, time-continuity is essential. These cases are really hard, and great care should be devoted to the analysis of those qualitative properties of the optimization problem that make time-continuity essential (such as boundary-layer effects, appearance of relaxed controls, etc.). These qualitative properties should be taken into account when looking for alternative formulations of the problem, for an appropriate space of control functions, when choosing finite-dimensional bases for a sequence of subspaces approximating the control space, when determining what is the reduced gradient expressed in terms of a finite-dimensional basis. A naive discretization of equations (61) ... (64) can lead to serious errors, when, say, a naively discretized gradient equation (63) produces numbers that are in no correspondence to the gradient that would be consistent with a chosen discretization of the control space.

We close this section with a simple example, when the continuity of time is important only because it facilitates almost fully the analytical solution. Although it does not illustrate computational issues, the example illuminates some other important aspects of applications of multiobjective trajectory optimization.

Consider a simple model of relations between inflation and unemployment, as analyzed by Snower and Wierzbicki (1980) when comparing various economic policies. The inflation rate, $x(t)$, is influenced by monetary policies, that influence also the unemployment, $u(t)$. An adaptive price expectation mechanism and a linearized Phillips curve result in the following equation:

$$\dot{x}(t) = rd(b - u(t)) \quad ; \quad x(0) = x_0 \tag{65}$$

where unemployment $u(t)$ is taken as a dummy control variable, b is a parameter of the linearized Phillips curve, rd is a composite coefficient. The social welfare function related to inflation and unemployment is assumed in the form:

$$U(x(t),u(t)) = 1 - \tfrac{1}{2}x^2(t) - \frac{q}{2}\,u^2(t) \tag{66}$$

where q is the weight attached to unemployment as compared to inflation. The intertemporal social welfare functional is assumed in the form

$$W(x,u) = \int_0^\infty e^{-rt}\, U(x(t),u(t))dt \quad . \tag{67}$$

The problem of maximizing (67) subject to (65) can be easily solved analytically to obtain:

$$\hat{u}(t) = \frac{\alpha_0}{d}\,(x_0 - \frac{bq}{d})e^{-r\alpha_0 t} + b \tag{68}$$

$$\hat{x}(t) = (x_0 - \frac{bq}{d})e^{-r\alpha_0 t} + \frac{bq}{d} \tag{69}$$

where

$$\alpha_0 = \tfrac{1}{2}((1 + 4\,\frac{d^2}{d})^{\frac{1}{2}} - 1) \quad . \tag{70}$$

However, if the initial inflation rate x_0 is high, the 'optimal' unemployment $\hat{u}(t)$ that results from this model for small t might be considered socially undesirable, too high. We could change the model by adding simply a constraint $u(t) \leq \bar{u}$. In this simple case, the constraining value $\bar{u}$ must be greater than b; otherwise, equation (65) would result in uncontrolled, increasing inflation. However, in more complicated models, it might be difficult to judge whether a control constraint is not too stringent. Therefore, it is reasonable to treat $\bar{u}$ as a desirable bound for trajectory rather than as a fixed constraint, and to formulate a multiobjective trajectory optimization problem: maximize the social welfare functional while, at the same time, trying to keep the unemployment smaller than $\bar{u}$.

Observe that, in this formulation, one of the outputs y^r of the model is just the input control u. However, such situations are quite frequent, when some important control variables appear directly as output variables in multiobjective trajectory optimization. Moreover, the unemployment u(t) is here only a dummy control variable; actually, the model should be controlled by a monetary policy that, after a transformation that was not included is the model for simplicity results in the unemployment u(t).

Suppose we apply the approximate achievement scalarizing functional (39) for this multiobjective trajectory optimization problem and choose the norm $\|u\|^2 = \int_0^\infty e^{-rt} u^2(t)dt$ for the control space. Then:

$$(71) \qquad s(Y(x,u) - \bar{y}) = \int_0^\infty e^{-rt}(U(x(t),u(t)) - \tfrac{1}{2}c(u(t) - \bar{u})_t^2)dt - \bar{y}^0$$

Suppose $u(t) > \bar{u}$ for $t\in[0;t_1)$, $u(t_1) = \bar{u}$, $u(t) < \bar{u}$ for $t\in(t_1; +\infty)$. Then (71) transforms to

$$(72)\qquad s(Y(x,u)-\bar{y}) = \int_0^{t_1} e^{-rt}(U(x(t)) - \tfrac{1}{2}\rho(u(t)-\bar{u})^2)dt +$$

$$+ e^{-rt_1}\hat{W}(x(t_1)) - \bar{y}^0$$

where

$$(73)\quad \hat{W}(x(t_1)) = \frac{1}{r}\left(1 - \frac{1}{2(1+\alpha_0)}x^2(t_1) - \frac{bd}{(1+\alpha_0)^2}x(t_1) - \right.$$

$$\left. - \frac{(2+\alpha_0)\alpha_0 qb^2}{2(1+\alpha_0)^2}\right)$$

is the minimal value of (67) depending on the initial state. The problem of minimizing (72) subject to (65) can be solved almost fully analytically to obtain:

$$(74)\quad \hat{u}(t) = \begin{cases} -\frac{rd}{q+\rho}\left(Ae^{-r\alpha_1 t} + Be^{-r\alpha_2 t}\right)+b, & t\in[0;t_1) \\ (\bar{u}-b)\,e^{-r\alpha_0(t-t_1)}+b, & t\in[t_1;+\infty) \end{cases}$$

$$(75)\quad \hat{x}(t) = \begin{cases} r\alpha_2 A e^{-r\alpha_1 t} + r\alpha_1 B e^{-r\alpha_2 t} - \frac{\rho}{d}(\bar{u}-b) + \frac{bq}{d}, & t\in[0;t_1) \\ \frac{d}{\alpha_0}(\bar{u}-b)e^{-r\alpha_0(t-t_1)} + \frac{bq}{d}, & t\in[t_1;+\infty) \end{cases}$$

where

$$(76)\qquad \alpha_1 = \tfrac{1}{2}\left(\left(1+4\frac{d^2}{q+\rho}\right)^{\frac{1}{2}} - 1\right);\ \alpha_2 = -\tfrac{1}{2}\left(\left(1+4\frac{d^2}{q+\rho}\right)^{\frac{1}{2}} + 1\right)$$

while the constants A,B and the time instant t_1 result from three conditions: the continuity of $\hat{u}(t)$ (implied by continuity of adjoint variable) and of $\hat{x}(t)$ at t_1 and the initial state $x_0=\hat{x}(0)$.

For example, the former two conditions determine A,B as functions of t_1:

$$(77) \qquad A = A(t_1) = e^{r\alpha_1 t_1} \cdot \frac{\bar{u}-b}{rd} \cdot \frac{q\alpha_0 - (q+\rho)\alpha_2}{\alpha_2 - \alpha_1} \quad ;$$

$$B = B(t_1) = e^{r\alpha_2 t_1} \cdot \frac{\bar{u}-b}{rd} \cdot \frac{q\alpha_0 - (q+\rho)\alpha_1}{\alpha_1 - \alpha_2}$$

while the latter condition results in the following equation for t_1 that does not admit analytical solutions (must be solved numerically)

$$(78) \qquad r\alpha_2 A(t_1) + r\alpha_1 B(t_1) - \frac{\rho}{d}(\bar{u}-b) + \frac{bq}{d} = x_0 \quad .$$

Nevertheless, (74),... (78) admit on easy interpretation of the influence of ρ and $\bar{u}$ on $\hat{u}(t)$ and $\hat{x}(t)$. The single-criterium solutions (68),(69) are compared with an example of solutions (74),(75) in Fig.1.

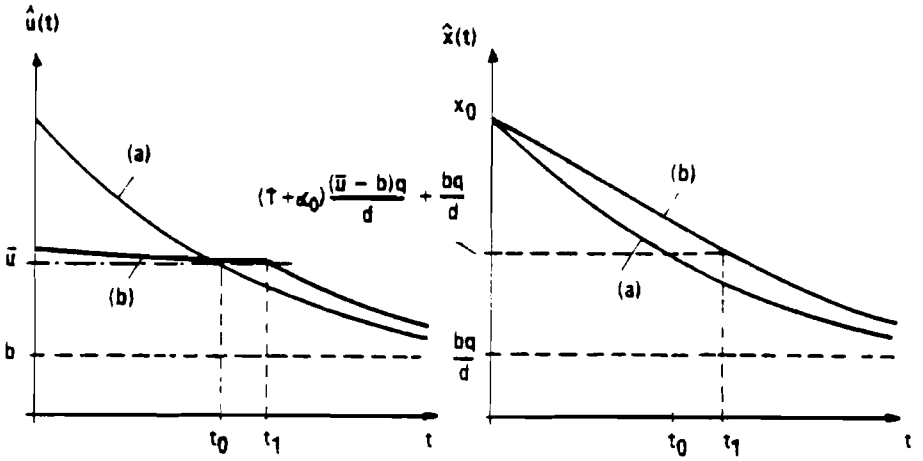


Figure 1. Examples of single-criterium 'optimal' solutions for unemployment $\hat{u}(t)$ and inflation $\hat{x}(t)$--case (a)-- compared with multicriteria D-maximal trajectories of these variables responding to a judgementably set reasonable level $\bar{u}$ of unemployment--case (b).

Observe that, if ρ is sufficiently large, the multicriteria D-maximal trajectory $\hat{u}$ has values $\hat{u}(t)$ only slightly greater than $\bar{u}$, and that the time t_1, at which $\hat{u}(t_1)=\bar{u}$ is also only slightly greater than the corresponding time t_0 for single-criterium case (the last observation follows from the fact that $\int_0^{\infty}(\hat{u}(t)-b)dt = \frac{1}{rd}(x_0 - \frac{bq}{d})$ for both cases). Thus, when applying multicriteria optimization, we can significantly reduce maximal unemployment while spreading the effects of this reduction over time. Clearly, in this simple example we could obtain similar results just by using an explicit constraint $u(t) \leq \bar{u}$. However, when using hard constraints, we must be careful not to specify $\bar{u} < b$; otherwise we would obtain $x(t) \to \infty$ as $t \to \infty$. When maximizing (71)--which is equivalent to a soft constraint on $u(t)$--we can assume $\bar{u} < b$ and still obtain well-defined results.

Observe also that one could interpret the achievement scalarizing functional (71) as just another form of welfare functional. This interpretation is correct; however, the modified welfare functional depends explicitly on judgementally set desirable bound $\bar{u}$ for unemployment, and in this aspect it differs basically from traditional welfare functionals. Moreover, it possesses the strong order preservation property. Thus, if $\hat{u}$ and $\hat{x}$ correspond to the maximum of this functional, then we cannot decrease the inflation $\hat{u}(t)$ at some t without increasing it at some other t or without decreasing the welfare functional $W(\hat{x},\hat{u})$.

6. COMPUTATIONAL ISSUES AND APPLICATIONS: A TIME-DISCRETE DYNAMIC LINEAR PROGRAMMING CASE

Many problems-- especially in economics (see, e.g. Kallio, Propoi, Seppälä 1980)--are formulated in terms of time-discrete dynamic programming models of the general form: maximize

$$(79) \qquad y^0 = Y^0(x,u) = \sum_{k=0}^{K=1} (c_k^* u_k + d_k^* x_k) + d_K^* x_K$$

subject to state equation constraints:

$$x_{k+1} = A_k x_k + B_k u_k \quad ; \quad x_0 \text{ - given} \tag{80}$$

and to additional constraints

$$(x_k, u_k) \in V_k \tag{81}$$

where V_k is a convex polyhedral set (described by linear inequalities), $u_k \in R^m$, $c_k^* \in R^{m*}$, $x_k \in R^n$, $d_k^* \in R^{n*}$, $A_k \in R^n x R^n$, $B_k \in R^n x R^m$. The trajectories x and u are, in this case, finite-dimensional, $u = \{u_0, \ldots u_k, \ldots u_{K-1}\} \in R^{mK}$, $x = \{x_0, \ldots x_k, \ldots x_K\} \in R^{n(K+1)}$, but can choose various norms in these trajectory spaces.

Various approaches have been devised to numerically solve this problem while taking advantage of its special structure (see, e.g. Kallio and Orchard-Hays 1980). For example, one of the efficient approaches is to solve this problem as a large scale static linear programming problem with the number of variables $(m+n)K$ (excluding x_0, which is a given parameter) and generating an initial feasible basic solution by choosing admissible u and solving state equation (80) for x.

It often happens that the solutions of this problem are practically non-unique (many admissible solutions correspond to almost maximal values of y^0) and that we are interested, in fact, not only in y^0 but also in some output trajectories $y^r = \{y_0^r, \ldots y_k^r, \ldots y_{K-1}^r\} \in R^{pK}$ of the model (80)

$$y_k^r = C_k^r u_k + D_k^r x_k \in R^p \tag{82}$$

where $C_k^r \in R^p \times R^m$, $D_k^r \in R^p \times R^n$. Suppose all output trajectories have to be maximized, thus the positive cone $D^r = R_+^{pK}$, $D = R_+^1 \times R_+^{pK}$.

A particularly convenient form of achievement scalarizing function for this class of problems has been developed by Wierzbicki (1978) and practically applied and further modified by Kallio, Lewandowski and Orchard-Hays (1980). The function corresponds to the choice of a maximum norm in the space

$E_y = R' \times E_y^r$ and has the form

$$(83)\quad s(y-\bar{y}) = \rho \min(y^0-\bar{y}^0,\ \min_{k,i}(y_k^{ri}-\bar{y}_k^{ri})) + y^{0*}(y^0-\bar{y}^0) + \sum_{k,i} y_k^{ri*}(y_k^{ri}-\bar{y}_k^{ri})$$

or, if we introduce the surplus variable $w = y-\bar{y} \in R^{pK+1}$, $w = \{w^j\}_{j=0}^{j=pK}$

$$(84)\quad s(w) = \rho \min_j w^j + y^* w$$

where $\rho > 0$ and y^* is a strictly positive linear function of unit norm in E_y^*. Because we have chosen maximum norm in E_y, E_y^* has the sum of absolute values norm, and y^* is simply a vector of positive weighting coefficients summing up to one, $y^* \in R^{(pK+1)*}$, $y^* = \{y^{j*}\}_{j=0}^{j=pK}$, $\sum_{j=0}^{pK} y^{j*} = 1$, $y^{j*} > 0$. Now, $\min_j w^j$ is strictly order preserving while y^*w is strongly order preserving, thus $s(w)$ is strongly order preserving. Moreover, if $D = R_+^{pK+1}$, then $D \subset S_0 = \{w \in R^{pK+1} : s(w) \geq 0\} = D_{\varepsilon 0} \subset D_\varepsilon$, where D_ε has the form (13) with $\varepsilon > \frac{1}{\rho}$, since $s(w) \geq 0$ and $\|y^*\| = 1$ imply together $\rho \operatorname{dist}(w,D) = -\rho \min_j w^j \leq y^* w \leq \|w\|$. Thus, $s(w)$ is order-approximating and a strong achievement scalarizing function.

The problem of maximizing $s(w)$, however, can be written equivalently as another large scale linear programming problem, by introducing $2(pK+1)$ or even only $(pK+1)$ additional linear constraints and $pK+1$ or even only 1 additional variables to the original problem. The modified problem is: maximize

$$(85)\quad \tilde{s}(w,v) = y^* w + \rho v$$

with $v \in R^1$, subject to:

$$(86)\quad v \leq w^j,\ j = 0,\ldots pK$$

$$(87) \qquad w^0 = \sum_{k=0}^{K-1} (c_k^* u_k + d_k^* x_k) + d_K^* x_K - \bar{y}^0$$

$$(88) \qquad \{w^j\}_{j=1}^{pK} = w^r = \{C_k^r u_k + D_k^r x_k - \bar{y}_k^r\}_{k=0}^{K-1}$$

and subject to (80),(81). Clearly we can set (87),(88) into (86), (85), thereby diminishing the number of additional constraints to (pK+1) and the number of additional variables to 1 (the variable v). An efficient algorithm for solving such problems has been developed by Orchard-Hays (see Kallio, Lewandowski, Orchard-Hays 1980).

According to the general theory from section 2, the choice of y^* and ρ does not affect principally the user of the model, who can obtain any desired D-maximal outputs of the model by changing the reference trajectory output $\bar{y}$. However, it might affect the easiness of interaction between the user and the model. This issue has been investigated in Kallio, Lewandowski and Orchard-Hays (1980) where $y^{j*} = \frac{1}{pK+1}$ and $\rho \geq 20$ resulted in good responses of the model. The particular model investigated was a Finnish forestry and forest industrial sector development model with maximized outputs representing the trajectory of the profit of the wood processing industries over time and the trajectory of income of the forestry from selling the wood to the industry over time (10 periods have been considered for each trajectory, hence the total number of objectives was 20; no intertemporal objective was included). Further improvements of the procedure have been also investigated, related to accumulating information about user's preferences revealed by the consecutive choice of reference trajectories $\bar{y}$ after a D-maximal trajectory $\hat{y}$ has been already proposed by the model. However, the main conclusions were the pragmatical and operational usefullness of the procedure; an example of trajectories $\bar{y}$ and $\hat{y}$ obtained in this model is shown in Fig.2.

It should be noted, finally, that achievement scalarizing function (83) is quite similar to functions used in goal programming techniques--see Charnes and Cooper (1961), Dyer (1972), Igmizio (1978), Kornbluth (1973). However, the use of function

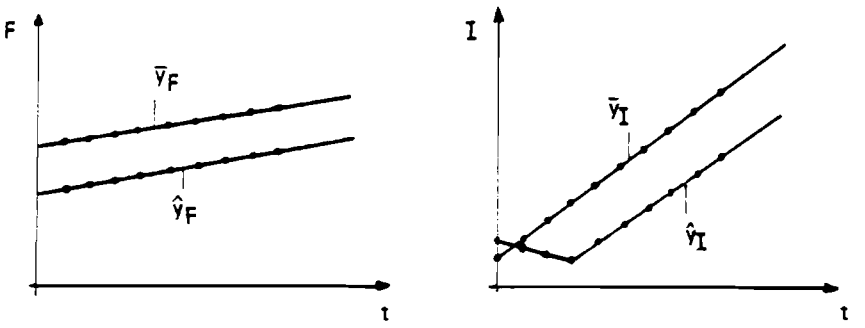


Figure 2. Forestry income trajectory (F) and forest industry profit trajectory (I) obtained in a multiobjective dynamic linear programming model: $\bar{y}$ - desired reference trajectories, $\hat{y}$ - corresponding D-maximal model outputs.

(83) is not related to some of the deficiencies known in applications of goal programming.

7. CONCLUSIONS

In many cases it is desirable and, as shown in this paper, both theoretically and practically possible to use multi-criteria trajectory optimization approaches to various dynamic system models. The approach is based on reference trajectories, when the user of the model specifies what are desirable output reference trajectories of a model and indicates what outputs would be even better than desirable ones, while the model responds with output trajectories that are not only attainable and nondominated in the sense of partial ordering in the output space as indicated by the user, but also correspond to the specified reference trajectories. On one hand, this approach is related to many interesting theoretical questions about the properties of achievement scalarizing functionals in normal spaces, their relations to augmented Lagrangian functionals, to regularization of solutions of ill-posed models; these questions have been investigated, to some extent, in the paper. On the

other hand, this approach is also eminently pragmatical; the author hopes that the examples presented show the reasonability and pragmatical values of using the seemingly abstract and untractable notions of infinite-dimensional or high-dimensional multicriteria trajectory optimization.

REFERENCES

Avenhaus, R., R. Beedgen, and L. Schrattenholzer. Handling Uncertainties in Linear Programming Models. WP-80- (in preparation). Laxenburg, Austria: International Institute for Applied Systems Analysis.

Bell, D.E., R.L. Keeney, and H. Raiffa. 1977. Conflicting Objectives in Decisions. New York: Wiley/IIASA International Series on Applied Systems Analysis.

Charnes, A., and W.W. Cooper. 1961. Management Models and the Industrial Applications of Linear Programming, Vols. 1 and 2. New York: Wiley.

Cohn. J.L., and D.H. Marks. 1975. A review and evaluation of multiobjective programming techniques. Water Resources Research 11:208-220.

Da Cunha, N.O., and E. Polak. 1967. Constrained minimization under vector-valued criteria in finite dimensional space. Journal of Mathematical Analysis and Application 19: 103-124.

Debreu, G. 1959. Theory of Value. New York: Wiley.

Dyer, J.S. 1972. Interactive goal programming. Management Science 19(1):62-70.

Everett, H. 1963. Generalized Lagrange multiplier method for solving problems of optimum allocation of resources. Operations Research 11:399-417.

Fortuna, Z. 1974. Properties of Conjugate Direction Method in Hilbert Spaces. Ph. D. thesis, Institute of Automatic Control, Technical University of Warsaw, Warsaw (in Polish).

Hesteness, M.R. 1969. Multiplier and gradient methods. JOTA 4:303-320.

Ignizio, J.P. 1978. Goal programming: a tool for multiobjective analysis. Journal of Operational Research 29(II):1109-1119.

Intriligator, M.D. 1971. Mathematical Optimization and Economic Theory. Englewood Cliffs, N.J.: Prentice-Hall.

Kallio, M., A. Lewandowski, and W. Orchard-Hays. 1980. An Implementation of the Reference Point Approach for Multi-objective Optimization. WP-80-35. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Kallio, M., and W. Orchard-Hays. 1979. Experiments with the Reduced Gradient Method for Linear Programming. WP-79-084. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Kallio, M., A. Propoi and R. Seppälä. 1980. A Model for the Forest Sector. WP-80-34. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Kalman, R., P.L. Falb, and M.A. Arbib. 1969. Topics on Mathematical System Theory. McGraw-Hill, New York.

Keeney, R.L., and H. Raiffa. 1976. Decisions with Multiple Objectives: Preferences and Value Trade-Offs. New York: Wiley.

Kornbluth, J.H.S. 1973. A survey of goal programming. Omega 1:193-205.

Kurcyusz, St. 1974. Necessary Conditions of Optimality for Problems with Constraints in a Function Space. Ph.D. thesis, Institute of Automatic Control, Technical University of Warsaw, Warsaw (in Polish).

Leitmann, G., and W. Schmittendorf. 1973. Some sufficient conditions for Pareto-optimal control Journal of Dynamical Systems, Measurement and Control 95(3).

Mangasarian, O.L. 1976. Equivalence of the complementarity problem to a system of nonlinear equations. SIAM J. Appl. Math. 31:89-92.

March, J.G., and H.A. Simon. 1958. Organizations. New York: Wiley.

Moreau, J.J. 1962. Decomposition orthogonale d'un espace hilbertien selon deux cones mutuellement polaires. C.R. Acad. Sci. Paris 225:238-240.

Robinson, S.M. 1976. Stability theory for systems of inequalities. SIAM J. Numeric. Anal. 23:497-513.

Rockefellar, R.T. 1974. Augmented Lagrangian multipliers functions and duality in nonconvex programming. SIAM J. Control 12:268-285.

Rolewicz, S. 1975. On Norm Scalarization in Infinite-Dimensional Banach Spaces. Working paper, Institute of Mathematics, Polish Academy of Sciences, Warsaw.

Sakluvadze, M.E. 1971. Optimization of vector functionals. Part I: Programming of optimal trajectories. Avtomatika i Telemekhanika 8:5-15. Part II: The analytic constructions of optimal controls. Ibidem 9:5-15 (in Russian).

Sakluvadze, M.E. 1974. On the existence of solutions in problems of optimization under vector-valued criteria. Journal of Optimization Theory and Applications 13(2).

Snower, D., and A.P. Wierzbicki. 1980. Notes on Optimal Policy Choice When Controlling Inflation and Unemployment. WP-80- (in preparation). Laxenburg, Austria: International Institute for Applied Systems Analysis.

Tikhonov, A.N., and W.J. Arsenin. 1975. Solutions of Ill-Posed Problems. Moscow: Nauka (in Russian); translated in 1977 by Fritz John and edited by V.H. Winston & Sons, Washington, D.C.

Wierzbicki, A.P. 1975. Penalty Methods in Solving Optimization Problems with Vector Performance Criteria. Proceedings of VI-th IFAC World Congress, Cambridge/Boston.

Wierzbicki, A.P. 1977a. Basic properties of scalarizing functionals for multiobjective optimization. Mathematische Operations-Forschung und Statistik, Ser. Optimization 8(1):55-60.

Wierzbicki, A.P. 1976. Models and Sensitivity of Control Systems. Warsaw, WNT (in Polish).

Wierzbicki, A.P., and St. Kurcyusz. 1977. Projection on a cone, penalty functionals and duality theory for problems with inequality constraints in Hilbert Space. SIAM J. Control and Optimization 15:25-26.

Wierzbicki, A.P. 1978. On the Use of Penalty Functions in Multiobjective Optimization. Proceedings of the International Symposium on Operations Research, Mannheim.

Wierzbicki, A.P. 1979. A Methodological Guide to Multiobjective Optimization. Proceedings of the 9-th IFIP Conference on Optimization Techniques, Warsaw. Also WP-79-122. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Wierzbicki, A.P. 1980. A Mathematical Basis for Satisficing Decision Making. WP-80-90. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Yu, P.L., and G. Leitmann. Compromise solutions, domination structures and Sakluvadze's solution. Journal of Optimization Theory and Applications 13:362-378.

NONCONVEX DUALITY PROPERTIES FOR MULTIOBJECTIVE OPTIMIZATION PROBLEMS

F. di Guglielmo

Centre Universitaire d'Avignon, France

We discuss in the following some non convex duality properties for multiobjective optimization problems. By using a characterization of Pareto optima by means of generalized Tchebycheff norms due to BOWMAN (1976) we reduce the multiobjective problem to a family of semi discrete minimax problems and we associate with them their duals. We then give estimates of the corresponding duality gaps in terms of the lack of γ-convexity of the objective functions. Finally we give a general characterization of Pareto optima in the non convex case and we apply it to the approximate determination of these optima by using some ideas due to DUTTA-VIDYASAGAR (1977) and HUARD (1967).

1/ Notations and definitions

Let X be a subset of $\mathbb{R}^n$ and $(f_j)_{j\in N}$ a finite set of real valued functions defined on X. There does not generally exist any $x^* \in X$ such that $f_j(x^*) = \inf_{x\in X} f_j(x)$ for all $j \in N$, but there may exist weakly minimal points $\bar{x} \in X$ such that there does not exist any $x \in X$ satisfying :

$$f_j(x) < f_j(\bar{x}) \quad \text{for all } j \in N.$$

Such points are called weak Pareto minima (or weakly efficient points).

It is well known that in the case of convex data the problem of determining the set of Pareto minima of the $(f_j)_{j\in N}$ in the set X is equivalent to solving the family of single objective optimization problems :

$$(Q_\lambda) \qquad \inf_{x\in X} \sum_{j\in N} \lambda_j f_j(x)$$

for all systems of non negative multipliers $(\lambda_j)_{j\in N}$ and that one can associate with each Pareto minimum $\bar{x} \in X$ a set of multipliers $\bar{\lambda}_j$ such that $\bar{x}$ is a solution of $(Q_{\bar{\lambda}})$.

On the other hand it has been shown by BOWMAN in the non convex case that, under appropriate assumptions, each Pareto minimum of the $(f_j)_{j\in N}$ in the set X can be considered as a solution of a problem of minimum distance :

$$(P_\beta) \qquad \inf_{x\in X} \; ||F(x) - \omega||_\beta$$

where F(x) denotes the vector of $\mathbb{R}^N$ with components $(f_j(x))_{j\in N}$, $||y||_\beta = \max_{i\in N} \beta_i |y_i|$ denotes a generalized Tchebycheff norm with weights $\beta_i > 0$ and ω is some vector of $\mathbb{R}^N$ such that $F(x) > \omega$ for all $x \in X$.

We shall study the relationship between problems (P_β) and (Q_λ) and we will see that (Q_λ) is connected with the following dual problem of (P_β) :

$$\sup_{\substack{\eta \geq 0 \\ \langle \eta, \mathbb{1} \rangle = 1}} \; \inf_{x\in X} \; \langle \eta, \beta \left[F(x) - \omega \right] \rangle$$

2/ The BOWMAN characterization of Pareto optima

Let us assume that the functions $(f_j)_{j\in N}$ are defined on a compact subset X of $\mathbb{R}^n$ and lower semi continuous. We assume moreover that there exists a point $\omega \in \mathbb{R}^N$ such that $f_j(x) > \omega_j$ for all $j \in N$ and $x \in X$. This means that all points $F(x) = (f_j(x))_{j\in N}$, $x \in X$ are contained in the interior of the non-negative orthant of $\mathbb{R}^N$ with origin in ω.

Then we have the following result of BOWMAN (1976)

Theorem $\bar{x} \in X$ is a weak Pareto minimum for the $(f_j)_{j\in N}$ in the set X if and only if there exists a parameter $\beta > 0$ such that :

$$||F(\bar{x}) - \omega||_\beta = \inf_{x\in X} ||F(x) - \omega||_\beta$$

It is to be noticed that problem (P_β) can be written as a semi discrete minimax problem of the form :

$$\min_{x\in X} \max_{j\in N} \varphi_j(x)$$

or explicitly :

$$(P'_\beta) \qquad \begin{cases} \text{Minimize } Z \\ \text{subject to :} \\ Z \geq \varphi_j(x) \quad \text{for all } j \in N, \quad x \in X, \quad Z \in \mathbb{R} \end{cases}$$

with $\varphi_j(.) = \beta_j \left[f_j(.) - \omega_j \right]$ for all $j \in N$.

3/ Some inequalities for semi discrete minimax problems.

Let us consider a family of real valued functions $(\varphi_j)_{j\in N}$ defined on a set X and let us introduce the vector valued function $\phi(x) = (\varphi_j(x))_{j\in N}$. We associate with ϕ the following quantities :

$$\alpha = \inf_{x\in X} \sup_{j\in N} \varphi_j(x)$$

$$\beta_o = \sup_{\eta\in \mathcal{M}^N} \inf_{x\in X} \langle \eta, \phi(x)\rangle$$

where $\mathcal{M}^N = \{\eta \in \mathbb{R}_+^N \mid \sum_{i\in N} \eta_i = 1\}$ denotes the unit simplex of $\mathbb{R}^N$.

We now define the lack of γ-convexity of the functions $(\varphi_j)_{j\in N}$ on the set X. We denote by $\mathcal{M}(X)$ the set of all discrete probability measures on X of the form $m = \sum_{i\in I} \alpha_i \delta(x^i)$ where I is a finite set of indices, the x^i are points of X and $\delta(x^i)$ denotes the Dirac measure at the point x^i.

If γ is any mapping from $\mathcal{M}(X)$ into X, then the lack of γ-convexity of a function f defined on the set X is given by

$$\rho_{\gamma,X}(f) = \sup_{m\in \mathcal{M}(X)} \left[f(\gamma m) - f^{\Delta}(m) \right]$$

where :

$$f^{\Delta}(m) = \sum_{i\in I} \alpha_i f(x^i).$$

We will say that f is γ-convex in the set X if and only if $\rho_{\gamma,X}(f) = o$. If X is convex and if γ is defined by $\gamma(\sum_{i\in I} \alpha_i \delta(x^i)) = \sum_{i\in I} \alpha_i x^i$ then $\rho_{\gamma,X}(f) = o$ means that f is convex in the usual sense in X.

Now we can prove the following :

Proposition 1 Let X be an arbitrary subset of $\mathbb{R}^n$ and $(\varphi_j)_{j\in N}$ a finite set of functions bounded on X . If γ is any mapping from $\mathcal{M}(X)$ into X the following estimate holds :

$$(1) \qquad o \leq \alpha - \beta_o \leq \sup_{j\in N} \rho_{\gamma,X}(\varphi_j).$$

We notice first that for arbitrary $\eta\in \mathcal{M}^N$ and $x \in X$ we have :

$$\langle \eta, \phi(x)\rangle = \sum_{j\in N} \eta_j \varphi_j(x) \leq (\sup_{j\in N} \varphi_j(x)) \sum_{j\in N} \eta_j = \sup_{j\in N} \varphi_j(x)$$

and this implies :

$$(2) \qquad \beta_o = \operatorname{Sup}_{\eta \in \mathcal{M}^N} \operatorname{Inf}_{x \in X} < \eta, \phi(x) > \leq \operatorname{Inf}_{x \in X} \operatorname{Sup}_{j \in N} \varphi_j(x) = \alpha.$$

Let us denote by $\mathbb{1}$ the vector of $\mathbb{R}^N$ with all components equal to one. We show that $\beta_o \mathbb{1} \in \overline{Co}\ (\phi(X) + \mathbb{R}^N_+)$(*). Let us suppose that this is not true. Then from the separation theorem there would exist $\epsilon > o$ and $\mu \in \mathcal{M}^N$ such that :

$$\beta_o = <\mu, \beta_o \mathbb{1}> \leq <\mu, \phi(x)> + <\mu, p> - \epsilon \text{ for all } x \in X,\ p \in \mathbb{R}^N_+ .$$

For fixed $x \in X$ this inequality implies that the linear form $p \mapsto <\mu, p>$ is bounded from below on $\mathbb{R}^N_+$ and hence that $\mu \geq o$. By taking $p = o$ we deduce that :

$$\beta_o \leq \operatorname{Inf}_{x \in X} <\mu, \phi(x)> - \epsilon \leq \operatorname{Sup}_{\mu \in \mathcal{M}^N} \operatorname{Inf}_{x \in X} <\mu, \phi(x)> - \epsilon = \beta_o - \epsilon$$

which yields a contradiction.
Hence $\beta_o \mathbb{1} \in \overline{Co}(\phi(X) + \mathbb{R}^N_+)$ and this implies that there exists coefficients $\alpha_i \geq o$ and points $x^i \in X$ such that :

$$(3) \qquad \beta_o \geq \sum_{i \in I} \alpha_i \varphi_j(x^i) \qquad \text{for all } j \in N.$$

If we set $m = \sum_{i \in I} \alpha_i \delta(x^i) \in \mathcal{M}(X)$ then the definition of the lack of γ-convexity of φ_j implies that for all $j \in N$ we have :

$$\beta_o \geq \varphi_j^{\Delta}(m) \geq \varphi_j(\gamma m) - \rho_{\gamma, X}(\varphi_j).$$

Since $\gamma m \in X$ these inequalities imply that :

$$\beta_o + \operatorname{Sup}_{j \in N} \rho_{\gamma, X}(\varphi_j) \geq \operatorname{Inf}_{x \in X} \operatorname{Sup}_{j \in N} \varphi_j(x) = \alpha.$$

It is possible to derive a sharper estimate for the difference $\alpha - \beta_o$ when $X = X_1 \times X_2 \times \ldots \times X_T$ and the functions φ_j have the following form :

$$(4) \qquad \varphi_j(x) = \frac{1}{T} \sum_{t=1}^{T} \varphi_{j,t}(x_t)$$

with $\varphi_{j,t}$ defined on X_t for $t = 1, 2, \ldots T$.

We shall use the Shapley-Folkman theorem which we first recall (see for instance

(*) We denote by Co(A) (resp. $\overline{Co}(A)$) the convex (resp. closed convex) hull of the set A.

AUBIN-EKELAND (1976)) :

Theorem (Shapley-Folkman)

Consider a finite family $(K_t)_{t=1,2,\dots T}$ of subsets of $\mathbb{R}^m$. For every $\xi \in Co(\sum_{\tau=1}^{T} K_t)$ there exists a subset of at most m indices $s(\xi) \subset \{1,2,\dots T\}$ such that :

$$\xi \in \sum_{t \notin s(\xi)} K_t + \sum_{t \in s(\xi)} co(K_t) \quad . \tag{5}$$

We can now prove the following result :

Proposition 2 Assume that the functions φ_j have the form (*) where the $\varphi_{j,t}$ are bounded functions defined on X_t for $j \in N$ and $t = 1,2,\dots T$. If for $t = 1,2,\dots T$ γ_t denotes an arbitrary mapping from $\mathcal{M}(X_t)$ into X_t, then the following estimate holds :

$$0 \leq \alpha - \beta_0 \leq \frac{1}{T} \sup_{1 \leq t \leq T} \sup_{j \in N} \rho_{\gamma_t, X_t}(\varphi_{j,t}). \tag{6}$$

For all $j \in N$ we have, as in the foregoing proposition :

$$\beta_0 = \sum_{i \in I} \alpha_i \varphi_j(x^i) + p_j = \frac{1}{T} \sum_{i \in I} \alpha_i \sum_{\tau=1}^{T} \varphi_{j,t}(x_t^i) + p_j \quad \text{with } p_j \geq 0$$

or equivalently :

$$\beta_0 - p_j \in Co(\sum_{\tau=1}^{T} E_{j,t}) \tag{7}$$

where $E_{j,t} = \{\frac{1}{T} \varphi_{j,t}(x_t)\}_{x_t \in X_t} \subset \mathbb{R}$ for $t = 1,2,\dots T$, $j \in N$.

By the Shapley-Folkman theorem there exists a set of indices containing at most one subscript τ_j such that :

$$\beta_0 - p_j \in \sum_{t \neq \tau_j} E_{j,t} + Co(E_{j,\tau_j}) \tag{8}$$

This means that there exists $x_t \in X_t$ for $t \neq \tau_j$ and $m_{\tau_j} \in \mathcal{M}(X_{\tau_j})$ with $m_{\tau_j} = \sum_{i \in I} \alpha_i \delta(x^i_{\tau_j})$ such that :

$$\beta_0 \geq \sum_{t \neq \tau_j} \frac{1}{T} \varphi_{j,t}(x_t) + \frac{1}{T} \sum_{i \in I} \alpha_i \varphi_{j,\tau_j}(x^i_{\tau_j}) = \sum_{t \neq \tau_j} \frac{1}{T} \varphi_{j,t}(x_t) + \frac{1}{T} \varphi^{\Delta}_{j,\tau_j}(m_{\tau_j})$$

which implies :

$$(9)\quad \beta_o \geq \sum_{t\neq\tau_j} \frac{1}{T}\varphi_{j,t}(x_t) + \frac{1}{T}\varphi_{j,\tau_j}(\gamma_{\tau_j} m_{\tau_j}) - \frac{1}{T}\rho_{\gamma_{\tau_j},X_{\tau_j}}(\varphi_{j,\tau_j})$$

By setting $x_{\tau_j} = \gamma_{\tau_j} m_{\tau_j}$ we obtain for all $j \in N$:

$$\beta_o + \frac{1}{T}\rho_{\gamma_{\tau_j},X_{\tau_j}}(\varphi_{j,\tau_j}) \geq \frac{1}{T}\sum_{t=1}^{T}\varphi_{j,t}(x_t) = \varphi_j(x)$$

and finally :

$$\beta_o + \frac{1}{T}\sup_{1\leq t\leq T}\sup_{j\in N}\rho_{\gamma_t,X_t}(\varphi_{j,t}) \geq \inf_{x\in X}\sup_{j\in N}\varphi_j(x) = \alpha.$$

■

By taking $\varphi_j(.) = \beta_j\left[f_j(.) - \omega_j\right]$ for all $j \in N$ we will now deduce from Propositions 1 and 2 duality results for Problems(P_β) and (Q_λ).

4/ Estimates for the duality gap.

We have first the following

Theorem 1 Let X be a compact subset of $\mathbb{R}^N$ and $(f_j)_{j\in N}$ a finite set of lower semi continuous functions defined on X. Assume that $\bar{x} \in X$ is a Pareto minimum satisfying :

$$||F(\bar{x}) - \omega||_\beta = \inf_{x\in X} ||F(x) - \omega||_\beta$$

Then there exists $\bar{\eta} \in \mathbb{R}^N$ such that :

$$(10)\quad \text{i)}\ 0 \leq \inf_{x\in X}||F(x)-\omega||_\beta - \inf_{x\in X}\sum_{j\in N}\bar{\lambda}_j\left[f_j(x)-\omega_j\right] \leq \sup_{j\in N}\beta_j\rho_{\gamma,X}(f_j)$$

$$(11)\quad \text{ii)}\ \sum_{j\in N}\bar{\lambda}_j f_j(\bar{x}) \leq \inf_{x\in X}\sum_{j\in N}\bar{\lambda}_j f_j(x) + \sup_{j\in N}\beta_j\rho_{\gamma,X}(f_j)$$

where $\bar{\lambda}_j = \bar{\eta}_j\beta_j$.

If $X = X_1 \times X_2 \times \ldots \times X_T$ and $f_j(x) = \frac{1}{T}\sum_{t=1}^{T} f_{j,t}(x_t)$ then :

$$(12)\quad \text{i')}\ 0 \leq \inf_{x\in X}||F(x)-\omega||_\beta - \inf_{x\in X}\sum_{j\in N}\bar{\lambda}_j\left[f_j(x)-\omega_j\right] \leq \frac{1}{T}\sup_{1\leq t\leq T}\sup_{j\in N}\beta_j\rho_{\gamma_t,X_t}(f_{j,t})$$

(13) ii') $\sum_{j\in N} \bar{\lambda}_j f_j(\bar{x}) \leq \inf_{x\in X} \sum_{j\in N} \bar{\lambda}_j f_j(x) + \frac{1}{T} \sup_{1\leq t\leq T} \sup_{j\in N} \beta_j \rho_{\gamma_t, X_t}(f_{j,t})$.

By the uppersemicontinuity of the map $\eta \longrightarrow \inf_{x\in X} \langle \eta, \phi(x)\rangle$ and the compactness of the unit simplex $\mathcal{M}^N$ there exists $\bar{\eta} \in \mathcal{M}^N$ such that :

(14) $$\sup_{\eta\in\mathcal{M}^N} \inf_{x\in X} \langle \eta, \phi(x)\rangle = \inf_{x\in X} \langle \bar{\eta}, \phi(x)\rangle$$

Now we notice that :

$$\alpha = \inf_{x\in X} \sup_{j\in N} \psi_j(x) = \inf_{x\in X} ||F(x) - \omega||_\beta = ||F(\bar{x}) - \omega||_\beta$$

$$\beta_o = \sup_{\eta\in\mathcal{M}^N} \inf_{x\in X} \langle \eta, \phi(x)\rangle = \sup_{\eta\in\mathcal{M}^N} \inf_{x\in X} \langle \eta, \beta[F(x)-\omega]\rangle = \inf_{x\in X} \langle \bar{\eta}, \beta[F(x)-\omega]\rangle$$

from which it follows by Proposition 1 that :

(15)
$$\langle \bar{\eta}, \beta[F(\bar{x})-\omega]\rangle \leq \sup_{\eta\in\mathcal{M}^N} \langle \eta, \beta[F(\bar{x})-\omega]\rangle = ||F(\bar{x})-\omega||_\beta = \inf_{x\in X} ||F(x)-\omega||_\beta =$$
$$= \alpha \leq \beta_o + \sup_{j\in N} \beta_j \rho_{\gamma,X}(f_j) = \inf_{x\in X} \langle \bar{\eta}, \beta[F(x)-\omega]\rangle + \sup_{j\in N} \beta_j \rho_{\gamma,X}(f_j)$$

which in turn implies i) and ii) by setting $\bar{\lambda} = \bar{\eta}\beta$.

A similar argument together with Proposition 2 can be used to prove i') and ii').

If the functions $(f_j)_{j\in N}$ are γ-convex, then $\rho_{\gamma,X}(f_j) = 0$ and we deduce from Theorem 1 the following :

Corollary 1 If the functions f_j are γ-convex and lower semi-continuous, then for any Pareto minimum $\bar{x}$ which is a solution of (P_β), there exists $\bar{\lambda} \in \mathbb{R}^N_+$ such that :

(16) i) $$\inf_{x\in X} ||F(x) - \omega||_\beta = \inf_{x\in X} \sum_{j\in N} \bar{\lambda}_j [f_j(x) - \omega_j]$$

(17) ii) $$\sum_{j\in N} \bar{\lambda}_j f_j(\bar{x}) = \inf_{x\in X} \sum_{j\in N} \bar{\lambda}_j f_j(x).$$

From Proposition 1 and Corollary 1 it follows that the problems

$$(P_\beta) \quad \inf_{x \in X} \; ||F(x)-\omega||_\beta$$

$$(S_\beta) \quad \sup_{\eta \in \mathcal{H}_+^N} \; \inf_{x \in X} \; \langle \eta, \beta[F(x)-\omega] \rangle$$

can be considered as dual problems, the duality gap being bounded by :

$$\sup_{j \in N} \; \beta_j \; \rho_{\gamma, X}(f_j) \qquad \text{resp.} \quad \frac{1}{T} \sup_{1 \leq t \leq T} \; \sup_{j \in N} \; \rho_{\gamma_t, X_t}(f_{j,t}).$$

When T tends to infinity while the lack of γ_t-convexity $\rho_{\gamma_t, X_t}(f_{j,t})$ remains bounded by a fixed constant, then this duality gap tends to zero. The multipliers $\bar{\lambda}_j = \bar{\eta}_j \beta_j$ can be interpreted in the non convex case as approximate Pareto multipliers associated with $\bar{x}$.

We conclude by giving a generalized characterization of Pareto optima in the non convex case.

5/ A generalized Characterization of Pareto optima

Let us denote by P a closed convex cone of $\mathbb{R}^N$ with non empty interior. We will use the following notation :

$u \underset{P}{\geq} v$ (resp. $u \underset{\bar{P}}{\geq} v$) if and only if $u-v \in P$ (resp. $u-v \in$ Int P).

We will assume that there exists $\omega \in \mathbb{R}^N$ such that $F(x) \underset{P}{\geq} \omega$ for any $x \in X$; then by replacing F(x) by $F(x)-\omega$ we can take $\omega = o$. Any $\tilde{x} \in X$ such that there does not exist any $x \in X$ for which $F(x) \underset{P}{\leq} F(\tilde{x})$ will be called a weak Pareto minimum for the $(f_j)_{j \in N}$ in the set X with respect to P.

We now consider a set of functions $(\psi_\beta)_{\beta \in B}$ defined on $\mathbb{R}^N$ and satisfying :

A_1) for any $b \in$ Int P there exists $\beta \in B$ such that :
$a \notin b -$ Int P implies $\psi_\beta(a) \geq \psi_\beta(b)$

A_2) for any $a, b \in P$ $b-a \in$ Int P implies $\psi_\beta(a) < \psi_\beta(b)$
for all $\beta \in B$.

We have now the following result which states that each Pareto minimum is a solution of some optimization problem of the form :

$$\inf_{x \in X} \quad \psi_\beta(F(x)) \qquad \text{with} \quad \beta \in B.$$

Theorem 2 Assume that the functions $(\psi_\beta)_{\beta \in B}$ satisfy assumption A_1). Then for any Pareto minimum $\bar{x} \in X$ there exists $\beta \in B$ such that :

(18) $$\psi_\beta(F(\bar{x})) = \inf_{x \in X} \psi_\beta(F(x)).$$

Conversely if the $(\psi_\beta)_{\beta \in B}$ satisfy assumption A_2) and $\bar{x} \in X$ is an optimal solution of $\inf_{x \in X} \psi_\beta(F(x))$ with $\beta \in B$, then it is a Pareto minimum of the $(f_i)_{i \in N}$.

Since $\bar{x} \in X$ is a Pareto minimum for the $(f_i)_{i \in N}$ there does not exist any $x \in X$ such that $F(x) \leqq_P F(\bar{x})$; this means that for any $x \in X$:

(19) $F(x) \notin F(\bar{x}) - \text{Int } P.$

According to A_1) this implies that there exists $\beta \in B$ (depending on $\bar{x}$) such that :

$$\psi_\beta(F(x)) \geqslant \psi_\beta(F(\bar{x})) \qquad \text{for any } x \in X$$

or equivalently :

$$\psi_\beta(F(\bar{x})) = \inf_{x \in X} \psi_\beta(F(x)).$$

Conversely let us assume that $x^1 \in X$ is an optimal solution of (P_β) and is not a Pareto minimum. Then there exists $\bar{x} \in X$ such that :
$F(\bar{x}) \leqq_P F(x^1)$ or equivalently $F(x^1) - F(\bar{x}) \in \text{Int } P$.

Then according to A_2 this implies that :

$$\psi_\beta(F(\bar{x})) < \psi_\beta(F(x^1)).$$

But since x^1 is an optimal solution of (P_β) we have :

$$\psi_\beta(F(x^1)) \leqslant \psi_\beta(F(\bar{x}))$$

which contradicts the preceding inequality.

■

We now give an example of a family of functions which satisfy assymptions A_1) and A_2).

Let us assume that there exists a finite set of non zero vectors $(\sigma_j)_{j\in J}$ which span the positive polar cone of P :

$$(20) \quad P^+ = \{ p^* \in \mathbb{R}^N \mid \langle p^*, p\rangle \geq o \quad \forall p \in P \} \quad .$$

Then for any $p^* \in P^+$ we have $p^* = \sum_{j\in J} \alpha_j \sigma_j$ with $\alpha_j \geq o$ for all $j \in J$. Moreover the interior of P admits the following representation :

$$(21) \quad \text{Int } P = \{ p \in \mathbb{R}^N \mid \langle p^*, p\rangle > o \quad \forall p^* \in P^+ - \{o\} \} \quad .$$

Denote by $(q_j)_{j\in J}$ a set of real valued strictly increasing functions defined on $\mathbb{R}$ such that $q_j(t) > o$ for all $t > o$ and define :

$$(22) \quad \psi_\beta(y) = \max_{j\in J} \beta_j q_j(\langle \sigma_j, y\rangle) \quad .$$

Then the $(\psi_\beta)_{\beta>o}$ satisfy assumptions A_1) and A_2).

i) Let $b \in \text{Int } P$, then for any $j \in J$ $\langle \sigma_j, b\rangle > o$ and we can take :

$$(23) \quad \beta_j = \frac{1}{q_j(\langle \sigma_j, b\rangle)} \quad .$$

If $a \notin b - \text{Int } P$, then there exists $p^* \in P^+ - \{o\}$ such that $\langle p^*, b-a\rangle \leq o$ or equivalently by taking $p^* = \sum_{j\in J} \alpha_j \sigma_j$ with $\alpha_j \geq o$ for all $j \in J$:

$$(24) \quad \sum_{j\in J} \alpha_j \langle \sigma_j, b\rangle \leq \sum_{j\in J} \alpha_j \langle \sigma_j, a\rangle \; .$$

This shows that there exists a subscript $j_o \in J$ such that :

$$\langle \sigma_{j_o}, b\rangle \leq \langle \sigma_{j_o}, a\rangle$$

and hence :

$$q_{j_o}(\langle \sigma_{j_o}, b\rangle) \leq q_{j_o}(\langle \sigma_{j_o}, a\rangle).$$

This in turn implies that :

$$(25) \quad \psi_\beta(a) = \max_{j\in J} \beta_j q_j(\langle \sigma_j, a\rangle) \geq 1 = \max_{j\in J} \beta_j q_j(\langle \sigma_j, b\rangle) = \psi_\beta(b) \quad .$$

Consider now $a,b \in P$ with $b-a \in \text{Int } P$. Then we have for all $j \in J$ $\langle \sigma_j, b\rangle > \langle \sigma_j, a\rangle$ which implies :

$$q_j(\langle \sigma_j, b\rangle) > q_j(\langle \sigma_j, a\rangle)$$

from which it follows that :

$$\psi_\beta(b) = \max_{j\in J} \beta_j q_j(\langle \sigma_j, b\rangle) > \max_{j\in J} \beta_j q_j(\langle \sigma_j, a\rangle) = \psi_\beta(a)$$

for all $\beta \in B$.

6/ Algorithms for the approximate determination of Pareto optima in the non convex case

We will now apply to problem (P_β) some ideas due to DUTTA VIDYASAGAR (1977) and HUARD (1967) to obtain algorithms for the approximate determination of Pareto optima. We will consider the following problem :

$$(P)\quad \left\{ \begin{array}{l} \text{Find a Pareto optimum of the functions } (f_i)_{i\in N} \text{ on the set :} \\ X = \{x \in \Sigma \mid g_k(x) \leq o \quad \forall k \in K\} \end{array} \right.$$

where Σ denotes a closed subset of $\mathbb{R}^n$ and $(g_k)_{k\in K}$ denotes a family of lower semi continuous bounded functions defined on $\mathbb{R}^n$. We assume that the positive polar cone of P is spanned by the vectors $(\sigma_j)_{j\in J}$ and we will set :

$$u_j^\beta(x) = \beta_j \langle \sigma_j, F(x)\rangle .$$

By applying Theorem 2 with the function ψ_β defined by :

$$\psi_\beta(y) = \max_{j\in J} \beta_j \langle \sigma_j, y\rangle \qquad \text{for } y \in \mathbb{R}^N$$

we see that any solution of (P) can be obtained by solving the problem :

$$(P'_\beta)\quad \left\{ \begin{array}{l} \text{Minimize } z \\ \text{subject to :} \\ z \geq u_j^\beta(x) \qquad \text{for any } j \in J \\ g_k(x) \leq o \qquad \text{for any } k \in K \\ x \in \Sigma, \quad z \in \mathbb{R} \end{array} \right.$$

for some vector $\beta \in \mathbb{R}^J$, $\beta > o$.

We associate with problem (P'_β) the following function :

$$(26)\qquad P_\beta(x,z) = \sum_{j\in J} \left[u_j^\beta(x) - z \right]_+^2 + \sum_{k\in K} \left[g_k(x) \right]_+^2 , \text{ where } y_+ = \max(o,y)$$

and we will assume that the problem $\inf_{x \in E} P_\beta(x,z)$ has a solution for any z (which is for instance the case if the functions $(f_i)_{i \in N}$ and $(g_k)_{k \in K}$ are continuous and if $\lim_{\|x\| \to +\infty} P_\beta(x,z) = +\infty$ for any z or if E is compact).

We will also assume that there exists a solution $\bar{x} \in X$ of problem (P'_β) and that the corresponding optimal value is given by :

$$\phi_\beta(\bar{x}) = \inf_{x \in X} \max_{j \in J} u_j^\beta(x).$$

We will consider the following algorithm :

Algorithm I

1) Choose $z_0 \leq \phi_\beta(\bar{x})$
2) Determine $x_\nu \in E$ minimizing $x \to P_\beta(x, z_\nu)$ on E.
3) Define $z_{\nu+1} = z_\nu + \left(\frac{1}{m} P_\beta(x_\nu, z_\nu)\right)^{1/2}$ where $m = |J|$.
4) Augment ν by 1 and return to 2.

We have the following convergence result :

Theorem 3 — The sequence z_ν converges to a limit z and is bounded from above by $\phi_\beta(\bar{x})$. If the functions $(f_i)_{i \in N}$ and $(g_k)_{k \in K}$ are lower semi continuous, then any cluster point x^* of the sequence (x_ν) is an optimal solution of (P'_β).

Let us assume that $z_\nu \leq \phi_\beta(\bar{x})$. Since $\bar{x}$ satisfies $g_k(\bar{x}) \leq 0$ for any $k \in K$ we have :

$$P_\beta(x_\nu, z_\nu) \leq P_\beta(\bar{x}, z_\nu) = \sum_{j \in J} \left[u_j^\beta(\bar{x}) - z_\nu\right]_+^2 \leq m \left[\phi_\beta(\bar{x}) - z_\nu\right]^2$$

and hence :

$$z_{\nu+1} = z_\nu + \left[\frac{1}{m} P_\beta(x_\nu, z_\nu)\right]^{1/2} \leq \phi_\beta(\bar{x}) . \tag{27}$$

Since by assumption $z_0 \leq \phi_\beta(\bar{x})$ this implies that z_ν is a non decreasing bounded sequence. Therefore it converges to a limit z^* and we have :

$$\lim_{\nu \to +\infty} P_\beta(x_\nu, z_\nu) = m \lim_{\nu \to +\infty} (z_{\nu+1} - z_\nu)^2 = 0$$

or equivalently :

$$\lim_{\nu \to +\infty} \sum_{j \in J} \left[u_j^\beta(x_\nu) - z_\nu\right]_+^2 = 0 \tag{28}$$

$$\lim_{\nu \to +\infty} \sum_{k \in K} \left[g_k(x_\nu)\right]_+^2 = 0 \tag{29}$$

We will set $g(x) = \sum_{k\in K} [g_k(x)]_+^2$, by assumption g is lower semi continuous. Denote by x^* a cluster point of the sequence (x_ν), then for any given $\varepsilon>0$ there exists ν_o large enough such that $\nu' \geq \nu_o$ implies :

$$(30) \qquad 0 \leq g(x^*) \leq g(x_{\nu'}) + \frac{\varepsilon}{2} \quad .$$

On the other hand for given $\delta>0$ there exists $\nu_1 \geq \nu_o$ such that $\nu' \geq \nu_1$ implies $0 \leq g(x_{\nu'}) \leq \frac{\varepsilon}{2}$ and therefore $0 \leq g(x^*) \leq \varepsilon$ with $\varepsilon > 0$ arbitrary. Therefore x^* satisfies the constraints of (P_β). It remains to verify that x^* is an optimal solution of this problem.

We notice first that :

$$(31) \qquad \phi_\beta(x^*) = \max_{j\in J} u_j^\beta(x^*) \geq \min_{x\in X} \max_{j\in J} u_j^\beta(x) = \phi_\beta(\bar{x}).$$

If we assume that $\phi_\beta(x^*) > \phi_\beta(\bar{x})$, then there exists $j_o \in J$ such that $u_{j_o}^\beta(x^*) > \phi_\beta(\bar{x})$. We will show that this leads to a contradiction with relation (28). Indeed if $\eta = u_{j_o}^\beta(x^*) - \phi_\beta(\bar{x})$ we have by (27) and by the lower semi continuity of u_{j_o} for ν' large enough :

$$(32) \qquad u_{j_o}^\beta(x_{\nu'}) - z_{\nu'} \geq u_{j_o}^\beta(x_{\nu'}) - \phi_\beta(\bar{x}) \geq u_{j_o}^\beta(x^*) - \phi_\beta(\bar{x}) - \frac{\eta}{2} = \frac{\eta}{2} .$$

which contradicts (28). Hence $\phi_\beta(x^*) = \phi_\beta(\bar{x})$. ■

We will now show how the "Method of Centres" introduced by HUARD can be used to compute Pareto optima. We assume that the functions $(f_i)_{i\in N}$ and $(g_k)_{k\in K}$ are continuous, that the set E is closed and that the subset $X = \{x \in E \mid g_k(x) \leq 0 \ \forall k\in K\}$ is bounded. Under these assumptions problem (P'_β) has a solution $(\bar{x},\bar{z})$ with $\bar{z} = \max_{j\in J} u_j^\beta(\bar{x})$ and one can formulate the following :

Algorithm II
---

1) Choose $x_o \in X$, $z_o \geq \max_{j\in J} u_j^\beta(x_o)$ and $\ell_o \geq z_o$.

2) Determine $x_{\nu+1}$ and $z_{\nu+1}$ such that :

(π_ν) $\qquad \psi_\nu(x_{\nu+1}, z_{\nu+1}) = \max_{(x,z)\in U_\nu} \psi_\nu(x,z)$

where :

$$\psi_\nu(x,z) = (\ell_\nu - z) \prod_{j\in J} (z - u_j^\beta(x)) \prod_{k\in K} (-g_k(x))$$

$$U_\nu = \{(x,z) \in E \times R \mid \ell_\nu \geq z \geq u_j^\beta(x) \quad \forall j \in J \ , \ g_k(x) \leq 0 \quad \forall k \in K\}$$

3) Define $\ell_{\nu+1} = \ell_\nu - r_\nu(\ell_\nu - z_{\nu+1})$ where r_ν is a given sequence of positive numbers satisfying $o < e \leq r_\nu \leq 1$.

4) Augment ν by one and return to 2).

We have the following convergence result for algorithm II :

Theorem 3 Assume that E is a closed convex set, that the functions $(f_i)_{i\in N}$ are locally lipschitzean , that the $(g_k)_{k\in K}$ are convex and continuous and that there exists $\overset{\vee}{x} \in E$ such that $g_k(\overset{\vee}{x}) < o$ for all $k\in K$. Then if the set $X = \{ x \in E \mid g_k(x) \leq o \quad \forall\, k\in K\}$ is compact, the sequence (z_ν) converges for $\nu \to +\infty$ to the optimal value $\bar{z}$ of (P_β) and any cluster point x^* of the sequence (x_ν) is a Pareto optimum of the $(f_i)_{i\in N}$ in X.

As X is compact and the functions u_j^β are continuous, the set U_ν is compact in $\mathbb{R}^{n+1}$ and therefore the problem $(\pi)_\nu$ has a solution $(x_{\nu+1}, z_{\nu+1})$ which satisfies all constraints of (P_β). Hence we have :

$$(33) \qquad \ell_\nu \geq z_{\nu+1} \geq \bar{z} .$$

Since by construction the sequence ℓ_ν is decreasing, it converges to a limit $\ell \geq \bar{z}$ which implies that $\lim_{\nu\to+\infty} (\ell_\nu - \ell_{\nu+1}) = o$ and we have finally $\lim_{\nu\to+\infty} \ell_\nu = \lim_{\nu\to+\infty} z_\nu = \ell$. We have now to prove that $\ell = \bar{z}$.

We will introduce the following perturbation function of problem (P_β) :

$$(34) \qquad \alpha(p) = \operatorname{Min} \{ \operatorname*{Max}_{j\in J} (u_j^\beta(x) + p_j) \mid x \in E, \; g_k(x) + p_k \leq o \quad \forall\, k \in K\}$$

where $p \in \mathbb{R}^J \times \mathbb{R}^K$.

Then one can show (see Lemma 1 below) that for any $\varepsilon > o$ there exists $h > o$ such that :

$$(35) \qquad \alpha(h) \leq \alpha(o) + \varepsilon.$$

Let us assume that $\bar{z} < \ell$ then since $\alpha(o) = \bar{z}$ there exists $h > o$ such that

$$\alpha(h) \leq \alpha(o) + \frac{\ell - \bar{z}}{2} = \bar{z} + \frac{\ell - \bar{z}}{2} < \ell$$

and this implies the existence of $\tilde{x} \in E$ and $\tilde{z} \in \mathbb{R}$ such that :

$$(36)\qquad \begin{cases} \tilde{z} < \ell \\ \tilde{z} - u_j^{\beta}(\tilde{x}) \geq h_j > 0 & \forall\, j \in J \\ g_k(\tilde{x}) \leq -h_k < 0 & \forall\, k \in K \end{cases},$$

But as $\ell \leq \ell_\nu$ for any ν, this implies that the point $(\tilde{x},\tilde{z}) \in U_\nu$ for any ν and therefore :

$$(37)\qquad \psi_\nu(x_{\nu+1},z_{\nu+1}) = \max_{(x,z)\in U_\nu} \psi_\nu(x,z) \geq \psi_\nu(\tilde{x},\tilde{z}) \geq (\ell-\tilde{z}) \prod_{j\in J} (\tilde{z}-u_j^{\beta}(\tilde{x})) \prod_{k\in K} (-g_k(\tilde{x})) > 0$$

which yields a contradiction, since by construction $\psi_\nu(x_{\nu+1},z_{\nu+1})$ tends to zero for $\nu \to +\infty$. Hence we have $\ell = \bar{z}$.

Let us now consider a subsequence $(x_{\nu'})$ of (x_ν) converging to a limit $x^* \in E$, by construction we have $(x_{\nu'},z_{\nu'}) \in U_{\nu'}$ and it results from the continuity of the functions u_j^{β} and g_k that :

$$(38)\qquad \begin{cases} \ell = \bar{z} \geq \max\limits_{j\in J} u_j^{\beta}(x^*) \\ g_k(x^*) \leq 0 \qquad \forall k \in K. \end{cases}$$

Since $\bar{z}$ is the optimum of (P_β) we have :

$$\bar{z} = \max_{j\in J} u_j^{\beta}(x^*)$$

which proves that the optimum is attained at x^* and, by Theorem 2, x^* is a Pareto optimum of the $(f_i)_{i\in N}$ in the set X.

It remains to prove :

Lemma 1 — Under the assumptions of Theorem 3 for any $\varepsilon > 0$ there exists $h > 0$ such that :

$$(35)\qquad \alpha(h) \leq \alpha(0) + \varepsilon \quad .$$

From the definition of $\alpha(0)$ for given $\varepsilon > 0$ there exists $x \in E$ such that :

$$(39)\qquad \begin{cases} \max\limits_{j\in J} u_j^{\beta}(x) \leq \alpha(0) + \frac{\varepsilon}{2} \\ g_k(x) \leq 0 \qquad \forall\, k \in K \end{cases}.$$

If we set $\rho = -\max\limits_{k\in K} g_k(x) > 0$, then by the convexity of g_k we have for any

$\theta \in [0,1]$:

$$(40) \quad g_k(\theta \overset{\vee}{x} + [1-\theta]x) \leq \theta g_k(\overset{\vee}{x}) + (1-\theta) g_k(x) \leq -\theta\rho \qquad \forall\, k \in K.$$

The vector $y = \theta \overset{\vee}{x} + (1-\theta)x \in E$ satisfies the constraints of the perturbed problem $\alpha(\theta\rho 1)$ and is such that $||y-x|| = \theta||\overset{\vee}{x} - x||$. Since the u_j^β are locally lipschitzean and y belongs to a bounded set we have for θ small enough :

$$(41) \quad \begin{aligned} \alpha(\theta\rho\, 1) &\leq \max_{j\in J} (u_j^\beta(y) + \theta\rho) \leq \max_{j\in J} (u_j^\beta(x) + L\,||x-y|| + \theta\rho) \\ &\leq \alpha(o) + \frac{\varepsilon}{2} + L\theta||\overset{\vee}{x} - x|| + \theta\rho \leq \alpha(o) + \varepsilon. \end{aligned}$$

■

References

AUBIN J.P. (1979)

Mathematical Methods of Game and Economic Theory.

North Holland.

AUBIN J.P., EKELAND I. (1976).

Estimates of the duality gap in Non convex Optimization

Mathematics of Operations Research1 p. 225-245.

BENSON H.P. (1977).

Efficiency and proper efficiency in Vector Maximization with respect to Cones.

General Motors Research Laboratories Report 2557.

BOWMAN J. Jr. (1976).

On the relationship of Tchebycheff norm and the efficient frontier of multiple criteria objectives.

in Multiple Criteria Decision Making

Lecture notes in Economics and Mathematical Systems n° 130.

Springer Verlag.

DUTTA R., VIDYASAGAR M. (1977).

New algorithm for constrained minimax optimization.

Mathematical Programming 13 p. 140-155.

HUARD P. (1967).

Resolution of mathematical programming with non linear constraints by the method of centres.

in Methods of non linear Programming North Holland (1967).

RECENT RESULTS IN THE THEORY OF MULTIOBJECTIVE OPTIMIZATION

K.-H. Elster and R. Nehse

Sektion Mathematik, Rechentechnik und ökonomische Kybernetik, Technische Hochschule Ilmenau, GDR

INTRODUCTION

In this paper, we consider the general vector optimization problem

$$f(x) \to \min \quad , \qquad \text{(VOP)}$$

$$x \in G := \{z \in R^n \mid g(x) \leq 0\} \quad ,$$

where $f: R^n \to R^p$, $g: R^n \to R^m$. The Euclidean spaces R^p and R^m are partially ordered in the usual sense, i.e., the order cones are R^p_+ and R^m_+, respectively. We want to characterize the so-called Pareto-optimal points $x^0 \in G$ of this problem, which are defined by

$$[f(G) + R^p_+ - f(x^0)] \cap R^p_- = \{0\}; \quad R^p_- := -R^p_+ \quad .$$

Clearly, from the mathematical point of view it is easier to describe weakly efficient solutions $x^0 \in G$ of the vector optimization problem (VOP), which are defined by

$$[f(G) + R^p_+ - f(x^0)] \cap \operatorname{int} R^p_- = \emptyset \quad .$$

However, from the practical point of view it is more convenient to restrict the set of Pareto-optimal points. For instance, we can consider proper efficient solutions $x^0 \in G$, for which

$$T(f(G) + R^p_+, f(x^0)) \cap R^p_- = \{0\}$$

is satisfied; here $T(V, v^0)$ denotes the tangent cone of $V = f(G) + R^p_+$ at $v^0 = f(x^0)$.

Clearly, more general cases may also be considered: for instance, problems in locally convex Hausdorf vector spaces. In

this case closed convex cones are used instead of R_+^p and R_+^m, respectively (see Craven (1980), Jahn (1980), and Minami (1981)).

OPTIMALITY AND DUALITY

In this section we shall give some theoretical results about optimality via duality theorems for certain vector optimization problems.

Definition

A vector optimization problem

$$d(z) \to \max , \qquad z \in D;\ D \subseteq R^q,\quad d: R^q \to R^p \tag{DVOP}$$

is dual to (VOP) if and only if

$$d(z) - f(x) \in R_+^p \setminus \{0\} \quad \forall\, x \in G,\ \forall\, z \in D \ .$$

This means that the weak duality relation

$$d(z) \not\geq f(x) \quad \forall\, x \in G,\ \forall\, z \in D$$

is fulfilled for the two problems (VOP) and (DVOP).

Let us denote the set of all Pareto-optimal solutions of (VOP) by arg min f(G) and the set of all Pareto-optimal solutions of (DVOP) by arg max d(D). Then we have.

Proposition 1. Let (VOP) and (DVOP) be dual problems.

(a) If there are $x^0 \in G$ and $z^0 \in D$ such that $f(x^0) = d(z^0)$, then $x^0 \in$ arg min f(G) and $z^0 \in$ arg max d(D).

(b) If there is $x^0 \in G$ such that $f(x^0) \in d(D)$, then $x^0 \in$ arg min f(G).

(c) If there is $z^0 \in D$ such that $d(z^0) \in f(G)$, then $z^0 \in$ arg max d(D).

It is clear that each of these assertions is a criterion for optimality.

The main problem is therefore to construct convenient duals for a given problem. Some examples are given by Bitran (1978), Elster and Nehse (1979a,b, 1980), Gros (1978, 1980), Isermann (1978), Nehse (1978, 1981), and Schönfeld (1970).

Let the following problem be associated with (VOP) for $\lambda \in \operatorname{int} R_+^p$:

$$L(V,y) \to \max, \qquad (DOP_\lambda)$$

$$(V,y) \in G_\lambda^* := \{(U,x) \in R_+^{pxm} \times R^n \mid \lambda^T \nabla_x L(U,x) = 0\} \quad ,$$

where $\nabla_x L(U,x)$ denotes the Jacobian of

$$L(U,x) := f(x) + U\, g(x), \quad U \in R_+^{pxm}, \quad x \in R^n \quad ,$$

with respect to x. We can prove the following theorem (see Nehse, 1981).

Theorem 1

Let f_i $(i=1,2,\ldots,p)$ and g_j $(j=1,2,\ldots,m)$ be convex and differentiable.

(a) For each $\lambda \in \operatorname{int} R_+^p$ the vector optimization problems (VOP) and (DOP_λ) are dual problems.

(b) If a regularity condition[+] is satisfied for (VOP) at $x^0 \in G$ and if x^0 is a proper efficient solution of (VOP), then there are $U^0 \in R_+^{pxm}$, $\lambda_0 \in \operatorname{int} R_+^p$ such that $(U^0,x^0) \in \arg\max L(G_\lambda^*)^+$ and $f(x^0) = L(U^0,x^0)$.

[+] As a regularity condition we may use

$$T^*(G,x^0) = K^*(G,x^0)$$

where

$$K(G,x^0) := \{z \in R^n \mid z^T \nabla g_j(x^0) \leq 0, \quad j = 1,2,\ldots,m\}$$

is the so-called linearization cone of G at x^0; T^* and K^* denote the polar cones of T and K, respectively.

OPTIMALITY CONDITIONS AND METHODS

Here we restrict ourselves to optimality conditions which are also useful methodologically. Most of them we formulate as versions of Kuhn-Tucker conditions using convenient surrogate problems for (VOP).

One way to obtain Pareto-optimal solution to (VOP) is to solve the following problem:

$$f_k(x) \to \min \quad , \qquad (P_k(u_k))$$

$$x \in G \cap G_k(u_k),$$

$$u_k \in B_k,$$

where

$$u_k := (u_1, u_2, \ldots, u_{k-1}, u_{k+1}, \ldots, u_p)^T \quad ,$$

$$G_k(u_k) := \{x \in R^n \mid f_i(x) \leq u_i, \quad i = 1, \ldots, p; \quad i \neq k\} \quad ,$$

$$E_k := \{u_k \in R^{p-1} \mid G_k(u_k) \neq \emptyset\} \quad .$$

Theorem 2

A unique solution of $(P_k(u_k))$, for any $1 \leq k \leq p$, is a Pareto-optimal solution of (VOP). Conversely, any Pareto-optimal solution of (VOP) solves $(P_k(u_k))$ for some $u_k \in E_k$ and for all $k = 1,2,\ldots,p$.

Using this theorem (Chankong and Haimes, 1978) and the classical theory of nonlinear programming, it is possible to construct interactive methods for solving (VOP).

We shall now give a theorem that is useful in solving general vector optimization problems, i.e., without convexity and continuity assumptions (see Ester and Schwartz, forthcoming).

Let (VOP) be given as

$$f(x) \to \min$$
$$x \in G,$$

where $G \subseteq R^n$ is any nonempty subset of R^n such that $f(G)$ is closed and $f(x) \geq 1$ for all $x \in G$ and any $1 \in R^p$.
Let

$$f_{s,r}(x,h) = sh + \sum_{i=1}^{p} r_i f_i(x) \to \min \qquad \text{(EP)}$$

$$x \in G \cap G_{a,b,g} \quad ,$$

where

$$G_{a,b,g} = \{x \in R^n \mid b \prod_{i=1}^{p} (f_i(x) - g_i h - k_i) = a \; ;$$

$$f_i(x) - g_i h - k_i \geq 0 \quad , \quad i = 1,2,\ldots,p\}$$

and

$$s > 0,\ a > 0,\ b > 0,\ r_i \geq 0,\ g_i \geq 0,\ k_i \in R,\quad i = 1,2,\ldots,p \quad .$$

Then we have

Theorem 3

(a) If (x^0, h^0) is a solution of (EP) for any parameters s,b,r_i,g_i,k_i, $i = 1,2,\ldots,p$, then $x^0 \in \arg\min f(G)$.

(b) If $x^0 \in \arg\min f(G)$ and $\varepsilon > 0$, then there are convenient parameters $s = b = g_i = 1$, $a^0 > 0$, $r_i = 0$, $k_i = f_i(x^0)$, such that for a solution $(x^0(a^0), h^0(a^0))$ of (EP)

$$|f_i(x^0(a^0)) - f_i(x^0)| < \varepsilon, \quad i = 1,2,\ldots,p \quad .$$

This means that $x^0 \in G$ is Pareto-optimal if and only if $f(x^0)$ may be approximated 'sufficiently exactly' by $f(x)$, $x \in G \cap G_{a,b,g}$, for convenient parameters.

We can adopt an approach developed by Ben-Israel et al. (1981) to derive an optimality condition which can be used to check the optimality of a feasible solution of (VOP).

Let f_i $(i = 1,2,\ldots,p)$ and g_j $(j = 1,2,\ldots,m)$ be convex functions and let, for any $h:R^n \to R$,

$$C_h^{<}(x^0) = \{d \in R^n \mid \exists\, \alpha_0 > 0 : h(x^0 + \alpha d) < h(x^0) \quad \forall \alpha \in (0,\alpha_0)\} \quad ,$$

$$C_h^{=}(x^0) = \{d \in R^n \mid \exists\, \alpha_0 > 0 : h(x^0 + \alpha d) = h(x^0) \quad \forall \alpha \in (0,\alpha_0)\} \quad ,$$

$$F(G,x^0) = \{d \in R^n \mid \exists\, \alpha_0 > 0 : x^0 + \alpha d \in G \quad \forall \alpha \in (0,\alpha_0)\} \quad ,$$

where

$C_h^{<}(x^0)$ is the cone of directions of descent of h at x^0,

$C_h^{=}(x^0)$ is the cone of directions of constancy of h at x^0,

$F(G,x^0)$ is the cone of feasible directions of G at x^0.

Firstly, we use the trivial fact that $x^0 \in G$ is Pareto-optimal for (VOP) if and only if x^0 is a solution of

$$f_0(x) = \sum_{i=1}^{p} f_i(x) \to \min \quad ,$$

$$x \in G_0 = G \cap G(x^0) \quad ,$$

where

$$G(x^0) = \{x \in R^n \mid f_i(x) \leq f_i(x^0) \quad , \quad i = 1,2,\ldots,p\} \quad .$$

Hence, we obtain immediately

Theorem 4

Let f_i $(i = 1,2,\ldots,p)$ and g_j $(j = 1,2,\ldots,m)$ be convex functions. $x^0 \in G$ is Pareto-optimal for (VOP) if and only if

$$C_{f_0}^{<}(x^0) \cap F(G_0,x_0) = \emptyset \quad .$$

From Dubovickij and Miljutin, this equation is equivalent to

$$[(C_{f_0}^{<}(x^0))^* \cap -F^*(G_0,x_0)] \setminus \{0\} \neq \emptyset \quad .$$

If $J(x^0)$ denotes the index set of the active constraints g_j at x^0 then we obtain explicitly

$$F(G_0,x^0) = \Big(\bigcap_{j\in J(x^0)} C^{\leq}_{g_j}(x^0)\Big)\cap\Big(\bigcap_{i=1} C^{\leq}_{f_i}(x^0)\Big) \quad .$$

Using

$$I(x^0) = \{1,2,\ldots,p\}\cup\{p+j \mid j\in J(x^0)\} \quad ,$$

we obtain

$$F(G_0,x^0) = \bigcup_{I'\subseteq I(x^0)} (C^{<}_{I(x^0)\setminus I'}(x^0)\cap C^{=}_{I'}(x^0))$$

where the right-hand side represents the reunion of intersections of the cones associated with $i\in I(x^0)\setminus I'$ and $i\in I'$, respectively. Therefore, taking $G_0 = G$ for simplicity, we obtain:

x^0 is Pareto-optimal for (VOP) if and only if for each subset $I'\subseteq\{1,2,\ldots,p\} = I$ we have

$$C^{<}_{I\setminus I'}(x^0)\cap C^{=}_{I'}(x^0) = \emptyset \quad .$$

In the dual case this becomes:

For each subset $I'\subseteq I$ there exist

$$y^i \in (C_{f_i}(x^0))^* \quad , \quad i\in I'' = I\setminus I' \quad ,$$

$$y^i \neq 0 \text{ for at least one } i\in I''$$

such that

$$\sum_{i\in I''} y^i \in -(C^{=}_{I'}(x^0))^* \quad .$$

In the differentiable case we have:

For each subset $I'\subseteq I$ there exist

$$\lambda_i \geq 0 \quad , \quad i\in I'' \quad ,$$

$$\lambda_i \neq 0 \quad \text{for at least one } i \in I''$$

such that

$$\sum_{i \in I''} \lambda_i \nabla f_i (x^0) \in (C_I^{=}, (x^0))^* \quad .$$

Theoretically, to check the optimality of x^0 we must solve $2^{\text{card } I}$ such systems, where the difficulty lies in the representation of the cone $(C_I^{=}, (x^0))^*$; however, in some cases (for example, if f_i, $i \in I$, is faithfully convex) we have $C^{=}$ as a subspace and, therefore, $(C^{=})^*$ is also a subspace.

REFERENCES

Ben-Israel, A., Ben-Tal, A., and Zlobec, S. (1981) Optimality in nonlinear programming. A feasible directions approach. Wiley, New York.

Bitran, G.R. (1978) Duality for nonlinear multiple-criteria optimization problems. Technical Report No. 155, Operations Research Center, Massachusetts Institute of Technology.

Chankong, V. and Haimes, Y.Y. (1978) On multiobjective optimization theory. A unified treatment. Technical Memorandum No. SEG-WRG-76-8, Systems Engineering Department, Case Western Reserve University, Cleveland, Ohio.

Craven, B.D. (1980) Lagrangian conditions, vector minimization, and local duality. Research Report No. 37-1980, University of Melbourne.

Elster, K.-H. and Nehse, R. (1979a) Dualität und Vektoroptimierungsprobleme. Jahrestagung "Math. Optimierung" 1979 in Vitte (Hiddensee); Vortragsauszüge 33-36.

Elster, K.-H. and Nehse, R. (1979b) Ergebnisse und Probleme der Vektoroptimierung. 24. Internat. Wiss. Koll. TH Ilmenau 1979, Vortragsreihe B1, 17-20.

Elster, K.-H. and Nehse, R. (1980) Zum derzeitigen Entwicklungsstand der Vektoroptimierung. Wiss. Z. TH Ilmenau 26 (6), 63-80.

Ester, J. and Schwartz, B. (forthcoming) Ein erweitertes Effizienztheorem. Math. Operationsf. Statist., Ser. Optimization.

Gros, C. (1978) Generalization of Fenchel's duality theorem for convex vector optimization. Eur. J. Oper. Res. 2, 368-376.

Gros, C. (1980) Approche de la dualité en optimisation multicritère. Cah. Centr. Etud. Rech. Opér. 22, 73-79.

Isermann, H. (1978) On some relations between a dual pair of multiple objective linear programs. Zeitschr. Oper. Res. 22 (1), 33-41.

Jahn, J. (1980) Duality theory for vector optimization problems in normed linear spaces. TH Darmstadt, Preprint-No. 534, February.

Minami, M. (1981) Weak Pareto-optimality of multiobjective problems in a locally convex linear topological space. J. Opt. Theor. Appl. 34 (4), 469-484.

Nehse, R. (1978) Dualitätsaussagen für Vektoroptimierungsprobleme. Akad. Wiss. DDR, ZKI-Informationen 2/78, 126-136.

Nehse, R. (1981) Duale Vektoroptimierungsprobleme vom WOLFE-Typ. Seminarber. Humboldt-Univ. Berlin 1981.

Schönfeld, P. (1970) Some duality theorems for the non-linear vector maximum problem. Zeitschr. f. Unternehmensforsch. 14, 51-63.

LEXICOGRAPHIC GOAL PROGRAMMING: THE LINEAR CASE

H. Isermann

Universität Bielefeld, FRG

1. INTRODUCTION

Among the various approaches to multicriteria problems which have been suggested, goal programming has proven to be widely adopted. In goal programming (cf., e.g. Charnes-Cooper [1], Ijiri [5], Ignizio [4], Lee [9], Spronk [12], instead of attempting to optimize the objective criteria directly, the deviations between goals and levels of achievement within the given set of system constraints are to be minimized. Thus the objective becomes the minimization of these deviations, based on the importance assigned to them. Goal programming models may be regarded as special compromise models for a satisficing decision maker (Isermann [6], [7]).

In many applications (cf. e.g. Lee [9], Lin [10], Nijkamp-Spronk [11]) preemptive priority based goal programming models are employed in order to reflect a compromise concept where a priority structure with respect to the achievement of the considered goal levels has to be taken into account. On the basis of this concept the achievement of those goals at any one priority level is immeasurably preferred to the achievement of the goals at any lower priority level.

In Section 2 we shall demonstrate that each preemptive priority based linear goal program can be equivalently represented by a lexicographic linear goal program and solved by lexicographic simplex method. The proposed lexicographic simplex method differs from

the ordinary simplex method with regard to the pivot column selection rule and differs from the solution methods (cf. Ignizio [4], pp. 31 - 71; Lee [9], pp. 93 - 125) which have been proposed in order to solve preemptive priority based linear goal programs.

To a preemptive priority based linear goal program a multidimensional "dual problem" has been formulated (cf. Ignizio [4], pp. 98 - 114), which, however, does not exhibit the duality relations which are commonly associated to a dual pair of mathematical programs. In Section 3 we shall introduce a proper dual problem and establish the duality relations which provide the theoretical basis for the proposed solution method, any post optimality analysis in preemptive priority based linear goal programming and related information gathering procedures in interactive goal programming.

Before going further, for convenience, let us introduce the following notation. Let $\mathbb{R}$ denote the set of the real numbers and $\mathbb{R}_o$ the set of the nonnegative real numbers. With regard to lexicographic vector and matrix inequalities the following convention will be applied: For $a, b \in \mathbb{R}^M$ the *strict lexicographic inequality* $a \succ b$ holds, if and only if, $a \neq b$ and $a_i > b_i$ holds for $i = \min \{1,\ldots,M \mid a_m \neq b_m\}$ and the *weak lexicographic inequality* $a \succsim b$ holds, if and only if, $a \succ b$ or $a = b$. For any two $(M \times N)$-matrices $C = (c_1,\ldots,c_N)$ $D = (d_1,\ldots,d_N)$, $C \succ D$ holds, if and only if, $c_n \succsim d_n$ for all $n = 1,\ldots,N$. Accordingly, $C \succsim D$ holds, if and only if, $c_n \succsim d_n$ for all $n = 1,\ldots,N$. The transpose of a vector or a matrix will be denoted by an upper index T.

2. The Lexicographic Linear Goal Program

In goal programming, instead of attempting to optimize the objective criteria directly, the deviations between goals and levels of achievement dictated by the set of system constraints are to be minimized. Thus the objective becomes the minimization of these deviations, based on the preferences assigned to them by the decision maker. In goal programming an ordinal as well as a cardinal assessment of preferences may be utilized: All relevant deviations between goals and levels of achievement are grouped according to their respective priority levels. The minimization of those deviations at anyone priority level is immeasurably preferred to the minimization of the deviations at any lower priority level. The priority levels will be

denoted by the index $k = 1,\ldots,K$, with $k = 1$ representing the highest priority level and $k = K$ representing the lowest priority level. To each priority level k a linear dispreference function z_k is associated. If deviations expressed in different measures have been assigned to the same priority level k the decision maker has to assign weights by which the different deviations can be expressed in terms of a common unit of measure. Applying the non-Archimedean preemptive priority weights $P_k >> P_{k+1}$ $(k = 1,\ldots,K-1)$, originally proposed by Charnes-Cooper ([1] pp. 756 - 757) the preemptive priority based linear goal programming model is the linear program (cf. e.g. Ijiri [5] pp. 49)

$$\min \phi = \sum_{k=1}^{K} P_k z_k (y^+, y^-) \qquad (1)$$

$$\text{s.t.} \quad Cx - I y^+ + I y^- = g \qquad (2) \qquad \text{(PPGP)}$$

$$Ax = b \qquad (3)$$

$$x, y^+, y^- \geq 0 \qquad (4)$$

Here $C = (c_1,\ldots,c_N)$ is the $J \times N$ matrix of the criteria coefficients, $x \in \mathbb{R}_o^N$ the vector of the instrumental variables, I is a $J \times J$ identity matrix, $g \in \mathbb{R}^J$ is the goal vector, $A = (a_1,\ldots,a_N)$ is the $M \times N$ technological matrix and $b \in \mathbb{R}^M$ is the vector of availabilities, $y^+ \in \mathbb{R}_o^J$ and $y^- \in \mathbb{R}_o^J$ measure respectively the positive deviations and negative deviations from the stated goal vector g.

However, the linear programming approach may cause considerable difficulties. In the context of a specific application the question arises, which numerical values are to be attached to the preemptive priority weights P_k $(k = 1,\ldots,K)$ in order to take care that low priority goals are considered only after higher priority goals are achieved as desired. There exists no foolproof method which translates the preemptive priority weights *a priori* into real-valued weights which can then be used within a linear programming format. These difficulties lead to the development of a solution procedure (cf. Ignizio [4] pp. 42 - 60; Lee [9] pp. 93 - 125, Kornbluth [8], pp. 199 - 203) which is designed to *successively* minimize the linear dispreference functions $z_k (y^+, y^-)$ $(k = 1,\ldots,K)$.

As the appropriate priority structure can be immediately represented by a vector-valued objective function which is to be minimized in

a lexicographic order, we can dispose of a solution method which *simultaneously* considers all J scalar-valued objective functions if necessary: a lexicographic simplex method. Recall that each z_k (y^+, y^-) $(k = 1,\ldots,K)$ is assumed to be a linear function. Let

$$z(y^+, y^-) = \begin{pmatrix} z_1(y^+, y^-) \\ \vdots \\ z_K(y^+, y^-) \end{pmatrix} := D^+ y^+ + D^- y^-$$

where $D^+ = (d_1^+,\ldots,d_J^+)$ and $D^- = (d_1^-,\ldots,d_J^-)$ are $K\times J$ matrices. Then to (PPGP) we can relate the following linear lexicographic goal program

$$\begin{aligned} &\textit{lexmin } z = D^+ y^+ + D^- y^- \qquad (5) \\ &\text{s.t.} \quad (2) - (4) \qquad\qquad\qquad\qquad\qquad\qquad (\text{LGP}) \end{aligned}$$

which can be solved by a lexicographic simplex method.

The lexicographic simplex method differs from the ordinary simplex method with regard to the check for optimality and the selection rule for the pivot column: Let $(\hat{y}^+, \hat{y}^-, \hat{x})$ be a basic feasible solution for (LGP). If all $K\times 1$ vectors of the reduced cost coefficients in the respective multiple objective simplex tableau are lexicographically smaller than or equal to the zero vector then $(\hat{y}^+, \hat{y}^-, \hat{x})$ is optimal for (LGP). Otherwise an entering basic variable has to be selected according to the following selection rule: From among all $K\times 1$ vectors of the reduced cost coefficients which are lexicographically greater than the zero vector select the lexicographically maximal vector. The corresponding nonbasic variable becomes the new entering basic variable.

In order to illustrate the lexicographic simplex method we consider the lexicographic linear goal program

$$\textit{lexmin } z = \begin{pmatrix} 2\,y_1^+ + y_1^- + 3\,y_2^- \\ y_2^+ + y_3^- \\ y_4^- \end{pmatrix}$$

$$\begin{aligned} \text{s.t. } & 3x_1 + 6x_2 - y_1^+ + y_1^- = 12 \\ & x_1 - y_2^+ + y_2^- = 5 \\ & 5x_1 + 2x_2 - y_3^+ + y_3^- = 30 \\ & 2x_1 + 2x_2 - y_4^+ + y_4^- = 10 \\ & x_1, x_2, y_1^+, \ldots, y_4^+, y_1^-,\ldots, y_4^- \geq 0 \end{aligned}$$

The corresponding initial simplex tableau in canonical form (cf. Dantzig [2], p. 195) is given in Table 1. For clarity, the zero elements of the tableau have been omitted.

	x_1	x_2	y_1^+	y_2^+	y_3^+	y_4^+	
y_1^-	3	6	-1				12
y_2^-	1			-1			5
y_3^-	5	2			-1		30
y_4^-	2	2				-1	10
z_1	6	6	-3	-3			27
z_2	5	2		-1	-1		30
z_3	2	2				-1	10

Table 1: Initial Multiple Objective Simplex Tableau

The initial basic feasible solution $y^{-(1)} = (12,5,30,10)^T$, $y^{+(1)} = 0$, $x^{(1)} = 0$ with $z^{(1)} = (27,30,10)^T$ is not optimal as by raising x_1 or x_2 to some positive level each component of $z^{(1)}$ can be reduced. If we consider the reduced cost coefficients of the objective associated with priority level one x_1 as well as x_2 are candidates for becoming the new basic variable. The algorithms proposed to solve the pre-emptive priority based goal program would perform an arbitrary selection (cf. e.g. Ignizio [4], p. 47; Lee [9], p. 105) whereas the lexicographic simplex method selects x_1 as the new entering basic variable: By applying the pivot column selection rule we obtain $(6,5,2)^T = \text{lexmax}\ \{(6,5,2)^T, (6,2,2)^T\}$.

The new tableau, with x_1 and y_1^- interchanged, is as given in Table 2.

	x_2	y_1^+	y_2^+	y_3^+	y_4^+	y_1^-	
x_1	2	$-1/3$				$1/3$	4
y_2^-	-2	$1/3$	-1			$-1/3$	1
y_3^-	-8	$5/3$		-1		$-5/3$	10
y_4^-	-2	$2/3$			-1	$-2/3$	2
z_1	-6	-1	-3			-2	3
z_2	-8	$5/3$	-1	-1		$-5/3$	10
z_3	-2	$2/3$			-1	$-2/3$	2

Table 2: Optimal Multiple Objective Simplex Tableau

The basic feasible solution $y^{-(2)} = (0,1,10,2)^T$, $y^{+(2)} = 0$, $x^{(2)} = (4,0)^T$ with $z^{(2)} = (3,10,2)^T$ is optimal for (LGP), as all vectors of the reduced cost coefficients are lexicographically smaller than or equal to the zero vector and thus satisfy the sufficient optimality condition. The *negative* vectors of the reduced cost coefficients are called *trade-off vectors* as they specify the rate of exchange among the considered objective functions, if the respective nonbasic variable is raised to a positive level. The trade-off vector associated to y_1^+ reads $\bar{d}_1^+ = (1,-5/3,-2/3)^T$ and thus indicates that z_1 will be increased by 1 unit, z_2 and z_3 will be decreased by $5/3$ units and $2/3$ units, respectively, if y_1^+ is raised from 0 to 1. In other words: The trade-off vector associated to y_1^+ indicates that a change in the priority assignment of the objective functions, such that to z_2 instead of z_1 the highest priority is assigned, the current basic feasible solution is no more optimal. A more intensive discussion of this point and related topics will be presented in the next section.

3. Duality and Its Relevance for the Multiple Goal Decision Maker

In this section we shall analyse some relations between the lexicographic linear program and its *dual*. The dual of (LGP) is also a lexicographic linear program which is based on the same given information, the matrices D^+, D^-, A and C and the vectors g and b. To (LGP), which we shall term the *primal*, we associate the dual

$$
\begin{array}{lrcll}
\textit{lexmax}\ w = & U\,g + V\,b & & (6) & \\
\text{s.t.} & U\,C + V\,A & \lesssim & 0 \quad (7) & \\
& -U & \lesssim & D^+ \quad (8) & \text{(DLGP)} \\
& U & \lesssim & D^- \quad (9) &
\end{array}
$$

In (DLGP) U is a K×J matrix of variables and V is a K×M matrix of variables. In restriction (7) 0 is a K×N zero matrix. Note that in (DLGP) all constraints represent weak lexicographic inequalities. In order to illustrate the set-up of the dual we consider the linear goal program we solved in Section 2. The respective dual problem reads

$$
\begin{array}{lrrrrrrrrl}
\textit{lexmax}\ w = & 12\,u_1 & + & 5\,u_2 & + & 30\,u_3 & + & 10\,u_4 & \\
\text{s.t.} & 3\,u_1 & + & u_2 & + & 5\,u_3 & + & 2\,u_4 & \lesssim 0 \\
& 6\,u_1 & & & + & 2\,u_3 & + & 2\,u_4 & \lesssim 0 \\
& -\,u_1 & & & & & & & \lesssim \begin{pmatrix} 2 \\ 0 \end{pmatrix} \\
& & - & u_2 & & & & & \lesssim \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
& u_1 & & & & & & & \lesssim \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\
& & & u_2 & & & & & \lesssim \begin{pmatrix} 3 \\ 0 \end{pmatrix} \\
& & & & & u_3 & & & \lesssim \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
& & & & & & & u_4 & \lesssim \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
& & & & & u_3, & & u_4 & \gtrsim 0
\end{array}
$$

and has the optimal solution $\hat{u}_1 = (-1,-5/3,-2/3)^T$, $\hat{u}_2 = (3,0,0)^T$, $\hat{u}_3 = (0,1,0)^T$, $\hat{u}_4 = (0,0,1)^T$ with $\hat{w} = (3,10,2)^T = z^{(2)}$.

Let S and T denote the feasible set of (LGP) and (DLGP), respectively. In the context of goal programming we can assume without loss of generality that in (LGP) each z_k $(k = 1,\ldots,K)$ is bounded from below in S.

LEMMA 1. *Consider problem* (LGP) *with each* z_k $(k = 1,\ldots,K)$ *being bounded from below in* $S \neq \emptyset$. *Then* (LGP) *has an optimal solution.*

Proof. In order to prove this assertion, let $(x^{(p)}, y^{+(p)}, y^{-(p)})$ denote the p-th basic feasible solution for the linear system (2) - (4). Suppose the set of basic feasible solutions for the system (2) - (4) has cardinality r. Consider $(\hat{x},\hat{y}^+,\hat{y}^-)$ with

$$z(\hat{x},\hat{y}^+,\hat{y}^-) = \text{lexmin}\,\{\, z(x^{(1)},y^{+(1)},y^{-(1)}),\ldots,\ z(x^{(r)},y^{+(r)},y^{-(r)})\,\}\,.$$

As z is a linear vector function and each z_k $(k = 1,\ldots,K)$ is bounded from below in S there exists no $(\bar{x},\bar{y}^+,\bar{y}^-) \in S$ such that $z(\bar{x},\bar{y}^+,\bar{y}^-) \lneq z(\hat{x},\hat{y}^+,\hat{y}^-)$ holds. Hence $(\hat{x},\hat{y}^+,\hat{y}^-)$ is optimal for (LGP).

An immediate implication of the proof of Lemma 1 is

LEMMA 2.

At least one basic feasible solution for (LGP) *is optimal.*

The duality relation between (LGP) and (DLGP) will be characterized by the following statements:

LEMMA 3. $z(x,y^+,y^-) \gtrsim w(U,V)$ *for all* $(x,y^+,y^-) \in S$ *and* $(U,V) \in T$.

Proof. Let $(U,V) \in T$ and $(x,y^+,y^-) \in S$. Because of the nonnegative restriction (4) we obtain the system

$$\begin{array}{rcl} UCx + VAx & \lesssim & O \\ -U\,y^+ & \lesssim & D^+\,y^+ \\ U\,y^- & \lesssim & D^-\,y^- \end{array}$$

and thus $z(x,y^+,y^-) = D^+\,y^+ + D^-\,y^- \gtrsim U\,g + V\,b = w(U,V)$.

LEMMA 4. *Let* $(\hat{x}, \hat{y}^+, \hat{y}^-) \in S$ *and* $(\hat{U},\hat{V}) \in T$ *with* $z(\hat{x},\hat{y}^+,\hat{y}^-) = w(\hat{U},\hat{V})$. *Then* $(\hat{x},\hat{y}^+,\hat{y}^-)$ *is optimal for* (LGP) *and* $(\hat{U},\hat{V})$ *is optimal for* (DLGP).

Proof. In order to prove this assertion assume, to the contrary, that $(\hat{U},\hat{V})$ is not optimal. Then there exists a $(\bar{U},\bar{V}) \in T$ with $w(\bar{U},\bar{V}) \gtrdot w(\hat{U},\hat{V}) = (z(\hat{x},\hat{y}^+,\hat{y}^-)$, which implies a contradiction to Lemma 3. A similar argument may be applied to $(\hat{x},\hat{y}^+,\hat{y}^-)$.

THEOREM 1. *Let* $(\hat{x},\hat{y}^+,\hat{y}^-)$ *be an optimal solution for* (LGP). *Then there exists an optimal solution* $(\hat{U},\hat{V})$ *for* (DLGP) *and* $z(\hat{x},\hat{y}^+,\hat{y}^-) = w(\hat{U},\hat{V})$.

Proof. According to Lemma 2 at least one basic feasible solution for (LGP) is optimal. Let $(\hat{x},\hat{y}^+,\hat{y}^-)$ be an optimal basic solution for (LGP). Denote by B and B^{-1} the associated $(J+M)\times(J+M)$ basic matrix and its inverse, respectively, and by R the respective nonbasic matrix of the linear system

$$\begin{pmatrix} C & -I & I \\ A & O & O \end{pmatrix} \begin{pmatrix} x \\ y^+ \\ y^- \end{pmatrix} = \begin{pmatrix} g \\ b \end{pmatrix}$$

where we assume without loss of generality rank $(A) = M$. Then the optimal basic solution $(\hat{x},\hat{y}^+,\hat{y}^-)$ is given by

$$\begin{pmatrix} \hat{x}_B \\ \hat{y}_B^+ \\ \hat{y}_B^- \end{pmatrix} = B^{-1} \begin{pmatrix} g \\ b \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \hat{x}_R \\ \hat{y}_R^+ \\ \hat{y}_R^- \end{pmatrix} = O.$$

Let (O_B,D_B^+,D_B^-) be the $K \times (J+M)$ matrix of the criteria coefficients corresponding to B and (O_R,D_R^+,D_R^-) be the $K \times (N+J-M)$ matrix of the criteria coefficients corresponding to R and $(\hat{U},\hat{V}) := (O_B,D_B^+,D_B^-)\,B^{-1}$.

Now we express z as a function of the nonbasic variables of $(\hat{x},\hat{y}^+,\hat{y}^-)$:

$$(O,D^+,D^-)\begin{pmatrix}x\\y^+\\y^-\end{pmatrix} = (O_B,D_B^+,D_B^-)\begin{pmatrix}x_B\\y_B^+\\y_B^-\end{pmatrix} + (O_R,D_R^+,D_R^-)\begin{pmatrix}x_R\\y_R^+\\y_R^-\end{pmatrix}$$

$$= (O_B,D_B^+,D_B^-)\left[B^{-1}\begin{pmatrix}g\\b\end{pmatrix} - B^{-1}R\begin{pmatrix}x_R\\y_R^+\\y_R^-\end{pmatrix}\right] + (O_R,D_R^+,D_R^-)\begin{pmatrix}x_R\\y_R^+\\y_R^-\end{pmatrix}$$

$$= (\hat{U},\hat{V})\begin{pmatrix}g\\b\end{pmatrix} + \left[(O_R,D_R^+,D_R^-) - (\hat{U},\hat{V})\,R\right]\begin{pmatrix}x_R\\y_R^+\\y_R^-\end{pmatrix}$$

The optimality of $(\hat{x},\hat{y}^+,\hat{y}^-)$ implies

$$(O_R,D_R^+,D_R^-) \gtrsim (\hat{U},\hat{V})\,R, \qquad (10)$$

as otherwise by rising a nonbasic variable to some positive level a lexicographic lower value for z could be determined. However, $(C_B,D_B^+,D_B^-) = (\hat{U},\hat{V})B$ and (10) imply that $(\hat{U},\hat{V})$ satisfies the dual constraints (7) - (9). Moreover, we obtain $z(\hat{x},\hat{y}^+,\hat{y}^-) = (\hat{U},\hat{V})\begin{pmatrix}g\\b\end{pmatrix} = w(\hat{U},\hat{V})$. Hence $(\hat{U},\hat{V})$ is optimal for (DLGP) according to Lemma 4.

In general, the optimal solution for (DLGP) cannot immediately be read from the final multiple objective simplex tableau. The dual solution which is presented in the final tableau is the solution for the dual to the lexicographic linear program which is found in the initial tableau in canonical form. In our earlier example of Section 2 the objective of the lexicographic linear program which is represented in the initial tableau of Table 1 is

$$\text{lexmin} \begin{pmatrix} 27 - 6x_1 - 6x_2 + 3y_1^+ + 3y_2^+ \\ 30 - 5x_1 - 2x_2 + y_2^+ + y_3^+ \\ 10 - 2x_1 - 2x_2 + y_4^+ \end{pmatrix}$$

and the corresponding dual objective is

$$\text{lexmax} \begin{pmatrix}27\\30\\10\end{pmatrix} + 12u_1 + 5u_2 + 30u_3 + 10u_4$$

The optimal solution for this dual problem is found in Table 2: $\hat{u}_1 = (-2,-6/3,-2/3)^T$, $\hat{u}_2 = \hat{u}_3 = \hat{u}_4 = 0$, while the optimal solution for the respective (DLGP) has to be determined by means of $(\hat{U},\hat{V}) = (O_B,\ D_B^+,D_B^-)\ B^{-1}$.

Ignizio ([4], pp. 98 - 113) formulates a multidimensional "dual" of a preemptive priority based linear goal program which he characterizes on p. 99 as follows: "Rather, the dual of a linear goal program is a traditional linear programming problem except that it has *multiple right-hand side values.*" In order to illustrate his formulation of the multidimensional "dual" we shall present an example due to Ignizio([4] pp. 99 - 101). Applying our notation, the primal problem reads

$$lexmin\ z = \begin{pmatrix} y_1^+ + y_2^- \\ y_1^+ + 0.5\ y_2^+ \end{pmatrix}$$

$$\begin{aligned} \text{s.t. } & x_1 + x_2 - y_1^+ + y_1^- = 10 \\ & x_1 - y_2^+ + y_2^- = 10 \\ & x_1, x_2, y_1^+, y_2^+, y_1^-, y_2^- \geq 0 \end{aligned}$$

and has the optimal solution $\hat{x}_1 = 10$, $\hat{x}_2 = \hat{y}_1^+ = \hat{y}_2^+ = \hat{y}_1^- = \hat{y}_2^- = 0$ with $\hat{z} = (0,0)^T$. To this primal problem Ignizio associates the following vector problem with $v, u_1, u_2 \in \mathbb{R}^2$:

$$\begin{aligned} max\ \ v = & -10\ u_1 - 10\ u_2 \\ \text{s.t. } & -u_1 - u_2 \leqq \begin{pmatrix} -1 \\ 0 \end{pmatrix} \\ & -u_1 \leqq 0 \\ & u_1 \leqq \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ & u_2 \leqq \begin{pmatrix} 1 \\ 0.5 \end{pmatrix} \end{aligned}$$

The optimal solution for this problem is $\hat{u}_1 = (1,0)^T$ and $\hat{u}_2 = 0$ with $\hat{v} = (-10, 0)^T$. Thus $\hat{z} \neq \hat{v}$. Moreover, the primal and "dual" problem are *not* based on the same set of information, the operator *max* does not express that v has to be maximized *in a lexicographic order* and the dual constraints, in general, do not hold as ordinary vector inequalities but as weak lexicographic inequalities.

While goal programming offers a great deal of flexibility in solving management problems its need for rather detailed *a priori* information on the decision maker's preferences implies a number of difficulties associated with its use. Goal programming requires the decision maker to specify goal levels, to partition the over- and under-achievement of goals into preemptive priority classes and to assign weights to the goal deviations within these classes. This fact calls

for information at the decision maker's disposal which allows him to reflect upon his a priori preference statements. In this context the duality theory provides a considerable potential in order to support the multiple goal decision maker in his decision process.

Let $(\hat{x},\hat{y}^+,\hat{y}^-)$ and $(\hat{U},\hat{V})$ be optimal solutions for (LGP) and (DLGP). Each coefficient $\hat{u}_{kj}$ of $\hat{U}$ indicates how much the current value of the k-th scalar-valued objective function changes when the j-th goal level g_j is augmented by one unit provided that the current basis B remains feasible. Accordingly, the coefficients $\hat{v}_{km}$ of $\hat{V}$ determine the change in the current value of z_k when the availability of the m-th resource is increased by one unit and the current basis B remains feasible. The sketched sensitivity analysis with respect to $g_j (j = 1,\ldots,J)$ and b_m $(m = 1,\ldots,M)$ can readily be extended to a parametric sensitivity analysis, as the proposed solution method immediately applies to parametric lexicographic goal programming. Note that if the value of at least one basic variable becomes negative a dual adapted simplex iteration has to be performed in order to regain primal feasibility while maintaining dual feasibility. Such an iteration is performed in the following way: First we select the pivot row (with the maximal violation of primal feasibility) in the current multiple objective simplex tableau. The pivot element has to be selected from the set of those coefficients in the pivot row which are negative. If all coefficients in the pivot row are non-negative, there exists no feasible solution. In each column where we have identified a negative coefficient in the pivot row we divide the respective vector of the reduced cost coefficients (which is lexicographically smaller than or equal to the zero vector) by the negative coefficient in the pivot row. Each of the obtained vectors is lexicographically greater than or equal to the zero vector. From the set of these vectors we select the lexicographically minimal one. The respective column becomes the pivot column and the associated nonbasic variable becomes the entering basic variable of the new basic solution which is dual feasible. If primal feasibility is not yet established the algorithm is continued with a further dual adapted iteration step.

The influence of changes in the weights of the deviational variables which are the coefficients of the matirces D^+ and D^- can also be

studied by means of the dual. The current basic solution remains optimal as long as the dual solution remains feasible, i.e. as long as the corresponding weak lexicographic inequality (10) holds. Note that when the variable(s) the weights of which are subject to a change are basic, the values of $\hat{U}$ and $\hat{V}$ also change. If the dual feasibility is violated the iteration steps of the presented lexicographic solution method are applied to reestablish dual feasibility. A parametric change of weights can be readily analysed by means of a parametric version of our algorithm which represents nothing else other than a lexicographic generalization of the corresponding parametric linear programming routine.

A modification of coefficients in the matrices A and C induces a change of coeffcients in the multiple objective simplex tableau. As long as coefficients associated to nonbasic variables are subject to change, the dual feasibility may be violated and readily reestablished by the lexicographic simplex method. However, if we modify coefficients which are associated to basic variables this involves a simultaneous consideration of primal and dual feasibility.

The addition of a new goal or a new decision variable involves an augmentation as well as a modification of coefficients of the current multiple objective simplex tableau which is actually performed in the same way as in linear programming if we take the multiplicity of objective functions into consideration. A thorough discussion of this point is e.g. found in Dinkelbach ([3] pp. 83 - 87). In the new multiple objective simplex tableau dual feasibility has to be checked and, if necessary, reestablished by means of the lexicographic simplex method.

Even though the decision maker will exercise considerable care in developing the priority structure, there still may be uncertainty regarding the assignment of priority levels in a manner that actually represents the preferences of the decision maker. Again, the effects of reordering these priorities can be investigated by means of the duality theory developed. If the decision maker wants to investigate a reordering of priority levels, this induces a rearranging of the objective functions in the multiple objective simplex tableau according to the new priority structure, and a check for dual feasibility: If all vectors of the reduced cost

coefficients are lexicographically smaller than or equal to the zero vector, the change in the priority assignment has no effect on the optimality property of the current basic solution. Otherwise, the effect of the changes in priority assignment can be investigated by re-establishing dual feasibility by means of the lexicographic simplex method.

REFERENCES

[1] CHARNES, A., and W.W. Cooper, *Management Models and Industrial Applications of Linear Programming*, vol. 1, Wiley, New York, 1961

[2] DANTZIG, G.B., *Linear Programming and Extensions*, University Press, Princeton, N.J., 1963

[3] DINKELBACH, W., *Sensitivitätsanalysen und parametrische Programmierung*, Springer-Verlag, New York, 1969

[4] IGNIZIO, J.P., *Goal Programming and Extensions*, Lexington Books, Heath, Lexington, Mass., 1976

[5] IJIRI, Y., *Management Goals and Accounting for Control*, North-Holland Publishing, Amsterdam, 1965

[6] ISERMANN, H., "Lösungsansätze zum Entscheidungsproblem des Satisfizierers bei mehrfacher Zielsetzung", in: P. GESSNER et al. (eds.), *Proceedings in Operations Research 3*, Physica Verlag, Würzburg, 1974, pp. 64 - 74

[7] ISERMANN, H., "Some Remarks on Optimizing and Satisficing in Multiple Criteria Decisions Problems", in: WHITE, D.J. and K.C. BOWEN, *The Role and Effectiveness of Theories of Decision in Practice*, Crane, Russak & Co., New York 1975, pp. 43 - 51

[8] KORNBLUTH, J.S.H., "A Survey of Goal Programming", *Omega*, vol 1, no. 2, 1973, pp. 193 - 205

[9] LEE, S.M., *Goal Programming for Decision Analysis*, Auerbach Publishers, Philadelphia, 1972

[10] LIN, W.T., "A Survey of Goal Programming Applications", *Omega*, vol 8, no. 1, 1980, pp. 115 - 117

[11] NIJKAMP, P. and J. SPRONK, Goal Programming for Decision Making, An Overview and a Discussion, Ricerca Operativa, Vol 9 (1979), pp. 3 - 49

[12] SPRONK, J., *Interactive Multiple Goal Programming Applications to Financial Planning*, Martinus Nijhoff Publishing, Boston, 1981.

RECENT RESULTS IN STOCHASTIC PROGRAMMING WITH MULTIPLE OBJECTIVE FUNCTIONS

I.M. Stancu-Minasian

Economic Cybernetics Department, Academy of Economic Studies, Bucharest, Romania

1. Introduction

In the classical problems of mathematical programming the coefficients are assumed to be exactly known. However this assumption is seldom satisfied in practice because these coefficients are either subject to errors of measurement or they vary with market conditions.

Many problems of mathematical programming related to several optimum criteria have a stochastic character because some or all of the data contained in such problems are random variables.

The problem under consideration is encountered in several contexts. For instance, H.S.Lau /25/ considers the newsboy problem and three objective functions: a) maximizing the expected profit, b) maximizing expected utility, c) maximizing the probability of achieving a budgeted profit (for more examples see also and /40/). For applications of probabilistic goal programming see /9/ and /21/.

We consider the following problem of stochastic linear programming with multiple objective functions:

$$(1)\quad \gamma_k(\omega)=\max_{x\in X(\omega)}\left\{ Z_k(x,\omega)=c_k'(\omega)x\right\} \qquad k=1,\dots,r$$

$$(2)\quad X(\omega)=\left\{x \mid A(\omega)x \leqslant b(\omega),\ x\gg 0\right\}, \qquad \omega\in\Omega$$

where $A(\omega)=(a_{ij}(\omega))$, $1\leqslant i\leqslant m$, $1\leqslant j\leqslant n$, $b(\omega)=(b_i(\omega))$, $1\leqslant l\leqslant m$ and $c_k(\omega)=(c_{kl}(\omega))$, $1\leqslant k\leqslant r$, $1\leqslant l\leqslant n$ are matrices whose elements are real-valued random variables defined on a probability space $\{\Omega,K,P\}$; Ω is a Borel set in R^s with $s=mn+m+rn$, K is the σ-algebra of all Borel subsets of Ω, P is a probability measure and x is an n-dimensional column vector.

We consider that $\gamma_k(\omega) = -\infty$, if $X(\omega) = \emptyset$ ($\emptyset$ is the empty set). If $P\{-\infty < \gamma_k(\omega) < \infty\} = 1$, then $\gamma_k(\omega)$ is a random variable. In this case we say that the stochastic linear program (1)-(2) has an optimal value. In Ref. /8/ necessary and sufficient conditions are provided for the existence of the optimal value of an stochastic linear program.

Because the optimal values $\gamma_k(\omega)$ $(k=1,\dots,r)$ are random variables, they can not be known in advance. However, we wish to find the probability distribution function and (or) some moments (say the expectation and variance) of the random vector

$$\gamma(\omega) = (\gamma_1(\omega), \gamma_2(\omega), \dots, \gamma_r(\omega))$$

subject to an arriori probability distribution of the state of nature i.e. the triple $\{A(\omega), b(\omega), (c_1(\omega), \dots, c_r(\omega))\}$.

In this paper we outline the recent results and developments of multiple criteria stochastic programming. A comprehensive bibliography is given.

Although the domain of stochastic programming and that of mathematical programming with multiple objective functions are fruitful nowadays (for edification see the research bibliographies /32/, /35/), however the field of stochastic programming with multiple objective functions was partly neglected in the European and American literature.

The methods discussed in the present paper are generalizations of the methods used for solving the deterministic programming problems. Due to this fact, in the following section, we shall first recall some ways of solving deterministic mathematical programming problems with several objective functions.

2. Points of View in Solving the Deterministic Mathematical Programming Problem with Several Objective Functions

Let us consider the following mathematical programming problem with several objective functions.

$$(3) \quad \max_{x \in X} \{Z_k(x) = c_k^t x\} \qquad k=1,\dots,r$$

where $X = \{x \mid x \in R^n,\ Ax \leqslant b,\ x \geqslant 0\}$, A, b, c_k, x are matrices of size $m \times n$, $m \times 1$, $1 \times n$, $n \times 1$, respectively.

This problem has not a sole solution, various authors ascribe different meanings to the vector $x^* \in R^n$ which should be "as a good as possible" from the viewpoint of all the objective functions $Z_k(x)$.

The main approaches to define the vector x^* are the following:

1°. x^* optimizes a synthesis-function of the r efficiency functions

$$h(Z)=h(Z_1,\dots,Z_r)$$

The function h may be defined in various manners /40/:

a) $h(Z_1,\dots,Z_r)=\min_{i=1,\dots,r}(\max)\{Z_i(x)\}$

b) $h(Z_1,Z_2)=Z_1(x)/Z_2(x)$

c) $h(Z_1,\dots,Z_r)=\sum_{i=1}^{r}\alpha_i\left[Z_i(x)\right]^{\beta_i}$ $\qquad \alpha_i \geqslant 0,\ \beta_i > 0$ etc.

2°. x^* is the vector minimizing function

$$h(x)=h(\Psi_1(x-X_1),\dots,\Psi_r(x-X_r))$$

where X_j optimizes function Z_j and $\Psi_j(x-X_j)$ is a function of distance between x and X_j.

3°. x^* is a feasible solution which is obtained by a seeking method according to certain criteria. Such methods are the POP Method (Progressive Orientation Procedure) /4/ and the STEM Method /5/.

4°. x^* is an optimal solution obtained by ordering criteria. We solve r problems, each time restricting the field X by turning into constraints the optimal solutions obtained by solving a certain problem with a single function.

5°. x^* is the vector belonging to a set of efficient points(also called nondominated solutions or Pareto-optimal solutions)that is defined like this: $x^*\in X$ is efficient if and only if there exist no $x'\in X$ such that $Z_h(x^*)\leqslant Z_h(x')$ for h=1,...,r and h_o exist so that $Z_{h_o}(x^*)<Z_{h_o}(x')$ (assuming that all the functions are maximand functions).

3. Stochastic Chebyshev Problem

In /38/ Stancu-Minasian considered the following multiple criteria stochastic programming problem.

$$(4)\qquad \min_{x\in X}\left\{Z_k(x,\omega)=c_k'(\omega)x\right\} \qquad k=1,\dots,r,$$

$$(5)\qquad X=\left\{x \mid Ax\leqslant b,\ x\geqslant 0\right\}$$

where the elements of the vectors c_k are stochastic variables with known (joint) probability distribution.

Taking into account 1°-a) this problem can be reduced to one with a single objective function as follows: instead of r objective functions $Z_k(x,\omega)$we consider a single objective function

$$\gamma(x,\omega)=\max_{1\leqslant k\leqslant r}\left\{Z_k(x,\omega)=c_k'(\omega)x\right\}$$

and minimize it.

In this manner, we obtain the so-called Chebyshev problem

$$(6)\quad \gamma(\omega) = \min_{x \in X} \max_{1 \leq k \leq r} \left\{ Z_k(x,\omega) = c_k'(\omega)\, x \right\}$$

In /38/ is determined the probability distribution function of $\gamma(\omega)$, when the cost coefficients of $Z_k(x,\omega)$ are linear functions of the same random variable $t(\omega)$, that is they are under the form

$$Z_k(x,\omega) = (c_k + t(\omega)\, d_k)'\, x$$

where c_k and d_k are n-dimensional vectors and $t(\omega)$ is a random variable having a distribution function $T(z)$ continuous and strictly increasing.

4. Stochastic Fractional Programming Problem

In the particular case of two objective functions we can draw up a function of the type 1°-b which leads us to a stochastic fractional programming problem

$$(7)\quad \gamma(\omega) = \operatorname*{optimum}_{x \in X} \frac{Z_1(x,\omega)}{Z_2(x,\omega)}$$

$$(8)\quad X = \{ x \mid Ax = b,\ x \geq 0 \}$$

It is considered that Z_1 and Z_2 are linear functions of the same random variable $t(\omega)$ having a distribution function $T(z)$ continuous and strictly increasing i.e.,

$$Z_1(x,\omega) = (c' + t(\omega)\, c_1')\, x,$$

$$Z_2(x,\omega) = (d' + t(\omega)\, d_1')\, x$$

For the problem (7)-(8) we make the following assumptions:

a) The set X is nonempty and bounded,

b) The denominator of the objective function preserves the same sign (let us assume it to be positive) on X, consequently $P\{\omega \mid (d' + t(\omega)\, d_1')\, x > 0\} = 1$,

c) All basic solutions are nondegenerate.

In /37/, Stancu-Minasian gives a representation of the probability distribution of the optimal value $\gamma(\omega)$ by means of the characteristic values of the corresponding parametric programs.

In the remainder of this section, we shall recall other two problems for stochastic fractional programming problem /33/, /36/.

a) The minimum risk solution with level k, i.e. the optimal solution of the program:

$$(9)\quad \max_{x \in X} P\left\{ \omega \,\middle|\, \frac{c'(\omega)\, x}{d'\, x} \leq k \right\}$$

where $c(\omega)$ has a normal non-degenerated distribution, with mean value $\bar{c}$ and covariance matrix V.

Theorem (Stancu-Minasian /33/, /36/) The minimum risk problem (9) is equivalent to the following non-linear programming problem

$$\max_{x \in X} \frac{k.d'x-\bar{c}'x}{(x'Vx)^{1/2}}$$

b) Kataoka's problem of the maximum level, i.e.

$$(10) \quad \max\left\{ k \mid P\left\{\omega \left| \frac{c'(\omega)x}{d'x} \leq k \right.\right\} = \alpha, x \in X \right\}$$

Theorem (Stancu-Minasian /33/, /36/) Kataoka's problem of the maximum level (10) is equivalent to the following non-linear programming problem

$$(11) \quad \max_{x \in X} \frac{\bar{c}'x - q(x'V x)^{1/2}}{d'x}$$

where $\Phi^{-1}(\alpha)=-q$ and $\Phi(.)$ is Laplace's function .

It should be noted that in some conditions the problem (11) has a finite optimal solution and that the objective functions is explicitely quasi-concave.

5. Stochastic Goal Programming

The goal programming problem is that of finding

$$\min_{x} \; d(x,\bar{x}) \text{ subject to } g_i(x,\bar{x}) \geq 0 \qquad i = 1,\ldots,n$$

where $\bar{x}$ is a given goal vector, d is a distance between x and $\bar{x}$, and g_i are given functions.

Because $d(x,x)=\|x-\bar{x}\|$ we shall further use instead of the distance between x and $\bar{x}$, the norm of the difference of vectors x and $\bar{x}$, as a measure of approach between x and $\bar{x}$. The best known is norm L_p or the Hölder norm $\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$, $p \geq 1$. In particular cases the following norms are obtained:

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| ; \quad \|x\|_2 = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2} \text{ (Euclidian distance)};$$

$$\|x\|_\infty = \max_i \{|x_i|\}$$

The aim of this section is to analyse four variants of goal-programming in the stochastic case in order to obtain some deterministic mathematical programming problems.

I. B. Contini /12/ considered that in relation between the objectives $F=(F_1,\dots,F_r)$ and variable $x\in X=\{x \mid Ax=b, x\geqslant 0\}$ there exists a vector $u=(u_1,\dots,u_r)$ of random variables i.e.:

$$F_k(x)=c_{k1}x_1+\dots+c_{kn}x_n+u_k=c_k'x+u_k \qquad k=1,\dots,r \tag{12}$$

If u is a normally distributed random vector, with mean vector 0 and covariance matrix V, then for each $x\in X$ $F=(F_1,\dots,F_r)$ is a random vector , with means $(c_1'x,\dots,c_r'x)$ and covariance matrix V; thus we have

$$f(y_1,\dots,y_r)=(2\pi)^{-r/2}\,|V|^{-1/2}\,e^{-Q/2}$$

Here Q is a quadratic form defined as

$$Q=(F-Cx)'V^{-1}(F-Cx) \qquad (C=(c_{ij})) \tag{13}$$

Let $\bar F=(\bar F_1,\dots,\bar F_r)$ be the levels to be attained by the objectives. Due to the random vector u, we have not $\bar F=Cx$. In this conditions we choose Y^* a region in R^r such that $\bar F=(\bar F_1,\dots,\bar F_r)\in Y^*$ and the problem consists in choosing of $x\in X$ for which F(x) defined by (12) has the highest probability of falling in Y^*.

We obtain the following model /12/

$$\max_{x\in X} \; P(F(x)\in Y^*) \tag{14}$$

Usually, due to the normality assumption of u, Y^* is taken an ellipsoid in R^r, centered at $\bar F$, of the form:

$$Y^*=\{\, y=(y_1,\dots,y_r) \mid Q^*=(y-\bar F)'\,V^{-1}(y-\bar F)\leqslant c^2\,\}$$

If V is nonsingular, then Q defined by (13) has a χ^2 distribution with r degree of freedom. According to this fact, Y^* can be interpreted as a confidence region for F(x) at level α, i.e., if for x^0 we have $EF(x^0)=Cx^0=\bar F$ (where E denotes expected value) then $P(F(x)\in Y^*)=\alpha$.

The model (14) is equivalent /12/ with the following quadratic programming model

$$\min_{x\in X} \; R(x) = (\bar F-Cx)'\,V^{-1}(\bar F-Cx)$$

With notations $k=\bar F'V^{-1}\bar F$, $B=C'V^{-1}C$, $p=-C'V^{-1}\bar F$ we have $R(x)=k+x'Bx+2p'x$

Theorem (B.Contini /12/). The optimal solution of the problem (14) is obtained by solving the following quadratic programming problem

$$\min_{x\in X} \; (k+x'Bx+2p'x)$$

General algorithms to solve such problems are available (see, for example, /41/).

2. We assume that in the relation $\bar{F}=Cx$ the elements of C are random variables. Then for each $x \in X$ there will be more or less deviations of Cx from $\bar{F}$(we denote them $d^{+}(x)$ and $d_k^{-}(x)$)and a convenient criterion is minimization of expected value of the norm $\|\bar{F}-Cx\|$.

The following model is obtained /10/

$$(15) \qquad \min_{x \in X} \; E(\|F-Cx\|_p)$$

For p=1 the model (15) becomes

$$\min_{x \in X} \; E\Big(\sum_{k=1}^{r} (d_k^{+}(x)+d_k^{-}(x)) \Big)$$

subject to

$$F_k(x)+d_k^{-}(x)-d_k^{+}(x)=\bar{F}_k \qquad k=1,\dots,r$$

$$Ax = b$$

$$x,\ d_k^{+},\ d_k^{-} \geqslant 0$$

For p=2 the model (15) becomes

$$(16) \qquad \min_{x \in X} \; E\Big(\sum_{i=1}^{r} |\bar{F}_i-C_i x|^2 \Big)$$

where $C_i=(c_{i1},\dots,c_{in})$

In order to find the deterministic equivalent of problem (16) we assume that the elements of C are independent random variables with means $\bar{c}_{ij}$ and variances σ_{ij}^2, i.e.,

$$\bar{c}_{ij} = Ec_{ij} \quad , \quad \sigma_{ij}^2 = D^2(c_{ij})$$

Immediately, it follows that for random variable

$$z_i=\bar{F}_i-C_i x=\bar{F}_i-\sum_{j=1}^{n} c_{ij}x_j$$

we have

$$Ez_i=\bar{z}_i=\bar{F}_i-\sum_{j=1}^{n} \bar{c}_{ij}x_j \qquad \text{and} \qquad \sigma_{z_i}^2 = \sum_{j=1}^{n} \sigma_{ij}^2 x_j^2$$

Hence

$$E\Big(\sum_{i=1}^{r} |F_i-C_i x|^2 \Big)=E\Big(\sum_{i=1}^{r} z_i^2\Big)= \sum_{i=1}^{r} E(z_i^2)= \sum_{i=1}^{r} (\bar{z}_i^2 + \sigma_{z_i}^2) =$$

$$= \sum_{i=1}^{r} (\bar{F}_i- \sum_{i=1}^{n} \bar{c}_{ij}x_j)^2 + \sum_{i=1}^{r} \sum_{j=1}^{n} \sigma_{ij}^2 x_i^2 x_j^2 = \sum_{i=1}^{r} (\bar{F}_i- \sum_{j=1}^{n} \bar{c}_{ij}x_j)^2 +$$

$$\sum_{j=1}^{n} \Big(\sum_{i=1}^{r} \sigma_{ij} \Big)^2 x_j^2$$

Theorem (Chobot M. /10/) The optimal solution of the problem (16) is derived by solving thefollowing quadratic programming problem

$$\min_{x \in X} \left[\sum_{i=1}^{r} \left(\bar{F}_i - \sum_{j=1}^{n} \bar{c}_{ij} x_j \right)^2 + \sum_{j=1}^{n} \left(\sum_{j=1}^{r} \sigma_{ij} \right)^2 x_j^2 \right]$$

3. Now, we consider the problem from pct.1) in another variant. We assume that the elements of C are random variables. It is choosen the domain Y^*

$$Y^* = \left\{ (y_1, \ldots, y_r) \in R^r \mid \xi_{1i} \leq y_i \leq \xi_{2i},\ i=1,\ldots,r \right\}$$

and two bounds $\alpha_1 = (\alpha_{1i})$, $\alpha_2 = (\alpha_{2i})$ $i=1,\ldots,r$ for the probability that $F(x) \in Y^*$; the problem consists in finding how much we can "diminich" Y^* such that the probability that $F(x) \in Y^*$ to be placed between the two bounds.

The obtained problem takes the form

$$\min \| \xi_2 - \xi_1 \|$$

subject to

$$P\{ C_i x \geq \bar{F}_i - \xi_{1i} \} \geq \alpha_{1i} \qquad i=1,\ldots,r$$

$$P\{ C_i x \geq \bar{F}_i + \xi_{2i} \} \geq \alpha_{2i} \qquad i=1,\ldots,r$$

$$\xi_{1i}, \xi_{2i} \geq 0$$

This problem is equivalent /10/ with the following problem

$$\min \| \xi_2 - \xi_1 \|$$

subject to

$$\bar{F}_i - \xi_{1i} - \sum_{j=1}^{n} \bar{c}_{ij} x_j \leq \beta^{-1}(1-\alpha_{1i}) \left(\sum_{j=1}^{n} \sigma_{ij}^2 x_j^2 \right)^{1/2} \qquad i=1,\ldots,r$$

$$\bar{F}_i + \xi_{2i} - \sum_{j=1}^{n} \bar{c}_{ij} x_j \geq \beta^{-1}(\alpha_{2i}) \left(\sum_{j=1}^{n} \sigma_{ij}^2 x_j^2 \right)^{1/2} \qquad i=1,\ldots,r$$

$$\xi_{1i}, \xi_{2i} \geq 0$$

4. In what follows we assume that the elements of the vector $\bar{F}$ are independent random variables with known distribution $F_i(.)$ $(i=1,\ldots,r)$. We can consider the problem /6/:

$$\min \| \xi \|$$

subject to

$$P\{ C_i x \geq \bar{F}_i - \xi_i \} \geq \alpha_{1i} \qquad i=1,\ldots,r \tag{17}$$

$$P\{ C_i x \leq \bar{F}_i + \xi_i \} \geq \alpha_{2i} \qquad i=1,\ldots,r$$

where $\xi = (\xi_1, \ldots, \xi_r)$, $\alpha_1 = (\alpha_{1i})$, $\alpha_2 = (\alpha_{2i})$, $i=1,\ldots,r$

For any i and $0 < \theta \leq 1$, let

$$F_i^{-1}(\theta) = \inf\{\eta : F_i(\eta) \geq \theta\} \quad , \quad \hat{F}_i^{-1}(\theta) = \sup\{\eta : F_i(\eta) \leq \theta\}$$

and the vectors $F^{-1}(\alpha_1) = (F_i^{-1}(\alpha_{1i}))$, $\hat{F}^{-1}(e-\alpha_2) = (\hat{F}_i^{-1}(1-\alpha_{2i}))$, $i=1,\dots,r$ where e is the vector of ones.

With this notations, the deterministic equivalent of the above problem is

$$\min \|\xi\|$$

(17′) subject to

$$(*) \qquad C_i x + \xi_i \geq F_i^{-1}(\alpha_{1i})$$

$$(**) \qquad C_i x - \xi_i \leq \hat{F}_i^{-1}(e-\alpha_{2i}) \qquad i=1,\dots,r$$

$$\xi_i \geq 0$$

Substracting $(**)$ from $(*)$, it results

$$(***) \qquad \xi \geq \frac{1}{2}\{F^{-1}(\alpha_1) - \hat{F}^{-1}(e-\alpha_2)\} = \xi^0$$

If $\xi^0 \geq 0$ then the constraint $\xi_i \geq 0$ is redundant. We have $\xi^0 \geq 0$ if $\alpha_1 > e-\alpha_2$ (i.e., $\alpha_{1i} + \alpha_{2i} > 1$ $i=1,\dots,m$) This condition is always satisfied because α_{1i} and α_{2i} are normally close to 1 . From $(***)$ it results $\xi = \xi^0 + \delta$ $(\delta \geq 0)$ which, when substituted in $(*)$ and $(**)$ results

$$-\delta \leq Ax - \delta^0 \leq \delta$$

where $\delta^0 = 1/2\{F^{-1}(\alpha_1) + \hat{F}^{-1}(e-\alpha_2)\}$

Theorem /6/. If $\alpha_1 > e-\alpha_2$ and if $\bar{F}_i$ are independent random variables , then the problem (17′) is equivalent to the following deterministic problem

minimize $\|\xi^0 + \delta\|$ subject to $-\delta \leq Ax - \delta^0 \leq \delta$ and $\delta \geq 0$

6. Group Decision Making in Stochastic Programming

We recall again the problem (4)-(5) and consider a weighted sum of the r objective functions.

$$F^* = \sum_{k=1}^{r} \alpha_k Z_k = \sum_{k=1}^{r} \alpha_k \sum_{j=1}^{n} c_{kj}(\omega) x_j = \sum_{j=1}^{n} \left(\sum_{k=1}^{r} \alpha_k c_{kj}(\omega) \right) x_j =$$

$$= \sum_{j=1}^{n} c_j(\alpha, \omega) x_j = c'(\alpha, \omega) x, \qquad \alpha = (\alpha_1, \dots, \alpha_r)$$

The weights $\alpha_k (k=1,\dots,r)$ reflect the importance attributed by the decision maker to the objective functions. Usually, the weights are normed by the conditions $0 \leq \alpha_k \leq 1$, $k=1,\dots,r$, $\sum_{k=1}^{r} \alpha_k = 1$.

Thus, we obtain the following problem

$$\min_{x \in X} \left\{ \sum_{k=1}^{r} \alpha_k Z_k(x,\omega) = \sum_{k=1}^{r} \alpha_k c_k'(\omega) x \right\}$$

From a practical point of view, we can not add the objective functions because they are incomensurable. However, functions $Z_k(x,\omega)$ can be turned into utility functions in von Neuman-Norgenstern's sense /28/. Thus, each function $Z_k(x,\omega)$ will undergo a transformation of the form

$$Z_k'(x,\omega) = \alpha_k Z_k(x,\omega) + \beta_k$$

where $X_k = \underset{x \in X, \omega \in \Omega}{\text{infimum}} Z_k(x,\omega)$, $Y_k = \underset{x \in X, \omega \in \Omega}{\text{supremum}} Z_k(x,\omega)$

and α_k and β_k are solutions of:

$$\alpha_k X_k + \beta_k = 1$$

$$\alpha_k Y_k + \beta_k = 0$$

The r objective functions $Z_k'(x,\omega)$ are summed up and we maximize the synthesis function

$$\max_{x \in X} \left\{ Z^*(x,\omega) = \sum_{k=1}^{r} \alpha_k Z_k'(x,\omega) \right\}$$

Another aspect of the problems is the follows . In a group decision the weights α_k are random variables because the members of the group will assign several weights to objective functions. Without loss of generality, we assume that α_k are random variables and c_k are deterministic. If the group is small enough, each member may effectively advance his preferences and then α_k represents the expected values of the weights assigned by each member. If the group is very large, as in the case of determining a welfare function, it is not possible to take into account the corresponding weight vectors reflecting the preferences of all members of the group. Thus, α_k can be considered as random variables.

Let be $c(\alpha) = \sum_{k=1}^{r} \alpha_k c_k$ and $F_\alpha(.)$ the probability distribution function of the random vector α .

B.Bereanu /7/ considered two problems:

a) Finding a vector x, before consulting the group members, which maximizes the probability that the value of the function shall exceed a certain level u, i.e.:

(18) $\max_{x \in X} P\{\omega \mid c'(\alpha)x \geq u\}$

The optimal vector represents minimum risk solution for multiple criteria problem.

b) Finding the probability distribution function of the random variable $\eta(\omega)$ defined by

(19) $\eta(\omega) = \max_{x \in X} c'(\alpha)x$

subject to $F_\alpha(.)$ and the probabilities of various basic feasible solutions of X to be optimal.

Definitions. (B.Bereanu /7/) If the minimum risk solution in (18) does not depend of $F_\alpha(.)$ we call it distribution free minimum risk solution (d.f.m.r.s) of the multicriteria problem. If it does not depend on u either, it is called free minimum risk solution (f.m.r.s).

A stable optimal basis (optimal with probability 1) independent of the $F_\alpha(.)$ is called distribution-free optimal basis. It provides a distribution-free optimal solution (d.f.o.s).

Theorem (Bereanu /7/) Let C be a nonempty convex set in R^n, $\mathcal{P}_C$ be the set of all probability measures on R^n such that $(P \in \mathcal{P}_C) \Rightarrow (P(C)=1)$ and $\mathcal{C}(C)$ be the set of corresponding random vectors. The following statements are equivalent with respect to $\mathcal{C}(C)$:

(i) $x^* \in X$ is a f.m.r.s of $\max_{x \in X} P\{\omega \mid c'(\alpha)x \geq u\}$

(ii) $x^* \in X$ is a d.f.o.s. of $\eta(\omega) = \max_{x \in X} c'(\alpha)x$

Also, we can use the simulation-method to establish the weights α_k /11/. It can be imagined that α_k are generated by a random vector of the Dirichlet type, having the probability density.

$$g(\alpha_1, \ldots, \alpha_r) = \frac{\Gamma(\nu_1 + \ldots + \nu_{r+1})}{\Gamma(\nu_1) \ldots \Gamma(\nu_{r+1})} \alpha_1^{\nu_1 - 1} \ldots \alpha_r^{\nu_r - 1} (1 - \alpha_1 \ldots - \alpha_r)^{\nu_{r+1} - 1}$$

if $(\alpha_1, \ldots, \alpha_r) \in S_r = \left\{ (\alpha_1, \ldots, \alpha_r) \mid \xi_i \geq 0,\ 1 \leq i \leq r;\ \sum_{i=1}^{r} \alpha_i \leq 1 \right\}$

and

$$g(\alpha_1, \ldots, \alpha_r) = 0 \text{ if } (\alpha_1, \ldots, \alpha_r) \notin S_r$$

We assume that we known the modes $M(\alpha_i)$ $(i=1, \ldots, r)$ (given by the user).

It can be shown that

$$M(\alpha_1) = \frac{\nu_1}{\nu_1 + \dots + \nu_{m+1}}, \quad \dots \quad M(\alpha_r) = \frac{\nu_r}{\nu_1 + \dots + \nu_{r+1}}$$

With help of ν_i it generates the Gamma variables δ_i having the same parameter $\lambda = 1$.

If $\delta_i (i=1,\dots,m+1)$ are independent random variables of the Gamma type with parameters λ and respectively $\nu_i (i=1,\dots,m+1)$ then, the vector $\alpha = (\alpha_1, \dots, \alpha_r)$ with the components

$$\alpha_1 = \frac{\delta_1}{\delta_1 + \dots + \delta_{r+1}}, \quad \dots, \quad \alpha_r = \frac{\delta_r}{\delta_1 + \dots + \delta_{r+1}}$$

is a random vector with a Dirichlet distribution. This vector is taken as the weights vector for solving the multiple criteria problem.

7. The Protrade Method

For solving the stochastic programming problem with multiple objective functions, L.Goicoechea, L.Duckstein and R.Bulfin /20/ proposed a method, labeled Protrade, which can be considered to be the analogue in the stochastic case of the POP and STEM Methods /4/,/5/.

It is an interactive method, the decision maker having the possibility to conduct the process of choosing efficient solutions by modification initial conditions of the problem, function of partial results obtained. Thus, this method alternates the phase of computation with the phase of decision, the decision maker trading level of achievement for each objective function directly against probability of achieving that level. For other interactive methods used in multiple criteria stochastic programming, see /19/ and /20/. Let us consider the following problem

$$\max \{ Z(x) = (Z_1'(x), \dots, Z_p'(x)) \}$$

subject to

$$x \in D_o = \{ x \mid x \in R^n, \ g_i(x) \leq 0, \ x \geq 0, \ i=1,\dots,r \}$$

where

$$Z_i' = \sum_{j=1}^{n} c_{ij} x_j, \quad Z_i(x) = E(Z_i'(x)), \quad c_{ij} \sim N\left[E(c_{ij}), \ Var(c_{ij}) \right]$$

and the functions $g_i(x)$ are differentiable and convex. The Protrade Method consists in the following steps:

Step 1. For each function $Z_k'(k=1,\dots,r)$ determine the values extreme on D_o, i.e.

$$Z_{k_{max}} = Z_k(x^{i*}) = \max_{x \in D_o} Z_k(x)$$

$$Z_{k_{min}} = Z_k(x^{i**}) = \min_{x \in D_o} Z_k(x)$$

and let $U_1 = (Z_1(x^{1*}), \ldots, Z_r(x^{r*}))$

Step 2. Define an initial surrogate objective function

$$F(x) = \sum_{k=1}^{r} G_k(x)$$

where

$$G_k(x) = \frac{(Z_k - Z_{k\ min})}{Z_{k\ max} - Z_{k\ min}}$$

Step 3. Maximize $F(x)$, $x \in D_o$ and with the optimal solution X^1 generate an initial, non-inferior goal vector $G_1 = (G_1(X^1), \ldots, G_r(X^1))$

Step 4. Estimate a multi-attribute utility function $u(G)$ reflecting the decision maker's preferences (for instance, the multiplicative form /15/, /24/ $1+kU(G) = \prod_{i=1}^{n} [1+kk_i u_i(G_i)]$, k and k_i are parameters) and define a new surrogate objective function

$$S_1(x) = \sum_{i=1}^{r} w_i G_i(x),$$

where $w_i = 1 + \frac{r}{G_i(x^1)} \left[\frac{\partial U(G)}{\partial G_i} \right]_{G_1}$

and r stands the step size required to yield a new goal vector in the direction of a desired increment $\Delta U(G)$.

Step 5. Maximize $S_1(x)$, $x \in D_o$ and with the optimal solution X^2 generate vectors

$$G_2 = (G_1(X^2), \ldots, G_r(X^2)) \text{ and } U_2 = (Z_1(X^2), \ldots, Z_r(X^2))$$

Then, the vector V_1, which expresses trade-off between goal value and its probability of achievement, is generated

$$V_1 = \begin{pmatrix} G_1(X^2) & G_2(X^2) & \ldots & G_r(X^2) \\ 1-\alpha_1 & 1-\alpha_2 & & 1-\alpha_r \end{pmatrix}$$

with $1-\alpha_i$ choosen such that

$$P(Z_i'(x) \geq Z_i(X^2)) \geq 1-\alpha_i$$

Step 6. Now, the decision maker aeswer tc the following question: "Are all the $Z_i(X^2)$ values satisfactory?" If Yes, U_2 represents a desired solution. Otherwise, go to the next step.

Step 7. Choose the function $Z_k(x)$ with the least satisfactory pair $\begin{pmatrix} G_k(X^2) \\ 1-\alpha_k \end{pmatrix}$ and find $e_k \in R^+$, $\alpha_k^{\circ} \in [0,1]$ such that

$P(Z_k'(x) \geq e_k) \geq 1-\alpha_k^\circ$

Step 8. Redefine the solution space as follows

$$D_1 = \{ x \mid x \in D_o,\ P(Z_k'(x) \geq e_k) \geq 1-\alpha_k^\circ \}$$

Step 9. Define the new surrogate objective function

$$S_2(x) = \sum_{i \neq k}^{r} w_i G_i(x)$$

and call to Step 5 for maximize $S_2(x)$ on D_1 etc. Finally, a satisfactory vector V_2 is achieved

$$V_2 = \begin{pmatrix} e_1 & e_2 & \cdots & e_r \\ 1-\alpha_1^\circ & 1-\alpha_2^\circ & \cdots & 1-\alpha_r^\circ \end{pmatrix}$$

8. Probabilities of Optima in Multi-Objective Linear Programmes

Let us now restrict ourselves to find the efficient solutions for the deterministic problem.

$$\max_{x \in X} \{ Z(x) = (Z_1(x) = c_1'x, \ldots, Z_r(x) = c_r'x) \}$$

where $X = \{ x \in R^n \mid Ax = b;\ x \geq 0 \}$

As it is known /23/, x is an efficient point of the linear vector maximization problem P: max Cx subject to $x \in X$ (C is a r x n matrix) if and only if there is some $\lambda > 0$ in the set $\Lambda = \{ \lambda = (\lambda_1, \ldots, \lambda_r) \in R^r : 0 < \lambda_i < 1, \sum_{i=1}^{r} \lambda_i = 1 \}$ such that x is an optimal solution of the associated linear problem whose objective function is a weighted combination of the r objective functions of the vector maximization problem $P_\lambda : \max_{x \in X} \lambda' Cx$

The parameter space Λ can be decomposed into a finite number of polyhedra $\Lambda(x_i)$ (which correspond to different nondominated extreme points of the feasible set of P) and open half-spaces $\tilde{\Lambda}_j$. These polyhedra are connected by their boundary hyperplanes. The polyhedral set $\Lambda(x_i)$ consists of all such that x_i is an optimal solution of P_λ, and the problem P_λ has no finite optimal solution for all $\lambda \in \tilde{\Lambda}_j$.

In connection with the structure of the space Λ of λ-values, in /29/, S.S.Sengupta, M.L.Podrebarac and T.D.H.Fernando put the following question "Given a basic feasible solution, what is the probability that it is also optimal? Or, equivalently, given that a solution is basic optimal, what is the probability that it may be associated with an assigned region in the space Λ?"

Of course, the notion of "probability" is the one of geometric probability, i.e., in terms of relative lenghts, volumes etc.

Because Λ is a subset of n-dimensional Euclidian space R^r we can define a σ-ring F of Lebesque- measurable subset of Λ. Let $\{\Lambda, F, m(.)\}$ be a measure space, where $m(\cdot)$ is the Lebesque measure.

We construct another measure $\mu(.) = m(.)/m(\Lambda)$ which give rise to a new measure space $\{\Lambda, F, \mu(.)\}$. This new measure space can serve as a probability space because the measure $\mu(.)$ is totally finite and $\mu(\Lambda) = 1$.

Theorem (/29/) Let Λ_G be an arbitrarily assigned Lebesque measurable subset of Λ, and let x_B be an arbitrary basic feasible solution of P_λ. Then the following exist, namely,

a) $P(x_B$ is optimal)
b) $P(x_B$ is optimal $|\lambda \in \Lambda_G)$
c) $P(\lambda \in \Lambda_G \mid x_B$ is optimal)

9. Multiple Minimum Risk Solution in Stochastic Programming

In several papers /31/, /33/, /34/, /39/, /40/ Stancu-Minasian considered a more general case of the minimum risk problem, i.e. the case when the probabilities that the values of r objective functions exceed some levels of performance are maximized

$$\max_{x \in X} \; P\{\omega \mid c_1'(\omega) x \geq u_1\}$$

. . . . . .

$$\max_{x \in X} \; P\{\omega \mid c_r'(\omega) x \geq u_r\}$$

with X defined as in (8).

For solving such problem, Stancu-Minasian proposes a method which consists in the sequential solving of some minimum risk problem (containing also quadratic constraints). Related to this method, the concept of the multiple minimum risk solution is introduced and a relaxation-type method for its obtaining is given.

We can introduce other models in a similar manner to that suggested in Ref./21/, /22/. Thus, a possible objective is to maximize the joint probability /22/

$$P\{c_1'x \geq u_1, \ldots, c_r'(\omega) x \geq u_2\}$$

or to maximize

$$\sum_{k=1}^{r} w_j P(c_k'(\omega) \geq u_k)$$

where w_j are weights associated with each objective function.

REFERENCES

/1/ ARMSTRONG R.D.and BALINFTY J.L.: A Chance-constrained Multiple Choice Programming Algorithm. Operations Res.23, 494-510, 1975

/2/.AVENHAUS R., BEEDGEN R., CHERNAVSKY S. and SCHRATTENHOLZER L. (with Annex A by Alois Hölzl): Handling Uncertainities in Linear Programming Models. Working Paper WP-80-170, November 1980, International Institute for Applied Systems Analysis, Austria

/3/ BARON D.P.: Stochastic Programming and Risk Aversion.in Multiple Criteria Decision Making, J.L.Cochrane and M.Zeleny (eds.), University of South Carolina Press, Columbia, S.C., 1974, 124-138

/4/ BENAYOUN R., TERGNY J. and KEUNEMAN D.: Mathematical programming with multi-objective functions: a solution by P.O.P. (Progressive Orientation Procedure). Metra 9(2), 279-299, 1970

/5/ BENAYOUN R., de MONTGOLFIER J., TERGNY J. and LARICHEV O.I.: Linear programming with multiple objective functions. Step method (STEM). Math.Programming 1(3), 366-375, 1971

/6/ BEN-ISRAEL A., CHARNES A. and KIRBY M.J.L.: On Stochastic Linear Approximation Problems. 18(3), 555-558, 1970

/7/ BEREANU B.: Large Group Decision Making with Multiple Criteria. Multiple criteria decision making. Proceedings of a conference held at Jouy-en-Josas 1975 Thiriez H. and Zionts S.(eds.). Lect.Notes in Economics and Math.Systems.Vol.130, Springer, Berlin-Heidelberg-New York, 1976, 87-101

/8/ BEREANU B.: The Continuity of the Optimum in Parametric Programming and Applications to Stochastic Programming.J.Optim.Theory Appl.18(3), 319-333, 1976

/9/ CHARNES A. and STEDRY A.C.: Search-Theoretic Models of Organization Control by Budgeted Multiple Goals. Management Sci. 12(5), 457-481, 1966

/10/CHOBOT M.: Stochastický pristup k cílovému programovaniu.Ekon. Mat.Obzor 3, 305-320, 1973

/11/CIOBANU GH.: Applying Digital Simulation in Solving Problems of Linear Programming with Several Objective Functions.Econom. Comp.Econom.Cybernet.Stud.Res.1, 63-68, 1976

/12/CONTINI B.: A Stochastic Approach to Goal Programming. Operations Res.16(3), 576-586, 1968

/13/DÜRR W. : Stochastische Programmierungsmodelle als Vektormaximumprobleme.in Proceedings Operations Research, Würzburg-Wien 1, 189-199, 1972

/14/FANDEL G. and JOCHEN W.: Das Vektormaximumproblem bei Sicherheit und Unsicherheit. Tech.Report, Institut für Gesellschafts und Wirtschaftswissenschaften der Universität Bonn, 1974

/15/FISHBURN P.C.: Utility Theory for Decision Making, Wiley, New York, 1970

/16/GEOFFRION A.M.: On the Relevance of the Vector Maximum Problem to Decision-Making under Uncertainty and Risk. Tech.Report, Stanford University, 1964

/17/ GEOFFRION A.M.: A Parametric Programming Solution to the Vector Maximum Problem, with Applications to Decision under Uncertainty. Tech. Report, Stanford University, 1965

/18/ GEOFRION A.M.: Stochastic Programming with Aspiration and Fractile Criteria. Management Sci.13, 672-679, 1967

/19/ GOICOECHEA A.: A Multi-Objective, Stochastic Programming Model in Watershed Management. Ph.D.disertation, Dept. of Systems and Industrial Engineering, Univ. of Arizona, Tucson, Arizona, 1977

/20/ GOICOECHEA A., DUCKSTEIN L. and BULFIN R.: The PROTRADE-method: a multiobjective approach to decision-making in Multiple Criteria decision making (Starr M.K. and Zeleny M. eds.) TIMS studies in the management sciences.Vol.6 North-Holland Publ.Comp., Amsterdam 1977

/21/ HENDRIX G.C. and STEDRY A.: The Elementary Redundancy-Optimization Problem: A Case Study in Probabilistic Multiple - Goal Programming.Operations Res.22, 639-653, 1974

/22/ KAPLAN R. and SODEN J.: On the Objective Function for the Sequential P-Model of Chance-Constrained Programming. Operations Res.19(1), 105-114, 1971

/23/ KARLIN S.: Mathematical methods and theory in games, programming and economics. Addison-Wesley, Cambridge, Mass., 1959

/24/ KEENEY R.L.: Multiplicative Utility Functions. Operations Res. 1, 22-34, 1974

/25/ LAU H.S.: The Newsboy Problem under Alternative Optimization Objectives.J.Opl.Res.Soc.31, 525-535, 1980

/26/ LECLERCQ J.P.: Résolution de programmes linéaires stochastiques par des techniques multicritères. Thèse de doctorat. Faculté de Sciences de l'Université de Namur, 1979

/27/ LECLERQ J.P.: La programmation lineaire stochastique: Une approache multicritere. Partie I: Formulation, Partie II: Un algorithme interactif adapte aux distributions multinormales. Cahiers Centre Études Recherche Opèr 23(1),/1981; ibidem 23(2), 121-132, 1981 31-41

/28/ NEUMANN VON J. and MORGENSTERN O.: Theory of Games and Economic Behavior. Princeton University Press, 1953

/29/ SENGUPTA,S.S., PODREBARAC M.L. and FERNANDO T.D.H.: Probabilities of Optima in Multi-Objective Linear Programmes.in Multiple Criteria Decision Making, J.L.Cochrane and M.Zeleny (eds.), University of South Carolina Press, Columbia,S.C., 1974, 217-235

/30/ STANCU-MINASIAN M.I.: Stochastic Programming with Multiple Objective Functions. Econom.Comp.Econom.Cybernet.Stud.Res.1 1974, 49-67

/31/ STANCU-MINASIAN M.I.: Notă asupra programării stohastice cu mai multe functii obiectiv, avînd vectorul c aleator. Stud.Cerc. Mat., 27(4), 453-459, 1975

/32/ STANCU-MINASIAN M.I.: A Selected Bibliography of Works Related to the Multiple Criteria Decision Making. Preprint nr.2, Catedra de Cibernetică Economică, A.S.E., Bucharest, Roumania, 1975

/33/ STANCU-MINASIAN M.I.: Criterii multiple în programarea stohastică Ph.D.dissertation. Centrul de Statistică Matematică, Bucuresti, 1976

/34/ STANCU-MINASIAN M.I.: Asupra problemei de risc minim multiplu I: cazul a două funcţii obiectiv. II: cazul a r($r>2$) funcţii obiectiv. Stud.Cerc.Mat., 28(5), 617-623, 1976, ibidem 28(6), 723-734, 1976

/35/ STANCU-MINASIAN M.I. and WETS M.J.: A Research Bibliography in Stochastic Programming. Operations Res. 24(6), 1078-1119, 1976

/36/ STANCU-MINASIAN M.I.: Asupra problemei lui Kataoka. Stud. Cerc.Mat. 28(1), 95-111, 1976

/37/ STANCU-MINASIAN M.I.: On Stochastic Programming with Mutiple Objective Functions. Proceedings of the Fifth Conference on Probability Theory. September 1-6, 1974, Braşov, Romania, Editura Academiei R.S.R., 428-436, 1977

/38/ STANCU-MINASIAN M.I.: Problema Cebîşev stocastică. Funcţia de repartiţie a optimului Stud.Cerc.Mat.30 (5), 567-577, 1978

/39/ STANCU-MINASIAN M.I.: On the Multiple Minimum Risk Problem. Bull.Math.Soc.Sci.Math.R.S.Roumanie (N.S.) Tome 23 (71), nr.4, 427-437, 1979

/40/ STANCU-MINASIAN M.I.: Stochastic Programming with Several Objective Functions (in Romanian). Editura Academiei R.S.R., Bucureşti, 1980

/41/ WOLFE P.: The Simplex Method for Quadratic Programming. Econometrica 27, 382-398, 1959

II

MULTIOBJECTIVE OPTIMIZATION AND DECISION MAKING

THE MANAGER'S DILEMMA: GOOD DECISIONS NEED NOT LEAD TO GOOD OUTCOMES

D.E. Bell

Harvard University, Boston, Massachusetts, USA

1. Introduction

The purpose of decision analysis is to aid the cognitive processes of a decision maker by creating an evaluation system that depends on a small number of rules. For decision making under uncertainty von Neumann and Morgenstern [3] suggested a number of rules that many people now believe should naturally be part of any person's decision making process. These rules imply that a decision maker need only specify a utility function incorporating tradeoffs between measures of outcome performance, and then select the alternative which maximizes the probability weighted expected value of that function.

The widespread acceptance of this model has led analysts to try and aid managers in this fashion. The manager will acknowledge, let us say, that his objective is to make money for the company. The analyst then asks questions of the form "Would you rather make $1 million profit for sure or take a 50-50 gamble between $0 and $3 million profit?". At this point the manager asks such questions as "What is the context? How did this choice arise? What can I do about the uncertainty?" and so on. The analyst explains patiently that the answers to these questions are of no consequence since all we are trying to do is establish the utility function. The manager knows these questions are of critical importance and the analysis is doomed from this point.

A traditional analysis of the management delegation problem is as follows. The organization, or higher level manager, has a utility function $u(x)$ for the profits X of the sector run by the manager. The manager perceives an economic benefit $e(x)$ accruing to himself (through promotions or direct bonuses) for achieving a level x and has a utility function $v(e)$ for those benefits. Thus the manager selects on the basis of $v(e)$, whereas the organization would wish him to use $u(x)$. The question is, does there exist an incentive system $i(x)$ such that $v(e(x) + i(x)) = u(x)$, where we assume $i(x)$ is negligible compared to x. If so, the manager should then make decisions exactly in accordance with company wishes.

But this is not the whole story. It would be the whole story if a manager's job were, each day, to select one of a number of alternatives presented to him, each explicitly described by a probability distribution of outcomes. However it is a manager's job to seek out and identify alternatives, also to estimate the probability distribution of outcomes for each alternative, and only then to evaluate them for desirability. The manager recognizes that he is being evaluated not only for his ability to reflect company objectives in his selections but also in his ability to create alternatives and forecast their prospects accurately.

To simplify the situation, suppose that an alternative always results in \$0 or \$1 and that a good manager has a probability p of picking a good alternative. Suppose that a bad manager can do this only with probability q (less than p). Suppose further that a newly hired manager is equally likely to be good or bad. If the manager's first decision proves to be bad, higher management will revise downwards its estimate of the likelihood of that manager being good from one-half to $(1-p)/[(1-p) + (1-q)]$. Thus a good manager who, by definition, is making good decisions may be condemned by bad outcomes.

It is natural, therefore, for a manager's behavior to be distorted by the knowledge that however loyal and accurate is his decision making process, his own interests may be better served by a different process, precisely because he will be judged by the quality of outcomes not by the quality of his decisions.

2. Examples of the Importance of Context

In order to show the need of the manager to know the context of a decision, it is useful to illustrate some contexts.

Illustration 1 You work for Company A, a regional toy manufacturer that earns 80% if its income during December. Due to long production and advertizing lead times your strategy for pricing and advertizing must be set well in advance, in fact in February. Your principal competitor is a national toy manufacturer, Company B. Their pricing is fairly standard and constant from year to year, but their advertizing strategy varies wildly. Some years they launch a major advertizing compaign, some years they advertize hardly at all. They do each about half the time and there is no way of prdeicting what they will do this year.

Company A has also kept its pricing strategy constant as too with its advertizing, which is quite modest. Company B is too big to care what strategy A follows, but A has seen no reason to change it.

You, however, have thought of two ideas for increasing profits. The first is to lower prices. You calculate that should B fail to advertize, customers seeking toys will see that A's are cheaper and buy them. This will lead to a net gain of $6 million in profit. However, if B does advertize, customers will be looking for B's products and A will only get its usual customers, but they will pay less so that in this case profits will be reduced by $4 million. However since each case is equally likely there is a net expected gain of $1 million. The question is do you really

want to take a decision which has a 50% chance of losing the company $4 million? The decision involves a considerable amount of personal risk.

The second idea is for A to advertize more itself. This strategy you calculate, will bring in a net gain of $1 million, no matter what B does. This is splendid; you achieve the gain of $1 million with no risk to yourself.

But is this decision in the company's best interests? Suppose that, in a normal year, if B advertizes, because of general increased interest in toys A makes $10 million more profit than it would otherwise. See Table 1.

Table 1

	B advertizes (.50)	B does not advertize (.50)
Status Quo	$ 15 million	$ 5 million
Lower Prices	$ 11 million	$ 11 million
Advertize	$ 16 million	$ 6 million

We can see from Table 1 that lowering prices reduces the profit uncertainty for company A. If top management were making the decision they would clearly want to lower prices. Because of the incremental evaluation system, a lower level manager would select the advertizing strategy.

Illustration 2 You invest Company A's pension funds in bonds, stocks and other vehicles such as mortgages and direct real estate. You believe that now is the time to be solidly in bonds but your counterparts at the at the rival companies B and C do not think so. Even after incorporating the fact that they think stocks will be superior, you remain convinced that bonds are the way to go. Table 2 summarizes your feelings on this.

Table 2

	Stocks go up (.50)	Stocks do not go up (.50)
Return on Stocks	20%	0%
Return on Bonds	12%	12%

What do you do? If stocks fail to go up you will be something of a hero in your own company and looked upon as lucky elsewhere. If stocks go up not only will you be dismissed for incompetence at A, companies B and C will not hire you either. Going with the consensus opinion is much the safer strategy.

Illustration 3 As a geologist you are assigned to decide which of two well sites to purchase. Well site A has a 50% chance of providing oil worth $5 million. Site B has a 25% chance of providing oil worth $10 million. The site you reject will be developed by your rivals.

Subcase 1 Sites A and B are adjacent. If well A is dry, then so too will B be dry.

Subcase 2 Sites A and B are hundreds of miles apart and have no bearing on each other.

Table 3

	Subcase 1		Subcase 2		
	A	B	A	B	
.50	0	0	0	0	.375
.25	5	0	5	0	.375
.25	5	10	5	10	.125
	-	-	0	10	.125

Table 3 shows that in Subcase 1 you have a 25% chance of being ahead $5 million by choosing A over B. You also have a 25% chance of being behind by $5 million by choosing A over B. Choosing B over A gives the same analysis. Since you are indifferent you select A on the grounds of risk aversion for the company. Now look at Subcase 2. The choice between A and B comes down to deciding whether a 25% chance of being $5 million ahead with A is worth a 12 1/2% chance of being behind $10 million. Perhaps this is a good tradeoff. Perhaps it is not. The point is merely that the analysis is different in the two subcases.

3. Performance Evaluation

The three illustrations show different ways in which managers may be evaluated:

a) Incremental Comparison If the company started the period in state x_o and finished it in state x_1, it is calculated in what state x_1' the company would have been had you not made the decision you did. This difference $x_1 - x_1'$ is a measure of your performance. You perceive $v(x_1 - x_1')$ to be the impact to you of this evaluation. The function v may be concave if you are worried about dismissal, convex if your job is secure but you are looking for promotion opportunities or s-shaped if you are required to achieve some goal of contribution (success if $x_1 - x_1' \geq g^*$, failure if $x_1 - x_1' < g^*$).

b) Comparison with the Competition If your division began the period in state x_o and ended in state x_1 and if during the same period your competitor (which might be internal or external to your company) went from y_o to y_1 then your performance is measured by $(x_1 - x_o) - (y_1 - y_o)$ or possibly by $x_1 - y_1$. Once again your perception of this evaluation is $v(x_1 - y_1)$.

This form of evaluation is quite widespread. Many companies judge their own performance by comparison with similar companies in their own business sector. A company with multiple divisions will investigate the management capabilities of the division performing, relatively, least well.

c) Comparison with Options not taken If it is your job to choose between two lucrative projects A and B you will get no credit just because the project you select does well. Instead your performance will be evaluated by how well your chosen project does (x) relative to the one you rejected (which results in y). Once again we have a measure x - y and an implication v(x - y).

This is the manner of decision evaluation used in the theory of decision regret [1] used to explain apparently paradoxical behavior in lottery style questions [2].

4. Implications for Management Decision Theory

The generic payoff function for a manager is thus $v(\tilde{x} - \tilde{y})$ where $\tilde{x}$ is a direct measure of the manager's contribution and $\tilde{y}$ is the level achieved by the benchmark. If the utility function of the company is u(x) it is clear that the manager will not make decisions consistent with it if he is using v(x - y). Moreover, no incentive system I(x) will correct this problem. Indeed even if an incentive system I(x,y) were developed to meet this need one would have to be sure that the particular choice of y (among those covered in Section 3) does not vary from problem to problem.

To take a specific example, let us take v to be concave and y to be the performance of a second division of the company.

Observation 1 The manager prefers projects $\tilde{x}$ which are positively correlated with $\tilde{y}$ over projects which are negatively correlated.

Proof $E[v(\tilde{x} - \tilde{y})] \approx (\bar{x} - \bar{y})\, v'(0) + 1/2 v''(0)[\sigma_x^2 + \sigma_y^2 - 2c\sigma_x\sigma_y + (\bar{x} - \bar{y})^2]$ where $\bar{x}$ and $\bar{y}$ are average returns and σ_x^2, σ_y^2 represent variances. The constant c is the correlation coefficient between $\tilde{x}$ and $\tilde{y}$. Note that if $v''(0) < 0$ then $c > 0$ is preferred to $c < 0$.

The problem is that if u is the utility function of the company we have

$$E[u(\tilde{x} + \tilde{y})] \approx (\bar{x} + \bar{y})u'(0) + 1/2\, u''(0)[\sigma_x^2 + \sigma_y^2 + 2c\sigma_x\sigma_y + (\bar{x} + \bar{y})^2].$$

If $u''(0) < 0$ the company prefers $c < 0$ to $c > 0$.

For the next example let $\tilde{y}$ be the project not taken.

Observation 2 The manager may pay a premium for situations in which y remains unresolved if not taken.

If y is not taken, evaluation of the manager can take place only relative to an estimate of how $\tilde{y}$ would have been resolved. Suppose the estimate is $\bar{y}$. For example in choosing between two R&D projects, it will never be known how the one not taken would have turned out. But the quality of the decision will be judged relative to the average performance of such projects.

We have

$$E[v(\tilde{x} - \tilde{y})] \approx (\bar{x} - \bar{y})v'(0) + 1/2\ v''(0)(\sigma_x^2 + (\bar{x} - \bar{y})^2)$$

which is greater than $E[v(\tilde{x} - \tilde{y})]$ if $v''(0)[\sigma_y - 2c\sigma_x] < 0$. If $v''(0) < 0$ this requires $\sigma_y \geq 2c\sigma_x$.

5. Conclusions

It is a fiction that a company has objectives that can be neatly summarized in a single attribute utility function. Just as the manager has complicated incentives and pressures so too has the organization. Therefore in its desire to aid organizations and their managers in making better decisions, management science must learn to recognize these pressures and to allow for them in an analysis. This may require a thorough understanding of the manager's environment and perspective in order to provide a useful decision support system.

6. References

[1] D.E. Bell, "Regret in Decision Making Under Uncertainty", Harvard University Business School Working Paper 82-15, September 1981.

[2] D. Kahneman and A. Tversky, "Prospect Theory: An Analysis of Decision Under Risk", Econometrica 47, 263-290, 1979.

[3] J. von Neumann and O. Morgenstern, Theory of Games and Economic Behavior, 2nd edition, Princeton University Press, Princeton, N.J., 1947.

ON THE APPLICABILITY OF THE NASH SOLUTION AND THE ZEUTHEN THEOREM TO WAGE BARGAINING PROCESSES

G. Fandel

Fernuniversität Hagen, FRG

Wage bargaining are decision processes among two parties with different utility functions. For treating them game and bargaining theoretical solution procedures (Fandel 1979a; Fandel 1979b) may be used. It is analysed how far the cooperative Nash-solution (Nash 1953) and the bargaining theorem of Zeuthen (Zeuthen 1930) are suitable to describe the decision behaviour of labour and management in the metal-processing industry of the Federal Republic of Germany during the wage bargaining processes between 1961 and 1979.

2. Wage bargaining in the metal-processing industry

2.1 Preliminary remarks

The subject of the analysis will be the wage disputes between "Gesamtmetall" as organization representing the employers and "IG Metall" (Metalworkers Union) as the employees' representation in the Federal Republic of Germany in 1961/62 to 1979. The years 1964, 1967 and 1972/73, however,had later to be excluded from the study, since in these years agreements were either reached unusually quickly on account of political or economic events or were simply taken over from other areas; in these cases there were not any signs of bargaining processes.

The metal-processing industry seems to lend itself particularly well to investigations of this kind since first of all, within the DGB (German Trade Union Federation) IG Metall represents by far the majority of the employees organized in this confederation and secondly, it shows a stronger tendency towards the strategy of cash wage than other trade unions (Zerche 1979, p. 25 and p. 117 ff.). The latter is often justified by the argument that the highest possible

nominal wage increase strengthens the power position of the union leaders more efficiently than any other strategies. Therefore, additional claims will not be taken into account because of their secondary role and the difficulties to quantify them adequately within the wage movements. In this context the difference between the greatly centralized bargaining strategy of the employers and the unionists' will to negotiate regionally has to be allowed for in a suitable way; for this reason we have chosen as bargaining processes the wage negotiations in the pilot areas which in the year concerned were regarded as trend-setting for the total wage movement within the branch of industry under review, and whose agreement proposals have been approved of by the members of Gesamtmetall and the representatives of IG Metall.

Table 1 specifies the initial claims of IG Metall $\tilde{w}_t$, the initial offers made by Gesamtmetall $\underline{w}_t$, the actual contracts w_t^* as well as the naive solutions $\tilde{w}_t$ for the years t from 1961/62 to 1979 expressed in percentages of the last basic wage. At this, the notion of the naive solution bases on a general rule of the economic practice saying that, usually, the later agreement with respect to the wage increase rate lies halfway between the union's initial claim and the employers' initial offer, that is to say approximately corresponds to the arithmetic mean of these two quantities.

2.2 Application of the cooperative Nash-solution

The close relation between the naive solution and the actual agreement calls for an inquiry of the question to what extent the agreement points of the wage negotiations are interpretable in the sense of the cooperative Nash-solution, since in case of linearly transferable utility this game-theoretic concept devides the bargaining cake available between two parties into two equal shares. For an analytic reconstruction of this possibility of interpretation let us assume in the following that the cake to be devided between IG Metall and Gesamtmetall is each time defined by the difference of the changes of the wage sums resulting from the initial claim of the union and the initial offer made by the employers, thus reads

$$(1)\quad K_t = (\tilde{w}_t - \underline{w}_t) \cdot L_t$$

Table 1: Survey of the wage negotiations 1961/62 to 1979

Year t	Initial claim of IG Metall $\bar{w}_t$	Initial offer made by Gesamtmetall $\underline{w}_t$	Actual contract achieved w_t^*	Naive solution $\tilde{w}_t = \frac{\bar{w}_t + \underline{w}_t}{2}$
1961/62	10,0	1,5	6,0	5,75
1963	8,0	3,0	5,67	5,5
1965/66	9,0	2,4	6,63	5,7
1968	7,0	4,0	5,5	5,5
1969	12,0	5,0	8,7	8,5
1970	15,0	7,0	11,0	11,0
1971	11,0	4,5	7,0	7,75
1974	18,0	8,5	12,25	13,25
1975	11,0	6,0	6,8	8,5
1976	8,0	4,0	5,4	6,0
1977	9,5	4,75	6,9	7,125
1978	8,0	3,0	5,4	5,5
1979	6,0	3,0	4,3	4,5

and is consequently fixed, L_t denoting the wage sum before the wage dispute in the year t, and workers and employers possessing linearly homogeneous utility functions with respect to the shares which they eventually obtain of this bargaining cake. If IG Metall and Gesamtmetall are assigned the indices n=1 and n=2 respectively, the utility functions remaining constant over the years with respect to the shares can be written as follows

(2) $u_{1t} = a \cdot \frac{w_t - \underline{w}_t}{\bar{w}_t - \underline{w}_t}$ and (3) $u_{2t} = b \cdot \frac{\bar{w}_t - w_t}{\bar{w}_t - \underline{w}_t}$ respectively,

the quantities a and b indicating the constant marginal utility. It is easy to see that the utility of IG Metall (Gesamtmetall) grows (declines) linearly with the rising wage increase rate w_t which the two parties have to agree upon in year t and for which $\underline{w}_t \leq w_t \leq \bar{w}_t$ holds in general. For $w_t = \bar{w}_t$ IG Metall reaches the highest utility, that of Gesamtmetall becoming equal to zero; correspondingly the

ratio is inverse for $w_t = \underline{w}_t$. Moreover, let $\bar{w}_t > \underline{w}_t$ be presupposed. The cooperative Nash-solution (Nash 1953) is characterized by the fact that the two negotiating parties involved in the case considered here agree upon a wage increase rate w_t^* or, which is equivalent, upon shares of the bargaining cake by which the product of their utility increases with respect to a disagreement vector $\bar{u} = (\bar{u}_1, \bar{u}_2)$ is maximized. To simplify matters the disagreement vector can be fixed by the zero utility levels of the negotiation parties. The following assumption which is fairly plausible is to justify this understanding: The bargaining cake corresponds to that part of the return of production for which in future they will have to work in common and the distribution of which to the factors work and capital must be agreed upon within the framework of the wage disputes; if one of the parties claims the total share this will be met by strike and lockout measures respectively by the other party; in such a case the burden for the fighting fund on the union's part on the one side is opposite to the capital expenditure for the plant facilities on the employers' part on the other.

Assuming a disagreement vector $\bar{u} = (\bar{u}_1, \bar{u}_2) = (0,0)$ in this sense, and considering (2) and (3) the cooperative Nash-solution can be determined as follows

$$\max u_t = (u_{1t} - 0)(u_{2t} - 0) = u_{1t} \cdot u_{2t} = a \cdot \frac{w_t - \underline{w}_t}{\bar{w}_t - \underline{w}_t} \cdot b \cdot \frac{\bar{w}_t - w_t}{\bar{w}_t - \underline{w}_t}$$

(4)

$$= \frac{a \cdot b}{(\bar{w}_t - \underline{w}_t)^2} (w_t - \underline{w}_t)(\bar{w}_t - w_t) = c(w_t - \underline{w}_t)(\bar{w}_t - w_t).$$

This expression is exclusively dependent on the wage increase rate w_t as variable which has to be determined optimally by both parties in the form of an agreement. As necessary condition for determining such an optimal w_t^o one obtains from (4)

$$(5) \quad \frac{du_t}{dw_t} = c[-(w_t - \underline{w}_t) + (\bar{w}_t - w_t)] = c(-2w_t + \bar{w}_t + \underline{w}_t) = 0$$

and from this, because of $c \neq 0$,

$$(6) \quad w_t^o = \frac{\bar{w}_t + \underline{w}_t}{2} .$$

Consequently, assuming the utility functions (2) and (3) to apply, and taking the assumptions concerning the bargaining object of the two parties and their disagreement vector as a basis, the cooperative Nash-solution w_t^o tallies with the naive solution $\bar{w}_t$, and its explanatory value with respect to the actual agreements reached w_t^*, as may be seen from Table 1, can be estimated accordingly. In this connection the cooperative Nash-solution has been derived from the initial claim $\bar{w}_t$ of the union and the initial offer $\underline{w}_t$ made by the employers; no statement, however, has thus been made about how these two initial values were obtained. As far as this is concerned it may be enough to say that both employers and employees probably take the data of the past or the future economic trend as a basis; due to its limited methodical performance the Nash concept does certainly not allow these data to be elucidated and verified. Similarly, the deviations of the actual contracts w_t^* from the analytically derived values w_t^o cannot be explained on the basis of the Nash approach. Since these deviations are not too important, as a rule, they could be ascribed to the differences which are usually to be found between the rationally postulated and the empirically observable decision behaviour. Assuming constant marginal utilities for both parties, however, seems to be unproblematic in view of the fact that wage movements are generally of special economic importance, and IG Metall as well as Gesamtmetall represent a very high number of persons interested.

For the wage negotiations of 1969 the above-mentioned analytic solution according to Nash is graphically represented in Figure 1. Here the coordinate axes are denoted by the utility arguments $(w_t-\underline{w}_t)$ and $(\bar{w}_t-w_t)$ or $[-(w_t-\bar{w}_t)]$ from the expression in (4), so that the utility values u_{1t} and u_{2t} of both parties increase positively with the direction of the coordinates. Between the points A and B line L marks all contracts w_t for which $\underline{w}_t=5\leq w_t\leq 12=\bar{w}_t$ holds.

While any points above L are not feasible since for them the union's claims are always higher than the offers made by the employers, which means that no agreements can be reached there, the points below L represent a waste of the cake to be divided. From the condition

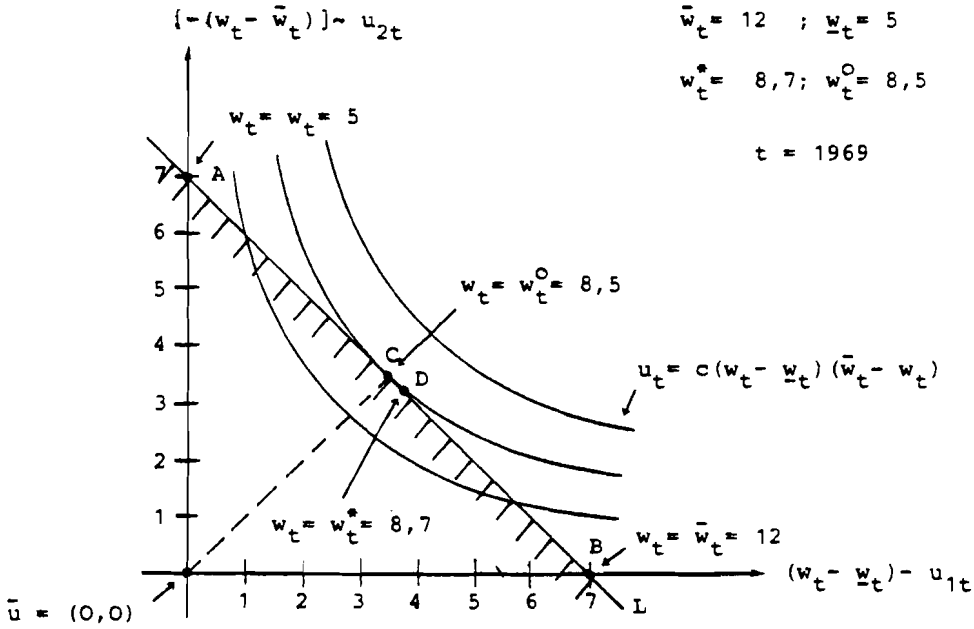


Figure 1: Nash solution w_t^o and actual agreement reached w_t^* for the wage negotiations in 1969.

of optimality $w_t^o - \underline{w}_t = \bar{w}_t - w_t^o$ according to (5) follows the Nash-solution $w_t^o = 8{,}5$ with $8{,}5 - 5 = 12 - 8{,}5 = 3{,}5$ which in Figure 1 is near point C; the actual contract signed $w_t^* = 8{,}7$, however, corresponds to point D on L. As can be seen from (5) and (6) w_t^o is independent of parameter c in (4) due to the invariance of the cooperative Nash-solution as to linear utility transformations; this parameter only represents a level constant with respect to the product of the utility increase which has to be maximized in common by both contrahents in comparison with the disagreement vector.

2.3 Verification of the wage bargaining processes with the aid of the theorem by Zeuthen

The bargaining theorem by Zeuthen (1930), which serves for rationalizing the concession behaviour of decision makers in conflicting decision situations (Fandel 1979b, p. 105 ff.), is identical with the axioms of the cooperative Nash-solution, so that

both concepts are equivalent with respect to determining an optimal agreement solution between the bargaining partners. But in comparison with the cooperative Nash-solution the bargaining theorem by Zeuthen represents a much more efficient instrument for verifying the behaviour-theoretic consistency of the decisions made by the bargaining partners. While according to Nash the optimal solution is determined statically, the bargaining theorem by Zeuthen, in order to reach this aim, requires a dynamic process consisting in proposals and counterproposals which consider the concessions made by the parties; in this way the optimal solution is interactively approximated step by step and eventually reached.

In order to be able to properly check the rationality of the concessions made by IG Metall and Gesamtmetall during the pay talks of 1961/62 to 1979 according to Zeuthen while strictly maintaining the utility functions introduced in (2) to (4) the respective claims and offers which were submitted by the bargaining partners until the agreement was reached, have been compiled in Table 2 for these different wage disputes in terms of bargaining rounds (Krelle 1976, p. 617 ff.). The claims and offers are given in wage increase rates and according to (4) can immediately be converted into the utility values that are necessary for applying the theorem of Zeuthen; to simplify matters the constant c can be neglected here, i.e. set equal to one. Simultaneously Table 2 shows that the wage negotiations are always to begin with an initial claim by the unions in the first round, the following claims are always to be made in the respective odd rounds, and the offers by Gesamtmetall are to be made only in the respective even rounds. Moreover, as far as those cases are concerned in which between two different claims and/or offers made by the one side there was no reaction by the other side, it has been presupposed that the other side has maintained its last claim and/or its last offer in the intermediate round. In 1963 the last two rounds and in 1965/66 the last four rounds are not taken into account in the following considerations, since in these two years agreements could have been reached earlier, but the negotiations were continued for the time being due to the fact that additional claims were dropped and finally a higher wage increase rate was agreed upon. Accordingly, it should be noted in 1968 that the wage claim of the union increases again in round 9, consequently no concession is made.

Table 2: Survey of the claims and offers $w_t^n(r_t)$ (in terms of bargaining rounds) submitted by IG Metall (n=1) and Gesamtmetall (n=2) during the wage bargaining processes 1961/62 to 1979.

r_t \ t, n	1961/62		1963		1965/66		1968		1969		1970		1971		1974		1975		1976		1977		1978		1979	
	1	2	1	2	1	2	1	2	1	2	1	2		2	1	2	1	2	1	2	1	2	1	2	1	2
1	10		8		9		7		13		15		11		18		11		8		9,5		8		6	
2		1,5		3		2,4		4		5		7		4,5		8,5		6		4		4,75		3		3
3	10		7		9		6,5		11		15		7,9		14		7		8		9,5		8		6	
4		3		3,5		5,8		4,57		5		10		6		11,33		6,5		4,5		5,25		4,6		4
5	8,5		6		5,4		6,5		10		11		7,5		12,25		6,8		5,4		7		5,7		4,3	
6		4		4,5		5,8		4,67		7		11		6,5		12,25		6,8		5,4		6,5		5		4,3
7	6		5,5		5		5		8,7				7								6,9		5,4			
8		6		5,5		6,63		4,83		8,7				7								6,9		5,4		
9			5,67		6,63		5,5																			
10				5,67				5,5																		
			1)		1)																					

1) These values are not considered in the verification according to the theorem by Zeuthen.

Let $R_t=\{r_t \mid r_t=1,2,\ldots,\bar{r}_t\}$ denote the set of round indices in year t, where $\bar{r}_t$ indicates the number of bargaining rounds in this year required for a (possible) agreement. Let $w_t^n(r_t)$ characterize the proposal made by partner n, $n\in\{1,2\}$ in the round $r_t\in R_t$ of year t; in this connection it should be pointed out that according to the arrangement described in the last paragraph index n=1 (IG Metall) can appear only in the case of odd and index n=2 (Gesamtmetall) only in the case of even round indices. Moreover, let $\hat{R}_t=\{\hat{r}_t \mid \hat{r}_t\in R_t\}$ be the set of round indices of year t in ascending order, for which one of the bargaining partners makes a concession, thus $w_t^1(\hat{r}_t)<w_t^1(\hat{r}_t-2)$ holds for the union, and $w_t^2(\hat{r}_t)>w_t^2(\hat{r}_t-2)$ holds for Gesamtmetall with $\hat{r}_t\in\hat{R}_t$ and $(\hat{r}_t-2)\in R_t$. Then let the mapping $\rho:\hat{R}_t\to\{1,2\}$ be defined by the fact that it indicates for each concession round $\hat{r}_t\in\hat{R}_t$ that bargaining partner $\rho(\hat{r}_t)\in\{1,2\}$ who made the concession in this round. The new suggestion $w_t^n(\hat{r}_t)$, $n\in\{1,2\}$ of that bargaining partner which results from this concession leads - according to (4) - to the product of the utility increase

$$(7)\quad u_t^n(\hat{r}_t)=c[w_t^n(\hat{r}_t)-\underline{w}_t][\bar{w}_t-w_t^n(\hat{r}_t)].$$

Moreover, in order that this concession may be rational according to the theorem by Zeuthen

$$(8)\quad u_t^n(\hat{r}_t-2)\leq u_t^{n'}(\hat{r}_t-1);\ n,n'\in\{1,2\},\ n\neq n';\ (\hat{r}_t-2),\ (\hat{r}_t-1)\in R_t$$

must hold (Fandel 1979b, p. 106).

If the fact whether such a concession was rational or not is then mapped by the binary attribute function $\psi(\hat{r}_t)$ with

$$(9)\quad \psi(r_t)=\begin{cases}1, & \text{if } \rho(\hat{r}_t)=n \text{ and } u_t^n(\hat{r}_t-2)\leq u_t^{n'}(\hat{r}_t-1)\\ 0 & \text{otherwise}\end{cases}$$

then the result of the analysis can be illustrated in a simple form by Table 3. Here index i indicates the concession steps of the years 1961/62 to 1979. In order to be able to partly reconstruct the evaluation of the results of Table 3, by way of explanation the derivation of the results for the year t=1969 is demonstrated

Table 3: Result of the verification with respect to Zeuthen rationality of the concessions made by IG Metall and Gesamtmetall during the wage negotiations 1961/62 to 1979

t	1961/62					1963						1965/66		1968						1969					1970		
$\hat{r}_t$	4	5	6	7	8	3	4	5	6	7	8	4	5	3	4	6	7	8	10	3	5	6	7	8	4	5	6
$\rho(\hat{r}_t)$	2	1	2	1	2	1	2	1	2	1	2	2	1	1	2	2	1	2	2	1	1	2	1	2	2	1	2
$\psi(\hat{r}_t)$	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1
i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27

t	1971						1974				1975				1976			1977					1978					1979		
$\hat{r}_t$	3	4	5	6	7	8	3	4	5	6	3	4	5	6	4	5	6	4	5	6	7	8	4	5	6	7	8	4	5	6
$\rho(\hat{r}_t)$	1	2	1	2	1	2	1	2	1	2	1	2	1	2	2	1	2	2	1	2	1	2	2	1	2	1	2	2	1	2
$\psi(\hat{r}_t)$	1	1	0	1	0	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1
i	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57

Table 4: Derivation of the results in Table 3 for the year t=1969

r_t	$w_t^1(r_t)$	$w_t^2(r_t)$	$\hat{r}_t$	$\rho(\hat{r}_t)$	$u_t^n(r_t)$	$\psi(\hat{r}_t)$	i
1	12		-	-	0	-	-
2		5	-	-	0	-	-
3	11		3	1	6	1	20
4		5	-	-	0	-	-
5	10		5	1	10	0	21
6		7	6	2	10	1	22
7	8,7		7	1	12,21	1	23
8		8,7	8	2	12,21	1	24

in Table 4, using the symbols introduced. The common utility values $u_t^n(r_t)$ resulting from the proposals and counterproposals of the partners can be marked as points on line L in Figure 1; for the sake of clarity, however, this has not been done in this case.

As can be seen from Table 3 47 out of 57 considered concession steps by the bargaining partners in the years 1961/62 to 1979 are rational according to Zeuthen; thus for the proportional value $\hat{f}$ of the Zeuthen steps

$$(10)\quad \hat{f}(\psi=1)=\frac{47}{57}=0,82$$

holds.

If one starts from the assumption that ψ is a random variable of quantitative binary quality and the concession steps of the years 1961/62 to 1979 represent a sample of the size m=57 from an infinite dichotomic parent entirety consisting of the set of all concession steps made between the bargaining partners IG Metall and Gesamtmetall in the Federal Republic of Germany and for whose elements only the property $\psi=1$ or $\psi=0$ is interesting, then the number of Zeuthen steps in such a sample of the size m is binomially distributed with the parameters m and f, where f may designate the percentage of Zeuthen steps of the infinite parent entirety. If on

this basis the zero hypothesis is tested, that is to say the percentage of Zeuthen steps in the parent entirety is at the most 70 %, i.e.

(11) $H_o: f \leq f_o = 0,70$

and if the sample proportion $\hat{f} = 0,82$ is chosen as suitable test variable which, due to $m \cdot f_o(1-f_o) = 57 \cdot 0,7 \cdot 0,3 = 11,97 > 9$ can be assumed to be approximately normally distributed, then the zero hypothesis is rejected with an error probability $\alpha = 0,05$ since the test variable $\hat{f}$ exceeds the upper limit c_o of the zero hypothesis (Wetzel 1973, p. 195 ff.), i.e.

(12) $\hat{f} = 0,82 > 0,81 = c_o$.

Thus by inverse conclusion it may be stated with 95 percent certainty that the percentage of Zeuthen steps in the parent entirety is higher than 70 %. Moreover, if on account of the sample one determines the confidence interval for the parameter f on the level of significance of $\alpha = 0,1$, where because of $m \cdot \hat{f}(1-\hat{f}) = 57 \cdot 0,82 \cdot 0,18 = 8,41 < 9$ an approximation of the binomial distribution by the normal distribution is not possible, but the F distribution must be used, then one obtains the interval

(13) $f \in [0,72;\ 0,90]$,

that is to say, with 90 percent certainty the percentage of the Zeuthen steps in the parent entirety is between 72 and 90 percent. The latter two statements which have been derived statistically on the basis of the empirical analysis suggest that the theorem by Zeuthen, together with the utility functions (2) and (3) assumed when applying the cooperative Nash-solution in the last section, is by all means acceptable as a methodical basis for substantiating the wage bargaining processes in the metal-processing industry under decision-theoretic aspects. Thus simultaneously the explanatory value of the cooperative Nash-solution and the naive solution is persistently backed.

References

[1] FANDEL, G.: Optimale Entscheidungen in Organisationen, Berlin-Heidelberg-New York 1979 (Fandel 1979 a).

[2] FANDEL, G.: Decision Concepts for Organizations, in: MORSE, J.N. (ed.): Organizations: Multiple Agents with Multiple Criteria, Berlin-Heidelberg-New York 1981, pp. 91-109 (Fandel 1979 b).

[3] KRELLE, W.: Preistheorie Teil II, 2. Edition, Tübingen 1976.

[4] NASH, J.F.: Two Person Cooperative Games, in: Econometrica, 1953, pp. 128-140.

[5] WETZEL, W.: Statistische Grundausbildung für Wirtschaftswissenschaftler, Volume 2: Schließende Statistik, Berlin 1973, pp. 195-220.

[6] ZERCHE, J.: Arbeitsökonomie, Berlin-New York 1979.

[7] ZEUTHEN, F.: Problems of Monopoly and Economic Warfare, London 1930.

STOCHASTIC DOMINANCE FOR MULTIPLE-CRITERIA DECISION MAKING

H. Nakayama[1], T. Tanino[2], T. Takeguchi[3], and Y. Sawaragi[4]

[1] *Department of Applied Mathematics, Konan University, Kobe, Japan*
[2] *Department of Mechanical Engineering, Tohoku University, Sendai, Japan*
[3] *Department of Precision Mechanics, Kyoto University, Kyoto, Japan*
[4] *The Japan Institute of Systems Research, Sakyo, Japan*

ABSTRACT:

In decision making under uncertainty, the best alternative may be obtained by maximizing the expected utility. To do this, many researchers have been paying much attention to the identification of utility functions. In practical situations, however, there are many cases where it is difficult to identify utility functions, for example, in group decisions, etc. Noting that probability distributions of alternatives play an important role in expected utilities as well as utility functions, the information on probability distributions give us a key for ranking alternatives, when only partial knowledge of utility functions is available. The notion of stochastic dominance is originated just from this idea. In this paper, stochastic dominance over a single attribute will first be surveyed and then it will be extended to cases with multiple attributes. Finally, the potential effectiveness of stochastic dominance will be discussed along with some examples.

1. INTRODUCTION

Let $\mathcal{A}$ denote the set of alternatives and let X be an attribute to measure possible outcomes caused by alternatives. x denotes a specific level of X and is supposed to be $x \geq 0$. We shall consider X scalar for a while. Let $F(x)$ and $G(x)$ be cumulative distributions corresponding to alternatives A and B, respectively. Throughout this paper, we follow the well-known axiom of maximizing the expected utility, that is,

$$A \succeq B \iff E(u,F) \geq E(u,G)$$

where

$$E(u,F) = \int_0^\infty u(x)dF(x).$$

Under the above axiom, the selection of the best alternative among $\mathcal{A}$ may be reduced to a kind of mathematical programming problems, when the full knowledge about the utility function is available. However, there are several controversial points in the practical assessment of utility functions such as violation of independence conditions, ambiguity due to subjective judgment by lotteries and difficulties of interpersonal comparison of utilities in group decisions, etc. Without identifying complete forms of utility functions, therefore, we try to rank alternatives based upon the information on probability distributions. The notion of stochastic dominance is introduced just to this aim.

Definition 1.1 Let u', u'' and u''' denote the first, the second and the third derivatives of the utility function u, respectively. Then the classes of utility functions $\mathcal{U}_1$, $\mathcal{U}_2$ and $\mathcal{U}_3$ are defined as follows:

(i) $\mathcal{U}_1 := \{u(x) \mid u(x) \in C^1,\ u'(x) > 0,\ \forall x \in [0,\infty)\}$.

(ii) $\mathcal{U}_2 := \{u(x) \mid u(x) \in C^2,\ u(x) \in \mathcal{U}_1,\ u''(x) < 0,\ \forall x \in [0,\infty)\}$.

(iii) $\mathcal{U}_3 := \{u(x) \mid u(x) \in C^3,\ u(x) \in \mathcal{U}_2,\ u'''(x) > 0,\ \forall x \in [0,\infty)\}$.

As is readily recognized, $\mathcal{U}_1$ is the class of utility functions for which the decision maker's preference is strictly increasing over outcomes. Furthermore, $\mathcal{U}_2$ is the class of utility functions which are of $\mathcal{U}_1$ and risk averse. Finally, it is easily seen that decreasing risk aversion utility functions belong to the class $\mathcal{U}_3$.

Definition 1.2 For i=1, 2 and 3,

$$F >_i G \text{ iff } E(u,F) > E(u,G), \quad \forall u \in \mathcal{U}_i.$$

We refer to $>_1$ as the first-degree stochastic dominance, or FSD; to $>_2$ as the second-degree stochastic dominance, or SSD; to $>_3$ as the third-degree stochastic dominance, or TSD.

These dominance relations induce some kinds of partial orderings over risky prospects. Therefore, although stochastic dominance may not be in general expected to lead to the best alternative by itself, it may be helpful to narrow down the alternative set.

Definition 1.3 Let $\mathcal{P}$ be the set of all right-continuous distribution functions with $F(0)=0$. For each $F \in \mathcal{P}$, we define F^n as follows:

$$F^1(x) := F(x)$$

$$F^{n+1}(x) := \int_0^x F^n(y)dy, \quad \forall x \in [0,\infty).$$

The right continuity of F and G of $\mathcal{P}$ implies that $F^2 \neq G^2$ when $F \neq G$. Therefore, $F = G$ follows whenever $F^n = G^n$ for some $n \geq 2$.

Theorem 1.1 (Fishburn, 1975)

(i) $F >_1 G$ iff $G(x) > F(x)$ $\forall x \in [0,\infty)$,

(ii) $F >_2 G$ iff $G^2(x) > F^2(x)$ $\forall x \in [0,\infty)$,

(iii) $F >_3 G$ iff $\mu_F \geq \mu_G$ and $G^3(x) > F^3(x)$ $\forall x \in [0,\infty)$,

where μ_F denotes the mean value of F given by

$$\mu_F := \int_0^\infty x dF(x).$$

Remark 1.1 As is well known, EV-dominance, for which alternatives with larger mean value and smaller variance are preferred, can not be necissarily adequate for decision problems, if we take the preference of the decision maker into account (Keeney and Raiffa, 1976). By Taylor expansion around the mean value μ_F, we have

$$u(x) = u(\mu_F) + u'(\mu_F)(x-\mu_F) + \frac{u''(\mu_F)}{2!}(x-\mu_F)^2 + \frac{u'''(\mu_F)}{3!}(x-\mu_F)^3 + \cdots\cdots\cdots .$$

Therefore, the expected utility is given by

$$E(u,F) = u(\mu_F) + \frac{u''(\mu_F)}{2!}\sigma^2 + \frac{u'''(\mu_F)}{3!}m^3 + \cdots\cdots\cdots \tag{1.1}$$

where the relation $E(x-\mu_F,F) = 0$ was used and σ^2 and m^3 denote the second moment $E((x-\mu_F)^2,F)$ and the third moment $E((x-\mu_F)^3,F)$, respectively. In general, there is no relationship between EV-dominance and stochastic dominance. However, the relation (1.1) implies that if more than second order derivatives of u vanish, for example, if u is of a quadratic form, then EV-dominance (EVD) induces second-degree stochastic dominance (SSD), namely, EVD $\subset$ SSD. Moreover, if more than second order moments vanish, we have immediately from (1.1) EVD $\subset$ TSD. Since for any two binary relations R_1 and R_2 with $R_1 \subset R_2$ we have $Ext[S|R_1] \supset Ext[S|R_2]$, stochastic dominance is more relevant for narrowing down the alternative set than EV-dominance . Here $Ext[S|R]$ denotes the extremal solution set of S with respect to the ordering R.

2. MULTIVARIATE STOCHASTIC DOMINANCE

The authors have discussed stochastic dominance for decision problems with multiple attributes in their previous paper (Nakayama and others, 1981). Repeating the result, two approaches to multivariate stochastic dominance are possible :

The one is based on a direct consideration of the decision maker's attitude toward risk in multiattribute cases. Let $X = X_1 \times X_2$ for simplicity. If for any x_1, x_2, y_1 and y_2 with $x_1 < y_1$ and $x_2 < y_2$, a decision maker prefers the risky alternative which gives an even chance for (x_1,y_2) and (y_1,x_2) to the one which gives an even chance for (x_1,x_2) and (y_1,y_2), then his preference is said to be <u>multivariate risk averse</u> (Richard, 1975). Then stochastic dominance for multivariate risk aversion $>_{MRA}$ is defined by

$$F >_{MRA} G \quad \text{iff} \quad E(u,F) > E(u,G) \quad \forall u \in \mathcal{U}_{MRA},$$

where $\mathcal{U}_{MRA}$ denotes the set of all multivariate risk averse utility functions. Under the terminology, we have the following:

<u>Theorem 2.1</u> Let $F(x_1,x_2)$ and $G(x_1,x_2)$ be two probability distribution functions absolutely continuous on $R_+^2 := \{(x_1,x_2) \in R^2 \mid x_1 \geq 0,\ x_2 \geq 0\}$ with $F_1(x_1)$, $F_2(x_2)$ and $G_1(x_1)$, $G_2(x_2)$ as marginals. Suppose that utility functions under consideration $u(x_1,x_2)$ are smooth enough and $\lim_{x_i \to \infty} u(x_1,x_2)$ exists for each $i=1,2$. Then if the decision maker's preference is increasing over outcomes,

$$F >_{MRA} G \quad \text{iff} \quad F(x_1,x_2) < G(x_1,x_2) \quad \text{for all} \quad (x_1,x_2) \in R_+^2 .$$

(proof) See Levy and Paroush (1974) and Nakayama and others (1981).

As a special case, if the utility function $u(x_1,x_2)$ is of an additive form $u(x_1,x_2) = w_1u_1(x_1) + w_2u_2(x_2)$, for example, if the attributes X_1 and X_2 are probability independent, then

$$F >_{MRA} G \quad \text{iff} \quad F_1(x_1) < G_1(x_1) \quad \forall x_1 \in [0,\infty)$$
$$\text{and} \quad F_2(x_2) < G_2(x_2) \quad \forall x_2 \in [0,\infty).$$

In the above approach to multivariate stochastic dominance, we have a drawback that it is not so easy to check the condition of multivariate

risk aversion in cases with more than two attributes. Now turn to the other approach to multivariate stochastic dominance via value functions.

Since it is often difficult for decision makers to answer their preferences over lotteries, several researchers have been recently trying to use value functions more actively. As is well known, measurable value functions, which are assessed by the information on the intensity of preference without using any lotteries, are originally valid for decision making under certainty. However, by making use of them as inputs to utility functions, several merits appear in risky situations (Bell,1981; Bodily, 1980; Keeney and Raiffa, 1976). When we have only partial knowledge about a multiattribute value function $v(x_1,\ldots,x_n)$ and a utility function over the value $u(v)$, stochastic dominance is introduced in a similar way to univariate stochastic dominance.

Definition 2.1 Letting $v(0,\ldots,0) = 0$ and $u(0) = 0$,

$$\mathcal{U}_1^* := \{u = u(v) \mid u' > 0,\ \partial v/\partial x_i > 0,\quad {}^{\forall}(x_1,\ldots,x_n) \in R_+^n \text{ and } i=1,\ldots n\}$$

$$\mathcal{U}_2^* := \{u = u(v) \mid u \in \mathcal{U}_1^*,\ u'' < 0,\ \partial^2 v/\partial x_i^2 < 0,\quad {}^{\forall}(x_1,\ldots,x_n) \in R_+^n,\ i=1,\ldots,n\}.$$

As is well known, the condition of $\partial^2 v/\partial x_i^2 < 0$ 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 function over v forms the class $\mathcal{U}_2^*$.

Although the followings will be concerned with $X = X_1 \times X_2$ for simplicity, the results can be easily extended to more general cases with $X = X_1 \times \ldots \times X_n$.

Definition 2.2 Cumulative distributions over $X = X_1 \times X_2$ are given by

$$\bar{F}(v) := \text{Prob}[v(x_1,x_2) \le v] = \int_0^\infty F_{2|1}(v_2(v,x_1)|x_1)dF_1$$

where $v_2(v,x_1)$ is the solution of x_2 to $v(x_1,x_2) = v$. Then the expected utility is written by

$$E(u(x_1,x_2),F(x_1,x_2)) = \int_0^{\infty}\int_0^{\infty} u(x_1,x_2)d^2F(x_1,x_2) = \int_0^{\infty} u(v)d\bar{F}(v).$$

Definition 2.3

$$F(x_1,x_2) \succ_{i*} G(x_1,x_2) \quad \text{iff} \quad E(u(x_1,x_2),F(x_1,x_2)) > E(u(x_1,x_2),G(x_1,x_2)), \quad \forall u \in \mathcal{U}_{i*}.$$

Note here that $F(x_1,x_2) \succ_{i*} G(x_1,x_2)$ is equivalent to $\bar{F}(v) \succ_i \bar{G}(v)$. Then multivariate stochastic dominance via value functions can be summarized as follows:

Theorem 2.2 (Huang and others, 1978; Takeguchi and Akashi, 1981)

If

(a) $G_1(x_1) > F_1(x_1)$ for all $x_1 \in [0,\infty)$

(b) $\partial G_{2|1}(x_2|x_1)/\partial x_1 < 0$ for all $(x_1,x_2) \in R_+^2$.

(c) $G_{2|1}(x_2|x_1) > F_{2|1}(x_2|x_1)$ for all $(x_1,x_2) \in R_+^2$,

then

$$F(x_1,x_2) \succ_{1*} G(x_1,x_2).$$

As to the second degree dominance, if

(a') $G_1(x_1) > F_1(x_1)$ for all $x_1 \in [0,\infty)$,

(b') $\partial G_{2|1}(x_2|x_1)/\partial x_1 < 0$ for all $(x_1,x_2) \in R_+^2$,

(c') $\int_0^{x_2} G_{2|1}(t|x_1)dt > \int_0^{x_2} F_{2|1}(t|x_1)dt$ for all $(x_1,x_2) \in R_+^2$,

then

$$F(x_1,x_2) \succ_{2*} G(x_1,x_2).$$

It is not so easy in general to check the conditions of Theorem 2.2 for joint distributions in multiattribute cases. For normal distributions,

which appear commonly in practical problems, we can invoke to the following equivalent conditions. Now we expand the domain X from R_+^2 to R^2.

Theorem 2.3 Let μ, σ, ρ denote respectively the mean value, the standard deviation and the corelation coefficient. If the distributions F and G assessed by the bivariate normal distributions $N(\mu_{F_1}, \mu_{F_2}, \sigma_{F_1}^2, \sigma_{F_2}^2, \rho_F)$ and $N(\mu_{G_1}, \mu_{G_2}, \sigma_{G_2}^2, \sigma_{G_2}^2, \rho_G)$, respectively, then we have

(i) the FSD conditions (a), (b) and (c) of Theorem 2.2 are equivalent to

$$\sigma_{F_1} = \sigma_{G_1}, \quad \sigma_{F_2} = \sigma_{G_2}, \quad \mu_{F_1} \geq \mu_{G_1}, \quad \rho_F = \rho_G \geq 0, \text{ and} \tag{2.1}$$

$$\sigma_{F_1}(\mu_{F_2} - \mu_{G_2}) + \rho_F \sigma_{F_2}(\mu_{G_1} - \mu_{F_1}) \geq 0,$$

(ii) the SSD conditions (a'), (b') and (c') of Theorem 2.2 are equivalent to

$$\sigma_{F_1} = \sigma_{G_1}, \quad \sigma_{F_2} \leq \sigma_{G_2}, \quad \mu_{F_1} \geq \mu_{G_1}, \quad \rho_F = \rho_G \geq 0, \text{ and} \tag{2.2}$$

$$\sigma_{F_1}(\sigma_{G_2}\mu_{F_2} - \sigma_{F_2}\mu_{G_2}) + \rho_F \sigma_{F_2}\sigma_{G_2}(\mu_{G_1} - \mu_{F_1}) \geq 0.$$

(proof) The marginal distribution function and the conditional distribution function of bivariate normal distribution G are respectively given by

$$G_1(x_1) = \int_{-\infty}^{(x_1-\mu_{G_1})/\sigma_{G_1}} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2} t^2)dt \tag{2.3}$$

and

$$G_{2|1}(x_2|x_1) = \int_{-\infty}^{\frac{x_2 - M_G(x_1)}{\sqrt{1-\rho_G^2}\,\sigma_{G_2}}} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2} t^2)dt \tag{2.4a}$$

where

$$M_G(x_1) = \mu_{G_2} + \rho_G \sigma_{G_2}(x_1 - \mu_{G_1})/\sigma_{G_1}. \tag{2.4b}$$

$F_1(x_1)$ and $F_{2|1}(x_2|x_1)$ are similarly given. It follows immediately from (2.4) that

$$\frac{\partial G_{2|1}(x_2|x_1)}{\partial x_1} = -\frac{\rho_G}{\sqrt{2\pi(1-\rho_G^2)}} \exp\{-\frac{(x_2 - M_G(x_1))^2}{2(1-\rho_G^2)\sigma_{G_2}^2}\}.$$

Hence, both conditions (b) and (b') of Theorem 2.2 are equivalent to

$$\rho_G \geq 0. \tag{2.5}$$

By the relation (2.3), we have

$$G_1(x_1) - F_1(x_1) = \int_{(x_1-\mu_{F_1})/\sigma_{F_1}}^{(x_1-\mu_{G_1})/\sigma_{G_1}} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}t^2)dt \tag{2.6}$$

In the right hand side of (2.6), the integrand is non-negative and both upper and lower bound of the interval of the integration are linear with respect to x_1, and therefore both conditions (a) and (a') of Theorem 2.2 are equivalent to

$$\frac{x_1 - \mu_{G_1}}{\sigma_{G_1}} - \frac{x_1 - \mu_{F_1}}{\sigma_{F_1}} \geq 0 \quad \text{for all} \quad x_1 \in R, \tag{2.7}$$

that is,

$$\sigma_{F_1} = \sigma_{G_1}, \quad \mu_{F_1} \geq \mu_{G_1}. \tag{2.8}$$

From the relation (2.4a), we have

$$G_{2|1}(x_2|x_1) - F_{2|1}(x_2|x_1) = \int_{\frac{x_2-M_F(x_1)}{\sqrt{1-\rho_F^2}\,\sigma_{F_2}}}^{\frac{x_2-M_G(x_1)}{\sqrt{1-\rho_G^2}\,\sigma_{G_2}}} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}t^2)dt. \tag{2.9}$$

The integrand of (2.9) is a non-negative and even function. Furthermore, both upper and lower bound of the interval of the integration are linear with respect to x_2. Therefore, the conditions (c) and (c') of Theorem 2.2 are respectively equivalent to the following relations:

$$(2.10)\quad \begin{cases} \sqrt{1-\rho_G^2}\,\sigma_{G_2} = \sqrt{1-\rho_F^2}\,\sigma_{F_2} \\ M_F(x_1) \geq M_G(x_1) \quad \text{for all} \quad x_1 \in R \end{cases}$$

and

$$(2.11)\quad \begin{cases} \sqrt{1-\rho_G^2}\,\sigma_{G_2} \geq \sqrt{1-\rho_F^2}\,\sigma_{F_2} \\ \sqrt{1-\rho_G^2}\,\sigma_{G_2} M_F(x_1) \geq \sqrt{1-\rho_F^2}\,\sigma_{F_2} M_G(x_1) \quad \text{for all} \quad x_1 \in R. \end{cases}$$

Moreover, from (2.4b), the relations (2.10) and (2.11) are respectively equivalent to

$$(2.10')\quad \begin{cases} \sqrt{1-\rho_G^2}\,\sigma_{G_2} = \sqrt{1-\rho_F^2}\,\sigma_{F_2} \\ \rho_G\sigma_{G_2}\sigma_{F_1} = \rho_F\sigma_{F_2}\sigma_{G_1} \\ \sigma_{F_1}\sigma_{G_1}(\mu_{F_2} - \mu_{G_2}) + \rho_G\sigma_{F_1}\sigma_{G_2}\mu_{G_1} - \rho_F\sigma_{G_1}\sigma_{F_2}\mu_{F_1} \geq 0 \end{cases}$$

and

$$(2.11')\quad \begin{cases} \sqrt{1-\rho_G^2}\,\sigma_{G_2} > \sqrt{1-\rho_F^2}\,\sigma_{F_2} \\ \sqrt{1-\rho_G^2}\,\rho_F\sigma_{G_1} = \sqrt{1-\rho_F^2}\,\rho_G\sigma_{F_1} \\ (\sqrt{1-\rho_G^2}\,\sigma_{G_2}\mu_{F_2} - \sqrt{1-\rho_F^2}\,\sigma_{F_2}\mu_{G_2}) + \sigma_{F_2}\sigma_{G_2}(\sqrt{1-\rho_F^2}\,\rho_G\mu_{G_1}/\sigma_{G_1} \\ \qquad - \sqrt{1-\rho_G^2}\,\rho_F\mu_{F_1}/\sigma_{F_1}) \geq 0. \end{cases}$$

Hence, the relation (2.1) is obviously derived from the relations (2.5), (2.8) and (2.10'). Similarly, the relation (2.2) follows from the relations (2.5), (2.8) and (2.11'). This completes the proof. □

Remark 2.1 In particular, if the attributes X_1 and X_2 are uncorelated, that is, $\rho_F = \rho_G = 0$, then the relations (2.1) and (2.2) are reduced respectively to the followings;

(i) [FSD] $\sigma_{F_1} = \sigma_{G_1}$, $\sigma_{F_2} = \sigma_{G_2}$, $\mu_{F_1} \geq \mu_{G_1}$, $\mu_{F_2} \geq \mu_{G_1}$,

and

(ii) [SSD] $\sigma_{F_1} = \sigma_{G_1}$, $\sigma_{F_2} \leq \sigma_{G_2}$, $\mu_{F_1} \geq \mu_{G_1}$, $\mu_{F_2}\sigma_{G_2} \geq \mu_{G_2}\sigma_{F_2}$.

In other words, the condition (i) implies that when the decision maker's preference is strictly increasing over outcomes, alternatives with larger mean values are more preferable in the set of alternatives with an equal standard deviation. The condition (ii) means that when the decision maker's preference is strictly increasing over outcomes and risk averse,

(a) if the ratio of the standard deviations $\sigma_{G_2}/\sigma_{F_2}$ is nearly equal to 1, alternatives with their standard deviations smaller and their mean values larger are more preferable,

(b) if the ratio $\sigma_{G_2}/\sigma_{F_2}$ is considerablylarger than 1, alternatives with their standard deviations smaller are more preferable, even if their mean values are more or less smaller.

3. EXAMPLES

The following example shows how effectively the set of alternatives can be narrowed down by stochastic dominance.

First, we selected 121 stocks from those traded on the Japanese Stock Exchange. Our aim is to find the optimal diversification investment for

a combination of some of these 121 stocks. The annual rate of return R_t for each stock is given by the following formula:

$$(3.1) \qquad R_t = \frac{(P_t - P_{t-1}) + D_t}{P_{t-1}} \quad \text{for} \quad t = 1969, \cdots, 1980,$$

where P_t denotes the stock price at the end of the period t and D_t represents the dividend in the period t. Let x_i $(i = 1,2,\cdots,12)$ be variables of R_t relabelled from smaller values to larger ones. Assume that the relative frequency of R_t is 1/12, that is, the frequency function $f(x_i)$ is given by

$$(3.2) \qquad f(x_i) = 1/12 \quad \text{for all} \quad i = 1,\ldots,12.$$

Therefore, the distribution function $F^1(x_k)$ $(k = 1,\ldots,12)$ is given by

$$(3.3) \qquad F^1(x_k) = \sum_{i=1}^{k} f(x_i) \quad \text{for} \quad k = 1,\ldots,12.$$

Moreover we have

$$(3.4) \qquad F^2(x_k) = \sum_{i=1}^{k} (x_i - x_{i-1})F^1(x_{i-1}) \quad \text{for} \quad k = 1,\ldots,12$$

$$(3.5) \qquad F^3(x_k) = \frac{1}{2} \sum_{i=1}^{k} (x_i - x_{i-1})[F^2(x_i) + F^2(x_{i-1})] \quad \text{for} \quad k = 1,\ldots,12.$$

$G^1(x_k)$, $G^2(x_k)$ and $G^3(x_k)$ are given in a similar fashion.

Stochastic dominance was applied in order to narrow down the original set of alternatives. The number of the efficient alternatives by FSD is 88 (reduction rate: 27.3 per-cent), that by SSD, 14 (reduction rate: 88.4 per-cent) and that by TSD, 11 (reduction rate: 90.9 per-cent). According to the result, we can see that FSD does not help the investor very much. It seems to be because the knowledge of his preference is too coarse. SSd and TSD, however, reduce the number of efficient alternatives effectively enough to our aim. The optimal diversification may be made among these reduced efficient alternatives. To do this, we are trying to use an interactive optimization technique. The details of our experiment will be presented in a separate paper.

4. CONCLUDING REMARKS

Stochastic dominance seems to play the most effective role in group decisions. As is well known, even if we accept the existence of the group utility function theoretically (see, for example, Keeney,1976), the interpersonal comparison of utilities is indispensable for practical assessment of the group utility function. This seems to be a fatal drawback of the utility approach to group decisions. However, note that there are many practical cases in which we can know some qualitative characteristics of the group utility function. Once the group utility function appears to be of some class discussed above, we can reduce the set of alternatives by using the corresponding stochastic dominance. It is clearly easier for members of the group to discuss their opinions based on the reduced set of alternatives. More detailed discussion on the applicability of stochastic dominance to group decisions can be seen in Nakayama and others (1981).

REFERENCES

Bell, D. E. (1981). Components of Risk, Proc. IFORS Meeting, Hamburg.

Bodily, S.E. (1981). Stability, Equality, Balance and Multivariate Risk, Proc. IV Conf. on MCDM, Delaware, Springer.

Fishburn, P.C. (1975). Stochastic Dominance: Theory and Applications, The Role and Effectiveness of Theories of Decision in Practice, ed. by D.J. White and K.C. Bowen, Hodder and Stoughton.

Huang, C.C., D. Kira and I. Vertinsky (1978). Stochastic Dominance Rules for Multiattribute Utility Functions, Rev. Economic Studies, XLV, 611-615.

Keeney, R.L. (1976). A Group Preference Axiomatization with Cardinal Utility, Management Sci., 23.

Keeney, R.L. and H. Raiffa (1976). Decisions with Multiple Objectives: Conflicts and Value Trade-offs. Wiley N.Y.

Levy, H. and J. Paroush (1974). Toward Multivariate Efficiency Criteria, J. Economic Th., 7., 129-142.

Nakayama, H., T. Tanino and Y. Sawaragi (1981). Stochastic Dominance for Decision Problems with Multiple Attributes and/or Multiple Decision-Makers, Preprints of IFAC Congress, Kyoto, Pergamon.

Richard, S.F. (1968). Multivariate Risk Aversion, Management Sci. 22. 12-21.

Takeguchi, T. and H. Akashi (1981). Decision Making by Stochastic Dominance, Proc. JAACE Meeting, Kyoto, (in Japanese).

INTERACTIVE MULTIOBJECTIVE DECISION MAKING BY THE FUZZY SEQUENTIAL PROXY OPTIMIZATION TECHNIQUE AND ITS APPLICATION TO INDUSTRIAL POLLUTION CONTROL

M. Sakawa[1] and F. Seo[2]

[1] *Department of Systems Engineering, Kobe University, Japan*
[2] *Kyoto Institute of Economic Research, Kyoto University, Japan*

Abstract

A new interactive multiobjective decisionmaking technique, which is called the sequential proxy optimization technique (SPOT), has been proposed by the authors. Using this technique, the preferred solution for the decisionmaker can be derived efficiently from among a Pareto optimal solution set by assessing his marginal rates of substitution and maximizing the local proxy preference functions sequentially. In this paper, considering the imprecise nature of decisionmaker's judgements, techniques based on fuzzy set theory are incorporated into the algorithm of SPOT. On the basis of the decisionmaker's marginal rates of substitution presented in a fuzzy form, which can be interpreted as type L-R fuzzy numbers, the revised version of SPOT called the fuzzy sequential proxy optimization technique (FSPOT) is presented. Based on the algorithms of FSPOT, a time-sharing computer program is also written in FORTRAN to implement man-machine interactive procedures. The industrial pollution control problem in Osaka City in Japan is formulated and the interaction processes are demonstrated together with the computer outputs.

1. Introduction

During the last few years, numerous methods related to multiobjective optimization and decisionmaking have been proposed [1-3], [7], [9-13], [17], [19], [20-25], [27], [31-32]. The excellent survey paper of Cohon and Marks [3] and, more recently, that of Wierzbicki [27] are devoted to a comparative evaluation of existing techniques; among them two competitive methods, namely, the multiattribute utility function (MUF) method [13] and the surrogate worth trade-off (SWT) method [10, 11] use global and local utility (preference) modelling respectively.

The MUF method developed by Keeney et al., global utility function modelling, uses two assumptions of preference independence and utility independence to limit the utility function to specialized forms. These global functions are mathematically simple and convenient, but they have a drawback. The assumptions are reasonable locally, but when assumed globally, they are very restrictive and may force the decisionmaker (DM) to fit a function not truly representing his or her preferences.

The SWT method developed by Haimes et al., based on local utility function modelling, provides an alternative approach that avoids the restrictive assumptions. Instead of specifying the utility function globally, their procedures construct a sequence of local preference approximations of it. The SWT method uses the ϵ-constraint problem as a means of generating Pareto optimal solutions. Trade-offs among objectives, whose values are expressed by values of strictly positive Lagrange multipliers, are used as a medium. The DM responds by expressing his degree of preference over the prescribed trade-offs and by assigning numerical values to each surrogate worth function. This method guarantees the generated solution in each iteration to be Pareto optimal and the DM can select his preferred solution from among Pareto optimal solutions.

However, the original version of the SWT method is noninteractive and several improvements, particularly in the way that the information from the DM is utilized, have been made [2], [25].

Recently, Sakawa [20] has proposed a new interactive multiobjective decisionmaking technique, which was called sequential proxy optimization technique (SPOT), by incorporating the desirable features of the conventional multiobjective decisionmaking methods. In his interactive on-line scheme, after obtaining a Pareto optimal solution, the marginal rates of substitution (MRS) assessed by the decisionmaker are used to determine the direction to which the utility funciton increases most rapidly. The local proxy preference function is updated to determine the optimal step size and Pareto optimality of the generated solution is guaranteed. The time-sharing computer program for this interactive procedure was also written in FORTRAN and was called ISPOT(interactive SPOT) [21, 22]. It was designed to facilitate the interactive processes for computer-aided decisionmaking and implemented on the ACOS-6 time-sharing system at the Kobe University, Japan.

SPOT requires a great number of precise MRS estimates of the DM, but it is a question whether the DM can respond with precise and consistent values of MRS through the whole searching processes because the DM's actions are often erratic, inconsistent due to the imprecise nature of human decision processes. Although by performing MRS consistency test some of this erratic behaviour is usually reduced, it is required to cope with the imprecise nature of DM's judgements.

In 1980, Baptistella and Ollero [1] proposed two different fuzzy interactive decisionmaking methods for multiobjective convex problem with linear constraints using the fuzzy set theory [4-6], [8], [28-30]. In their method, after reformulating the multiobjective convex problem into the equivalent scalar optimization problem using the classical

weighting method, the weighting vector is updated through the interaction with the DM in a fuzzy form. In the first method, they assumed that the DM can estimate in an approximate numerical way his local MRS between objectives; in the second one they assumed that the only disposable iterms of information are linguistic ones. Their methods enable us to consider in an efficient way inaccuracies, inherent in the DM's judgements, but concerning their interactive algorithm called basic algorithm some improvements, particularly in the way of generating Pareto optimal solutions as well as updating the weighting vector, must be made.

In this paper, considering the imprecise nature of DM's judgements, techniques based on fuzzy set theory which are similar to the first method of Baptistella and Ollero are incorporated into the algorithm of SPOT. On the basis of the DM's MRS presented in a fuzzy form, which can be interpreted as type L-R fuzzy numbers [4-6], the revised version of SPOT called the fuzzy sequential proxy optimization technique (FSPOT) is presented. Based on the algorithm of FSPOT, a time-sharing computer program is also written in FORTRAN to implement man-machine interactive procedures. Then the industrial pollution control problem in the industrialized areas near Osaka City is formulated. The problem is to allocate production factors (capital and labour) to each industry in such a way that industrial output, chemical oxygen demand (COD) and sulphur dioxide (SO_2) are optimized subject to reaource, technical and frictional constraints. The interaction processes using the time-sharing computer program based on FSPOT to solve this problem are also demonstrated along with the corresponding computer outputs.

2. Multiobjective Decisionmaking by the Fuzzy Sequential Proxy Optimization Technique

The multiobjective optimization problem (MOP) is represented as:

MOP

$$\min_{x} f(x) \triangleq (f_1(x), f_2(x), \ldots, f_n(x)) \tag{1}$$

subject to

$$x \in X = \{x | x \in E^N, g_j(x) \leq 0, \quad j=1, \ldots, m\} \tag{2}$$

where x is an N-dimentional vector of decision variables, $f_1, \ldots, f_n$ are n distinct objective functions of the decision vector x, $g_1, \ldots, g_m$ are inequality constraints and X is the feasible set of constrained decisions.

Fundamental to the MOP is the Pareto optimal concept, also known as a noninferior solution. Qualitatively, a Pareto optimal solution of the MOP is one where any improvement of one objective function can be achieved only at the expense of another. Usually, Pareto optimal solutions consist of an infinite number of points, and some kinds of subjective judgements should be added to the quantitative analyses by the DM. The DM must select his preferred solution from among Pareto optimal solutions.

The multiobjective decisionmaking problem (MDMP) we wish to solve is:

MDMP

$$\max_{x} U(f_1(x), f_2(x), \ldots, f_n(x)) \tag{3}$$

subject to

$$x \in X^P \tag{4}$$

where X^P is the set of Pareto optimal solutions of the MOP. $U(\cdot)$ is the DM's overall utility function defined on $F \triangleq \{f(x) | x \in E^N\}$ but his local imprecise knowledge of $U(\cdot)$ is assumed.

One way of obtaining Pareto optimal solutions to the MOP is to solve ε-constraint problem $P_1(\varepsilon_{-1})$ [2], [10-12]:

$\underline{P_1(\varepsilon_{-1})}$

$$\min \quad f_1(x) \tag{5}$$

$$\text{subject to} \quad x \in X \cap X_1(\varepsilon_{-1}) \tag{6}$$

$$\text{where } \varepsilon_{-1} \triangleq (\varepsilon_2, \varepsilon_3, \ldots, \varepsilon_n) \tag{7}$$

$$X_1(\varepsilon_{-1}) \triangleq \{x | f_i(x) \leq \varepsilon_i, \quad i=2, \ldots, n\} \tag{8}$$

$$\varepsilon_{-1} \in E_1 \triangleq \{\varepsilon_{-1} | X_1(\varepsilon_{-1}) \neq \phi\}. \tag{9}$$

Of course, any objective could be chosen as f_1. Note that (9) is a necessary condition for $P_1(\varepsilon_{-1})$ to have a feasible solution.

Let us assume that $x^*(\varepsilon_{-1})$, an optimal solution to the $P_1(\varepsilon_{-1})$, be unique for the given $\varepsilon_{-1} \in E_1$. And let AE_1 be a set of ε_{-1} such that all the ε-constraint (8) are active, that is

$$AE_1 \triangleq \{\varepsilon_{-1} | \varepsilon_{-1} \in E_1, \ f_i(x^*(\varepsilon_{-1})) = \varepsilon_i, \quad i=2, \ldots, n\} \tag{10}$$

Then the following theorem, which is essentially same as in Payne et al. [19] shows that the Pareto optimal solution set of the MOP coincides with the solution set of $P_1(\varepsilon_{-1})$ under suitable assumptions.

Theorem 1. $x^* \in X$ is a Pareto optimal solution of the MOP if and only if $x^* \in X$ is a unique solution of $P_1(\varepsilon_{-1})$ for some $\varepsilon_{-1} \in AE_1$.

If the Kuhn-Tucker condition for problem $P_1(\varepsilon_{-1})$ is satisfied, the Lagrange multiplier $\lambda_{1i}(\varepsilon_{-1})$ associated with the ith active constraint can be represented as follows:

$$\lambda_{1i} = -\{\partial f_1(\varepsilon_{-1})\}/\{\partial f_i(\varepsilon_{-1})\} \quad i=2, \ldots, n \tag{11}$$

By taking account of Theorem 1, if the unique optimal solutions of the $P_1(\varepsilon_{-1})$, $x^*(\varepsilon_{-1})$, are substituted to the MDMP given desired levels of $\varepsilon_{-1} \in AE_1$, the MDMP can be restated as the following ε_{-1}-parametric utility maximization problem.

$$\max_{\epsilon_{-1} \in AE_1} \bar{U}(\epsilon_{-1}) \triangleq U(f_1[x^*(\epsilon\text{-}1)], \epsilon_{-1}) \qquad (12)$$

To proceed we introduce the concept of marginal rates of substitution (MRS) of the DM.

Definition 1. At any point of $f=(f_1,\ldots, f_i,\ldots, f_n)$, the amount of f_1 that the DM is willing to acquire for sacrificing an additional unit of f_i is called the MRS. Mathematically, the MRS is the negative slope of the indifference curve at the f:

$$m_{1i}(f) = [\partial U(f)/\partial f_i]/[\partial U(f)/\partial f_1] = -df_1/df_{i|dU=0,\ df_r=0,\ r\neq 1,\ i} \qquad (13)$$

where each indifference curve is a locus of points among which the DM is indifferent.

Usually, the decision analyst assesses MRS by presenting the following prospects to the DM.

$$f = (f_1,\ldots, f_i,\ldots, f_n), \quad f' = (f_1-\Delta f_1,\ldots, f_i+\Delta f_i,\ldots, f_n) \qquad (14)$$

for a small fixed Δf_1, small enough so the indifference curve is approximately linear but large enough so the decrement is meaningful. The analyst varies Δf_i untill the DM is indifferent between f and f'. At this level, $m_{1i}(f) \cong \Delta f_1/\Delta f_i$.

Dyer proposed algorithmic procedures for approximating the MRS through a series of ordinal paired comparisons, where the DM is required to indicate a preference for one of two possible choices, or to indicate his indifference between the two [7]. For the fixed decrement Δf_1, a modification of the bisection algorithm is applied to the value of Δf_i, decreasing Δf_i if the DM prefers f and increasing Δf_i if he prefers f'. The process is continued until a value Δf_i is obtained at which the DM is indifferent between f and f'.

In this paper, considering the imprecise nature of DM's judgements, we assume that the DM can estimate in an approximate numerical way his local MRS as follows: "For the fixed decrement Δf_1, Δf_i should be an approximate number between Δf_{imin} and Δf_{imax} with the mean in the vicinity of $\Delta \bar{f}_i$." Denoting

$$\bar{m}_{1i} = \Delta f_1 / \Delta \bar{f}_i, \quad m_{1imin} = \Delta f_1 / \Delta f_{imax}, \quad m_{1imax} = \Delta f_1 / \Delta f_{imin} \tag{15}$$

and using the concept of fuzzy numbers and its L-R representation [4-6], the DM's MRS can be interpreted as type L-R fuzzy numbers with mean $\bar{m}_{1i}$, left and right spread parameters η_{1i} and θ_{1i}:

$$\tilde{m}_{1i} = (\bar{m}_{1i}, \eta_{1i}, \theta_{1i})_{LR} \quad i=2,\ldots, n \tag{16}$$

where the membership function $\mu_{\tilde{m}_{1i}}$ which describes $\tilde{m}_{1i}$ is

$$\mu_{\tilde{m}_{1i}}(t) = \begin{cases} L[(\bar{m}_{1i}-t)/\eta_{1i}], & t \leq \bar{m}_{1i} \\ R[(t-\bar{m}_{1i})/\theta_{1i}], & t \geq \bar{m}_{1i} \end{cases} \quad i=2,\ldots, n \tag{17}$$

and

$$\left.\begin{aligned} L[(\bar{m}_{1i}-m_{1imin})/\eta_{1i}] &= 0 \\ R[(m_{1imax}-\bar{m}_{1i})/\theta_{1i}] &= 0 \end{aligned}\right\} \quad i=2,\ldots, n \tag{18}$$

L is for left reference function and R for right reference function. Examples of L and R functions are :

$$\left.\begin{aligned} &y = \max(0, 1-|x|^p) \quad p \geq 1 \\ &y = \exp(-|x|^p) \quad p \geq 1 \\ &y = 1/(1+|x|^p) \quad p \geq 1 \\ &y = 1 \text{ in } [-1, +1], \quad y = 0 \text{ elsewhere} \end{aligned}\right\} \tag{19}$$

Throughout this paper we make the following assumptions.

Assumption 1: The DM's local imprecise knowledge of U is assumed. Moreover, U is assumed to be concave, a strictly decreasing and continuously differentiable function on F.

Assumption 2: All f_i, i=1,..., n and all g_j, j=1,..., m are convex and twice continuously differentiable in their respective domains and constraint set X is compact.

Assumption 3: ε_{-1} is an interior point of AE_1.

From Assumptions 1-2, the following theorem holds [20].

Theorem 2. Under Assumptions 1-2, the utility function $\bar{U}(\varepsilon_{-1})$ is concave with respect to $\varepsilon_{-1} \in AE_1$.

Now, we can formulate the gradient $\partial\bar{U}(\varepsilon_{-1})/\partial\varepsilon_i$(i=2,..., n) of utility function $\bar{U}(\varepsilon_{-1})$. Applying the chain rule

$$\partial\bar{U}(\cdot)/\partial\varepsilon_i = \partial U(\cdot)/\partial\varepsilon_i + [\partial U(\cdot)/\partial f_1][\partial f_1/\partial\varepsilon_i] \quad i=2,\ldots, n \qquad (20)$$

Using the relations (11) and (13), we have the following:

$$\partial\bar{U}(\cdot)/\partial\varepsilon_i = [\partial U(\cdot)/\partial f_1]\,(m_{1i}-\lambda_{1i}) \quad i=2,\ldots, n. \qquad (21)$$

From the strict monotonicity of U with respect to f_1, $\partial U(\cdot)/\partial f_1$ is always negative. Therefore $\lambda_{1i}-m_{1i}$ (i=2,..., n) decide a direction improving the values of $\bar{U}(\cdot)$ at a current point.

Under Assumptions 1-3, if the maximum is reached at an interior point of AE_1, the optimality conditions for a maximization point ε_{-1} are $\partial\bar{U}(\cdot)/\partial\varepsilon_{-i} = 0$, that is

$$m_{1i} = \lambda_{1i} \quad i=2,\ldots, n. \qquad (22)$$

This is a well-known result that at the optimum the MRS of the DM must be equal to the trade-off rate.

The optimality condition (22) can be rewritten by using the square of their normalized scalar product C defined by:

$$C = (\textstyle\sum_i \lambda_{1i}m_{1i})^2/(\sum_i \lambda_{1i}^2 \sum_i m_{1i}^2) \qquad (23)$$

where $\sum_i$ implies the sum over i=2,..., n. Obviously $0 \leq C \leq 1$ and the alternative form of the optimality condition is then

$$C = 1 \tag{24}$$

However, when the DM's MRS are presented in a fuzzy form as in (15), which can be interpreted as type L-R fuzzy numbers $\tilde{m}_{1i} = (\bar{m}_{1i}, \eta_{1i}, \theta_{1i})_{LR}$ (i=2, 3,..., n), C in (23) becomes a fuzzy number $\tilde{C}$ as follows:

$$\tilde{C} = (\sum_i \lambda_{1i}\tilde{m}_{1i})^2/(\sum_i \lambda_{1i}^2 \sum_i \tilde{m}_{1i}^2) \tag{25}$$

Operations of (25) can be greatly simplified by using L-R representations. The following formulae were already demonstrated by Dubois and Prade [4-6]. Given the L-R fuzzy numbers $\tilde{m} = (m, \eta, \theta)_{LR}$ and $\tilde{n} = (n, \mu, \nu)_{LR}$, the sum $\tilde{m} \oplus \tilde{n}$ is an L-R fuzzy number, and

$$(m, \eta, \theta)_{LR} \oplus (n, \mu, \nu)_{LR} = (m+n, \eta+\mu, \theta+\nu)_{LR} \tag{26}$$

The approximation formulae for product $\tilde{m} \odot \tilde{n}$, inverse of $\tilde{n}$ and quotient $\tilde{m} \oslash \tilde{n}$ are

$$(m, \eta, \theta)_{LR} \odot (n, \mu, \nu)_{LR} \simeq (mn, m\mu+n\eta, m\nu+n\theta)_{LR} \tag{27}$$

$$(n, \mu, \nu)_{LR}^{-1} \simeq (n^{-1}, \nu n^{-2}, \mu n^{-2})_{RL} \tag{28}$$

$$(m, \eta, \theta)_{LR} \oslash (n, \mu, \nu)_{RL} \simeq (m/n, (m\nu+n\eta)/n^2, (m\mu+n\theta)/n^2)_{RL} \tag{29}$$

Using these formulae we get

$$(\sum_i \lambda_{1i}\tilde{m}_{1i})^2 \simeq (p, \alpha, \beta)_{LR} \tag{30}$$

$$(\sum_i \lambda_{1i}^2 \sum \tilde{m}_{1i}^2)^{-1} \simeq (q, \gamma, \delta)_{RL} \tag{31}$$

where

$$p = (\sum_i \lambda_{1i}\bar{m}_{1i})^2 \tag{32}$$

$$\alpha = 2 \sum_i \lambda_{1i}\bar{m}_{1i} \sum_i \lambda_{1i}\eta_{1i} \tag{33}$$

$$\beta = 2 \sum_i \lambda_{1i}\bar{m}_{1i} \sum_i \lambda_{1i}\theta_{1i} \tag{34}$$

$$q = 1/(\sum_i \lambda_{1i}^2 \sum_i m_{1i}^2) \tag{35}$$

$$\gamma = (\sum_i \lambda_{1i}^2 \sum_i 2\bar{m}_{1i}\theta_{1i})/(\sum_i \lambda_{1i}^2 \sum_i \bar{m}_{1i}^2)^2 \tag{36}$$

$$\delta = (\sum_i \lambda_{1i}^2 \sum_i 2\bar{m}_{1i}\eta_{1i})/(\sum_i \lambda_{1i}^2 \sum_i \bar{m}_{1i}^2)^2 \tag{37}$$

If we assume L(x) = R(x), an approximate L-R representation to $\tilde{C}$ in (25) becomes

$$\tilde{C} = (\bar{C}, \rho, \sigma)_{LR} \simeq (p-q, p\gamma+q\alpha, p\delta+q\beta)_{LR} \tag{38}$$

When comparing the fuzzy numbers $\tilde{C}^{\ell-1}$ and $\tilde{C}^{\ell}$, the following question may arise: what is the true-value of the assertion; " Is $\tilde{C}^{\ell}$ greater than $\tilde{C}^{\ell-1}$? " In other words, what is the possibility for $\tilde{C}^{\ell}$ to be greater than $\tilde{C}^{\ell-1}$? The truth value of the assertion $\tilde{C}^{\ell}$ is greater than $\tilde{C}^{\ell-1}$, which we write $\tilde{C}^{\ell} > \tilde{C}^{\ell-1}$, defined as [4-6]

$$v(\tilde{C}^{\ell} > \tilde{C}^{\ell-1}) = \max_{p \geq q} \min \left(\mu_{\tilde{C}^{\ell}}(p), \mu_{\tilde{C}^{\ell-1}}(q) \right) \tag{39}$$

where $\mu_{\tilde{C}^{\ell}}$ and $\mu_{\tilde{C}^{\ell-1}}$ are respective membership functions of $\tilde{C}^{\ell}$ and $\tilde{C}^{\ell-1}$. This formula is an extension of the inequality $p \geq q$ according to the extension principle. It can be easily checked that [4-6]

$$v(\tilde{C}^{\ell} > \tilde{C}^{\ell-1}) = 1 \quad \text{if and only if} \quad \bar{C}^{\ell} > \bar{C}^{\ell-1} \tag{40}$$

$$v(\tilde{C}^{\ell} < \tilde{C}^{\ell-1}) = \text{height} (\tilde{C}^{\ell} \cap \tilde{C}^{\ell-1}) = \mu_{\tilde{C}^{\ell}}(d) = \mu_{\tilde{C}^{\ell-1}}(d) \tag{41}$$

where d is the ordinate of the highest intersection point D between $\mu_{\tilde{C}^{\ell}}$ and $\mu_{\tilde{C}^{\ell-1}}$.

Thus, the answer to the question "Is $\tilde{C}^{\ell}$ greater than $\tilde{C}^{\ell-1}$?" is a fuzzy set A of the universe {yes, no}:

$$A = v(\tilde{C}^{\ell} > \tilde{C}^{\ell-1})/\text{yes} + v(\tilde{C}^{\ell} < \tilde{C}^{\ell-1})/\text{no} \tag{42}$$

Now, at the ℓth iteration, if we adopt the mean value $\bar{m}_{1i}^{\ell}$ of the DM's MRS, the possible direction of search S_{-1}^{ℓ} is given by:

$$S_{-1}^{\ell} = (\lambda_{12}^{\ell} - \bar{m}_{12}^{\ell}, \ldots, \lambda_{1n}^{\ell} - \bar{m}_{1n}^{\ell}) \tag{43}$$

Then we adopt $(f_1(\varepsilon_{-1}^{\ell} + \alpha S_{-1}^{\ell}), \varepsilon_{-1}^{\ell} + \alpha S_{-1}^{\ell})$ as a search point in the process of linear search for the step size α. Our search point becomes a Pareto optimal solution by solving the ε-constraint problem and the DM can select his preferred solution from among the Pareto optimal solutions. In order to determine the step size, we introduce the following local proxy preference functions like oppenheimer's method [17].

(1) sum-of-exponentials

$$\text{If } [-\partial m_{ij}(f)/\partial f_j]/m_{ij}(f) = \omega_j \text{ then } P(f) = -\sum a_i \exp(-\omega_i f_i). \quad (44)$$

(2) sum-of-powers ($\alpha_j \neq 0$)

$$\text{If } [-\partial m_{ij}(f)/\partial f_j]/m_{ij}(f) = (1+\alpha_j)/(M_j+f_j) \text{ then } P(f) = -\sum a_i (M_i+f_i)^{\alpha_i} \quad (45)$$

where M_i is a constant such that $M_i+f_i > 0$, i=1,..., n

(3) sum-of-logarithms

$$\text{If } [-\partial m_{ij}(f)/\partial f_j]/m_{ij}(f) = 1/(M_j-f_j) \text{ then } P(f) = \sum a_i \ln(M_i-f_i) \quad (46)$$

where M_i is a constant such that $M_i-f_i > 0$, i=1,..., n.

Although these utility functions are very restrictive globally, they are reasonable when assumed locally. We use one of these utility functions only as a mechanism to guide the search for the best step size.

Following the above discussions we can now construct the algorithm of the fuzzy sequential proxy optimization technique (FSPOT) in order to obtain the preferred solution of the DM for the MDMP.

Step 1. (Initialization) Choose initial point $\epsilon_{-1} \in E_1$ and set $\ell = 1$.

Step 2. (Pareto optimal solution) Set $\epsilon_{-1} = \epsilon_{-1}^{\ell}$, solve an ϵ-constraint problem $P_1(\epsilon_{-1}^{\ell})$ for ϵ_{-1}^{ℓ} and obtain a Pareto optimal solution $x^*(\epsilon_{-1}^{\ell})$, a Pareto optimal value $f^{\ell} = (f_1^{\ell}[x^*(\epsilon_{-1}^{\ell})], \epsilon_{-1}^{\ell})$ and corresponding Lagrange multiplier λ_{1i}^{ℓ} (i=2,..., n). If all the ϵ-constraints are active, go to the next step. Otherwise, replace ϵ_i for inactive constraints by $f_i[x^*(\epsilon_{-1}^{\ell})]$ (i=2,..., n) and solve the ϵ-constraint problem to obtain the corresponding Lagrange multipliers.

Step 3. (Assessment) Assess the MRS of the DM at f^{ℓ} in a fuzzy form and choose the form of L and R functions in a subjective manner.

Step 4. (Termination) If $v(\tilde{C}^{\ell} > \tilde{C}^{\ell-1}) \cong 1$, $v(\tilde{C}^{\ell} < \tilde{C}^{\ell-1}) \cong 1$ and $\tilde{C} \cong 1$, stop. Then a Pareto optimal solution $(f_1^{\ell}[x^*(\epsilon_{-1}^{\ell})], \epsilon_{-1}^{\ell})$ is the preferred solution of the DM.

Step 5. (Direction vector) Determine the direction vector S_{-1}^{ℓ} by $S_i^{\ell} = \lambda_{1i} - \bar{m}_{1i}^{\ell}$ $(i=2,\ldots, n)$.

Step 6. (Local proxy) Select the form of the proxy function that will be used at each iteration, and calculate the parameters using the mean values of the DM's MRS.

Step 7. (Step size) Change the step size, obtain corresponding Pareto optimal values and search for three α values α_A, α_B and α_C which satisfy $\alpha_A < \alpha_B < \alpha_C$ and $P(\alpha_A) < P(\alpha_B) > P(\alpha_C)$, where $P(\alpha) \triangleq P(f_1[x^*(\epsilon_{-1}^{\ell} + \alpha S_{-1}^{\ell})], \epsilon_{-1}^{\ell} + \alpha S_{-1}^{\ell})$. This step operates either doubling or halfing the step size until the maximum is bracketed. Then a local maximum of $P(\alpha)$ is in the neighbourhood of $\alpha = \alpha_B$. Set $\ell = \ell+1$ and return to step 2.

Remark. Reduction of some of the inconsistent behaviour of the DM is usually accomplished by performing MRS consistency test at the various iteration points. Two types of consistency tests may be performed using the mean value $\bar{m}_{1i}$ of the DM's MRS; the first testing MRS consistency at a single point, and the second testing the consistency at successive points.

The single point test requires a second set of assessments at each point and checks whether the MRS of the DM satisfies the chain rule, i.e., $m_{kj} = m_{ki}\, m_{ij}$, $i, j=1,\ldots, n$, $i\neq k$, $k\neq i$, $k\neq j$.

The second test checks for decreasing marginal rates of substitution of the proxy, which is based on the following theorem [20].

Theorem 3.

(1) The sum-of-exponentials proxy P(f) is concave and strictly decreasing if and only if

$$a_i > 0 \text{ and } \omega_i > 0, \quad i=1,\ldots, n \tag{47}$$

(2) The sum-of-powers proxy P(f) is concave and strictly decreasing if and only if

$$a_i > 0 \text{ and } \alpha_i > 1, \quad i=1,\ldots, n \tag{48}$$

(3) The sum-of-logarithms proxy P(f) is concave and strictly decreasing if and only if

$$a_i > 0, \quad i=1,\ldots, n. \tag{49}$$

3. An Application to Environmental Problems

Consider an application of the proposed method to an industrial area in Japan. The middle part of Osaka Prefecture is one of the most highly industrialized areas in Japan. Osaka City, which is the second largest industrial and commercial area in Japan, contains many small rivers which are branches of the Yodo River. The Yodo River is an important source of drinking water for Osaka's residents, but water pollution in the Yodo River basin has become increasingly serious because of the rapid industrial development since 1960. Air pollution is also at critical level in the greater Osaka area. In addition, the water supply capacity is limited in this area. Although the sixth expansion program for Osaka's public water supply increased the supply by more than 11 percent in 1975, water shortages are still predicted for the future. The limitations of land use in this area are obvious since it is one of the most populous areas in Japan.

Here we formulate the industrial pollution control problem for Osaka City as the following three objective optimization problems:

$$\text{maximize} \quad f_1 = \sum_{j=1}^{n} A_j K_j^{1-b_j} L_j^{b_j} \tag{50}$$

$$\text{minimize} \quad f_2 = \sum_{j=1}^{n} (\omega_{1j}/k_j) K_j \tag{51}$$

$$\text{minimize} \quad f_3 = \sum_{j=1}^{n} (\omega_{2j}/k_j) K_j \tag{52}$$

subject to

$$\sum_{j=1}^{n} (\gamma_{ij}/k_j) K_j \leq \Gamma_i \quad (i=1,2) \tag{53}$$

$$q_2 \leq (\sum_{j=1}^{n} K_j)/(\sum_{j=1}^{n} L_j) \leq q_1 \tag{54}$$

$$\alpha K_{j0} \leq K_j \leq \beta K_{j0} \tag{55}$$

$$\alpha' L_{j0} \leq L_j \leq \beta' L_{j0} \tag{56}$$

where

j an industry ($j=1,\ldots, 20$),

K_j capital value (book value of tangible fixed assets) in industry j,

K_{j0} actual capital value in industry j,

L_j number of employees in industry j,

L_{j0} actual number of employees in industry j,

ω_{ij} unit load of chemical oxygen demand (COD) ($i=1$) or sulphur dioxide (SO_2) ($i=2$) per industrial shipments in industry j,

γ_{ij} resource coefficient for land ($i=1$) or water ($i=2$) per industrial shipments in industry j,

k_j capital coefficient, namely capital value per unit of shipments in industry j,

Γ_i restriction for land ($i=1$) or water ($i=2$),

q_i upper (i=1) or lower (i=2) bound for the overall capital intensity (ratio of total capital value to total number of employees), b_j, A_j parameters of the production function for each industry j, and α, β, α', β' parameters which represent friction (resistance) in the transfer of capital and labour.

The objective function f_1 is a Cobb-Douglas type of production function which is homogeneous of degree one and thus if to each factor the value of its marginal product is paid, total output is distributed between capital and labour in the production $1-b_j$ and b_j, respectively. This value should be maximized so as to increase the total production. The objective function f_2 is the total amount of COD and should be minimized so as to decrease the water pollution. The objective function f_3 is the total amount of SO_2^- and should be minimized so as to decrease the air pollution. Constraints (53) are resource constraint, each of which is land or water resource constraint. Constraint (54) is the technical constraint which shows capital intensity as a whole. This has been utilized to indicate the direction of technological changes occurring as a result of the reformation of the industrial structure in each region. Constraints (55) and (56) are frictional constraints: because drastic changes in the industrial structure are not desirable, frictional coefficients are imposed to provide upper and lower bounds for each decision variable. The problem is to find the optimal allocation of production factors (capital and labour) to each industry under constraints (53) to (56).

The resource restrictions Γ_1 and Γ_2 in the constraints (53) were assumed to be $\Gamma_1 = 232{,}200$, $\Gamma_2 = 200{,}000$. The parameters q_1 and q_2 were supposed to be 1.4 and 0.9 respectively. The parameters for capital and labour, α, α', and β, β', were assumed to be $\alpha = \alpha' = 0.903$, $\beta = \beta' = 1.070$. The parameters for A_j, b_j, k_j and ω_{ij}, γ_{ij} are shown in Table 1

and Table 2, respectively. The code numbers of the industrial classification are explained in Table 3. The sources for these data have been obtained mainly from the Statistical Office of Osaka Prefecture [26], the Ministry of International Trade and Industry [16], and the Osaka Bureau of Trade and Industry [18].

Table 1. Calculated Values of Parameters A_j, b_j and k_j

Industry	A_j	b_j	k_j
1	10.9000	0.1145	0.1195
2	8.6200	0.1391	0.1160
3	15.3900	0.1566	0.0716
4	6.1000	0.1779	0.1599
5	9.9900	0.1723	0.0926
6	5.4600	0.1540	0.1868
7	7.2200	0.2291	0.1824
8	7.9100	0.1294	0.1400
9	6.7300	0.1479	0.1735
10	9.5200	0.1737	0.1125
11	15.2200	0.1445	0.0670
12	6.1300	0.1865	0.1926
13	6.4900	0.1216	0.1746
14	8.1800	0.0870	0.1077
15	6.8500	0.1981	0.1486
16	7.4300	0.2000	0.1659
17	9.6700	0.1588	0.1020
18	7.3600	0.1841	0.1491
19	7.0000	0.2107	0.1394
20	8.4700	0.1677	0.1228

Table 2. Calculated Values of parameters ω_{ij} and γ_{ij}

Industry	COD	SO_2	Land	Water
1	0.07875	0.00822	0.0244	0.0407
2	0.03111	0.02235	0.0718	0.1292
3	0.03110	0.02235	0.0219	0.0072
4	0.00142	0.00076	0.1024	0.0324
5	0.00142	0.00076	0.0244	0.0121
6	0.21680	0.06751	0.0487	0.1564
7	0.07133	0.05218	0.0105	0.0154
8	0.07133	0.05218	0.0429	0.0599
9	0.03466	0.01505	0.1461	0.0212
10	0.02592	0.00413	0.0553	0.0549
11	0.02592	0.00413	0.0468	0.0542
12	0.00198	0.07963	0.1087	0.0617
13	0.00587	0.02136	0.0773	0.0562
14	0.00084	0.03055	0.0354	0.0373
15	0.00116	0.00778	0.0589	0.0293
16	0.00083	0.00340	0.0464	0.0129
17	0.00105	0.00243	0.0235	0.0133
18	0.00073	0.00116	0.0702	0.0267
19	0.00367	0.00228	0.0451	0.0324
20	0.00864	0.00228	0.0354	0.0258

Table 3. Classification of Industries

Code	Industries	Code	Industries
1	Foods	11	Leather products
2	Textile mill products	12	Clay and stone products
3	Apparel products	13	Iron and steel
4	Lumber and products	14	Nonferrous metals
5	Furnitures	15	Fabricated metal products
6	Pulp and paper products	16	Machinery
7	Printing and publishing	17	Electrical machinery
8	Chemicals and products	18	Transportation equipment
9	Coal and petroleum products	19	Precision machinery
10	Rubber products	20	Miscellaneous

Let us now choose f_1 as our primary objective and formulate the corresponding ϵ-constraint problem $P_1(\epsilon_{-1})$, where the sign of f_1 is changed to reflect the maximization process.

$\underline{P_1(\epsilon_{-1})}$

$$\min \quad -f_1 = -\sum_{j=1}^{n} A_j K_j^{1-b_j} L_j^{b_j} \tag{57}$$

subject to

$$\epsilon_i - \sum_{j=1}^{n} (\omega_{ij}/k_j)\cdot K_j \geq 0 \quad i=2,3 \tag{58}$$

$$\Gamma_i - \sum_{j=1}^{n} (\gamma_{ij}/k_j)\cdot K_j \geq 0 \quad i=1,2 \tag{59}$$

$$q_1 \sum_{j=1}^{n} L_j - \sum_{j=1}^{n} K_j \geq 0 \tag{60}$$

$$\sum_{j=1}^{n} K_j - q_2 \sum_{j=1}^{n} L_j \geq 0 \tag{61}$$

and (55), (56).

For illustrative purposes, we shall assume that the DM's structure of preference can be accurately represented by the utility function $U(f_1, f_2, f_3)$ where

$$U(f) = -(f_1+5350000)^2-1000(f_2-140000)^2-5000(f_3-101000)^2 \tag{62}$$

However, it should be stressed that the explicit form of utility function as in (62) is used purely for simulating the responses of the DM. To be more specific, for the assessment decrement Δf_1, the values for $\Delta\bar{f}_i$, Δf_{imin} and Δf_{imax} are simulated by solving the following equations.

$$U(f_1+\Delta f_1,\ldots, f_i-\Delta\bar{f}_i,\ldots, f_n) = U(f_1, f_2,\ldots, f_n) \tag{63}$$

$$U(f_1-\Delta f_1,\ldots, f_i+\Delta f_{imin},\ldots, f_n) = U(f_1, f_2,\ldots, f_n) + \delta_4 \tag{64}$$

$$U(f_1-\Delta f_1,\ldots, f_i+\Delta f_{imax},\ldots, f_n) = U(f_1, f_2,\ldots, f_n) - \delta_4 \tag{65}$$

Where the tolerance δ_4 is a prescribed positive number. Concerning L and R functions, max (0, 1-|x|) is choosed in a subjective manner.

In Fig. 1, the interaction processes using the time-sharing computer program under TSS of ACOS-6 digital computer in the computer center of Kobe University in Japan are explained especially for the first iteration through the aid of some of the computer outputs. In this interaction, the initial values of decision variables, $x \triangleq (K_1, \ldots, K_{20}, L_1, \ldots, L_{20})$, are set to be the lower bounds of them, the initial values of $\varepsilon_{-1}^1 = (\varepsilon_2^1, \varepsilon_3^1)$ are choosed to be (144000, 103000) by taking account of the calculated individual minimum and maximum of f_2 and f_3. Furthermore, the assessment decrement Δf_1 meaningful to the DM is assumed to be $\Delta f_1 = 1000$ and the value of δ_4 is set to be 20000000.

Pareto optimal solutions are obtained by solving the ε-constraint problems using the revised version of the generalized reduced gradient (GRG) [14] program called GRG2[15]. In GRG2 there are two optimality test, i.e.:

(i) to satisfy the Kuhn-Tucker optimality conditions,

(ii) to satisfy the fractional change condition

$$|FM - OBJTST| < EPSTOP \times |OBJTST|$$

for NSTOP times consecutive iterations. FM is the current objective value and OBJTST is the objective value at the start of the previous one dimentional search. NSTOP has a default value of 3.

In Fig. 1, it is shown that one of these conditions are satisfied.

After testing whether the trial point obtained at ITERATION 1 is optimal or not, the direction vector is determined and the same procedure continues in this manner, where the sum-of-logarithms proxy preference function is selected through the whole searching process.

In this example, at the 3rd iteration the termination criteria is satisfied and the preferred values of the objectives and decision variables are shown in Fig. 2. CPU time required in this interaction processes was 79.56 seconds under TSS of ACOS-6 digital computer.

Fig. 1. Interactive Decision Making Processes

```
COMMAND?
=SPOT

INPUT NUMBER OF OBJECTIVES :
=3

LOWER BOUNDS ON VARIABLES ARE
   1) 0.285790E+05   2) 0.207490E+05   3) 0.913200E+04   4) 0.121700E+05
   5) 0.774800E+04   6) 0.334030E+05   7) 0.682540E+05   8) 0.780470E+05
   9) 0.180900E+04  10) 0.466000E+04  11) 0.339900E+04  12) 0.140290E+05
  13) 0.975460E+05  14) 0.259580E+05  15) 0.680240E+05  16) 0.736230E+05
  17) 0.276980E+05  18) 0.327640E+05  19) 0.413300E+04  20) 0.237160E+05
  21) 0.203490E+05  22) 0.158200E+05  23) 0.163320E+05  24) 0.743700E+04
  25) 0.747100E+04  26) 0.144830E+05  27) 0.396820E+05  28) 0.308440E+05
  29) 0.747000E+03  30) 0.378800E+04  31) 0.497700E+04  32) 0.764900E+04
  33) 0.261520E+05  34) 0.916200E+04  35) 0.476270E+05  36) 0.472740E+05
  37) 0.241400E+05  38) 0.167910E+05  39) 0.374500E+04  40) 0.204970E+05

UPPER BOUNDS ON VARIABLES ARE
   1) 0.338560E+05   2) 0.245800E+05   3) 0.108180E+05   4) 0.144170E+05
   5) 0.917800E+04   6) 0.395710E+05   7) 0.808560E+05   8) 0.924560E+05
   9) 0.214300E+04  10) 0.552000E+04  11) 0.402600E+04  12) 0.166190E+05
  13) 0.115555E+06  14) 0.307510E+05  15) 0.805830E+05  16) 0.872160E+05
  17) 0.328120E+05  18) 0.388130E+05  19) 0.489600E+04  20) 0.280940E+05
  21) 0.257830E+05  22) 0.187400E+05  23) 0.193470E+05  24) 0.881000E+04
  25) 0.885100E+04  26) 0.171570E+05  27) 0.470080E+05  28) 0.365390E+05
  29) 0.885000E+03  30) 0.448700E+04  31) 0.589600E+04  32) 0.906200E+04
  33) 0.309800E+05  34) 0.108530E+05  35) 0.564200E+05  36) 0.560020E+05
  37) 0.285970E+05  38) 0.198910E+05  39) 0.443700E+04  40) 0.242800E+05

NO UPPER BOUNDS ON INEQUALITY CONSTRAINTS


INITIAL X IS
   1) 0.285790E+05   2) 0.207490E+05   3) 0.913200E+04   4) 0.121700E+05
   5) 0.774800E+04   6) 0.334030E+05   7) 0.682540E+05   8) 0.780470E+05
   9) 0.180900E+04  10) 0.466000E+04  11) 0.339900E+04  12) 0.140290E+05
  13) 0.975460E+05  14) 0.259580E+05  15) 0.680240E+05  16) 0.736230E+05
  17) 0.276980E+05  18) 0.327640E+05  19) 0.413300E+04  20) 0.237160E+05
  21) 0.203490E+05  22) 0.158200E+05  23) 0.163320E+05  24) 0.743700E+04
  25) 0.747100E+04  26) 0.144830E+05  27) 0.396820E+05  28) 0.308440E+05
  29) 0.747000E+03  30) 0.378800E+04  31) 0.497700E+04  32) 0.764900E+04
  33) 0.261520E+05  34) 0.916200E+04  35) 0.476270E+05  36) 0.472740E+05
  37) 0.241400E+05  38) 0.167910E+05  39) 0.374500E+04  40) 0.204970E+05

                INDIVIDUAL MINIMUM AND MAXIMUM

          I          MIN          I          MAX
     ---------+--------------------+--------------------
      F(1)    I  -0.50209829E+07   I  -0.45931085E+07
      F(2)    I   0.14327542E+06   I   0.16291024E+06
      F(3)    I   0.10174714E+06   I   0.11416334E+06

INPUT INITIAL VALUES OF EPSILONS ( EP(I),I=2,3 )  :
=144000. 103000.
```

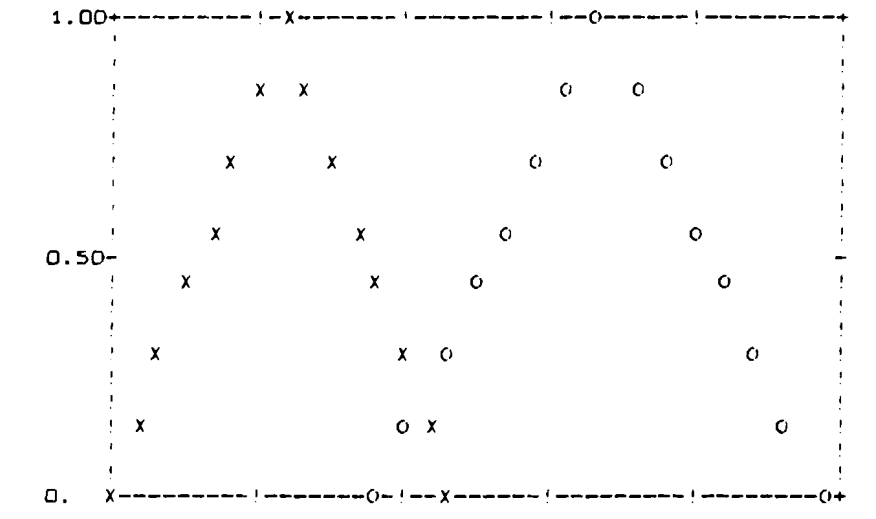
```
DO YOU USE IDEAL DM ? (YES OR NO) :
=YES
WHICH METHOD DO YOU USE TO SIMULATE MRS ? ( 1, 2 OR 3 ) :
 1  DIRECTLY BY DEFINITION
 2  MRS SUBROUTINE THROUGH A SERIES OF ORDINAL COMPARISON
 3  IN A FUZZY FORM
=3

PARETO OPTIMAL SOLUTION FOR INITIAL EPSILONS
( KUHN-TUCKER CONDITIONS SATISFIED )

          1      F(1)  = -0.48677483E+07
          2      F(2)  =  0.14400000E+06
                 EP(2) =  0.14400000E+06
                 LAGRANGE MULTIPLIER =  0.30927132E+02
          3      F(3)  =  0.10300000E+06
                 EP(3) =  0.10300000E+06
                 LAGRANGE MULTIPLIER =  0.32025648E+02

INPUT DECREMENT DF(1) :
=1000.

INPUT DELTA4 :
=20000000.

INPUT INCREMENT DF(2) IN A FUZZY FORM.
NAMELY, DF(2) IS A FUZZY NUMBER BETWEEN DF(2)MIN AND DF(2)MAX
WITH THE MEAN IN THE VICINITY OF DF(2)MEAN
 =       116.       119.       121.

INPUT INCREMENT DF(3) IN A FUZZY FORM.
NAMELY, DF(3) IS A FUZZY NUMBER BETWEEN DF(3)MIN AND DF(3)MAX
WITH THE MEAN IN THE VICINITY OF DF(3)MEAN
 =        47.        48.        49.


L-R FUZZY NUMBER C(1) IS
    C(1)MEAN                =  0.85794807E+00
    LEFT SPREAD PARAMETER   =  0.70396781E-01
    RIGHT SPREAD PARAMETER  =  0.70396187E-01

----------------------------<ITERATION 1>----------------------

DIRECTION VECTOR TO UPDATE EPSILONS IS
     S(2) =       0.22500927E+02
     S(3) =       0.11021009E+02

INPUT INITIAL STEP SIZE ( ALFO ) :
=1.
YOUR MARGINAL RATES OF SUBSTITUTION ARE :
        M(1,2) =  0.84262046E+01
        M(1,3) =  0.21004639E+02
```

```
SELECT LOCAL PROXY PREFERENCE FUNCTION ( 1,2 OR 3 ) :
1    SUM OF EXPONENTIALS
2    SUM OF POWERS
3    SUM OF LOGARITHMS
=3


INPUT VALUE OF M(I) SUCH THAT M(I)-F(I)>0 (I=1,3) :
=0. 147000. 107000.

LOCAL PROXY PREFERENCE FUNCTION
P(F)=
     + 0.10000000E+01*LOG( 0.              -F(1))
     + 0.51930816E-02*LOG( 0.14700000E+06-F(2))
     + 0.17260251E-01*LOG( 0.10700000E+06-F(3))

PARATE OPTIMAL SOLUTION (ALF=0.100E+01)
( KUHN-TUCKER CONDITIONS SATISFIED )

          1      F(1)   = -0.48687964E+07
          2      F(2)   =  0.14402250E+06
                 EP(2)  =  0.14402250E+06
                 LAGRANGE MULTIPLIER =  0.30850401E+02
          3      F(3)   =  0.10301102E+06
                 EP(3)  =  0.10301102E+06
                 LAGRANGE MULTIPLIER =  0.32043529E+02

PARATE OPTIMAL SOLUTION (ALF=0.200E+01)
( KUHN-TUCKER CONDITIONS SATISFIED )

          1      F(1)   = -0.48698429E+07
          2      F(2)   =  0.14404500E+06
                 EP(2)  =  0.14404500E+06
                 LAGRANGE MULTIPLIER =  0.30774908E+02
          3      F(3)   =  0.10302204E+06
                 EP(3)  =  0.10302204E+06
                 LAGRANGE MULTIPLIER =  0.32061074E+02

INPUT THE MAXIMUM STEP SIZE (ALFMAX)
=1000.

P(F) IS INCREASING ; STEP SIZE WILL BE DOUBLED
PARETO OPTIMAL SOLUTION (ALF=0.400E+01)
( KUHN-TUCKER CONDITIONS SATISFIED )

          1      F(1)   = -0.48719316E+07
          2      F(2)   =  0.14409000E+06
                 EP(2)  =  0.14409000E+06
                 LAGRANGE MULTIPLIER =  0.30627486E+02
          3      F(3)   =  0.10304408E+06
                 EP(3)  =  0.10304408E+06
                 LAGRANGE MULTIPLIER =  0.32095189E+02
```

```
P(F) IS INCREASING ; STEP SIZE WILL BE DOUBLED
PARETO OPTIMAL SOLUTION (ALF=0.800E+01)
( KUHN-TUCKER CONDITIONS SATISFIED )

          1      F(1)   = -0.48760508E+07
          2      F(2)   =  0.14418001E+06
                 EP(2)  =  0.14418001E+06
                 LAGRANGE MULTIPLIER =  0.28823283E+02
          3      F(3)   =  0.10308817E+06
                 EP(3)  =  0.10308817E+06
                 LAGRANGE MULTIPLIER =  0.32578238E+02

P(F) IS INCREASING ; STEP SIZE WILL BE DOUBLED
PARETO OPTIMAL SOLUTION (ALF=0.160E+02)
( KUHN-TUCKER CONDITIONS SATISFIED )

          1      F(1)   = -0.48840408E+07
          2      F(2)   =  0.14436001E+06
                 EP(2)  =  0.14436001E+06
                 LAGRANGE MULTIPLIER =  0.27952277E+02
          3      F(3)   =  0.10317634E+06
                 EP(3)  =  0.10317634E+06
                 LAGRANGE MULTIPLIER =  0.32792177E+02

P(F) IS INCREASING ; STEP SIZE WILL BE DOUBLED
PARETO OPTIMAL SOLUTION (ALF=0.320E+02)
( KUHN-TUCKER CONDITIONS SATISFIED )

          1      F(1)   = -0.48953261E+07
          2      F(2)   =  0.14472003E+06
                 EP(2)  =  0.14472003E+06
                 LAGRANGE MULTIPLIER =  0.10534637E+02
          3      F(3)   =  0.10335267E+06
                 EP(3)  =  0.10335267E+06
                 LAGRANGE MULTIPLIER =  0.37522127E+02

P(F) IS INCREASING ; STEP SIZE WILL BE DOUBLED
PARETO OPTIMAL SOLUTION (ALF=0.640E+02)
( KUHN-TUCKER CONDITIONS SATISFIED )

          1      F(1)   = -0.49161106E+07
          2      F(2)   =  0.14544006E+06
                 EP(2)  =  0.14544006E+06
                 LAGRANGE MULTIPLIER =  0.10490666E+02
          3      F(3)   =  0.10370534E+06
                 EP(3)  =  0.10370534E+06
                 LAGRANGE MULTIPLIER =  0.37420754E+02
```

```
P(F) IS INCREASING ; STEP SIZE WILL BE DOUBLED
PARETO OPTIMAL SOLUTION (ALF=0.128E+03)
( KUHN-TUCKER CONDITIONS SATISFIED )

          1      F(1)   = -0.49550901E+07
          2      F(2)   =  0.14688012E+06
                 EP(2)  =  0.14688012E+06
                 LAGRANGE MULTIPLIER =  0.98469370E+01
          3      F(3)   =  0.10441069E+06
                 EP(3)  =  0.10441069E+06
                 LAGRANGE MULTIPLIER =  0.31313919E+02

MAXIMUM OF P(F) IS BRACKETED ( NEAR OPTIMAL STEP SIZE IS 0.640E+02 )


PARETO OPTIMAL SOLUTION FOR OPTIMAL STEP SIZE
( KUHN-TUCKER CONDITIONS SATISFIED )

          1      F(1)   = -0.49161106E+07
          2      F(2)   =  0.14544006E+06
                 EP(2)  =  0.14544006E+06
                 LAGRANGE MULTIPLIER =  0.10490666E+02
          3      F(3)   =  0.10370534E+06
                 EP(3)  =  0.10370534E+06
                 LAGRANGE MULTIPLIER =  0.37420754E+02



INPUT DECREMENT DF(1) :
=1000.

INPUT DELTA4 :
=20000000.

INPUT INCREMENT DF(2) IN A FUZZY FORM.
NAMELY, DF(2) IS A FUZZY NUMBER BETWEEN DF(2)MIN AND DF(2)MAX
WITH THE MEAN IN THE VICINITY OF DF(2)MEAN
 =         77.        79.        81.

INPUT INCREMENT DF(3) IN A FUZZY FORM.
NAMELY, DF(3) IS A FUZZY NUMBER BETWEEN DF(3)MIN AND DF(3)MAX
WITH THE MEAN IN THE VICINITY OF DF(3)MEAN
 =         31.        32.        33.

L-R FUZZY NUMBER C(2) IS
    C(2)MEAN                =  0.98805211E+00
    LEFT SPREAD PARAMETER   =  0.90695848E-01
    RIGHT SPREAD PARAMETER  =  0.90695957E-01
```

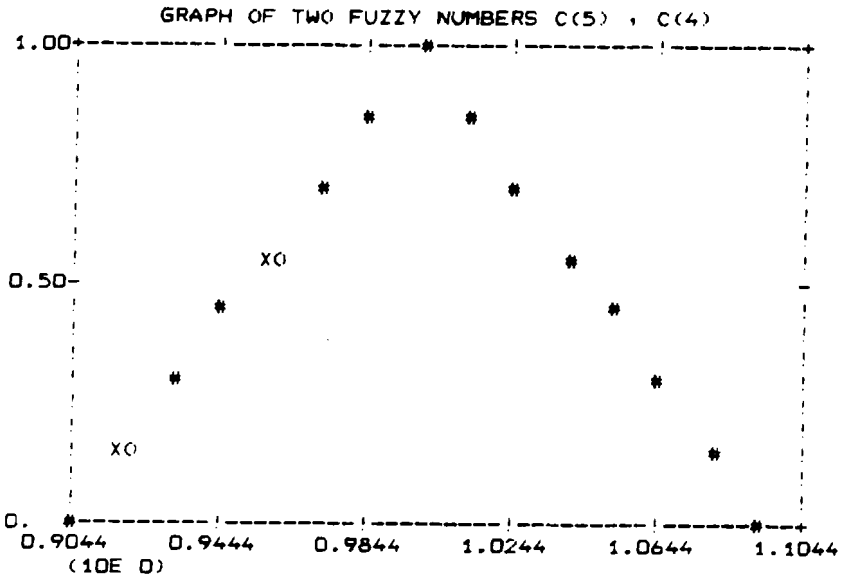
```
COMPARISON BETWEEN TWO FUZZY NUMBERS C(2) AND C(1)

THE ANSWER TO THE QUESTION [ IS C(2) GREATER THAN C(1) ? ]
IS A FUZZY SET A OF THE UNIVERSE (YES OR NO) :
  A =  0.100E+01/YES + 0.192E+00/NO

            GRAPH OF TWO FUZZY NUMBERS C(2) , C(1)
```




```
                O  :  C(2)
                X  :  C(1)
                #  :  C(2) ; C(1)  (SAME VALUE)


COMMAND?
=PROCEED

-----------------------------<ITERATION 2>-----------------

DIRECTION VECTOR TO UPDATE EPSILONS IS
     S(2) =       -0.21529407E+01
     S(3) =        0.60255409E+01

INPUT INITIAL STEP SIZE ( ALFO ) :
=1.
YOUR MARGINAL RATES OF SUBSTITUTION ARE :
        M(1,2) =  0.12643606E+02
        M(1,3) =  0.31395213E+02
```

Fig. 2. The Preferred Solution of the DM

```
COMMAND?
=STOP

THE FOLLOWING VALUES ARE YOUR PREFERRED SOLUTION

PREFERRED VALUES OF OBJECTIVES
     F(1) =       -0.49253282E+07
     F(2) =        0.14483814E+06
     F(3) =        0.10412229E+06

PREFERRED VALUES OF VARIABLES
X( 1)= 0.28841357E+05   X( 2)= 0.20749000E+05   X( 3)= 0.91320000E+04
X( 4)= 0.14417000E+05   X( 5)= 0.91780000E+04   X( 6)= 0.33403000E+05
X( 7)= 0.68254000E+05   X( 8)= 0.78047000E+05   X( 9)= 0.18090000E+04
X(10)= 0.55200000E+04   X(11)= 0.40260000E+04   X(12)= 0.14029000E+05
X(13)= 0.10623457E+06   X(14)= 0.25958000E+05   X(15)= 0.80583000E+05
X(16)= 0.87216000E+05   X(17)= 0.32812000E+05   X(18)= 0.38813000E+05
X(19)= 0.48960000E+04   X(20)= 0.28094000E+05   X(21)= 0.25783000E+05
X(22)= 0.18740000E+05   X(23)= 0.19347000E+05   X(24)= 0.88100000E+04
X(25)= 0.88510000E+04   X(26)= 0.17157000E+05   X(27)= 0.47008000E+05
X(28)= 0.36539000E+05   X(29)= 0.88500000E+03   X(30)= 0.44870000E+04
X(31)= 0.58960000E+04   X(32)= 0.90620000E+04   X(33)= 0.30980000E+05
X(34)= 0.10853000E+05   X(35)= 0.56420000E+05   X(36)= 0.56002000E+05
X(37)= 0.28597000E+05   X(38)= 0.19891000E+05   X(39)= 0.44370000E+04
X(40)= 0.24280000E+05
```




```
DIRECTION VECTOR IS
     S(2) =       -0.93272592E+00
     S(3) =        0.24395976E+00
```

The preferred allocation of capital and labour to each industry corresponding to the results obtained by interaciton is summarized in Table 4 together with the values in 1975.

Table 4. The preferred Allocation of Capital and Labour

Industry	1975		Proposal	
	Capital	Labour	Capital	Labour
1	31653	22527	28841	25783
2	22981	17521	20749	18740
3	10114	18088	9132	19347
4	13479	8237	14417	8810
5	8581	8275	9178	8851
6	36996	16041	33403	17157
7	75595	43494	68254	47008
8	86440	34161	78047	36539
9	2004	827	1809	885
10	5161	4195	5520	4487
11	3764	5512	4026	5896
12	15538	8472	14029	9026
13	108036	28964	106235	30980
14	28750	10147	25958	10853
15	75339	52749	80583	56420
16	81541	52358	87216	56002
17	30677	26736	32812	28597
18	32687	18597	38813	19891
19	4577	4148	4896	4437
20	26266	22701	28094	24280

The obtained result compares favorable with the values obtained by solving $\max_{x \in X} U(f_1, f_2, f_3)$ directly based on (62) using GRG2. These values are $(f_1, f_2, f_3) = (4924409, 144676, 104144)$.

The preferred values for the objective functions can be interpreted as the compromised values of the DM between the conflicting objectives which are the maximization of the production function and the minimization of two environmental factors (COD and SO_2). The preferred solutions for the decision vatiables K_j and L_j show the preferred allocation of capital and labour to each industry. These results show that capital values in industry as a whole is reduced compared with the values in 1975. Especially, capital formation in the coal and petroleum industry and in the chemicals and related products industry is severely reduced and the nonferrous metals industry and the fabricated metal product industry decrease their capital formation. On the other hand, in consumer industries such as the lumber and furniture industries, as well as in machine industries such as the electical machinery industry, capital formation is promoted.

4. Conclusion

In this paper, we introduced the revised version of SPOT called fuzzy sequential proxy optimization technique (FSPOT) in order to deal with inaccuracies of the DM's judgements in interactive multiobjective optimization problems. In our interactive on-line scheme, after solving the ε-constraint problem the values of DM's MRS assessed in a fuzzy form were interpreted as type L-R fuzzy numbers and the mean value of the MRS were used to determine the direction and the local proxy preference function was updated to determine the step size. Pareto optimality of the generated solution in each iteration is also guaranteed in our technique. Based on the algorithm of FSPOT, the time-sharing computer program has been written to facilitate the interactive processes.

An application to the industrial pollution control problem in Osaka City demonstrated the feasibility and efficiency of both the proposed technique and its interactive computer program by simulating the responses

of the hypothetical DM. Although the actual DM for the formulated problem would of course select other values of the three objectives than the ones which were selected by the hypothetical DM used in this paper, the way to iterate and calculate is essentially the same. However, further applications must be carried out in cooperation with a person actually involved in decisionmaking. From such experiences the proposed technique and its computer program must be revised. We hope that the proposed technique and its extension will become efficient tools for man-machine interactive decisionmaking under multiple conflict objectives.

Acknowledgements

The authors are indebted to Professor Yoshikazu Sawaragi of Kyoto Sangyo University for his constant encouragement. The authors also wish to thank Mr. H. Yano for his cooperation in this study.

References

1. Baptistella, L.F.B. and Ollero, A., "Fuzzy Methodologies for Interactive Multicriteria Optimizations," IEEE Trans. Syst., Man, Cyber., Vol. SMC 10, (1980), PP.355-365.
2. Chankong, V. and Haimes, Y.Y., "The Interactive Surrogate Worth Trade-off (ISWT) Method for Multiobjective Decision-making," In S. Zionts (Ed.), Multiple Criterion Problem Solving, New York: Springer-Verlag (1978).
3. Cohon, J.L. and Marks, D.H., "A Review and Evaluation of Multiobjective Programming Technique," Water Resources Research, Vol. 11, (1975), PP.208-220.
4. Dubois, D. and Prade, H., "Operations on Fuzzy Numbers," Int. J. Syst. Sci., Vol. 9, (1978), PP. 613-626.
5. ——, "Fuzzy Real Algebra: Some Results," Fuzzy Sets and Syst., Vol. 2, (1979), PP.327-348.

6. ——, "Systems of Linear Fuzzy Constraints," Fuzzy Sets and Syst., Vol. 3, (1980), PP.37-48.

7. Dyer, J.S., "An Empirical Investigation of a Man-machine Interactive Approach to the Solution of the Multiple Criteria Problem," In J.L. Chochrane and M. Zeleny (Ed.), Multiple Criteria Decision Making, Columbia, SC: Univ. of South Carolina, (1973).

8. Gaines , B.R., "Foundations of Fuzzy Reasoning," Int. J. Man-Machine Studies, Vol. 8, (1976), PP.623-668.

9. Geoffrion, A.M., Dyer J.S. and Feinberg A., "An Interactive Approach for Multi-criterion Optimization with an Application to the Operation of an Academic Department," Management Sci., Vol. 19, (1972), PP.357-368.

10. Haimes, Y.Y., Hall, W.A. and Freedman, H.T., Multiobjective Optimization in Water Resources Systems: The Surrogate Worth Trade-off Method, Elsevier, New York, 1975.

11. Haimes, Y.Y., Hierarchical Analyses of Water Resources Systems, McGraw-Hill, New York, 1977.

12. Haimes, Y.Y. and Chankong, V., "Kuhn-Tucker Multipliers as Trade-offs in Multiobjective Decision-making Analysis," Automatica, Vol. 15, (1979), PP.59-72.

13. Keeney, R.L. and Raiffa H., Decisions with Multiple Objectives: Preferences and Value Trade-offs, Wiley, New York, 1976.

14. Lasdon, L.S. Fox, R.L. and Ratner, M.W., "Nonlinear Optimization Using the Generalized Reduced Gradient Method," Revue Francaise d'Automatique, Informatique et Recherche Operationnelle, Vol. 3, (1974), PP.73-103.

15. Lasdon, L.S., Waren, A.D. and Ratner, M.W., "GRG2 User's Guide," Technical Memorandum University of Texas, 1980.

16. Ministry of International Trade and Industry, Census of Manufactures 1973 Report on Industrial Land and Water, Research and Statistics Department, Minister's Secretariat, Ministry of International Trade and Industry, Tokyo, 1973. (in Japanese).

17. Oppenheimer, K.R., "A Proxy Approach to Multi-attribute Decision Making," Management Sci. Vol. 24, (1978). PP.675-689.

18. Osaka Bureau of Trade and Industry, Input-Output Table for Industrial Pollution Analysis in the Kinki Area in 1973, Osaka Bureau of Trade and Industry, Osaka, 1973. (in Japanese).

19. Payne, H.J., Polak, E., Collins, D.C. and Meisel W.S., "An Algorithm for Bicriteria Optimization Based on the Sensitivity Function," IEEE Trans. Automat. Contr., Vol. AC-20, (1975), PP.546-548.

20. Sakawa, M., "Interactive Multiobjective Decision Making by the Sequential Proxy Optimization Technique: SPOT," Working Paper WP-80-66, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1980.

21. ——, "A Computer Program for Multiobjective Decision Making by the Interactive Sequential Proxy Optimization Technique," Working Paper WP-80-77, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1980.

22. ——, "An Interactive Computer Program for Multiobjective Decision Making by the Sequential Proxy Optimization Technique," Int. J. Man-Machine Studies, Vol. 14, (1981), PP.193-213.

23. Sakawa, M. and Seo, F., "Interactive Multiobjective Decision Making in Environmental Systems Using Sequential Proxy Optimization Techniques (SPOT)," Automatica, Vol. 18, (1982) (to appear).

24. ——, "Interactive Multiobjective Decisionmaking for Large-Scale Systems and Its Application to Environmental Systems," IEEE Trans. on Syst. Man, Cyber., Vol. SMC-10, (1980), PP.796-806.

25. Shimizu, K., Kawanabe, H. and Aiyoshi, E., "A Theory for Interactive Preference Optimization and Its Algorithm - Generalized SWT Method -" Trans. Institute of Elec., Commun. Eng. Japan, Vol. J61-A, (1978), PP.1075-1082 (in Japanese).

26. Statistical Office of Osaka Prefecture, Industrial Statistics Survey Result Table of Industrial Statistics in 1975, Statistical Office of Osaka Prefecture, Osaka, 1973 (in Japanese).

27. Wierzbicki, A.P., "A Methodological Guide to Multiobjective Optimization," Working Paper WP-79-122, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1979.

28. Zadeh L., "Fuzzy Sets," Inform. Contr., Vol. 8, (1965), PP.338-353.

29. Zadeh, L., "Fuzzy Algorithms," Inform. Contr., Vol. 19, (1968), PP.94-102.

30. ——, "Outline of a New Approach to the Analysis of Complex Systems and Decision Processes," IEEE Trans. Syst., Man, Cyber., Vol. SMC-3, (1973), PP.28-44.

31. Zeleny, M., "The Theory of the Displaced Ideal," In M. Zeleny (Ed.), Multiple Criteria Decision Making, Kyoto 1975, Springer-Verlag, New York, 1976.

32. Zionts, S. and Wallenius, J., "An Interactive Programming Method for Solving the Multiple Criteria Problem," Management Sci. Vol. 22, (1976), PP.652-663.

MATHEMATICAL OPTIMIZATION FOR MULTIOBJECTIVE DECISION MAKING

F. Seo[1] and M. Sakawa[2]

[1]*Kyoto Institute of Economic Research, Kyoto University, Japan*
[2]*Department of Systems Engineering, Kobe University, Japan*

I. Introduction

Multiobjective optimization techniques have been developed during the past ten years. Evaluation criteria for the solution techniques have been also presented. Loucks (1975) paid attention to conflict and choice problems among objectives in an uncertain environment. In particular he pointed out the importance of recognizing uncertainty in trade-offs and preferences. However he was rather in line with pervasive economic literature and emphasized the importance of simulation and prediction techniques for the bargaining and the decision-making processes. Cohon and Marks (1975), due to the undeveloped situations of the above techniques, proposed the following criteria: (1) computational feasibility and efficiency for practical use, (2) explicit quantification of the trade-offs among objectives, (3) sufficient information about the noninferior solution set. Those criteria are all valid and useful. In particular, explicit inclusion of a responsive decision-making process for quantifying value judgement is of great value. However, from the operational point of view, "the more the better" rule for information available to the noninferior solutions set is in need of examination. In this paper, the author proposes alternative criteria for multiobjective optimization, and presents a method for multiobjective decision making under uncertainty.

The multiobjective decision-making process for a purposeful system has two phases, analytical and judgemental. In the analytical phase, mathematical optimization can be executed along with modeling and simulation. In the judgemental phase, a responsive decision-making process for intervention and coordination should be carried out. Generally the decision-making process has a gap between these two phases, and this paper is concerned with presenting a method to bridge this gap. Duality of mathematical programming is used for basic systems evaluation and combined with the decision analysis technique. Devices for treating uncertain evaluation are also suggested.

II. Characteristics of Multiobjective Optimization

1. Multiobjective decision problem

In general, an overall multiobjective optimization problem is considered in the following form:

MOP

$$\underset{x \in X}{\text{Maximize}} \quad \{f_1(x), f_2(x), \ldots, f_m(x)\} \tag{1}$$

where $f_i : R^n \rightarrow R^1$, $i = 1, \ldots, m$, is a criterion function (or objective function) of an n-dimentional decision vector x. X is a constraints set of feasible decisions.

$$x \in X = \{x \mid x \in R^n, g_j(x) \leq 0, j = 1, \ldots, \ell\} \tag{2}$$

In problem (1), m-objective functions are usually noncommensurate and in conflict with each other. Thus it is impossible to find directly the superior solution for the problem (1). Instead the preferred solution x^* of the multiobjective optimization problem is obtained from among the Pareto-optimal or noninferior solution set X^P. The noninferior solution set X^P is defined as follows:

$$X^P = \{x^P \mid x^P \in X, \ \nexists\, x \in X, \ x \neq x^P, \text{ such that } f_r(x) \geq f_r(x^P) \text{ for } \forall r \in I = [1, 2, \ldots, m], \text{ and } f_k(x) > f_k(x^P) \text{ for } k \in I, \ k \neq r\}. \tag{3}$$

For solving the overall optimization problem (1), an overall criteria function should be maximized. Thus the MOP is converted to the multiobjective decision problem in the following form.

MDP

$$\underset{x \in X}{\text{Maximize}} \quad U\{f_1(x), f_2(x), \ldots, f_m(x)\} \tag{4}$$

The function $U: R^m \rightarrow R^1$ in problem (4) is an overall preference function defined on all the values of the multiple criteria function $\{f_i(x)\}$. The preference function U is usually unknown, and so we are concerned with finding a rational procedure to derive the preference function. Thus procedural as well as substantive rationality should be considered (Simon 1978).

2. Brief review for the solution techniques

During recent years, several approaches to solve the MDP (4) have been presented. Those approaches are classified into three categories.

(1) Generating non-inferior set techniques.

The first is the generating techniques for identifying the noninferior or Pareto-efficiency set. Although the preference function is generally unknown, the decision-maker can find the preferred solution if he knows the preferred marginal rates of substitution between objectives from among those on the Pareto-efficiency frontier.

The marginal rate of substitution (MRS) for the preference function U is defined as follows:

$$m_{ij}(f) = [\partial U(f)/\partial f_j]/[\partial U(f)/\partial f_i] = -\frac{df_i(x)}{df_j(x)} \Big| du = 0, \ df_r = 0, \ r \neq i, j \tag{5}$$

where $f = \{f_1, \ldots, f_m\}$.

For the prefered solution, the MRS m_{ij} corresponds to the trade-off rate, $T_{ij} = -df_i^P(x)/df_j^P(x)$, on the Paretian frontier. The Paretian frontier can be mathematically derived.

(i) The weighting or parametric method formulates the MDP (4) as follows.

$$\underset{x \in X}{\text{Maximize}} \quad W(f) = w_1 f_1(x) + w_2 f_2(x) + \ldots + w_m f_m(x) \tag{6}$$

From the extremal condition,

$$dW\Big|_{\substack{df_r = 0 \\ r \neq i,\, j}} = w_i df_i + w_j df_j = 0 . \tag{7}$$

At optimal,

$$-\frac{df_i}{df_j} = \frac{w_j^*}{w_i^*} . \tag{8}$$

Thus the weighting or parametric method finds the prefered solution with the weighting coefficient w_i, $i = 1, \ldots, m$, for determining the preferred marginal rates of substitution. However this method has no device for finding if the Paretian trade-off rate T_{ij} corresponds to the MRS m_{ij} on the preference function. The weighting parameter w_i should be set as given. For example, in the conventional multi-sectoral optimization problem for economic activity such as maximization of a gross national income, market prices are used as the weighting coefficient for each sectoral income. Dorfman (1972) treated the weighting coefficient as the "political weight" in his net benefit maximization problem for regional water quality management.

(ii) The constraint method is a counterpart of the weighting method. The ε-constraint method with parametric variation of the constraint values can derive the noninferior set without the requirement of convexity for preference and constraint functions. The ε-constraint method formulates the MDP in the following form.

$$\underset{x \in X}{\text{Maximize}} \quad f_r(x) \tag{9}$$

$$\text{subject to} \quad f_k(x) \geq \varepsilon_k, \quad k = 1, \ldots, m, \quad k \neq r \tag{10}$$

where $\varepsilon_k = \bar{f}_k + \varepsilon'_k$, $\varepsilon'_k > 0$ and $\bar{f}_k$ is an minimized value of an objective function f_k.

In this method, although the weighting parameters corresponding to the Pareto optimality are given analytically as a result of mathematical optimization, a process of searching for the prefered weights corresponding to the preferred solution is not provided.

Marglin (1962, 1967) used this method for his bi-criteria problem for economic efficiency and income redistribution. In the place of market prices and "political weights," he used the pre-assigned marginal opportunity cost or Lagrange multiplier as the weighting parameter.

(2) Prior articulation of preferences and its extension

Goal programming techniques have been long developed based on prior articulation of preferences, and recently interactive versions have been presented (Dyer 1972, 1973). A device for combining goal programming techniques with multiattribute utility functions for pre-assigning the desirable goal is also provided (Dyer 1977). His study is based on assumption of the additively separable utility function. Those devices are intended to articulate directly the desirable goal with the preference function without any regard to deriving the noninferior solution set, and to seek the most efficient path to the ideal goal.

(3) Interactive derivation of the preference function

The methods for deriving the preference function U in the MDP (4) with interactive processes have two directions. One is to obtain the non-inferior solution set and, based on evaluation of the trade-off rate functions corresponding to it, to assess the surrogate worth functions for seeking the most preferable solution (Haimes 1974, 1975). An interactive

version for assessing the marginal rates of substitution has also been presented (Chankong and Haimes 1978). This method is called Surrogate Worth Trade-Off (SWT) method. The main characteristic of this method is composed of two steps. In step 1, a scalar optimization problem is solved in the forms of the ε-constraint or Lagrange-type weighting method. The dual optimal solution $\lambda_{ij}(f_j(x)) = -\dfrac{\partial f_i(x)}{\partial f_j(x)} > 0$ is used as the trade-off rate function between objectives f_i and f_j. In step 2, the surrogate worth function $w_{ij}(\lambda_{ij})$ is assessed as an ordinal number and the preferred solution is chosen which corresponds to the preferable trade-off rate function to which the decision maker will be indifferent. Thus, in the SWT method, the analytical process of mathematical optimization is combined with a judgemental process. However any device for manipulating the "phase gap" has not been provided. Evaluation of the marginal rates of substitution is straightforward, and, even though it is interactive, wholly depends on the subjective appraisal of the DM.

Another direction for interactive derivation of the preference function is multiobjective decision analysis (Keeney 1974, Ostrom and Gros 1975, Keeney and Raiffa 1976). In the following section, the main idea of multiattribute decision analysis is discussed.

3. Multiobjective decision analysis

Multiobjective decision analysis proceeds in the following way. First, the overall multiobjective optimization problem (MOP) (1) is decomposed to m-subsystems.

DMOP

$$\underset{x_i \in X}{\text{Maximize}} \quad \{f_1(x_1), f_2(x_2), \ldots, f_m(x_m)\} \tag{11}$$

where x_i is an n_i-dimensional decision vector in a subsystem i, $i = 1, 2, \ldots, m$.

To manipulate the noncommensurateness and conflict in problem (11), consider a decomposed multiobjective decision problem in the following form:

DMDP

$$\underset{x_i \in X}{\text{Maximize}} \quad \{u_1(f_1(x_1)),\ u_2(f_2(x_2)),\ldots,\ u_m(f_m(x_m))\} \tag{12}$$

Decision problem in each subsystem i is described in terms of utility function $u_i(f_i(x_i))$, and the argument $f_i(x_i)$ is defined as the measure of effectiveness that indicates the degree to which the objective f_i is achieved. This measurable quantity is called an attribute and redefined as $X_i = f_i(x_i)$. $u_i(X_i)$ is an uniattribute utility function. In this section hereafter, the notation x_i is used to express the systems attribute for convenience in place of X_i.

Then the DMDP (12) is converted to an overall multiattribute decision problem in the following form, where X shows a feasible attribute set.

MADP

$$\underset{x_i \in X}{\text{Maximize}} \quad U(x_1, x_2, \ldots, x_m) \tag{13}$$

The overall preference function $U:R^m \rightarrow R^1$ is called the multiattribute utility function (MUF). Arguments for the MUF, x_i, can also be the MUF. The procedure of sequentially embedding those component utility functions into the multiattribute utility function is called nesting. The overall preference function U expresses a preference or objectives hierarchy in the following form ($q < m$):

$$\underset{x_i \in X}{\text{Max:}} \quad U[u^1(x^1),\ u^2(x^2),\ldots,\ u^q(x^q)] \tag{14}$$

$$= \underset{x_i \in X}{\text{Max}} \quad U[u^1(u_1(x_1),\ u_2(x_2),\ldots,\ u_t(x_t)),\ u^2(u_{t+1}(x_{t+1}),\ldots,$$

$$u_r(x_r)),\ldots,\ u^q(u_{s+1}(x_{s+1}),\ldots,\ u_m(x_m))] \tag{15}$$

Expression (15) shows the nesting of the m-subsystems into q-subsystems where u_i can also be a multiattribute utility function. The nesting procedures can be executed one after another in the objectives hierarchy of the stratified systems.

Now the problem is to specify a functional form of formulation (13). Keeney and Raiffa, under the assumptions of preferential independence and utility independence, show that function (13) is assessed in the following forms (Representation Theorems).

Additive utility function

$$U(x_1, x_2, \ldots, x_m) = \sum_{i=1}^{m} k_i u_i(x_i), \quad \text{if } \sum k_i = 1, \tag{16}$$

Multiplicative utility function

$$U(x_1, x_2, \ldots, x_m) = \frac{1}{K}\left[\prod_{i=1}^{m} (1 + Kk_i u_i(x_i)) - 1\right], \quad \text{if } \sum k_i \neq 1 \tag{17}$$

i) U and u_i are utility functions scaled from 0 to 1,

ii) $0 < k_i < 1$, $i = 1, 2, \ldots, m$, and

iii) if $\sum_{i}^{m} k_i \neq 1$, $k > -1$ is the non-zero solution to

$$1 + K = \prod_{i=1}^{m} (1 + Kk_i). \tag{18}$$

Parameters k_i and K are called scaling constants, and x_i is the attribute. Identification of the u_i and k_i is executed by the chance lottery technique and the indifference experiment. In the identification process of the MUF, value trade-offs among objectives are assessed and the systems coordination to cope with incompatibility among them is also performed. By reducing the overall decision problem (4) to the heuristic identification and maximization problem (13) of the MUF, noncommensurability in the original multiobjective optimization problem (1) is manipulated in commensurated terms. The preference structure of large-scale systems

is elucidated in a hierarchical configuration with the nesting procedure. However, with this method the optimization process for each subsystem i included in problem (12) is disregarded. In other words, systems evaluation is wholly disjuncted from optimization of the $f_i(x_i)$. In addition, diversification and ambiguity of the systems evaluation are also neglected.

4. Criteria for multiobjective optimization techniques

Now we enumerate the criteria which a desirable multiobjective optimization technique should meet. The MDP is concerned with a Complex Problematique whose characteristics are as follows: (i) largeness of scale, (ii) noncommensurateness, (iii) conflict and (iv) uncertainty. Systems to be optimized include many attributes related to various disciplines. Naturally the systems attributes are not measurable quantitatively in a commensurated unit and usually are incompatible with one another. In addition, systems evaluation is usually under uncertainty or faces fuzziness.

Because of those characteristics, a desirable technique should meet the following requirements.

(i) For coping with the large scale, systems should be structured to correspond to the level of complexty for decision-making including modeling and evaluation. Systems decomposition and coordination in a hierarchical configuration have been well-developed (Mesarovic and other 1970, Haimes 1977) and are recommended for effective structuring.

(ii) For manipulating the noncommensurability, a scalar-valued criterion function which is transformed from the vector criteria function should be introduced, and an operational identification procedure for deriving such a preference function should be developed with the procedural rationality.

(iii) For coping with the incompatibility of objectives, conflict management processes should be embedded, and value trade-offs among the conflicting objectives should be articulated.

(iv) For treating uncertain quantities, probabilistic or fuzzy assessment techniques should be included.

In short, the MDP is constructed and solved for supporting decisions that treat real problems in today's world. Thus the method should be problem-finding and problem-solving, and it should be able to determine priority among alternative policies. For raising the acceptability of the preferred solutions, interactive or learning and adaptation processes for achieving the "best compromised" solution should be implemented.

III. Multiobjective Decision Making and Mathematical Optimization

In the preceding section, it has been pointed out that the original multiobjective decision analysis technique has disjuncted the optimization phase from the judgemental phase of multiobjective decision-making. However, in the first step toward solving the DMDP(12), it is possible to optimize each subsystem $f_i(x_i)$ independently with mathematical programming techniques, and in the second step, the coordination process is executed with judgemental decisions. Thus two-layer systems are configured (Figure 1), which correspond to systems characteristics such as incompleteness of information (i.e., degree of uncertainty), modeling difficulty (i.e., feasibility of quantification) and complexty of decision making (i.e., level of abstraction). Now the problem is to bridge those two phases in this system. In other words, we are concerned with how to coordinate the systems and, based on this coordination, how to construct an overall systems evaluation preserving analytical results from the independently executed optimization process. For this purpose, we have proposed direct utilization of dual optimal solutions as the basic systems evaluation factor (Seo 1977, 1980, Seo and Sakawa 1979A, 1979B). This device is based on an interpretation of mathematical programming formulated in a multilevel system.

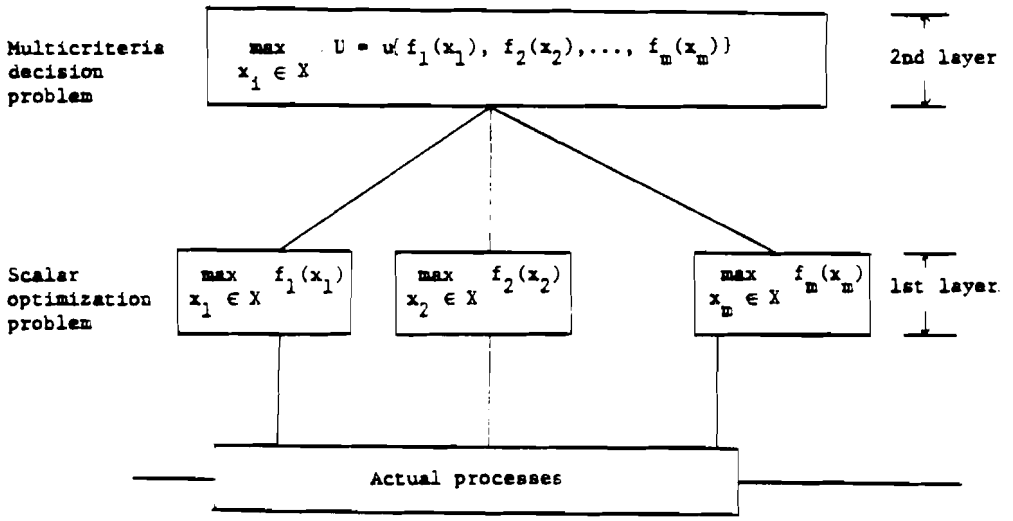


Figure 1. Systems decomposition and coordination in two-layers optimization

Consider the following mathematical programming problem.

$$(P) \quad \text{Maximize} \quad f_i(x_i) \tag{19}$$

$$\text{subject to} \quad h_{ij}(x_i) \leq d_{ij}, \quad j = 1,\ldots, p_i \tag{20}$$

$$g_{is}(x_i) \leq p_{is}, \quad s = p_i + 1,\ldots, \ell_i \tag{21}$$

where $x_i \in R^n$, $f_i: R^n \to R^1$, $h_{ij}: R^n \to R^1$, $g_{is}: R^n \to R^1$. Constraints are partitioned into policy constraints and technical constraints. The problem (P) is constructed independently for each subsystem, but also considered in a hierarchical systems configuration. The objective function $f_i(x_i)$ is regarded as the lower-level objective and the policy constraint d_{ij} is the upper-level objective indicated by the upper-level decision unit. The decision vector x_i is the lowest level objective. The problem (P) is solved in the lower-level decision unit. The dual optimal solution

λ^*_{ij} is used as the basic evaluation factor and information on it is sent up to the upper-level decision unit. The primal optimal solution x^*_i is sent down to actual activity process. The overall decision problem (DMDP)(12) is solved, via the nesting procedure (14) - (15) into the q subsystems, in the following form ($q < m$), where the dual optimal solution is used an inverse image of the utility function. We call this the nested Lagrangian multiplier problem.

NLMP (NDMDP)

$$\underset{x^r \in X}{\text{Maximize}} \quad U[u^1(\lambda^1(x^1|d^1)),\ldots, u^q(\lambda^q(x^q|d^q))] \tag{22}$$

where u^r, λ^r, x^r and d^r are all vectors included in the nested subsystem r, $r = 1,\ldots, q$. With this device, the hierarchical configuration of the two-layer optimization system shown in Figure 1 is converted to the two-layer decision system shown in Figure 2.

$$U = U\left[u^1(\lambda^1(d^1)),\ldots,u^q(\lambda^q(d^q))\right]$$

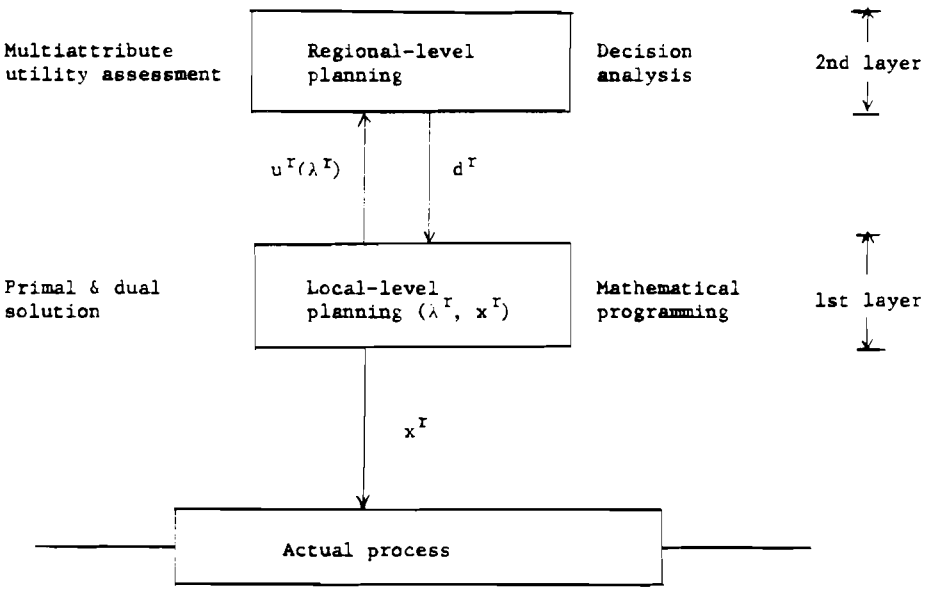


Figure 2. Structure of two layer decision system.

For example, the problem (P) is formed as a local environment management plan within a comprehensive regional planning. The problem is independently constructed without any regard to other subsystems. The objective function $f_i(x_i^*)$ is a local industrial production function and the policy constraint d_{ij} is an environmental requirement such as COD or SO_2 effluent discharge. The decision variables $x_i = \{x_{i1}, \ldots, x_{in}\}$ are sectoral capital formation and etc. The problem (P) is solved in each local administration agency. Information on the dual solution λ_{ij}^* is sent up to the regional authority as an evaluation medium for local-level planning and local industrial activity units (firms etc.) are informed of the primal solution x_i^*. Evaluation (22) for overall regional planning is used for assessing alternative environmental management plans and selecting the "best compromised" policy (Figure 3). In the next section, some theoretical points regarding using the dual optimal solution as the basic systems evaluation factor are examined.

IV. Optimality and the Kuhn-Tucker Multiplier

Consider a general mathematical programming problem.

$$(\bar{P}) \quad \text{Maximize} \quad f_i(x_i) \tag{23}$$

$$\text{subject to} \quad g_{ij}(x_i) \leq d_{ij}, \quad j = 1, \ldots, \ell_i, \tag{24}$$

where $x_i \in S = \{x_i \mid a_i \leq x_i \leq b_i, \; x_i \in R^n, \; a_i, b_i \in R^n\}$.

Now, formulate the following Lagrangian function:

$$\bar{L}(x_i, \lambda_i) = f_i(x_i) - \sum_{j=1}^{\ell_i} \lambda_{ij}\left[g_{ij}(x_i) - d_{ij}\right] - \mu_i(x_i - b_i) + \nu_i(x_i - a_i) \tag{25}$$

Suppose the $(\bar{P})$ has the optimal solution x_i^*. Then the Kuhn- Tucker theorem gurantees, under the condition $f_i \in C^1$, $g_{ij} \in C^1$ and holding the constraint qualification, the exsistence of the Lagrangian multiplier vectors $\lambda_i^* \geq 0$, $\lambda_i^* \in R^{\ell_i}$, $\mu_i^*, \nu_i^* \geq 0$, $\mu_i^*, \nu_i^* \in R^n$, satisfising the Kuhn-Tucker conditions (26)-(27).

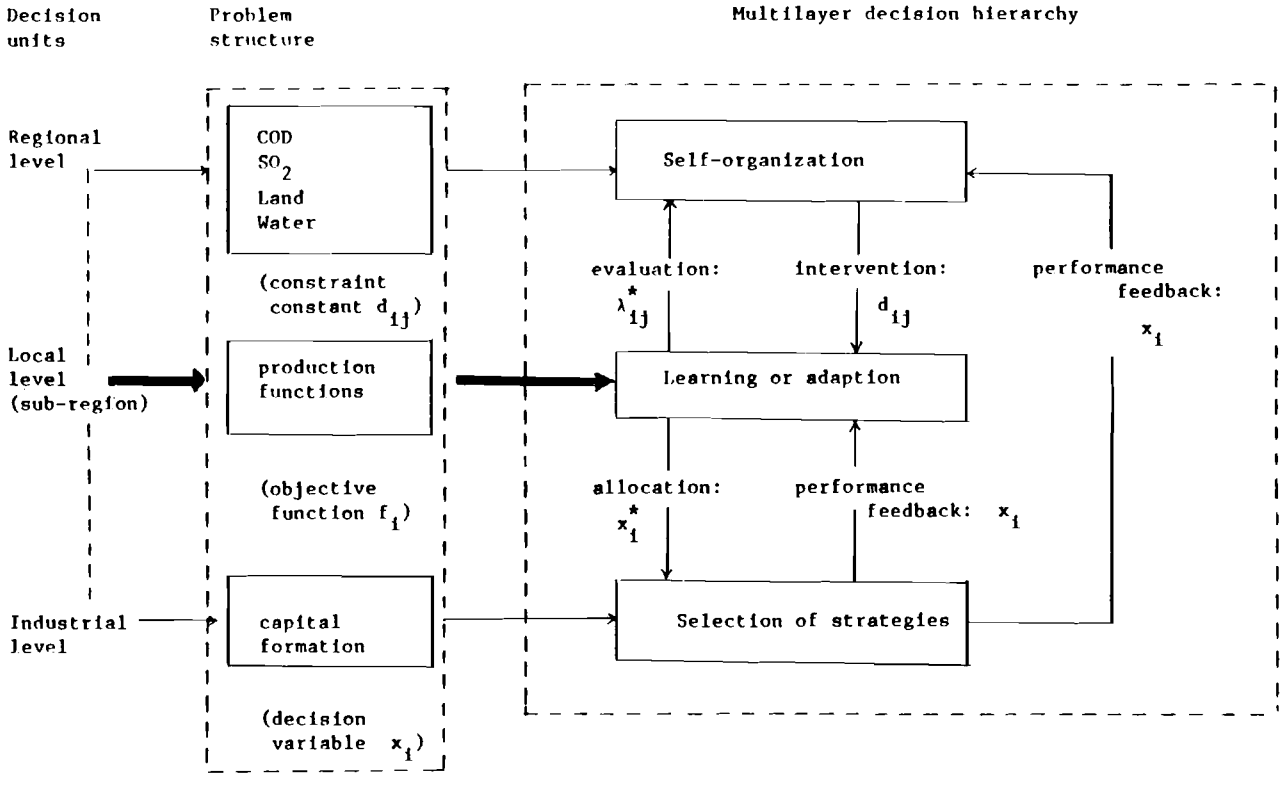


Figure 3. Functional hierarchy of local-level planning.

Kuhn-Tucker conditions

$$\nabla f_i(x_i^*) - \sum_{j=1}^{\ell_i} \lambda_{ij}^* \nabla g_{ij}(x_i^*) - \mu_i^* + \nu_i^* = 0 \tag{26}$$

$$\lambda_{ij}^* \geq 0, \quad \lambda_{ij}^* [g_{ij}(x_i^*) - d_{ij}] = 0, \qquad j=1,\ldots,\ell_i$$
$$\mu_i^* \geq 0, \quad \mu_i^*(x_i - b_i) = 0, \quad \nu_i^* \geq 0, \quad \nu_i^*(x_i - a_i) = 0 \tag{27}$$

The theorem shows that the existence of the Kuhn-Tucker multiplier is the first-order necessary condition for the x_i^* to be locally optimal in the (P) and only differentiability is assumed. Thus exsistence of the Kuhn-Tucker vetor λ_{ij}^* corresponding to the x_i^* is guranteed as the necessary condition of optimality in non-convex as well as convex problem.

Under the approapriate conditions, we can perform meaningful differenciation of the $(\bar{L})$ for $x_i(d_i)$ and $\lambda_i(d_i)$, where $x_i \in R^n$, $\lambda_i \in R^{\ell_i}$ and $d_i \in R^{\ell_i}$. in the vector form,

$$\frac{\partial \bar{L}}{\partial d_i} = \left[\frac{\partial f_i}{\partial x_i} - \lambda_i \frac{\partial g_i}{\partial x_i} - \mu_i + \nu_i\right] \frac{\partial x_i}{\partial d_i} - \left[g_i(x_i) - d_i\right] \frac{\partial \lambda_i}{\partial d_i} + \lambda_i \tag{28}$$

Especially when $\lambda_i^* > 0$, from (26) and (27), $\frac{\partial \bar{L}}{\partial d_i} = \lambda_i^* = \frac{\partial f_i^*}{\partial d_i}$.

Now the $\lambda_i^* > 0$ has an interpretation as an evaluator. First, the component of $\lambda_i^* = \partial f_i^* / \partial d_i$ shows a ratio of a marginal variation of the criterion function f_i (system output) expressed in terms of value to marginal variation of the constraint constant d_i (system input) expressed in terms of quantity. In other words, the inverse of λ_i^* is imputed prices or shadow prices of the constraints measured in terms of the value of criterion function (Luenberger 1973). Second, λ_i^* is the dual optimal solution. Consider the following minimax dual problem (Lasdon 1970).

$$(\bar{P}^*) \qquad \underset{\lambda_i \in D}{\text{Minimize}} \quad \max_{x_i \in S} \bar{L}(x_i, \lambda_i) \tag{29}$$

where $D = \{\lambda_i \mid \lambda_i \geq 0, \ \lambda_i \in R^{\ell_i}, \ \max_{x_i \in S} \bar{L}(x_i, \lambda_i) \text{ exists}\}$.

$$h(\lambda_i) = \max_{x_i \in S} \bar{L}(x_i, \lambda_i) \tag{30}$$

is the dual function. Domain D does not have to be convex. When f_i and g_{ij} are differentiable and convex, the $(\bar{P})^*$ corresponds to Balinski-Baumol's dual problem (1968) which is a variation of Wolfe's dual problem (1961). Thus the Kuhn-Tucker multiplier vector λ_i^* is the solution of the dual problem $(\bar{P})^*$ combined with the primal problem $(\bar{P})$. When the solutions (x^*, λ^*) are the saddle point of the $(\bar{L})$, $f(x_i^*) = h(\lambda_i^*)$. Even in the case of a non-convex problem, the saddle point can be achieved locally with an heuristic algorithm if proper initial values are selected.

Now the above discussion is applied to the original problem (19) - (21). Form the Lagrangian function:

$$(L) \qquad L_i(x_i, \lambda_i) = f_i(x_i) - \sum_{j=1}^{p_i} \lambda_{ij}(h_{ij}(x_i) - d_{ij}) - \sum_{s=p_i+1}^{\ell_i} \lambda_{is}(g_{is}(x_i) - b_{is}) \qquad (31)$$

When the Kuhn-Tucker multiplier $\lambda_{ij}^* > 0$, then $\lambda_{ij}^* = \partial f_i^*/\partial d_{ij}$. In the hierarchical system, the λ_{ij}^* inversely shows a marginal variation of the upper-level objective evaluated in terms of a marginal variation of the lower-level objective. The larger the λ_{ij}^* is, the smaller the opportunity cost of the d_{ij} measured in terms of the marginal variation of the f_i is. This means that the degree of satisfaction of the d_{ij} (quantity) measured in terms of the f_i (value) is already high. The λ_{ij}^* is the opportunity cost or shadow price. The λ_{ij}^* represents proportionally the degree of satisfaction with the lower-level objective in the presentation of the upper-level objective. Note that the market price expresses the degree of satisfaction with the commodity inversely. In contrast, the λ_{ij}^* can express proportionally the degree of satisfaction with the d_{ij} because the inverse $1/\lambda_{ij}^*$ is the opportunity cost or the shadow price.

In the case of our local-level planning (Section III), the λ^*_{ij} is used as the basic evaluation factor for the environmental constraint imposed by the regional authority, which is examined from the point of view of maximizing local industrial output in subregion i.

V. Conversion of the Kuhn-Tucker Multiplier to a Quasi-utility Function and its Nesting

1. Derivation of the component utility function

Now the basic utility function $u_{ij}(\lambda_{ij}(x_i|d_{ij}))$, $j = 1,\ldots, p_i$, should be derived from the above discussions. The problem is to convert the Kuhn-Tucker multiplier into a utility index function. Along the line of von Neumann and Morgenstern's theorem, it can be shown that positive linear transformation of the λ^*_{ij} to u_{ij} is admissible.

Define a relation $A = (\Omega, R)$ and call it preference relation A. Here Ω is a nonempty set and R is a binary relation defined on elements of Ω.

Definition 1 (preference relation A). If R is a binary relation on the set Ω and if $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z} \in \Omega$, then preference relation A on individual choice satisfies the following axioms:

(i) Transitivity: if $\mathcal{X} R \mathcal{Y}$, $\mathcal{Y} R \mathcal{Z}$, then $\mathcal{X} R \mathcal{Z}$,

(ii) Weak connectivity: $\mathcal{X} R \mathcal{Y}$, or $\mathcal{Y} R \mathcal{X}$,

(iii) Nonsatiety: if $\mathcal{X} I \mathcal{Y}$, then $\mathcal{X}' P \mathcal{Y}$ for $\mathcal{X}' = \mathcal{X} + \Delta\mathcal{X}$, $\Delta\mathcal{X} > 0$.

(iv) Continuity: if $\mathcal{X} R \mathcal{Y}$ and $\mathcal{Y} R \mathcal{Z}$, then there is a real number α such that $0 \leq \alpha \leq 1$ and $(\alpha \mathcal{X} + (1 - \alpha)\mathcal{Z}) I \mathcal{Y}$.

Here R shows "prefer to" (P) or "indifferent to" (I).

The von Neumann-Morgenstern theorem is restated as follows (Luce and Suppe 1965).

Theorem 1. Under preference relation A, there exists a real-valued function S defined on Ω such that for every $\mathcal{X}$ and $\mathcal{Y}$ in Ω and a parameter α in [0, 1]

(i) $\mathcal{X} R \mathcal{Y}$ if and only if $S(\mathcal{X}) \geq S(\mathcal{Y})$

(ii) $S(\alpha \mathcal{X} + (1 - \alpha)\mathcal{Y}) = \alpha S(\mathcal{X}) + (1 - \alpha) S(\mathcal{Y})$

Moreover, if S' is any other function satisfying (i) and (ii), then the S' is related to S by a positive linear transformation.

According to the interpretation of the Kuhn-Tucker multiplier as the shadow price, the λ_i^* can replace the S in theorem 1. Consider real valued functions $\lambda_i(\mathcal{X})$ and $S(\mathcal{X})$ defined on the decision set $D(\mathcal{X}, \mathcal{Y} \in D)$. From now on, * is omitted.

Proposition. Two real valued functions $\lambda_i(\mathcal{X})$ and $S(\mathcal{X})$ defined on the set D are in an equivalence class. Namely

(i) A binary relation R for numerical magnitudes of λ_i and S on the set D is reflexive, or $\lambda_i R \lambda_i$ for every $\lambda_i \in D$.

(ii) The binary relation R for the λ_i and S is symmetric, or if $\lambda_i RS$ then $SR\lambda_i$ for every S, $\lambda_i \in D$.

Thus the above S in theorem 1 is replaced with the λ_i.

Theorem 2. For every $\mathcal{X}$ and $\mathcal{Y}$ in the set D defined under the preference relation A, the following properties are preserved for the function λ_i.

(i) $\mathcal{X} R \mathcal{Y}$ if and only if $\lambda_i(\mathcal{X}) \geq \lambda_i(\mathcal{Y})$,

(ii) $\lambda_i\{\alpha \mathcal{X} + (1 - \alpha)\mathcal{Y}\} = \alpha\lambda_i(\mathcal{X}) + (1 - \alpha)\lambda_i(\mathcal{Y})$

The $\mathcal{X}$ and $\mathcal{Y}$ are regarded as some implicit evaluations for the system's constraints d_{ir} and d_{is} respectively. We can write $\lambda_i(d_{ir}) = \lambda_{ir}$ and $\lambda_i(d_{is}) = \lambda_{is}$ where $s, r \in I = [1, \ldots, j, \ldots, p_i]$.

The function λ_i can be linearly transformed to the function u_i which satisfies (i) and (ii). We can describe this in the following way.

Theorem 3. (derivation of quasi-utility function) A Kuhn-Tucker multiplier λ_{ij} can be positive-linearly transformed to a numerical utility u_{ij} defined on a value between 0 and 1.

Thus we derive the quasi-utility function in alternative forms.

$$u_i = u_i(x_i) \tag{32}$$

$$= u_i(\lambda_i(x_i|d_i)) \tag{33}$$

$$= -a_i + b_i\, \lambda_i(x_i|d_i)\,, \tag{34}$$

where $\lambda_i = \{\lambda_{ij} | \lambda_{ij} > 0\,,\ j = 1,\ldots, p_i\}$.

The basic idea for deriving the quasi-utility function is shown in a more general vector form in Figure 4. The procedure of converting the λ_{ij} into the u_{ij} is shown in Figure 5. In practice, we choose lower and upper bounds, $\underline{\lambda}_{ij}$ and $\overline{\lambda}_{ij}$, of the λ_{ij} such as $0 < \underline{\lambda}_{ij} < \lambda_{ijmin}$ at $u_{ij}(\underline{\lambda}_{ij}) = 0$ and $\overline{\lambda}_{ij} > \lambda_{ijmax} > 0$ at $u_{ij}(\overline{\lambda}_{ij}) = 1$.

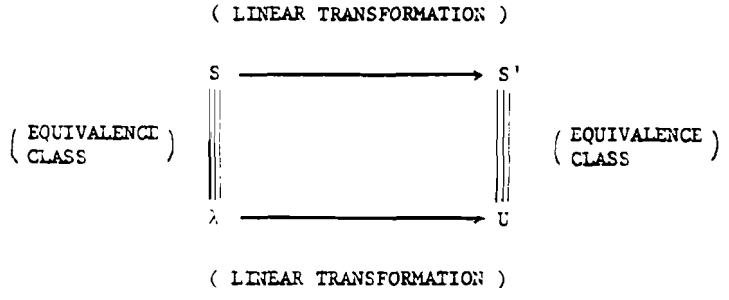


Figure 4. Deviation of a basic utility function.

For the numerical utility, although differences between the utilities are numerically measurable, the position of origin and the unit of a numerical scale for the utilities can be arbitrarily decided. This type of scale is called an interval scale. Thus, the cardinal utility functions are derived.

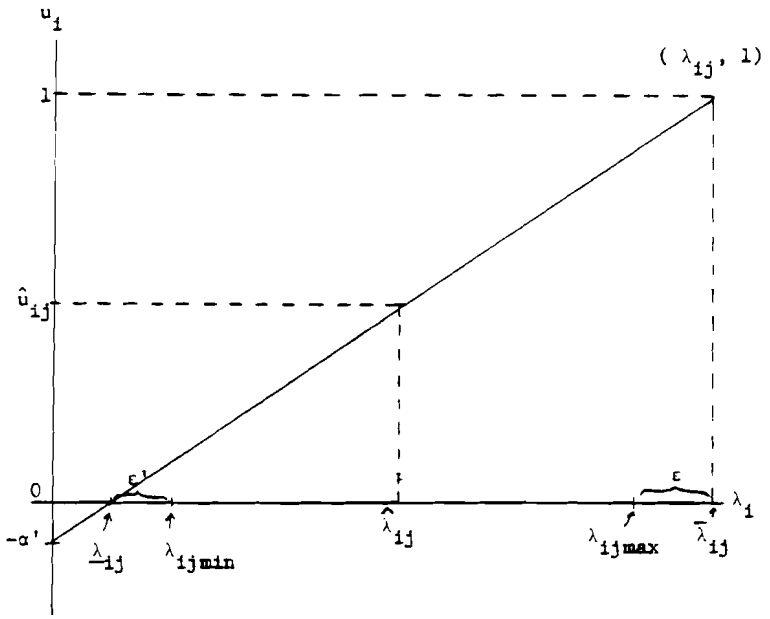


Figure 5. Conversion of the shadow prices into quasi-utility functions

In our local-level planning, the $u_{ij}(\lambda_{ij}(x_i|d_{ij})$ is used as a component utility function which is related to the environmental restriction (target constraint) d_{ij} indicated for subregion i.

2. Nesting into the multiattribute utility functions

Now, using the component utility function, the multiattribute utility functions (MUF) are constructed and nested. The procedure for deriving the MUFs is similar to the technique of multiattribute utility analysis by Keeney and Raiffa, except that here the assessment of trade-offs between attributes is executed on normalized utility values.

In local-level planning for each subsystem i, multiattribute utility functions are constructed in the following forms.

Additive

$$U_i(u_1, u_2, \ldots, u_{p_i}) = \sum_{j=1}^{p_i} k_{ij} u_{ij}(\lambda_{ij}), \quad \sum_{j=1}^{p_i} k_{ij} = 1 \tag{34}$$

or multiplicative

$$U_i(u_1, u_2, \ldots, u_{p_i}) = \frac{1}{K_i} \left[\prod_{j}^{p_i} (1 + K_i k_{ij} u_{ij}(\lambda_{ij})) - 1 \right],$$

$$\sum_{j=1}^{p_i} k_{ij} \neq 1, \quad K > -1 \tag{35}$$

When constructing an overall MUF via nesting into the q-subsystem,

additive

$$U = (U_1, U_2, \ldots, U_q) = \sum_{i=1}^{q} k_i U_i(u_1, u_2, \ldots, u_{p_i}) \tag{36}$$

or multiplicative

$$U = (U_1, U_2, \ldots, U_q) = \frac{1}{K} \left[\prod_{i}^{q} (1 + K k_i U_i(u_1, \ldots, u_{p_i})) - 1 \right] \tag{37}$$

The NDMDP (22) is expressed, using formulations (34) - (37), in the following form.

$$\text{Maximize} \quad U[U_1(u_{11}(\lambda_{11}(x_1)), u_{12}(\lambda_{12}(x_1)), u_{1p_1}(\lambda_{1p_1}(x_1))), \ldots,$$

$$U_q(u_{q1}(\lambda_{q1}(x_q)), u_{q2}(\lambda_{q2}(x_q)), \ldots, u_{qp_q}(\lambda_{qp_q}(x_q)))] \tag{38}$$

The nesting procedures can be carried out one after another in more complex forms of an objectives hierarchy.

In the process of assessing the MUFs, a coordination procedure in the judgemental phase of decision making is explicitly introduced with the trade-off experiments. The component utility functions are weighted by the DM and compromised with each other.

A drawback of the MUF method lies in the strong assumption of preference independence and utility independence among the attributes. However in our method the uniattribute utility function u_{ij} is based only on the λ_{ij} and simply transformed from it. Thus, it is not necessary for the assessor to be bothered with an independence check.

VI. Summary of the technique

The technique for multiobjective decision-making described in Sections III to V is summarized in the following.

(1) Problems which a purposeful system includes are structured in a hierarchical modeling of multi-level systems, mainly according to the levels of abstraction of objectives. It is called an objectives hierarchy. This system is composed of two-layer decision systems corresponding to the complexity of the decision making.

(2) Mathematical modeling is constructed for each subsystem in the first layer, and mathematical programming is solved independently as single-objective optimization problems.

(3) Using the Kuhn-Tucker multiplier, the oportunity cost or the shadow price for the systems constraint is assessed. The Kuhn-Tucker multiplier is directly transformed to the quasi-utility function, which is used as the basic uniattribute utility function.

(4) The basic utility functions are nested into the MUFs. This nesting procedure is executed sequentially with interactive processes. In this nesting process, systems coordination of the decomposed multilevel systems is carried out.

(5) Finally an overall MUF for the overall decision system is derived. Using this value, alternative policy programs are examined and compared. Priority for selecting the most desirable programs (a set of normative values of decision variables) is determined.

We call this procedure the Nested Lagrangian Multiplier (NLM) method. Using this procedure, the multiobjective optimization problem (MOP) is reduced to a set of scalar optimization problems (p) in the first step, and then these are coordinated into an overall decision problem (MDP) in the second step. This procedure is primarily based on the duality of mathematical programming and the multiattribute decision analysis.

Now the strong and weak characteristics of the NLM method should be examined.

(1) The method configures a hierarchical structure of objectives. Thus a problem structure can be clearly specified according to the properties of the systems objectives.

(2) Due to utilization of the duality of mathematical programming. The evaluation problem and the optimal resource allocation problem combined with it are simultaneously solved.

(3) The method directly utilizes the Kuhn-Tucker multipliers for basic systems evaluation. By this device, ambiguity included in decision analysis is excluded in the first step of the evaluation.

(4) The noncommensurable objectives are commensurated with the quasi-utility function as a medium converted from the Kuhn-Tucker multiplier. The quasi-utility function is used as the preference function(cardinal utility function) without any loss of generality.

(5) The function of the DM for systems coordination is explicitly introduced by using decision analysis. Thus the analytical and judgemental phases of the decision making process are combined.

(6) Quantitative evaluation of alternative systems designs or programs is carried out with a single evaluation standard. Thus the ordering of priorities for selecting the alternatives is determined.

(7) For applying this method to empirical problems, unit measures selected for data bases should be reasonable and practically meaningful,

namely, numerical results are not free from the magnitude or dimensions of the unit measures of data bases. However because of this property of the method, differences in relative scale between systems are reflected in the evaluation results.

(8) This method carries out optimization of systems and evaluation of preferences with heuristic procedures, depending only on local information for the systems' functions. Thus devices to correct possible biases in information are required even if this characteristic may avoid risky u-values.

(9) The systems coordination ultimately depends on the single DM who is assumed to be a knowledgeable person. This Platonic assumption should be mitigated and methods for introducing variety of evaluation and for forming consensus should be developed.

VII. Example

In this section, this method is applied in a case study of regional planning.

Figure 6 is the hierarchical systems configuration, where regional decomposition has been performed on three levels in the second layer and functional decomposition has been done on two levels in the first layer.

Mathematical programming formulations are as follows.

Local residential problem (Seo 1977)

$$\text{Maximize} \quad f(W) = \sum_{j=1}^{n} (C - AB^{W_j}) \tag{39}$$

subject to

$$W_j \geq \underline{W}_j , \qquad j = 1, \ldots, n , \tag{40}$$

and

$$\sum_{j=1}^{n} |W_j - \bar{W}| \leq \nu , \tag{41}$$

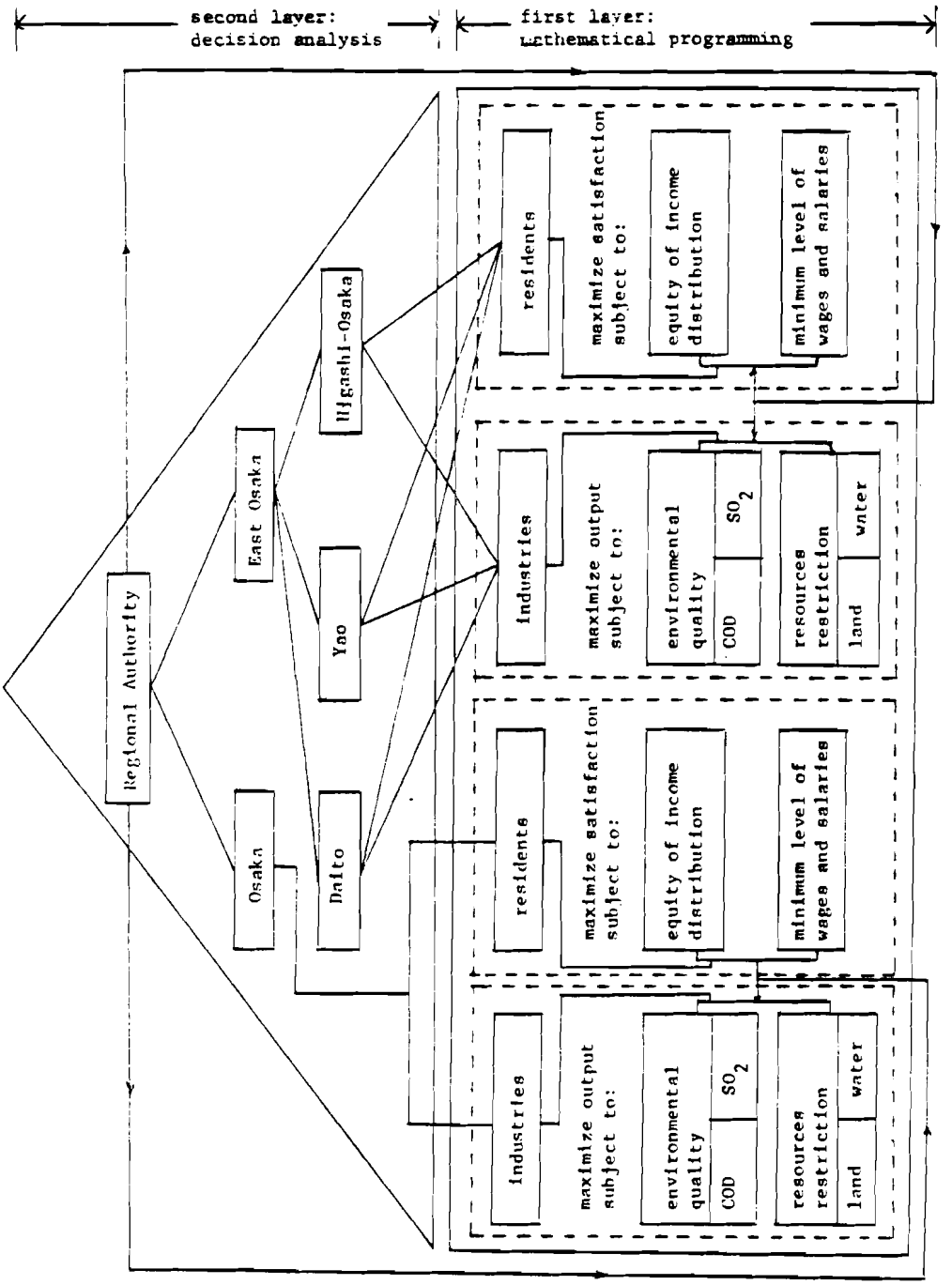


Figure 6. A hierarchical environmental system in the industrialized Yodo River basin

where

j denotes an industry,

n is the number of industries,

W_j denotes the wages and salaries per employee per year in industry j (decision variable),

W is the vector $(W_1, \ldots, W_n)$,

$\tilde{W}_j$ denotes the actual value of wages and salaries in 1974 in industry j,

$\bar{W}$ is the actual mean value of the unit payment in all industries,

v is the sum of the deviations from the actual mean value of wages and salaries, and A, B, C are parameters.

Constraint (40) represents the minimum requirements for wages and salaries in each industry, and constraint (41) is an equity requirement among all the industries. For the objective function, a modified exponential curve has been estimated:

$$f(W) = \sum_{j=1}^{n} (1.0102 - 0.9926 \times 0.76808^{W_j}) . \qquad (42)$$

Both primal and dual optimal solutions have been obtained, and the Lagrangian multipliers have been converted into single-attribute utility functions.

As an example of deriving the quasi-utility function, utility values for the equity requirement are shown in Table 1.

Table 1. Assessment of the equity requirement

Region (i)	Osaka	Yao	Daito	Higshi-Osaka
λ	1.6209×10^{-3}	1.6659×10^{-3}	1.6452×10^{-3}	1.6886×10^{-3}
u	0.2099	0.6593	0.4524	0.8864
σ	41.53	34.60	38.60	34.15

Here σ is a standard derivation of average wages and salaries among industries and is used for checking the results of the u-value assessment. Ordering of the u-values for equity corresponds well to that of the σ-values.

Local MUFs for the residential problem

Osaka City

$$U_O^R = \frac{1}{0.61}\left[(1 + 0.43\, u_O^m)(1 + 0.13\, u_O^e) - 1\right] \tag{43}$$

East Osaka:

$$U_Y^R = \frac{1}{21.1}\left[(1 + 19.0\, u_Y^m)(1 + 0.95\, u_Y^e) - 1\right] \tag{44}$$

$$U_D^R = \frac{1}{0.31}\left[(1 + 0.25\, u_D^m)(1 + 0.05\, u_D^e) - 1\right] \tag{45}$$

$$U_H^R = \frac{1}{0.12}\left[(1 + 0.11\, u_H^m)(1 + 0.01\, u_H^e) - 1\right] , \tag{46}$$

where u^m and u^e are the utility functions for the minimum-wages requirement and the equity requirement respectively, and the subscripts O, Y, D, and H denote Osaka, Yao, Daito, and Higashi-Osaka respectively.

Local industrial pollution control problem (Seo and Sakawa 1979A)

$$\text{Maximize} \quad \sum_r f_r(K_r, L_r) = \sum_r c_r K_r^{a_r} L_r^{b_r} \tag{47}$$

subject to

$$\sum_r \frac{\omega_{jr}}{\kappa_r} K_r \leq \tau_j , \quad j = 1, \ldots, 4 , \tag{48}$$

$$\sum_r K_r / \sum_r L_r \leq \gamma , \tag{49}$$

$$\alpha\tilde{K}_r \leq K_r \leq \beta\tilde{K}_r , \quad r = 1, \ldots, n , \tag{50}$$

$$\alpha'\tilde{L}_r \leq L_r \leq \beta'\tilde{L}_r , \quad r = 1, \ldots, n , \tag{51}$$

where

j denotes an environmental factor [chemical oxygen demand (COD), sulphur dioxide (SO_2), land, or water] ($j = 1, \ldots, 4$)

K_r is the capital value (book value of tangible fixed assets) of industry r (decision variable),

$\tilde{K}_r$ is the actual capital value of industry r,

L_r is the number of employees in industry r (decision variable),

$\tilde{L}_r$ is the actual number of employees in industry r,

ω_{jr} is the requirement or discharge of environmental factor j per unit of shipments in industry r,

κ_r is the capital coefficient, namely capital value per unit of shipments in industry r.

τ_j is the restriction or target level for the environmental factor j,

γ is the actual overall capital intensity (ratio of total capital value to total number of employees),

a_r, b_r, c_r are parameters of the production function for each industry r, and

α, β, α', β' are parameters which represent friction in (resistance to) the transfer of capital and labor ($0 < \alpha, \alpha' < 1$; $1 < \beta, \beta'$).

The objective function (47) is a Cobb-Douglas type of production function which is homogeneous of degree one ($a_r + b_r = 1$), and thus if each factor is paid its marginal product, total output is distributed between labor and capital in the respective proportions a_r and b_r. Constraint (48) is the target constraint which shows that the total amount of any environmental factor required or discharged by each industry must not exceed a limit imposed by the decision-maker. Constraint (49) is a technical constraint which shows capital intensity as a whole. Constraints (50) and (51) are frictional constraints for avoiding drastic changes in the industrial structure.

The problem is to find the optimal allocation of production factors (capital and labor) to each industry under constraints (48) to (51). The augmented Lagrangian method proposed by Pierre and Lowe (1975) has been utilized in solving the nonlinear optimization problems.

The assessment with the u-value is shown in Table 2.

Table 2. Assessment of the u-values for environmental control.

Area (j)	COD		SO_2		Land		Water	
(i)	λ	u	λ	u	λ	u	λ	u
Osaka								
case 1	2.1642	0.0323	0.9474	0.0012	0.0000	-0.0230	37.5866	0.9383
case 2	0.2275	0.0008	14,4730	0.4101	4.7484	0.1307	30.3245	0.8656
Yao	0.0000	0.0000	11.5712	0.7714	8.1213	0.5414	6.1050	0.4070
Daito	10.9544	0.7121	0.9618	0.0008	7.2767	0.4503	0.0000	-0.0676
Higashi-Osaka	15.5528	0.7760	0.1605	0.0005	7.0646	0.3483	9.1199	0.4519

Here a negative sign of the u-value is used to indicate slackness of inactive constraints.

Local MUFs for the industrial pollution control problem

Osaka City

$$U_O^{IPC} = \frac{1}{-0.8268}\left[(1 - 0.6614\, u_O^{COD})(1 - 0.2646\, u_O^{SO_2}) \times (1 - 0.1984\, u_O^{land})(1 - 0.1323\, u_O^{water}) - 1\right] \quad (52)$$

East Osaka

$$U_Y^{IPC} = \frac{1}{-0.8871}\left[(1 - 0.2395\, u_Y^{COD})(1 - 0.07984\, u_Y^{SO_2}) \times (1 - 0.7984\, u_Y^{land})(1 - 0.1996\, y_Y^{water}) - 1\right] \quad (53)$$

$$U_D^{IPC} = \frac{1}{-0.6704}\left[(1 - 0.5363\, u_D^{COD})(1 - 0.1073\, u_D^{SO_2}) \times (1 - 0.1341\, u_D^{land})(1 - 0.0804\, u_D^{water}) - 1\right] \quad (54)$$

$$U_H^{IPC} = \frac{1}{-0.5855}\left[(1 - 0.1317\, u_H^{COD})(1 - 0.4391\, u_H^{SO_2}) \times (1 - 0.1098\, u_H^{land})(1 - 0.0439\, u_H^{water}) - 1\right] \quad (55)$$

Regional MUFs are composed of the above local MUFs.

Regional MUFs

Osaka City :U_O

$$U_O = \frac{1}{0.1020}\left[(1 + 0.02856\ U_O^R)(1 + 0.714\ U_O^{IPC}) - 1\right] \tag{56}$$

East Osaka :U_E

$$U_Y = \frac{1}{0.3125}\left[(1 + 0.05\ U_Y^R)(1 + 0.25\ U_Y^{IPC}) - 1\right] \tag{57}$$

$$U_D = \frac{1}{0.2076}\left[(1 + 0.0265\ U_D^R)(1 + 0.1765\ U_D^{IPC}) - 1\right], \tag{58}$$

$$U_H = \frac{1}{0.1481}\left[(1 + 0.0333\ U_H^R)(1 + 0.1111\ U_H^{IPC}) - 1\right] \tag{59}$$

$$U_E = \frac{1}{0.6406}\left[(1 + 0.5125\ U_Y)(1 + 0.05765\ U_D)(1 + 0.0256\ U_H) - 1\right]; \tag{60}$$

Region :

$$U_R = U(U_E\ ,\ U_O) = \frac{1}{0.3125}\left[(1 + 0.25\ U_O)\ (1 + 0.05\ U_E) - 1\right]; \tag{61}$$

With the numerical u-values and MUF values, diagnosis for the industrial pollution control program is provided in Table 3.

Table 3. Diagnosis

Region	Satisfaction level minimum	maximum	Slackness	Technological change (capital-saving)
Osaka				
case 1	SO_2	water	land	drastic
case 2	COD	water	none	drastic
Yao	COD	SO_2	none	medium
Daito	SO_2	COD	water	medium
Higashi-Osaka	SO_2	COD	none	medium

VIII. Concluding remarks: extension for uncertainty

The multiobjective optimization method thus for presented still depends on deterministic procedures, even though subjective decision-making processes are included. Deterministic evaluation based on mathematical programming

and the MUF method needs to be modified with explicit consideration given to uncertainty and diversification of information. Here are several approaches for managing these in the judgemental phase of decision-making.

(1) Probability assessment and derivation of expected utility functions

For including randomness in the u-value assessment, the values of the utility functions are treated as uncertain quantities. Expected values of the component utility functions are assessed with judgemental or hypothetical probability distributions. The probability distribution is assessed according to the method developed by Schlaifer (1969). The value u_{ij}^{s} for the cumulative distribution function

$$F_{ij}^{s}(u_{ij}^{s} \leq \hat{u}_{ij})$$

is assessed for several fractiles of the distribution. Computer programs can effectively convert the cumulative function to the density function $f_{ij}^{s}(u_{ij}^{s})$ and calculate the expected values of the component utility function,

$$\bar{u}_{ij} = E\, u_{ij} = \sum_{s} f_{ij}^{s}(u_{ij}^{s})\, u_{ij}^{s} \quad . \tag{62}$$

Based on the MANECON collection by Schlaifer (1971), the ICOPSS computer package (Interactive Computer Program for Subjective Systems) developed by the authors (Sakawa and Seo 1979C, 1980) is used in this process. Probability characteristics and diagramic representations are also provided in this package. Using the expected u-values, the MUF-values are derived in additive and multiplicative forms,

$$\tilde{U}_i(\lambda_i(x_i)) = \sum_{j} k_{ij}\tilde{u}_{ij}(\lambda_{ij}(x_i)) \quad , \qquad \sum_{j} k_{ij} = 1 \tag{63}$$

$$\tilde{U}_i(\lambda_i(x_i)) = \frac{1}{K}\left[\prod_{j}(K\, k_{ij}\tilde{u}_{ij}(\lambda_{ij}(x_i)) + 1) - 1\right] \quad , \tag{64}$$

$$\sum_{j} k_{ij} \neq 1, \quad K > -1$$

where $\tilde{u}_{ij}$ is the expected value of the component utility function. This technique has been presented by the authors (Seo and Sakawa 1980).

(2) Entropy modeling for alternative policy-making

The selected attributes whose values are included in the alternative policy program are regarded as random variables. The u-values, $u_{i1},\ldots, u_{ip_i}$, are supposed to occur independently in an alternative senario with probability $p_{i1},\ldots, p_{ip_i}$. The following bicriteria problem is solved.

(A) bottom-up policy

To minimize the average current u-value to be included in the possible attribute set X^h for the alternative policy program h:

$$\bar{u}_{ij} = p_{i1}u_{i1} + p_{i2}u_{i2} +\ldots + p_{ip_i}u_{ip_i}, \tag{65}$$

(B) maximization of diversification

To maximize the entropy:

$$H_i = -p_{i1} \log p_{i1} -\ldots - p_{ip_i} \log p_{ip_i} \tag{66}$$

The entropy problem is to find the probability distribution $\{p_{i1}, p_{i2},\ldots, p_{ip_i}\}$, $\sum_{j=1}^{p_i} p_{ij} = 1$ and $0 \le p_{ij} \le 1$, which satidfies the criteria (A) and (B) (Kunisawa 1975).

After rearranging the u-values, $u_1, u_2,\ldots, u_r$, as a relatively prime integer ratio, $\hat{u}_{i1}, \hat{u}_{i2},\ldots, \hat{u}_{ip_i}$, the following equation is solved and a positive root w^* is obtained.

$$w^{-\hat{u}_{i1}} + w^{-\hat{u}_{i2}} + \ldots + w^{-\hat{u}_{ip_i}} = 1 \tag{67}$$

The entropy problem (A) (B) is solved by obtaining the p_{ij} in the following form:

$$p_{ij} = w^{*-\hat{u}_{ij}}, \quad j = 1,\ldots, p_i \tag{68}$$

This technique has been used tentatively by the authors (1981).

(3) Fuzzification procedure for the u-value assessment

(i) Operations on fuzzy numbers extrapolated from fuzzy set theory have been recently developed (Dubois and Prade 1978, 1979). Let $\tilde{m}_{ij}$ be the fuzzy number, and consider the L-R fuzzy numbers with the mean value $\bar{m}_{ij}$ and the left and right spreads α_{ij} and β_{ij}:

$$\tilde{m}_{ij} = (\bar{m}_{ij}, \alpha_{ij}, \beta_{ij}) \quad . \tag{69}$$

The L-R type fuzzy number is defined in terms of a mumbership function $\mathcal{F}_{(\tilde{m}_{ij})}(y_{ij})$ as follows.

$$\mathcal{F}_{-(\tilde{m}_{ij})}(y_{ij}) = L_{-}[(y_{ij} - \bar{m}_{ij})/\alpha_{ij}] \quad , \quad \alpha_{ij} > 0 \tag{70}$$

$$\mathcal{F}_{+(m_{ij})}(y_{ij}) = R_{+}[(y_{ij} - \bar{m}_{ij})/\beta_{ij}] \quad , \quad \beta_{ij} > 0 \tag{71}$$

where $f_{\tilde{m}_{ij}}(y_{ij})$ is the truth value of the assertion "the value $\tilde{m}_{ij}$ is y_{ij}" and called numbership function. Now, in the process of deriving the MUF-value, the k_{ij} is assessed as the fuzzy number:

$$\tilde{k}_{ij} = (\bar{k}_{ij}, \alpha_{ij}, \beta_{ij})_{LR=RL} \quad . \tag{72}$$

Actually, in the indifference experiment with a canonical lottery ("all best or all worst") for deriving the (standard) k_{ij}-value, the probability p_{ij} for the chance fork is regarded as the fuzzy number. In the case of $\sum_{j} \bar{k}_{ij} = 1$, the additive form is applied; otherwise the multiplicative form is used. Using the addition and production calculus, the fuzzy MUF-value is derived sequentially.

$$\tilde{U}_i(u_{i1}, \ldots, u_{ip_i}) = \{(\bar{u}_{ij}, \alpha_{ij}, \beta_{ij})\}_{LR=RL} \tag{73}$$

$$\tilde{U}(U_1, \ldots, U_q) = \{(\bar{U}_i, \delta_{ij}, \gamma_{ij})\}_{LR=RL} \tag{74}$$

(ii) In addition, fuzzification of max-min operations is also used for determining what is the value of the smaller or grater of two fuzzy numbers, $\tilde{k}_{is}$ and $\tilde{k}_{ir}$. The fuzzy max and fuzzy min of the $\tilde{k}_{ij}$-value is determined respectively in terms of membership functions $\mathcal{Q}(z_{ij})$ and $\mathcal{P}(z_{ij})$.

fuzzy max $(\tilde{k}_{is}, \tilde{k}_{ir})$

$$\mathcal{Q}(z_{ij}) = \max_{\max(y_{is}, y_{ir}) = z_{ij}} \min\left(\mathcal{G}_{is}(y_{is}), \mathcal{G}_{ir}(y_{ir})\right) \tag{75}$$

fuzzy min $(\tilde{k}_{is}, \tilde{k}_{ir})$

$$\mathcal{P}(z_{ij}) = \max_{\min(y_{is}, y_{ir}) = z_{ij}} \min\left(\mathcal{G}_{is}(y_{is}), \mathcal{G}_{ir}(y_{ir})\right) \tag{76}$$

where $\mathcal{G}_{is}(y_{is}) = \mathcal{G}_{is(\tilde{k}_{is})}(y_{is})$ and $\mathcal{G}_{ir}(y_{ir}) = \mathcal{G}_{ir(\tilde{k}_{ir})}(y_{ir})$ are membership functions.

In operation, the calculation procedures for actual application of definition (75)-(76) is troublesome and tedious work, so empirical studies along this line should be carried out with consideration to the effectiveness of this method of augumenting information relative to the burden of calculation procedures necessary.

Reference

Balinski, M.L. & W.J. Baumol, (1968), The dual in nonlinear programming and its economic interpretation, Review of Economic Studies, Vol. 35, pp.237-256.

Chankong, V. and Y.Y.Haimes, (1978), The interactive surrogate worth trade-off (ISWT) method for multiobjective decision-making, in S. Zionts (ed.) Multiple Criteria Problem Solving, Springer-Verglag, New York.

Charnes, A. and W.W. Cooper, (1961), Management Models and Industrial Applications of Linear Programming, John Wiley, New-York.

Cohon, J.L. and D.H. Marks, (1975), A Review and Evaluation of Multiobjective Programming Techniques, Water Resources Research, Vol. 11, 208-220.

Dorfman, R., Samuelson P.A. and R.M. Solow, (1958), Linear Programming and Economic Analysis, McGraw-Hill, New York.

Dorfman, R., (1972), Conceptual model of a regional water quality authority, in Dorfman, R., Jacoby, H.D. and H.A. Thomas (ed), Models for Managing Regional Water Quality, Harvard University Press, Cambridge, Massachusetts.

Dubois, D. and H. Prade, (1978), Operations on fuzzy numbers, International Journal of Systems Science, Vol.9, pp. 613-626.

Dubois, D. and H. Prade, (1980), Systems of linear fuzzy constraints, Fuzzy Sets and Systems, Vol.3 , pp. 37-48.

Dyer, J.S., (1972), Interactive goal programming, Management Science, Vol.19, pp.62-70.

Dyer, J.S., (1973), An empirical investigation of a man-machine interactive approach to the solution of the multiple criteria problem, in Cochrane, J.L. and M. Zeleny (ed.), Multiple Criteria Decision Making, University of South Carolina Press, South Carolina.

Dyer, J.S., (1977), On the relationship between goal programming and multiattribute utility theory, Discussion paper No.69, Management Science Study Center, Graduate School of Management, University of California, Los Angeles.

Fishburn, P.C. (1970), Utility Thoery for Decision Making, John Wiley, New York.

Geoffrion, A.M., (1968). Proper efficiency and theory of vector maximization, Journal of Mathematical Analysis and Application, Vol. 22, pp.618-630.

Haimes, Y.Y., and W.A. Hall, (1974), Multiobjectives in water resources systems analysis, The Surrogate Worth Trade off method, Water Resources Research Vol. 10, pp. 615-624.

Haimes, Y.Y., Hall, W.A. and H.T. Freedman, (1975), Multiobjective Optimization in Water resources Systems: The Surrogate Worth Trade-off method, Elsevier, New York.

Haimes, Y.Y., (1977), Hierarchical Analyses of Water Resources Systems: Modeling and Optimization of Large Scale Systems, McGraw Hill, New York.

Haimes, Y.Y. and V. Chankong, (1979), Kuhn-Tucker multipliers as trade-offs in multiobjective decision-making analysis, Automatica, Vol. 15, pp. 59-72.

Haimes, Y.Y. and D. Macko, (1973), Hierarchical structures in water resources systems management, IEEE Transactions on Systems, Man amd Cybernetics. Vol. SMC-3, pp. 396-402.

Ijiri, Y., (1965), Management Goals and Accounting for Control, North-Holland Pub. Co., Amsterdam.

Keeney, R., (1972), Utility functions for multiattributed consequences, Management Sciences, Vol. 18, pp. 276-287.

Keeney, R., (1974), Multiplicative utility function, Operations Research, Vol. 22, pp. 22-34.

Keeney, R. and H. Raiffa, (1976), Decisions with Multiple Objectives: Preferences and Value Trade-offs, John Wiley, New York.

Kuhn, H.W. and A.W. Tucker, (1951), Nonlinear Programming, in Neyman, J. (ed.), Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, California, pp. 481-492.

Kumisawa, K., (1975), Entropy Model (in Japanese), Nikka-Giren, Tokyo.

Lasdon, L.S., (1968), Duality and decomposition in mathematical programming, IEEE Transactions on Systems Science and Cybernetics, Vol. 4, pp. 86-100.

Lasdon, L.S., (1970), Optimization Theory for Large Systems, Macmillan, London.

Locks, D.P., (1975), Conflict and choice: planning for multiple objectives, in Blitzer, C., Clark, P. and L. Tayler (ed.), Economic Wide Models and Development Planning, Oxford University Press, New York.

Luce, R.D. and P. Suppes, (1965), Preference, utility and subjective probability, in Luce, R.D., Bush, R.R. and E. Galanter (ed), Handbook of Mathematical Psychology, Vol. III., John Wiley, New York, pp. 249-410.

Luenberger, D.G., (1973), Introduction to Linear and Nonlinear Programming, Addison-Wesley, Reading, Massachusetts.

Marglin, S.A., (1962), Objectives of water-resource development: a general statement, in Maass, A., Hufschmidt, M.M., Dorfman, R., Thomas, H.A., Marglin, S.A. and D.M. Fair (ed.), Design of Water-resource Systems, Harvard University Press, Massachusetts.

Marglin, S.A., (1967), Public Investment Criteria, George Allen & Unwin, London.

Mesarovic, M.D., Macko, D. and Y. Takahara, (1970), Theory of Hierarchical, Multilevel, Systems, Academic Press, New York.

Ostron, A.R. and J.G. Gros, (1975), Application of Decision Analysis to Pollution Control: The Rhine River Study, IIASA RM-75-45, Laxenburg, Austria.

Pierre, D.A. and M.J. Lowe, (1975), Mathematical Optimization via Augmanted Lagrangians, Addison-Wesley, Reading, Massachusetts.

Pratt, J.W., H. Raiffa and R. Schlaifer, (1964), The foundations of decision under uncertainty: an elementary exposition, Journal of American Statistical Association, Vol. 59, pp. 353-375.

Raiffa, H., (1968), Decision Analysis, Addison-Wesley, Reading, Massachusetts.

Schlaifer, R., (1969), Analysis of Decision under Uncertainty, McGraw-Hill, New York.

Schlaifer, R., (1971), Computer Programs for Elementary Decision Analysis, Division of Research, Harvard Business School, Boston.

Sakawa, M. and F. Seo, (1980), An Interactive Computer Program for Subjective Systems and Its Application, IIASA WP-80-64, Laxenburg, Austria.

Seo, F., (1978), Evaluation and control of regional environmental systems in the Yodo River basin: socio-economic aspects, in Sawaragi, Y. and H. Akashi (ed.), Proceedings of IFAC Symposium on Environmental Systems Planning, Design and Control, pp.601-608, Pergamon, Oxford.

Seo, F., (1980), An integrated approach for improving decision-making processes, Behavioral Science, Vol. 25, pp. 387-396.

Seo, F. and M. Sakawa, (1979A), An evaluation method for environmental-systems planning: an alternative utility approach, Environment and Planning A, Vol. 11, No. 2, pp. 149-168.

Seo, F. and M. Sakawa, (1979B), A methodology for environmental system management: dynamic application of the nested Lagrangian multiplier method, IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-9, No. 12, pp. 794-804.

Seo, F. and M. Sakawa, (1979C), Technology Assessment and Decision Aid Utility Analysis in Fossil Fuel-to-Fuel Conversion, IIASA WP-79-35, Laxenburg, Austria.

Seo, F. and M. Sakawa, (1980), Evaluation for Industrial Land-use Program Related to Water Quality Management, IIASA WP-80-48, Laxenburg, Austria.

Seo, F. and M. Sakawa, (1981), Fuzzy Evaluation of Regional Planning with Conflicting Objectives, Paper presented at the Second Tsukuba Conference on World Regional Development and Planning, Tsukuba, Japan.

Shannon, C.E. and W. Weaver, (1949), The Mathematical Theory of Communication, University of Illinois Press, Urbana, Illinois.

Sicherman, A., (1975), An Interactive Computer Program for Assessing and Using Multiattribute Utility Functions, Technical Report No. 111. Operations Research Center, MIT, Cambridge, Massachusetts.

Simon, H.A., (1978), Rationality as process and as product of thought, American Economic Review, Vol. 68, No. 2, pp. 1-16.

Wismer, D.A. (ed.), (1971), Optimization Methods for Large-scale Systems, McGraw-Hill, New York.

Wolfe, P., (1961), A duality theorem for non-linear programming, Quarterly of Applied Mathematics, Vol. 19, No. 3, pp. 239-244.

Zadeh, L.A., (1963), Optimality and non-scalar valued performance criteria, IEEE Transactions on Automatic Control, Vo. AC-8, pp. 59-60.

Zadeh, L.A., (1965), Fuzzy sets, Information and Control, Vol. 8, pp. 338-353.

MULTIOBJECTIVE DECISION MAKING AND THE VOLTERRA EQUATIONS

M. Peschel and W. Mende

Academy of Sciences of the GDR, Berlin, GDR

INTRODUCTION

It has recently been shown that most ordinary differential equations can be transformed into Volterra equations, which have become well known because of their importance in mathematical ecology [1]. This means that we can obtain a one-to-one correspondence between the trajectories of the original ordinary differential equations and a higher-dimensional Volterra representation in the form

$$dx_i/dt = x_i\left(\sum_j G_{ij}x_j\right)$$

with constant parameters G_{ij}. The way in which this transformation may be performed is described in [2,3]. The main idea is to apply the operator $F = (d/dt)\,\ln$ consecutively to the state variables of the original ordinary differential equations.

It is very important to ensure that the Volterra representation obtained is regular, i.e., to guarantee that the Volterra trajectory is restricted to the positive cone $x_i \geq 0$. Under this assumption, equivalence transformations can be found between different Volterra representations for a given system of ordinary differential equations; this is described in more detail in the next section [4].

The replication equation [5] is very convenient for simulating and discussing the qualitative properties of Volterra representations, and is used here to solve hierarchical clustering or assignment problems. The replication equation of a Volterra representation has the following form:

$$dy_i/dt = y_i\left(\sum_j G_{ij}y_j - \sum_{k,j} y_k G_{kj} y_j\right) \quad .$$

With Volterra equations, it may be possible to unify the different concepts used to describe growth functions. A very important concept here is that of the generalized logistic growth given by a power-product driven differential equation

$$dx/dt = Kx^k(B-x^w)^\ell \quad .$$

The significance of this equation for modeling growth processes is considered in detail in [6] and [7].

Every optimization task is characterized by the growth of the benefit, and this is especially true in multiobjective decision-making. Therefore the searching processes used to find optimal solutions can be considered as growth functions, where the growth (or decline) is in the value of the objective function. A natural extension of this approach is to assume that the known properties of natural growth processes are also reflected in the growth processes leading to optimal solutions.

This paper does not present any results derived from the connection between growth dynamics and multiobjective decision-making (MOD); it simply points out, on the one hand, the possibility of finding MOD problems in the field of Volterra analysis, and, on the other, the possible application of the Volterra approach to MOD problems.

VOLTERRA REPRESENTATION OF THE TRAJECTORIES OF ORDINARY DIFFERENTIAL EQUATIONS

Consecutive application of the operator $(d/dt)\ln$ to the solution of an ordinary differential equation yields expressions

consisting of signals that have already been constructed and new signals. All new signals perform in the same way as the signals given originally [2,3]. To demonstrate this procedure we shall apply it to the power-product driven differential equation

$$dx/dt = Kx^{k}(B-x^{w})^{\ell}$$

$$x(t) = x(0)x_0$$

$$Fx_0 = Kx(0)^{k+w-1}x_0^{k-1}(B/x(0)^{w} - x_0^{w})^{\ell} = K_0x_1$$

$$Fx_1 = (k-1)K_0x_1 - \ell wK_0'x_0^{k-1+w}(B'-x_0^{w})^{\ell-1} = K_1x_1 + L_1x_2$$

$$Fx_2 = (k-1+w)K_0x_1 - (\ell-1)wK_0'x_0^{k-1+w}(B'-x_0^{w})^{\ell-1} = K_2x_2 + L_2x_1 \quad .$$

This is a regular representation with the unified initial condition $x_i(0) = 1$. We cannot find the equivalence transformations for the Volterra equations directly -- we have to generalize them slightly to the multinomial differential equations

$$Fx_i = \sum_J A_{iJ}\prod_r x_r^{a_{Jr}} \quad ,$$

with arbitrary real matrices a and A, and a finite number of power-product terms. We consider these equations only in the positive cone $x_i \geq 0$.

The following term-consistent equivalence transformations are now introduced:

$$y_j = \prod x_r^{t_{jr}^{-1}}$$

$$x_j = \prod y_r^{t_{jr}}$$

with a regular transformation matrix $T = (t_{jr})$.

It can easily be shown that multinomial differential equations preserve their form under a transformation of this type. This means that the class of matrix pairs $(T^{-1}A,aT)$ defines a huge set of equivalent multinomial differential equations. Every multinomial differential equation can be transformed into a Volterra representation if we introduce new variables for all terms. We have therefore also implicitly found a broad class of equivalence transformations for Volterra representations.

MOD PROBLEMS IN THE VOLTERRA APPROACH

First we must introduce some new notation. Let $B = (b_{ij})$ be any $m \times n$ matrix. Then we denote the row vectors by $b_{i.}$ and the column vectors by $b_{.j}$. The product of the matrix B and the vector b is then $Bb = ((b_{i.},b))$, using the scalar product (a,b) for vectors a and b. The product of two matrices B and C can then be written in the form

$$BC = ((b_{i.},c_{.j})) \quad .$$

Using this notation, the class of equivalent matrix pairs $(T^{-1}A,aT)$ can be written as follows:

$$(((t^{-1}_{i.},a_{.J})),((a_{J.},t_{.j}))) \quad .$$

Let $e_j = t_{.j}$ be any vector base in the n-dimensional state space of a Volterra system. The corresponding dual base f^i is then given by the condition of biorthogonality

$$(f^i,e_j) = \delta^i_j = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}$$

In this case obviously $f^i = t^{-1}_{i.}$. The class of equivalent matrix pairs can then be characterized in the following way:

$$((f^i,A_{.J}),(a_{J.},e_j)) \quad .$$

The ambiguity is given by the choice of any base (e_i) in the n-dimensional space.

When developing a normal form theory of multinomial differential equations we have a number of natural objectives:

1. The number of variables occurring in the multinomial differential equation should be as small as possible, i.e., $(a_{J.}, e_i)$ should contain as many zeros as possible.

2. The number of terms occurring in the multinomial differential equation should be as small as possible, i.e., $(f^j, A_{.J})$ should contain as many zeros as possible.

3. We should transform in such a way that the number of state variables with constant values (first integrals) should be as large as possible, i.e., the number of zero rows $(f^i, A_{.J})$ for all J should be as large as possible.

4. We should transform in such a way that the number of redundant variables (those found on the left-hand side but not on the right-hand side of the multinomial differential equations) is as large as possible, i.e., the number of zero columns $(a_{J.}, e_i)$ for all J should be as large as possible.

To achieve these aims using the common resource (e_i) we have to solve a polyoptimization problem. This kind of problem is obviously strongly connected with sparse matrix techniques, since to store (A,a) it is necessary to find a representation which contains as many zeros as possible.

Another important problem is the development of structure through hierarchical clustering. This is a polyoptimization situation because the greater the aggregation of particles into clusters on the first level of a hierarchy, the lower the binding forces between the clusters for combining them into superclusters on the next level of the hierarchy. To deal with the problem of hierarchical clustering, we use the Volterra approach in the form of replication equations. Let us examine such a process from an ecological viewpoint.

We consider an ecological system containing a finite number of species with relative weights x_i such that $\sum x_i = 1$ (barycentric coordinates of the species in the ecosystem). The interaction matrix G_{ij} must play an important role in the determination of the weights x_i; in this "battle of life" we use the replication form because it preserves the normalization condition $\sum x_i = 1$. This means that we model the battle between the species by the following equation:

$$dx_i/dt = x_i\Big(\sum_j G_{ij}x_j - \sum_k x_k G_{kj}x_j\Big) \quad .$$

If an equilibrium exists, it can be determined by

$$x = G^{-1}e/(e,G^{-1}e) \quad \text{with} \quad e = (1,1,\ldots,1) \quad .$$

This is something like a fair distribution between the different species of the ecosystem.

To achieve greater stability and to avoid the danger of extinction, it is useful for the species to gather into ecological subsystems, into clusters. The building of such clusters is a dynamic process for which we again use the replication equation as a modeling tool. When clustering has taken place we say that part x_{ir} of species x_i belongs to cluster r. The normalization condition is now $x_i = \sum_r x_{ir}$.

Because of the real interaction between the species given by the matrix G_{ij}, species i now has the following affinity with cluster r:

$$A_{ir} = \sum_j G_{ij}x_{jr} \quad .$$

The interaction matrix $\tilde{G}_{rs}$ between the clusters r,s is naturally defined by:

$$\tilde{G}_{rs} = \sum_i x_{ir}G_{ij}x_{js} \quad .$$

We can now see the polyoptimal character of the cluster-building process. If we have constructed well-established clusters taking into account the interaction matrix G_{ij}, we may expect high concentration values $\tilde{G}_{rr}$ and small cluster interactions $\tilde{G}_{rs}$ for $r \neq s$. Such small values of $\tilde{G}_{rs}$ are not conducive to higher aggregation to superclusters.

To form the clusters on the first level we once again use the Volterra approach through the replication equation. We have the following conditions:

$$\sum_r x_{ir} = x_i$$

$$\sum_i x_{ir} = X_r \qquad \text{(mass of cluster r)}$$

$$Fx_{ir} = \left(\sum_s \tilde{G}_{rs} x_{is} - \frac{1}{y_i} \sum_{r,s} x_{ir} \tilde{G}_{rs} x_{is} \right) .$$

In the same way we can proceed to build up superclusters, fixing the cluster mass X_r when partitioning each cluster and recombining the parts to form superclusters κ, λ. We now have the following conditions:

$$\sum_\kappa u_{r\kappa} = X_r$$

$$\sum_r u_{r\kappa} = U_\kappa \quad .$$

The interaction matrix $\bar{G}_{\kappa,\lambda}$ for the superclusters is then

$$\bar{G}_{\kappa,\lambda} = \sum_{r,s} u_{r\kappa} \tilde{G}_{rs} u_{s\lambda}$$

and the building of the superclusters is modeled by

$$Fu_{r\kappa} = \left(\sum_\lambda \bar{G}_{\kappa\lambda} u_{r\lambda} - \frac{1}{X_r} \sum_{\kappa,\lambda} u_{r\kappa} \bar{G}_{\kappa\lambda} u_{r\lambda} \right) .$$

To solve this polyoptimization problem we have to find the optimal way of coordinating the formation of clusters on the different

levels, either by stopping the process on one level and starting it on the next (higher) level or by parallel processing and control of the interaction matrices on all levels without G_{ij}.

THE VOLTERRA APPROACH TO POLYOPTIMIZATION PROBLEMS

To find efficient solutions to polyoptimization problems we suggest the following *ecological idea:*

> *The competition between different objectives in multiobjective decision-making is similar to the competition between species in an ecosystem. The seeking of efficient points resembles the aim-oriented behavior of populations.*

The consequence of this idea is:

> *Growth dynamics should be applied to multiobjective decision-making.*

This idea can be systematized in a number of ways:

1. The searching process is very often organized in the control space X of a polyoptimization problem $Q = f(x)$. A continuous seeking process, for example a gradient-like method, can be described by an ordinary differential equation D_x. This equation can be transformed into a Volterra representation V_x and vice versa. In other words, we can start with an interaction in control space (using local utilities) and arrive at a description V_x.

2. We can start with a searching process in objective space which can be modeled by a differential equation D_Q taking into account the global preference rule (for example, Pareto optimality). This can be transformed into a differential equation D_x in control space using the transformation $Q = f(x)$. We can also transform D_Q into a Volterra representation V_Q in objective space.

3. We could begin with a Volterra representation V_Q in objective space which reflects the global preference rule and maybe also utilities; this could be transformed into

a differential equation D_Q by backward integration, or into a Volterra representation $\tilde{V}_x$ in control space by making use of the transformation $Q = f(x)$, possibly through an intermediate step $\tilde{D}_x$.

Certain properties of ecological searching processes, for example initial hyperbolic growth and parabolic saturation in the neighborhood of an equilibrium, may also be useful in the search for efficient points in multiobjective decision-making [6].

REFERENCES

1. Volterra, V. (1931) Leçon sur la théorie mathématique de la lutte pour la vie. Gauthier-Villars, Paris.
2. Mende, W. and Peschel, M. (1981) Strukturentwurf für instationäre und nichtlineare Systeme, msr. H.10.
3. Peschel, M. and Mende, W. (1981) Structure building with power product driving forces, in H. Haken, Structure and Chaos in Nature, Series on Synergetics. Springer-Verlag, Berlin.
4. Peschel, M. and Mende, W. (1981) Volterra-Modelle für nichtlineare und instationäre Systeme und deren Äquivalenztransformationen.
5. Hofbauer, J. (1981) On the occurrence of limit cycles in the Volterra-Lotka equation. Nonlinear Analysis: Theory, Methods and Applications, Vol. 5 (9), pp. 1003-1007.
6. Mende, W. and Peschel, M. (1981) Probleme der mathematischen Modellierung von Evolutionsprozessen, msr. 11.
7. Mende, W. and Peschel, M. (1977) Problems of fuzzy modelling, control and forecasting of time-series and some aspects of evolution. IFAC Symposium on Control Mechanisms in Bio- and Ecosystems, Sept. 1977, Leipzig.
8. Peschel, M. and Mende, W. (1982) Do we live in a Volterra world?'-- an ecological approach to applied systems analysis. Working Paper, International Institute for Applied Systems Analysis, Laxenburg, Austria.

GOALS AND CONSTRAINTS IN FINANCIAL PLANNING

J. Spronk

Erasmus University, Rotterdam, The Netherlands

1. Introduction

In this paper we will discuss financial planning in the private sector. In spite of the fact that financial theory is still one of the strongholds of single criterion decision making, we will argue that financial planning in the private firm should be considered as a multiple criteria decision problem, the solution of which can be brought closer by employing multiple criteria decision methods.

Financial theory explicitly states that the firm is and should be maximizing the wealth of its current stockholders. Since, for quoted companies, this wealth is being determined on the capital market through supply and demand for the firm's stock, much attention has been and is being devoted to the development of models describing the capital market. In summary, these models intend to describe the determination of stock prices or - in other words - the determination of prices of claims on future and uncertain income streams.

In our opinion, the determination of prices on the capital market does not (at least not directly) depend on the goal(s) of the firm. What matters to the capital market is the quality of the income stream that can be expected from the firm. As such, the empirical evidence on the fact that the firm takes account of multiple goals (see for a discussion Spronk, 1981) should not be disturbing to financial theorists. Indeed, there seems to be a growing awareness among financial theorists that the classical assumption of stockholders' wealth maximization as the firm's one and only goal should be rejected (see a.o. Amihud and Lev, 1981, and Jensen and Meckling, 1976). Unfortunately, it is generally assumed that goals other than shareholders' wealth maximization can either be translated into 'cost' factors which can be adopted in the stockholders' wealth formula, this re-establishing a single criterion decision problem or into hard constraints, subject to which the shareholders' wealth can be maximized again. Serious objections can be made to both approaches. These will be discussed in more detail in the next section.

2. Goals and constraints in financial planning

In the remainder of this paper we assume there is a single decision-maker, who has to choose from among a set of implicitly described financial plans. Thus we have a series of instrumental variables and a number of constraints limiting the admissable values of the instruments. Together, the instruments and the constraints determine the set of admissable plans. Within the set of admissable plans, the most preferred solution should be chosen. Which alternative is most preferred, depends on the goals strived for by the decision-maker and on his perception of the goals of others. By using the term 'goal' we assume the existence of a 'goal variable' depending on the instruments, together with a specification of what is wanted with respect to this goal variable (e.g. maximization, attainment of some minimum value, etc.).

Although financial planning problems are usually formulated as single criterion optimization problems, the corresponding planning models generally contain - apart from the already mentioned cost factors - a series of 'policy constraints' [1] which are intended to take account of policy considerations other than wealth maximization. In our opinion, these policy constraints should be explicitly considered as goals rather than as constraints. To clarify this point, we will distinguish between three different kinds of policy constraints:

I. constraints which are self-imposed, i.e. the decision-maker can determine the exact position of the constraint. For instance, the decision-maker may have several goals, of which all but one are formulated as constraints subject to which the remaining goal variable is to be maximized. The main objection against such an approach is that the goals formulated as constraints have pre-emptive priority over the one which is to be maximized. A better approach is to deal with all goal variables simultaneously, because then the trade-offs between all goal variables can be considered explicitly, while the positioning of the constraints becomes a result rather than an input of the planning process.

---

1) Apart from policy constraints, a planning model may include definitional constraints and logical constraints. The present discussion is limited to policy constraints.

II. constraints which are externally imposed, which means that the decision-maker himself cannot choose the position of the constraint. In that situation, the decision-maker may or may not know the position of the constraint. Examples of the first case are existing governmental regulations, existing production capacity, etc. Of course, these are alle examples of hard constraints. Alternatively, the position of the constraint may be uncertain to the decision-maker, for instance, because this position has not yet been chosen. An example is a governmental regulation, of which the precise influence is yet unknown, but which is known to be announced in a future planning period: A fisherman who wants to buy a new ship, may have to take account of future catch limits set by the government. A more or less traditional way of dealing with such constraints is to formulate them as chance constraints. Nevertheless, situations might arise in which it makes sense to replace such a constraint by a goal variable. In our example, the fisherman might formulate 'the catch limit to be set by the government' as a goal variable - thus being able to influence the risk that the plan chosen will be (in)feasible in the future.

III. constraints which result from a game situation, where the positioning of the constraint is determined both by the actions of the decision-maker and by the actions of other participants, and where neither the decision-maker nor any of the other participants have complete control over the game. In his planning model, the decision-maker may insert estimates of the ultimate position of the constraint, although game theory learns that this estimation is a hazardous if not impossible task [1]. For example, the decision-maker may set certain limits on the firm's market share or on the magnitude of the firm's labor force. However, competitors of the firm will anticipate and react on the firm's actions, thus influencing the market share that will be realized. Similarly, trade-unions may anticipate and react on

---

1) See for a more detailed discussion Spronk, 1981, p.26.

> the firm's actions which influence its employment level. Instead of trying to predict the outcome of these games a priori, without considering the complete profile of the firm's plans, it is better to translate the variables, which are the subject of the game, as goal variables. In principle, every plan can then be described by means of a coherent vector (profile) of values attained for each of the goal variables respectively, after which it is the decision-maker's task to judge whether a particular profile is desirable and whether it has a reasonable chance of realization.

In summary, most of the above types of policy constraints should be viewed as goal variables rather than as rigid constraints. Of course, on basis of the above classification of policy constraints, a corresponding classification of goal variables can be made. In a particular financial planning problem, any of these different types of goal variables can be present, separately or in combination with other types.

Similarly, it is generally very hard - if not impossible - to translate policy consideration by means of cost factors, which should be included in the formula of the firm's single goal variable. If calculated on an ex post basis, the difference between the stockholders' wealth in case that the goal has been adopted and in case 'it would not have been adopted' should be calculated. On an ex ante basis the problem becomes even more difficult, because then the 'cost factor' may depend on the decision to be taken. Unfortunately, it is in the very nature of (financial) planning to deal ex ante with these different goals/cost factors. In more technical terms, the latter complication arises when the functional form of one goal variable depends on the value of other goal variables. Unfortunately, this is a situation which is not irrelevant. For instance, the goal variable 'wealth maximization' may depend on such entities as the quality of the products of the firm, the job security of the management, the well-being of the workers, the firm's public relations, etc. Sometimes it is possible to specify more or less exactly how one goal variable depends on another. For instance, one might try to express the stockholders' wealth as a function of the quality of the firm's products. More often however, it is practically impossible to describe how one goal variable depends on the other. In

our opinion this is among others the case for the dependence of the stockholders' wealth on private goals of managers (and other workers) such as their own job security (Amihud and Lev, 1981, provide empirical evidence that this goal indeed plays a role in practice), although financial theory almost silently assumes that these private goals can be translated ex ante into so-called agency-costs [1], independent of the decisions to be made and thus also independent of the profile of the financial plan to be chosen.

In practical planning problems, one should of course be aware of the fact that the value of one goal variable may influence the functional form of other goal variables. However, it will seldomly be possible - and if so, probably be too costly - to find out a priori how the functional form of each of the goal variables concerned precisely depends on the values of the other goal variables. Instead, one may proceed by selecting financial plans on basis of profiles which do not completely reflect all possible interdependencies between the goal variables. Then, once a plan seems attractive enough to be adopted, a more detailed study of these interdependencies can be made. Within the interactive framework to be discussed in the following section, such an approach can be integrated in a straightforward manner.

3. Financial planning as a multiple criteria decision problem

In the preceding section we showed not only that in financial planning there is generally a multiplicity of goal variables, but also that these goal variables may be of different kinds. The first point is an invitation to use multiple criteria decision methods in financial planning. The second point leads us to assume that it is practically impossible to find an overall preference function defined over the set of admissable combinations of goal variables. Among others to avoid the need of constructing such an overall preference function, it is best to choose for one of the interactive multiple criteria decision methods. To select from among the set of interactive methods, several rather common characteristics of financial planning models may be helpful (cf. Spronk, 1981):

---

1) The costs the stockholders have to make to 'persuade' management to work in the best interest of the stockholders.

- financial planning models generally include several policy constraints.
- in financial planning it is not unusual to use goal variables that are defined as ratios.
- sometimes, financial planning models include chance constraints
- in many financial planning models, goal variables occur which are neither to be maximized nor minimized (e.g. 'growth of earnings over time' and 'amount of cash to be held'). Instead, the values of these goal variables will have to meet certain upper and lower limits or will have to be as close as possible to a given target value.
- financial planning models very often include discrete (0,1) variables to represent the yes-no options inherent in the selection of capital investment projects.

Given these characteristics, it is not surprising that in literature, very often goal programming models have been proposed to deal with the financial planning problem. Because goal programming requires the specification of a detailed (dis)preference function by the decision-maker, we have developed an interactive procedure which can handle the same class of problems as goal programming does, called IMGP (Interactive Multiple Goal Programming).

IMGP needs no more a priori information on the decision maker's preferences than other multiple objective programming methods. Nevertheless, if available, such a priori information can be used in a fruitful manner, while giving the opportunity to the decision-maker to reconsider his a priori information. Since IMGP has been described in detail elsewhere (see Nijkamp en Spronk, 1980, and Spronk, 1981) we give only a brief sketch of the procedure.

In IMGP, the decision-maker has to provide information about his preferences on the basis of a pessimistic solution and an ideal solution1) presented to him. A pessimistic solution is a vector of lower limits (for the goal variables to be maximized and upper limits (for the goal variables to be minimized) for the respective goal variables. These limits are either defined directly by the decision-maker or, in cases in which this is possible and useful, derived mathematically from known properties of his preference structure and the set of alternatives. The

---

1) Possibly accompanied by other information on the state of the current solution, that may be required by the decision-maker.

ideal solution shows the optimal value, for each of the goal variables separately, given the goal values of the pessimistic solution concerned. The decision-maker merely has to indicate whether or not a solution is satisfactory, and if not, which of the pessimistic goal values should be changed. He does not have to specify how much these goal values should be changed. Nor is there any need to specify weighting factors (however, if available, these kinds of information can be used fruitfully within the procedure). On basis of the above information, a new pessimistic solution together with the corresponding ideal solution is calculated and presented to the decision-maker. Then, the decision-maker has to indicate whether the shifts in the pessimistic solution are outweighed by the shifts in the ideal solution (and in possible other indicators). If not, the goal values in the current pessimistic solution are imposed as constraints on the set of admissable alternatives. Within the thusly reduced set of admissable alternatives, the pessimistic solution can further be improved upon, and so on, until the decision-maker is satisfied with the remaining set of alternatives. This remaining set may be an ε-region around an optimal or Pareto-optimal solution, but may also - depending on the decision-maker's preferences - be a much larger set of alternatives, which for instance can be investigated further on basis of policy considerations that could not be accounted for within the model.

4. A Financial Planning Model with Multiple Goals: an Example

In this section we give an example of a financial planning model with multiple goals, and we show how a financial planning model might be selected by using IMGP1). The described model includes some features of financial planning, which are in practice rather usual. Nevertheless, every financial planning in reality has its own peculiarities. The purpose of the present section is merely to show how IMGP might help to cope with some of those peculiarities occurring frequently.

---

1) A more detailed description of the example has been given in Spronk, 1981.

Assumptions

We assume that management has to choose a combination of capital investment projects. Furthermore, it has to consider a series of financing decisions, which - assuming no equity issues and no dividend payments during the planning period - consist of a series of debt issues. Next to these decisions, management has to decide upon the amount of cash to be held in each period.

One of the most important goal variables is assumed to be the firm's total market value, defined as the sum of the unlevered present values of the projects adopted plus the market value of the tax savings implied by debt financing plus a correction term for the opportunity costs of cash holdings (cf. also Myers, 1974, and Myers and Pogue, 1974). Among others, the costs of a possible bankruptcy have been ignored. The unlevered present values are assumed to be determined by the capital market which also takes account of the risks associated with the projects. Independently of possible other goal variables, however. Of course, the negligence of certain goal variables might have a negative effect on the firm's total market value. Instead of accounting for such goal variables either by translating them into correction terms of the market value formula or by imposing hard constraints on the goal values, we deal with these goal variables explicitly and separately. As such we have chosen goal variables relating to the stability of the earnings over time, the amount of debt outstanding, the minimum and the maximum cash level, and the firm's employment level.

The model

The definitional equations of the model are given by

$$(4.1) \quad E_t = \sum_{n=0}^{20} E_t^n \cdot x^n \qquad \text{for } t = 1,\ldots,10;$$

$$(4.2) \quad C_t = \sum_{n=0}^{20} C_t^n \cdot x^n \qquad \text{for } t = 1,\ldots,10;$$

$$(4.3) \quad W_t = \sum_{n=0}^{20} W_t^n \cdot x^n \qquad \text{for } t = 1,\ldots,10; \text{ and}$$

$$(4.4)\quad C_t + D_{t+1} + [r.T-(1+r)].D_t + L_t - L_{t+1} + [(1+r)-r.T].Z_t - Z_{t+1} = 0$$

for $t = 0,\ldots,10$; where x^n, $n = 0,\ldots,20$; are (0,1) decision variables representing the possibility of adopting or rejecting each of the twenty available projects ($x^0=1$ represents the existing firm).

E_t stands for the firm's total accounting earning's before interest and taxes in period t and E_t^n is the part of the period t earnings, contributed by project n. C_t represents the firm's total after tax cash flow in period t and C_t^n is the period t cash flow associated with project n. W_t is the firm's average number of employees in period t and W_t^n is the average number of employees demanded by project n in period t. The sources and uses of funds contraints are given by (4.4), where D_t represents a one-period loan (borrowed at the beginning and repaid at the end of period t), r is the interest rate, T is the tax rate, L_t is the amount of cash hold during period t and Z_t are the firm's lendings during period t.

One of the firm's most important goal variables is its total market value, defined as

$$(4.5)\quad V = \sum_{n=0}^{20} A^n.x^n + \sum_{t=0}^{10} r.T.D_t/(1+r)^t - D_0 - \sum_{t=1}^{10} r.L_t/(1+r)^t + L_0,$$

where A^n is the unlevered present value of project n. Clearly, this goal variable is to be maximized.

A second desire of the firm is to smoothen the time path of its total accounting earnings before interests and taxes. To this end, the following time path has been defined:

$$(4.6)\quad E_t^* = E_0.(1+g_E)^t \qquad \text{for } t = 1,\ldots,10;$$

where E_t^* is the target for earnings in period t and g_E is the target growth rate. The second goal variable is defined as the maximum deviation from this growth path, which is to be minimized. Thus we have:

(4.7) Min! e^-_{max} , s.t.

$$e^-_{max} \geq e^-_t /(1+g_E)^t \qquad \text{for } t = 1,\ldots,10;$$

$$E_t - e^+_t + e^-_t = E^*_t \qquad \text{for } t = 1,\ldots,10.$$

In order to limit the risk of being unable to pay interests on debt out of current earnings, the firm in our example employs a ratio which is often used in practice: the interest cover (or times interest earned), defined as the ratio of earnings before interest and taxes to the interest charges. This ratio can be defined for each period of the planning horizon. However, in order to avoid a too large number of goal variables, we assume that the firm wants to maximize the smallest value of the ratio during the planning horizon. Or, in other words:

(4.8) Max! μ , s.t.

$$\mu \leq E_t /(r.D_t) \qquad \text{for } t = 1,\ldots,10.$$

To solve this non-linear problem, we adopted an approach suggested by Charnes and Cooper, 1977. Assuming $r.D_t > 0$ for $t = 1,\ldots,10$; the variable μ is fixed at a value $\mu = \bar{\mu}$. Then we solve

(4.9) Max! v , s.t.

$$v \leq E_t - \bar{\mu}.r.D_t \qquad \text{for } t = 1,\ldots,10;$$

and v unrestricted in sign.

If the maximum v^* is negative, $\bar{\mu}$ is infeasible and should be lowered. A positive value of v^* means that $\bar{\mu}$ if feasible and can be raised. A zero value of v^* corresponds with the optimal value $\bar{\mu} = \mu^*$. Notwithstanding its heuristic nature, this procedure worked quite well in our computations. Besides (4.9), a set of lower limits

(4.10) $E_t /r.D_t \geq 1.0 \qquad \text{for } t = 1,\ldots,10;$

was defined. These lower limits have been changed systematically during the interactive process.

The fourth goal variable concerns the minimum cash level, which is desired to approximate 5 per cent of the per period cash flow plus 5 per cent of the per period interest charges. As with the preceding goal variable, this was formalized as

(4.11) Min! λ , s.t.

$$L_t / (C_t + r.D_t) + \lambda \geq 0.05 \qquad \text{for } t = 1,\ldots,10.$$

Using the same heuristic as above, λ can be fixed at $\lambda = \bar{\lambda}$, after which

(4.12) Min! ξ , s.t.

$$L_t + (\bar{\lambda} - 0.05).(C_t + r.D_t) + \xi \geq 0 \quad \text{for } t = 1,\ldots,10;$$

gives either a positive or a zero value for ξ. If $\xi > 0$, the 'tolerance' $\bar{\lambda}$ in meeting the target ratio value 0.05 is too tight and should be relaxed. If $\xi = 0$, $\bar{\lambda}$ can be tightened. The best value of $\bar{\lambda}$ that can be tought of is $\bar{\lambda} = 0$, which occurs in the case that $L_t \geq 0.05 \cdot (C_t + r.D_t)$ for all t. Together with (4.12), the set of limits

$$(4.13) \qquad L_t \geq \lambda^{\min} \cdot (C_t + r.D_t) \qquad \text{for } t = 1,\ldots,10;$$

is formulated, where $\lambda^{\min}$ can be raised systematically.

The fifth goal value is defined as

(4.14) Max! $\sum_{t=1}^{10} \eta_t$, s.t.

$$L_t - 0.15 \cdot (C_t + r.D_t) + \eta_t = 0 \quad \text{for } t = 1,\ldots,10;$$

in order to keep the amount of cash as low as possible, while assuring that the amount of cash in each period is less than or equal to 15 per cent of the cash flow plus interest charges.

The last goal variable has been introduced to minimize the maximum number of dismissals of workers per period, in order to promote the conditions of employment and to avoid work disputes.

This desire can be formalized as:

$$(4.15) \quad \text{Min!} \; W^-_{max} \text{ , s.t.}$$

$$W^-_{max} \geq W^-_t \qquad \text{for } t = 1,\ldots,10;$$

$$W_t - W^+_t + W^-_t = W_{t-1} \qquad \text{for } t = 1,\ldots,10;$$

Results

We have done several experiments with the above model, which are described in detail in Spronk, 1981. Here, it suffices to give a brief outline. We did two kinds of experiments. First, we assumed the variables x^n, $n = 0,\ldots,20$; to be continuous (0,1)-variables. In the next stage, we assumed these variables to be discrete (0,1)-variables. In both cases we made the decisions, necessary to reach a final solution, by ourselves. In Figure 1, it is shown how this final solution was reached in the continuous case. First, minimal values for the amount of cash hold, the interest cover and the market value are required, respectively. Then, the earnings goal is formulated as a constraint in iteration 5, after which the maximum number of dismissals is limited and the constraint on the interest cover is further tightened. In iterations 8 and 9, the minimally desired market value is raised. However, given the solutions in iteration 9, the minimally desired value is pushed back in iteration 10. Finally, in iteration 11, the interest cover is raised further.

In the discrete case, we started with the pessimistic solution values of the fifth iteration in the continuous case. Of course, the choice of this starting point is somewhat arbitrary. Therefore, in practical problems, this starting point should be selected with more care. Furthermore, in the discrete case, we dropped the goal to keep the amount of cash as low as possible, since the experiments in the continuous case had shown that this goal variable was less relevant[1].

---

1) Because the maximum cash level is defined as a percentage of cash flow plus interest charges (see (4.14)), the underattainment of this maximum cash level can be enlarged by simply raising the 'antagonistic' variables D_t and Z_t simultaneously by equal amounts.

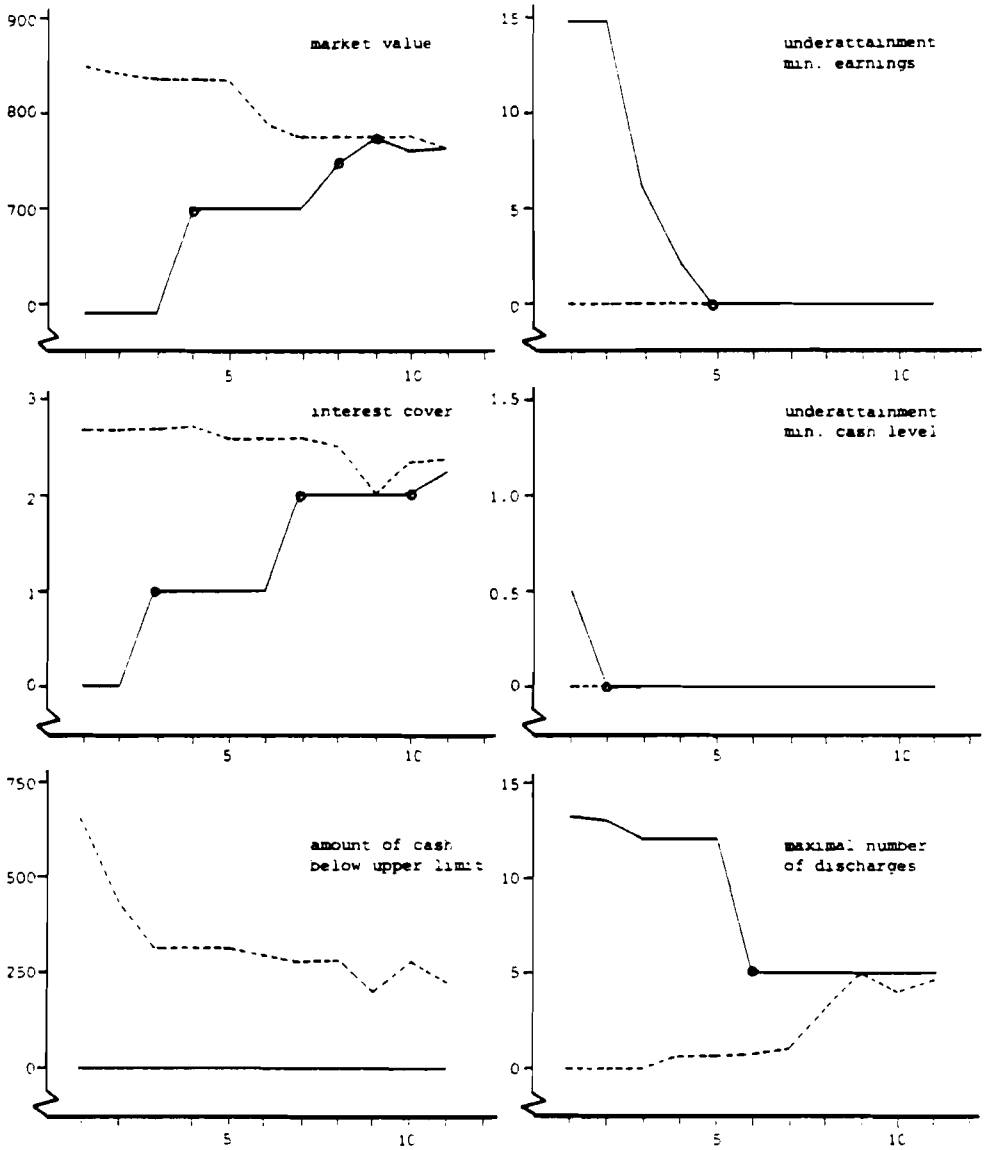


FIGURE 1. Development of the goal values from iteration to iteration

——— minimal goal values required

– – – – ideal goal values given the minimal goal values

• minimal goal value altered at this iteration

However, the constraint limiting the amount of cash to 15 per cent of the cash flow plus interest charges remained in the model. The results of the discrete experiments are not very different from those in the continuous case. Pessimistic goal values, differing only slightly from those in the final iteration of the continuous case, could be found. There are four different solutions satisfying these minimally required goal values.

We can be quite possitive about the computational aspects of IMGP in this integer case. The standard package for mixed integer programming offered by IBM, MIP, was used. The branch and bound procedure used in MIP offers several options which appear to be quite useful. For instance, if bounds on the goal variable to be optimized are known (which is one of the features of IMGP), these bound can be inserted - generally simplifying the branching process. Another possibility that was used to lighten the computational burden is to start each iteration of IMGP with an integer solution which is known to be feasible, subject to the newly added constraint. As such, the preceding optimal solution of the goal variable, of which the minimum value is raised in the new iteration, can be used. Using an IBM 370/158 computer, the CPU-time needed for each iteration was on average less than one minute, notwithstanding an incidentally limited amount of real storage. Finally, the similarity of the optimization problems, both within and between each iteration, might offer additional opportinuties to simplify the computational process.

5. Concluding Remarks

The main purpose of this paper is to stress the idea of considering financial planning as a decision problem involving multiple goals. Furthermore, we want to show that IMGP can a useful tool in dealing with this kind of problems, because (a) it can be handle many of the peculiarities inherent in financial planning, (b) it is a simple method and (c) it does not require very sophisticated information on the decision-maker's preferences. The fact that IMGP uses target goal values is, in our opinion, attractive from both a technical point of view and - since

the use of targets is rather common in practice - also from the users' point of view.

For a successful implementation it is necessary to verify not only whether the proposed procedure is not in conflict with the organizational framework of the financial planning process, but also whether its implementation is really desired and at least not counteracted by the participants involved in the planning process. For instance, managers may be quite reluctant to even mention their goals, let alone specify them in such a way that these goals can be incorporated within the model. To overcome this type of resistence, it is necessary to:

- inform all participants concerned about the aim, the assumptions, and the operation of the method.
- build the financial planning model in close co-operation with the participants.
- pay very much attention to the formulation of the goals.
- give sufficient opportunity to change and rechange the model and the goal variables, and to revise earlier choices made during the interactive process.

Some practical consequences of these desirata are a.o. that the necessary computer software should be as user-friendly as possible, and that attention should be paid to the development of an adequate information system.

Once it is recognized, that financial planning should be viewed as a multiple criteria decision problem, many areas for further research can be formulated. We mention the problem of how to deal with large numbers of goal variables, the problem of the (0,1) instrumental variables, the problem of uncertainty and fuzziness, and the problem of interdependencies between goal variabels (cf. also Spronk, 1981, Chapter 9). Another interesting area for further study concerns the phenomenon that in practice, not all potential capital investment projects are known at the beginning of the planning period, but are often proposed in the course of the period. Furthermore, if one is willing to accept the fact that the set of available investment projects is not completely given and fixed, a next step might be to include the design of new projects within the planning procedure (cf. Zeleny, 1981). Finally, we would like to mention the problem of financial planning in decentralized organizations. The procedure proposed in this paper offers interesting

opportunities to deal with the decentralization problem, because it uses lower and upper limits on the goal variables, which can be used as guidelines or budgets for different decision levels. Although it may not be expected that all of these problems can easily be solved, we believe that the use of multiple criteria decision methods in general, and of the described interactive procedure in particular, can already be beneficial to financial planning in practice.

Literature

Amihud, Y. and B. Lev (1981), Risk reduction as a managerial motive for conglomerate mergers, Bell Journal of Economics, pp.602-617.

Charnes, A. and W.W. Cooper (1977), Goal Programming and Multiple Objective Optimizations - Part I, European Journal of Operational Research, pp.39-54.

Jensen, M.C. and W.H. Meckling (1976), Theory of the firm: managerial behavior, agency costs, and ownership structure, Journal of Financial Economics, pp.305-360.

Myers, S.C. (1974), Interactions of Corporate Financing for Capital Budgeting, Journal of Finance, pp.1-25.

Myers, S.C. and G.A. Pogue (1974), A Programming Approach to Corporate Financial Management, Journal of Finance, pp.579-599.

Nijkamp, P. and J. Spronk (1980), Interactive Multiple Goal Programming: an Evaluation and Some Results, in G. Fandel and T. Gal, Multiple Criteria Decision Making Theory and Application, Springer, Berlin, pp.278-293.

Spronk, J. (1981), Interactive Multiple Goal Programming: Applications to Financial Planning, Martinus Nijhoff, Boston.

Zeleny, M. (1981), A Case Study in Multiobjective Design: de Novo Programming, in P. Nijkamp and J. Spronk (eds), Multiple Criteria Analysis: Operational Methods, Gower, Aldershot, pp.37-52.

III

MULTIOBJECTIVE PROGRAMMING AND UNCERTAINTY

PORTFOLIOS WITH STOCHASTIC BETAS: THEORY AND HEURISTICS FOR A MIXED INTEGER QUADRATIC PROGRAMMING PROBLEM

J.N. Morse

Department of Finance, Loyola College, Baltimore, Maryland, USA

Abstract

Capital market theory holds that all investors will hold a linear combination of the market portfolio and the risk-free asset, depending on their risk preferences. In fact, many portfolios are not so constructed. Due to an inability to hold all the securities in the market porfolio, a smaller number of risky assets are chosen. Current financial theory counsels the investor to maximize his return while minimizing his risk. That risk is measured by "beta", a regression coefficient which relates each security's volatility to that of the whole market.

Computed betas suffer from sampling error and the presence of the unique risk which is not explained by the regression line. Since individual betas have variance, so will the portfolio beta. The investor who wishes to maximize return given a risk level ex ante may pick a set of securities whose portfolio beta is far from the target, ex post facto. This paper attempts to reduce the probability that this will occur by constructing portfolios that minimize the variance of the portfolio beta.

This problem is formulated as a mixed integer programming problem with a quadratic objective function. Such a problem cannot be solved by using existing optimization methods, no easy solution method has been found. The problem was transformed to a series of linked Lagrangian multiplier problems. Coding of the algorithm is complete. Results show that the variance of the portfolio beta approaches zero shortly after six securities are included in the portfolio. This result is very interesting because previous research has found that a greater number of securities were necessary to obtain the maximal benefits of diversification.

The Lagrangian approach suffers from the curse of dimensionality. Because we actually enumerate the power set of a list of risky securities, the exact solution technique becomes unwieldy. To generate solutions of realistic problems, which involve large number of securities, we propose a scheme to calculate portfolio weights directly from the raw security attributes. This closed form equation is stated without proof; no computational experience is available at this time.

This paper is part of a collaborative research effort. The financial thrust of Sections II and III was developed with E. Saniga and T. McInish of the University of Delaware. The proof (forthcoming) of the conjecture in Section IV was a collaboration with A. Ben-Israel, A. Ben-Tal, both of the University of Delaware, and Dennis Karney, of Georgia Institute of Technology.

Further research will attempt to identify, a priori, the characteristics of securities which are most useful for achieving a portfolio "target" beta. For example, do the individual betas (of assets which are included in the final portfolio) tend to straddle the target beta? Do they lie near or far from the target? We also intend to examine the conditions under which short-selling (i.e. negative portfolio weights) is recommended.

I. Statement of the Problem

A. Introduction to the Portfolio Selection Problem

Modern Portfolio Theory (MAT) is a quantitative version of the proverb, "Nothing ventured, nothing gained." The particulars of MAT are articulated by the Capital Asset Pricing Model (CAPM). Developed by Sharpe [1964] and Lintner [1965] the CAPM holds that

$$r_i = r_f + \beta_i (r_m - r_f) \tag{1}$$

where r_i is the holding period return (HPR) on the i^{th} security, r_f is the risk-free rate, r_m is the HPR on the market, and β_i is an indicant of the covariance between r_i and $(r_m - r_f)$.

As explained in standard finance texts (for example, Myers and Brealey [1981, Chs. 8 - 9]),

$$\beta_i = \frac{\text{Cov}\ (r_i, r_m)}{\sigma^2_{r_m}} \tag{2}$$

This is called systematic, or market risk. Whatever variation in r_i that remains unexplained by (1) is called unique risk.

This paper is concerned primarily with systematic risk. In practice β_i is estimated using ordinary least square methods. Because we are sampling from a population, and because regression coefficients such as β_i may be non-stationary (one of the earliest articles on this topic was Sharpe and Cooper [1972]), the investor should be cautioned that the systematic risk of his portfolio is not simple to assess. That risk, β_p, has a positive standard deviation. If the investors propose a target beta (β_p) of 1.5 for example, the realized β_p might be 1 or 2. Since speculative markets are volatile, this observation warrants a more detailed treatment.

B. Managing Systematic Risk

Assume that the portfolio manager (PM) is a modern portfolio theory type (i.e. he believes, as promulgated in the citations above, that he will only be compensated for bearing systematic risk). A good procedure is to diversify away unsystematic (unique) risk, and to measure systematic risk by setting a target for the beta of the portfolio. This will be notated as β_p, where

$$\beta_p = a_1\beta_1 + a_2\beta_2 + \ldots + a_n\beta_n \tag{3}$$

in which a_1, a_2, ..., a_n are the portfolio weights (i.e. the percentage of the funds that are invested in the i^{th} security) and β_1, β_2, ..., β_n are the individual security betas. β_p indicates the PM's risk tolerance; the assessment of that quantity is not included in this paper. The present motivation is the observation that the naive choice of securities to reach a target β_p fails to recognize that the variance of the realized (or ex post facto) β_p can be controlled by a judicious selection of the portfolio's component securities.

The objective is to choose a set of portfolio weight, a_j, $j = 1,2, \ldots, n$ which minimize the variance of β_p ex ante. Note that in the sections which follow

$$\text{Var}(\beta_p) = \sigma^2_{\beta_p} = \sum_{j=1}^{n} a_j^2 \sigma^2_{\beta_j} \tag{4}$$

This variance exists because (1) is estimated by regression analysis, as

$$r_i = r_f + \beta_i (r_m - r_f) + \mu_i \tag{5}$$

where r_i, r_f, β_i and r_m are as above, and μ_i is the error term of the regression.

Note that the approach taken here is not to solve the classical Markowitz [1952] portfolio problem

$$\begin{aligned} &\text{Minimize} && aQa^T - \lambda aE(r) \\ &\text{subject to} && \sum_j a_j = 1 \\ & && a_j \geq 0,\ j = 1,2,\ldots, n \end{aligned} \tag{6}$$

in which the heretofore undefined terms are Q, a matrix of the covariances between securities, λ is the market-determined trade-off ratio between risk and reward, and E(r) is the expected return. This sort of portfolio selection generates a set of efficient (nondominated) portfolio; by specifying a risk level or a return level, a final portfolio is selected. The formulation of this paper assumes that all the relevant (market) risk of a security is expressed by stating its beta (this is the essence of Sharpe's [1971] "diagonal" model). It is further assumed that the PM defines his risk preference in terms of a beta (β_p) for the entire portfolio.

To motivate the detailed presentation of the model in section II, let us simulate the results of the naive PM, that is one who does not use the proposed optimization method. The PM wishes to choose a portfolio of two securities from a list of fourteen candidate securities. Leaving the exact description of those securities until later in the paper, 91 combinations of two securities must be screened. If the PM sets a target beta of 1, he can achieve that goal on an <u>expected</u> basis with any one of 49 portfolios. Due to regression formulation (5), $\sigma^2_{\beta_p}$, taken <u>ex post facto</u> for these portfolios, has a startling effect.

Suppose that the portfolio chosen consisted of 0% in security 1 (from Appendix I) and 100% in security 7. The quantity $\sigma(\sigma^2_{\beta_p})$ would equal $[(1^2)(.413^2)]^{\frac{1}{2}} = .412$. The β_p could end up as low as -.236 or as high as 2.236. Perhaps I have exaggerated my case by illustrating a degenerate solution. By parametrically varying the weights a_j for all 91 pairs of securities in order to achieve the minimum $\sigma^2_{\beta_p}$ portfolio (hereafter referred to simply as the minimum variance portfolio), it can be shown that $\sigma^2_{\beta_p}$ has a minimum of .015, a maximum of .139, and a standard deviation of .023. What this means is that the actual β_p might turn out to be more than 1.069 or less than .931 about 1% of the time, using the normalty assumptions of ordinary least squares

theory. For the minimum variance portfolio with $\sigma(\sigma^2_{\beta_p}) = .015$, the $\pm 3\sigma$ confidence interval would be $.955 < \beta_p < 1.045$.

Section II explains the optimization model which achieves this tighter distribution of β_p, a task whose purpose is to reduce the investor's vulnerability.

II. The Mixed Integer Quadratic Programming Problem

To gain the benefits described above we exploit the portfolio weights a_j, and the estimates $\hat{\beta}_j$, $\hat{\sigma}^2(\hat{\beta}_j)$. After this point the notation is simplified by omitting the "hats." The master problem is

$$\begin{aligned} \text{Minimize} \quad & \sigma^2_{\beta_p} = \sum_{j=1}^{n} a_j^2 \sigma^2_{\beta_j} \\ \text{Subject to} \quad & \sum_{j=1}^{n} a_j = 1 \\ & \sum_{j=1}^{n} a_j \beta_j = \beta \\ & \sum_{j=1}^{n} x_j \leq k, \; k = 2,3,\ldots,n \\ & a_j \leq x_j, \; j = 1,2,\ldots,n \\ & x_j = 0 \text{ or } 1 \\ & a_j \geq 0 \end{aligned} \tag{7}$$

with unsubscripted β as the target beta, and x_j acting as a binary (or switching) variable that facilitates the complete enumeration of the k-sized, $k \leq n$, portfolios which can be constructed from n candidate securities.

This is really a set of nested sub-problems; therefore the whole approach can be understood more readily by formulating that "typical" sub-problem successively for subsets $k = 2,3,\ldots,n$ of the n securities:

Minimize

$$\sigma^2_{\beta_p} = \sum_{j=1}^{n} a_j^2 \sigma^2_{\beta_j} \tag{8}$$

Subject to $\sum_{j=1}^{n} a_j = 1$

$\sum_{j=1}^{n} a_j \beta_j = \beta$

$a_j \geq 0$

For the k = 2 case I will be even more concrete:

Minimize

$$\sigma^2_{\beta_p} = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 \tag{9}$$

Subject to

$$a_1 + a_2 = 1$$

$$a_1\beta_1 + a_2\beta_2 = \beta$$

$$a_1, a_2 \geq 0$$

The master problem (7) is not readily solvable by standard algorithms. An approach is to reformulate (9), for all k up to a size that becomes too big too solve. The tactic is to form a series of Lagrangian multiplier problems. The k = 2 case, for example, is

$$\text{Minimize} \quad a_1^2 \sigma^2_{\beta_1} + a_2^2 \sigma^2_{\beta_2} - \lambda_1(a_1 + a_2 - 1) - \lambda_2(a_1\beta_1 + a_2\beta_2 - \beta) \tag{10}$$

Next differentiate to obtain the system of equations

$$\begin{aligned} 2a_1\sigma_1^2 - \lambda_1 - \lambda_2\beta_1 &= 0 \\ 2a_2\sigma_2^2 - \lambda_1 - \lambda_2\beta_2 &= 0 \\ a_1 + a_2 - 1 &= 0 \\ a_1\beta_1 + a_2\beta_2 - \beta &= 0 \end{aligned} \tag{11}$$

Then we can add the constants 1 and β to the last two equations and rewrite (11) in detached coefficient form as a product of a vector and a matrix each of which has structure. Again illustrating with the case of k = 2, let

$$A = [a_1 \; a_2 \; \lambda_1 \; \lambda_2]$$

$$B = \begin{bmatrix} 2\sigma_1^2 & 0 & 1 & \beta_1 \\ 0 & 2\sigma_2^2 & 1 & \beta_2 \\ -1 & -1 & 0 & 0 \\ -\beta_1 & -\beta_2 & 0 & 0 \end{bmatrix} \tag{12}$$

$$C = [\, 0 \quad 0 \quad 1 \quad \beta \,]$$

Perform the matrix operations

$$AB = C \tag{13}$$

$$ABB^{-1} = CB^{-1} \tag{14}$$

$$A = CB^{-1} \tag{15}$$

A is the desired set of portfolio weights; when (12) - (15) is solved for all desired subsets k, the minimum variance portfolio has been discovered.

III. The Results of the Analysis

A. The Data Base

Candidate securities for the portfolios were from a list of Fama [1976, p. 123]. Of his 30 securities, each described by a β_j and a σ_{β_j}, 14 were chosen by drawing random numbers from a uniform distribution. The algorithm was coded in APL on a DEC-10 computer; machine limits dictated the sample size. Three perturbations of the original data helped to test the implications of our algorithm. First, the fourteenth security was changed to the risk-free, or zero-beta asset. This was done because portfolio theory holds that investors should hold a mixture, depending on their risk preference, of the market portfolio and the riskless asset (this is the separation theorem described in Fama [op. cit.] and Myers and Brealey [op. cit.]). We wished to see if the calculated portfolios confirmed this result. Next, to test the algorithm on a wider range of data, the first four betas were first doubled, and subsequently halved. To estimate what the related σ_{β_j} would be, a regression analysis was performed on the original 14 securities, using σ_{β_j} as the

independent variable. These manipulations of the data base are described in Appendix I, along with the relevant descriptions of the securities. The algorithm was run on the four data sets, using in each case, target betas of .4, 1, and 1.6.

B. Empirical Results

First I found that each minimum variance portfolio of size k is a subset of the portfolio of size k + v, v > 1. Probably this nesting phenomenon is not surprising, given the hierarchical nature of the formulation in section II.

The second observation is best introduced by remembering that the original problem can be viewed as finding out how many securities are needed to achieve most of the benefits of diversification, given this particular optimizing paradigm. To date, research had indicated, using a naive approach to constructing portfolios, that diversification benefits leveled off as the number of securities in the portfolio reached ten to twelve (see Wagner and Lau [1971]). The present results are not directly comparable because while Wagner and Lau looked at total risk, we only looked at systematic risk.

Nevertheless, a gratifying set of results surfaced. Appendix II demonstrates conclusively that a rapid reduction of our rather narrowly defined risk measure $\sigma^2_{\varepsilon_j}$. This appendix shows that the rate of change (negative) in both the minimum and the mean approaches zero with about six (optionally chosen) securities included. Appendix III strengthens these findings by documenting a similar decline in the range and the standard deviation.

Also included in Appendix III are measures of the skewness and the kurtosis of $\sigma^2_{\varepsilon_p}$. In all but three cases (see Appendix III, Data Base #4 for the exceptions), the distributions of our figures of interest were fat-tailed with positive skewness. The results were not monotone, or even unimodal; this might be due to the feasibility constraints implied by setting a target beta with a very limited set of candidate securities. The kurtosis figures are reported for completeness only. As yet no interpretation is available.

IV. A Conjecture for Heuristic Portfolio Analysis

Because of the combinatorial nature of the proposed algorithm, an extremely long list of indices and related matrix inverses must be processed. Of course more efficient coding, switching from APL to Fortran, or migrating to a bigger machine would be fruitful. But, even these efforts, when applied to problems of practical size, will be plagued by the curse of dimensionality. For problems with many more candidate securities we have a conjecture; a closed form solution which bypasses the complete enumeration which has been described in earlier sections of this paper.

To the problem indicated by the equations (10), add two more constraints:

$$\min_j \{\beta_j\} < \beta < \max_j \{\beta_j\} \tag{16}$$

and

$$\min_j \{\beta_j\} \neq \max_j \{\beta_j\}$$

Naturally these conditions were implicit above. But they were not stated because the coding included controls that allowed only feasible solutions, in the sense of (16), to be printed. The optional solution a*, the analog of A in (15),

$$a^* = ||a_j^*|| = \frac{1}{\Delta}\left(\frac{\alpha + \gamma\beta_j}{d_j}\right) \tag{17}$$

where

$$\Delta = \left(\sum_{j=1}^{n} \frac{1}{d_j}\right)\left(\sum_{j=1}^{n} \frac{\beta_j^2}{d_j}\right) - \left(\sum_{j=1}^{n} \frac{\beta_j}{d_j}\right)^2 \tag{18}$$

$$\alpha = \sum_{j=1}^{n} \left(\frac{\beta_j^2 - \beta\beta_j}{d_j}\right) \tag{19}$$

and

$$\gamma = \sum_{j=1}^{n} \left(\frac{\beta - \beta_j}{d_j}\right) \tag{20}$$

and d_j is defined as $\sigma^2_{\beta_j}$.

The non-negativity of the portfolio weights a_j^* is indicated by observing that the Cauchy - Schwartz Inequality forces $\Delta > 0$. Since each $d_j > 0$ by definition (or observation),

$$a_j^* \geq 0 \quad \text{iff} \quad -\frac{\alpha}{\gamma} \geq \beta_j \tag{21}$$

and

$$a_j^* > 0 \quad \text{iff} \quad -\frac{\alpha}{\gamma} > \beta_j \tag{22}$$

Suppose, to further our understanding of this conjecture, that a deterministic analog of our problem were constructed, such that

$$\beta_j = md_j + c \quad , \quad m > 0 \tag{23}$$

then

$$a_j^* = \frac{1}{\Delta} \left(\frac{\alpha + \gamma \left(md_j + c\right)}{d_j}\right) \tag{24}$$

$$= \frac{1}{\Delta} \left(m\gamma + \frac{\alpha + \gamma c}{d_j}\right) \tag{25}$$

We observe that if $\alpha + \gamma c < 0$, then the larger d_j is, the larger a_j^* is. In other words they are positively correlated. Additionally, if $\alpha + \gamma c > 0$, then as d_j becomes smaller, a_j^* becomes larger.

Computational experimentation and a proof of the conjecture will be reported in the future.

V. Summary and Directions for Future Research

We hope to utilize the skewness and kurtosis data to establish estimates of the conditions under which the naive PM will do poorly, or do well, in the sense of this paper.

Using both the optimization algorithm and the conjectural heuristic, we would like to identify the types of securities that are most useful for achieving portfolio goals. Are they, for example, ones with β_j over or under β, or straddling it "nearby," or straddling it "far away?" Another study route begins by determining the optional solution for a list of candidate securities, and the associated value of $\sigma_{\varepsilon_p}^2$ for the minimum variance portfolio.

Next pick several securities with the lowest $\sigma^2_{\beta_j}$ (Set 1). Pick several securities with the highest $\sigma^2_{\beta_j}$ (Set 2). Solve for the associated weights a^*_j for Set 1 and Set 2. It is possible that one of the two solutions will be very close to the optional solution.

Short selling (i.e. negative weights) is a fact of life in securities markets. Under what conditions would this be optional? Additionally, what synthetic securities, such as can be constructed with puts and calls, would improve portfolio performance?

This paper has defined a particular portfolio selection approach in which the objective is to minimize the chance that the target beta (chosen ex ante) will be missed (ex post) by some margin. Data requirements are minimal. Proceeding from each security's beta, and the variance of that quantity, a mixed integer quadratic programming problem was formulated. This was then solved by an extended series of related Lagrangian multiplier problems.

Preliminary results show that this optional diversification does reduce the variability of the portfolios beta. This may be a promising approach to portfolio selection and revision, especially where fiduciary responsibility is stressed.

References

E. F. Fama: Foundations of Finance, Basic Books, Inc: New York, 1976.

J. Lintner: "The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," Review of Economics and Statistics, 47: 13-37 (February 1965).

H. M. Markowitz: "Portfolio Selection," Journal of Finance, 7: 77-91 (March 1952).

R. Brealey and S. Myers: Principles of Corporate Finance, McGraw-Hill Book Company, New York (1981).

W. F. Sharpe: "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," Journal of Finance, 19: 425-442 (September 1964).

W. F. Sharpe: Portfolio Theory and Capital Markets, McGraw-Hill Book Company, New York, 1971.

W. F. Sharpe and G. M. Cooper, "Risk-Return Classes of New York Stock Exchange Common Stocks, 1931-1967," Financial Analysts Journal, 28: 46-54, (March-April 1972).

W. H. Wagner and S. C. Lau: "The Effect of Diversification on Risk," Financial Analysts Journal, 27: 48-53 (November-December 1971).

APPENDIX I

THE 14 SECURITIES FROM FAMA*

SECURITY # FROM FAMA'S LIST	β_i	c_{β_i}	SECURITY # USED IN THIS PAPER
5	.69	.245	1
7	1.11	.191	2
8	1.14	.202	3
10	1.30	.201	4
12	.66	.140	5
13	.87	.177	6
15	2.24	.413	7
16	1.01	.180	8
17	1.22	.334	9
18	.58	.145	10
19	.67	.196	11
20	.14	.028	12
23	.83	.227	13
30	1.34	.522	14

The Four Data Sets

1. The original 14 securities, randomly chosen from Fama
2. Data Base #1, with 14th security changed to $\beta_i = 0$ (i.e. the risk-free asset)
   $c_{\beta_i} = 0$
3. Data Base #1, with β_i doubled for securities 1-4*
4. Data Base #1, with β_i halved for securities 1-4*

*$c_{\beta_i} = .0961 + .1307\ (\beta_i)$, for i = 1,2,3,4, as computed by regression analysis of Data Base #1

	# of combinations of k = 2,3,4,5,6 securities
2 case	91 portfolios
3 case	364 portfolios
4 case	1001 portfolios
5 case	2002 portfolios
6 case	3003 portfolios

Characteristics of Data Bases

Data Base	Average β_i	Average c_{β_i}
1	.985	.228
2	.89	.191
3	1.288	.275
4	.834	.215

* The subscript i is equivalent to j above.

APPENDIX II

ANALYSIS OF $\sigma^2_{\delta_p}$ FOR THE FOUR DATA BASES*

	$\beta = .4$		$\beta = 1.$		$\beta = 1.6$	
DATA BASE	MINIMUM	MEAN	MINIMUM	MEAN	MINIMUM	MEAN
	.0025015	.0063834	.0157088	.0361512	.0361016	.0659072
	.0015926	.0028549	.0102517	.0226994	.0309026	.0484600
1	.0012498	.0019159	.0077577	.0158739	.0251122	.0413519
	.0010617	.0014930	.0063518	.0118418	.0195146	.0355996
	.0009549	.0012530	.0054104	.0093147	.0161225	.0307846
	.0025015	.0073110	.0157088	.0302448	.0361016	.0609073
	.0015926	.0032699	.0102517	.0199686	.0309026	.0483945
2	.0012498	.0021562	.0077577	.0143628	.0243760	.0394674
	.0010617	.0016349	.0063518	.0109313	.0187037	.0323360
	.0009503	.0013301	.0054104	.0087204	.0152059	.0266745
	.0027497	.0056193	.0190743	.0367489	.0389811	.0680255
	.0017185	.0026188	.0119556	.0228260	.0258416	.0472319
3	.0013417	.0018123	.0084820	.0160254	.0204213	.0362077
	.0011511	.0014435	.0066553	.0120224	.0163204	.0293208
	.0010302	.0012318	.0055008	.0094537	.0136919	.0244517
	.0032165	.0116167	.0199706	.0482796	.0480177	.0728003
	.0018645	.0084985	.0122926	.0331548	.0423902	.0578929
4	.0014639	.0068067	.0093342	.0246316	.0322976	.0515968
	.0012346	.0056365	.0082943	.0190565	.0269379	.0459622
	.0011109	.0047295	.0074503	.0151856	.0238301	.0408969

*In each cell, the 5 numbers refer to the k portfolios, k = 2,3,4,5,6.

APPENDIX III

FURTHER ANALYSIS OF $\sigma^2_{\beta_p}$ FOR THE FOUR DATA BASES*

DATA BASE		range	standard deviation	skewness	kurtosis
1	β = .4	.0111302	.0038135	.8092466	2.2758500
		.0052941	.0011151	1.3922215	4.4995588
		.0028613	.0005025	1.5295972	5.6155747
		.0017402	.0002764	1.4396225	5.8829128
		.0011840	.0001706	1.2462705	5.4187258
	β = 1.0	.1243048	.0230184	2.4297246	9.9163552
		.0661959	.0118695	2.3177439	9.1331932
		.0622272	.0073145	3.1285751	17.7288094
		.0591481	.0044868	3.6347901	27.0688799
		.0419173	.0027317	3.3880793	27.6873396
	β = 1.6	.1159228	.0289246	2.0244906	6.7226320
		.0380943	.0099542	.1556519	1.9843573
		.0379291	.0082936	.4299027	2.4875503
		.0421261	.0075231	.5451865	3.0416177
		.0419317	.0070487	.5616002	3.1439061
2	β = .4	.0176706	.0043742	1.2215193	4.0022574
		.0065635	.0012780	1.3498642	5.0033259
		.0036046	.0006162	1.2551957	5.0853985
		.0023918	.0003574	1.1474560	4.7399084
		.0016166	.0002265	1.0874350	4.4636235
	β = 1.0	.0592414	.0122589	1.9009895	6.9259092
		.0374804	.0071060	1.6521596	6.1376190
		.0354387	.0047058	2.2115268	10.5236618
		.0346871	.0030660	2.7781773	17.5270060
		.0275169	.0019381	2.8029624	21.7052852
	β = 1.6	.0509233	.0151228	.0347605	2.0311148
		.0380943	.0099655	.2236347	1.9956217
		.0386654	.0082719	.5891961	2.8385157
		.0429370	.0076898	.7972654	2.5018768
		.0428484	.0070878	.9901556	2.7829236

*In each cell, the 5 numbers refer to the k portfolios, k = 2,3,4,5,6

APPENDIX III - continued

FURTHER ANALYSIS OF $\sigma^2_{\hat{\beta}_p}$ FOR THE FOUR DATA BASES*

DATA BASE		range	standard deviation	skewness	kurtosis
3	$\beta = .4$	.0105229	.0031836	1.0927047	3.1870972
		.0039526	.0008565	1.5380816	4.9963812
		.0020583	.0003696	1.6453896	6.3178043
		.0012903	.0001964	1.4699549	6.2218930
		.0008879	.0001179	1.1853246	5.2656060
	$\beta = 1.0$	.1209394	.0216688	2.7871741	2.3334132
		.0638184	.0101983	2.0808971	3.8710734
		.0600668	.0064783	2.3626165	3.9323939
		.0417952	.0043496	2.6963079	1.8943872
		.0318000	.0029648	3.2431066	1.1263251
	$\beta = 1.6$	.1407428	.0321537	2.3334132	7.3298971
		.1284540	.0186634	3.8710734	21.7459955
		.1220266	.0114137	3.9323939	36.1256434
		.1172542	.0079472	1.8943872	19.3457845
		.0386526	.0062613	1.1263251	4.1881324
4	$\beta = .4$	.0157218	.0049104	-.2479338	1.8372566
		.0166445	.0052294	.1970323	1.3914850
		.0162201	.0049903	.3730358	1.3824717
		.0147793	.0045691	.5022293	1.4683263
		.0142611	.0041176	.6289642	1.6006274
	$\beta = 1.0$	.1200430	.0277915	1.5377652	4.8522963
		.0881366	.0179380	1.4242418	4.4822654
		.0769155	.0130754	1.8002116	6.6256228
		.0692957	.0095713	2.2?78834	9.9542564
		.0629559	.0068776	2.5300755	13.1652474
	$\beta = 1.6$	.1040067	.0253118	2.4381145	8.1011198
		.0338643	.0073960	- .1457451	2.2386921
		.0338149	.0071419	- .0467723	2.3651551
		.0375383	.0075295	.0435641	2.6137719
		.0393020	.0078994	.1425508	2.4834050

UNCERTAINTY AND MULTIPLE OBJECTIVES IN STORAGE CONTROL PROBLEMS

S. Orlovski[1], S. Rinaldi[2], and R. Soncini-Sessa[2]

[1] *International Institute for Applied Systems Analysis, Laxenburg, Austria* and *Computing Center of the USSR Academy of Sciences, Moscow, USSR*
[2] *Centro Teoria dei Sistemi, CNR, Politecnico di Milano, Milan, Italy*

Abstract

The problem under analysis is to construct a multistage decision-making procedure for controlling releases of water from a reservoir. Two goals are considered: flood protection and satisfaction of known agricultural demands. The approach and its application to flood protection in a multireservoir hydrosystem were developed at the Computing Center of the USSR Academy of Sciences, Moscow. This paper describes the approach in detail and illustrates its validity by various computational experiments performed at IIASA using the model of Lake Como (Italy).

1. Introduction

We consider a storage system designed for covering shortages of an uncontrolled supply of a commodity with respect to demands by releasing necessary amounts of the commodity previously stored in the system during periods when the supply was abundant. A problem of controlling this type of a system consists in determining rational policies for releasing the commodity from the system in order to satisfy the demand under conditions of limited storage capacity and also of uncertainty with regard to future supplies.

As an example of such a system we consider a lake with regulated releases

* Research carried out at IIASA (Laxenburg, Austria) when the authors participated in the summer study "Real-Time Forecast versus Real-Time Management of Hydrosystems", 1981.

of water from it. Regulation of the lake should provide for rational degrees of satisfaction of demands for water from multiple downstream users and also for avoiding too high floods at the lake site, and must be based on limited information with regard to uncertain inflows of water into the lake.

Storage control problems of this type have long since been analyzed (see, for instance, Moran [3]) and many applications can be found in the fields of economics, natural resources management and industrial engineering. One essential feature of the approach suggested in this paper lies in that the information with regard to future supplies or inflows is modeled not in terms of probability distributions but in the form of explicitly described sets of sequences of all possible future inflows. Another feature is that the rationality of the control policies is based upon the trade-off's not among the actual values of the multiple objectives, which are in fact uncertain due to the uncertain nature of the inflows, but among their values which can be in some sense guaranteed. The problems analyzed in this paper are essentially based on the explicit description of a dynamic information structure which enables to give precise and practically significant interpretation to solutions of those problems. Although this approach is original, some similarities can be found in the early work due to Rippl [5] and in the extensions carried out by other authors (for a review, see Klemes [2]), which nevertheless deal only with single-objective problems of demand satisfaction.

The basic idea of the approach suggested in this paper has been developed at the Computing Center of the USSR Academy of Sciences and implemented there for flood protection problems in the multireservoir Volga River System. More recently the method has been successfully applied to the multiobjective control problem for the lake Como in Northern Italy (see Guariso et al. [1], and Orlovski et al. [4]). Some computational results of this analysis are illustrated in

Appendix to this paper.

2. General description of lake regulation problems

Let us consider a water reservoir and denote by x_t water storage (state variable) in the reservoir at the beginning of t-th time-interval with t being an integer valued time variable. We also denote by a_t (supply) the inflow of water into the reservoir during the t-th time-interval, and as r_t (control variable) - release of water from the reservoir during the same interval. We assume that an initial instant of time is fixed and is equal to 0 so the time-horizon for controlling the system is $[0,\infty]$.

Using this notation *the state equation* of the reservoir can be written in the form:

$$x_{t+1} = x_t + a_t - r_t,\ t \geq 0. \tag{2.1}$$

Constraints on the control variables r_t have the form:

$$0 \leq r_t \leq \mathbf{N}(x_t),\ t \geq 0 \tag{2.2}$$

with $\mathbf{N}(x)$ being a given function of the state variable. In the following we assume that this function possesses the properties:

$$\mathbf{N}(0) = 0,\ 0 \leq \frac{\mathbf{dN}(x)}{\mathbf{d}t} < 1,\ any\ x \geq 0.$$

Note that these properties imply $\mathbf{N}(x) \leq x$ for all $x \geq 0$ and this in turn together with the nonnegativeness of the inflows and the initial storage guarantees that $x_t \geq 0$, for any $t > 0$.

Information structure. We consider a multistage decision-making process of controlling the system in question during which the decision-maker (DM) at any current instant of time t chooses a value of the corresponding control variable r_t basing on the information available to him at that instant. We assume that in the course of the control process at any current instant of time t (beginning of t-th time interval) the DM knows the value x_t of the state variable and also the

value a_t of the inflow during t-th interval.

We also assume that at any current instant t the DM models his knowledge with regard to possible future values of the inflows from instant t+1 up to infinity in the form of a set of infinite sequences of inflows and thus considers only the elements of this set as possible future patterns or scenarios of the inflows. We shall denote this set by $\mathbf{A}_{t+1}^{\infty}$ and its elements by $a_{t+1}^{\infty}=(a_{t+1},a_{t+2},\ldots)$. In the sequel we shall use a similar notation $\mathbf{A}_t^p$ $(p\geq 0)$ for a set of finite sequences $a_t^p=(a_t,\ldots,a_p)$.

The basic element of the structure of the sets $\mathbf{A}_t^{\infty}$ is a "reference set" $\mathbf{A}_0^{*T-1}$ used for "generating" elements of the sets $\mathbf{A}_t^{\infty}$. This set consists of sequences of inflows, each of length T, where T is a time period corresponding to "seasonal" fluctuations of the inflows. In water management problems for example, the time period T is usually one year and the set $\mathbf{A}_0^{*T-1}$ may contain past observations of the inflows for a number of years together with sequences generated using statistical techniques, and also possibly hypothetic sequences added into this set by the DM to make it more representative from his/her viewpoint.

We denote by k (k=1,2,...) the number of the periods (years) counted from the initial instant t=0. Using this notation the structure of the sets $\mathbf{A}_t^{\infty}$ can be described as follows. For any instant t of the form t=(k-1)T, (k=1,2,...) (beginning of a year) the corresponding set $\mathbf{A}_{(k-1)T}^{\infty}$ consists of all infinite concatenations of the sequences from the "reference" set $\mathbf{A}_0^{*T-1}$. This means essentially that at the beginning of any new year the DM does not use any new information to model the future inflows except the a priori information that has been described in the form of the "reference" set. Note that as it follows from this assumption we have

$$\mathbf{A}_{(k-1)T}^{kT-1} = \mathbf{A}_0^{*T-1}, \quad \mathbf{A}_{(k-1)T}^{\infty} = \mathbf{A}_0^{\infty} \tag{2.3}$$

for any k=1,2,... .

On the other hand we assume that the observations made by the DM from the beginning of a year up to a current instant t can reduce his/her ambiguity with regard to the future inflows *for the remaining part of the same year*. To account for that we assume that for any current $k \geq 0$ and $t \in [(k-1)T, kT-1)$ the set $\mathbf{A}_{t+1}^{kT-1}$ may contain only some of the subsequences corresponding to the set $\mathbf{A}_0^{*T-1}$.

In conclusion, for any k=1,2,... and $t \in [(k-1)T, kT-1)$ the set $\mathbf{A}_{t+1}^{*}$ consists of all concatenations each beginning with a sequence from the set $\mathbf{A}_{t+1}^{kT-1}$ and followed by some sequence from the set $\mathbf{A}_0^{*}$.

Having in mind this structure and assuming that the set $\mathbf{A}_0^{*T-1}$ is known to the DM from the beginning of the control process, we can denote by the tuple

$$(t, x_t, a_t, \mathbf{A}_t^{kT-1}) \tag{2.4}$$

the information available to the DM at a current instant t where k is related to t as follows

$$(k-1)T \leq t \leq kT-1$$

A control law is a rule for calculating a value of the control variable for any instant of time $t \in [(k-1)T, kT-1]$, $k \geq 1$ for any tuple of information of the form (2.4). We shall describe this rule by a function

$$r_t = r(t, x_t, a_t, \mathbf{A}_{t+1}^{kT-1}). \tag{2.5}$$

In the sequel we shall also use the notation

$$\varphi(t, p, x, r, a_t^{p-1})$$

for the state of the system at instant p with x being the state at instant t, r being the control law and a_t^{p-1} sequence of inflows from t up to p-1. We assume that $\varphi(t, p, x, r, a_t^{p-1}) = x$ for p=t.

The goals of controlling the reservoir consist in preventing floods at the reservoir site and also in satisfying agricultural demands as to the releases of water from the reservoir. Formally these goals can be expressed in terms of

inequalities which are to be observed when choosing an appropriate control of the system. In our case these inequalities have the following form:

$$r_t \geq \alpha r^*_t, \quad t \in [0,\infty), \tag{2.6a}$$

$$x_t \leq \beta x^*_t, \quad t \in [0,\infty) \tag{2.6b}$$

with r^*_t, x^*_t, $t \in [0,\infty)$ being prespecified reference values. Coefficients α ($\alpha \leq 1$) and β ($\beta \geq 1$) are introduced to make these constraints more flexible. Clearly, the ideal case would be to find a control of the system for which the inequalities (2.6) hold with $\alpha=1$ and $\beta=1$. But if such a control does not exist then the constraints (2.6) can be somewhat relaxed by introducing some values $\alpha<1$ and/or $\beta>1$. In these cases the problem of controlling the system is of the two objective nature and consists in providing greater possible values of α and lower possible values of β. We assume that both r^*_t and x^*_t are periodic in the sense that

$$r^*_t = r^*_{t+T}, \; x^*_t = x^*_{t+T}$$

for any $t \geq 0$, with T being the time period introduced above.

The problems considered in this paper consist in determining couples (x_0, r) (x_0 - initial state, r - control law) which provide for the satisfaction of constraints (2.6) for given values of α, β. To be more precise we shall introduce the following

Definition 1 (feasibility of (x_t, r)). *Given system (2.1) a pair (x_t, r) is called (α,β)-feasible iff the storage $\varphi(t, p, x_t, r, a_t^{p-1})$ and the corresponding values of the control variable*

$$r_p = r(p, \varphi(t, p, x_t, r, a_t^{p-1}), a_p, A_{p+1}^{\infty})$$

satisfy constraints of the type (2.2) and (2.6) for all $p \geq 0$ and all $a_t^{p-1} \in A_t^{p-1}$.

Similarly a pair (x_t, r) will be called α-feasible [β-feasible] if the control constraints (2.2) and goal constraints (2.6a) [(2.6b)] are satisfied for all $p \geq 0$. Thus a pair (x_0, r) is (α,β)-feasible iff it is both α- and β-feasible.

In the following we shall be using notations $\mathbf{F}_t^{\alpha}$, $\mathbf{F}_t^{\beta}$ and $\mathbf{F}_t^{\alpha,\beta}$ for the sets of all

α-, β- and (α,β)-feasible pairs respectively, and consider three respective problems of determining reasonably large subsets of the sets $\mathbf{F}_0^\alpha$ (problem of satisfaction of demands), $\mathbf{F}_0^\beta$ (problem of flood protection) and $\mathbf{F}_0^{\alpha,\beta}$ (corresponding two-objective problem)

3. Problem of satisfaction of demands (problem D)

In this section we suggest a means for determining a subset $\mathbf{F}^{*\alpha}_0$ of the set $\mathbf{F}_0^\alpha$ of solutions to problem D of the following form

$$\mathbf{F}^{*\alpha}_0 = \mathbf{X}_t^\alpha \times \mathbf{R}^\alpha$$

where the set $\mathbf{X}_0^\alpha$ is described as follows

$$\mathbf{X}_t^\alpha = \{x_t \mid \textit{there exists } r \textit{ such that } (x_t, r) \in \mathbf{F}_t^\alpha\}$$

and the set $\mathbf{R}^\alpha$ is defined later in this section. Therefore, to determine the set $\mathbf{F}^{*\alpha}_0$ we can separately determine the sets $\mathbf{X}_0^\alpha$ and $\mathbf{R}^\alpha$.

Description of sets $\mathbf{X}_t^\alpha$. In what follows we shall be using notation $r^{\min}$ for the control law which specifies the value αr^*_p for the control variable at any current instant of time p. It can be shown that any set $\mathbf{X}_t^\alpha$, $(t \in [(k-1)T, kT-1],\ k \geq 1)$ can be described as follows:

$$\begin{aligned} \mathbf{X}_t^\alpha = \{x \mid & \mathbf{N}(\varphi(t, p, x, r^{\min}, a_t^{p-1})) \geq \alpha r^*_p, \\ & \varphi(t, kT, x, r^{\min}, a_t^{kT-1}) \in \mathbf{X}_0^\alpha, \\ & p \in [t, kT-1],\ a_t^p \in \mathbf{A}^{tp}\} \end{aligned} \tag{3.1}$$

To obtain an explicit description of the set $\mathbf{X}_t^\alpha$ (t, α - fixed) it suffices to find the value

$$x_t^\alpha = \min_{x \in \mathbf{X}_t^\alpha} x.$$

Then the set $\mathbf{X}_t^\alpha$ can be described as

$$\mathbf{X}_t^\alpha = \{x \mid x \in \mathbf{R}^1,\ x \geq x_t^\alpha\} \tag{3.2}$$

To obtain x_t^α the following problems may be solved

Problem-0.

$$x \to \min$$
$$\mathbf{N}(\varphi(0,p,x,r^{\min},a_0^{p-1})) \geq \alpha r^*_p$$
$$\varphi(0,T,x,r^{\min},a_0^{T-1}) \geq x$$
$$any\ p \in [0,T-1],\ a_0^p \in \mathbf{A}_0^p.$$

The solution to this problem will be denoted as x_0^α.

Problem-t, $(t \in ((k-1)T,kT-1])$

$$x \to \min$$
$$\mathbf{N}(\varphi(t,p,x,r^{\min},a_t^{p-1})) \geq \alpha r^*_p$$
$$\varphi(t,kT,x,r^{\min},a_t^{kT-1}) \geq x_0^\alpha,$$
$$any\ p \in [t,kT-1],\ a_t^p \in \mathbf{A}_t^p.$$

We denote by x_t^α the solution to this problem.

Description of the class $\mathbf{R}^\alpha$. The description of the class $\mathbf{R}^\alpha$ is given by the following

Definition 2. *A control law*

$$r_t = r(t,x_t,a_t,\mathbf{A}_{t+1}^{kT-1})$$

belongs to the class $\mathbf{R}^\alpha$ iff it possesses the property

$$\min\{\mathbf{N}(x_t),\alpha r^*_t\} \leq r_t \leq$$
$$\leq \min\{\max\{x_t+a_t-x_{t+1}^\alpha,\alpha r^*_t\},\mathbf{N}(x_t)\}, \qquad (3.3)$$
$$t \geq 0,\ x_0 \in \mathbf{X}_0^\alpha$$

The following theorem can be proved.

Theorem 1. *If a control law $r \in \mathbf{R}^\alpha$ then the couple (x_0,r) is α-feasible for any $x_0 \in \mathbf{X}_0^\alpha$.*

As follows from Theorem 2 the multistage control process providing the satisfaction of the inequalities (2.6a) for some fixed α, for all $t \geq 0$ and all sequences of inflows from the set $\mathbf{A}_0^\infty$ consists in the following. At any instant of time $t \in [(k-1)kT,kT-1]$, $k \geq 1$ in the course of the process the DM uses the information available:

$$(t,x_t,a_t,\mathbf{A}_{t+1}^{kT-1})$$

to calculate the value x_{t+1}^α and then chooses any value of r_t (release during

time-interval t) satisfying the inequalities of the type (3.3). The initial state of the system at initial instant t=0 must belong to the set $\mathbf{X}_0^\alpha$ obtained by solving Problem-0.

We can give another useful interpretation of this behavior of the DM. Let us introduce for any $x \in \mathbf{R}^1$ and a nonempty set $\mathbf{X} \subseteq \mathbf{R}^1$ a distance $\mathbf{d}$ from x to the set $\mathbf{X}$ as follows

$$\mathbf{d}(x,\mathbf{X}) = \min_{y \in \mathbf{X}} |x-y|$$

Now for any fixed $t \geq 0$ we formulate the problem:

$$\mathbf{d}(\varphi(t,t+1,x,r_t,a_t),\mathbf{X}_{t+1}^\alpha) \to \min_{r_t}$$

$$s.t. \quad \min\{\mathbf{N}(x_t),\alpha r^*_t\} \leq r_t \leq \mathbf{N}(x_t),$$

with x_t, a_t, r^*_t, α being fixed values of the respective variables.

It can easily be seen that the set of all solutions r_t to this problem is described by the inequalities (3.3). From here it follows that at any current instant t the behavior of the DM consisting in choosing a value r_t satisfying the inequalities (3.3) is equivalent to the tendency of bringing the state x_{t+1} into the "target" set $\mathbf{X}_{t+1}^\alpha$ determined on the basis of the current information available to the DM.

4. Problem of flood protection (problem F)

The analysis of problem F is quite similar to the analysis of problem D in sec. 3 and we shall use this similarity here to make the presentation more compact. In this section we determine a set $\mathbf{F}^{*\beta}_0 \in \mathbf{F}^\beta_0$ of β-feasible couples (x_0, r) that has the form similar to that of $\mathbf{F}^{*\beta}_0$ (sect. 3)

$$\mathbf{F}^{*\beta}_0 = \mathbf{X}^\beta_0 \times \mathbf{R}^\beta$$

The definitions of the sets are similar to those of $\mathbf{X}^\alpha_0$ and $\mathbf{R}^\alpha$ with the difference that here we consider the inequalities of the form (2.6b) for some fixed β and replace $r^{\min}$ with the *maximal release policy* $r^{\max}$ given by

$$r_t^{max} = r(t, x_t, a_t, \mathbf{A}_{t+1}^{kT-1}) = \mathbf{N}(x_t)$$

Similarly to what we discussed in sec. 3 it can be shown that to obtain an explicit description of the set $\mathbf{X}_t^\beta$, $(t, \beta$ – fixed) it suffices to find the value

$$x_t^\beta = \max_{x \in \mathbf{X}_t^\beta} x.$$

Then the set $\mathbf{X}_t^\beta$ can be described as

$$\mathbf{X}_t^\beta = \{x \mid x \in \mathbf{R}^1,\ x \le x_t^\beta\}.$$

To obtain x_t^β the following problems may be solved

Problem 0.

$$x \to \max$$
$$s.t.\quad \varphi(0, p, x, r^{max}, a_0^{p-1}) \le \beta x^*_p$$
$$\varphi(0, T, x, r^{max}, a_0^{T-1}) \le x$$
$$any\ p \in [0, T-1],\ a_0^p \in \mathbf{A}_0^p.$$

Solution to this problem will be denoted as x_0^β.

Problem t, $(t \in ((k-1)T, kT-1],\ k \ge 1)$

$$x_t \to \max$$
$$s.t.\quad \varphi(t, p, x, r^{max}, a_t^{p-1}) \le \beta x^*_p$$
$$\varphi(t, kT, x, r^{max}, a_t^{kT-1}) \le x_0^\beta,$$
$$any\ p \in [t, kT-1],\ a_t^p \in \mathbf{A}_t^p$$

The solution to this problem is denoted as x_t^β.

Description of the class $\mathbf{R}^\beta$. Using the definition of the distance $\mathbf{d}(x, \mathbf{X})$ from sec. 3 we formulate the following problem for any $t \ge 0$:

$$\mathbf{d}(\varphi(t, t+1, x_t, r_t, a_t), \mathbf{X}_{t+1}^\beta) \to \min_{r_t}$$
$$s.t.\quad 0 \le r_t \le \mathbf{N}(x_t),$$

with x_t, a_t, β being fixed values of the respective variables. We denote by $\mathbf{L}_t^\beta(x_t, a_t)$ a set of all solutions r_t to this problem. It can be shown that any set $\mathbf{L}_t^\beta(x_t, a_t)$ can be described by the inequalities:

$$0 \le r_t \le \min\{\max\{x_t + a_t - x_{t+1}^\beta, 0\}, \mathbf{N}(x_t)\} \qquad (4.1)$$

Using this notation we can now define the class $\mathbf{R}^\beta$ as follows:

Definition 3. *The class* $\mathbf{R}^{\beta}$ *consists of all control laws*

$$r_t = r(t, x_t, a_t, \mathbf{A}_{t+1}^{kT-1})$$

possessing the property

$$r_t \in \mathbf{L}_t^{\beta}(x_t, a_t), \quad t \geq 0$$

The following theorem can be proved which is similar to Theorem 2:

Theorem 3. *If a control law* $r \in \mathbf{R}_{\beta}$ *then the couple* (x_0, r) *is* β*-feasible for any* $x_0 \in \mathbf{X}_0^{\beta}$

Theorem 3 implies that the multistage control process providing the satisfaction of the inequalities (2.6b) (flood protection) for some fixed β, for all $t \geq 0$ and all possible future inflows consists in the following. At any instant of time $t \in [(k-1)T, kT-1]$, $k \geq 1$ in the course of the control process the DM uses the information available:

$$(t, x_t, a_t, \mathbf{A}_{t+1}^{kT-1})$$

to calculate the value x_{t+1}^{β} and chooses any value r_t (release during time-interval t) satisfying the inequality of the type (4.1). The initial state at initial instant t=0 must belong to the set $\mathbf{X}_0^{\beta}$ obtained by solving Problem-0 formulated earlier in this section.

5. Two-objective problem. Classes of semi-efficient control laws

In this section we consider problems D and F jointly or in other words we consider a problem of satisfaction of both the goal constraints (2.6) for some fixed values of α and β. In sec. 2 we have denoted the set of all solutions to this problem by $\mathbf{F}_0^{\alpha,\beta}$. Similarly to what we did in sec. 3 and 4 we determine here a subset $\mathbf{F}^{*\alpha,\beta}_0 \subseteq \mathbf{F}_0^{\alpha,\beta}$, having the following form

$$\mathbf{F}^{*\alpha,\beta}_0 = \mathbf{X}_0^{\alpha,\beta} \times \mathbf{R}^{\alpha,\beta} \tag{5.1}$$

where

$$\begin{aligned} \mathbf{X}_0^{\alpha,\beta} &= \mathbf{X}_0^{\alpha} \cap \mathbf{X}_0^{\beta}, \\ \mathbf{R}^{\alpha,\beta} &= \mathbf{R}^{\alpha} \cap \mathbf{R}^{\beta}. \end{aligned} \tag{5.2}$$

and sets $\mathbf{X}_0^{\alpha}$, $\mathbf{X}_0^{\beta}$, $\mathbf{R}^{\alpha}$, $\mathbf{R}^{\beta}$ have been defined in the previous sections.

As can be seen from (5.1) and (5.2) the set $\mathbf{F}_0^{*\alpha,\beta}$ may be determined by solving separately problems D and F for the respective values of α and β as it has been suggested in sec. 3 and 4. Combining the results from sec. 3, 4 we obtain the following

Theorem 3. *A control law*

$$\tau_t = \tau(t, x_t, a_t, \mathbf{A}_{t+1}^{kT-1})$$

belongs to the class $\mathbf{R}^{\alpha,\beta}$ *iff it possesses the property*

$$\min\{\max\{x_t + a_t - x_{t+1}^{\beta}, \alpha r_t^*\}, \mathbf{N}(x_t)\} \leq \tau_t$$
$$\leq \min\{\max\{x_t + a_t - x_{t+1}^{\alpha}, \alpha r_t^*\}, \mathbf{N}(x_t)\}$$

Another problem of interest in this context consists in obtaining in some sense "the best" values of α and β which can be guaranteed using control laws from the classes of the type $\mathbf{R}^{\alpha,\beta}$. To discuss this problem we introduce the following

Definition 4. *A pair* (α^0, β^0) *is called semi-efficient iff the set* $\mathbf{F}_0^{*\alpha^0,\beta^0}$ *is not empty and for any other pair* (α,β) *such that* $\alpha > \alpha^0$ *and* $\beta < \beta^0$ *the corresponding set* $\mathbf{F}_0^{*\alpha,\beta}$ *is empty. For any semi-efficient pair* (α^0, β^0) *we refer to control laws from the class* $\mathbf{R}^{\alpha^0,\beta^0}$ *as semi-efficient control laws.*

The problem we consider now consists in determining a class of semi-efficient control laws and the corresponding set of initial states for the system in question. The procedure for solving this problem suggested here includes the following stages:

a) obtaining a semi-efficient pair (α^0, β^0) of the values of the indicators considered,

b) obtaining the set $\mathbf{X}_0^{\alpha^0,\beta^0}$ and the class $\mathbf{R}_0^{\alpha^0,\beta^0}$ of (α^0, β^0)- feasible control laws

To obtain a semi-efficient pair (α^0, β^0) we use the following procedure: first

we fix some value α° of α and then obtain the minimal value β° of β such that the inequalities

$$r_t \geq \alpha^\circ r^\circ_t \tag{5.3}$$

$$x_t \leq \beta^\circ x^\circ_t \tag{5.4}$$

can be satisfied for all $t \geq 0$ and all sequences $a_0^\infty \in \mathbf{A}_0^\infty$. Since this analysis should be performed prior to the start of the control process the only information with regard to the possible future inflows which is available at this stage is the set $\mathbf{A}_0^\infty$.

Let us fix some value α°. Then by solving problem D we can obtain the set $\mathbf{X}_0^{\alpha^\circ}$ and the class of control laws of the type $\mathbf{R}^{\alpha^\circ}$ providing for the satisfaction of the inequalities (5.3) for all $t \geq 0$ and all sequences $a_0^\infty \in \mathbf{A}_0^\infty$. Let us introduce the notation

$$r_t^{max} = \min\{\max\{x_t - a_t - x_{t+1}^{\alpha^\circ}, \alpha^\circ r^\circ_t\}, \mathbf{N}(x_t)\}$$

for the control law corresponding to the right-hand side of the inequalities of the type (3.3) describing the class of control laws obtained at this stage. Then the value β° may be obtained as the solution to the following problem

$$\beta \to \min_{x \in \mathbf{X}_0^{\alpha^\circ}}$$

$$\varphi(0,p,x,r^{max},a_0^{p-1}) \leq \beta x^\circ_p$$

$$x_0^{\alpha^\circ} \leq \varphi(0,T,x,r^{max},a_0^{T-1}) \leq x$$

$$any\ p \in [0,T-1],\ a_0^{T-1} \in \mathbf{A}^{\circ T-1}_0.$$

Having thus obtained β° we come to the second stage to find the class $\mathbf{R}^{\alpha^\circ,\beta^\circ}$ of semi-efficient control laws together with the set of initial states $\mathbf{X}_0^{\alpha^\circ,\beta^\circ}$. The procedure for that consists in solving a problem of the type D for $\alpha=\alpha^\circ$ as in sec. 3 and a problem of the type F for $\beta = \beta^\circ$ as in sec. 4. Using (5.1) and (5.2) and Theorem 3 we can obtain the desired set $\mathbf{X}_0^{\alpha^\circ,\beta^\circ}$ and the class $\mathbf{R}^{\alpha^\circ,\beta^\circ}$

6. Concluding remarks

The formulation of the problem and its analysis here are strongly based upon the formal assumptions as to the structure of the information available to

the DM in the course of the control process. The successful implementation of this approach depends upon how adequate is that structure to the information available in real situations. In other words, the reliability of the results obtained using the approach suggested in this paper depends on how complete from DM's viewpoint is the description of all possible future inflows in the form of the reference sets $\mathbf{A}^{T-1}_{\circ}$ used to perform the analysis. If the DM considers the given set $\mathbf{A}^{T-1}_{\circ}$ not to be sufficiently representative (and therefore the results obtained using this set not sufficiently reliable) then he may apply some means to collect more information, and thus to enlarge the set $\mathbf{A}^{T-1}_{\circ}$, for example using statistical techniques (provided of course, that he considers those techniques capable of enchancing the reliability of the analysis).

References

[1] Guariso G., Orlovski S.A., Rinaldi S., "A min-max approach to the management of the lake Como", IIASA Research Report, in preparation, 1982.

[2] Klemes V., "Storage mass-curve analysis in a systems-analytic perspective", Water Resources Research, Vol. 15, No. 2, pp. 359-370, 1979.

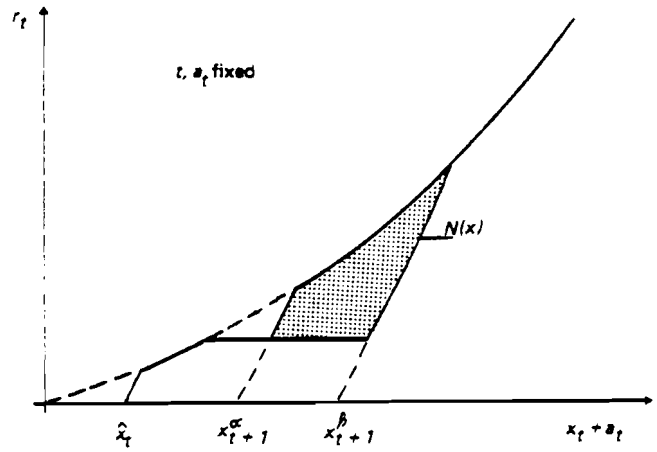
[3] Moran, P.A.P., The theory of Storage, Methuen, London, 1959.

[4] Orlovski, S.A., Rinaldi S., Soncini-Sessa R., "Uncertain inflows and multiple objectives in reservoir management (A min-max approach)", IIASA, Laxenburg, Austria, CP-82-17.

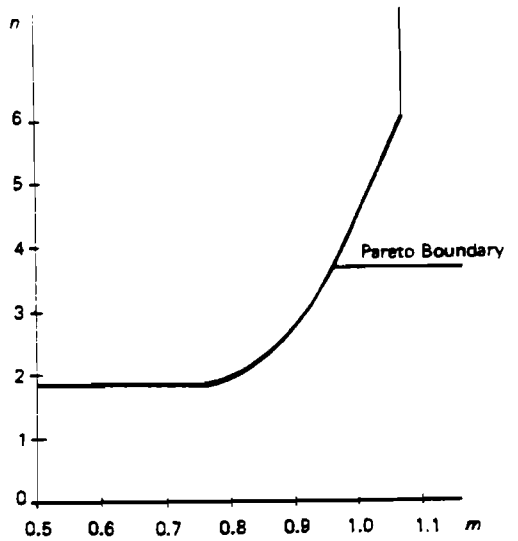
[5] Rippl W., "The capacity of storage reservoirs for water supply", Minutes Proc. Inst. Civil Eng., LXXI, pp. 270-283, (1873).

Appendix: Computational experience

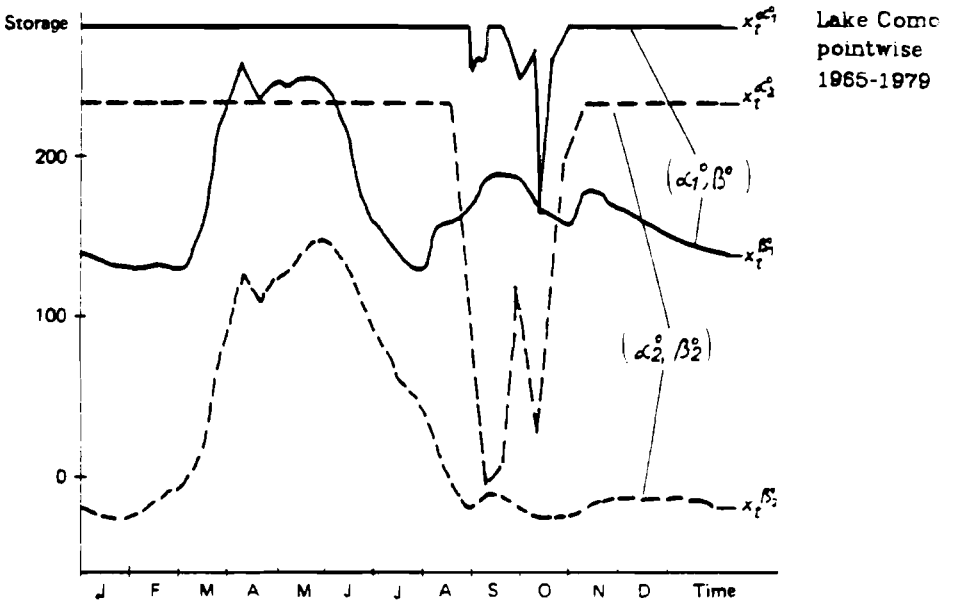
The approach outlined in this paper has been implemented for the analysis of the water management problem for the lake Como system, Northern Italy. The information on the inflows had the form of a 15-year-long sequence of daily observations of the inflows to the lake. For the analysis January 1 was chosen as the initial instant and a year (365 days) was chosen as the length of the time period T. Thus the 15-years-long sequence was "cut" into 15 one-year-long sequences each starting from January 1, and all those sequences were assumed to form the reference set $\mathbf{A}^{T-1}_{\bullet 0}$ introduced in this paper. This particular choice of the initial instant of time and of the value of T was justified by the fact that the one-year-long sequences obtained by the "cutting" procedure could adequately be considered as equally probable samples of the same stochastic 365-dimensional vector.



1. Graphical representation of the inequalities describing the class $\mathbf{R}^{\alpha^\circ,\beta^\circ}_{t}$, see sec. 5)
The shadowed region on the graph together with the solid line correspond to feasible values of the control variable r_t for given t and a_t



2 Using the procedure outlined in this paper a set of semi-efficient pairs $(\alpha^\circ, \beta^\circ)$ was obtained which is illustrated in Fig. 2.



3. The graphs in Fig. 3 illustrate 2 sets of constraints of the values of the control variable corresponding to the following semi-efficient pairs: $(\alpha_1^0 = 0.847;\ \beta_1^0 = 8.528)$, $(\alpha_2^0 = .748;\ \beta_2^0 = 1.836)$.

OPTIMIZATION UNDER UNCERTAINTY VIEWED AS A MULTICRITERIA PROBLEM

S.K. Stoyanov

Department of Automation, Higher Institute of Chemical Technology, Sofia, Bulgaria

I. INTRODUCTION

The numerical solution of optimization problems of technological objects statics and the practical realization of optimal values of control parameters presupposes the availability of an adequate mathematical model for the object.

Very often, the objects are characterized by several parameters which are subject to significant uncertainty. The main groups of uncertainties are related to parameters of the inlet flow (compositions, concentrations, etc.), to some not well known constants (technological or commercial) or to operating conditions which are sometimes unknown prior to the object design. The existance of such uncertainties for the object makes the optimization problem more complicated since the mathematical model becomes then fuzzier and the optimization solution unstable.

The need for a solution to the optimization problem under uncertainties has led to considerable development in recent years [1 + 12]. The main optimization strategies are based on the sensitivity theory or on the theory of statistical decisions. For different problem formulations different methods are recommended [1, 14]. The main aim of these problems is to find the optimization solution which is least sensitive to changes in the unknown parameters. Also the criterion used for optimization should be maximum (or minimum) in an integral sense.

The optimization problem becomes more complicated in real time. Many of the technological parameters cannot be measured at a given instant of time or information on them is essentially delayed (due to the need for analysis of raw materials, compositions, concentrations, etc.) and cannot therefore be used directly for optimal control purposes. However, this information can be accumulated, analyzed and used for subsequent optimization, particularly by using an on-line computer.

In the present paper a new algorithm is proposed where the optimization problem under uncertainties is solved as a double criteria problem.

II. FORMULATION OF THE OPTIMIZATION PROBLEM UNDER UNCERTAINTIES

The optimization of objects under uncertainties at steady state conditions can be represented by the general formulation of the problem of nonlinear programming:

$$\underset{\bar{U},\ \bar{D}}{\text{MAX}}\ F\ (\bar{U},\ \bar{P},\ \bar{D}) \tag{1}$$

subject to

$$\bar{U} \in U,\ \bar{D} \in D \tag{2}$$

$$\underline{G}(\bar{U},\ \bar{P},\ \bar{D}) \leq 0, \tag{3}$$

$$\underline{H}(\bar{U},\ \bar{P},\ \bar{D}) = 0. \tag{4}$$

For an already constructed object the design parameters will be constants.

The parameter uncertainty ($\bar{P}$) is usualy in:

i) <u>the input parameters.</u> Sometimes material, technological or economical parameters, which must be initially defined deviate from the normal values. Such parameters as kinetic constants, heat- and mass-transfer coefficients etc. experimentally defined are approximate. Other parameters deviate during exploitation (pheno-

mena of aging, the raw materials, environment). In the problems of design of technological objects a great number of the constructive parameters take part in the mathematical model as uncertain parameters if taking into account the necessity for optimal reservation.

ii) the output parameters. Some parameters being arguments of the objective functions of participating in constraint inequalities (such as prices, exploitation costs etc.) are not exactly known. This brings to the admittance of some tolerance about the functional constraints of the optimization problem.

Usually the search for the optimal control $\overline{U}^*$ and for the maximum of (1) is carried out for given nominal values of uncertainty parameters. But this solution will only be the correct one if $\overline{P} = \overline{P}_N$.

III. REVIEW OF THE OPTIMIZATION STRATEGIES UNDER UNCERTAINTIES

Optimization problems under uncertainty formulated in concern with operating objects as well as objects in design stage may be stated in various forms and consequently different strategies may be utilized for their solving.

Still a unique universal strategy does not exist, which may surve well in all cases. The different strategies proposed in literature need to be tested on real technological problems and the results obtained must be compared in order to make a correct decision about their fitness.

The strategies based on the sensitivity theory give a solution which is a compromise between the optimum of the objective function and its uncertain parametrical sensitivity.

They are used mainly in the following cases:

- the variations of the uncertain parameters are small;

- linearization of the objective function is supposed in the variation range of the uncertain parameters;

- an exact information about the boundary values of varying parameters is requested;

- a fast going optimization subroutine is requested;

- the number of uncertain parameters is not very large.

The starategies based on the statistical decisions theory give the most probable solution, which is most invariant to the critical variations of the uncertain parameters.

These strategies are used in the cases of:

- big deviations of the uncertain parameters;

- known distribution function;

- unnecessary linearization of the objective function in the area of uncertain parameters;

- requirements for high accuracy and reliability of the solution.

III.1. Strategies based on the sensitivity theory:

*including sensitivity function in the objective function [2]

$$\max_{\substack{\bar{U}\in U \\ \bar{D}\in D}} \left[F(\bar{U}, \bar{P}, \bar{D}) - \sum_{I=1}^{M} \lambda_{P_I} \left| \frac{\partial F(\bar{U}, \bar{P}, \bar{D})}{\partial P_I} \cdot \frac{P_I}{F} \right| \right] \qquad (5)$$

The use of this strategy is effective in cases when the weight coefficients λ_{P_I} of the uncertain parameters P_i are known by which the objective function is modified.

The main difficulty in using this method is the determination on the base of expert estimates test of statistically significant coefficients λ_{P_I} which indicate the importance of uncertain parameters.

The method suits well the optimization of operating technological objects.

*including sensitivity constraints [2]

$$\max_{\bar{U}\in U,\ \bar{D}\in D} F(\bar{U}, \bar{P}, \bar{D}) \tag{6}$$

subject to

$$\left|\frac{\partial F(\bar{U}, \bar{P}, \bar{D})}{\partial P_I}\cdot\frac{P_I}{F}\right| \le \xi_I, \qquad I=1, 2, \ldots, M. \tag{7}$$

It is used in cases when a range for the sensitivity ξ_I of the objective function to variations of the uncertain parameters can be stated appriory.

The solution of the problem is rather difficult in cases when in the objective functionthere exist appriory some other range constraints. It may turn out that part of these constraints exclude each other.

The utilization of this strategy is also recommended for operating objects.

*calculation of optimal design margin [3, 6, 7]

$$\min_{\substack{\Delta\bar{U}\\ \Delta\bar{D}}} \left[\sum_{I=1}^{\mathrm{DIM}\,\bar{D}} \frac{\partial F}{\partial D_I}\Delta\bar{D}I + \sum_{J=1}^{\mathrm{DIM}\,\bar{U}} \frac{\partial F}{\partial U_J}\Delta\bar{U}J\right]. \tag{8}$$

subject to

$$A_I \le \Delta Y_I(\Delta\bar{U}, \Delta\bar{D}, \Delta\bar{P}) \le B_I, \quad I=1, 2, \ldots, \mathrm{DIM}\,\bar{Y} \tag{9}$$

The method is applicable in design problems with known appriory deviations of constructive parameters $\Delta\bar{D}$ from their nominal values and ranges of variations of the objective parameters ($A_I \le \Delta Y_I\ (\Delta\bar{U}, \Delta\bar{D}, \Delta\bar{P}) \le B_I$).

In the last case the modified objective function includes the sensitivities of F to the constructive parameters.

If an optimal control is demanded then the sensitivities with respect to the control parameters are added.

*mean value optimization taking in account the sensitivity and the distribution function of the uncertain parameters.

$$\underset{\substack{\bar{U}\in U\\ \bar{D}\in D}}{MAX}\left[F(\bar{U},\bar{P},\bar{D})+\frac{1}{2}\sum_{I=1}^{M}\frac{\partial^2 F}{\partial P_I^2}\sigma_{P_I}^2+\sum_{\substack{I\\ I<J}}^{M}\sum_{J}^{M}\frac{\partial^2 F}{\partial P_I \partial P_J}\sigma_{P_I P_J}^2\right]. \quad (10)$$

It can be applied for operating objects, in cases when the objective function may be represented as a Taylor series in the area of nominal values of the uncertain parameters [5]. The multipliers in such a representation are the derivatives of the sensitivity coefficients. This strategy is a mixed one since it is necessary to know appriory the sensitivities on one hand and the distribution functions of the uncertain parameters on the other.

III.2. Strategies based on the statistical decisions theory

*minimization of the solution risk [1]

$$\underset{\substack{\bar{U}\in U\\ \bar{D}\in D}}{MIN}\left\{\int\int\cdots\int_{\bar{P}\in P}\left[\underset{\bar{U}\in U}{MAX}\,F(\bar{U},\bar{P},\bar{D})-F(\bar{U},\bar{P},\bar{D})\right]\varphi(\bar{P})\,dP\right\} \quad (11)$$

The method is used in case when the probability density function of the uncertain parameters $\varphi(\bar{P})$ is known as well as requirements are stated for a minimum deviation from the main optimal solution, generated for the nominal values of the uncertain parameters.

The method is recomanded for operating object optimization and a small number of uncertain parameters since the ammount of computations greatly increases.

*mean value optimization taking in account the distribution of the uncertain parameters [1, 8].

$$\max_{\substack{\bar{U}\in U\\ \bar{D}\in D}}\left[\int\int\cdots\int F(\bar{U},\bar{P},\bar{D})\,\varphi(\bar{P})\,dP\right]. \tag{12}$$

It is used in design problems, when a second order polynomial approximation of the objective function is possible in the variation area of the uncertain parameters and the distribution functions of the last are known.

For the approximation of the mean value of the objective function a numerical experiment is proposed realized according to an orthogonal central compositional design. This gives an independance of the estimates of influence of the uncertain parameters on the objective function.

*minimax strategy [8, 9]

$$\max_{\substack{\bar{U}\in U\\ \bar{D}\in D}}\left[\min_{\bar{P}\in P} F(\bar{U},\bar{P},\bar{D})\right]. \tag{13}$$

The minimax strategy is applied if the number of uncertain parameters P_I does not exceed 4 and their probability density functions are known.

The strategy is recommended in designing of technological objects. Since all the design boundary states are analysed they must be appriory known.

*relative sensitivity optimization [10].

$$\max_{\substack{\bar{U}\in U\\ \bar{D}\in D}}\left\{\min_{\bar{P}\in P}\left[\frac{\max\limits_{\bar{U}\in U,\ \bar{D}\in D} F(\bar{U},\bar{P},\bar{D}) - F(\bar{U},\bar{P},\bar{D})}{\max\limits_{\bar{U}\in U,\ \bar{D}\in D} F(\bar{U},\bar{P},\bar{D})}\right]\right\}. \tag{14}$$

This method can be applied when more strict requirements about the accuracy of derived optimal solution are stated. That is why a relative sensitivity S_R is introduced instead of the

objective function $F(\bar{U}, \bar{P}, \bar{D})$.

Having in mind the necessity of multiple calculation of S_R this strategy can be effectively used when the number of uncertain parameters is small and the objective function is simple.

It can be used for optimization of operating objects and objects in the design stage.

*optimal design with stochastic modelling [11].

$$\underset{\bar{D}\in D}{\text{MIN}} \left[\sum_{K=1}^{DIM\ \bar{Y}} \lambda_K \left| Y_{K\circ} - (M\{Y_K\} - \gamma \sigma_{YK}) \right| \right]. \tag{15}$$

This strategy is applied when the values of the output parameters $Y_{K\circ}$ of the object which is being designed are know appriory, and a minimum deviation from them caused by the uncertain parameters is requested.

The often used method for operating objects is to optimize the objective function, subjected to regional constraints (7) on the sensitivity of the objective function to changes in the unknown parameters. The basic difficulty however is in the proper choice of the sensitivity limits γ_I.

IV. FORMULATION OF THE NEW STRATEGY

In fact the optimization problem under uncertainties is a double criteria problem, that is, finding the control parametrs which maximize the objective function and simultaneously minimize the sensitivity function.

For the operating objects (D_I =const), the optimization problem (1) to (4) can be reformulated as follows

$$\underset{\bar{U}\in U}{\text{MAX}}\ F(\bar{U}, \bar{P}), \tag{16}$$

$$\underset{\bar{U}\in U}{\text{MIN}}\ S(\bar{D}_P^{\circ}, \bar{\lambda}_P, \bar{P}_N) \tag{17}$$

subject to (3) and (4), where

$$D_P^* = \Phi\left(\frac{\partial F(\bar{U}, \bar{P})}{\partial P_I}\right), \tag{18}$$

$$\lambda_P = F(\Delta\bar{P}, \bar{P}_N, \sigma_{PI}^2, \bar{D}_P^*). \tag{19}$$

For the modified optimization problem it is necessary to build a new objective function

$$\Psi = \Psi(F(\bar{U}, \bar{P}), S(\bar{U}, \bar{P}), W_F, W_S), \tag{20}$$

for which a maximum or minimum must be found, depending on its formulation, while also satisfying conditions (2), (3) and (4).

V. ALCORITHM OF THE METHOD

1. Formulation of the basic objective function

$$F(\bar{U}, \bar{P}) \tag{21}$$

subject to (2), (3) and (4).

2. Formulation fo the new objective function

$$S = S\left(\frac{\partial F(\bar{U}, \bar{P})}{\partial P_I}, \lambda_{P_I}, P_{NI}\right), \quad I = 1, 2, \ldots, M. \tag{22}$$

The sensitivity function (22) can be built in different ways:

2.1.

$$S = \sqrt{\sum_{I=1}^{M} \lambda_{P_I}\left(\frac{\partial F(\bar{U}, \bar{P})}{\partial P_I}\right)^2 P_I = P_{NI}}, \tag{23}$$

2.2.

$$S = \sum_{I=1}^{M} \lambda_{P_I}\left|\frac{\partial F(\bar{U}, \bar{P})}{\partial P_I}\right| \Big/ \sum_{I=1}^{M}\left[\frac{\partial F(\bar{U}, \bar{P})}{\partial P_I}\right]^2, \tag{24}$$

2.3.

$$S = \sum_{I=1}^{M} \lambda_{P_I}\left|\frac{\partial F(\bar{U}, \bar{P})}{\partial P_I} \cdot \frac{P_{NI}}{F(\bar{U}, \bar{P}_N)}\right|. \tag{25}$$

We recomand usage of (23) or (24). It is generally more convenient to use normalized sensitivities such as (25) but when $P_{NI} \approx 0$ or $F(\bar{P}_N) \approx 0$ the solution becomes unstabile.

3. Formulation of the weighted coefficients λ_{pl} for $\bar{P}$.

4. Finding the $\bar{U}_1^*$ for which

$$F_{MAX} = \max_{\bar{U}\in U} F(\bar{U}, P_N). \tag{26}$$

5. Finding the $\bar{U}_2^*$ for which

$$S_{MIN} = \min_{\bar{U}\in U} S(\bar{U}, P_N). \tag{27}$$

6. Formulation of the new objective function

$$6.1.\quad \Psi_1(\bar{U},\bar{P}) = \left|\frac{F_{MAX} - F(\bar{U},\bar{P})}{F_{MAX}}\right| W_F + \left|\frac{S_{MIN} - S(\bar{U},\bar{P})}{S_{MIN}}\right| W_S, \tag{28}$$

or

$$6.2.\quad \Psi_2(\bar{U},\bar{P}) = \frac{(F(\bar{U},\bar{P}) - F_{MIN})W_F}{F_{MAX} - F_{MIN}} + \frac{(S_{MAX} - S(\bar{U},\bar{P}))W_S}{S_{MAX} - S_{MIN}}. \tag{29}$$

7. Finding the optimum of (28) or (29)

$$7.1.\quad \min_{\bar{U}\in U} \Psi_1(\bar{U},\bar{P}_N), \tag{30}$$

or

$$7.2.\quad \max_{\bar{U}\in U} \Psi_2(\bar{U},P_N). \tag{31}$$

In this algorithm the function of losses of optimal solutions (26) and (27) can be used when $S_{MIN} > 0$. When $S_{MIN} \approx 0$ the solution (30) becomes unstabil and the function (28) loses its meaning. For these cases use of useful function (29) can be recommended. Its main disadvantage is the need a four-fold optimization problem solution in order to find maximum and minimum for both $F(\bar{U}, \bar{P})$ and $S(\bar{U}, \bar{P})$.

One of the main difficulties in using algorithm is to obtain the weifhted coefficients W_F and W_S for the two objects functions. Except for cases where $W_F = W_S$, the weighted coefficients can be taken to confirm to differences between S_{MAX} and S_{MIN} or F_{MAX} and F_{MIN} within the feasible region. When the relative differences in F or S are large it is advisable to increase the cor-

responding weighted coefficient.

For different practical problems, the sensitivity function (17) can be of different complexity. Thus, in the simplest cases it is possible for this function to be conctant; otherwise it can be a linear or non-linear function of the control parameters. The proposed algorithm is recommended for use with non-linear sensitivity function.

VI. ALGORITHM FOR THE DEFINITION OF THE WEIGHTED COEFFICIENTS OF THE UNCERTAIN PARAMETERS

The weighted coefficients λ_{P_I} in (17) can be defined according to the available information about the uncertain parameters P_I or according to the opinion of specialists (expert evaluation).

The available information about some of the uncertain parameters might be:

- variations range

$$\Delta P_i = P_{MAXI} - P_{MINI} \tag{32}$$

- dispersion $G^2_{P_I}$ and rated value P_{NI}

- sensitivity of the objective function to the variations of the uncertain parameters with regard to their rated value:

$$S^N_{P_i} = \left(\frac{\partial F(\bar{u},\bar{p})}{\partial p_i} \Big/ |\nabla F|\right)_{p_i = P_{NI}} \tag{33}$$

When the variations range or the dispersion of the uncertain parameters are large, or when the objective function reveals high sensitivity, the weighted coefficients λ_{P_I} must be increased, λ_{P_I} can be calculated by the formulae (34) ÷ (38), the proper choice depending on the available information about P_I:

$$\lambda_{P_I} = \frac{D_{P_I}}{\sum_{I=1}^{M} D_{P_I}}, \tag{34}$$

where

$$D_{P_I} = \frac{2\Delta P_I}{P_{MAX_I} + P_{MIN_I}}, \qquad I = 1, 2, \ldots, M, \tag{35}$$

or

$$\lambda_{P_I} = \frac{A_{P_I}}{\sum_{I=1}^{M} A_{P_I}}, \tag{36}$$

where

$$A_{P_I} = \frac{\sqrt{G_{P_I}^2}}{P_{N_I}}, \qquad I = 1, 2, \ldots, M, \tag{37}$$

or

$$\lambda_{P_I} = \frac{S_{P_I}^N}{1 - S_{P_I}^N}, \quad I = 1, 2, \ldots, M. \tag{38}$$

Table 1.

Uncertain parameters / Specialists	P_1	P_2	...	P_I	...	P_M
1	A_{11}	A_{12}	...	A_{1I}	...	A_{1M}
2	A_{21}	A_{22}	...	A_{2I}	...	A_{2M}
3	A_{31}	A_{32}	...	A_{3I}	...	A_{3M}
. . .						
J	A_{J1}	A_{J2}	...	A_{JI}	...	A_{JM}
. . .						
R	A_{R1}	A_{R2}	...	A_{RI}	...	A_{RM}
$\sum_{J=1}^{R} A_{JI}$						
δ_I						
V_I						
λ_{P_I}						

If there is not the above mentioned characteristics, the weighted coefficients λ_{P_I} might be defined by the experts' evaluations, related to the degree of the influence of a given parameter P_I upon the problem under consideration and the optimal solution.

Applying some of the assumptions of the rank correlation approach [16], we would like to propose the following algorithm:

1. "R" specialists were asked to present their opinion about the influence of "M" uncertain parameters, ranking them by their importance from 1 to M.

2. The derived results are formed in a weighted matrix, where each element A_{JI} defines the weight(position), which was prescribed to the uncertain parameter P_I by the expert J (Table 1).

The calculations of the weighted coefficients λ_{P_I} can be performed only in case the experts' opinions coincide. The coincidence is defined by the concordance coefficient W_K.

3. W_K is calculated as

$$W_K = \frac{12 \sum_{I=1}^{M} \delta_I^2}{MR^2(M^2 - 1)}, \tag{39}$$

where

$$\delta_I = \sum_{J=1}^{R} A_{JI} - S_c \tag{40}$$

and

$$S_c = \frac{R(M+1)}{2}. \tag{41}$$

4. The adequacy of the concordance coefficient W_K is evaluated.

4.1. If $M \geq 7$, calculate χ^2_{CALC}

$$\chi^2_{CALC} = R(M-1)W_K \tag{42}$$

The concordance coeffecient is adequate if

$$\chi^2_{CALC} > \chi^2_{TAB(\alpha,\nu)}, \tag{43}$$

where the degrees of freedom are defined as

$$\nu = M - 1. \tag{44}$$

4.2. If $M<7$, calculate

$$F_{CALC} = \frac{1}{2} \ell n \frac{(R-1) W_K}{1 - W_K}. \tag{45}$$

In order W_K to possess sufficient adequacy, it is necessary:

$$F_{CALC} > F_{TAB}(\alpha, \nu_1, \nu_2), \tag{46}$$

at

$$\nu_1 = M - 1 - \frac{2}{R}, \tag{47}$$

at

$$\nu_2 = (R-1).\nu_1. \tag{48}$$

5. If the experts' evaluations exibit expressed concordance, i.e., if W_K is adequate, the weighted coefficients are calculated as follows

$$\lambda_{p_I} = \frac{V_I}{\sum_{I=1}^{M} V_I}, \qquad I = 1, 2, \ldots, M, \tag{49}$$

where

$$V_I = \frac{RM - \sum_{J=1}^{R} A_{JI}}{R(M-1)}. \tag{50}$$

For the derived weighted coefficients stands the condition

$$\sum_{I=1}^{M} \lambda_{p_I} = 1. \tag{51}$$

The described algorithm provides the possibility to define reliable weighted coefficients, granted that the inquired experts are pretty well acquainted with the object in question and the contents of the formulated objective function $F(\bar{U}, \bar{P})$.

VII. CONCLUSION

The optimization algorithm under uncertainties expresses as a double criteria problem can be used in two basic cases:

1. In off-line, with a given objective function (1) under constraints (2), (3) and (4), and for given primary information for uncertainty parameters: nominal values, dispersion or limits. Availability of every of those parameters is not required. It should be kept in mind that, for multicriteria optimization problems, not only generalized function (28) and (29) can be used also other schemes for compromise, following the analysis of compromising solutions [13, 15].

2. In on-line application, it is possible to accumulate information for uncertainty technological parameters, using real time computer, and to employ this algorithm with either a statistical model or an analytical model with experimentally adjustable constants.

The alforithm has been realized in FORTRAN 4 and Real Time FORTRAN for both cases, and successfully applied in off-line for the solution of optimization problems in the control of an ammonia - water desorption colomn as well as heat exchanges for dusted gases.

NOTATION

$\bar{U}$	- vector of control parameters
$\bar{U}^*$	- vector of optimal values of control parameters
$\bar{Y}$	- vector of output parameters
$\bar{P}$	- vector of parameters under uncertianties
M	- number of parameters under uncertainties
P_N	- vector of nominal values of uncertainty parameters
$\bar{D}$	- vector of design parameters
$\bar{\lambda}_P$	- vector of weighted coefficients for uncertainty parameters

$F(\bar{U}, \bar{P})$ - basic objective function for technological object

$S(\bar{U}, \bar{P})$ - sensitivity function

σ_P^2 - dispersion for uncertainty parameters

P_{MIN}, P_{MAX} - limits for incertainty parameters

W_F - weighted coefficient for basic objective function

W_S - weighted coefficient for sensitivity function

S_{MAX}, S_{MIN} - maximum and minimum values of sensitivity function

F_{MAX}, F_{MIN} - maximum and minimum values for basic objective function

$\Psi_1(\bar{U}, \bar{P})$ - generalized function of losses

$\Psi_2(\bar{U}, \bar{P})$ - generalized useful function

R - number of experts

S_C - mean weighted sum

δ_I - deviation of the sum of the weights for each uncertain parameter from the mean sum

W_K - concordance coefficient

F_{TAB} - tabulated value of Fisher's criterion

α - probability level

REFERENCES

1. Dittmar, R., G.Reining, K.Hartmann, Chem.Technik, v.28, No.5, No.8, 1976, p.264, p.460.

2. Chen, M.S.K., L.E.Erickson, L.T.Fan, I & EC Proc. Des.Devel. v.9, No.1, 1970, p.53, and No.4, 1970, p.514.

3. Takamatsu, T., I.Hashimoto, S.Shoiya, J.Chem.Eng. Japan, v.6, No.5, 1973, p.453.

4. Grossmann, I.E., R.W.H.Sargent, AIChE Journal, v.24, No.6, 1978, p.1021.

5. Seinfeld, J.H., W.L. McBridge, I & EC Proc. Des. Devel., v.9, No.1, 1970, p. 53.

6. Takamatsu, T., I.Hashimoto, H.Ohno, I & EC Proc. Des. Devel., v.9, No.3, 1970, p.368.

7. Dittmar, R., K. Hartmann, Chem. Engrg. Sci., v.31, No.7,

1976, p.563.

8. Watanabe, N., Y.Nishimura, M.Matsubara, Chem. Engrg. Sci., v. 28, No.3, 1973, p.905.

9. Nishida, N., A.Ichikawa, T.Tazaki, AIChEJ, v.18, No.3, 1972, p.561.

10.Rohrer, R.A., M.Sobral, IEEE Trans. on AC, v.10, No.1, 1965, p.43.

11.Berryman, J.E., D.M.Himmelblau, I & EC Proc.Des. Devel., v.10, 1971, p.441.

12.Freeman, R.A., J.L.Gaddy, AIChEJ, v.21, No.3, 1975, p.436.

13.Stoyanov, S.K., Ph.D. Thesis, Higher Inst. of Chem.Techn., Sodia, 1975.

14.Stoyanov,S.K., R.B.Pavlova, Pros.Chemplant '80, Heviz, Hungary, 1980.

15.Nonova, A.N., Ph.D.Thesis, Higher Inst. of Chem. Techn. Sofia, 1978.

16.Kendall, M.G.Rank correlation methods, London, Griffin, 1957.

FUZZY MULTIOBJECTIVE PROGRAMMING WITH COMPOSITE COMPROMISES

C. Carlsson

Department of Business Administration, Abo Academy, Abo, Finland

Abstract

The linear multiobjective programming technique developed by Zeleny is based on the results on cone convexity and nondominated solutions derived by Yu. Takeda and Nishida developed fuzzy counterparts of Yu's results, and several authors have developed various forms for fuzzy programming. Here we will draw upon these results to develop a method for fuzzy multiobjective programming which offers means for composite compromises. The method is intended for tackling decision problems in which we have a hierarchical structure of explicitly interdependent aspirations, and the decision maker is initially uncertain about his preference structure . The FMOP-C model is evolved from a fuzzy domination structure and applies some results on cone convexity, relative distances and hierarchical structures to obtain a fuzzy subset of the fuzzy set of nondominated solutions. This subset represents both feasible, "almost optimal" and relevant compromises in relation to all the aspirations of the decision maker. The method is illustrated with a fairly simple example.

1. INTRODUCTION

There are normally a few prerequisites we take as given when working with mathematical programming models - both when we develop algorithms and when we apply the models in support of planning and/or decision making. These prerequisites are (i) the model covers essential aspects of the planning and/or decision situation at hand, as it is perceived by some interested party; (ii)

the context covered by the model is reasonably well controlled by the interested party; (iii) the relationships which can be identified within the context are adequately enough represented in a mathematical programming model; (iv) the interested party recognizes active constraints and is able to formulate aspirations to be met in the context, and (v) it is possible to translate relationships, constraints and aspirations into the elements of a mathematical programming model. If these prerequisites are met the model may be used for guiding the interested party in how to meet both aspirations and constraints in the context at hand. This process is referred to as a problem-solving process, and its outcome is not seldom regarded as normative for the interested party, as the mathematical programming technique is aimed at finding some "best" or "optimal" solution.

But should we really take these prerequisites as given? Peters [10] points to the fact that a CEO (whom we shall take to be our "interested party") in reality cannot solve problems, because (i) his attention is too fragmented; (iii) the issues come too late to him, and (iii) the information on which he should react has become too imprecise due to excessive filtering and aggregation. Peters (cf [10], p 170) finds that the only options for a CEO is to "generally shape business values" in his organization and to "educate by example". This kind of reality is, however, a far cry from the context of a mathematical programming model, and seems to indicate that the methodology and the thinking behind the mathematical programming technique does not address the relevant issues (cf also Zeleny [21]).

That is probably true if we take the normative aspect of the mathematical programming model too literally - which is too often done by the critics of O.R.. If we, on the other hand, emphasize the planning and decision supporting aspects of the model-aiming more at a tool for description and experiments - the technique is not so irrelevant. The reasons for that are (i) the technique can be developed and adaptated to the relevant context; (ii) if some elements in a decision and/or planning context can be treated with some techniques, it is easier to handle the rest by "shaping business values" and "educating by

example" (providing that the techniques do not reduce the degrees of freedom too much), and (iii) the CEO cannot handle to much detail, but his staff should carry out some evaluation, screening, filtering and aggregation, for which efficient and operational techniques are needed.

Here we will not deal with all the indicated aspects of reality, but emphasize two aspects, which seem to be crucial both for efficient and relevant modelling, and for carrying out decision and planning tasks in a managerial context: (i) imprecision and (ii) multiple, interdependent aspirations. If these two aspects are present a managerial context is described as "complex" or "messy" (cf [2], [5],[6]), which according to Peters [10] would be a fairly accurate description of reality. Here we will now adapt the mathematical programming technique to imprecision and multiple, interdependent aspirations.

The methodological tools for handling multiple, interdependent aspirations are to be found in the field of research called MCDM. This field has in recent years matured into a scientific endeavour in its own right (cf [4], [7], [11], [12], [22]) as it has been possible to show that many of the standard decison and planning models are special cases of some more developed MCDM-models (cf [22]). Here we will apply some results on domination structures among multiple criteria which have been presented by Yu [18] and Zeleny [20]; we will also apply some elements of the ideal point (cf Zeleny [19]) and reference point techniques (cf Wierzbicki [14]); the framework in which these elements are put together is one of fuzzy, multiobjective programming (cf Carlsson [4], [6], Orlovsky [9], Yager [15], [17]).

The reason for using fuzzy programming is that we wanted to deal with imprecision. The traditional way to deal quantitatively with imprecision is to employ the concepts and techniques of probability theory. But Bellman and Zadeh [1] emphasize that imprecision is not the same as randomness; this is even more the case in a planning and decision making context in which imprecision may come from knowledge composed of qualitative elements, aggregated data, uncertainty, etc (cf [10]). The methodology for dealing with imprecision is developed

within the theory of fuzzy sets (cf [1] for an introduction); a fuzzy set is one for which there is no sharp transition from non-membership to membership of its elements - there is even a membership function (defined for the [0,1]-interval) indicating for each element its degree of membership in the set. The concepts and methods developed in the theory of fuzzy sets have also been introduced in the mathematical programming technique, by - among others - Carlsson [4], Orlovsky [9], Widey - Zimmermann [15], Yager [16], [17]. Takeda and Nishida [13] developed a conceptual framework for fuzzy domination structures, which generalizes the results obtained by Yu [13]. We will apply these results to formulate a method for fuzzy, multiobjective programming in which we propose to handle interdependent aspirations through compromises which - in some optimal fashion - are composed of elements from several aspirations. Hence the method is called the FMOP-C method: for fuzzy multiobjective programming with composite compromises.

2. AN ILL-STRUCTURED PRODUCTION PLANNING PROBLEM.

Consider given the following production planning problem: to develop tactical (= one-year) production plans for three divisions of a divisionalised company, later to be specified into two-week operational plans, and a coordinative plan for the CEO, which would enable him to evaluate the divisional plans in an objective and systematic fashion. He wants, furthermore, to experiment with a few options for an overall policy, and requires that,

(2.1) i. the k_1, k_2 and k_3 goals (representing several interdependent aspirations) defined for divisions 1-3 can be attained within the scope of the tactical plans (k_1, k_2, k_3 all > 2),

ii. all the $p = k_1 + k_2 + k_3$ goals can be evaluated against his possible options for an overall policy,

iii. an operational scheme for assessing the relevance and relative importance of the goals is developed,

iv. conflicting aspirations are resolved or absorbed, and "positive" interdependences among the aspirations are explored and utilized,

v. degrees of goal attainment, or satisfaction of aspirations, can be evaluated against different options for an overall policy.

A previous study [3] indicated that a composition/decomposition scheme for linear and goal programming models in a systems framework was sufficient for meeting at least two of these requirements) ((2.1) i and ii). It was even possible to establish the conditions for the existence of a composite optimal solution in the scheme (cf MP_0 - MP_3 in fig. 1), in the sense that an optimal solution in MP_0 (which is a multi-objective, linear- and goal programming model) could be decomposed

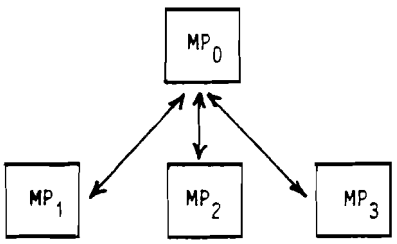


Fig. 1 Linking of MP-models.

to locally optimal solutions in MP_1 - MP_3. But it turned out that the requirements (2.1) iii - v were not dealt with in any efficient way, and some other approach is needed; this is the case for multiple, interdependent aspirations we introduced in the previous section.

This far we have introduced the production planning problem in a tidy, textbook manner. The reality facing the CEO was somewhat different; it may be described in rather a similar manner as the reality described by Peters (cf [10] p 165):

1. "Senior managers will usually receive for review what amounts to a single option ... rather than a set of fully developed choices."
2. "It is unusual for senior management to get a look at proposals when the options are still wide open."
3. "Senior managers will be shielded from most bad news" - i.e. most of the information they get is filtrated and aggregated.
4. "Most really important decisions emerge only after top managers have vacillated for months or years, apparently or actually; and the solution they choose at the end may well be indistinguishable from that proposed at the beginning of the search." - i.e. the manager will probably not on the spot turn an optimal solution from a model into a real life policy.

This description means that the element of imprecision is introduced into the

production planning problem. Imprecision can be introduced from a wide variety of sources: (i) filtering and/or aggregation of data; (ii) bias, discrepancies and stochastic influences on the data; (iii) the use of "knowledge" based on intuition, guesses and experience; (iv) uncertainty; (v) incompatible aspirations and (vi) conflicting interests - to mention a few.

Then what is needed is a set of techniques which could help the CEO, or his staff of trusted assistants, in a learning process, i.e. the process of forming his own aspirations in interaction with other aspirations, defining a common course for growth and development, and evaluating and assessing the constraints imposed on him (or them) and the process. Such a set of techniques should be able to handle imprecision, should be more supporting than normative, and should allow the user to tackle complex and difficult situations quickly and efficiently, i.e. with computer support.

In the next section we will find out what the proposed FMOP-C method has to offer on these requirements.

3. THE FMOP-C MODEL

We will first outline the principles on which a multiobjective programming model is built, find out the principles of non dominated extreme points and then turn to the fuzzy multiobjective programming model.

Consider a set of alterntive activities $X \sqsubset E^n$ which a decision maker (DM) has found to be relevant to consider for some planning horizon t_p, and let us assume that there exist p criteria $C(X) = \{c_1(X), c_2(X), \ldots, c_p(X)\}$ which are acceptable representations of DM's aspirations - described by a set $A \sqsubset E^n$ - in a specific planning situation $\langle X, A, t_p \rangle$. Let us, furthermore, assume that

(3.1) $X = X_1 \sqcup X_2$, where,

(3.2) $X_2 = Z \sqcup Y \sqcup W$,

and $X_1 \sqsubset E^n$, $X_2 \sqsubset E^{n_1}$, $Z \sqsubset E^{n_2}$, $Y \sqsubset E^{n_3}$, $W \sqsubset E^{n_4}$, where $n \geq n_1 = n_2 + n_3 + n_4$ and $n_1 \geq n_2 \geq n_3 \geq n_4$. It is necessary that these sets, pairwise, have at least

one dimension in common, as we are dealing with multidimensional descriptions of the activities; analogously we have p criteria dimensions, which we - for the time being - will assume to be theoretically independent and appropriate for linking DM's aspirations to the set X of activities.

A criterion $c_j(X)$, $j \in [1,p]$, may be linked to more than one aspiration, but, for simplicity, we will assume that $c_j(X)$ is related to only one, identifiable aspiration: $A_k^{(j)} \sqsubset A$, $k \in [1,q]$, which is described in up to n dimensions. Then we propose that the criterion $c_j(X)$ should be a vector-representation, which constitutes a relevant description of the aspiration, and that all the criteria, intially, are theoretically independent.

The traditional way of solving this type of multicriteria problems (cf [20]) is to formulate them as a vector maximum problem: let $z(x) = (c_1(x), c_2(x),\ldots, c_p(x))$ be a set of p objective functions, which now are representations of the p criteria, and $z(x)$ thus a vector-valued function; let $X \sqsubset E^n$ be the set of all feasible solutions; then the vector-maximum problem is,

(3.3) v-max $z(x)$ subject to $x \in X$

which is the problem of finding all the points $\bar{x} \in X$ which produce nondominated criteria values, i.e. there is no other $x \in X$ such that $z(x) \geq z(\bar{x})$ and $x \neq x$. We will refer to the points x as nondominated; let the set of nondominated points (or solutions) be N. Then it can be shown (cf [20]) that,

(3.4) $(L) \sqsubset N \sqsubset L$

where (L) and L may be called "approximating sets". These sets are defined as follows:

(3.5) $L = \{x \mid x \in X, x \text{ solves } P(\gamma) \text{ for some } \gamma \in \Gamma\}$

$(L) = \{x \mid x \in X, x \text{ solves } P(\gamma) \text{ for some } \gamma \in \text{Int } \Gamma\}$

where Γ is a set of vectors γ defined as,

(3.6) $\Gamma = \{\gamma \mid \gamma \in E^p,\ \gamma_i \geq 0,\ \sum \gamma_i = 1,\ i = [1, p]\}$

and the problem $P(\mathcal{V})$ is to find a point $\hat{x} \in X$ such that $\mathcal{V}\, z(\hat{x}) \geq \mathcal{V}\, z(x)$ for all $x \in X$. It has been proved several times (cf [20]) that,

(3.7) .i $(L) \sqsubseteq N$

.ii if $z[X]$ is convex, then $N \sqsubseteq L$

.iii for a closed and convex X; if $z(x)$ is concave for all $x \in X$ and one component of $z(x)$ strictly concave for $x \in X$, then $N \sqsubseteq L$.

We will use these results, but apply them to a formulation which involves a fuzzy domination structure (cf also [6]).

3.1. Fuzzy domination structures.

Let us now recall the criteria set $C(X)$; let $D(C) \sqsubset E^P$ be a family of fuzzy sets, where each fuzzy set $D(c_j)$, $j \in [1, p]$, is generated with a membership function $\mu_{c_j}(d)$; where $d = d(c',c)$ is the grade for $c' \in C(X)$ to be dominated by $c \in C(X)$, $c' \neq c$. We will refer to the family $D(C)$ of fuzzy sets as a fuzzy domination structure (cf [10]); Yu [18] has a similar formulation, but his "domination set" refers to some given preference structure.

The fuzzy set M, which is defined by the following membership function,

(3.8) $\mu_M(c') = 1 - \sup \mu_D(d)$, for $d = d(c',c)$

is the fuzzy set of all nondominated points in $C(X)$ (cf [10]). M is derived from a fuzzy domination structure, and its elements are characterized through membership values in the $[0,1]$-interval, which indicate how "dominant" various grades of dominance are.

For solving a multicriteria problem we need identifiable extreme points. For that purpose it is convenient to assume that $D(C)$ can be described as a fuzzy convex cone. We have the following definition: let $\Lambda \sqsubset E^P$ be a fuzzy set; it is a convex cone if its α-level sets Λ_α are convex cones for any $\alpha \in [0,1]$; an α-level set is defined by $\Lambda_\alpha = \{ d \in E^P \mid \mu_\Lambda(d) \geq \alpha \}$; $\Lambda_\alpha \sqsubset E^P$ is a cone iff $\mu_{\Lambda_\alpha}(d) = \mu_{\Lambda_\alpha}(\mathcal{V} \cdot d)$, $\mathcal{V} > 0$, $\forall\, d \neq 0$ and $d \in E^P$, and $\mu_{\Lambda_\alpha}(d) = 1$

for $d \neq 0$ (i.e. c' and c are both nondominated). Let us assume that $\Lambda = \{\Lambda_\alpha\}$ and that each Λ_α is a convex cone. Then $[D(C) \mid \Lambda]$ is called a fuzzy domination structure Λ, and the corresponding set of nondominated points is $M[C(X) \mid \Lambda]$, which now replaces M (cf (3.8)).

Then it is known (cf [18]) that a domination structure based on an ordinary convex cone, which can be defined by a membership function of the following type,

$$(3.9) \qquad \mu_{\Lambda^{\leq}}(d) = \begin{cases} 1 \text{ if } d \leq 0, \text{ for all } d \in E^p \\ 0 \end{cases}$$

produces an ordinary set $M[C(X) \mid \Lambda^{\leq}]$; this set contains the Pareto-optimal or efficient points, i.e. the individual, optimal values of each of the p objective functions $z(x) = (c_1(x), c_2(x), \ldots, c_p(x))$.

We could also apply another domination structure defined by the following membership function,

$$(3.10) \qquad \mu_{\Lambda^o}(d) = \begin{cases} 1 \text{ if } d = 0 \text{ or } \lambda^o \cdot d < 0, \text{ for all } d \in E^p \\ 0 \end{cases}$$

where $\lambda^o = \{\lambda_1^o, \lambda_2^o, \ldots, \lambda_p^o\}$ are trade-offs between the p criteria; the corresponding ordinary set of nondominated points is denoted by $M[C(X) \mid \Lambda^o]$, cf [13], [18], [20].

A fuzzy domination structure establishes some grade of dominance between any two points (c', c) in the criteria space, which is derived from DM's assertion of to what extent the two points satisfy his aspirations. Then the two domination structures $\Lambda^{\leq}$ and Λ^o represent extreme cases in relation to Λ; for $\Lambda^{\leq}$ there are no trade offs established between the p criteria, but for Λ^o the trade offs are assumed to be completely known, and a problem in multiple criteria is solved simply by maximizing some linear combination $\lambda \cdot z(x) = \lambda_1 \cdot c_1(x) + \lambda_2 \cdot c_2(x) + \ldots + \lambda_p \cdot c_p(x)$ over C(X), where the trade offs are determined relative to, for instance some reference point (cf [14]), or some ideal point (cf [19]), etc. Then Λ, which is "between" $\Lambda^{\leq}$ and Λ^o, should allow for fuzzy trade-offs through the membership functions.

From these observations follows that the fuzzy domination structure Λ should be found from,

$$(3.11) \quad \Lambda^{\leq} \sqsubset \Lambda \sqsubset \Lambda \sqsubset H \cup \{0\}$$

where H is an open half space in E^p (cf [13]; a similar result is given by Zeleny [20]). This relationship is intuitively obvious, and from it follows that,

$$(3.12) \quad M[C(X) \mid \Lambda^{\leq}] \sqsubset M[C(X) \mid \Lambda] \sqsubset M[C(X) \mid \Lambda^{0}]$$

which gives an initial delimitation of the fuzzy set of nondominated points.

Then the next step is aimed at a more precise characterization of $M[C(X) \mid \Lambda]$ which is achieved by determining gradually more precise inner and outer approximations of the set (cf [13], [20]). The outer approximation should obviously be derived from $M[C(X) \mid \Lambda^{\leq}]$: it is shown by Takeda and Nishida [13] that if the criteria imply maximization and C(X) is $\Lambda^{\leq}$-convex, then we have for any fuzzy domination structure Λ ,

$$(3.13) \quad M[C(X) \mid \Lambda] \sqsubset O[C(X) \mid \Lambda^{*}]$$

where Λ^* is the fuzzy polar cone of Λ , and the fuzzy set $O[C(X) \mid \Lambda^*]$ comprises the solutions which maximize the weighted sum of criteria, when each weighting vector constitutes the fuzzy polar cone Λ^*. The relationship in (3.13) can be proved, and the fuzzy set $O[C(X) \mid \Lambda^*]$ can be shown to be a good approximation of $M[C(X) \mid \Lambda]$; for details cf [13].

In much the same way it is also possible to find an inner approximation of the fuzzy set of nondominated points, which should be derived from $M[C(X) \mid \Lambda^{0}]$. This inner approximation is given by $O[C(X) \mid \Lambda^{+}]$, where Λ^{+} for a given Λ is defined by,

$$(3.14) \quad \Lambda^{+} = \bigsqcup \alpha \, \mathrm{Int} \, \Lambda^{*}_{1-\alpha}, \text{ for } 0 < \alpha \leq 1,$$

and the fuzzy set $O[C(X) \mid \Lambda^{+}]$ by,

$$(3.15)\qquad \mu_O(c) = \begin{cases} \sup \mu_{\Lambda^+}(\lambda), & \text{if } S^*(c) \sqcap S \neq \emptyset \\ 0 & \text{, if } S^*(c) \sqcap S = \emptyset \end{cases}$$

for $\lambda \in S^*(c) \sqcap S$ and $c \in C(X)$. Here S is the simplex in E^p, i.e.

$$(3.16)\qquad S = \{ \lambda \in E^p \mid \Sigma \lambda_j = 1,\ \lambda_j \geq 0;\ j \in [1, p] \}$$

Now, for an ordinary convex cone $\mathrm{Int}\,\Lambda^*_\alpha$ (Λ^* is the polar cone of Λ, but both Λ^* and $\mathrm{Int}\,\Lambda^*$ can be shown to be ordinary convex cones),

$$(3.17)\qquad O[C(X) \mid \mathrm{Int}\,\Lambda^*_\alpha] = \{ C^O(\lambda) \mid \lambda \in \mathrm{Int}\,\Lambda^*_\alpha \}$$

where $C^O(\lambda) = \{ c^O \in C(X) \mid \lambda \cdot c^O = \max \lambda \cdot c,\ c \in C(X) \}$, so that $C^O(\lambda^O) = M[C(X) \mid \Lambda^O]$, where Λ^O is the domination structure based on the known trade offs between criteria. Then, as for a domination structure Λ,

$$(3.18)\qquad O[C(X) \mid \Lambda^+] = \bigsqcup_\alpha \alpha\, O[C(X) \mid \mathrm{Int}\,\Lambda^*_{1-\alpha}], \text{ for } 0 < \alpha \leq 1, \text{ and}$$

$$(3.19)\qquad \{ C^O(\lambda) \mid \lambda \in \mathrm{Int}\,\Lambda^*_\alpha \} \sqsubset M[C(X) \mid \Lambda]$$

we have (for details and proofs, cf [13]),

$$(3.20)\qquad O[C(X) \mid \Lambda^+] \sqsubset M[C(X) \mid \Lambda]$$

If we combine (3.13) and (3.20) we get the following inner and outer fuzzy approximates of the fuzzy set of nondominated points,

$$(3.21)\qquad O[C(X) \mid \Lambda^+] \sqsubset M[C(X) \mid \Lambda] \sqsubset O[C(X) \mid \Lambda^*]$$

The final step is to decide the form of Λ; this is done by transforming the fuzzy sets of (3.21) from the C(X)-space to the activity space X. The results are the fuzzy sets $O[X \mid \Lambda^+]$, $M[X \mid \Lambda]$ and $O[X \mid \Lambda^*]$. Takeda and Nishida [13] give a theorem to the effect that the grade of membership in $O[C(X) \mid \Lambda^*]$ of a given $x^O \in X$, for a fuzzy domination structure Λ, is equal to the objective value of an optimum solution to the problem,

(3.22) $\max_{\lambda} \mu_{\Lambda^*}(\lambda)$, subject to normal Kuhn-Tucker constraints

This is the way in which we will find out which $X^o \sqsubset X$ is going to produce solutions which are nondominated in a fuzzy sense. and represent feasible compromises in relation to all the criteria, i.e. the aspirations of DM.

Let us take a closer look at these results: an intuitive and simplified graphical representation is given in fig. 2 (a twodimensional projection of a part of an n-dimensional space):

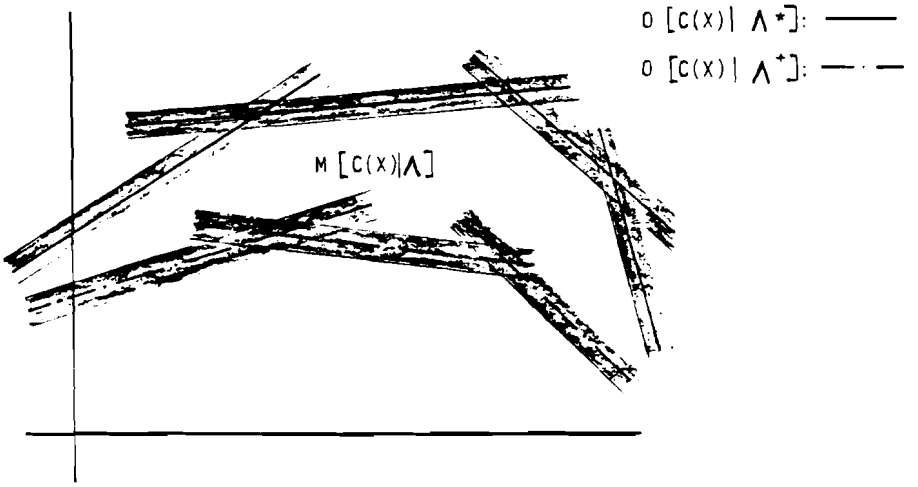


Fig. 2 M[C(X)|Λ] and fuzzy approximates.

There is, however, no easy way to utilize these results for algorithmic purposes, in the sense that we do not have any readily available algorithm for finding any of the three sets. We will, in other words, have to develop the principle of a fuzzy domination structure a bit more, in order to find some construction which can be linked to other existing algorithmic devices - or to modifications of such. Then, we will use the described "squeezing" of M[C(X)|Λ] between the fuzzy approximate sets as a way to get anchoral points for finding and evaluating compromises among the criteria, which then can be characterized both in terms of distance from the anchoral points and in terms of grade of dominance.

3.2 Fuzzy compromises

The fuzzy set $0[C(X) \mid \Lambda^*]$ describes conflicting aspirations: the extreme points indicated in fig. 2 are maximum points for a few of the p criteria of C(X). Zeleny [22] refers to these as competive criteria, i.e. the underlying aspirations have to be fulfilled per se, and have to "bargain" with each other over the set of alternative activities. Zeleny recommends that the compromises - which are necessary, as no extreme point could fulfill all the conflicting aspirations - are formulated on the basis of distance from an ideal but infeasible point, i.e. the one which results in a simultaneous maximum for all the criteria.

Let $\overset{*}{c} \in C(X) \sqsubset E^p$ be the infeasible ideal solution, and $\overset{*}{x}$ the corresponding ideal point in X space; let further ϑ_i^k be the distance of a point x_i^k from $\overset{*}{x}_i$, i.e. the distance is measured in relation to criterion i. The distance ϑ_i^k is preferably given a fuzzy interpretation (cf [22]) as $\overset{*}{x}_i$ is a fuzzy point,

$$(3.23) \qquad \vartheta(x_i^k, \overset{*}{x}_i) = \vartheta_i^k$$

where $\vartheta_i^k = 1$ if $x_i^k = \overset{*}{x}_i$ and otherwise $0 \leq \vartheta_i^k < 1$. Then it is possible to define a family of distance membership functions (cf [22]),

$$(3.24) \qquad \mu^q(\beta, k) = \left[\sum_{i=1}^{p} \beta_i^q (1 - \vartheta_i^k)^q\right]^{1/q}, \ \forall k \quad \text{for } x^k \in X,$$

where $\beta = (\beta_1, \ldots, \beta_p)$ is a vector which allows a DM to give different importance to the various criteria ($\sum_i \beta_i = 1$), and q is the distance parameter, $1 \leq q \leq \infty$. Zeleny [22] discusses the effects of alternative values of q on the form of the membership function. For some selected $\hat{q}$ every point x^k is then associated with a membership value $\mu^{\hat{q}}(\beta, k)$ representing its "distance" from an ideal point $\overset{*}{x}$ in relation to all the criteria. Covering all $x^k \in X$ we get a set of membership values $\Delta_\beta^{\hat{q}}$ which then characterizes all the possible compromises.

The fuzzy set $0[C(X) \mid \Lambda^+]$ contains the points which are representations of mutually and unilaterally supportive aspirations: the extreme points are established through known trade off-functions defined for the criteria. Zeleny (cf [22]) refers to these as compensatory criteria, i.e. there is an established structure

of trade offs covering the compensations of low fulfillment of certain aspirations with a high fulfillment of other aspirations.

Let $c^o \in C(X) \sqsubset E^p$ be the ideal solution with supportive or compensatory solutions, and x^o the corresponding (probably existing) ideal point: let further ε_i^k be the distance of a point x_i^k from x_i^o, a distance which is defined by the appropriate trade off-functions for criterion i in relation to the other p-1 criteria. Analogously with δ_i^k there is a family of distance membership functions (cf (3.24))

$$(3.25) \qquad \mu^q(\theta, k) = \left[\sum_{i=1}^{p} \theta_i^q (1 - \varepsilon_i^k)^q\right]^{1/q}, \forall k \text{ for } x^k \in X$$

where the vector $\theta = (\theta_1, \ldots, \theta_p)$ indicates the importance of the various criteria ($\sum_i \theta_i = 1$), and q is the distance parameter. As these trade offs probably do not exist for all criteria, we will get associated membership values $\mu^q(\theta, k)$ only for some x^k, but which then will indicate the grade of support or compensation relative to each criterion (with trade offs) associated with each compromise point. Covering all $x^k \in X$ we get a set of membership values $\theta_\theta^{\hat{q}}$ which then caracterizes all the compromises which are possible on the basis of the trade offs.

Then, finally, the complete model with the fuzzy domination structure Λ, and the two compromise structures $\Delta_\beta^{\hat{q}}$ and $\theta_\theta^{\hat{q}}$ - which we will call the fuzzy multiobjective programming model with composite compromises - or the FMOP- C model - can be written in the following form:

$$(3.22) \qquad \text{v-max } [C(X) \mid \Lambda, \Delta_\beta^{\hat{q}}, \theta_\theta^{\hat{q}}]$$
$$\text{subject to } x \in X$$

This is, of course, still far from being an operational model, as we should develop an algorithm which could accomodate all the various problem-solving elements introduced above. In the next section we will indicate how at least part of the model can be handled with existing techniques.

4. A SIMPLE EXAMPLE

In order to be able to handle the FMOP-C model with existing techniques we will have to give up the idea of a vector-maximum and resort to a scalar formulation. Let X be constrained by,

$$(4.1) \qquad \begin{aligned} Ax &\cong b \\ x &\geq 0 \\ x &\in X \end{aligned}$$

and the problem to be solved,

$$(4.2) \qquad \max [C(X) \mid \Lambda] = \max [c_1(x), c_2(x), \ldots, c_p(x) \mid \Lambda]$$

where ($\cong$) denotes a fuzzy constraint and $(c_1(x), c_2(x), \ldots, c_p(x))$ are the appropriate objective functions for the p criteria. A fuzzy constraint indicates that there are degrees of allowable violations of the constraints; the corresponding membership functions are,

$$(4.3) \qquad \mu_{b_i} = \begin{cases} 1 & \text{if } (Ax)_i \leq b_i \\ 1 - \dfrac{(Ax)_i - b_i}{\vartheta_i} & \text{if } b_i < (Ax)_i \leq b_i + \vartheta_i \\ 0 & \text{if } (Ax)_i > b_i + \vartheta_i \end{cases}$$

where $(Ax)_i$ refers to the ith row of the matrix of constraints. Let μ_b denote the combination $\mu_b = \mu_{b_1} \circ \mu_{b_2} \circ \ldots \circ \mu_{b_k}$, where k = the number of rows in the matrix of constraints; the combination is usually obtained with the min-operator (cf [1], [16]), and represents the fuzzy set of feasible alternative activities.

In this formulation it is fairly cumbersome to implement the fuzzy domination structure Λ; it is perhaps possible to "open up" the model at each iteration, work out the structure of fuzzy dominances and "redirect" the next iteration through modifications of the objective functions (if the parameters are fuzzy numbers (cf [1])). In this way we could get a set of points, which when reevaluated through some scheme, would give us the distances ϑ_i^k and ε_i^k ($\forall k$,

($\forall k$, $\forall i \in [1, p]$) and the sets of compromise points. This is, nevertheless, not a systematic process and definitely not one which would result in any definable, optimal solution.

Consider the following simplified formulation of a production planning problem (based on a plywood production planning problem; adapted from [4]), which is an application of the fuzzy programming formulation presented in [15], [16] and [17]. We have a problem in 3 planning periods, 7 products, 24 constraints and 3 objectives:

(4.4) max λ

s.t. (1) (2) (3)

$$.022x_1 + .030x_2 + .028x_4 + .015x_5 + .045x_7 - e_1 \leq 80,\ 90,\ 110$$
$$.039x_1 + .041x_3 + .043x_4 + .045x_6 + .052x_7 - e_2 \leq 160,\ 150,\ 140$$
$$.041x_1 + .044x_2 + .036x_3 + .042x_5 + .062x_7 - e_3 \leq 160,\ 160,\ 180$$
$$.062x_1 + .059x_3 + .070x_4 + .082x_6 + .095x_7 - e_4 \leq 280,\ 300,\ 300$$

which are capacity constraints for 4 production lines for the 3 planning periods; x_i is the number of units of product i produced per planning period; e_k is an overutilization of capacity on line k; the parameters give the capacity required on line k to produce one unit of product i; the rhs-vectors give available capacities of line k per planning period. The following production constraints are dictated by demand/planning period (all 3 periods) and a trimming problem,

(4.5) s.t.

$$1320 \leq x_1 \leq 1386 \quad ; \quad x_i - x_{i1} - x_{i2} = 0, \ \forall \ i \in [1, 7]$$
$$550 \leq x_2 \leq 577$$
$$1285 \leq x_3 \leq 1349$$
$$730 \leq x_4 \leq 766$$
$$433 \leq x_5 \leq 454$$
$$225 \leq x_6 \leq 236$$
$$113 \leq x_7 \leq 120$$

$$.056x_{11} + .111x_{21} + .125x_{31} + .250x_{41} + .083x_{51} + .167x_{61} + .250x_{71} \leq 330$$
$$.125x_{12} + .250x_{22} + .167x_{32} + .500x_{42} + .125x_{52} + .250x_{62} + 1.000x_{72} \leq 385$$

where x_{im} denotes the number of units of product i to be cut from standard sheet m (= 1, 2); the rhs-vector gives the number of standard sheets available for the 3 periods.

We have 3 objective functions, which are given as fuzzy constraints,

(4.6) s.t.

$$74.8x_1 + 58.8x_2 + 68.7x_3 + 24.6x_4 + 77.5x_5 + 41.2x_6 - 18.3x_7 - 15000\lambda \geq 270000$$

which is a profit function, and shows that profit should exceed 270000 and be raised by as much as 15000. The cutting should be done in such a way that waste is less than 3000, and decreased by up to 200,

(4.7) s.t.

$$\begin{aligned} &.2740x_{11} + .6229x_{12} + .2740x_{21} + .6929x_{22} + 1.4971x_{31} + .2531x_{32} \\ &+ 1.4971x_{41} + 1.0545x_{42} + .9338x_{51} + .1328x_{52} + .9338x_{61} + .1328x_{62} \\ &+ .2740x_{71} + 1.9429x_{72} + 200\lambda \leq 3000 \end{aligned}$$

and overtime should not exceed 5, 10 and 20, respectively, and should even be decreased by 5 if possible:

(4.8) s.t. (1) (2) (3)

$$e_1 + e_2 + e_3 + e_4 + 5\lambda \leq 5, \quad 10, \quad 20$$

This problem was solved with IFPOS (Interactive Financial Planning Optimization System), a high-level programming language, on a DEC System 20. IFPOS includes facilities for both linear and nonlinear programming - the coming Version 2 will also offer mixed integer programming - in an interactive, highly user-friendly form, and is very well suited for the trial and error mode we have used for solving the multiobjective problem (IFPOS was developed by Execucom Inc. in Austin, Texas; the actual model is given in Appendix I).

We obtained the following 5 alternative programs for the 3 planning periods:

	Program 1			Program 2			Program 3			Program 4			Program 5		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
x_1	1320	1320	1320	1386	1386	1386	1386	1386	1386	1386	1386	1386	1386	1386	1386
x_2	577	577	577	577	577	577	577	577	577	577	577	577	577	577	577
x_3	1349	1349	1349	1349	1317	1317	1349	1341	1341	1349	1349	1349	1349	1349	1349
x_4	730	730	730	730	730	730	762.4	730	730	766	747	747	730.6	730.6	730.6
x_5	439.3	439.3	439.3	449.7	454	454	454	454	454	454	454	454	454	454	454
x_6	236	236	236	225	225	225	216	225	225	236	236	236	236	236	236
x_7	113	113	113	113	113	113	113	113	113	113	113	113	113	113	113
profit	285000	285000	285000	290000	288392	288392	291871	290181.5	290181.5	291960	291369	291369	291285.5	291285.5	291285.5
waste	2800	2800	2800	2000	2756.44	2756.44	2460.72	2488.22	2488.22	2228.20	2355.00	2355.00	2042.86	2042.86	2042.86
λ	1.0000	1.0000	1.0000	1.0000	0.8928	0.8928	0.7914	0.6289	0.6289	0.4640	0.4246	0.4246	0.0857	0.0857	0.0857
s_1	0	0	0	0	0	0	1.04	0	0	1.14	0	0	0.18	0	0
s_2	0	8.79	14.79	0	5.54	15.54	0	6.61	16.61	0	7.08	17.08	0	7.71	17.71
s_3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s_4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

A λ close to 1.0000 indicates that there is little conflict between the goals; a λ close to 0.0000 that conflicts prevail between the goals. DM's membership functions should be defined for (i) the profit, (ii) the amount of waste, (iii) the use of overtime and, perhaps, (iv) the number of units produced of products 1 - 7. These membership functions would then form the basis for compromises, which will be necessary for the choice of a program; that is not a trivial task as can be seen from the results above (the reader is invited to test a few membership functions).

5. SUMMARY AND CONCLUSION

MCDM is becoming a field of research in its own right, due to the fact that it addresses one of the fundamental problems in management research: the systematic application of multiple criteria in managerial planning and decision making. Here we have discussed a special case of multiple criteria: the situation in which DM has several interdependent aspirations, and it is not clear how some alternative activities will help him in the fullfillment of these aspirations.

We developed a fuzzy multiobjective programming model with composite compromises - the FMOP-C model - and found that this approach may be useful for tackling problems with interdependent aspirations. We also found that we still do not have any simple algorithm for solving the FMOP-C model, but that simplified versions could be handled with a standard, fuzzy LP-technique; this last statement with the reservation that a fuzzy domination structure is fairly complex to implement. Nevertheless, the FMOP-C model seems to hold some promises for further development.

6. REFERENCES

1. Bellman, R.E. - L.A. Zadeh — Decision Making in a Fuzzy Environment, Management Sci., Vol. 17 (1970), 141-164

2. Carlsson, C. — Adaptive Multigoal Control. On the Principles for Problem Solving in a Complex Environment, Ph. D dissertation (in Swedish), Abo 1977

3. Carlsson, C. — Linking MP Models in a Systems Framework, IEEE Transactions, Vol. SMC-9 (1979), 840-849

4. Carlsson, C. - A. Törn - M. Zeleny (eds.) — Multiple Criteria Decision Making: Selected Case Studies, (McGraw Hill, New York 1981)

5. Carlsson, C. — Solving Comples and Ill-Structured Problem: An MCDM-Approach, in: Nijkamp, P. and Spronk, J. (eds.): Multiple Criteria Analysis. Operational Methods (Grower Press, Aldershof 1981)

6. Carlsson, C. — Solving Ill-Structured Problems through Well-Structured Fuzzy Programming, In: J.P. Brans (ed.): Operational Research'81 (North-Holland, Amsterdam 1981)

7. Cohon, J.L. — Multiobjective Programming and Planning, (Academic Press, New York 1978)

8. Leberling, H. — On Finding Compromise Solutions in Multicriteria Problems Using the Fuzzy Min-Operator, Fuzzy Sets and Systems, Vol 6. (1981), 105-118

9. Orlowsky, S.A. — On Formalization of a General Fuzzy Mathematical Problem, Fuzzy Sets and Systems, Vol. 3 (1980), 311-321

10. Peters, T.J. — Leadership: sad facts and silver limings, Harvard Business Review, Nov-Dec (1979), 164-172

11. Spronk, J. — Interactive Multiple Goal Programming, (Martinus Nijhoff Publishing, Boston 1981)

12. Starr, M.K. - M. Zeleny (eds.) — MCDM-State and Future of the Arts, in: Starr, M.K. and Zeleny, M. (eds.): Multiple Criteria Decision Making, TIMS Studies in the Management Sciences, Vol 6 (North-Holland Publishing Company, Amsterdam 1977)

13. Takeda, E. - T. Nishida — Multiple Criteria Decision Making with Fuzzy Domination Structures, Fuzzy Sets and Systems, Vol 3. (1980), 123-136

14. Wierzbicki, A.P.
A Methodological Guide to Multiobjective Optimization,
WP-79-122 (IIASA, December 1979)

15. Widey, G. - H.-J. Zimmermann
Media Selection and Fuzzy Linear Programming,
J. Operational Research Society, Vol 29 (1979), 1071-1084

16. Yager, R.R.
Mathematical Programming with Fuzzy Constraints and a Preference on the Objective,
Kybernetes, Vol. 8 (1979), 285-291

17. Yager, R.R.
Decisions with Ordinal Preferences and Importances,
Kybernetes, Vol. 9 (1980), 109-114

18. Yu, P.L.
Introduction to Domination Structures in Multicriteria Decision Problem,
in: Cochrane, J.L. and Zeleny, M. (eds.), Multiple Criteria Decision Making (Univ. South Carolina Press, Columbia, SC, 1973), 249-261

19. Zeleny, M.
Compromise Programming,
in: Cochrane, J.L. and Zeleny, M. (eds.): Multiple Criteria Decision Making (Univ. South Carolina Press, Columbia, SC. 1973), 262-301

20. Zeleny, M.
Linear Multiobjective Programming,
Lecture Notes in Economics and Mathematical Systems, Nr 95 (Springer Verlag, New York 1974)

21. Zeleny, M.
A Case Study in Multiobjective Design: de Novo Programming,
in: Nijkamp, P. and Spronk, J. (eds.): Multiple Criteria Analysis. Operational Methods
(Grower Press, Aldershof 1981)

22. Zeleny, M.
Multiple Criteria Decision Making,
(McGraw-Hill, New York 1981)

APPENDIX I

```
  DIRECTIVE MCOPT
READY FOR EDIT, LAST LINE IS 135
? LIST
5    COMPILEGET MCDM
10   OBJECTIVE
15    MAXIMIZE SLAMBDA (3)
20   *
25   DECISIONS
30   ALL DECISIONS POSITIVE
35    OVT1 THRU OVT4 (1-3)
40    PLW11 THRU PLW72 (1-3)
45    LAMBDA (1-3) BETWEEN 0 AND 1
50   *
55   CONSTRAINTS
60    PROD1 (1) .LE. 80 (2) .LE. 90 (3) .LE. 110
65    PROD2 (1) .LE. 160 (2) .LE. 150 (3) .LE. 140
70    PROD3 (1-2) .LE. 160 (3) .LE. 180
75    PROD4 (1) .LE. 280 (2-3) .LE. 300
80    PROFIT (1-3) .GE. 290000
85    OVERUSE (1) .LE. 5 (2) .LE. 10 (3) .LE. 20
90    TRIM1 (1-3) .LE. 330
95    TRIM2 (1-3) .LE. 385
100   TRIMLOSS (1-3) .LE. 3000
105   PLW1 (1-3) BETWEEN 1320 AND 1386
110   PLW2 (1-3) BETWEEN 550 AND 577
115   PLW3 (1-3) BETWEEN 1285 AND 1349
120   PLW4 (1-3) BETWEEN 730 AND 766
125   PLW5 (1-3) BETWEEN 433 AND 454
130   PLW6 (1-3) BETWEEN 225 AND 236
135   PLW7 (1-3) BETWEEN 113 AND 120
? QUIT

END OF EXECUTION
CPU TIME: 49.53 ELAPSED TIME: 34:55.36
EXIT
```

```
MODEL MCDM  VERSION OF  02/04/82  16:32
5  *
10 *   PLYWOOD PRODUCTION PLANNING
15 *
20    COLUMNS 1,2,3
25    PLW1=PLW11+PLW12
30    PLW2=PLW21+PLW22
35    PLW3=PLW31+PLW32
40    PLW4=PLW41+PLW42
45    PLW5=PLW51+PLW52
50    PLW6=PLW61+PLW62
55    PLW7=PLW71+PLW72
60    LAMBDA=0
65    SLAMBDA=LAMBDA+PREVIOUS
70    PROD1=PROD11+PROD12
75    PROD11=0.022*PLW1+0.030*PLW2+0.028*PLW4+0.015*PLW5
80    PROD12=0.045*PLW7-OVT1
85    PROD2=PROD21+PROD22
90    PROD21=0.039*PLW1+0.041*PLW3+0.043*PLW4+0.045*PLW6
95    PROD22=0.053*PLW7-OVT2
100   PROD3=PROD31+PROD32
105   PROD31=0.041*PLW1+0.044*PLW2+0.036*PLW3+0.042*PLW5
110   PROD32=0,062*PLW7-OVT3
115   PROD4=PROD41+PROD42
120   PROD41=0.062*PLW+0.059*PLW3+0.070*PLW4+0.082*PLW6
125   PROD42=0.095*PLW7-OVT4
130   PROFIT=PCOST1+PCOST2
135   PCOST1=C11*PLW1+C12*PLW2+C13*PLW3+C14*PLW4+C15*PLW5+C16*PLW6
140   PCOST2=C17*PLW7-15000*LAMBDA
145   OVERUSE=OVT1+OVT2+OVT3+OVT4+5*LAMBDA
205   TRIM1=TRIM11+TRIM12
210   TRIM11=0.056*PLW11+0.111*PLW21+0.125*PLW31+0.250*PLW41
215   TRIM12=0.083*PLW51+0.167*PLW61+0.250*PLW71-20*LAMBDA
220   TRIM2=TRIM21+TRIM22
225   TRIM21=0.125*PLW12+0.250*PLW22+0.167*PLW32+0.500*PLW42
230   TRIM22=0.125*PLW52+0.250*PLW62+1.000*PLW72-20*LAMBDA
235   TRIMLOSS=ALFA+BETA+GAMMA+DELTA
240   ALFA=0.274*PLW11+0.6229*PLW12+0.274*PLW21+0.6929*PLW22
245   BETA=1.4971*PLW31+0.2351*PLW32+1.4971*PLW41+1.0545*PLW42
250   GAMMA=0.9338*PLW51+0.1328*PLW52+0.9338*PLW61
255   DELTA=0.1328*PLW62+0.274*PLW71+1.9429*PLW72+200*LAMBDA
260   C11=74.8
265   C12=58.8
270   C13=68.7
275   C14=24.6
280   C15=77.5
285   C16=41.2
290   C17=-18.3
295   OVT1=2
300   OVT2=2
205   OVT3=2
210   OVT4=2
215   PLW11=10
220   PLW12=10
225   PLW21=10
230   PLW22=10
235   PLW31=10
```

```
340   PLW32=10
345   PLW41=10
350   PLW42=10
355   PLW51=10
360   PLW52=10
365   PLW61=10
370   PLW62=10
375   PLW71=10
380   PLW72=10
END OF MODEL
? QUIT
STOP

END OF EXECUTION
CPU TIME: 2.23 ELAPSED TIME: 2:30.88
EXIT
 LOGOUT
Killed Job 29, User FEI.CCCARL, Account , TTY 100,
  at 4-Feb-82 17:24:41, Used 0:00:55 in 0:38:34
```

IV

MULTIOBJECTIVE PROGRAMMING AND APPLICATIONS

AN INTERACTIVE MULTIPLE-OBJECTIVE LINEAR PROGRAMMING METHOD FOR A CLASS OF NONLINEAR OBJECTIVE FUNCTIONS

S. Zionts[1] and J. Wallenius[2]

[1] *State University of New York at Buffalo, New York, USA*
[2] *Purdue University, West Lafayette, Indiana, USA* and
University of Jyväskylä, Finland

1. Introduction

In 1976 we published an article in this journal (Zionts and Wallenius [18]) describing a method that finds an optimal solution to a multiple criteria linear programming problem which assumes an unknown underlying linear utility function. In that paper we devoted a paragraph to describe how the method can be generalized to handle a larger class of utility functions. That method has since evolved and been used in several decision making situations. Numerous requests for a formal description of it have been received. Since it has never been fully described in the literature, and has evolved from the original method outlined, it seems appropriate to describe it together with relevant theory.

The problem under consideration involves a set of n non-negative decision variables represented by the vector x constrained by m linear inequality constraints. We, therefore, represent the constraints algebraically as in linear programming

$$\begin{aligned} Ax &\leq b \\ x &\geq 0 \end{aligned} \tag{1}$$

where A is an $m \times n$ matrix, and b is an m-vector.

Our decision situation involves a single decision maker who has p linear objectives. We may write these objectives as $u = Cx$. Without loss of generality, we assume that the objectives are all to be maximized. Further, we assume that the objectives are nonsatiable (more of each objective is preferred to less). Concave objectives may be considered by using piecewise linearizations. The utility function is assumed to be an unknown underlying pseudo concave function of the objectives. Denote this function as $g(Cx)$. We assume that g is differentiable and that the first derivatives of g are continuous.

Our problem, therefore, is to

$$\begin{aligned} \text{Maximize} \quad & g(u) = g(Cx) \\ \text{subject to} \quad & Ax \leq b \\ & x \geq 0 \end{aligned} \tag{2}$$

with g unknown. Were g known, the problem would be a nonlinear programming problem with linear constraints. With g unknown, we use decision maker responses to questions regarding pairwise comparisons between alternatives and tradeoffs to construct local linear approximations to the utility function. In that way we generate a sequence of improved extreme point solutions to problem (2). The method terminates with a locally optimal extreme point solution which is in some circumstances globally optimal. In that case, the method declares the solution to be a global optimum. Otherwise further work is required to determine a globally optimal solution. Several search methods are possible. Deshpande [2] proposes one possible approach

that involves searching on facets of the polyhedron corresponding to the constraint set.

In this section we have presented the problem to be addressed, and have outlined the method to be used. In Section 2 we present an algorithm. In Section 3 we present the necessary theory and some remaining problems and possible ways of resolving them, and in Section 4 we discuss certain computational aspects of our method. Section 5 contains a discussion of the practical experience with the method. The appendix presents an example illustrating that the method need not find an optimal extreme point solution prior to finding a global optimum.

2. An Algorithm and Its Rationale

We now present the algorithm in a step by step manner, interspersing comments and explanations between the steps. The algorithm consists of sequential solutions to a linear programming problem, each involving maximizing a weighted sum of objectives. We denote the weights as a vector $\lambda > 0$.

Step 1. Choose a starting set of weights $\lambda > 0$. Find a solution maximizing $\lambda'Cx$. Denote this solution as the maximizing solution x^*.

In our experience with the method and its variations, we have not found it useful to ask the decision maker for his weights. We can envision situations where it would be useful to start with a set of weights supplied by a user or found in a previous study, but we have not had any experience with this idea. To start, we normally scale the objectives so that the average absolute value of nonzero coefficients is one. Then we use

equal weights to combine the objectives into a composite objective which we maximize subject to the constraints. The resulting current maximizing solution is efficient or nondominated with respect to the set of feasible solutions.

Step 2. Identify all adjacent efficient edges from x^* and the corresponding adjacent efficient extreme point solutions.

To find all adjacent efficient edges, we use the procedure described in Zionts and Wallenius [18,19]. Explained intuitively, we use a variation of linear programming to determine whether, given the solution x^*, a vector of weights $\lambda > 0$ can be found so that one nonbasic variable is attractive for entry into the basis, whereas none of the other nonbasic variables are attractive for entry. If such a vector can be found, the associated edge is efficient, and otherwise not. To find the adjacent extreme point, we need only determine the level at which a variable corresponding to an efficient edge enters the basis. Some edges may possibly be unbounded.

Step 3. Determine which of the efficient edges are also efficient with respect to the set of (active) previous responses of the decision maker. Call the set of such edges set A. Call its complement with respect to the set of efficient edges set B. Let I equal A.

For the case of a linear utility function and a consistent decision maker, we are concerned only with set A. For a pseudo concave utility function, we also need set B because the weights provide only a linear approximation. We will have to discard responses in a later step. I is an indicator which will either be A or B, depending upon which set is under consideration.

Step 4. With respect to set I, ask the decision maker to choose between x^* (objective function values Cx^*) and a distinctly different adjacent efficient extreme point solution. He may express a preference for one of the two points, or the inability to express a preference. If he prefers at least one alternative to x^*, save one such alternative, denote it as x^2, and go to step 8.

The purpose of this step is to ask for pairwise comparisons between distinct alternatives expressed as scenarios in terms of their objective function values Cx^* where possible. (Any efficient edges which are infinite should be left for step 5.) The use of scenarios was based on our practical experience with the earlier method. Managers seem to prefer choosing between alternatives to evaluating tradeoffs. In order to have them choose between alternatives, some threshhold must exist between two solutions. In our computer programs we use an arbitrary threshhold. For example, given two solutions there must be at least a ten percent difference in one objective function value. If the decision maker likes one or more alternatives, we terminate the questioning process, and revise the weights. Note that indifference is an instance of not being able to express a preference.

Step 5. With respect to set I, ask the decision maker whether beginning at point x^* he likes any efficient tradeoff not leading to a point asked about in step 4. He may respond that he likes a tradeoff, that he does not like the tradeoff, or that he is unable to decide. If he likes at least one such tradeoff, go to step 8. If he likes any unbounded tradeoffs, terminate with an unbounded solution.

The purpose of this step is to ask the decision maker to evaluate as tradeoffs efficient edges not considered in step 4. Because the objective function values of the adjacent points are too close or even identical (as, for example, in the case of degeneracy), questions must be put to the decision maker in the form of tradeoffs. Care must be taken in presenting tradeoffs to be sure that the concept is correctly understood. Here too, if the decision maker likes one or more tradeoffs, we terminate the questioning process.

Step 6. With respect to set I, ask the decision maker whether beginning at x^* he likes any tradeoffs leading to adjacent efficient points which he did not prefer to x^* in the most recent step 4. If he likes at least one such tradeoff, go to step 8.

This step involves asking as a tradeoff every paired comparison for which the decision maker preferred x^* in the most recent step 4. This helps establish our stopping criterion.

Step 7. If I is equal to A, let I equal B and go to step 4. Otherwise stop; the solution x^* is optimal.

Set A consists of efficient tradeoffs which may be attractive and consistent with active previous responses. We first check to see whether any of these are attractive to the decision maker (in steps 4 through 6). Only if none of them are do we consider set B, that is the set of efficient tradeoffs which are not consistent with active previous responses. We must consider these as part of our check for optimality.

For an extreme point solution to be optimal, no adjacent efficient extreme point solution should be preferred to it, and

no tradeoff corresponding to an efficient edge should be liked. We shall explore this condition more in the next section.

Step 8. Write inequalities on the weights based on the decision maker responses, and add them to the set.

If the decision maker prefers x^* to an adjacent solution x^0, we write a constraint $\lambda'Cx^* - \lambda'Cx^0 \geq 1$. If the decision maker likes a tradeoff vector w, we write the constraint $\lambda'w \geq 1$. (The vector λ is assumed to be bounded from above and below.) If x^0 is preferred to x^*, we write a similar inequality. We currently do not use the responses of I don't know, although it seems reasonable to try to set up equalities analogous to the above inequality. Early experience led us to abandon this strategy, but it may be worth further consideration.

Step 9. Find a set of positive weights λ consistent with all previous responses. If no set of consistent weights exist, drop the oldest set of active constraints and repeat step 9.

This step involves solving the constraints added in step 8 and earlier together with the constraints $\lambda_i \geq 1$ (see, for example, Ecker, Hegner and Kouada [3].) Earlier we had used sufficiently small numbers in all constraints rather than ones, and required that the sum of the λ's be one. The use of ones overcomes the problem of determining a sufficiently small epsilon, though the problem is transformed into finding a sufficiently large bound on the λ vector.

In the event that no consistent vector of weights can be

found, the oldest set of constraints, or one of the oldest constraints (we have not yet decided which), is dropped. The step is repeated until a consistent feasible set of weights is found.

Step 10. Using the new weighting vector λ, solve the linear programming problem. Maximize $\lambda'Cx$ subject to constraints (1). Denote the solution as x^1.

This step determines the optimal solution for the new linear approximation.

Step 11. If solution x^2 is not null, go to step 13. Otherwise have the decision maker choose between solutions x^* and x^1. If the decision maker chooses x^1, add a constraint based on the choice of x^1 over x^*, designate the solution x^1 as x^* and go to step 2.

This step asks if the new solution is preferred to the old. If so, it must have a higher utility and the procedure repeats using the new solution.

Step 12. Solution x^* is a local optimum, but there exist better solutions, some of which are not extreme points. A search procedure (not part of the method) should be used to find the optimum. The method terminates at the local optimum x^*.

Though x^* is a local optimum, we know that there are one ore more efficient edges emanating from x^* which are desirable to the decision maker, but the corresponding adjacent efficient extreme point is not preferred to x^*. Therefore, there exists a solution point along each such edge that is preferred to x^*. A search method may then be used to find an optimal solution.

Step 13. Ask the decision maker to choose between x^1 and x^2. If he chooses x^1 or x^2, add a constraint based on the preference, denote the preferred solution as x^*, make x^2 null, and go to step 2. If he is unable to choose, denote x^1 as x^*, make x^2 null and go to step 2.

This step makes use of the information that x^2 is preferred to x^*. Therefore, we have the decision maker compare x^2 with x'. If x^2 is preferred, then it becomes the new reference solution. Otherwise x' becomes the new reference solution.

3. Theory

The method is constructed in such a way that it generates a sequence of improved solutions. Each solution is preferred to its predecessor. This will occur until the procedure stops in steps 7 or 12. At that point, solution x^* is at least as preferred as every adjacent efficient solution. It may be that some of the adjacent efficient solutions are virtually indistinguishable from x^*. In any event, x^* must be regarded as a locally optimal extreme point. If, in addition, the decision maker does not like any efficient edge emanating from x^*, we may state and prove that x^* is a global optimum.

Theorem: An extreme point solution x^* is optimal if the decision maker does not like any efficient edge emanating from x^*.

Proof: Let the true utility function be $f(x_1,\ldots,x_n)$, which is psuedo concave by assumption. Let the efficient edges at x^* be given by $\{x_j\}$, $j \in E$ where E is the set of indices

of the variables corresponding to efficient edges. By assumption we have $\frac{\partial f}{\partial x_j} \leq 0 \quad (j \notin E)$. Consider an arbitrary efficient feasible direction y. By convexity, we may write $y = \sum_{j \in E} \alpha_j x_j$, $\alpha_j \geq 0$. From elementary calculus using the continuity of the partial derivatives $\frac{\partial f}{\partial x_j}$, we may write $\frac{\partial f}{\partial y} = \sum_{j \in E} \alpha_j \frac{\partial f}{\partial x_j}$, $\alpha_j \geq 0$. Since $\frac{\partial f}{\partial x_j} \leq 0$, $j \notin E$, we may, therefore, conclude $\frac{\partial f}{\partial y} \leq 0$ thereby proving the theorem.

The algorithm does not just check tradeoffs exclusively. It first asks the decision maker to compare a solution x^* with each of its adjacent efficient extreme points. If the decision maker prefers an adjacent solution to x^*, then the method will not terminate at x^*. If the decision maker prefers x^* to an adjacent solution, he is then asked about the corresponding tradeoff, and the results of the above theorem hold.

The above theorem says that so long as the decision maker does not like any of the efficient tradeoffs from a solution x^*, then solution x^* is optimal. We would like to relax the condition in practice to the decision maker not liking, or being uncertain about any of the efficient tradeoffs, but strictly speaking, we cannot. In that case, we can, however, argue that some part of the corresponding edge from x^* is contained in a half-space defined by a supporting hyperplane tangent to the underlying utility function at x^*.

Allowing that the "I don't know" or "uncertain" response is a fuzzy response of indifference, the supporting hyperplane is approximately tangent to the utility function at x^*. Allowing for inaccuracies or errors in responses, the hyperplane tangent to the utility function at x^* may intersect the

feasible region. Moreover, the solutions which are preferred to x^* are contained in the intersection of the feasible region and the constraint $u(x) \geq u(x^*)$. This set is a convex set which is a subset of the set defined by the feasible region and the hyperplane tangent to $u(x) \geq u(x^*)$ at x^*. (Note that the worst case is if the utility function is linear and the two sets coincide.) This latter set should generally not be large for two reasons:

1. The solution x^* is an extreme point solution, and therefore there is an infinite number of supporting hyperplanes containing x^*.
2. Any tradeoff responses which are answered negatively (none are answered affirmatively if the method terminates at x^*!) are edges containing x^* which cannot be included in the constructed convex set.

We cannot be more precise about the size of the region of better solutions than x^*. We are planning some computer simulation experiments to investigate questions related to how large the region is in practice. At any rate, our contention is that the region will normally be quite small. Therefore, we believe that choosing x^* as optimal will generally result in little or no error.

An alternative procedure which will not leave the indeterminacy outlined above is to do as follows:

1. Assume that the current solution x^* is such that the decision maker does not like any of the tradeoffs obtained along efficient edges moving from x^*, and is uncertain about one or

more tradeoffs. (Given the method, this is the only way we can have the uncertainty.)

2. For each efficient edge about which the decision maker is uncertain, construct the adjacent efficient extreme point. Denote it as x^1. Then we should consider efficient edges emanating from x^1 (except the one to x^*). Using convexity considerations we may argue that the decision maker should not like any of the edges emanating from x^1. However, if he is uncertain about any of these, we should construct the corresponding adjacent efficient solutions and repeat the above procedure.

Unfortunately, this procedure sounds cumbersome, to say the least.

A procedure such as the above need be used only if an answer of uncertain or I don't know is given with respect to tradeoffs offered at the last extreme point solution. If the decision maker does not like any of the tradeoffs offered at the last extreme point solution, such a procedure, of course, need not be utilized.

Now, if the procedure terminates in step 12, all we know is that x^* is a local optimum, and that there exist superior solutions. Another extreme point solution may even be optimal! Deshpande [2] has developed a method for using search techniques on facets of the polyhedral set, thereby finding an optimum in such cases. However, his method is sufficiently cumbersome that it does not appear to be of practical value at this time. We are currently conducting experiments on randomly generated problems to see how close our procedure stopping at an extreme point comes to the true optimum of a concave utility function.

By virtue of the large number of efficient extreme points of large multiple criteria linear programming problems (see, for example, Evans and Steuer [4] and Yu and Zeleny [15]) and the relative proximity that their large number implies, we anticipate that the difference between the utility of the last extreme point solution found and the true optimum will generally be relatively small. If this turns out not to be the case, we shall investigate Deshpande's search procedure further.

4. Some Comparisons with Other Methods

In order to put our method in perspective, it may be useful to compare it with several of the other methods which are relatively prominent. The method of Geoffrion, Dyer, and Feinberg [5] in its various forms begins with a feasible solution x^1, and has the decision maker provide information from which the partial derivatives (using our notation) $\partial G/\partial u_i$, $i = 1,\ldots,p$ can be estimated. Then using the vector of partials as weights, a composite objective is optimized over the feasible solution space to obtain a solution x^2. A single dimensional search is carried out by the decision maker to determine the best solution on the straight line segment connecting x^1 and x^2. The new solution becomes solution x^1 and the process is repeated until solution x^1 does not change from one iteration to the next.

The method of White [13] is related to that of Steuer although it was derived in part from our earlier method. It allows for prior restrictions on the weights as well as an unspecified form of utility function. In the same way that Steuer generates sets of solutions, this method generates sets,

except that the number of solutions per set increases as the procedure is used. Sequential pairs of solutions are presented to the decision maker for his choice and the results of the comparisons lead to the shrinking of the convex cone. Though the method looks rather interesting, no empirical testing of the method is cited, and the increasing number of solutions presented could be a problem.

Two other methods that should be discussed are the method of goal programming (see, for example, Charnes and Cooper [1]) and the scalarizing method of Wierzbicki [14]. Both involve the choice of an ideal point solution and use linear programming methods to find a suitably defined "closest point". Wierzbicki uses the ideal point solution as a trial-and-error mechanism by which points in the neighborhood of the optimal solution to the problem may be generated, whereas the traditional goal programming approach requires that the user specify weights to use in the finding of the closest point. Goal programming has been rather widely used; the scalarizing technique is rather recent in origin and has relatively few applications to date.

Our method asks pairwise alternative comparisons of the decision makers. It avoids the problem of line search and does not lead to large numbers of solution points. We next consider certain computational aspects of our method.

5. Computational Aspects of the Method

In the process of developing and applying our method, we have written several computer codes of it. Thus far, our prime interest was in tackling specific applied problems. In the

early developmental phases, we used simple linear programming codes to solve appropriate subproblems. We were, therefore, able to solve small problems in a piecemeal fashion using the method. So that we could draw upon programming skills of others all of our subsequent programming work built upon existing linear programming packages. Our first program, written strictly as an experiment, was written in FORTRAN for an IBM 370/158 and built upon the IBM Mathematical Programming package MPSX. We used it in several projects. Even though we had never intended it as an end-user program, we have sent listings of our code and a terse set of instructions to about thirty organizations. Several of these organizations are now using the program -- some with our help, some without. One large Dutch company got the program running by themselves, and is now involved in a multi-million dollar strategic planning problem using the method. Two programs based on the first were written for the Brookhaven National Laboratory. The first was of the original method. The second approximates what we have described here. They were also written in FORTRAN, but were designed and implemented on the CDC 6600-7600 computer at Brookhaven. The codes interact with the CDC APEX system, with the 6600 operating interactively and the 7600 being used to solve large linear programming problems with APEX. A third program was written in Finland on the Univac 1108 at the Helsinki School of Economics. It too is an interactive program, but was not used to solve problems as large as the other systems. See Wallenius, Wallenius, and Vartia [10] for more information.

A fourth computer program that is now running experimentally has been written for the SUNYAB CDC Cyber 173 in FORTRAN and it

utilizes Roy Marsten's XMP [6], which is a high technology user-oriented linear programming FORTRAN computer package. We plan to fully document this program, and to make it available to whoever wants it. Users will, of course, have to obtain XMP from Marsten. Since Marsten's code is available for several different computers and does not require a special computer configuration nor an expensive linear programming package such as MPSX or APEX, it should be a worthwhile code available to almost anyone.

We are currently experimenting with this last code. There are two changes in particular which seem worthy of testing and implementation. Both are based on our experience with integer multiple criteria linear programming (Villareal, Karwan, and Zionts [9]), and are changes which seem to reduce computation time. Though that may not be as great a concern in linear programming as in integer programming, the results seem sufficiently worthwhile as to make them worth incorporating. The first involves testing the set of previous responses, and eliminating redundant constraints from that set. This had a remarkable effect in accelerating the problem solution times and has enabled us to solve much larger integer problems than we were able to solve otherwise in a given amount of time. The second change is one that seems to reduce the number of questions asked of the decision maker. The current step 9 says to find any feasible set of weights λ consistent with all previous responses. Instead, in our integer programming method we find what we call a "middle-most" set of weights. That is, we use an objective function which satisfies the response constraints as well as the bound constraints on the weights "as much as possible" in that

the slack of the least satisfied constraint is maximized. As pointed out by Jack Elzinga, Department of Industrial and Systems Engineering, University of Florida, this is equivalent to finding the center of the largest hypersphere which may be inscribed within the feasible solution space. In our integer programming work, this handsomely accelerated the determination of the optimal solutions.

6. Practical Experience with the Method

Numerous problems have been solved with variations of the method. After solving numerous small problems which assumed a linear utility function, we worked on a substantial long-range planning problem for S. A. Cockerill, a large Belgian integrated steel company. The problem was a time-phased investment model consisting of four objectives, 143 constraints, and 248 variables. See Wallenius and Zionts [11] for further information. Our method is also being used by the Philips Company in Eindhoven, The Netherlands to solve a strategic management problem involving seven objectives. A form of the general concave method has been used for national economic planning in Finland. Four objectives were used (for more information, see Wallenius, Wallenius, and Vartia [10]). In addition, another rather large problem has been solved in various forms by several decision makers at the Brookhaven National Laboratory and at the Department of Energy in the U.S. That model is an energy planning model consisting of six objectives and several hundred constraints (for more information, see Zionts and Deshpande [17]).

The computational requirements are on the order of one linear program solution for each setting or revision of weights

(steps 1 and 10). The maximum number of setting or revision of weights has always been less than ten. The total number of questions asked of the decision maker has always been less than 100, and generally less than 50. We have ways of reducing the number of questions, if that should ever be a problem. Our users have generally been quite satisfied with the model.

Though used in several situations, the method may suffer from slow convergence properties of nonlinear programming methods. See also Wehrung [12].

The approach of Oppenheimer [7] is an adaptation of that of Geoffrion, Dyer, and Feinberg (GDF). Whereas the GDF method uses a linear approximation, Oppenheimer's approach uses a nonlinear proxy function which may be one of several types. The procedure works in the following general manner: The proxy function is maximized over the feasible solution space. If the resulting solution is an improvement over the previous solution, it becomes the starting point for the next iteration. If not, a point on the straight line segment between the other two solutions which is preferred to the starting solution is used as the starting point for the next iteration. The procedure continues until two successive solutions are identical. Even though the proxy function need not be a close approximation to the underlying unknown utility function, the method generally provides a superior convergence to the GDF method.

The method of Steuer [8] is somewhat different from the other methods. It assumes a linear utility function; the methods described so far assume nonlinear utility functions. It generates a sequence of sets of efficient solutions and asks the decision maker to choose the most-preferred solution of each set.

The set of vectors of weights used to generate the solutions are generators of a convex cone. The method then shrinks the cone about the most-preferred solution and the process is repeated. Though the procedure is not guaranteed to find the optimal solution (see Zionts [16] for a counterexample), it generates good solutions, and has been applied to several practical problems.

APPENDIX

In this appendix we present two examples to illustrate what we perceive to be anomalies that can occur. Problems from these anomalies can fortunately be overcome. We wish to thank Pekka Korhonen, Helsinki School of Economics, and Tim Lowe of Purdue University for suggestions leading to the construction of these examples. We indicate the step numbers of the algorithm where appropriate.

The first example is:

$$\begin{aligned} 3x_1 + 2x_2 + 3x_3 &\leq 18 \\ x_1 + 2x_2 + x_3 &\leq 10 \\ 9x_1 + 20x_2 + 7x_3 &\leq 96 \\ 7x_1 + 20x_2 + 9x_3 &\leq 96 \\ x_1, x_2, x_3 &\geq 0 \end{aligned}$$

The objective functions are $u_1 = x_1$, $u_2 = x_2$, $u_3 = x_3$. All are to be maximized. A (concave) utility function to be maximized is Maximize $u = \underset{x}{\text{Minimum}} \{3x_1, 5x_2, 3x_3\}$. This function does not have continuous derivatives, but a function $u = -\left((25 - 3u_1)^k + (25 - 5u_2)^k + (25 - 3u_3)^k\right)^{1/k}$ with k suffi-

ciently large is an almost equivalent concave utility function having continuous derivatives. In our example we shall assume the second utility function. All efficient extreme point solutions to the problem are A(0, 4.8, 0), B(1, 4, 1), C(4, 3, 0), D(0, 3, 4), E(0, 0, 6), and F(6, 0, 0). (Step 1) Using the method beginning with equal weights yields either solution C or D. (The problem is symmetric with respect to x_1 and x_3.) (Step 2) Assuming that C is the first solution, solutions A, B, D, and F will be offered to the decision maker as alternatives. (Steps 3, 4) The decision maker should prefer B only, but should not like solutions A, D, or F compared to C. (Step 8) Constraints reflecting the preferences are added. (Step 9) A new consistent set of weights (1/8 7/16 7/16) is found. (Step 10) The new solution is solution D. (Steps 11, 13) The manager chooses between solutions D and B. He prefers B. We then add a constraint reflecting this preference. (Step 2) The solutions adjacent to B are A, C, and D. (Step 3) The sets are $\underset{\sim}{A} = \{A\}$ and $\underset{\sim}{B} = \{C,D\}$. (Step 4) For the first set the decision maker is asked to choose between alternatives A and B. He prefers B. (Step 5) No questions are asked. (Step 6) The manager is asked about the tradeoff leading to A. He does not like it. (Step 7) Let $\underset{\sim}{I}$ equal $\underset{\sim}{B}(\{C,D\})$. (Step 4) The manager is asked to compare B with C, and B with D. He prefers B in each case. (Step 5) No questions are asked. (Step 6) The manager is asked to consider the questions asked as tradeoffs. He likes both. (Step 8) Constraints are added. (Step 9) Constraints are dropped. (Step 10) The new solution is solution C. (Step 11) The decision maker chooses B. (Step 12) We must terminate and proceed with a search procedure.

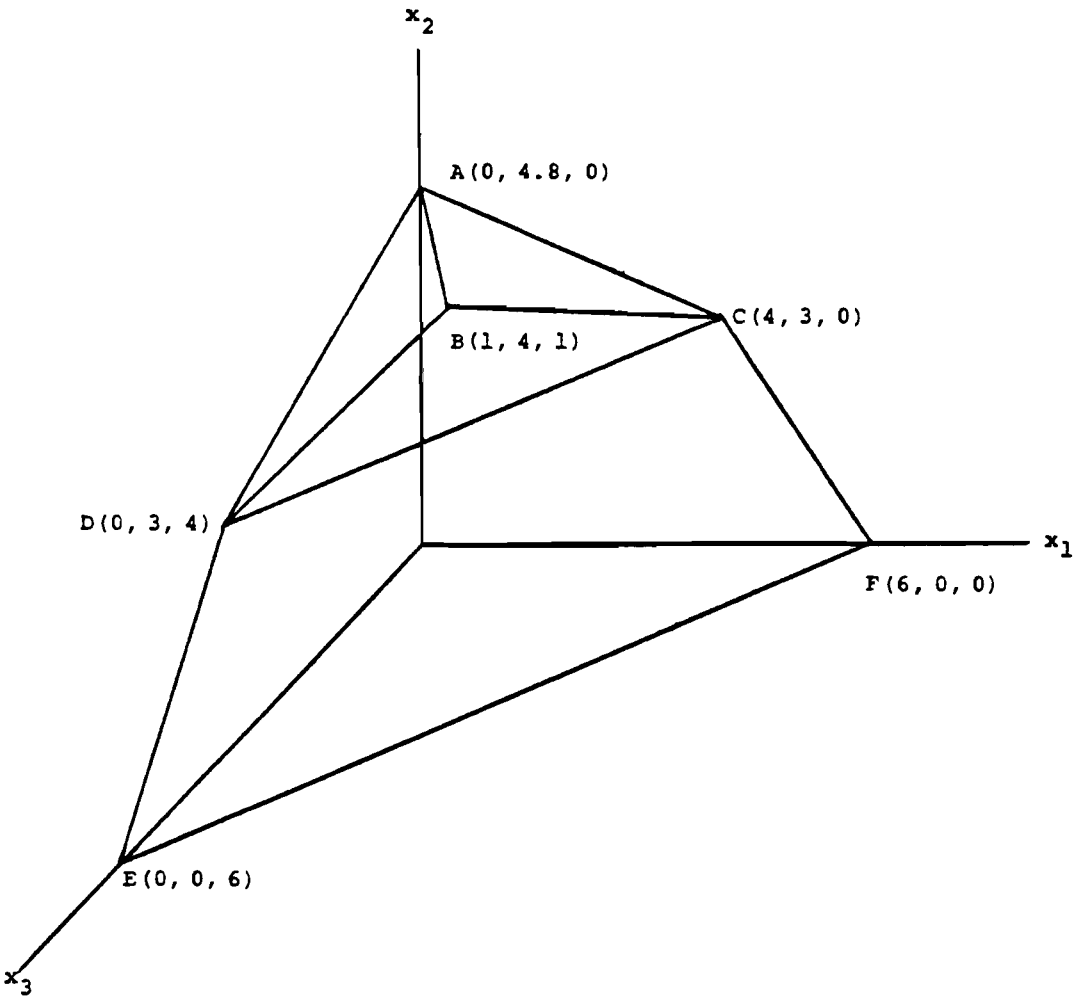


FIGURE 1

A plot of the example.

The search procedure yields the true optimal solution not on the first facet B C D but on an adjacent facet C D E F. The optimal solution is the point 2.5, 1.5, 2.5. Were we to have used the first utility function, we would have terminated (incorrectly) with solution B in step 7.

Suppose we replace the first constraint above by the following constraints:

$$\begin{aligned} 2x_1 + x_2 + 2x_3 &\leq 12 \\ x_1 + x_2 + x_3 &\leq 7 \\ 39x_1 + 30x_2 + 45x_3 &\leq 270 \\ 45x_1 + 30x_2 + 39x_3 &\leq 270 \end{aligned}$$

Then assuming that we terminated as above at point B using the second utility function, the search proceeds as above, but it terminates at an extreme point (2.5, 2, 2.5) which extends approximately from the center of the facet C D E F in Figure 1. In this case the true optimal is at an extreme point which is not in a facet containing the locally optimal solution, solution B.

REFERENCES

1 Charnes, A., and Cooper, W. W., "Goal Programming and Multiple Objective Optimization - Part 1," European Journal of Operations Research, 1, pp. 39-54 (1977).

2 Deshpande, D., "Investigations in Multiple Objective Linear Programming - Theory and an Application," unpublished Doctoral Dissertation, State University of New York, Buffalo (1981).

3 Ecker, J. G., Hegner, N. S., and Kouada, I. A., "Generating All Maximal Efficient Faces for Multiple Objective Linear Programs," Journal of Optimization Theory and Applications, Vol. 30, No. 3, (1980), pp. 353-381.

4 Evans, J. P. and Steuer, R. E., "Generating Efficient Extreme Points in Linear Multiple Objective Programming: Two Algorithms and Computing Experience," in Cochrane, J. L., and Zeleny, M. (Eds.) Multiple Criteria Decision Making, University of South Carolina Press (1973).

5 Geoffrion, A. M., Dyer, J., and Feinberg, A., "An Interactive Approach for Multicriterion Optimization with an Application to the Operation of an Academic Department," Management Science, Vol. 19, No. 4 (1972), pp. 357-368.

6 Marsten, R. E., "XMP: A Structured Library of Subroutines for Experimental Mathematical Programming," Technical Report No. 351, Management Information Systems, University of Arizona, Tucson (1979).

7 Oppenheimer, K. R., "A Proxy Approach to Multiattribute Decision Making," Management Science, Vol. 24, No. 6 (1978), pp. 675-689.

8 Steuer, R. E., and Schuler, A. T., "An Interactive Multiple-Objective Linear Programming Approach to a Problem in Forest Management," Operations Research, Vol. 26, No. 2 (1978), pp. 254-269.

9 Villareal, B., Karwan, M. H., and Zionts, S., "A Branch and Bound Approach to Interactive Multicriteria Integer Programming," paper presented at the Joint National Meeting of TIMS/ORSA, Washington, D.C., May 4-7, 1980.

10 Wallenius, H., Wallenius, J., and Vartia, P., "An Approach to Solving Multiple Criteria Macroeconomic Policy Problems and an Application," Management Science, Vol. 24, No. 10 (1978), pp. 1021-1030.

11 Wallenius, J. and Zionts, S., "Some Tests of an Interactive Programming Method for Multicriterion Optimization and An Attempt at Implementation," in H. Thiriez and S. Zionts (Eds.), Multiple Criteria Decision Making, Jouy-en-Josas, France, 1975, Springer-Verlag, Berlin, (1976), pp. 319-330.

12 Wehrung, D. A., "Interactive Identification and Optimization Using a Binary Preference Relation," Operations Research, Vol. 26, No. 2 (1978), pp. 322-331.

13 White, D. J., "Multi-Objective Interactive Programming," Journal of the Operational Research Society, Vol. 31 (1980), pp. 517-523.

14 Wierzbicki, A. P., "The Use of Reference Objectives in Multiobjective Optimization," Working Paper 79-66, International Institute for Applied Systems Analysis, Laxenburg, Austria (1979).

15 Yu, P. L. and Zeleny, M., "The Set of All Nondominated Solutions in the Linear Cases and a Multicriteria Simplex Method," Journal of Mathematical Analysis and Applications, 49, pp. 430-468.

16 Zionts, S., "Multiple Criteria Decision Making: An Overview and Several Approaches," Working Paper No. 454, State University of New York at Buffalo (1980).

17 Zionts, S. and Deshpande, D., "A Time Sharing Computer Programming Application of a Multiple Criteria Decision Method to Energy Planning — A Progress Report," in S. Zionts (Ed.), Multiple Criteria Problem Solving, Proceedings, Buffalo, NY, 1977, Springer-Verlag, Berlin (1978), pp. 549-560.

18 Zionts, S. and Wallenius, J., "An Interactive Programming Method for Solving the Multiple Criteria Problem," Management Science, Vol. 22, No. 6 (1976), pp. 652-663.

19 Zionts, S. and Wallenius, J., "Identifying Efficient Vectors: Some Theory and Computational Results," Operations Research, Vol. 28, No. 3 (1980), pp.

ON SAMPLING THE EFFICIENT SET USING WEIGHTED TCHEBYCHEFF METRICS

R.E. Steuer

College of Business Administration, University of Georgia, Athens, Georgia, USA

Abstract

An interactive procedure for solving a multiple objective program is described. The procedure operates by first computing an ideal criterion vector and then sampling the set of nondominated criterion vectors using variously weighted Tchebycheff metrics. To geometrically explore the procedure in all its nuances, a series of graphical examples is presented to better understand the computational behavior of the method. One advantage of the Tchebycheff procedure is that it can converge to non-extreme optimal solutions. In addition to the linear case, the procedure is applicable to integer and nonlinear multiple objective problems.

1. Introduction

Let us consider the multiple objective program

$$\begin{aligned} &\max \quad \{f_1(x) = z_1\} \\ &\max \quad \{f_2(x) = z_2\} \\ &\quad \vdots \\ &\max \quad \{f_k(x) = z_k\} \\ &\text{s.t.} \quad\quad x \in S \end{aligned}$$

where the f_i may be nonlinear and S may be nonconvex. Also, let us assume that S is bounded.

In Steuer and Choo [6] an interactive weighted Tchebycheff algorithm is described for solving the above multiple objective program. The purpose of this paper is to provide, through a series of graphical examples, a basis for a thorough geometrical understanding of the weighted Tchebycheff approach.

The paper is organized as follows. Section 2 introduces notation and Section 3 specifies the interactive algorithm. Section 4 discusses weighted Tchebycheff metrics along with the augmented and lexicographic weighted Tchebycheff program formulations. Guidelines for calibrating the algorithm are provided in Section 5. Six graphical examples comprise Section 6. Section 7 ends the paper with concluding remarks.

2. Notation

It is necessary to establish notation:

1. Let $Z \subset R^k$ be the set of all feasible criterion vectors where Z is the set of images of all $x \in S$ under the f_i.
2. Then $\bar{z} \in Z$ is a nondominated criterion vector if and only if there does not exist another $z \in Z$ such that $z_i \geq \bar{z}_i$ for all i and $z_i > \bar{z}_i$ for at least one i.
3. Let $N \subset Z$ denote the set of all nondominated criterion vectors.
4. Then $\bar{x} \in S$ is an efficient point if and only if $\bar{x}$ is a preimage of a $\bar{z} \in N$.
5. Let $E \subset S$ denote the set of all efficient points.
6. $z^\circ \in Z$ is an optimal criterion vector if and only if z° maximizes the decision-maker's (mathematically unknown) utility function $U: Z \to R$. Clearly $z^\circ \in N$.
7. Then, if $x^\circ \in S$ is preimage of an optimal criterion vector z°, x° is an optimal solution of the multiple objective program. Clearly $x^\circ \in E$.

Note that the term "nondominated" is only applied to vectors in criterion space and the term "efficient" is only applied to points in decision space.

3. Interactive Procedure

The strategy of the interactive Tchebycheff procedure is to initially sample N (the set of all nondominated criterion vectors) and then sample successively smaller neighborhoods in N. At each iteration the decision-maker is asked to select his most preferred of the criterion vectors presented to him. The iterative process continues until a criterion vector $\bar{z} \in N$ close enough to being optimal (if not actually optimal) is obtained to terminate the decision process. Then by taking the preimage of $\bar{z}$ (or searching among the preimages of $\bar{z}$ if there are more than one), we have the final solution achieved by the algorithm.

Within a more advanced version of the interactive framework used in Steuer and Schuler [9], the weighted Tchebycheff approach synthesizes ideas gleaned from

(a) the weighted Tchebycheff results of Bowman [1], Choo and Atkins [3], and Choo [2];

(b) the compromise programming work of Zeleny [11];

(c) the scalarizing function methods of Wierzbicki [10]; and

(d) the "filtering" techniques of Steuer and Harris [7].

Other recently completed research utilizing the Tchebycheff (minimax) philosophy is found in Ecker and Shoemaker [4].

Let

p ∿ the sample size (the size of the sample of criterion vectors that is presented to the decision-maker at each iteration)

t ∿ the number of iterations the interactive procedure is to run

r ∿ the convergence factor whose purpose is to sequentially reduce weighting vector space $\bar{\Lambda} = \{\lambda \in R^k | \lambda_i \geq 0, \sum_{i=1}^{k} \lambda_i = 1\}$.

With values for p, t and r calibrated as in Section 5, the interactive procedure is as follows:

Step 1: Solve for the ideal criterion vector $z^* \in R^k$ such that

$$z_i^* = \max \{f_i(x) \mid x \in S\} + \varepsilon_i$$

where a given $\varepsilon_i = 0$ unless

(i) there is more than one nondominated criterion vector that maximizes the i^{th} objective, or

(ii) the only nondominated criterion vector that maximizes the i^{th} objective also maximizes one of the other objectives

in which case ε_i is set to a positive scalar.

Step 2: Let $\bar{\Lambda}^{(1)} = \{\lambda \in R^k \mid \lambda_i \in [0, 1], \sum_{i=1}^{k} \lambda_i = 1\}$

Step 3: Let $h = 0$

Step 4: Let $h = h + 1$

Step 5: Randomly generate, for instance, 50 x k weighting vectors from $\bar{\Lambda}^{(h)}$.

Step 6: Filter the randomly generated λ-vectors of Step 5 to obtain, for instance, 2 x p maximally dispersed representatives.

Step 7: For each of the 2p λ-vectors of Step 6, solve the augmented (or lexicographic) weighted Tchebycheff program (see Section 4).

Step 8: Filter the resulting criterion vectors to obtain a sample of p nondominated criterion vectors.

Step 9: Let $z^{(h)}$ designate the criterion vector selected by the decision-maker as the most preferred from the sample of Step 8.

Step 10: If the decision-maker wishes to prematurely stop iterating, go to Step 15. Otherwise, proceed to Step 11.

Step 11: Let $\lambda^{(h)}$ designate the λ-vector whose components are given by

$$\lambda_i^{(h)} = \begin{cases} \dfrac{\dfrac{1}{(z_i^* - z_i^{(h)})}}{\left[\sum_{i=1}^{k} \dfrac{1}{(z_i^* - z_i^{(h)})}\right]} & \ldots \text{ if } z_i^{(h)} \neq z_i^* \text{ for all } i \\ 1 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \text{if } z_i^{(h)} = z_i^* \\ 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \text{if } z_i^{(h)} \neq z_i^* \text{ but } \exists j \neq i, z_j^{(h)} = z_j^* \end{cases}$$

Step 12: With $\bar{\Lambda}^{(h)}$ being the λ-vector computed in Step 11, define

$$\bar{\Lambda}^{(h+1)} = \{\lambda \in R^k | \lambda_i \in [\ell_i, u_i], \sum_{i=1}^{k} \lambda_i = 1\}$$

where

$$[\ell_i, u_i] = \begin{cases} [0, r^h] \ldots\ldots\ldots\ldots\ldots\ldots \text{if } \lambda_i^{(h)} - \frac{r^h}{2} \leq 0 \\ [1-r^h, 1] \ldots\ldots\ldots\ldots\ldots\ldots \text{if } \lambda_i^{(h)} + \frac{r^h}{2} \geq 1 \\ [\lambda_i^{(h)} - \frac{r^h}{2}, \lambda_i^{(h)} + \frac{r^h}{2}] \ldots \text{otherwise} \end{cases}$$

in which r^h is r raised to the h^{th} power.

Step 13: If $h < t$, go to Step 4. If $h \geq t$, go to Step 14.

Step 14: If the decision-maker wishes to keep iterating, go to Step 4. Otherwise, go to Step 15.

Step 15: Compute a preimage of the decision-maker's final criterion vector selection. Stop.

4. Weighted Tchebycheff Programs

As specified in Section 3, the interactive procedure first computes an ideal criterion vector z^*. By computing the $z \in Z$ closest to z^* (using differently weighted Tchebycheff metrics), the algorithm samples the set of all nondominated criterion vectors in an ever more concentrated fashion until termination.

With $\lambda \in \bar{\Lambda} = \{\lambda \in R^k | \lambda_i \geq 0, \sum_{i=1}^{k} \lambda_i = 1\}$, let

$$\| z^* - z \|_\infty^\lambda = \max_{i=1,\ldots,k} \{\lambda_i | z_i^* - z_i|\} + \rho \sum_{i=1}^{k} |z_i^* - z_i|$$

define the augmented weighted Tchebycheff metric for measuring the distance between any $z \in Z$ and the ideal criterion vector z^*. The term "augmented" is applied because of the presence of the term $\rho \sum_{i=1}^{k} |z_i^* - z_i|$ where ρ is a nonnegative scalar.

For $\lambda = (2/5, 3/5)$, illustrated in Figure 1 is

(a) the point $\bar{z} \in Z$ closest to z^* according to the augmented weighted Tchebycheff metric, and

(b) the lowest valued $|||\ z^* - z\ |||_\infty^\lambda$ isoquant that intersects Z

where

$$\theta_1 = \tan^{-1}\left(\frac{\rho}{1 - \lambda_1 + \rho}\right)$$

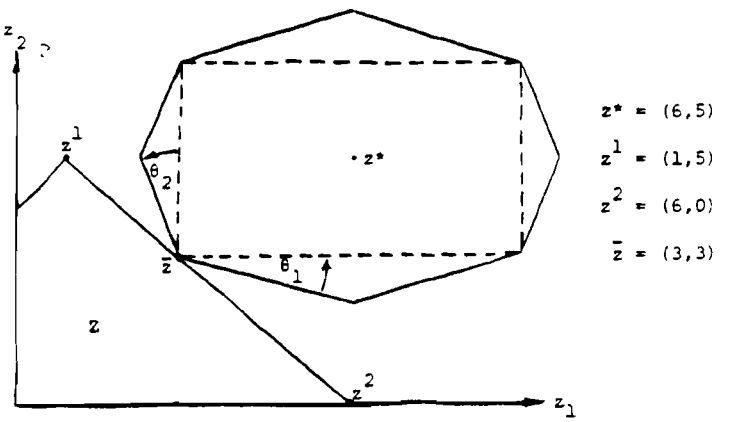
$$\theta_2 = \tan^{-1}\left(\frac{\rho}{1 - \lambda_2 + \rho}\right)$$

The corner point of the $|||\ z^* - z\ |||_\infty^\lambda$ isoquant in the negative orthant translated to z^* (e.g., $\bar{z}$ in Figure 1) is called the vertex of the isoquant. For any vertex, the direction of the line from the vertex to z^* is given by

$$(1/\lambda_1,\ 1/\lambda_2,\ \ldots,\ 1/\lambda_k)$$

Figure 1

Minimizing $|||\ z^* - z\ |||_\infty^\lambda$ Isoquant Intersecting Z



Note that the only relevant portion of the $|||\ z^* - z\ |||_\infty^\lambda$ isoquant is in the nonpositive orthant translated to z^* (because intersections with Z cannot occur anyplace else).

Utilizing the augmented weighted Tchebycheff metric to compute the $\bar{z} \in Z$ closest to z^*, we solve the augmented weighted Tchebycheff program

$$
(4.1)\qquad
\left\{
\begin{array}{ll}
\min & \{\alpha + \rho e^{T}(z^{*} - z)\} \\
\text{s.t.} & \alpha \geq \lambda_i (z_i^{*} - z_i) \qquad 1 \leq i \leq k \\
& f_i(x) = z_i \qquad 1 \leq i \leq k \\
& x \in S
\end{array}
\right.
$$

in which e^{T} is the sum vector of ones. For use with discrete and polyhedral feasible regions S, ρ is to be a sufficiently small positive scalar for which theoretical upper bounds are provided in [6]. In practice, however, a computationally significant value of $\rho = .001$ should suffice in most cases. When using the augmented weighted Tchebycheff metric, (4.1) is the program that is solved 2p times in Step 7 of the algorithm.

As demonstrated in [6], the usefulness of (4.1) is that $\bar{z} \in N \Leftrightarrow$ there exists a $\lambda \in \bar{\Lambda}$ such that $\bar{z}$ uniquely minimizes the augmented weighted Tchebycheff program.

In place of the augmented weighted Tchebycheff metric, we can compute the z's in Z closest to z^{*} using differently weighted lexicographic Tchebycheff metrics

$$
P_1 \left[\max_{i=1,\ldots,k} \{\lambda_i |z_i^{*} - z_i|\} \right] + P_2 \left[\sum_{i=1}^{k} |z_i^{*} - z_i| \right]
$$

where $\lambda \in \bar{\Lambda}$ and $P_1 >>> P_2$ (as in goal programming).

Being a non-Archimedian measure, it is not possible to draw an isoquant of a lexicographic weighted Tchebycheff metric. However, the $z \in Z$ closest to z^{*} according to the lexicographic weighted Tchebycheff metric is the point on the $\rho = 0$ $\|z^{*} - z\|_{\infty}^{\lambda}$ isoquant (dashed rectangle in Figure 1) closest to z^{*} according to L_1-metric. Note that the P_2 term of the metric is only invoked when there are ties with respect to the P_1 term.

Utilizing the lexicographic weighted Tchebycheff metric to compute the $\bar{z} \in Z$ closest to z^{*}, we solve the lexicographic weighted Tchebycheff program

$$
(4.2)\qquad
\begin{cases}
\min & \{P_1\alpha + P_2 e^T(z^* - z)\} & \\
\text{s.t.} & \alpha \geq \lambda_i(z_i^* - z_i) & 1 \leq i \leq k \\
& f_i(x) = z_i & 1 \leq i \leq k \\
& x \in S &
\end{cases}
$$

As shown in [6], for any S (discrete, polyhedral, nonpolyhedral-continuous), $\bar{z} \in N \Leftrightarrow$ there exists a $\lambda \in \bar{\Lambda}$ such that $\bar{z}$ uniquely minimizes the lexicographic weighted Tchebycheff program. When using the lexicographic weighted Tchebycheff metric, (4.2) is the program that is solved 2p times in Step 7 of the algorithm.

5. Calibrating the Algorithm

Based upon the motivation drawn from the unit hypercube in [6] in conjunction with personal computational experience, it is recommended that p, t and r be calibrated in accordance with the following relationships

$$
(5.1)\qquad
\begin{cases}
\sqrt[k]{1/p} \leq r \leq \sqrt[t-1]{w} \\
1/2k \leq w \leq 3/2k \\
p \geq k
\end{cases}
$$

where w is the estimated $[\ell_i, \mu_i]$ interval width of the iteration in which the decision-maker will become satisfied with his final solution.

From the above relationships we have, for instance, (a) the larger p, the fewer iterations that are required, (b) the smaller r, the faster the algorithm will converge, and (c) the narrower w, the longer it is anticipated that it will be before we find and recognize the final solution.

A surprising result that the author cannot fully explain is that the interactive Tchebycheff algorithm tends to converge to a final solution faster than expected (that is, in fewer than the t iterations as calibrated from (5.1)). This has been the author's experience on the 8 objective problem in [8] and the 11 objective problem in [5]. Because of this, Step 10 is included in the algorithm.

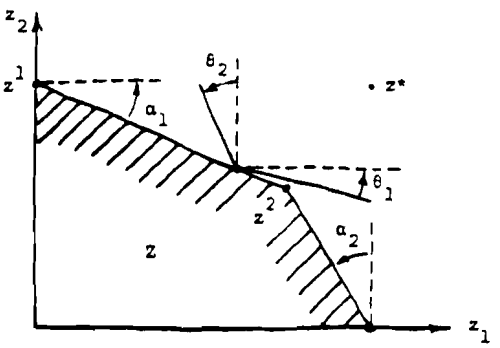
6. Graphical Examples

In this section is used the γ notation where $\gamma(x^1, x^2, \ldots, x^q)$ denotes the set of all convex combinations of the x^i $1 \leq i \leq q$.

Example 6.1: Consider the multiple objective program whose Z is indicated by the shaded area. What is the least upper bound $\bar{\rho}$ such that for $0 < \rho < \bar{\rho}$ each member of N is uniquely computable by the augmented weighted Tchebycheff program?

Figure 2

Graph of Example 6.1



Solution 6.1: In this problem, $N = \gamma(z^1, z^2, z^3)$. To determine $\bar{\rho}$, we note that for each member of N to uniquely minimize the augmented weighted Tchebycheff program, both of the relationships

$$\theta_1 = \tan^{-1}\left(\frac{\rho}{1 - \lambda_1 + \rho}\right) < \alpha_1$$

and

$$\theta_2 = \tan^{-1}\left(\frac{\rho}{1 - \lambda_2 + \rho}\right) < \alpha_2$$

must hold. Otherwise members of N in the relative interior of $\gamma(z^1, z^2)$ and/or $\gamma(z^2, z^3)$ could not be generated by the

vertex of the minimizing $|||z^* - z|||_\infty^\lambda$ isoquant that intersects Z for any $\lambda \in \bar{\Lambda}$.

Example 6.2: Let $z^* = (10, 10, 10)$. Draw the $|||z^* - z|||_\infty^\lambda = 1$ isoquants where $\rho = 0$ for

(a) $\lambda = (.4, .2, .4)$

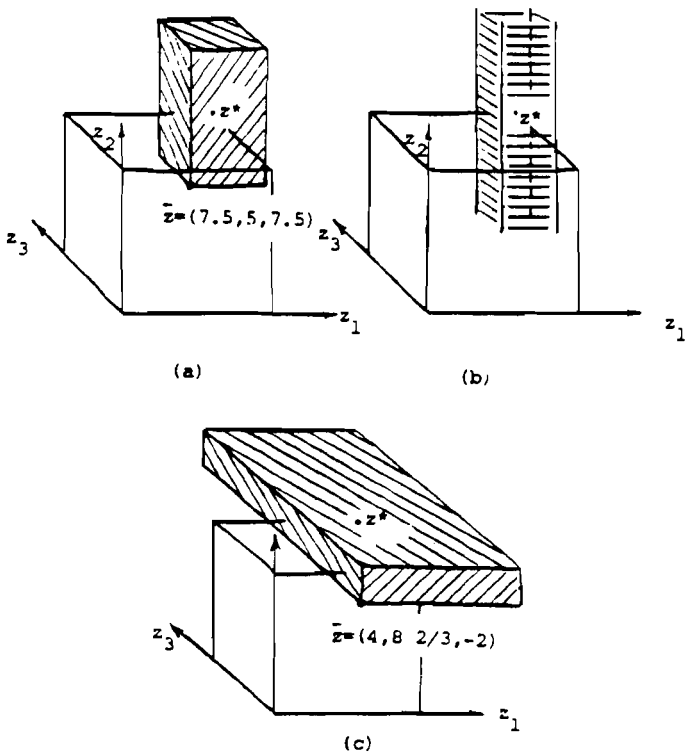
(b) $\lambda = (1/2, 0, 1/2)$

(c) $\lambda = (1/6, 3/4, 1/12)$

Solution 6.2: With vertices at $\bar{z}$, the $|||z^* - z|||_\infty^\lambda = 1$ isoquants are portrayed in Figure 3.

Figure 3

Graphs of Example 6.2



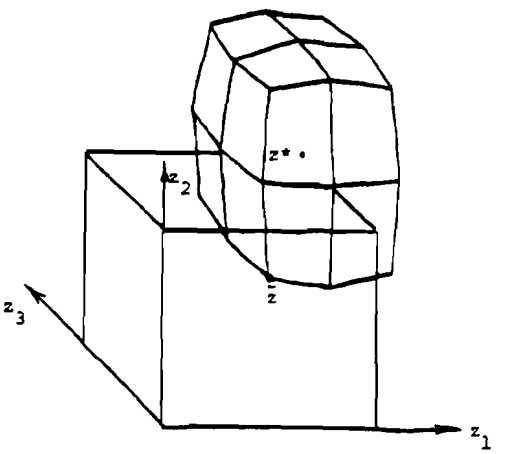
Whereas the isoquants of (a) and (c) have vertices, the isoquant of (b) is unbounded from above and below (because $\lambda_2 = 0$) with the edge closest to the origin being given by $\{z \in R^3 \mid z = (8, \beta, 8), \beta \in R\}$. If for instance we had case (d) where $\lambda = (0, 1, 0)$, the $|||z^* - z|||_\infty^\lambda = 1$ isoquant would be the horizontal plane intersecting the z_2 axis at (0, 9, 0). With $\rho > 0$, the faces of the $|||z^* - z|||_\infty^\lambda$ isoquant would have slight slopes.

Example 6.3: With $\rho = .01$, $z^* = (10, 10, 10)$ and $\lambda = (.4, .2, .4)$, draw the $|||z^* - z|||_\infty^\lambda = 1$ isoquant of the augmented weighted Tchebycheff metric.

Solution 6.3:

Figure 4

Graph of Example 6.3



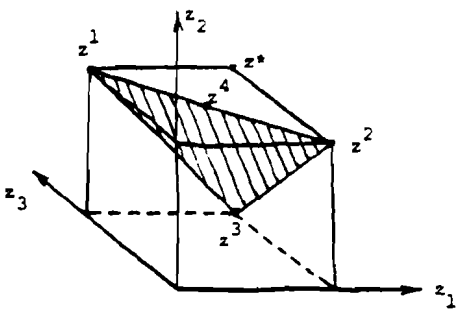
Note that in this example the vertex of the $|||z^* - z|||_\infty^\lambda = 1$ isoquant occurs at $\bar{z} = (85/11, 60/11, 85/11)$ instead of (7.5,5,7.5) when $\rho = 0$.

Example 6.4: What might be the consequence of setting an $\varepsilon_i = 0$ when, by the conditions in Step 1 of the algorithm, the ε_i should be set to a positive scalar?

Solution 6.4: Consider the MOLP (multiple objective linear program) whose Z is the shaded area in Figure 5.

Figure 5

Graph of Example 6.4



$z^1 = (0,3,3)$

$z^2 = (3,3,0)$

$z^3 = (3,0,3)$

$z^4 = (3/2,3,3/2)$

In this problem, $N = \gamma(z^1, z^2, z^3)$. By the conditions in Step 1, all ε_i's should be set to positive scalars. Suppose, however, $\varepsilon_i = 0$ for all i, in which case $z^* = (3, 3, 3)$. Consider $z^4 \in N$. The only weighting vector for which z^4 minimizes the augmented weighted Tchebycheff program is $\lambda = (0, 1, 0)$. Since all criterion vectors in $\gamma(z^1, z^2)$ minimize the augmented weighted Tchebycheff program for $\lambda = (0, 1, 0)$, z^4 would never be generated because it is not extreme in the augmented program. Thus, if z^4 were optimal in the MOLP, the Tchebycheff algorithm could never converge to it. However, if the ε_i were set to positive scalars as stipulated in Step 1, there would exist a $\lambda \in \bar{\Lambda}$ for which z^4 uniquely minimizes the augmented weighted Tchebycheff program. Hence, the Tchebycheff algorithm could

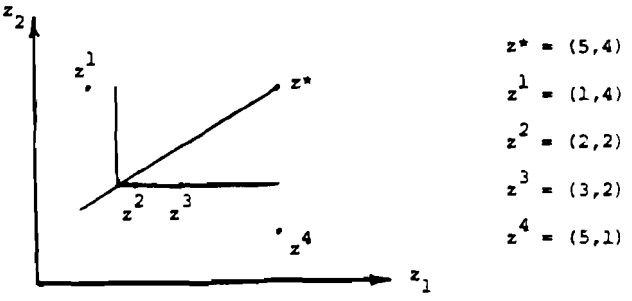
converge to it. When an ε_i is to be positive, it is not necessary to set it to a miniscule value. A value of 1 to 5% of its optimal $f_i(x)$ value should work very nicely.

Example 6.5: What might be the consequence of using the augmented metric with $\rho = 0$?

Solution 6.5: Consider the multiple objective integer program whose $Z = \{z^1, z^2, z^3, z^4\}$ as shown in Figure 6.

Figure 6

Graph of Example 6.5



In this problem $N = \{z^1, z^3, z^4\}$. With $\rho = 0$ and $\lambda = (3/8, 5/8)$, the minimizing $|||z^* - z|||_\infty^\lambda$ isoquant intersecting Z is as drawn. Thus, both z^2 and z^3 minimize the augmented weighted Tchebycheff program, but only z^3 is nondominated. Therefore, the undesirable consequence of ρ being set to 0 is that the augmented program may generate dominated criterion vectors. Note, however, that with a sufficiently small $\rho > 0$, z^3 would be the only criterion vector minimizing the augmented weighted Tchebycheff program. Note that z^3 is an unsupported nondominated criterion vector. Thus we see that the Tchebycheff approach can generate unsupported nondominated criterion vec-

tors such as z^3 as easily as it can generate supported nondominated criterion vectors such as z^1 and z^4.

Example 6.6: Is it possible for the augmented weighted Tchebycheff program (with $\rho > 0$) to generate the same nondominated criterion vector for different λ's?

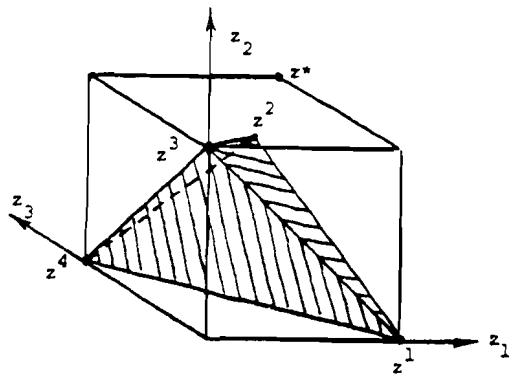
Solution 6.6: Yes. Consider the MOLP

$$\begin{aligned}
\max \quad & \{x_1 = z_1\} \\
\max \quad & \{x_2 = z_2\} \\
\max \quad & \{x_3 = z_3\} \\
\text{s.t.} \quad & x_1 + x_2 + x_3 \geq 4 \\
& -2x_1 + 3x_2 + 3x_3 \leq 12 \\
& 3x_1 - 2x_2 + 3x_3 \leq 12 \\
& 3x_1 + 3x_2 - 2x_3 \leq 12 \\
& x_1, x_2, x_3 \geq 0
\end{aligned}$$

whose Z is as indicated in Figure 7.

Figure 7

Graph of Example 6.6



$z^* = (4,4,4)$

$z^1 = (4,0,0)$

$z^2 = (3,3,3)$

$z^3 = (0,4,0)$

$z^4 = (0,0,4)$

In this problem $N = \gamma(z^1, z^2) \cup \gamma(z^2, z^3) \cup \gamma(z^2, z^4)$ and z^2 is a nondominated criterion vector that can be generated by the augmented program for many weighting vectors. Although $\lambda = (1/3, 1/3, 1/3)$ is the weighting vector for which the $|||z^* - z|||_\infty^\lambda$ isoquant intersects Z at z^2 with its vertex, other weighting vectors such as

$$\lambda = (3/7, 3/7, 1/7)$$

$$\lambda = (1/7, 3/7, 3/7)$$

$$\lambda = (1/2, 0, 1/2)$$

also cause z^2 to minimize the augmented program. Such can happen whenever the nondominated criterion vector in question intersects the minimizing $|||z^* - z|||_\infty^\lambda$ isoquant at places other than its vertex. Thus, the number of different criterion vectors to be filtered in Step 8 of the algorithm may be less than the number of weighted Tchebycheff programs solved in Step 7. Also, if the minimizing $|||z^* - z|||_\infty^\lambda$ isoquant intersects Z with one of its sides, the weighting vector $\lambda^{(h)}$ in Step 11 may be different from the weighting vector that produced $z^{(h)}$ in Step 7.

7. Concluding Remarks

The following are remaining comments about the weighted Tchebycheff interactive procedure:

1. 50 x k weighting vectors are generated in Step 5 in order to create a large enough pool to obtain a good covering of $\bar{\lambda}^{(h)}$ in Step 6.
2. Just because the λ-vectors of Step 6 may be maximally dispersed does not mean there will not be any distortion in the criterion vectors produced in Step 7. Thus, extra weighted Tchebycheff programs

are solved for the purpose of smoothing out the most redundant of the resulting criterion vectors to assure that the p criterion vectors presented to the decision-maker in Step 9 are indeed representative of the neighborhood of N being searched.

3. In Step 15 we may find that there is more than one preimage of the final criterion vector. If such turns out to be the case, it will be necessary to search the region of preimages before the final solution to the problem can be obtained.

References

[1] Bowman, V. J. (1976). "On the Relationship of the Tchebycheff Norm and the Efficient Frontier of Multiple-Criteria Objectives," Lecture Notes in Economics and Mathematical Systems, No. 135, pp. 76-85.

[2] Choo, E.-U. (1980). "Multicriteria Linear Fractional Programming," Ph.D. Dissertation, University of British Columbia.

[3] Choo, E.-U. and D. R. Atkins. (1980). "An Interactive Algorithm for Multicriteria Programming," Computers and Operations Research, No. 1-2, Vol. 7, pp. 81-87.

[4] Ecker, J. G. and N. E. Shoemaker. (1981). "Selecting Subsets from the Set of Nondominated Vectors in Multiple Objective Linear Programming," SIAM Journal on Control and Optimization, Vol. 19, No. 4, pp. 505-515.

[5] Klimberg, R. and R. E. Steuer. (1982). "Managing the Backlog of New Drug Applications at the Bureau of Drugs," in preparation.

[6] Steuer, R. E. and E.-U. Choo. (1981). "An Interactive Weighted Tchebycheff Procedure for Multiple Objective Programming," College of Business Administration, University of Georgia.

[7] Steuer, R. E. and F. W. Harris. (1980). "Intra-Set Point Generation and Filtering in Decision and Criterion Space," Computers and Operations Research, No. 1-2, Vol. 7, pp. 41-53.

[8] Steuer, R. E. and M. Ruwe. (1982). "Zero-One Multiple Objective Programming Applied to Hospital Capital Budgeting," in preparation.

[9] Steuer, R. E. and A. T. Schuler. (1978). "An Interactive Multiple Objective Linear Programming Approach to a Problem in Forest Management," Operations Research, Vol. 26, No. 2, pp. 254-269.

[10] Wierzbicki, A. P. (1980). "The Use of Reference Objectives in Multiobjective Optimization," Lecture Notes in Economics and Mathematical Systems, No. 177, pp. 468-486.

[11] Zeleny, M. (1973). "Compromise Programming," in Cochrane, J. L. and M. Zeleny (eds.), Multiple Criteria Decision Making, University of South Carolina Press, pp. 262-301.

THE REFERENCE POINT OPTIMIZATION APPROACH – METHODS OF EFFICIENT IMPLEMENTATION

A. Lewandowski[1] and M. Grauer[2]

[1] *International Institute for Applied Systems Analysis, Laxenburg, Austria* and *Technical University of Warsaw, Poland*
[2] *International Institute for Applied Systems Analysis, Laxenburg, Austria* and *Technical University of Leuna-Merseburg, GDR*

ABSTRACT

This paper first outlines the reference point optimization approach introduced by Wierzbicki. A software package called DIDASS (Dynamic Interactive Decision Analysis and Support System) which is based on this approach has been developed at IIASA to deal with linear and nonlinear multiple-criteria programming problems - the structure and implementation of DIDASS are discussed.

1. INTRODUCTION

The *reference point* approach introduced by Wierzbicki [1] has already been described in a series of papers and reports. This method is a generalization of the well-known goal programming method [2] and of the method of displaced ideals developed by Zeleny [3]. The basic idea of this method is as follows:

(I) The *decision-maker* (DM) thinks in terms of *aspiration levels*, i.e., he specifies acceptable values for given objectives.

(II) He works with the computer in an *interactive way* so that he can change his aspiration levels during the course of the analysis.

Practical experience with the DM has shown that these requirements are both realistic, which makes the approach very useful in practice. Other methods require the DM to provide rather unnatural information, e.g., the

methods based on the Morgenstern utility theory require the DM to compare the lotteries and to express his preferences in terms of probabilities [4]. It is also unreasonable to expect the DM to carry out a pairwise comparison of several alternatives. The *reference point* approach, in contrast, has proven its applicability in a number of practical problems [5, 6]. This approach has also been used in a study of the optimal structure of the chemical industry [7] and in a work dealing with the generation of efficient energy supply strategies [8].

In the authors' opinion, work on the reference point approach has now reached a stage at which efficient software based on this method can be developed. This paper will concentrate on the software package DIDASS (Dynamic Interactive Decision Analysis and Support System) being developed at IIASA to deal with linear and nonlinear multiple-criteria programming problems. There are several ways of increasing the efficiency of the software, in terms of both computing power and interaction between the user and the computer. Some of these improvements will also be discussed in this paper.

2. REFERENCE POINT OPTIMIZATION

The use of the *reference point* approach in the linear case has been discussed in an earlier paper [5]

Let A be in $R^{m \times n}$, C in $R^{p \times n}$, and b in R^m and consider the multicriteria linear program :

$$Cx = q \rightarrow \max \qquad \text{(MCLP-1)}$$

$$Ax = b \qquad \text{(MCLP-2)}$$

$$x \geq 0 \qquad \text{(MCLP-3)}$$

where the decision problem is to determine an n-vector x of decision variables satisfying $x \geq 0$ while taking into account the p-vector of objectives defined by $Cx = q$. We will assume that each component of q should be as large as possible.

An objective vector value $q = \bar{q}$ is *attainable* if there is a feasible x for which $Cx = \bar{q}$. A point $\bar{q}$ is *strictly Pareto inferior* if there is an attainable point q for which $q > \bar{q}$. If there is an attainable q for which $q \geq \bar{q}$ and the inequality is strict in at least one component, then q is *Pareto inferior*. An attainable point $\bar{q}$ is *weakly Pareto optimal* if it is not strictly Pareto inferior and it is *Pareto optimal* if there is no attainable point q such that $q \geq \bar{q}$ with a strict inequality for at least one component. Thus, a Pareto optimal point is also weakly Pareto optimal, and a weakly Pareto optimal point may be Pareto inferior. For brevity, we shall sometimes refer to a Pareto optimal point as a *Pareto point* and to the set of all such points as the *Pareto set*.

A *reference point* or *reference objective* is a suggestion $\bar{q}$ by the DM which reflects in some sense the "desired level" of the objective. According to Wierzbicki [9], an achievement scalarizing function $s(q - \bar{q})$ defined over the set of objective vectors q may be associated with reference point $\bar{q}$. The general forms of function s which result in Pareto optimal (or weakly Pareto optimal) minimizers of s over the attainable points q is given by Wierzbicki [10].

If we regard the function $s(q - \bar{q})$ as the "distance" between the points q and $\bar{q}$, then, intuitively, the problem of finding such a minimum may be interpreted as the problem of finding from within the Pareto set the point $\hat{q}$ "nearest" to the reference point $\bar{q}$. (However, as will be made clear later, the function s is not necessarily related to the usual notion of distance.) With this interpretation in mind, reference point optimization may be viewed as a way of guiding a sequence $\{\hat{q}^k\}$ of Pareto points generated from a sequence $\{\bar{q}^k\}$ of reference objectives. These sequences are generated in an interactive procedure and this should result in an interesting set of attainable points $\{\hat{q}^k\}$. If the sequence $\{\hat{q}^k\}$ converges, the limit may be seen as the solution to the decision problem.

The decision maker may be provided with initial information by maximizing

all objectives separately. A matrix D_S which yields information on the range of numerical values of the objectives is then constructed. We shall call this the ***decision support matrix***:

$$D_S = \begin{bmatrix} q_1^* & q_2^1 & \cdots & q_p^1 \\ q_1^2 & q_2^* & \cdots & q_p^2 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ q_1^i & q_2^i & \cdots & q_p^i \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ q_1^p & q_2^p & \cdots & q_p^* \end{bmatrix}$$

Row i corresponds to the solution vector x_i which maximizes objective q_i. The vector with elements $q_i^i = q_i^*$, i.e., the diagonal of D_S, represents the ***utopia (ideal) point***. This point is not attainable (if it were, it would be the solution of the proposed decision problem), but it can be used and presented to the decision maker as an upper limit to the sequence $\{\bar{q}^k\}$ of reference objectives . Let us consider column i of the matrix D_S . The maximum value in the column is q_i^*. Let q_n^i be the minimum value, where

$$\min_{1 \le j \le p} \left\{ q_j^i \right\} = q_n^i$$

We shall call this the ***nadir*** value. The vector with elements $q_n^1, q_n^2, \ldots, q_n^p$ represents the ***nadir point***, and can be seen as a lower limit to the values of the decision maker's objectives.

In the following analysis we shall use the notation $w \equiv (q - \bar{q})$. A practical form of the achievement scalarizing function $s(w)$, where minimization results in a linear programming formulation, is then given by:

$$s(w) = - \min \left\{ \rho \min_i w_i \; ; \; \sum_{i=1}^{p} w_i \right\} - \varepsilon w \qquad (1)$$

Here ρ is an arbitrary coefficient which is greater than or equal to p and $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p)$ is a nonnegative vector of parameters. In the special case $\rho = p$, (1) reduces to

$$s(w) = -\rho \min_i w_i - \varepsilon w \tag{2}$$

In our experience, eqn (1) has proven to be the most suitable form of achievement scalarizing function. Other practical forms are given in Wierzbicki [9].

For any scalar $\hat{s}$, the set $S_{\hat{s}}(\bar{q}) = \left\{ q \mid s(w) \geq \hat{s}, w = (q - \bar{q}) \right\}$ is called a level set. Some level sets for function (1) are illustrated in Figure 1 for the cases $\rho = p$, $\rho > p$ and $\rho >> 0$ with $\varepsilon = 0$. In each case, if $w \neq 0$, then $s(w)$ is given by (2), i.e., the functional value is proportional to the worst component of w. If $\rho = p$, the same is also true for $w \geq 0$. If $w > 0$, then for large enough ρ (see the case $\rho \gg p$) $s(w)$ is given by $\sum w_i$. In the general case when $\rho > p$, the situation is as shown in the central part of Figure 1. When $w \geq 0$ and its components are sufficiently close together (that is, $(\rho - 1)w_1 \geq w_2$ and $(\rho - 1)w_2 \geq w_1$ for $p=2$), then $s(w)$ is given by $\sum w_i$. All other cases are represented by eqn. (2).

For $\varepsilon = 0$, scalarizing function (1) guarantees only weak Pareto optimality for its minimizer. However, as will be shown in Lemma 1 below, if $\varepsilon > 0$, Pareto optimality is guaranteed.

The problem of minimizing $s(q - \bar{q})$ as defined by (1) over the attainable points q can be formulated as a linear programming problem, as mentioned above. In particular, making the substitution $w = (q - \bar{q}) = (Cx - \bar{q})$ and introducing an auxiliary decision variable y, this minimization problem may be restated as follows (P):

find y, w, and x to

$$\min \ (y - \varepsilon w) \tag{P-1}$$

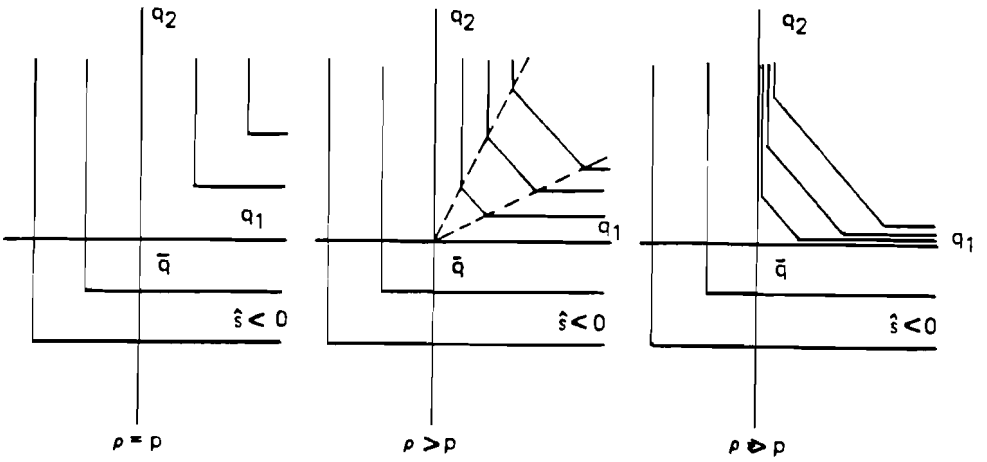


Figure 1 Level sets for achievement scalarizing functions (1) and (2) for $\varepsilon = 0$.

$$E\,y + D\,w \leq 0 \qquad \text{(P-2)}$$

$$-w + C\,x = \bar{q} \qquad \text{(P-3)}$$

$$A\,x = b \qquad \text{(P-4)}$$

$$x \geq 0 \qquad \text{(P-5)}$$

where D and E are appropriate vectors and matrices. Furthermore, $D \leq 0$, and if $w = \bar{w}$ and $y = \bar{y}$ are optimal for (P), then $s = \bar{y} - \varepsilon\,\bar{w}$ is the minimum value attained for the achievement scalarizing function s.

In the detailed formulation of (P) let $W = \left\{ w \mid -w + Cx = \bar{q},\ Ax = b,\ x \geq 0 \right\}$ denote the feasible set for vector w. Then, using the achievement scalarizing function (1), the reference point optimization problem (P) becomes:

$$= \min_{w \in \mathbb{P}} \left\{ -\min \left\{ \rho \min_i w_i \ , \ \sum_i w_i \right\} - \varepsilon w \right\}$$

$$= \min_{w \in \mathbb{P}} \left\{ \max \left\{ \max_i (-\rho w_i) \ , \ -\sum_i w_i \right\} - \varepsilon w \right\}$$

$$= \min_{w \in \mathbb{P}} \left\{ \max \left\{ \max_i (-\rho w_i - \varepsilon w) \ , \ -\sum_i w_i - \varepsilon w \right\} \right\}$$

$$= \min_{\substack{w \in \mathbb{P} \\ z \in R}} \left\{ z \mid z \geq -\rho w_i - \varepsilon w \ ; \text{for all } i ; z \geq -\sum_i w_i - \varepsilon w \right\}$$

$$= \min_{\substack{w \in \mathbb{P} \\ y \in R}} \left\{ y - \varepsilon w \mid -y - \rho w_i \leq 0 \ ; \ \text{for all } i \ ; \ -y - \sum_i w_i \leq 0 \right\} \qquad (3)$$

where we have substituted $y = z + \varepsilon w$.

The optimal solution of this problem is characterized by the following result:

LEMMA 1. *Let* $(y, w, x) = (\hat{y}, \hat{w}, \hat{x})$ *be an optimal solution and let* δ, μ, *and* π *be the corresponding dual vectors related to constraints (P-2), (P-3), and (P-4), respectively. Denote by* $\hat{q} = C\hat{x}$ *the corresponding objective vector, by* $\hat{s} = \hat{y} - \varepsilon\hat{w}$ *the optimal value of the achievement scalarizing function, and by* Q *the attainable set of objective vectors* q. *Then* $\hat{q} \in Q \cap S_{\hat{s}}(\bar{q})$ *and the hyperplane* $H = \left\{ q \mid \mu(\hat{q} - q) = 0 \right\}$ *separates* Q *and* $S_{\hat{s}}(\bar{q})$. *Furthermore*, $\mu \geq \varepsilon$ *and* $q = \hat{q}$ *maximizes* μq *over* $q \in Q$; *i.e.*, $\hat{q}$ *is Pareto optimal if* $\varepsilon > 0$ *and* $\hat{q}$ *is weakly Pareto optimal if* $\varepsilon \geq 0$.

Remark : As illustrated in Figure 2, the hyperplane H approximates the Pareto set in the neighborhood of $\hat{q}$. Thus the dual vector μ may be viewed as a vector of tradeoff coefficients which tells us roughly how much we have to give up in one objective in order to gain a given small amount in another objective.

This Lemma is proved in [5].

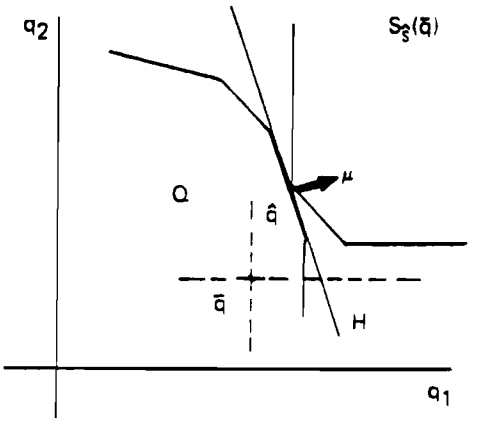


Figure 2 An illustration of Lemma 1.

3. COMPUTER IMPLEMENTATION

3.1 The Basic LP Version

The basic computer LP implementation of DIDASS consists of three parts. These are:

- The interactive "editor" for manipulating the reference point and the objectives (lpmod)
- The preprocessor, which converts the input model file containing the model description in standard MPSX format into its single-criterion equivalent (lpmulti)
- The postprocessor, which extracts the information from the LP system output file, computes the values of the objectives, and displays the necessary

information (lpsol).

This pre- and postprocessing of the LP problems makes the LP (DIDASS) system both flexible and portable. The only machine-dependent point is the format of the output file, which differs between LP packages.

All of the programs work in the interactive mode; however, the efficiency of interaction depends on the size of the LP model. Currently, one session of a 150×100 model on the VAX with the MINOS LP system (see[11]) takes about five to ten minutes CPU time. This makes the interactive analysis of quite nontrivial decision problems possible. The structure of the system and the information flow between components are presented in Figure 3.

3.2 The Extended Version - Approximation of the Pareto Set

Experience with this basic version of the software has shown that it is efficient enough to solve quite complex practical problems. However, this version has one disadvantage - if the DM changes the reference point components it is necessary to solve the LP problem again. For medium-sized LP models this usually takes at least 10 minutes of CPU time. After a brief analysis of the solution, the DM may conclude that the proposed reference point was evidently unacceptable, return to the previous solution and make a new trial. There is a simple way of avoiding such losses of time - instead of calculating a new solution corresponding to the new reference point, this solution could be estimated approximately. If this approximate solution is acceptable to the DM the exact solution can then be calculated.

The procedure for calculating the approximate solution (in the objective space) can be formulated on the basis of Lemma 1. In essence, the hyperplane H separating sets Q and $S_\varepsilon(\bar{q})$ can be used as the local approximation of the Pareto set.

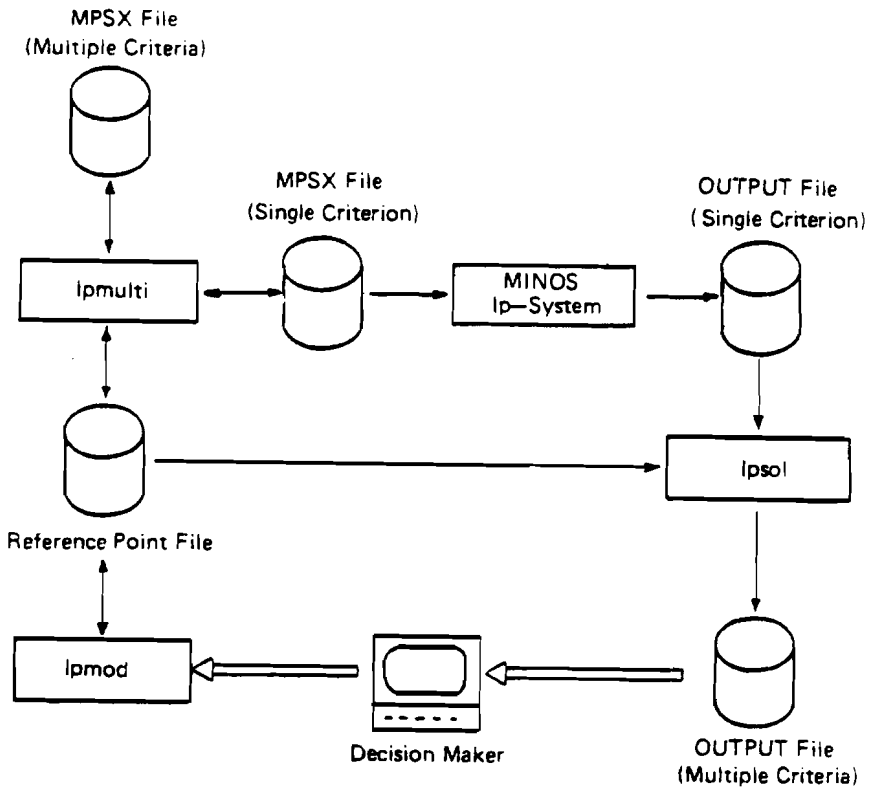


Figure 3 The structure of the multiple-criteria LP package DIDASS.

Let us assume that after the sequence of sessions we have collected the hyperplanes corresponding to each reference point

$$H_i = \left\{ q \mid \mu_i (\hat{q}_i - q) = 0 \right\} \tag{4}$$

The approximate solution can be computed as follows (AP problem):

$$\min_q S(q - \bar{q}) \tag{AP-1}$$

$$\mu_i (\hat{q}_i - q) \leq 0 \tag{AP-2}$$

Repeating the MCLP procedure, we can reformulate the problem described above as a standard LP problem. It should be noted that this problem is much simpler than the original one - the dimensionality of the decision space in this case is equal to the dimensionality of the objective space. Moreover, in view of the special structure of this problem, a simple computational procedure can be formulated; use of the LP algorithm is not necessary.

The simplest version of this algorithm has been implemented by extending the lpmod program of the package so that the DM can obtain approximate values for the objectives immediately after specifying the new reference point This version of the program has been used in [7], experience shows that even such a simplified approach reduces the computational effort significantly

The procedure for calculating the approximate solution for $\rho = p$ (i.e., the scalarizing function takes the form of eqn.(2)) is simple - it is sufficient to project the vector

$$r = \bar{q}_N - \bar{q}_O \tag{5}$$

(where $\bar{q}_N$ is the new reference point and $\bar{q}_O$ is the old reference point) onto the hyperplane

$$\mu (\hat{q}_1 - q) = 0. \tag{6}$$

(see Figure 4).

The result of this projection is

$$\hat{\bar{q}}_N - \bar{q}_O = \left[1 - \frac{w > < \mu}{< \mu, w >} \right] (\bar{q}_N - \bar{q}_O) \tag{7}$$

where $*> <*$ denotes the outer product and $w = \bar{q}_O - \hat{q}_O$.

The solution for $\rho > p$ is more complicated; in this case the standard LP algorithm must be used.

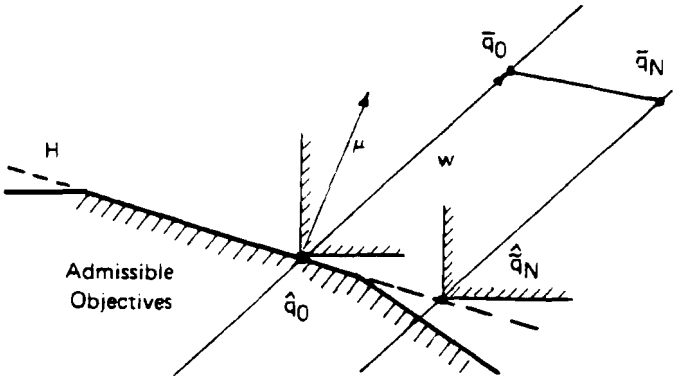


Figure 4 Estimation of the solution corresponding to a new reference point $\bar{q}_n$ starting from an "old" one $\bar{q}_0$.

Some other useful information can be obtained from the above formula, e.g., it is possible to calculate the sensitivity coefficients

$$S_{ij} = \frac{\partial \hat{q}_{Ni}}{\partial \bar{q}_{Nj}} \tag{8}$$

It is also possible to use this equation to solve the AP problem. This may be done in the following steps:

- apply eqn.(7) and calculate q
- calculate the vector $\xi = \bar{q}_N - \hat{\bar{q}}_N$
- find the smallest nonnegative number k such that

$$q = \hat{q}_N + k\,\xi \tag{9}$$

satisfies the set of inequalities (AP.2)

$$\mu_i (\hat{q}_i - q) \leq 0 \qquad (10)$$

However this approximation procedure can sometimes give results that are obviously wrong (see Figure 5).

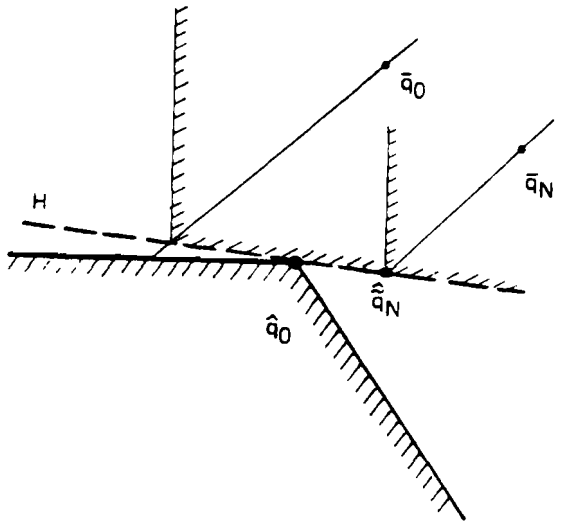


Figure 5 A case in which the hyperplane H gives an inaccurate estimate of the new solution q_N.

To avoid this situation, it is possible to use the convex hull of $\{ \hat{q}_i \}$ to approximate the set of attainable objectives. We could propose algorithms based on this technique, but as yet we do not have much numerical experience with this approach.

The above method is very simple to implement. It is only necessary to extend the lpsol program in order to generate the file containing the history of the session ($\bar{q}$, $\hat{q}$, and μ) and to modify the lpmod program in order to calculate the approximate solution (see Figure 6).

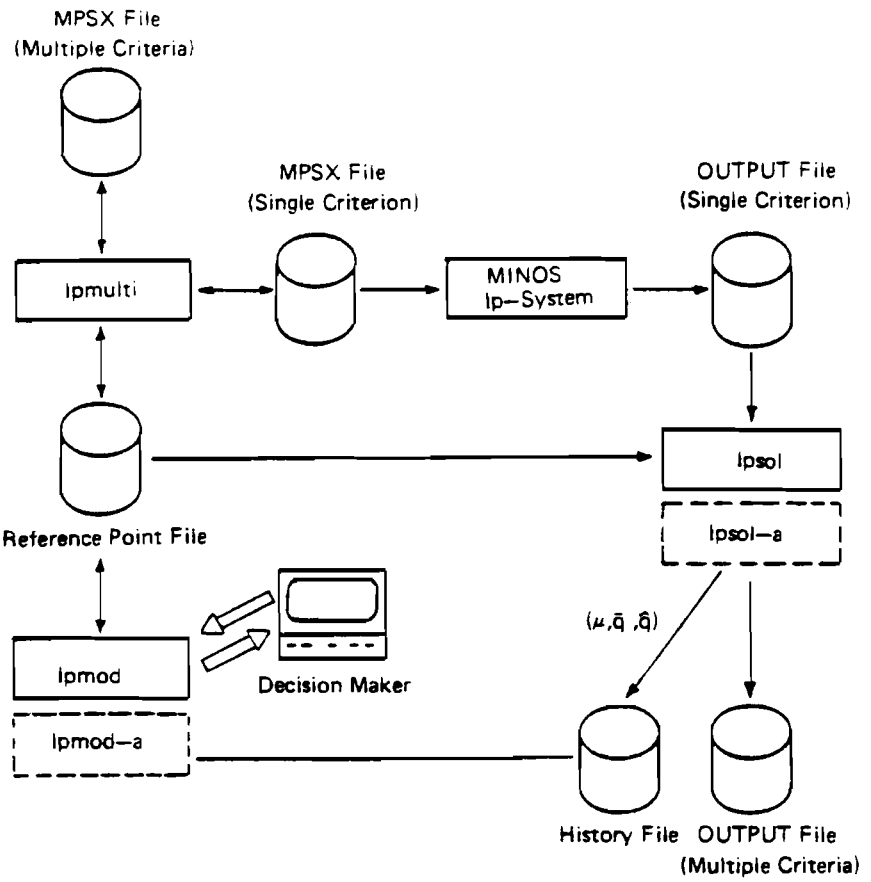


Figure 6 The structure of the extended multiple-criteria LP package with approximation of the Pareto set.

3.3 Parametric Programming Approach

Another useful approach is based on the parametric programming technique. It is easy to see that the reference point appears on the right-hand side of the constrained set in the equivalent LP problem (P). The transition from old to new reference point can therefore be parameterized as follows:

$$\bar{q} = \bar{q}_O + \xi (\bar{q}_N - \bar{q}_O) \quad , \quad 0 \leq \xi \leq 1 \tag{11}$$

Let us assume that we have computed the solution $\hat{q}_O$ corresponding to a given point $\bar{q}_O$. The following procedure is carried out:

- Starting from the basis corresponding to the obtained solution, adjust the parameter ξ in the direction of ξ_1 - the value for which the perturbed problem becomes infeasible
- Calculate the values of the objectives without changing the basis
- Ask the DM whether the direction of change is acceptable. If the answer is yes, calculate the new basis and continue; if no, return to $\bar{q}_O$ and ask the DM to generate a new $\bar{q}_N$.

The basic advantage of this method lies in the fact that if the value of $\bar{q}_N$ is obviously wrong we can interrupt the calculations as soon as possible.

The parametric approach also has another advantage - by changing the parameter ξ from one basis to the other we can simultaneously collect information about approximation hyperplanes. In this way we can obtain a much more detailed approximation of the Pareto set with virtually no additional computational effort.

The basic disadvantage of the method is that it is necessary to have a specially adapted LP system (Figure 7). In many cases when the source code of the existing system is not available it will be impossible to make the necessary changes. However, even in this case the parametric LP algorithm could be used to improve the quality of the local approximation of the Pareto set.

3.4 Incorporating Constraints for Objectives

Some other programs based on the modified reference point approach are being developed and tested. One of these approaches allows the DM to force or "amplify" his preferences using the penalty function technique. In this

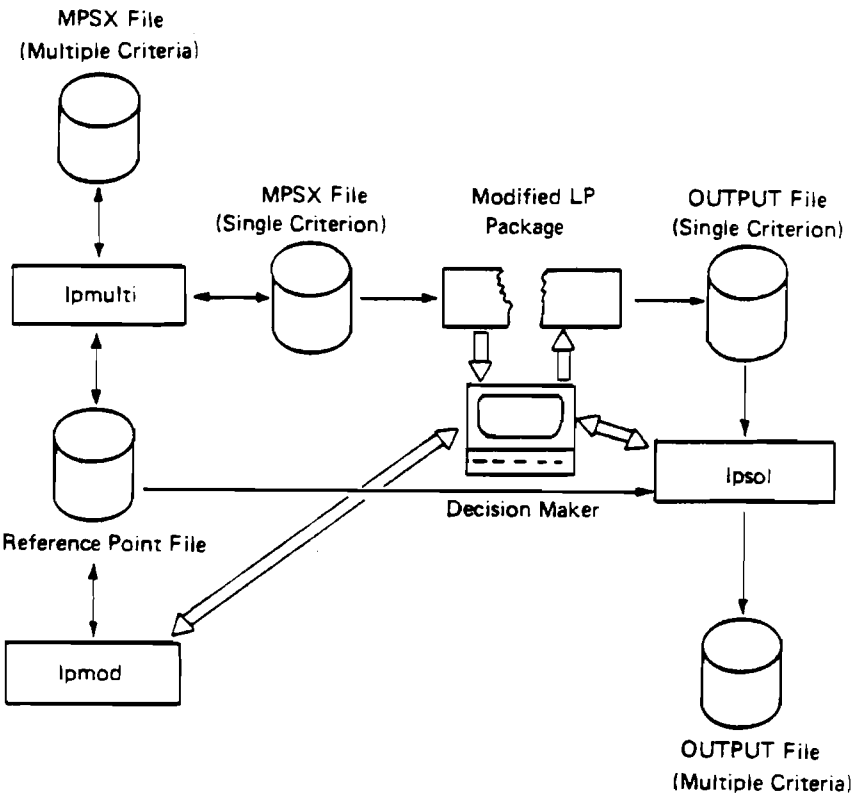


Figure 7 The structure of the multiple-criteria LP package based on the parametric approach.

approach, if the DM wishes to prevent the value of the objective changing in the wrong direction (becoming too large in a case of minimization or too small in a case of maximization), he can add a penalty function to the achievement scalarizing function.

Let J be a set of objectives for which the penalty term has been added. The modified (or nonsymmetric) scalarizing function has the following form (using (2) for simplicity):

$$s(w) = -\rho \min_i w_i - \varepsilon w + \max_{i \in J} (0 , -\rho_i w_i) \tag{12}$$

This problem can be transformed to the equivalent LP problem

$$\min s(w) = \min_{w \in W} \left\{ \max_i (-\rho w_i) - \varepsilon w + \max_{i \in J} (0, -\rho_i w_i) \right\} =$$

$$\min_{\substack{w \in W \\ y, p \in R}} \left\{ y - \varepsilon w + p \mid y \geq -\rho w_i , p \geq -\rho_j w_j , p \geq 0 , j \in J \right\} \tag{13}$$

The coefficients ρ_i in the formula express the "power" of the DM to keep the constraints

$$w_i \geq 0 \tag{14}$$

unviolated. In other cases, it is necessary to introduce two-sided constraints for the selected objectives. This type of problem arises frequently in cases of trajectory optimization when we want to ensure that a certain (reference) trajectory will be traced. In this case the achievement scalarizing function has the following form:

$$s(w) = -\rho \min_i w_i - \varepsilon w + \max_{i \in J} (0, -\rho_i w_i) + \max_{i \in M} (0, -\rho_i w_i) + \max_{i \in M} (0, \rho_i w_i) , \tag{15}$$

where M is the set of objectives for which two-sided constraints have been introduced. Transformation of this function into the equivalent LP problem is straightforward.

Programs based on these concepts have been written and testing has begun; further work on development and testing will be necessary.

3.5 Reference Point Approach With a Partly Nonlinear Objective Function

The LP approach presented in previous sections can be extended to the nonlinear case. If we consider the performance vector as an extesion of (MCLP-1)

$$f(x) + Cx = q \tag{16}$$

the equivalent nonlinear programming problem can be formulated as follows

$$\max_{\substack{w \in W \\ y \in R}} \left\{ y - \varepsilon w \mid y - \rho w_i \leq 0 \,,\; -y - \sum_i w_i \leq 0 \right\} \tag{17}$$

where

$$W = \left\{ w \mid -w + f(x) + Cx = \bar{q} \,,\; Ax = b \,,\; x \geq 0 \right\}. \tag{18}$$

Implementation in this case is quite straightforward – the standard version of the package can be used, the only difference being the need to write a FORTRAN procedure to calculate $f(x)$. The resulting nonlinear programming problem can be solved using the MINOS system or a similar package without any changes in the system.

3.6 The General Nonlinear Version

The basic nonlinear version of DIDASS also uses the idea of pre- and post-processing described in Section 3.1 (see Figure 8). In the nonlinear version of DIDASS, the decision support matrix D_S is calculated in the first step (Utopia) and the information about the utopia point and the nadir point is used to help the DM to choose the reference points. The interactive editor (NLPmod), the preprocessor(NLPmulti) and the postprocessor (NLPsol) operate similar as in the linear case.

The nonlinear constrained multiple-criteria problem to be solved must be expressed in the following standard form:

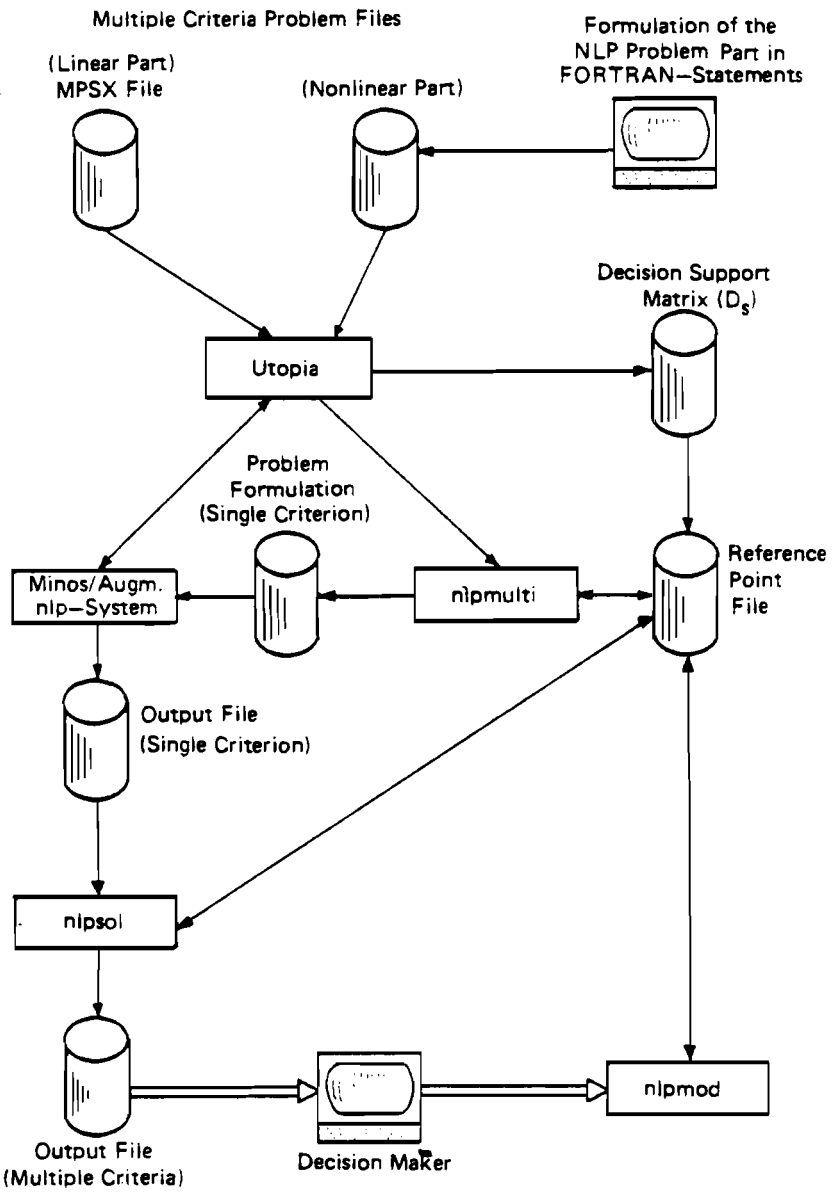


Figure 8 The structure of the nonlinear multiple-criteria package DIDASS.

$$\max_{x_{nl},x_l} \begin{bmatrix} f_1(x_{nl}) + c_1^T x_{nl} + d_1^T x_l = q_1 \\ f_2(x_{nl}) + c_2^T x_{nl} + d_2^T x_l = q_2 \\ \cdots \quad \cdots \quad \cdots \\ f_p(x_{nl}) + c_p^T x_{nl} + d_p^T x_l = q_p \end{bmatrix} \tag{19}$$

subject to

$$g(x_{nl}) + A_1 x_l \leq b_1 \tag{20}$$

$$A_2 x_{nl} + A_3 x_l \leq b_2 \tag{21}$$

$$l \leq \begin{bmatrix} x_{nl} \\ x_l \end{bmatrix} \leq u \tag{22}$$

where $g(x_{nl}) = \left[g_1(x_{nl}), g_2(x_{nl}), \cdots, g_m(x_{nl})\right]^T$ is the vector of nonlinear constraints. The independent variables are divided into two subsets: (x_{nl}) - a vector of "nonlinear" variables and (x_l) - a vector of "linear" variables.

The following two achievement scalarizing functions have undergone preliminary testing with positive results

$$s(w) = -\sum_{i=1}^{p} w_i^2 \tag{23}$$

where $w_i = (q_i - \bar{q}_i)/\bar{q}_i$ and $\bar{q}$ is not attainable and further

$$s(w) = -\frac{1}{\rho} \ln \left[\frac{1}{p} \sum_{i=1}^{p} (w_i)^{\rho} \right] \tag{24}$$

where $w_i = (\tilde{q}_i - q_i)/(\tilde{q}_i - \bar{q}_i)$, and $\tilde{q}_i$ is an upper limit for the sequence of reference points.

However further testing of the numerical features of suitable achievement scalarizing functions for the nonlinear case is necessary.

The nonlinear and linear versions of DIDASS differ in that the user must write FORTRAN statements for the nonlinear parts of the performance criteria $f_1(x_{nl}), f_2(x_{nl}), \ldots, f_p(x_{nl})$ in (19) and the nonlinear parts of the constraints $g(x_{nl})$ in (20) in the nonlinear case. The resulting single-criterion nonlinear pro-

gramming problem obtained using (23) or (24) is solved using the MINOS/AUGMENTED system [11].

4. RELATED PROBLEMS

One of the crucial points in designing interactive multiple-criteria optimization systems is that the interaction between the DM and the computer should be as simple as possible.

A number of important points should be taken into account:

- The DM is usually not a computer specialist, and for this reason the dialogue should be as simple as possible, free of technical details and easy to interpret. In particular, error messages should be self-explanatory. The command language should be as close to the natural language as possible. An interesting outline of this problem can be found in [12], and a more general discussion is given in [13].
- A special effort should be made to present the information in a simple form, preferably graphically. In the simplest case, two-dimensional projections of the Pareto point in the objective space can be very useful [7]; the cuts (or slices) of the Pareto set can give valuable information that is easy to understand.
- Special software must be designed to obtain results from the LP system output file quickly and easily. If the DM is obliged to go through hundreds of pages of computer printout to find the required information, the interaction is not efficient enough. Software systems such as PERUSE [14] can help to overcome this problem.
- Experience with DMs shows that they can usually remember only the results obtained during the last 5 - 10 iterations. In many cases the DM specifies a reference point which has already been specified or which is very close to

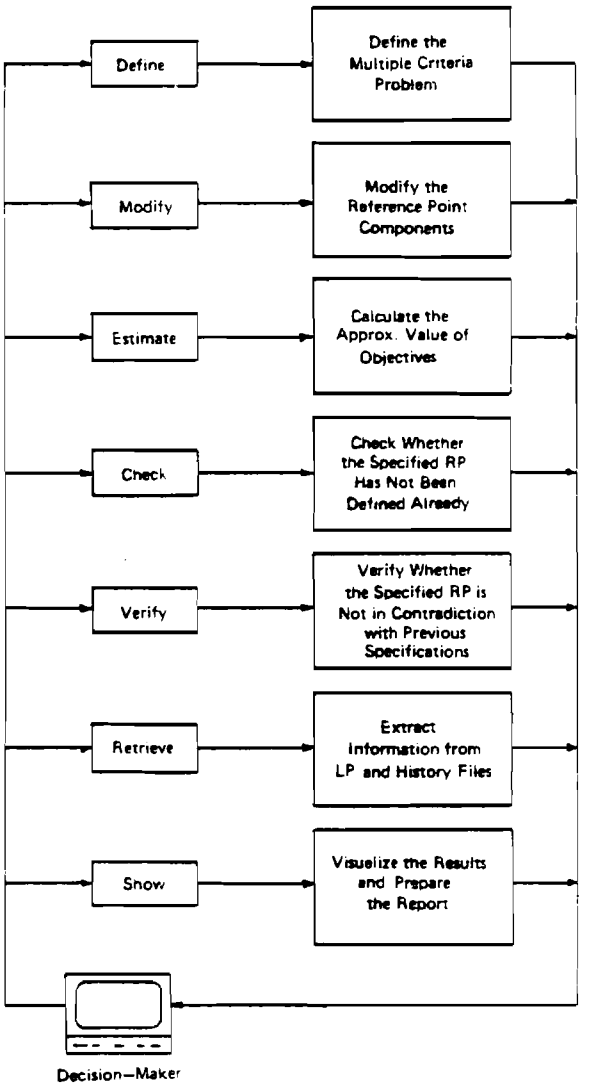


Figure 9 Structure of the DM-computer interface.

one specified in the past, in other cases the DM is not self-consistent and the preferred directions of change contradict those expressed in previous sessions. These situations should be detected and the DM informed.

The general structure of the DM-computer interface is displayed in Figure 9.

It should be pointed out that a number of multiple-criteria packages with a reasonably good interface already exist [15] This paper represents only an initial stage of development of a Decision Support System from an existing Multiple Criteria Optimization package - much work still remains to be done.

References

1. A. Wierzbicki, "A mathematical basis for satisficing decision making ," WP-80-90, IIASA (1980).

2. J.P. Ignizio , "A review of goal programming: a tool for multiobjective analysis," *J. Opl. Res. Soc.* **29**(11) pp. 1109-1119 (1978).

3. M. Zeleny , *Linear multiobjective programming* , Springer-Verlag, Heidelberg, Berlin, New York (1974).

4. R.L. Keeney and A. Sicherman, "An interactive computer program for assessing and analyzing preferences concerning multiple objectives," RM-75-12, IIASA (1975).

5. M. Kallio, A. Lewandowski , and W. Orchard-Hays , "An implementation of the reference point approach for multiobjective optimization.," WP-80-35, IIASA (1980).

6. J.P. Kindler, P. Zielinski, and L. de Mare, "An interactive procedure for multiobjective analysis of water resources allocation," WP-80-35, IIASA (1980).

7. G. Dobrowolski, J. Kopytowski, A. Lewandowski, and M. Zebrowski, "Generating the efficient alternatives for the development process of the chemical industry," CP-82, IIASA (1982).

8. M. Grauer , L. Schrattenholzer, and A. Lewandowski, "Use of the reference level approach for the generation of efficient energy supply strategies," WP-82-19, IIASA (1982).

9. A.P. Wierzbicki, *On the Use of Penalty Functions in Multiobjective Optimization*, Institute of Automatics, Technical University of Warsaw. (1978).

10. A. Wierzbicki, "The use of reference objectives in multiobjective optimization. Theoretical implications and practical experiences.," WP-79-66, IIASA (1979).

11. B.A. Murtagh and M.A Saunders, "Minos/Augmented," Technical Report SOL-80-14, Systems Optimization Laboratory, Stanford University (1980).

12. B Melichar, "Nonprocedural communication between users and application software," RR-81-22, IIASA (1981).

13. R.A. Guedj, "Methodology of Interaction," *Proceedings of the IFIP, Workshop on Methodology of Interaction*, North-Holland Publ. Comp., (1980).

14. W.G. Kurator and R.P. O'Neill, "PERUSE: An interactive system for mathematical programs," *ACM Trans. Math. Softw.* **6**(4) pp. 489-509 (1980).

15 M. Sakawa and F. Seo, "An Interactive Computer Program for Subjective Systems and Its Applications," WP-80-64, IIASA (1980).

A FLEXIBLE DECISION AID METHOD FOR LINEAR MULTICRITERIA SYSTEMS

H.-M. Winkels

Institut für Unternehmensführung und Unternehmensforschung,
Ruhr-Universität Bochum, FRG

Abstract

This paper presents the MULTIPAR method (MULTIPARametric method) for man-machine dialogue in the context of evolutive decision aid based on linear models. The problem consists in determining a satisficing solution or an ordered finite subset of solutions of a system $d^- \leq D \cdot z \leq d^+$, $e^- \leq z \leq e^+$. Therein each subset of variables and/or row values can be considered as a set of criteria such that there is no difference between constraints and objectives.

The method is based on a multicriteria simplex algorithm which incorporates the upper bounding technique. Several multicriteria approaches can now be integrated: Improvement by indication of desired aspiration levels, goals, or substitution rates (with an integration of efficiency projections), restriction of the set of alternatives by global aspiration levels, analysis of the neighbourhood of a goal by visualization or discretization, and implicit determination of criteria weights (by an automatic adaptation to the answers of the user). In the case of half-baked models MULTIPAR can be used for evolutive modelling: There are determined efficient modifications of the model in the case of infeasibility and the set of unlimited criteria is given in the case of feasibility without efficient solutions. For analysing the influcence of data inaccuracy free ranging sensitivity analysis is available.

The enormous flexibility of MULTIPAR makes it easy for a user to adapt himself both to the development of a model and to the manner how decision makers are used to formulate their preferences. To appeal to man's visual perception, the data output is given by means of one-, two-, or three-dimensional computer graphics.

1. Introduction

During the past years numerous methods have been suggested for finding a satisficing solution of a linear multicriteria system (see, e.g., [9], [12], [19]). The trend of these methods was to assume some decision maker willing to answer (simple) questions on his preference structure and to determine a solution by an evolutive and a heuristically or analytically convergent process. There do, however, exist many decision situations where such an assumption is illusive. In fact, we do not know of non-academic decision processes where this assumption happens to be reasonable. According to our experience, managers or decision makers of large companies do not have the time or the willingness to be troubled by the questions of some scientific analyst. They generally do not believe in such algorithms because they simply presume that it is the analyst who would be making the decision instead of themselves. All they are really interested in is to obtain a set of reasonable and well-illustrated decision proposals so that they can make their own choice (according to their own way of reasoning and to their own intuition).

But if this is the sole purpose of decision aid, any suitable method should satisfy two conditions

- the incorporation of interpretations of the decision maker's requirements should be easy
- the method should provide a multitude of information.

On the other hand, such a method does not need to be extremely simple, since only the analyst (and not the decision maker) will be working with the method and he ought to be sufficiently qualified to handle the method by means of a computer. (Obviously by improved techniques of man-machine dialogue in the future such requirements can be reduced.)

In this paper we present the MULTIPAR method (MULTIPARametric method) as a prototyp of such an approach. It is a method for a man-machine dialogue in the context of evolutive decision aid based on linear models. In its underlying philosophy, MULTIPAR is closely related to the MIRACLE method [2], whose purpose, however, is decision aid for general multicriteria systems. Furthermore, it can even be integrated into the discretization module of MIRACLE.

MULTIPAR can be used to determine

- o a satisficing solution or
- o an ordered set of "good" solutions

as well as to answer

- o all structural questions

to a linear model of the following type:

$$\begin{array}{l} d^- \leq D \cdot z \leq d^+ \\ e^- \leq \quad z \quad \leq e^+. \end{array} \qquad \text{(GLS)}$$

Therein each subset of variables and/or row-values can be considered as a set of criteria such that there is no difference between constraints and objectives. The criteria may have the orientation of maximization or minimization or both (i.e. it is desired to get well-balanced values with respect to those criteria). The compact form (GLS) of a linear model is used because the algorithmic part of MULTIPAR is based on a generalized upper-bounding technique of the simplex algorithm in order to enable the implementation of MULTIPAR into advanced LP-packages.

The special feature of the method is its construction as an integrated "drawer system" of different multicriteria approaches in conjunction with visualization techniques. By each of the several options, information can be gained helping the user to explore the set of alternatives and to analyse preferences. The rich global structure of MULTIPAR enables a user to choose that particular multicriteria technique which - in his opinion - is best suited to the special problem or allows him to make various uses of the decision maker's information. Computer graphics are provided as far as possible since this improves the human capacity for perceiving information in complex situations.

Besided this, MULTIPAR is an approach for man-machine dialogue which favours the interactivity with the decision system and takes into consideration the evolutive character both of a model and of the preference structure. To see the differences between the notion of "interactivity" in this paper and its meaning in other papers ([1], [3], [9], [12], [13], [15], [19]), some explanations will be necessary. By the man-machine dialogue we understand the communication between a user and the computer, in which MULTIPAR is implemented. This

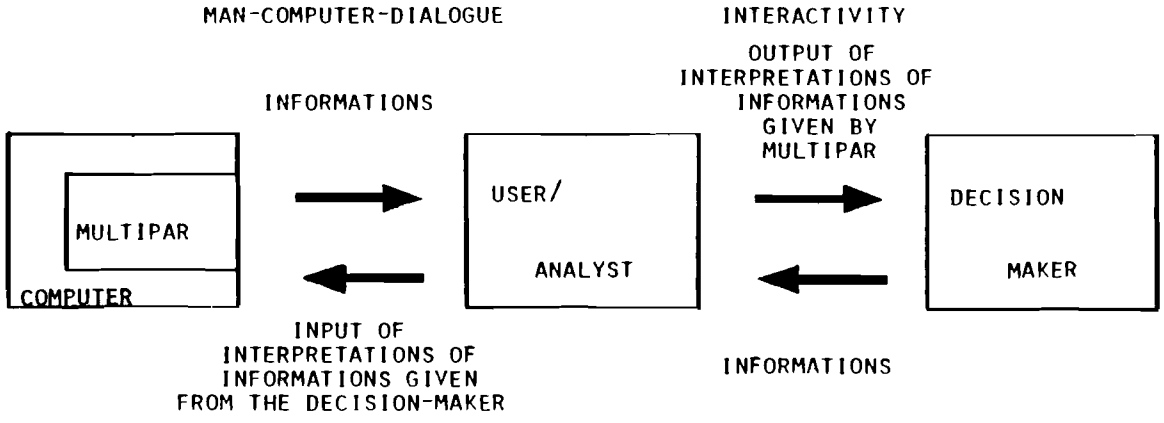


FIG. 1: APPLICATION OF MULTIPAR FOR THE CASE THAT THE USER IS NOT THE DECISION MAKER BUT THE ANALYST

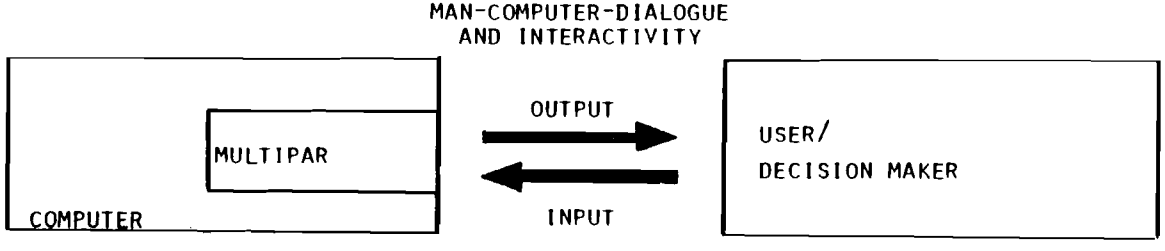


FIG. 2: APPLICATION OF MULTIPAR FOR THE CASE THAT THE USER IS THE DECISION MAKER

man-machine relation should not be confused with the interactivity, established in the framework of a decision process, between the user of MULTIPAR (the analyst) on one side and the decision maker or the decision system on the other side. As we said it above, MULTIPAR is conceived for a user who is accustomed to multicriteria approaches and linear models. In contrast to most other methods, it is principally constructed for an analyst who acts in a decision process as some kind of "human compiler", i.e. who interprets or models the informations or the requirements given from a decision maker and who determines decision proposals or provides information by means of the man-machine dialogue (see Figure 1). Hence this function corresponds exactly to the role operational research departments are playing in most companies.

Obviously, MULTIPAR could also be used by a decision maker himself after he has made himself familiar with the method, for example by means of a user's guide. But we judge such a case to be rather unrealistic (see Figure 2).

A further difference to other methods is the fact that by MULTIPAR the user is not strongly guided. On the contrary, MULTIPAR is a flexible method which through the medium of the decision analyst gives answers to the various questions a decision maker may have in order to get a better understanding of his problem. By this flexibility, the process of modelling can be integrated into the process of problem solving. Hence it is possible:

- to treat complete, non-evolutive models
- to treat a model which has voluntarily been left incomplete in order to allow an evolutive definition of criteria or to get a reduced solution set
- to start with a provisional model which will be successively improved during the interactive decision process.

Summarizing, the function of MULTIPAR within a decision process can be illustrated by means of Figure 3. The upper algorithmic plane symbolizes MULTIPAR. The method consists of the two parts: modelling and analysis of the solution set. The last part, however, allows always feed-back for model modifications. Information on preferences is automatically analysed with respect to criteria weights. On the right side, there is a module called HELP, which explains the subroutines of the algorithm and which indicates the means of their ac-

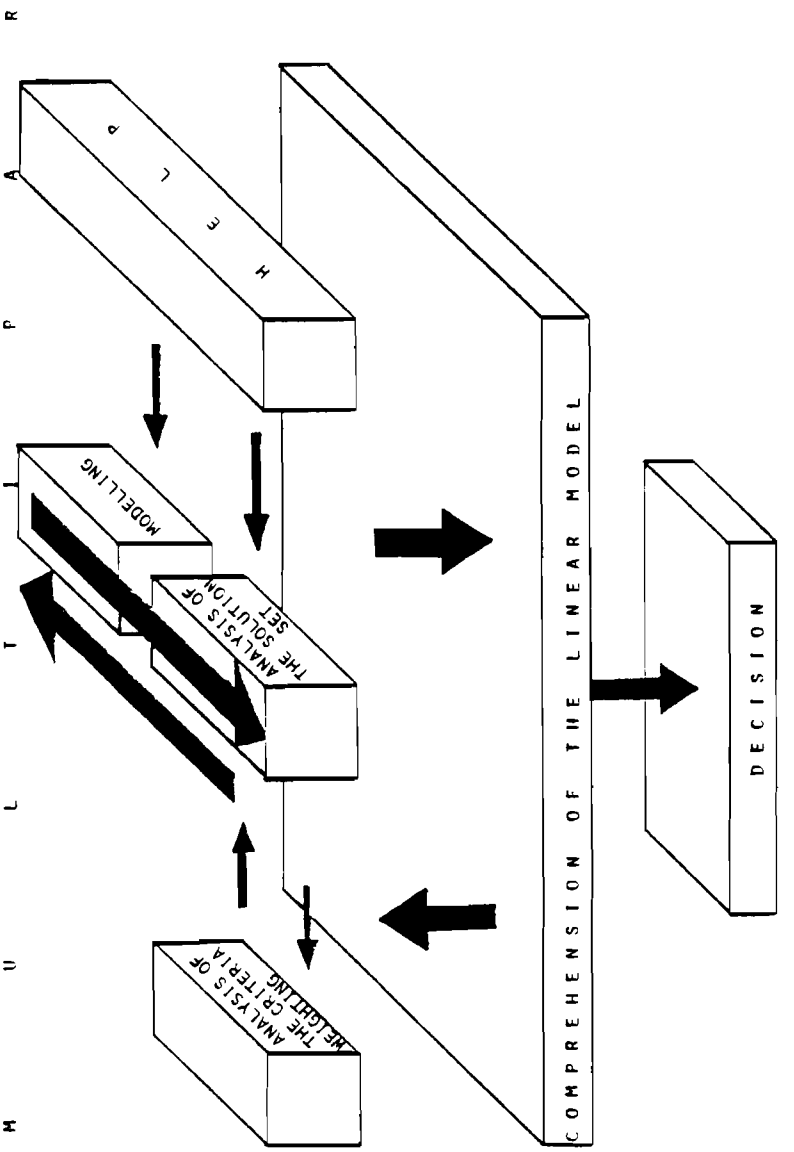


FIG. 3: INTEGRATION OF MULTIPAR INTO THE DECISION PROCESS

tivation on the screen. The use of the method leads to a better comprehension of the linear model. On the other hand, this understanding naturally influences the way of using the method or perhaps leads to a modification of the model. This evolutive process is terminated by a decision.

Notation:

Throughout this paper an (M×N)-matrix A is denoted by $A = ({}_mA_n)_{(m,n)\in M\times N}$, where M and N stand for finite, non-empty index sets. The (I×J)-submatrix of A is denoted by ${}_IA_J$, ${}_IA := {}_IA_N$, and $A_J := {}_MA_J$. The J-subtuple of the N-tuple x is denoted by x_J. |I| is the cardinality of the set I.

2. The Linear Model

The basis for the application of MULTIPAR is a general linear model of the type

$$\begin{array}{l} d^- \leqq D\cdot z \leqq d^+ \\ e^- \leqq z \leqq e^+, \end{array} \qquad \text{(GLS)}$$

where D denotes an (R×S)-matrix, d_i^-, $e_j^-\in\mathbb{R}\cup\{-\infty\}$ and d_i^+, $e_j^+\in\mathbb{R}\cup\{+\infty\}$ for $i\in R$, $j\in S$. (R and S are non-empty finite index sets, for simplicity suppose $R\cap S = \emptyset$.)

By $Z := \{z\in\mathbb{R}^S : d^- \leqq D\cdot z \leqq d^+,\ e^- \leqq z \leqq e^+\}$ we denote the solution set of (GLS).

To simplify the analysis of Z it is reasonable to restrict the consideration of solutions to the consideration of certain components, which will be called criteria:

$$\begin{array}{rl} {}_iD\cdot z \rightarrow \text{sat}_i! & : i\in R_c \\ z_j \rightarrow \text{sat}_j! & : j\in S_c \end{array} \qquad \text{(CGLS)}$$

(with $R_c \subseteq R$ and $S_c \subseteq S$).

Therein, each of these criteria has an orientation $\text{sat}_i\in\{\text{max},\text{min},\text{bal}\}$, where "max" means that the values of ${}_iD\cdot z$ (i-th row of $D\cdot z$) and z_i, respectively, should be as large as possible, "min" means the opposite, and "bal" means well balanced, i.e. not too large and not too small. (The criteria ${}_iD\cdot z \rightarrow \text{bal}!$ and $z_j \rightarrow \text{bal}!$ can be used instead of the two criteria ${}_iD\cdot z \rightarrow \text{max}$, ${}_iD\cdot z \rightarrow \text{min}$ and $z_j \rightarrow \text{max}!$, $z_j \rightarrow \text{min}!$, respectively.)

(GLS) combined with (CGLS) is a general linear multicriteria system, which extends the common notion of a vector maximum system is a natural way.

Since it is important to consider the row-values $D \cdot z$ as well as the solutions z, we extend the solution set Z to the set

$$\bar{Z} := \{(z, D \cdot z) : z \in Z\}.$$

A linear restriction system representing $\bar{Z}$ is given by

$$\begin{aligned} (-E|D) \cdot \bar{z} &= 0 \\ d^- \leqq \bar{z}_R &\leqq d^+ \\ e^- \leqq \bar{z}_S &\leqq e^+ \end{aligned} \qquad \text{(ELS)}$$

(E : $(R \times R)$-identity matrix) with the transformation $\bar{z}_R := D \cdot z$ and $\bar{z}_S := z$. Then (CGLS) becomes:

$$\bar{z}_j \to sat_j \quad \forall j \in R_c \cup S_c \qquad \text{(CELS)}$$

Finally, in order to get a normalized linear model which fits the simplex algorithm best, row- and column-scaling as well as codings of the index sets lead to the system

$$\begin{aligned} A \cdot x &= a \\ \Theta \leqq x &\leqq b \end{aligned} \qquad \text{(NLS)}$$

and to the vectorial criteria orientation

$$x_L \to Sat! \qquad \text{(CNLS)}$$

Here, M and N denote finite and disjoint index sets of natural numbers, $L \subseteq N$, A is an $(M \times N)$-matrix with rank $A = |M|$, $a \in \mathbb{R}^M$; $\Theta \in \overline{\mathbb{R}}^N$, $b \in \overline{\mathbb{R}}^N$ ($\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$), $Sat = (sat_l)_{l \in L}$, and the bound conditions can be written as

$$\begin{aligned} 0 = x_i &= 0 && \forall i \in N_0 : \text{"type 0"} \\ 0 \leqq x_i &\leqq b_i > 0 && \forall i \in N_1 : \text{"type 1"} \\ -\infty \leqq x_i &\leqq b_i && \forall i \in N_2 : \text{"type 2"} \\ 0 \leqq x_i &\leqq +\infty && \forall i \in N_3 : \text{"type 3"} \\ -\infty \leqq x_i &\leqq +\infty && \forall i \in N_4 : \text{"type 4"} \end{aligned}$$

with $N = \bigcup_{i=0}^{4} N_i$.

(NLS) is the linear system, all the computational operations of MULTIPAR are based on. The solution set X of (NLS) is isomorph to $\bar{Z}$ with respect to geometrical and topological aspects. We do not want to describe this isomorphism, here (see [25]). But note, that it is always necessary to translate the computational results (of NLS) into a data language (of (GLS)) (see Figure 4).

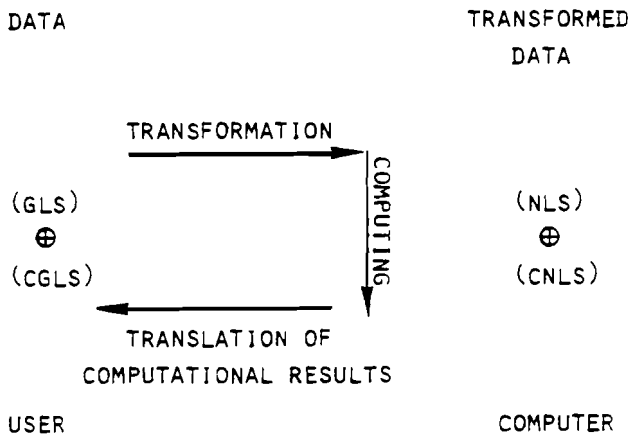


FIG. 4: PRINCIPLE OF DATA TRANSFORMATION AND DATA TRANSLATION

The use of (NLS) is due to the fact, that such a system is the standard system of advanced LP-packages, and not a linear system of the classical type

$$A \cdot x = a$$
$$x \geq 0$$

which has merely didactical advantages. Note that we do not use the classical variable types of Orchard-Hays [14], who has variables of types 0, 1, 3, and 4, only. Obviously, a type 2 variable could be transformed into a type 3 variable. But the explicit treatment of such type 2 variables has considerable advantages if one is interested in reducing the model in so far that redundant constraints have been discovered. (This generalized upper-bounding technique has been developed in [25]).

The analysis of the whole solution set X (and thus of $\bar{Z}$) would be rather complex if we do not restrict our consideration to the efficient solutions. In the context of general multicriteria systems a feasible solution x of (NLS) is said to be efficient with respect to (CNLS), if there is no solution $x' \in X$ with $x' \neq x$ such that

$$x_l \leq x_l' \quad \forall l \in L_{max} := \{l \in L \text{ with } sat_l = max\}$$
$$x_l \geq x_l' \quad \forall l \in L_{min} := \{l \in L \text{ with } sat_l = min\}$$
$$x_l = x_l' \quad \forall l \in L_{bal} := \{l \in L \text{ with } sat_l = bal\}.$$

This is nothing else but the natural extension of the notion of efficiency with respect to vector maximum systems. (x_l - bal! is considered as the double criterion x_l - max! and x_l - min!) An analogous definition could be given for solutions of (GLS) with respect to (CGLS).

3. The Multiparametric Auxiliary System

The name of our method (MULTIPAR) was influenced by the following multiparametric linear model, which constitues the core of the algorithm:

$$x_o - \sum_{l \in L} w_l \cdot (u_l - v_l) = 0$$

$$\begin{aligned} x_l - u_l + v_l &= g_l \quad \forall l \in L_{max} \\ x_l - u_l - v_l &= g_l \quad \forall l \in L_{min} \\ x_l - u_l + v_l &= g_l \quad \forall l \in L_{bal} \end{aligned} \qquad (\bar{N}(g,w))$$

$$A \cdot x = a$$

$$\theta \leq x \leq b, \; u \geq 0, \; v \geq 0, \; -\infty \leq x_o \leq +\infty$$
$$u \cdot v = 0$$

where

- x_o denotes a global objective variable
- $g \in \mathbb{R}^L$ denotes a vector parameter of goals
- $w \in \mathbb{R}^L$ denotes a vector parameter of weights : $w_l \geq 0 \quad \forall l \in L_{max} \cup L_{min}$
- $u \in \mathbb{R}^L$ denotes a distance variable of surplus
- $v \in \mathbb{R}^L$ denotes a distance variable of short falling.

This multiparametric system has some significant properties. Let $w^* \geq 0$ be some value of the weighting parameter w.

(3.1) If (x_o, x^*, u, v) minimizes x_o with respect to $(\bar{N}(g^*, w^*))$ for some goal g^*, then x^* is a feasible solution of (NLS) which minimizes the "distance" to g^* with respect to: $dist(x,x') := \sum_{l \in L} w_l^* \cdot |x_l - x_l'|$.

Let $(\bar{N}_o(g,w))$ denote the multiparametric linear system which is derived from $(\bar{N}(g,w))$ by fixing $v := 0$, $u_l := 0$ $\forall l \in L_{bal}$ and adding the objective $x_o \to \max!$:

$$
\begin{aligned}
& x_o \to \max! \\
x_o - & \sum_{l \in L_{max} \cup L_{min}} w_l \cdot u_l = 0 \\
& x_l - u_l = g_l \quad \forall l \in L_{max} \\
& x_l + u_l = g_l \quad \forall l \in L_{min} \\
& x_l \quad\; = g_l \quad \forall l \in L_{bal} \qquad\qquad (\bar{N}_o(g,w)) \\
& A \cdot x \;\; = a \\
-\infty \leq x_o \leq +\infty,\ & \ominus \leq x \leq b;\ u_l \geq 0 \ \forall l \in L_{max} \cup L_{min} \\
& v = 0;\ u_l = 0 \ \forall l \in L_{bal}\ .
\end{aligned}
$$

Furthermore let w^* denote a fixed value of the weighting parameter w with $w_l^* > 0$ $\forall l \in L_{max} \cup L_{min}$. Then we have the following statements with respect to $(\bar{N}_o(g,w^*))$:

(3.2) x^* is an efficient solution (of (NLS) with respect to (CNLS)) if and only if the solution $(x_o,x,u,v) = (0,x^*,0,0)$ is optimal with respect to $(\bar{N}_o(x_L^*,w^*))$.

(3.3) If (x_o^*,x^*,u^*,v^*) is an optimum of $(\bar{N}_o(g^*,w^*))$ for some goal g^*, then x^* is efficient with

$$
\begin{aligned}
x_l^* &\geq g_l^* \quad \forall l \in L_{max} \\
x_l^* &\leq g_l \quad \forall l \in L_{min} \\
x_l^* &= g_l \quad \forall l \in L_{bal}.
\end{aligned}
$$

(3.4) x^* is efficient if and only if there exists some goal $g^* \in \mathbb{R}^L$ such that (x_o^*,x^*,u^*,v^*) with

$$
\begin{aligned}
x_o^* &= w^* \cdot u^* \\
u_l^* &= x_l^* - g_l^* \quad \forall l \in L_{max} \\
u_l^* &= g_l^* - x_l^* \quad \forall l \in L_{min} \\
u_l^* &= 0 \qquad\;\; \forall l \in L_{bal} \\
v_l^* &= 0 \qquad\;\; \forall l \in L
\end{aligned}
$$

is an optimum of $(\bar{N}_o(g^*,w^*))$.

(3.5) If x^o is a feasible solution such that x_o has no maximum (i.e. x_o is unbounded from above) with respect to $(\bar{N}_o(x_L^o,w^*))$, then there are no efficient solutions of (NLS) with respect to (CNLS).

The first of the properties above is well known, the others can be proven analogously to [20].
These statements constitute the key to perform efficiency analysis of (NLS) with respect to (CNLS) by means of mono- and multiparametric programming techniques. (The most advanced text book of multiparametric programming is Gal [8], for extending his methods to upper-bounding technique see [25].) Some important results are listed below, for more details and proofs, the reader is referred to [20].

(3.6) If a basis of $(\bar{N}_o(g,w^*))$ is given which is dual feasible and primal feasible for at least one value g^* of the goal parameter g, then for each g of the critical parameter domain the corresponding basic feasible solution is efficient. It maximizes x_o, i.e. the weighted sum of the surplus variables with respect to the aspiration levels given by g (see (3.3)).

(3.7) If a basis of $(\bar{N}_o(g,w^*))$ is given which is dual feasible and primal feasible for at least one value g^* of the goal parameter g, then this basis induces an efficient face of (NLS). To get this face, for each non-basic variable x_j of the types 1, 2, or 3 do the following: "Fix x_j in (NLS) on its upper bound, if x_j is in the basis on its upper bound, otherwise on its lower bound."
By means of multiparametric programming technique it is possible to determine a covering of the complete critical parameter region. Hence we can get

$$\mathrm{Eff}_L(X) := \{x_L \in \mathbb{R}^L : x \text{ is an efficient solution}\},$$

the set of efficient criteria profiles, and for each of these criteria profiles a corresponding efficient solution by taking the union of the maximal faces, calculated as described above. Note that each of these faces is completely determined by the index set of its relevant bounds, which is given by the basis.

(3.8) A feasible solution of $(N_o(g,w^*))$ (i.e. a solution which is feasible with respect to some value g^* of the goal parameter g) can directly be derived from a feasible solution of (NLS) (and vice versa). Hence, if we discover that (NLS) has no feasible solution and if we see which of the constraints is too strong we can relax some bounds in (GLS) in order to guarantee feasibility. By adding new criteria to (CGLS) we can get a model which is as "close" as possible to the original one. In terms of [20],[21], we change to an analysis of efficient modifications of the model.

(3.9) A further indicator of an ill-structured model would be the case of feasibility without efficient solutions, i.e. the values of some (unlimited) criteria can jointly be improved up to infinity without reducing the values of the other criteria. The maximal set of such unlimited criteria can be determined by means of (3.5). This would be a useful information for a necessary modification of the original model ([20], [21]).

4. Efficiency Projections

A crucial subroutine of MULTIPAR consists in projecting the line segment between two feasible points x_L, $y_L \in X_L := \{x_L : x \in X\}$ on the set Eff(X) of efficient solutions. This can be done by applying (3.6) to the one-parameter system $(\tilde{N}_o(x_L + \lambda \cdot (y_L - x_L), w^*))$ with $0 \leq \lambda \leq 1$. Figure 5 illustrates the function of this subroutine PROJECT (with respect to a vector maximum system).

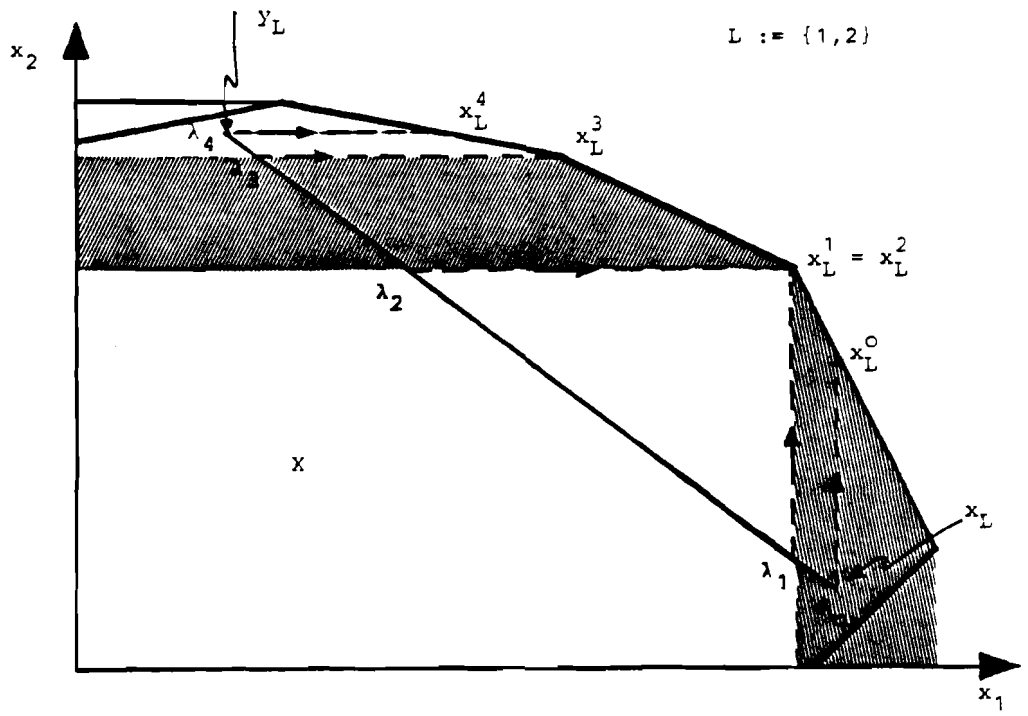


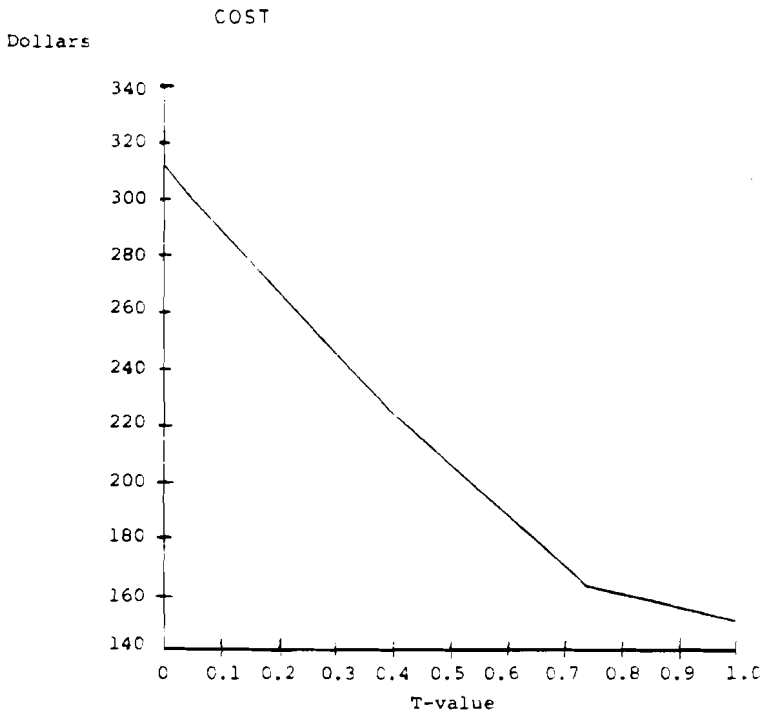
FIG. 5: PRINCIPLE OF THE SUBROUTINE PROJECT.

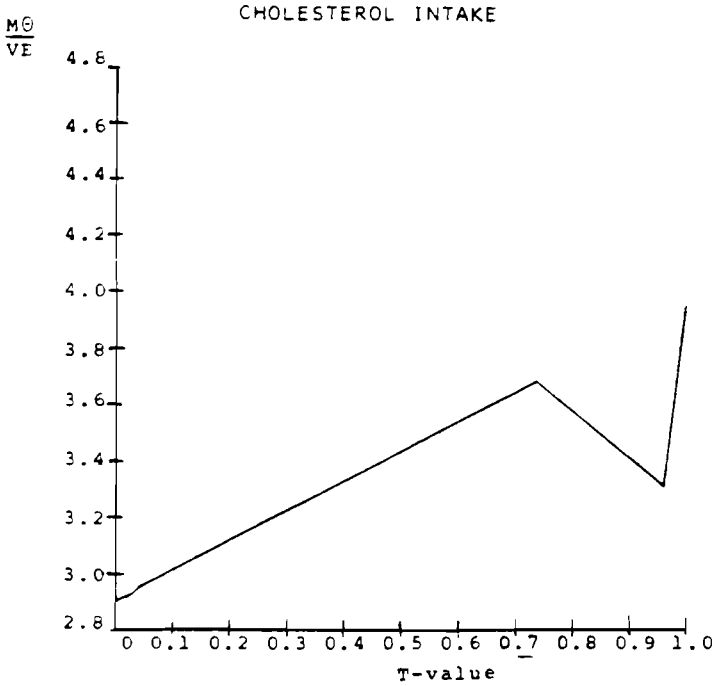
From x_L to y_L one passes through the index bases $I^0,\ldots,I^3$. Basis changes have to be performed for the parameter values λ_1, λ_2 and λ_3. The shaded regions indicate the critical regions belonging to the index bases. To each point of these regions a corresponding efficient solution is assigned (in the sense of (3.6)). The different arrows represent the different directions of the projections, depending on the critical region. Note that all goals in the region belonging to I^1 are projected on x_L^1.

Using this technique, we get a mapping

$$\text{proj} : \leqq 0,1 \geqq \to \bar{Z},$$

($\leqq 0,1 \geqq := \{\lambda \in \mathbb{R} : 0 \leqq \lambda \leqq 1\}$) such that proj($\lambda$) corresponds to the efficient solution, the point $x_L + \lambda \cdot (y_L - x_L)$ is projected on. This mapping is continuous and piecewise linear. A compact representation of this effiency projection is given by its graph with respect to interesting components, indicated by the user (usually the criteria). Figure 6 is an example for such a graphical computer printout. A rather nice representation is obtained by using different colours in a single figure. For some more details on efficiency projections the reader is referred to [22] and [26].





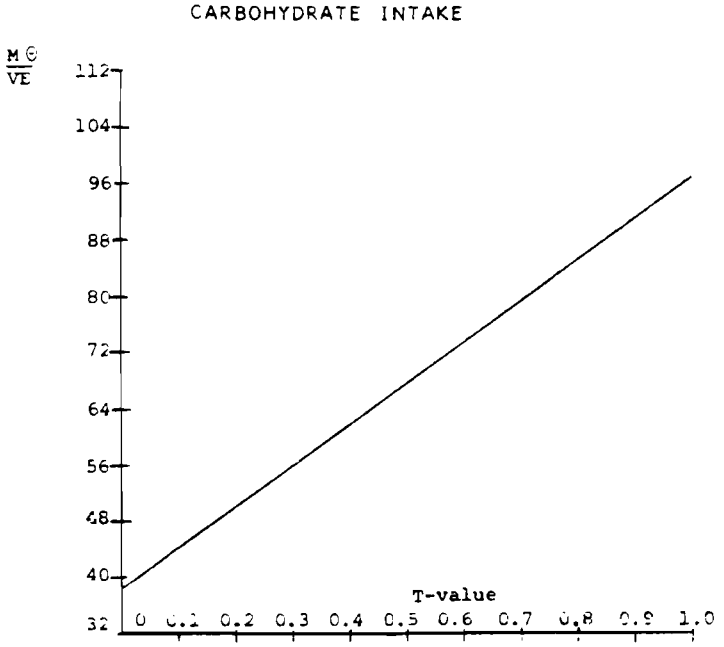


FIG. 6: EXAMPLE OF AN EFFICIENCY PROJECTION

5. Discretization and Visualization

Two other complex important subroutines are DISCRET and VISUAL, on which sensitivity analysis and local efficiency analysis of MULTIPAR are based.

(5.1) DISCRET is an algorithm for determining a finite, well distributed subset of some polyhedron, say X (to simply the illustration we use the solution set of (NLS)). The idea of discretizing a convex polyhedral set is quite natural: We merely have to define a lattice covering X and to determine all points of the lattice which meet X (see Figure 7).

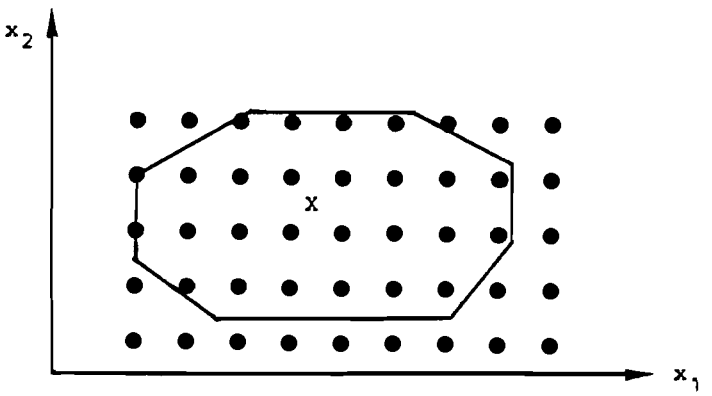


FIG. 7: THE NATURAL IDEA OF DISCRETIZATION

However, in the case of dependent variables, this procedure will not work, as we can see from Figure 8.

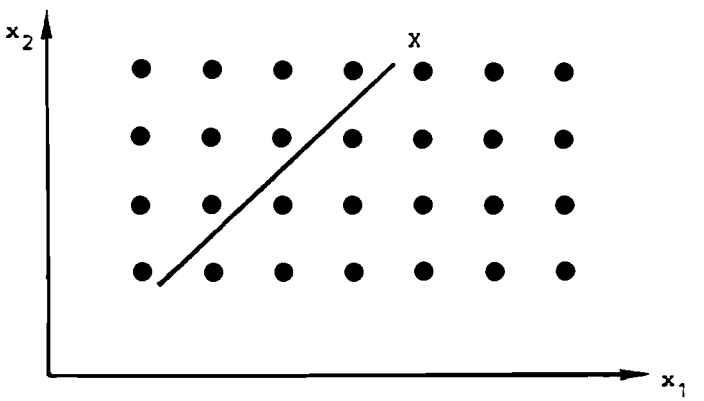


FIG. 8: DIFFICULTIES IN THE CASE OF DEPENDENT VARIABLES

We first have to explore the "hidden dependencies" between the variables and then to transfer the discretization to a "full-dimensional" subpolyhedron. This can be done by determining the slackless constraints of (NLS), i.e. all bounds b_i and θ_j such that for all $x \in X$ we have $x_i = b_i$ and $x_j = \theta_j$, respectively. Collecting all these equations we get a linear system

$$\begin{aligned} A \cdot x &= a \\ x_j &= b_j \quad \forall j \in equ^+(N) \\ x_i &= \theta_i \quad \forall j \in equ^-(N) \end{aligned} \qquad (NLS^{=})$$

whose solution set is the affine hull of X ([17], [18]).

Now, using the GAUSS-elimination procedure, we can derive a new system

$$x_{N \setminus U} = -D^{\Delta} \cdot x_U + d^{\Delta} \qquad (NLS^{=}_U)$$

($U \subseteq N$; D^{Δ} : $((N \setminus U) \times U)$-matrix, d^{Δ}:$(N \setminus U)$-tuple) which is equivalent to $(NLS^{=})$. This equation is called dependency equation (with respect to U) and x_i : $i \in U$ is said to be a system of independent variables.

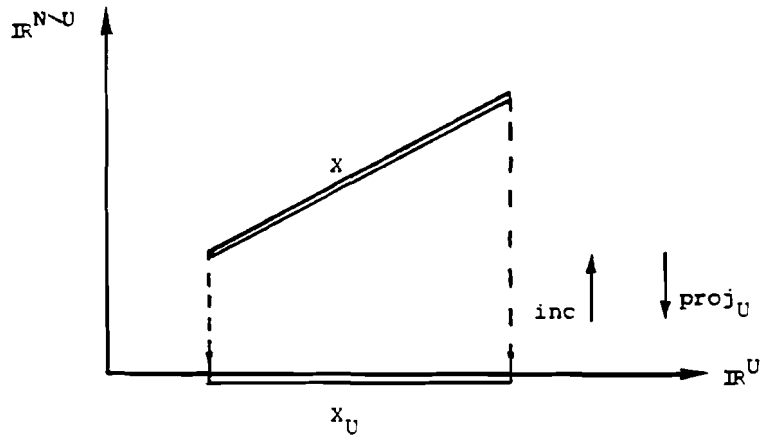


FIG. 9: PRINCIPLE OF THE RESOLUTION PROJECTION

Let $X_U := \{x_U : x \in X\}$ denote the projection of X on the independent variables. Then X_U and X are isomorph with respect to the convex, conical, affine linear and topological structure. Further, dim $X_U = |U|$, and X_U has interior points with respect to the natural euclidean topology on $\mathbb{R}^U$. Hence X_U has the desired properties for the discretiza-

tion procedure: We eliminate the dependent variables x_j : $j \in N \setminus U$ (and perhaps some hidden redundancy) in (NLS) and get a linear restriction system

$$\begin{array}{l} a^- \leqq A^{\Delta} \cdot x_U \leqq a^+ \\ b^- \leqq \quad x_U \leqq b^+ \end{array} \qquad (NLS_U)$$

of which the solution set is X_U. ($(NLS_U) \Leftrightarrow (NLS_U^=)$ is equivalent to (NLS).) Then we define for each independent variable a finite subset of values (usually integers between the maximum and the minimum of x_i on X_U) and take the cartesian product of these component sets as a lattice. Finally, the discretization of X is given by all points $(x_U, -D^{\Delta} \cdot x_U + d^{\Delta})$ such that x_U is an element of the lattice which satisfies (NLS_U).
Since it can be rather time consuming to check all points of the lattice with respect to (NLS_U) and since such an algorithm can produce too many points, it is reasonable to apply the "simulation trick": For each independent variable define a (usually uniform) distribution on its finite value set and check a certain number of lattice points x_U, derived by a simulation on the lattice with respect to the product distribution. For more details on discretizations, see [23], [25].

(5.2) VISUAL is an algorithm for determining 1-, 2-, or 3-dimensional figures of a convex polyhedron (say X, again), in order to understand its high dimensional structure. In the case of dim X > 3, we necessarily have to use projections of X. Now, if we fix the values of all but three of the variables on some feasible value and if we let the three variables vary, in the case of hidden dependencies between the variables it can happen that the so found projection is merely a singleton. The next problem arises with "neighbouring" projections: Suppose that we have a projection in the sense mentioned above but would like to change the projection parameters a little in order to see what happens for projections near by. If we apply our "usual sense topology", i.e. going a little bit to the right or to the left in some components of our feasible reference solution, there is the danger of "dropping from the manifold", generated by X.
As we see, for visualizing X we first have to determine a system x_i : $i \in U$ of independent variables and a corresponding resolution projection X_U. Then we can define projection parameters by fixing the

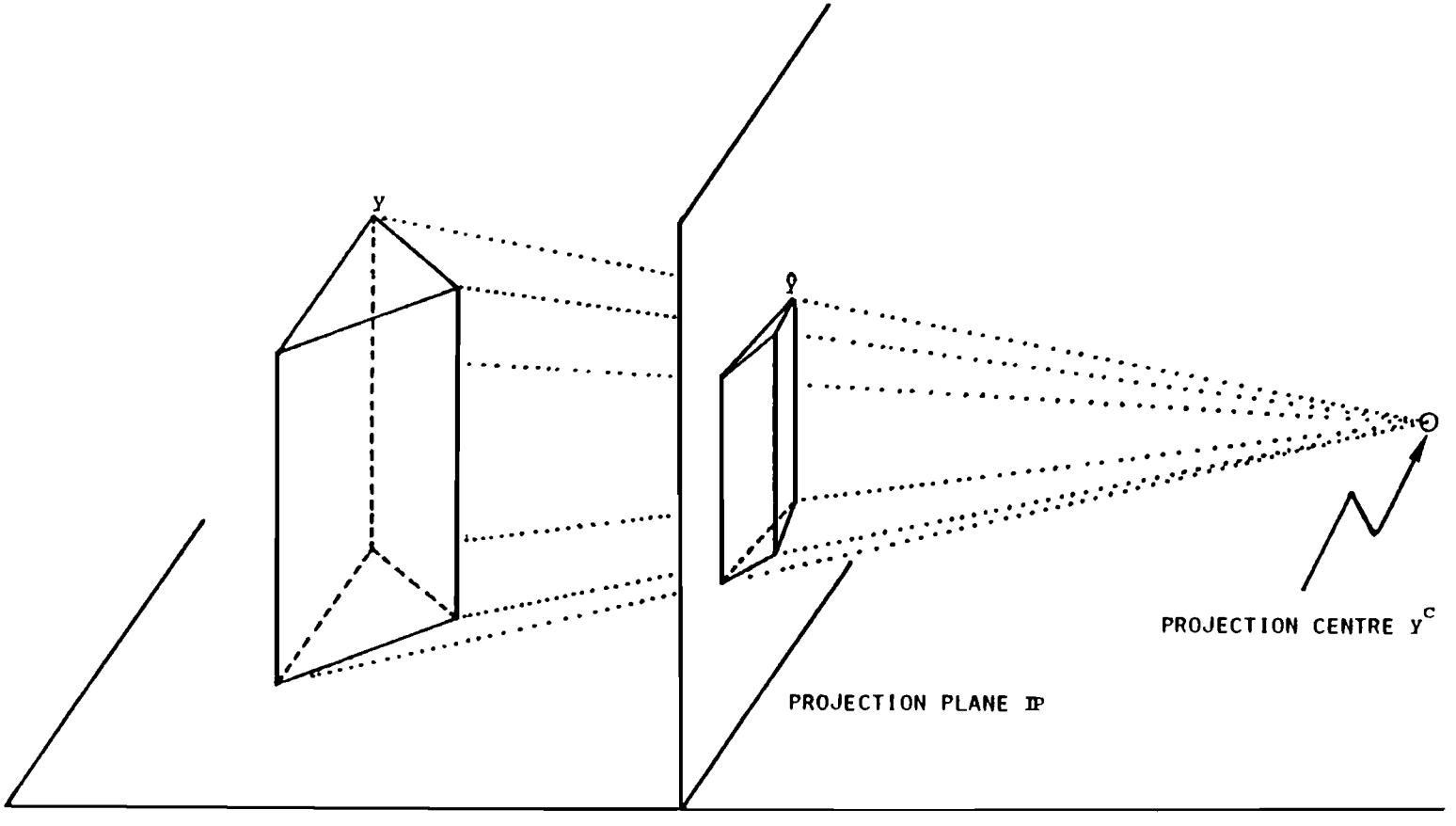


FIG. 10: PRINCIPLE OF THE PROJECTION IN THE 3-DIMENSIONAL CASE

variables x_i : $i \in U \setminus W$ for some subset $W \subseteq U$ with $|W| = 1,2,3$ to the value of some feasible solution $x^*_U \in X_U$. Substituting $x_{U \setminus W} = x^*_{U \setminus W}$ in (NLS_U) we get a linear restriction system

$$F \cdot x_W \leq f \qquad (PLS_{W,x^*})$$

of which the solution set Y is the projection of X on $\mathbb{R}^W$ with respect to $x_{U \setminus W} = x^*_{U \setminus W}$. But the problem has not been solved yet. Y is still a "ghost", for which we have to calculate the "appearance". Note that F has, at most, three columns but usually many rows. To take advantage of this special structure, a three-dimensional geometrical version of the simplex algorithm has been developed in [23]. This algorithm determines a "scaffold" for Y, i.e. it determines all extreme points and all extreme rays of Y, and, as well, for every pair of these vectors it yields the information whether the pair constitutes an edge of Y or not. In comparison to general methods which yield the same result, it can be seen that in the geometrical simplex algorithm the computational and the "bookkeeping" effort are reduced in an enormous way. Now, having a scaffold of Y, we can utilize the techniques of computer graphics (see e.g. [5], [6], [7], [23]) to generate a figure of Y. In the interesting three dimensional case the principle of producing a perspective figure of Y is illustrated by Figure 10: The projection centre y^c corresponds to the eyes of the user, the projection plane $\mathbb{P}$ to the screen of the computer, and $\hat{y}$ to the projection of $y \in Y$ on the screen. Hidden edges are represented by dotted lines, non-hidden edges by bright lines. Some possible graphical computer printouts for $|W| = 1,2,3$ are given by Figures 11 - 13.

A rather nice representation is obtained if colours are available. Then the non-hidden edges can be drawn in red and the hidden edges in green. It is even possible to determine the position of some (blue painted) point in the interior of Y or to "move" it through the polyhedron. This can be done by using cross-cuts (see Figure 14) or blinking techniques (if the moving point leaves Y). A further perfectionizing is obtained by letting the polyhedron "turn", in order to consider it from all sides.

Until now we only had a single figure in $\mathbb{R}^W$. But, using $(NLS^{=}_U)$ this figure can immediately be transformed into other coordinate systems. Suppose now (and this is the usual case if the polyhedron to visualize is an efficient face) that W and $U \setminus W$ are index sets of criteria

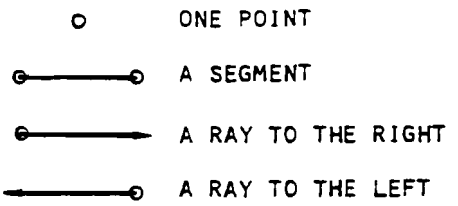


FIG. 11: SOME POSSIBLE COMPUTER PRINTOUTS IN THE 1-DIMENSIONAL CASE

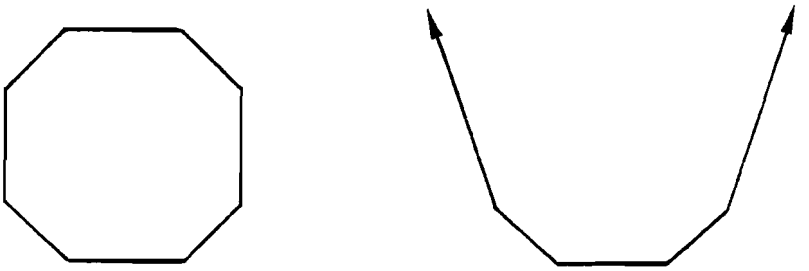


FIG. 12: SOME POSSIBLE COMPUTER PRINTOUTS IN THE 2-DIMENSIONAL CASE

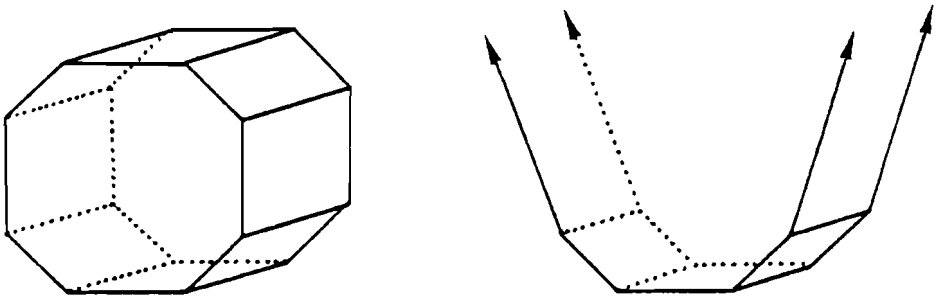
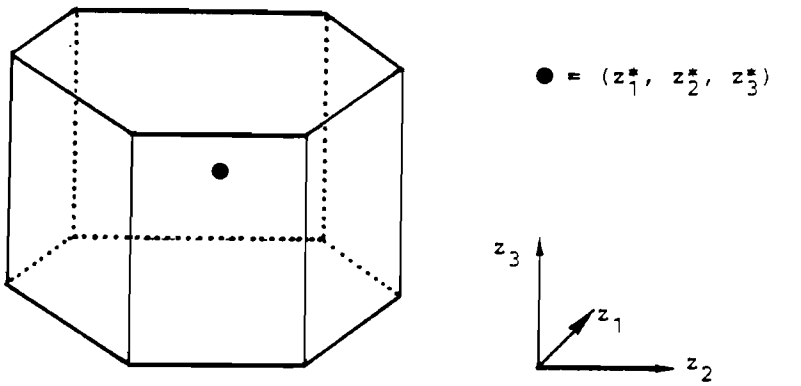


FIG. 13: SOME POSSIBLE COMPUTER PRINTOUTS IN THE 3-DIMENSIONAL CASE

and that $|U \setminus W|$ is 3 or not too much greater. Then we can use a discretization of $X_{U \setminus W}$ and determine one or several figures for every point $x^*_{U \setminus W}$ of this discretization. Doing so, it is possible to discover "qualitative changes" in the figures and hence to understand even the structure of higher dimensional convex polyhedra. For more technical details of visualizations the reader is referred to [23], [25], and more generally to [5], [6], [7], [13], [16].



CROSS-CUTS:

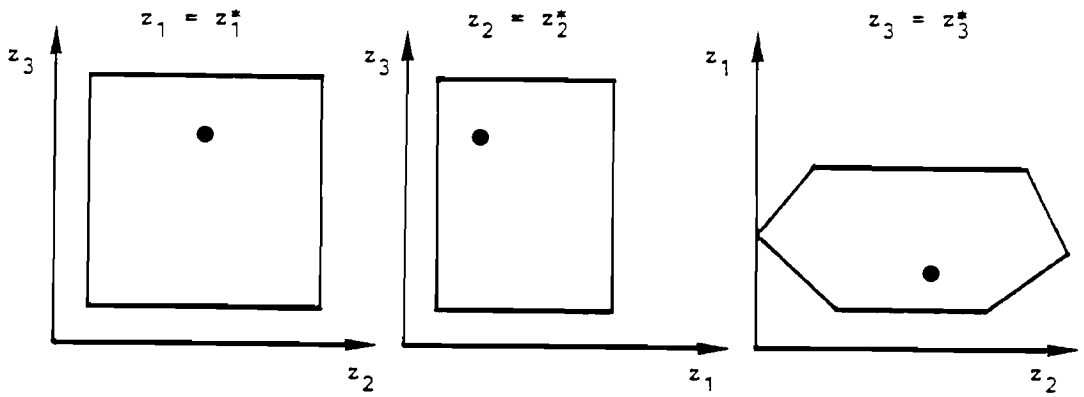


FIG. 14: DETERMINATION OF THE POSITION OF A MOVING POINT IN A 3-DIMENSIONAL POLYHEDRON

6. General Organization of MULTIPAR

The algorithmic aspect of MULTIPAR is summarized by Flow-Chart 1. Therein each module has to be considered as a special subroutine. The general organization of the method has been influenced by four basic principles:

(6.1) Modelling

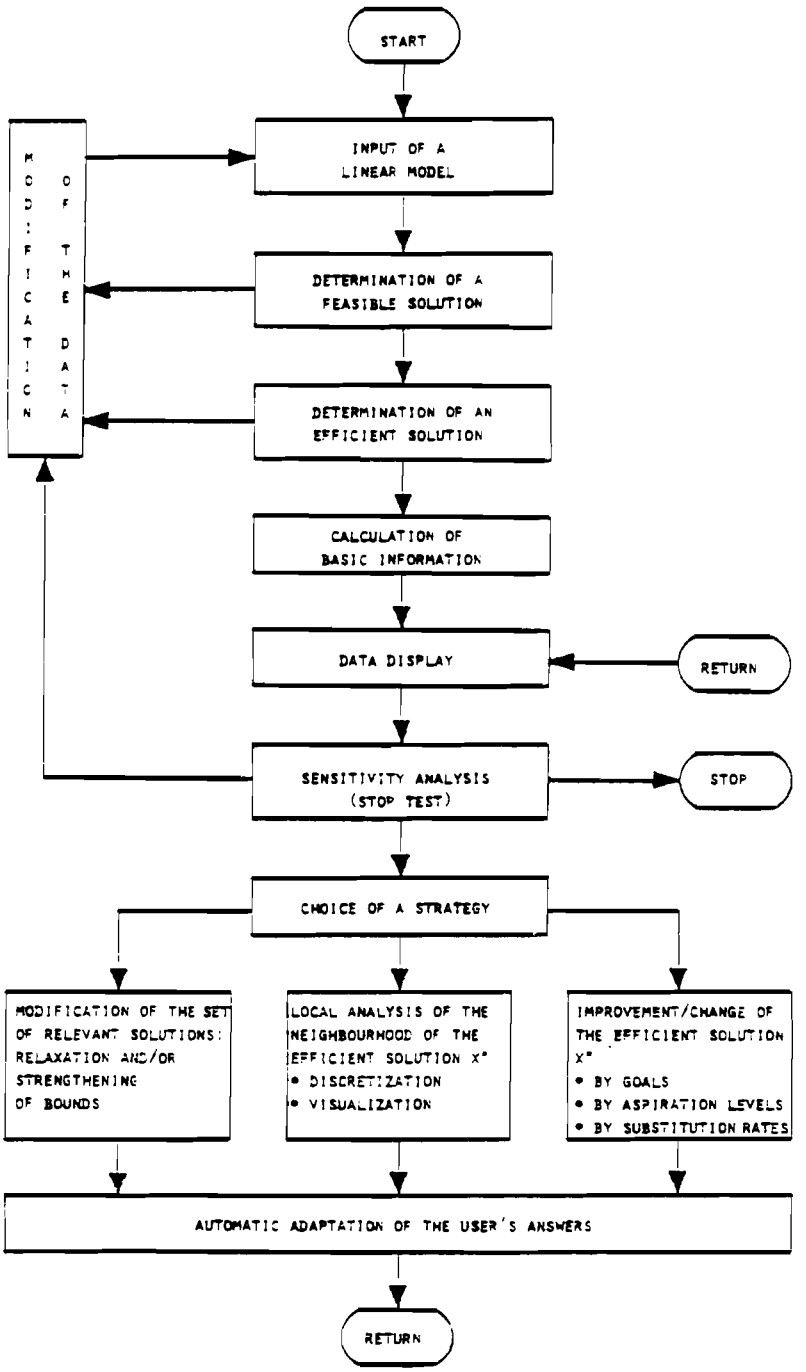
As input MULTIPAR needs a linear model like (GLS) in conjunction with (CGLS). Nowhere it is assumed that this multicriteria model remains stable. To the contrary, there is always the opportunity for modification. In the cases of ill-structured models (infeasibility or feasibility without efficiency) information is even given which helps to modify the original model. Sensitivity analysis simplifies the estimation of the influence of data (in)accuracy on a selected solution.

(6.2) Availability of information

After the initalization phase the user can always call a multitude of basic information which summarizes the actual model and the actual results of the preference analysis. Essentially, these data are: An ordered finite subset of good alternatives, an actual best compromise, the efficiency projections on the range of every criterion (between a minimizing and a maximizing solution with respect to this criterion), a vector of weights which fits best the given preference comparisons, an optimal solution with respect to those weights (the comparison of the actual best compromise and this particular optimal solution yields information on the preference structure and can be used as a convergence mechanism), and finally information on the geometrical structure of the actual set of relevant solutions (slackless and redundant constraints, independent variables).

(6.3) Active preference analysis

On the basis of the available information the user has numerous options to improve the actual compromise. Here, the most important strategies are: indication of goals, hierarchical aspiration levels, or substitution rates. In all of these cases an efficiency projection of the line segment between the actual compromise and the calculated new compromise is obtained. Another option is the visualization or the discretization of the efficient faces in the neighbourhood of the actual compromise.



FLOW-CHART 1: MULTIPAR: A FLEXIBLE DECISION AID METHOD FOR MULTICRITERIA ANALYSIS OF LINEAR MODELS

(6.4) Passive preference analysis

During the active preference analysis the user is asked to compare solutions. At least a comparison of the actual compromise and the optimal solution with respect to the approximated weights should be given. On the basis of these comparisons a new family of weights is calculated and a new solution which optimizes the corresponding weighted sum of the criteria.

7. A Brief Description of the Modules

In this section we finally come to the different elements of MULTIPAR. Since this would necessitate to write a book, we can not give a detailed description, here. For more technical representations, the reader is referred to [20] - [26] and to a monograph in preparation.

(7.1) Input of a linear model

The user indicates a linear model (GLS) in conjunction with criteria (CGLS). These data are transformed to (NLS) and (CNLS). Psychological investigations have shown that the human capacity of comparing data is restricted to a number of less than ten [11]. Hence the number of criteria should not exceed the magical number seven.

(7.2) Determination of a feasible solution

Starting with the natural basic solution, one tries to get a feasible solution by applying the generalized pivot steps of [25]: In each pivot step one changes to a more feasible solution in a lexicographical sense. With respect to some constraint which is actually not satisfied the new solution is more feasible or feasible, and each feasible component remains feasible. By doing so, the simplex algorithm adapts itself to the feasibility domain with respect to some linear order on the constraints (which can be indicated by the user). This procedure has three advantages in comparison to the common approaches of introducing artificial variables:

- the bookkeeping effort is reduced
- an initial basic feasible solution can be obtained which corresponds better to the importance of constraints (Sometimes there is empirical experience of such importances.)

- in the case of infeasibility, a set of satisfied constraints as well as a critical constraint of (GLS) are determined. This information can be used to relax the corresponding bounds. If we extend the criteria set by the relaxed constraints (with the orientation "→ min!" if an upper bound was relaxed and the orientation "→ max!" if a lower bound was relaxed) we can manipulate to get a new model whose solutions are as close to the original model as possible.

(7.3) Determination of an efficient solution

Starting with the basic feasible solution x^0 of (7.2) we next try to get an optimal solution of $(\bar{N}_0(g^\infty,w^*))$ with $g_l^\infty := -\infty \ \forall l \in L_{max}$, $g_l^\infty := +\infty \ \forall l \in L_{min}$, $g_l^\infty := x_l^0 \ \forall l \in L_{bal}$, and with a weighting $w_l^* > 0$. $\forall l \in L_{max} \cup L_{min}$ (which can be indicated by the user, otherwise $w_l^* = 1 \ \forall l$). If $(\bar{N}_0(g^\infty,w^*))$ has an optimal solution, it is efficient. If $(\bar{N}_0(g^\infty,w^*))$ has an unbounded objective value, we try to get an optimal solution of $(\bar{N}_0(g^0,w^*))$ where $g^0 = x_L^{00}$ (x^{00} is the last feasible solution). An optimum of $(\bar{N}_0(g^0,w^*))$ is efficient. If $(\bar{N}_0(g^0,w^*))$ has an unbounded objective value, the simplex tableau of the last feasible solution indicates an extremal ray and hence a set $H \subseteq L_{max} \cup L_{min}$ of unlimited criteria (i.e. in terms of (NLS) with respect to (CNLS)):

$$\forall x \in X \ \exists x' \in X : \begin{array}{ll} x_l' > x_l & \forall l \in H \cap L_{max} \\ x_l' < x_l & \forall l \in H \cap L_{min} \\ x_l' = x_l & \forall l \in L \setminus H \end{array}).$$

A maximal set H^* of such unlimited criteria can be determined if we successively replace L by $L \setminus H$ and repeat the procedure described above until the reduced system as an efficient solution (which is said to be reduced efficient with respect to (CNLS)). If L^* is the last reduced criteria set, we have $H^* = L \setminus L^*$. Obviously, it is a shortcoming of real world models to have unlimited criteria. Now, the decoding of H^* (i.e. H^* in terms of (CGLS)) is a useful information how to add new constraints or how to modify (GLS).

(7.4) Calculation of basic information

This subroutine is used to calculate the information which is necessary for a first data display (see (7.5)). The actual compromise is an efficient solution with minimal distance from the ideal point, the weighting w^* is the original weighting of (7.3), the solution with maximal utility is the efficient solution of (7.3), and the set of relevant solutions is the whole solution set.

(7.5) Data display

At any time (after the initialization), the user can call this subroutine. It is automatically presented after each iteration and provides characteristic data for the actual state of the model and the preference analysis. The user has the following information concerning preference analysis to his disposal:

- o an efficiency projection for the line segment between the actual compromise z^{com} and the solution with maximal utility z^{max} (see(7.11))
- o the actual relative weighting $\alpha_l \cdot w_l^*$ (see(7.11))
  - where α_l is an appropriate normalization factor for the absolute weighting w_l^* : $l \in L$.
- o the reliability of the weighting: ε^* (see (7.11))
- o the number of stored solutions
- o a listing of the stored solutions (ordered with respect to the approximated utility function)

- o the bounds $\bar{a}^-$, $\bar{a}^+$, $\dot{e}^-$, $\dot{e}^+$ which describe the set $\dot{Z}$ of relevant alternatives
  (i.e. $\dot{Z} = \{z \in \mathbb{R}^S : \bar{a}^- \leqq D \cdot z \leqq \bar{a}^+, \dot{e}^- \leqq z \leqq \dot{e}^+\}$
  with $\bar{a}^- \leqq \dot{a}^- \leqq \dot{a}^+ \leqq \bar{a}^+$ and $\dot{e}^- \leqq \dot{e}^- \leqq \dot{e}^+ \leqq \dot{e}^+$) (see (7.8))
- o the ranges for each criterion: $(\bar{z}_l^{min}, \bar{z}_l^{max})$
  with $\bar{z}_l^{opt} = opt\{\bar{z}_l : \bar{z} \in \bar{\dot{Z}}\}$, $opt \in \{min, max\}$.
  ($\bar{\dot{Z}}$: Extension of $\dot{Z}$, analogous to section 2)
- o an efficiency projection for each criterion, i.e. an efficiency projection of the line segment between two solutions z^1 and z^2 with $\bar{z}_i^1 = \bar{z}_i^{min}$ and $\bar{z}_i^2 = \bar{z}_i^{max}$, $i \in R_c \cup S_c$.

Concerning the geometrical structure of the set $\dot{Z}$ of relevant alternatives (and hence properties of the modelling), the user can get:

- o the discovered (or all) slackless constraints
- o the discovered (or all) redundant constraints
- o a system of independent variables (which fits best some indicated set of variables)
- o the corresponding dependency equation.

Finally, with respect to an intended modification of the model, the user can request:

- o the data of (GLS) and (CGLS).

(7.6) Sensitivity analysis

By this module the user gets a means to estimate the influence of data inaccuracy on z^{com} and/or z^{max} (or another stored solution z^*). Let x^* be z^* with respect to (NSL). First an optimal basic solution of $(\bar{N}_O(x^*_L,w^*))$ is calculated. Then, free ranging sensitivity analysis is based on the corresponding index basis (see [4] and [8]).

The user has the following options:

- o free ranging of a column of D
- o free ranging of the bounds and the goals g
- o free ranging of a row of D
- o free ranging of the weights w.

Here, sensitivity analysis is performed by visualizing the data regions. The subroutine terminates by choosing one of the following strategies how to continue the procedure:

(1) STOP
    (i.e. output of relevant data and end of the process)
(2) Modification of the data
    (i.e. modification of (GLS) and/or (CGLS)
(3) Choice of a strategy
    (i.e. continuation of preference analysis)

(7.7) Choice of a strategy

The user has to compare z^{com} and z^{max} with respect to the preferences (of the decision maker).

There are the following options

(1) z^{com} is preferred to z^{max} $(z^{com} \succ z^{max})$
(2) z^{max} is preferred to z^{com} $(z^{max} \succ z^{com})$
(3) z^{max} and z^{com} are indifferent $(z^{max} \sim z^{com})$
(4) z^{max} and z^{com} are incomparable $(z^{max} ? z^{com})$.

Then, the best (or one) of them becomes the new actual compromise (it is also possible to make one of the stored solutions to the new actual compromise, z^{max} and z^{com} are stored, automatically). The subroutine terminates by choosing one of the options for improving (the new) z^{com}.

(7.8) Modification of the set of relevant solutions

This is the most simple method for preference analysis. The user restricts the whole solution set Z to a non-empty subset $\dot{Z}$ of relevant solutions by indicating new bounds $\dot{d}^-$, $\dot{d}^+$, $\dot{e}^-$, $\dot{e}^+$ (with $d^- \leqq \dot{d}^- \leqq \dot{d}^+ \leqq d^+$ and $e^- \leqq \dot{e}^- \leqq \dot{e}^+ \leqq e^+$) for (GLS). The non-emptyness of $\dot{Z}$ is obtained by requiring $z^{com} \in \dot{Z}$.

Such a modification naturally induces a modification of (NLS), too. It is even possible that the structure of the set of efficient solutions changes. But, usually, such a modification consits in strenghtening aspiration levels of criteria. Then the new set of efficient solutions is a subset of the original one. If this is not the case, the new actual compromise will be determined by an optimum of $(\bar{N}_o(x_L^{com}, w^*))$ (where x_L^{com} corresponds to the old actual compromise z^{com}).

(7.9) Local analysis of the neighbourhood of z^{com}

The user defines a neighbourhood U of z^{com} by indicating aspiration levels for the criteria. Then all efficient faces of $Eff_L(X) \cap U$ are determined by means of (3.7). These faces will either be visualized or discretized. By doing so, it is possible to discover new interesting solutions. These can be stored and one of them can even define the new actual compromise. During this process, solutions can be compared with respect to the preferences (of the decision maker) in order to get information for a better approximation of the weights (see (7.11)).

(7.10) Direct improvement of the actual compromise z^{com}

This subroutine offers three different options for calculating a new solution z^*, trying to improve z^{com}.

(7.10.1) Indication of a goal

The user indicates a goal z^o. This can be done either directly or by means of a direction d^o and a multiplicative factor $\alpha > 0$ such that $z^o = z^{com} + \alpha \cdot d^o$. (Because of the efficiency of z^{com} it is reasonable to assume that z^o should not dominate z^{com}.) Then a solution z^* is determined which minimizes the distance from z^o (see (3.1)).

(7.10.2) Indication of hierarchical aspiration levels

The user indicates an ordered family of constraints:

$$\bar{z}_{i_1} \; \square_{i_1} \; \bar{z}^*_{i_1}, \ldots, \bar{z}_{i_k} \; \square_{i_k} \; \bar{z}^*_{i_k}$$

where $\square_i$ is $\geq$ for a criterion with orientation "max", $\leq$ for a criterion with orientation "min", and $\square_i$ is $\geq$ and/or $\square_i$ is $\leq$ for a criterion with orientation "bal". These requirements can be translated into a value g^* of the goal paramter g of $(\bar{N}(g,w^*))$. For doing so it is necessary to modify $(\bar{N}(g^*,w^*))$ by introducing upper bounds for u_l, v_l : $l \in L_{bal}$ and fixing $v_l = 0 \; \forall l \in L_{max} UL_{min}$. Let this modification of $(\bar{N}(g^*,w^*))$ be denoted by $(\bar{N}_*(g^*,w^*))$. The next step is to calculate a solution x^* of $(\bar{N}_*(g^*,w^*))$ which fits the desired aspiration levels in the given hierarchical order (analogous to (7.2)) best. Finally, z^* is the transformation of x^* (in terms of (GLS)).

(7.10.3) Indication of substitution rates

The user indicates a reference criterion $i^* \in R_c US_c$, a reasonable amount $\Delta_{i^*} \neq 0$ for improving $\bar{z}^{com}_{i^*}$ ($\Delta_{i^*} > 0$ if $sat_{i^*} = max$, $\Delta_{i^*} < 0$ if $sat_{i^*} = min$, $\Delta_{i^*} \neq 0$ if $sat_{i^*} = bal$), and for each $i \in R_c US_c$ an amount $\Delta_i \neq 0$ for relaxing $\bar{z}^{com}_i$ ($\Delta_i < 0$ if $sat_i = max$, $\Delta_i > 0$ if $sat_i = min$, $\Delta_i \neq 0$ if $sat_i = bal$), such that $(\bar{z}^{com}_i)_{i \in R_c US_c}$ and the criterion profile $\bar{z}^{(i)}$ are equivalent (with respect to the preferences of the decision maker):

$$\bar{z}^{(i)}_k := \begin{cases} \bar{z}^{com}_k + \Delta_k & \text{for } k \in \{i, i^*\} \\ \bar{z}^{com}_k & \text{for } k \notin \{i, i^*\} \end{cases}, \; k \in R_c US_c$$

The absolute amounts $|\Delta_i|$ induce a value w^o of the weighting vector w in $(\bar{N}(g,w))$. Depending on the sign on Δ_i either a balanced criterion will now be considered to have the orientation max or to have the orientation min. This yields a multicriteria system (without balanced criteria) and a corresponding weighting. (It corresponds to $(\bar{N}(g^\infty,w^o))$ with $g^\infty = \pm\infty$, $u_l = 0 \; \forall l \in L_{max} UL_{min}$, u_l or $v_l := 0$ for $l \in L_{bal}$). Maximization of x_o finally yields a new solution z^*.

Since z^* only approximates the preferences, it is reasonable to see what happens between z^{com} and z^*. That is why the user gets an efficiency projection of the line segment between z^{com} and z^*.

By using this subroutine, the user gets the opportunity to change some of the indicated data and to repeat the procedure. By doing so,

it is possible that he discovers interesting solutions (which will be stored) or even a better actual compromise. Preference comparisons of solutions can be given in order to improve the approximation of the weights in (7.11).

(7.11) Automatic adaptation to the user's answers

During the previous subroutines the user was asked to indicate comparisons of solutions with respect to the preferences of the decision maker. Under the hypothesis that there is a linear utility function

$$uti(\bar{z}_{i_1},\ldots,\bar{z}_{i_n}) = \sum_{k=1}^{n} w_{l_k} \cdot uti_{i_k}(\bar{z}_{i_k})$$

$$\text{with } uti_{i_k}(\bar{z}_{i_k}) := \begin{cases} x_{l_k} & \text{if } sat_{i_k} = max \\ -x_{l_k} & \text{if } sat_{i_k} = min \\ -\left|x_{l_k} - \frac{1}{2}(x^{max}_{l_k} - x^{min}_{l_k})\right| & \text{if } sat_{i_k} = bal \end{cases}$$

$$\text{and } w_{i_k} > 0 \ \forall k = 1,\ldots,n$$

such that

$$\bar{z} \succ \bar{z}' \Rightarrow uti(\bar{z}) > uti(\bar{z}') + \varepsilon$$
$$\bar{z} \sim \bar{z}' \Rightarrow |uti(\bar{z}) - uti(\bar{z}')| \leq \varepsilon$$

with a threshold $\varepsilon > 0$. (l_k corresponds to i_k with respect to (NLS).)

Now every preference comparison can be transformed into a linear constraint so that we can get a linear restriction system with variables w and ε. By means of the simplex algorithm it is possible to calculate a weighting $w^* \gg 0$ and a threshold ε^*, such that w^* and ε^* solve this linear system with a minimal threshold value ε.

Next, consider $(\bar{N}(g^o,w^*))$, where $g^o_l = -\infty \ \forall l \in L_{max}$, $g^o_l = +\infty \ \forall l \in L_{min}$, $g^o_l = \frac{1}{2}(x^{max}_l - x^{min}_l) \ \forall l \in L_{bal}$, $v_l = 0 \ \forall l \in L_{max} \cup L_{min}$. Optimization of x_o with respect to this system yields the new solution z^{max}, which maximizes utility. (In this case $-w^*_l$ is used instead of $w^*_l \ \forall l \in L_{bal}$.)

8. Conclusion

MULTIPAR is an advanced technique for multicriteria decision aid in the context of linear decision models. The method has been constructed to satisfy requirements desired by sophisticated users, i.e. users who have experience in multicriteria philosophy and who need a very

flexible instrument for immediate interaction in the progress of a decision process. It is not easy to implement the method because of its complex visualization techniques. Our experience, however, has convinced us that geometric figures (rather than infinite lists of data print-outs) are strongly appealing means for letting the decision makers grasp the essentials and hence for promoting the acceptance of operations research by decision makers in general. The use of computer graphics in O.R. is just at the beginning ([10], [16]). The powerful development of graphical systems even for micro-computers, however, will have increasing influence on the application of decision aid techniques. MULTIPAR should strengthen this promising new trend. If the implementation of the whole algorithm should appear to be too troublesome to a potential user, the technique of efficiency projections alone is extremely valuable for improving solutions. In a forthcoming paper we will report on some experiences with MULTIPAR for real world problems.

References

1. Alter, S., "Why is Man-Computer Interaction Important for Decision Support Systems?" Interfaces, Vol. 7, 2 (1977), pp. 109-115.
2. Bernabeu, Y.-A. and Winkels, H.-M., "MIRACLE: Une Méthode Interactive de Réduction, d'Amélioration et de CLassification Evolutive," Document du LAMSADE No. 16, Université Paris IX-Dauphine, Paris, France, 1981.
3. Bode, G., "Was ist interaktive Programmierung?" Elektronische Rechenanlagen, 20, 5 (1978), pp. 238-241.
4. Ciriani, T., "The Free Parametrization for Linear Programming with Bounded Variables," Rendiconti del Seminario Matematico della Universitá di Padova, Vol. 47 (1972), pp. 77-107.
5. Crow, F.C., "Three-Dimensional Computer Graphics, Part 1," Byte, Vol. 6, 3 (1981), pp. 54-82.
6. -----, "Three Dimensional Computer Graphics, Part 2," Byte, Vol. 6, 4 (1981), pp. 290-302.
7. Encarnação, J.L., Computer-Graphics: Programmierung und Anwendung von graphischen Systemen, R. Oldenbourg-Verlag, München-Wien, 1975.
8. Gal, T., Postoptimal Analyses, Parametric Programming, and Related Topics, McGraw-Hill, New York, 1979.

9. Hwang, C.-L. and Masud, A.S.M., Multiple Objective Decision Making - Methods and Applications - A State of the Art Survey, Lecture Notes in Economics and Mathematical Systems No. 164, Springer Verlag, Berlin-Heidelberg-New York, 1979.
10. Johnson, L.E. and Loucks, D.P., "Interactive Multiobjective Planning Using Computer Graphics," Computers and Operations Research, Vol. 7, 1-2 (1980), pp. 89-97.
11. Miller, G., "The Magical Number Seven Plus or Minus Two: Some Limits on Our Capacity for Processing Information," Psychological Review, Vol. 63 (1956), pp. 81-97.
12. Nasini, A.C., "Guide pour les Recherches Futures sur les Méthodes Interactives pour la Résolution de Problèmes Multicritères," Mémoire DEA 103, U.E.R. Sciences des Organisations, Université Paris IX-Dauphine, Paris, France, 1980.
13. Newman, W.M. and Sproull, R.F., Principles of Interactive Computer Graphics, Second Edition, McGraw-Hill, London, 1979.
14. Orchard-Hays, W., Advanced Linear-Programming Computing Techniques, McGraw-Hill, New York, 1968.
15. Pollack, M., "Interactive Models in Operations Research - An Introduction and Some Future Research Directions," Computers and Operations Research, Vol. 3, 4 (1976), pp. 305-312.
16. Robinson, P., "Trends in Computer Graphics," Journal of the Operational Research Society, Vol. 32 (1981), pp. 303-316.
17. Stoer, J. and Witzgall, Ch., Convexity and Optimization in Finite Dimensions I, Springer-Verlag, Berlin-Heidelberg-New York, 1970.
18. Winkels, H.-M., "Die Bestimmung aller Lösungen eines linearen Restriktionssystems und Anwendungen," Arbeitsbericht Nr. 15, Institut für Unternehmungsführung und Unternehmensforschung, Ruhr-Universität, D-4630 Bochum, 1979.
19. -----, "Interaktive Lösungsverfahren für lineare Probleme mit mehrfacher Zielsetzung," in P. Henn, B. Schips and P. Stähly, eds., Quantitative Wirtschafts- und Unternehmensforschung, Springer-Verlag, Berlin-Heidelberg-New York, 1980, pp. 560-585.
20. -----, "Complete Efficiency Analysis for Linear Vector Maximum Systems: Theoretical Background and an Algorithm," Document du LAMSADE No. 12, Université Paris IX-Dauphine, Paris, France, 1980, submitted for publication in Linear Algebra and Applications.

21. -----, "How to Analyse Half-Baked Linear Vector Maximum Systems," Operations Research Verfahren, 40 (1981), pp. 193-196.

22. -----, "Efficiency Projections", Working Paper on Economathematics No. 8103, Abteilung für Wirtschaftswissenschaft und Abteilung für Mathematik, Ruhr-Universität, D-4630 Bochum, 1981, submitted for publication in OR-Spektrum.

23. -----, "Visualization and Discretization for Convex Polyhedral Sets," Working Paper on Economathematics No. 8104, Abteilung für Wirtschaftswissenschaft und Abteilung für Mathematik, Ruhr-Universität, D-4630 Bochum, 1981, submitted for publication in Zeitschrift für Operations Research.

24. -----, "Local Efficiency Analysis," Working Paper on Economathematics No. 8110, Abteilung für Wirtschaftswissenschaft und Abteilung für Mathematik, Ruhr-Universität, D-4630 Bochum, 1981.

25. -----, Primal and Dual Upper Bounding Techniques: A Way to Modernize Classical Simplex Algorithms, Monograph to appear, 1982.

26. ----- and Meika, M., "An Interactive Gradient Algorithm for Generating Efficient Solutions," Working Paper on Economathematics No. 8105, Abteilung für Wirtschaftswissenschaft und Abteilung für Mathematik, Ruhr-Universität, D-4630 Bochum, 1981, submitted for publication in European Journal of Operational Research.

SOME IMPROVEMENTS TO THE REFERENCE POINT APPROACH FOR DYNAMIC MULTICRITERIA LINEAR PROGRAMMING

M. Kallio and M. Soismaa

Helsinki School of Economics, Finland

1 Introduction

Among desired properties of interactive approaches to multiple criteria linear programming /3/ are the following: (i) applicable to the size of problems which appear in practice, (ii) ability to deal with a large number of objectives (in particular for handling trajectory objectives /5/), (iii) easy to implement on a computer (given an LP code as a starting point), (iv) user friendliness taking into account that decision makers work at the terminal, and (v) satisfactory convergence in a reasonable number of interactive iterations (which does not necessarily mean convergence in a strict mathematical sense under an assumed utility function but convergence to a satisfying solution /5/).

This paper deals with the reference point approach of Wierzbicki /4/ which directly satisfies properties (i) and (ii). Kallio et al /1/ discuss a user oriented approach to implementation which is only a minor modification to a standard LP code. They provide also a basic modification needed for convergence. The aim of this paper is to provide some improvements to user friendliness and convergence.

After reviewing the reference point approach in Section 2 we present in Section 3 a method for constructing a feasible initial solution for a new reference point. The purpose of this method is to decrease the user's waiting time for solutions of the interactive iterations. Results from numerical experiments are also presen-

ted. In Section 4 we deal with a problem of trajectory optimization. A general undesired property of dynamic LP is that the trajectories of various quantities tend to fluctuate. In particular this phenomen may result in unsatisfactory objective trajectories when the reference point approach in applied. Interactive iterations resulting in such unacceptable trajectories are in fact useless, and if they are many the method may fail. We present two approaches to accomplish smooth trajectories. Results from numerical experiments are presented as well.

2 The reference point approach

We shall now briefly review the reference point approach for multicriteria linear programming (MCLP) as presented in Kallio et al /1/. Let A be in R^{mxn}, C in R^{pxn}, and b in R^m and consider a multicriteria linear program

$$Cx = q \tag{1}$$

$$Ax = b \tag{2}$$

$$x \geq 0 \tag{3}$$

where the decision problem is to determine a n-vector x of decision variables satisfying (2) and (3) and taking into account the p-vector q of objectives defined by (1). We assume that each component of q is desired to be as large as possible.

What we call a **reference point** or **reference objective** is a suggestion $\bar{q}$ by the decision maker reflecting in some sense an aspiration level for the objectives. According to Wierzbicki /4/, we consider for a reference point $\bar{q}$ a penalty scalarizing function $s'(q-\bar{q})$ defined over the set of objective vectors q. Characterization of functions s, which result in Pareto optimal (or weakly Pareto optimal) minimizers of s over attainable points q is given by Wierzbicki.

If we regard the fuction $s(q-\bar{q})$ as the "distance" between the points q and $\bar{q}$, then, intuitively, the problem of finding such a minimum point means finding among the Pareto set the **nearest** point $\hat{q}$ to the reference point $\bar{q}$. However, as it will be clear later, our function s is not necessarily related to the usual

notion of distance. Having this interpretation in mind, the use of reference points optimization may be viewed as a way of guiding a sequence $\{\bar{q}^k\}$ of Pareto points generated from the sequence $\{\bar{q}^k\}$ of reference objectives. These sequences will be generated in an interactive process and such interference should result in an interesting set of attainable points $\hat{q}^k$. If the sequence $\{\hat{q}^k\}$ converges, the limit point may be seen as a solution to the decision problem.

We shall apply a practical form of the penalty scalarizing function s(w), where minimization results in a linear programming formulation. We denote $w \equiv q - \bar{q}$, for brevity. Our function is given as follows:

$$(4) \qquad s(w) = -\min\{\rho \min_i w_i \ , \ \sum w_i\} - \varepsilon w.$$

Here ρ is an arbitrary penalty coefficient which is greater than or equal to p and $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots \varepsilon_p)$ is a nonnegative vector of parameters. If $\rho = p$, then (4) reduces to

$$(4') \qquad s(w) = \rho \max_i -w_i - \varepsilon w;$$

i.e. the minimum of s is attained according to the worst case criterion max $-w_i$ corrected with a term $-\varepsilon w/\rho$ (which is small if ε_i is small for each i).

The reference point optimization problem when the scalarizing function (4) is applied is the following linear programming problem /1/:

find s_o, s, y, w, and x to

$$(5) \qquad \text{minimize} \quad y - \varepsilon w$$

$$(6) \qquad \text{subject to} \quad s_o - y - \gamma w = 0$$

$$(7) \qquad s - \gamma^T y - \rho w = 0$$

$$(8) \qquad -w + Cx = \bar{q}$$

$$(9) \qquad Ax = b$$

$$(10) \qquad s_o, \ s \text{ and } x \text{ nonnegative.}$$

Here $\gamma = (1, 1, \ldots 1)$ and $s = (s_1, s_2, \ldots, s_p)^T$ are p-vectors, and s_0 and γ are scalars. One may show /1/ that if $\varepsilon > 0$ then the optimal solution $q = Cx$ is Pareto optimal.

3 A feasible initial solution

Consider an interactive iteration. After the new reference point has been inserted, a natural idea for solving the resulting problem (5) - (10) is to start (the simplex method) with the optimal basis obtained from a preceeding interactive iteration. (Now only the right hand side has been changed.) The drawback of this approach is that this initial basis usually is infeasible. Therefore Phase I of the simplex method would be needed, at the end of which the resulting feasible solution may be rather far from optimal. To avoid the difficulty caused by initial infeasibility, several approaches may be considered. First, if an (optimal) feasible basis is available for reference point $\bar{q}^1$, and if $\bar{q}^2$ is the reference point currently under consideration, then we may parametrize the reference point $\bar{q} = \bar{q}^1 + \theta(\bar{q}^2 - \bar{q}^1)$ letting the parameter θ increase from 0 to 1. Standard parametric programming starting with $\theta=0$ would be applicable to solve the problem for $\theta=1$ (i.e. for $\bar{q} = \bar{q}^2$). The tests (with a model mentioned below) concerning this approach, however, showed sometimes unsatisfactorily slow convergence (i.e. a very large number of basis changes were needed for parametrization). Second, the dual simplex method could have been applied for reoptimization, but software was not available. For these reasons an alternative approach was considered. In this case a previously obtained (optimal) feasible basic solution is employed for constructing for the present problem a feasible (typically nonbasic) solution and a related basis. Standard procedures are then applied to find a feasible basic solution and then Phase II of the simplex method is executed. We shall now describe, how the initial solution can be constructed, and thereafter, we demonstrate the performance of this approach by numerical experiments.

3.1 Construction of an initial solution

Let us assume that we have an optimal basic solution $(s_0^*, s^*, y^*, w^*, x^*)$ for a reference point q^* of problem (5) - (10), and consider the new reference point $\bar{q}$.

Let $\bar{x} = x^* \geq 0$ so that $A\bar{x} = b$. Define $\bar{w}$ to satisfy (8) for $x = \bar{x}$ so that $\bar{w} = C\bar{x} - \bar{o}$. In order to satisfy the constraints (6) and (7), given $x = \bar{x}$ and $w = \bar{w}$, we define the values for $\bar{\gamma}$, $\bar{s}_0$ and $\bar{s}$ as follows:

$$(11) \qquad \bar{y} = \max \{ -\sum \bar{w}_i, \max_i - \rho \bar{w}_i \},$$

$$(12) \qquad \bar{s}_0 = \bar{y} + \gamma \bar{w},$$

$$(13) \qquad \bar{s} = \gamma^T \bar{y} - \rho \bar{w}.$$

One may readily check that $\bar{s}_0 \geq 0$ and $\bar{s} \geq 0$. In summary, our construction has resulted in a solution $(\bar{s}_0, \bar{s}, \bar{y}, \bar{w}, \bar{x})$ which is feasible for problem (5) - (10) when the current reference point $\bar{q}$ is applied.

Obviously this solution may not be a basic solution, and our next task is to define a related initial basis. This of course may be done in several ways. A natural choice is the basis which we obtained at the optimum for reference point q^*. If the variables y and w are in this basis (which is a condition easy to satisfy, because y and w are free variables) then the only nonbasic variables being at a nonzero level at the initial solution $(\bar{s}_0, \bar{s}, \bar{y}, \bar{w}, \bar{x})$ are among the slacks s_0 and s. Thus only the values of these nonbasic slacks have to be set to start solution.

3.2 Numerical results

Because we are working interactively with the decision maker at the terminal it is very important that runs are completed as fast as possible. The purpose of the numerical runs was to investigate the savings in computing time (and in the decision maker's waiting time for solution at each interactive iteration) when the approach above was applied and compared with the naive approach where an initial basic solution is employed.

The model used for tests is a dynamic linear programming model of a forest sector (i.e. a model of forestry and forest industries). There are approximately 700 variables and 550 rows in the 8 period model which has two trajectory objectives: industrial profit and forestry profit. To study the efficiency of our approch we performed experiments with five different reference points. These reference trajectories have been illustrated in Figure 1. An optimal basic solution corresponding to trajectory q^* (of Figure 1) was employed to construct the initial feasible solution (and basis) for each of the five trajectories. The same basis was used to start the naive approach for comparison.

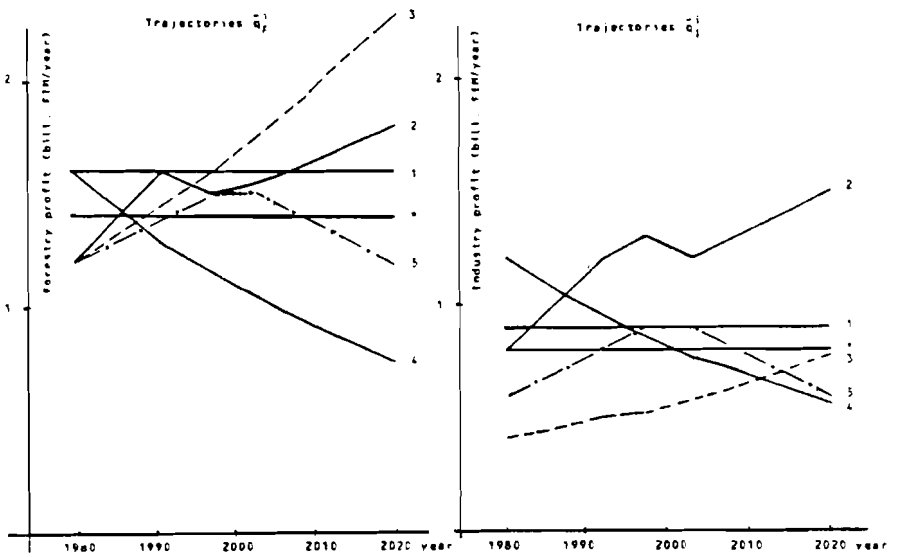


Figure 1. The reference trajectories $\bar{q}^i = (\bar{q}_F^i, \bar{q}_I^i)$ used in experiments i = 1,2,3,4,5 and *.

Table 1 shows the results of the experiments. It reveals both the total CPU time (on a VAX/780 computer when the MINOS code /2/ was used) to find an optimal solution. Also the number of simplex iterations is indicated. The CPU time includes a problem set up time which in each case (and for both approaches) was between 32 and 33 seconds.

Table 1. Solution time (in CPU-seconds) and the number of iterations for the naive approach (N) and the construction approach (C).

Reference	CPU seconds		Iterations	
trajectory	N	C	N	C
q^1	38	43	13	19
q^2	215	112	273	114
q^3	152	97	193	92
q^4	98	41	117	12
q^5	94	80	99	68
Total	597	373	695	405

We may conclude from Table 1 that the naive approach takes over 50 percent more CPU time than does the approach of Section 2.1. If we consider only the time required for iterations then the naive approach more than doubles the requi-

red CPU time. In only one case (the first one) the naive approach was superior. In this case (due to similarities in the reference trajectories between q^* and $\bar{q}^1$) the initial basis was also feasible for the naive approach.

4 Smoothing the trajectories

Due to properties of extreme point solutions, a drawback of dynamic LP is that the optimal trajectories tend to fluctuate over time. Because of the existence of multiple optima the trajectories even can be randomly placed. This does not appear desirable to a decision maker who has chosen a reference point with smooth growth for instance. A standard way to avoid this problem is to restrict the derivatives of the trajectories. This may be done by setting bounds either on relative or absolute change from one period to the next. If both lower and upper bounds are used then for a T-period model 2T additional constraints (which are not the type of simple upper bounds) are needed for each trajectory to be smoothed.

Another approach for smoothing is to restrict a trajectory to a linear combination of a finite number of predetermined smooth trajectories. Such linear combinations are supposedly smooth as well. The generating trajectories may be drawn individually for each application or some general approaches (e.g. the one presented in Section 4.1 below) may be used. The number of additional constraints needed for each trajectory is T; i.e. the increase is only one half compared with the alternative above. Furthermore, the loss in optimality (due to additional constraints imposed for smoothing) can be negligible given that the number and shape of generating trajectories is properly chosen. On the other hand, bounds on derivatives may lead to an undesired loss in optimality when these bounds are tight enough to guarantee reasonable smoothness.

In the following we apply smoothing to the trajectories of objectives. First, we apply an approach which in fact does not restrict but relaxes the original problem. Second, we present one way of applying generating trajectories for smoothing. Finally, we discuss some numerical experiments with these approaches.

4.1 Adjustment approach

When the trajectory represents income (as is the case for forestry and for industry profit) an efficient approach for smoothing is to allow adjustment through saving

and borrowing. We can thus save part of the profit from one period to the next one or vice versa. Accordingly, equation (8) becomes

$$(14) \qquad -w + Cx + Dv = q$$

Here the components of decision vector v refer to borrowing and investments over consequent periods. The matrix D transforms the effects of these activities into income trajectories simultaneously accounting for interest rates. If the borrowing and saving interest rates are assumed equal (as we do in experiments below) then a single non-sign-constrained component of v may represent both saving and borrowing. Otherwise, interest rate for borrowing has to be greater than than for saving (to avoid unboundedness), and separate nonnegative components of v are needed for these two activities.

4.2 Generating approach

An alternative for representing an objective trajectory as a linear combination of smooth generating trajectories is an expansion similar to the Fourier series. If $q_k(t)$ is the kth (endogeneous) objective trajectory we impose the following restriction for each t (t = 0, 1, ..., T-1) and k (suppressing k):

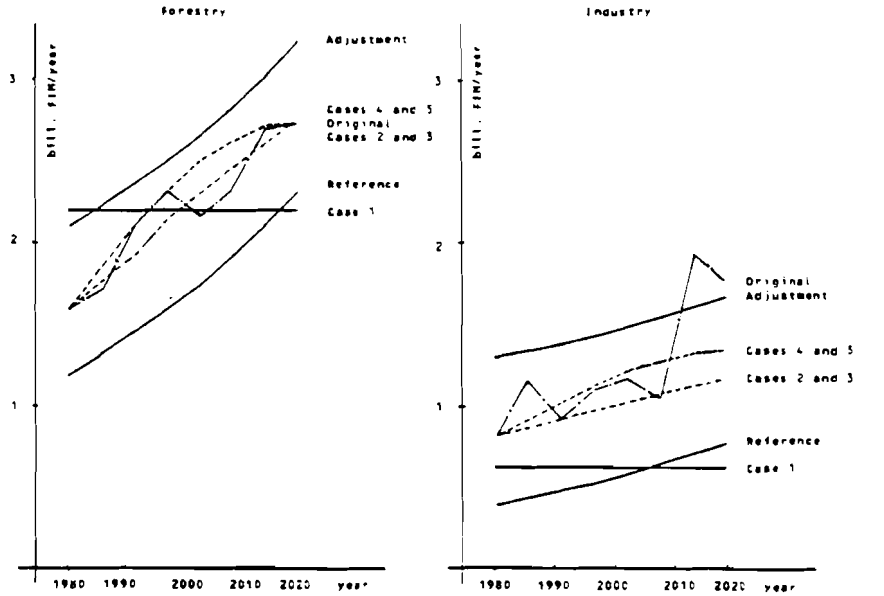
$$(15) \qquad q(t) = d + \sum\{a_j \sin(jt\pi/4(T-1)) + b_j \cos(jt\pi/4(T-1))\}.$$

Here d, a_j and b_j (for each j) are free variables to be determined by the optimal solution, and T is the number of time periods in the model. Truncation of this expansion (which may be dependent on trajectory k) determines the number of generating trajectories, for each k. The larger this number is the less restrictive is (15), and on the other hand, the less smooth may the (optimal) trajectory $q_k(t)$ be.

4.3 Numerical experiments

The approaches of Sections 4.1 and 4.2 were tested in six different cases for the forest sector model of eight periods. In the generating approach we used five different cases varying in the number of generating trajectories. Case i, for i = 1, 2, ..., 5, involves the first i terms (including the time-independent term d) of

RUN 1



RUN 2

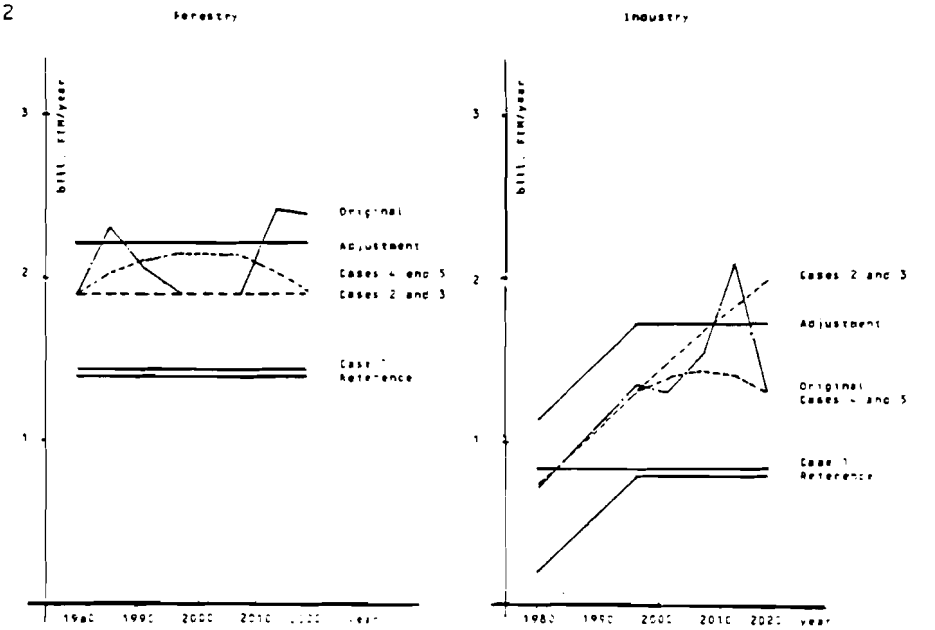
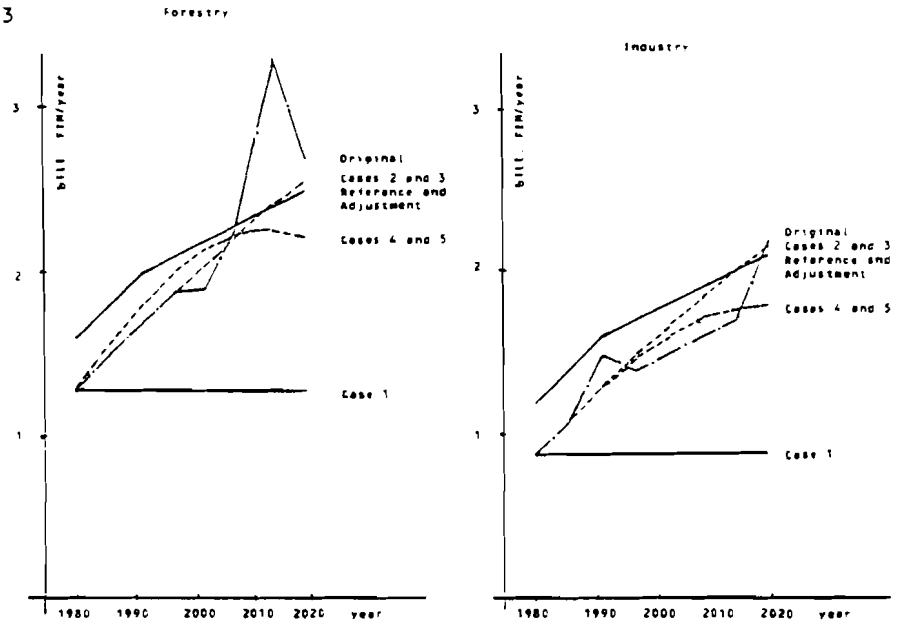


Figure 2. The trajectories of the adjustment approach and the generating approach (Cases 1,..., 5) comapred to the nonsmoothed trajectories.

RUN 3



RUN 4

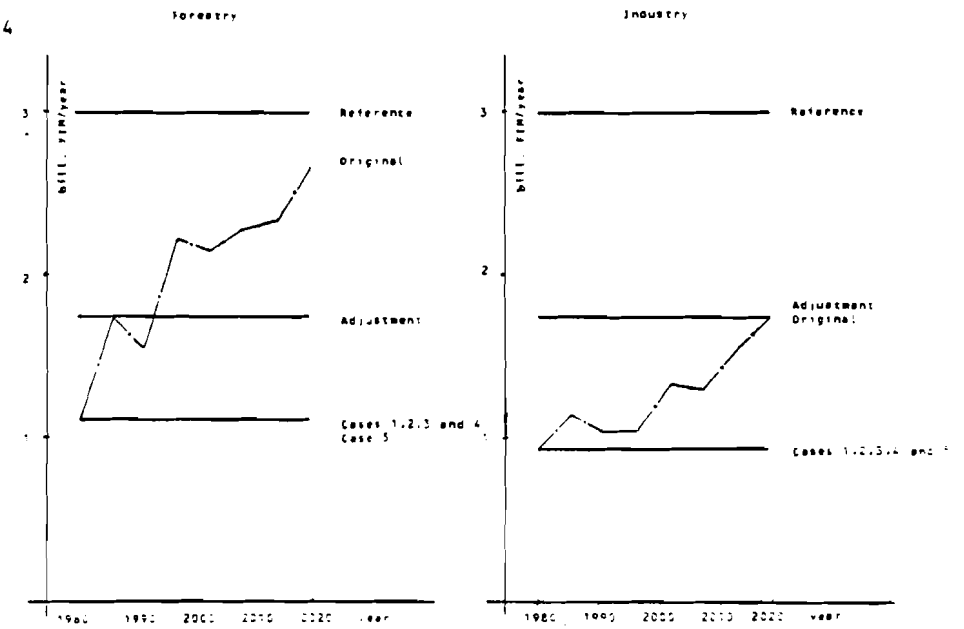
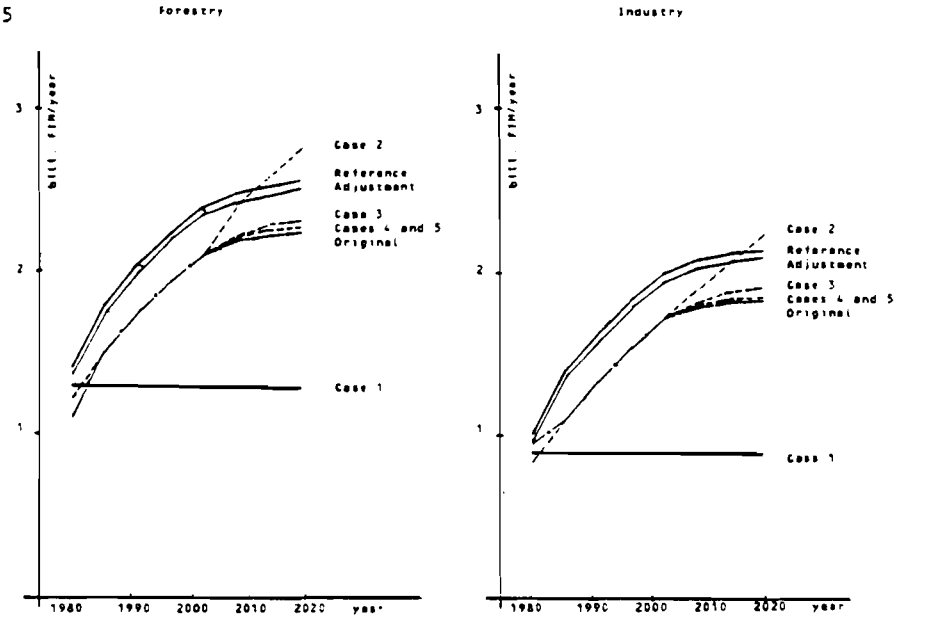


Figure 2. (Continued)

RUN 5



RUN 6

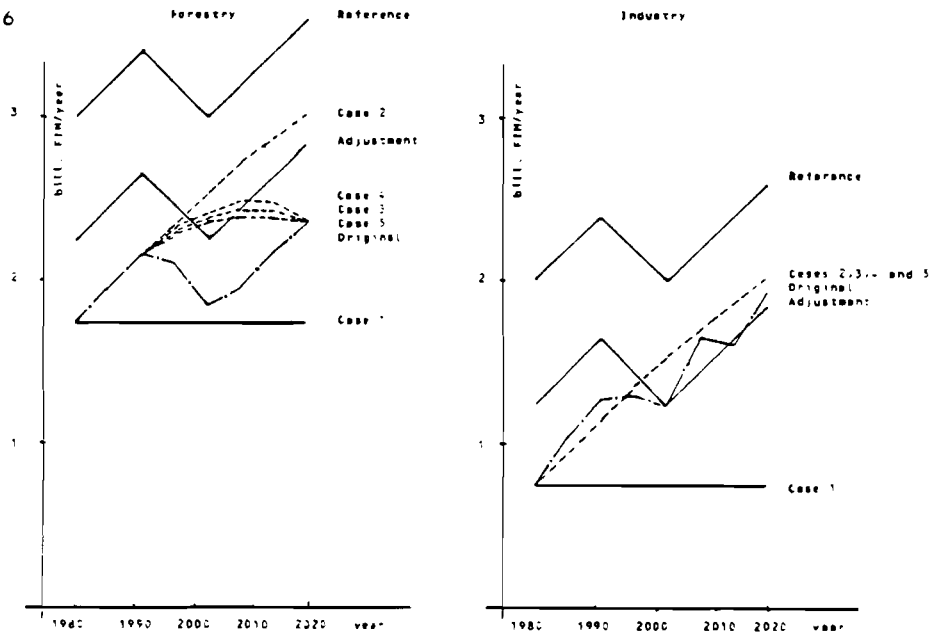


Figure 2. (Continued)

the expansion (15). We also present the original (non-smoothed) trajectories as well as the results from the experiments with the adjustment approach in Figure 2. Table 2 shows the loss of optimality resulting from the smoothing approach.

From these figures we may conclude that usually the original optimal trajectories are rather unsmooth with spikes. The trajectories from the adjustment approach behave ideally and (due to relaxation) they are superior to other trajectories. However, the generating approach yields variable results depending on the number of terms included in the truncation. The effects of the fluctuations in sin and cos terms become unsatisfacory when the number of terms becomes large. Such effects may be recognized in some of the Cases 4 and 5. Summarizing the results, three terms seem to be sufficient for the generating approach to smooth trajectories.

Table 2. The maximum deviations max - w_i for six runs and for three approaches: original nonsmoothed, generating (case 3) and adjustment approach.

Run	Original	Generating approach	Adjustment approach
1	-400	-400	-903
2	-512	-512	-926
3	310	310	6
4	2060	2060	1256
5	296	304	52
6	1244	1244	751

5 Summary and conclusions

Related to the reference point approach (in particular, for trajectory optimization), we have explored two questions: (i) how to reduce the waiting time for a new Pareto optimal solution when the reference point is changed, and (ii) how to obtain smooth objective trajectories without significant loss in Pareto optimality.

Our approach for the first question is to employ Pareto solutions obtained in preceeding iterations for constructing feasible initial solutions to the reference point optimization problem. Compared with a naive approach starting with a previous Pareto optimal basic solution considerable savings in computing time were obtained.

For smoothing we applied two approaches: one, where adjustment on trajectories is done via borrowing and saving and another where the objective is restricted to a linear combination of predetermined smooth trajectories. The latter approach is generally applicable, whereas the former one may be applied only when saving and borrowing have a meaningful real life interpretation. Both approaches were successfully demonstrated with a number of numerical runs.

References

/1/ **Kallio, M., A. Lewandowski and W. Orchard-Hays** (1980) An Implementation of the Reference Point Approach for Multiobjective Optimization. WP-80-35. Laxenburg, Austria: International Institute for Applied Systems Analysis.

/2/ **Murtagh, B. and M. Saunders** (1978) Large-Scale Linearly Constrained Optimization. Mathematical Programming 14 41-72.

/3/ **Roy, B. and P. Vincke** (1981) Multicriteria Analysis: Survey and New Directions. European Journal of Operational Research 8 207-218.

/4/ **Wierzbicki, A.** (1979) A Methodological Guide to Multiobjective Optimization. WP-79-122. Laxenburg, Austria: International Institute for Applied Systems Analysis.

/5/ **Wierzbicki, A.** (1980) A Mathematical Basis for Satisficing Decision Making. WP-80-90. Laxenburg, Austria: International Institute for Applied Systems Analysis.

USE OF THE REFERENCE LEVEL APPROACH FOR THE GENERATION OF EFFICIENT ENERGY SUPPLY STRATEGIES

M. Grauer[1], A. Lewandowski[2], and L. Schrattenholzer[3]

[1] *International Institute for Applied Systems Analysis, Laxenburg, Austria* and *Technical University of Leuna-Merseburg, GDR*
[2] *International Institute for Applied Systems Analysis, Laxenburg, Austria* and *Technical University of Warsaw, Poland*
[3] *International Institute for Applied Systems Analysis, Laxenburg, Austria*

INTRODUCTION

The reference level approach [1] has been shown to be an appropriate tool for studying conflicting objectives in practical decision situations [2]. A software package (Dynamic Interactive Decision Analysis and Support System or DIDASS) based on this approach has been developed at IIASA to analyze linear and non-linear multiple-criteria optimization problems.

This paper describes another experiment with the reference level approach, this time with the energy supply model MESSAGE [3]. In its original form, MESSAGE is a dynamic linear programming model with the (single) objective of minimizing the total discounted costs of meeting a set of energy demands over a given time horizon. The experiment described here shows that it is possible to take into account more than one objective and thus to study the interplay between costs and other factors such as import dependence, the need to develop infrastructure, and so on.

The main purpose of this paper is to describe the use of a new methodology; the data defining the MESSAGE run serve only to illustrate the method and their policy implications are therefore not discussed here.

PROBLEM FORMULATION

To test whether the reference level approach could be used to generate efficient energy policies, we used the energy supply model MESSAGE to study energy supply policies for the countries of the European Economic Community (EEC) [4] over the period 1980-2030. The main aim of the model is to meet the predicted demand for secondary energy by manipulating the vector of annual consumption of resources, the vector of energy production, and the annual increase in energy-producing capacity. The feasible set is determined mainly by strategies for the supply of primary energy resources via a menu of possible technologies (see Figure 1).

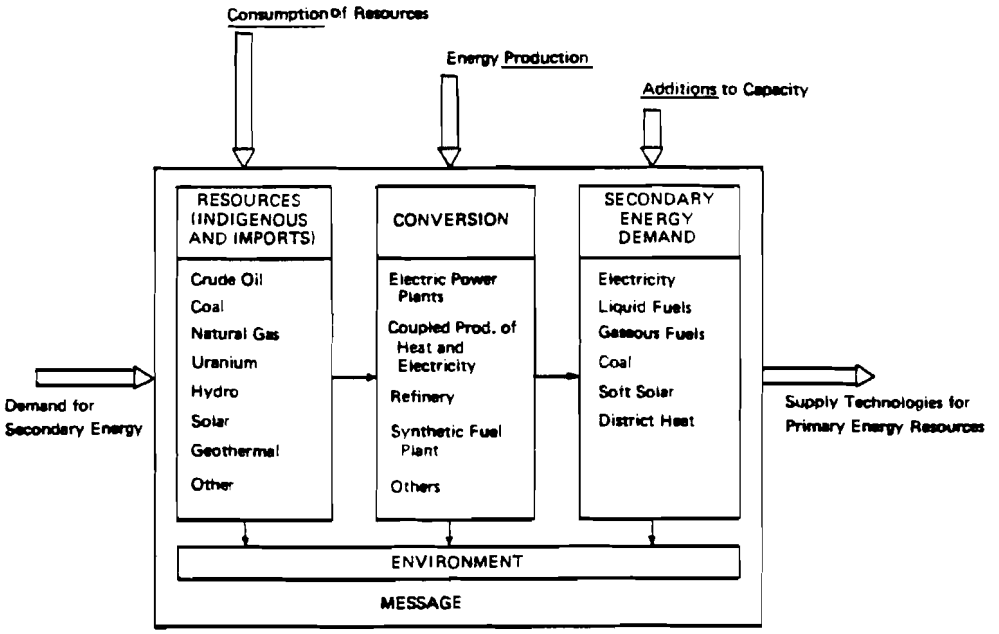


Figure 1. Schema of the energy supply model MESSAGE.

The resulting optimization problem can be formulated as a standard dynamic linear program. A detailed description of the formulation is given in the Appendix.

1. State Equations:

$$y(t+1) = \sum_{i=1}^{\nu} \bar{A}(t-n_i)\, y(t-n_i) + \sum_{j=1}^{\mu} \bar{B}(t-m_j)u(t-m_j) \qquad (1)$$

where:

$t = 0,1,\ldots,T-1$

y is a vector of state variables

u is a vector of control variables

$\bar{A}, \bar{B}$ are matrices of input data,

$(n_1, \ldots, n_\nu), (m_1, \ldots, m_\mu)$ are sets of integers which characterize time lags in state and/or control variables

T is the length of the planning period

2. Constraints:

$$\bar{G}(t)y(t) + \bar{D}(t)u(t) \leq f(t) \qquad (2)$$

where:

$t = 0,1,\ldots,T$

$\bar{G}, \bar{D}$ are matrices of input data,

f is a vector of input data

3. Bounds:

Upper and lower bounds on the control variables $u(t)$ and on the state variables $y(t)$ are also specified.

$$L(t) \leq \begin{bmatrix} u(t) \\ y(t) \end{bmatrix} \leq U(t) \qquad (3)$$

where $t = 0, 1, \ldots, T$.

4. Planning Period:

The planning period is fixed (T) and the initial state of the energy system is also given:

$$y(0) = y^0 \tag{4}$$

5. Criteria Functions:

The performance function for the scalar case has the general form:

$$J(u) = (a(T), y(T)) + \sum_{t=0}^{T-1} \left[(a(t), y(t)) + (b(t), u(t)) \right] \tag{5}$$

where a and b are input vectors.

The following scalar objective function, which reflects the total discounted costs of energy supply, was originally used in MESSAGE [5]:

$$J(u(t)) = \sum_{t=1}^{T} \left\{ \beta_1(t)(\alpha_1(t), x(t)) + \beta_2(t)(\alpha_2(t), z(t)) + \beta_3(t)(\alpha_3(t), r(t)) \right\} \rightarrow \min \tag{6}$$

where:

$T = 11$

$J(u(t)) = J(x(t), z(t), r(t))$

$x(t)$ is the vector of energy production

$z(t)$ is the vector of annual increase in energy-producing capacity

$r(t)$ is the vector of annual consumption of resources

β_i are discount factors

α_i are vectors of annual cost coefficients

The solution of eqn. (6) under conditions (1) - (4) will be described as problem S.

To improve our analysis of the decision situation we decided not simply to minimize a single aggregated function at the end of the planning period but to

minimize the trajectory of certain interesting criteria. As a test we considered the problem of simultaneous minimization of the undiscounted costs $J_{cost}(t)$, the amount of coal extracted $r_{coal}(t)$, and the volume of oil imported $r_{oil}(t)$ for each time period. This leads to the following vector of 33 criteria:

$$\begin{bmatrix} J_{cost}(t) \; ; \; t=1,2,\ldots,11 \\ r_{coal}(t) \; ; \; t=1,2,\ldots,11 \\ r_{oil}(t) \; ; \; t=1,2,\ldots,11 \end{bmatrix} \tag{7}$$

where:

$$J_{cost}(t) = \left\{ (\alpha_1(t), x(t)) + (\alpha_2(t), z(t)) + (\alpha_3(t), r(t)) \right\}$$

$r_{coal}(t)$ and $r_{oil}(t)$ are subvectors of the vector $r(t)$.

The minimization of vector (7) under constraints (1)-(4) will be described as problem M1. This represents a situation in which we wish to minimize both current costs and the use of fossil fuels in the production of energy. We also analyzed a slightly different problem in which both the overall costs (6) and the amount of coal extracted and oil imported are minimized. This gives an objective vector with 23 components:

$$\begin{bmatrix} J(u) \\ r_{coal}(t) \; ; \; t=1,2,\ldots,11 \\ r_{oil}(t) \; ; \; t=1,2,\ldots,11 \end{bmatrix} \tag{8}$$

The minimization of vector (8) under constraints (1)-(4) will be denoted as problem M2.

The general mathematical formulation of the linear multiple-criteria problems M1 and M2 discussed above is as follows:

Let A be in $R^{m \times n}$, C in $R^{p \times n}$, and b in R^m. If q is the vector of criteria (such as (7) or (8)) and x the joint vector of states y and controls u:

$$\begin{aligned} Cx &= q \rightarrow \min \\ Ax &= b \\ x &\geq 0 \end{aligned} \tag{9}$$

The reference or aspiration level approach is then used to analyze problem (9)

REFERENCE LEVEL APPROACH

The reference (or aspiration) level or trajectory is a suggestion $\bar{q}$ made by the decision maker reflecting in some sense the outcomes desired by him, in this case the trajectory of oil imported, coal extracted, and costs over the planning period 1980-2030 According to Wierzbicki [1], we must first define a partial ordering in the objective space that corresponds to the nature of the problem This means that for two trajectories q_A and q_B we may say for example that trajectory q_A is not worse than q_B if, $q_A(t) \leq q_B(t)$ for all $t \in [0,T]$ When specifying the reference trajectory $\bar{q}$ we introduce a relative ordering in the objective space - we can determine which trajectories are better or worse as than given reference trajectory $\bar{q}$ (see Figure 2) There are, of course, trajectories that are neither better nor worse

The reference trajectory optimization problem can then be formulated as follows

Given the reference trajectory $\bar{q}$, find a Pareto-optimal trajectory $\hat{q}$ which is attainable and in some sense related to the reference trajectory $\bar{q}$

In principle , two situations can arise :

(a) *Reference trajectory $\bar{q}$ is attainable , i.e., there is an admissible decision q for which $q = \bar{q}$ (i.e., there is a feasible x for which $Cx = q$)*

(b) *Reference trajectory $\bar{q}$ is not attainable, i.e., for every admissible decision q is unequal to $\bar{q}$*

Figure 3 illustrates the two situations (a) and (b) for the static two-dimensional case In problems (7) and (8), the dimensionality of the problem is increased according to the number of time steps

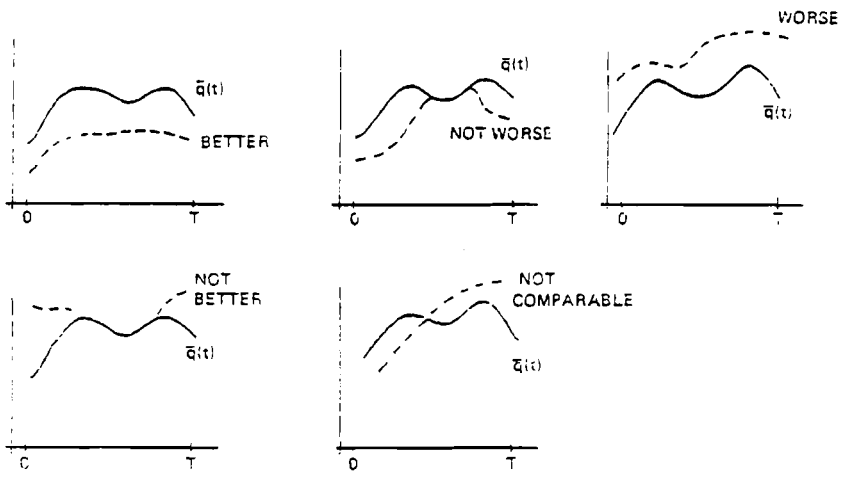


Figure 2. Ordering in the trajectory space

Since the reference trajectory expresses the outcomes that are desirable for the decision maker, it is reasonable to propose the following solution:

(I) *It is reasonable to require that the method proposes only non-improvable decisions , i.e only such objectives $\hat{q}$ that the set of attainable objectives better than $\hat{q}$ is empty (objectives in the Pareto-set, dashed line in Figure 3)*

(II) *In the case (a) it is reasonable to improve all components of the perfomance vector as much as possible but in a sense equitably , that is to maximize a "utility" $-s(q-\bar{q})$ of improving q over $\bar{q}$*

(III) *In the case (b) it is reasonable to find the attainable objective in the Pareto-set that is in a sense "nearest" to $\bar{q}$, that is to minimize a "distance" $s(q-\bar{q})$ for all $q \in \Omega_p$.*

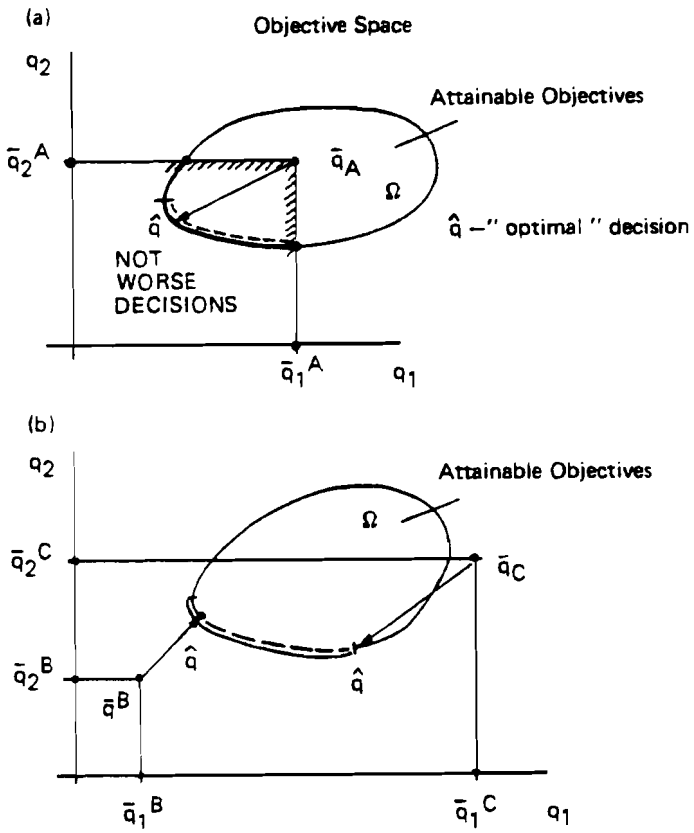


Figure 3. Interpretation of the reference level approach in the objective space ($\bar{q}_A$ is an attainable reference point, $\bar{q}_B$ and $\bar{q}_C$ are unattainable reference points).

The basic technical problem is to determine whether the situation is represented by case (a) or case (b) for a specified $\bar{q}$. To avoid this difficulty, the concept of an achievement scalarizing function has been introduced by Wierzbicki. The properties of the achievement scalarizing function are such that the result of the minimization

$$\min_{q \in \Omega} \; s(q - \bar{q})$$

satisfies all the requirements (I)-(III) specified above. The general properties of such functions are discussed by Wierzbicki elsewhere [1], [6] and [7].

The following form of the achievement scalarizing function $s(q-\bar{q})$ has the advantage that minimization results in a linear programming formulation [8]:

$$s(q-\bar{q}) = -\min_{q \in Q}\left\{ \rho \min_i (q_i-\bar{q}_i) + \sum_{i=1}^{p}(q_i-\bar{q}_i) \right\} - \sum_{i=1}^{p}\varepsilon_i(q_i-\bar{q}_i) \qquad (10)$$

Here ρ is an arbitrary penalty coefficient which is greater than or equal to p and $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p)$ is a nonnegative vector of parameters (this guarantees strict Pareto-optimality if $\varepsilon > 0$).

We also define $w_i = (q_i - \bar{q}_i)/\gamma_i$ for $i = 1, 2, \ldots, p$ where $\gamma_i \approx \bar{q}_i$ and γ_i (different from 0) is a scaling factor, chosen by the user. This scaling factor is introduced not in order to weight different objectives, but to make their influence independent of their physical units and their scale.

The set $S_{\hat{s}}(\bar{q}) = \left\{ q : s(w) \geq \hat{s}, \; w = (q-\bar{q})\gamma^{-1} \right\}$, for a given scalar $\hat{s}$, is called the level set of the scalarizing function, here γ is a diagonal matrix of scaling factors γ_i. The influence of scaling factors is illustrated in Figure 4 for function (10) and the case $\rho = p$, $\varepsilon = 0$.

Using these definitions, the problem of minimizing of (10) over the attainable points q can be formulated as a linear programming problem. For this we denote $w = (q-\bar{q})\gamma^{-1} = (Cx-\bar{q})\gamma^{-1}$ and introduce an auxiliary decision variable $y = z + \varepsilon w$. The resulting LP is

$$\min s(w) = \min_{\substack{w \in W \\ y \in R}} \left\{ y - \varepsilon w : -y + \rho w_i \leq 0, \text{ for all } i, \; -y - \sum_i w_i \leq 0 \right\} \qquad (11)$$

where $W = \left\{ w : -\gamma w + Cx = \bar{q}, \; Ax = b, \; x \geq 0 \right\}$ is the feasible set for w. This problem can be solved using any commercial LP system.

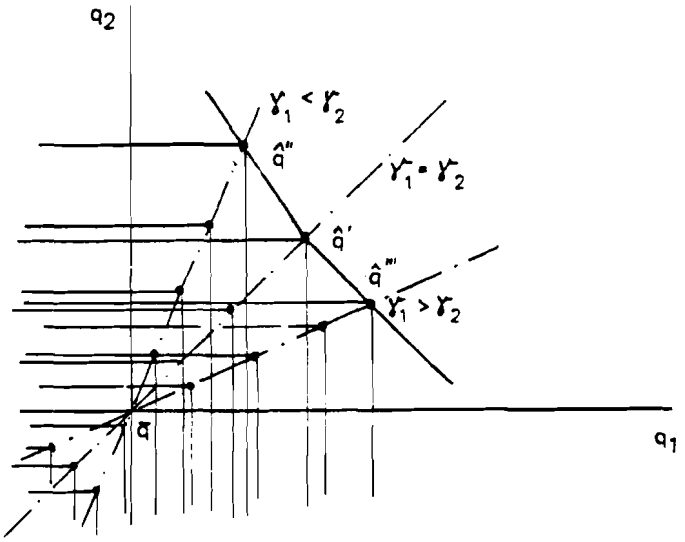


Figure 4. Level sets for achievement scalarizing function (10) for $\varepsilon=0$, $\rho=p$, and various scaling factors.

COMPUTER IMPLEMENTATION

The software for the energy supply model MESSAGE [3] has been combined with the DIDASS package for linear multiple-criteria reference point optimization to produce a system capable of solving the problems outlined above. The combined structure of the energy model and the multiple-criteria software is given in Figure 5.

The aim of Figure 5 is to explain how a model (e.g., the energy supply model) may be used in conjunction with an interactive multiple-criteria analysis procedure. The left-hand side of Figure 5 gives the usual stages in a computer

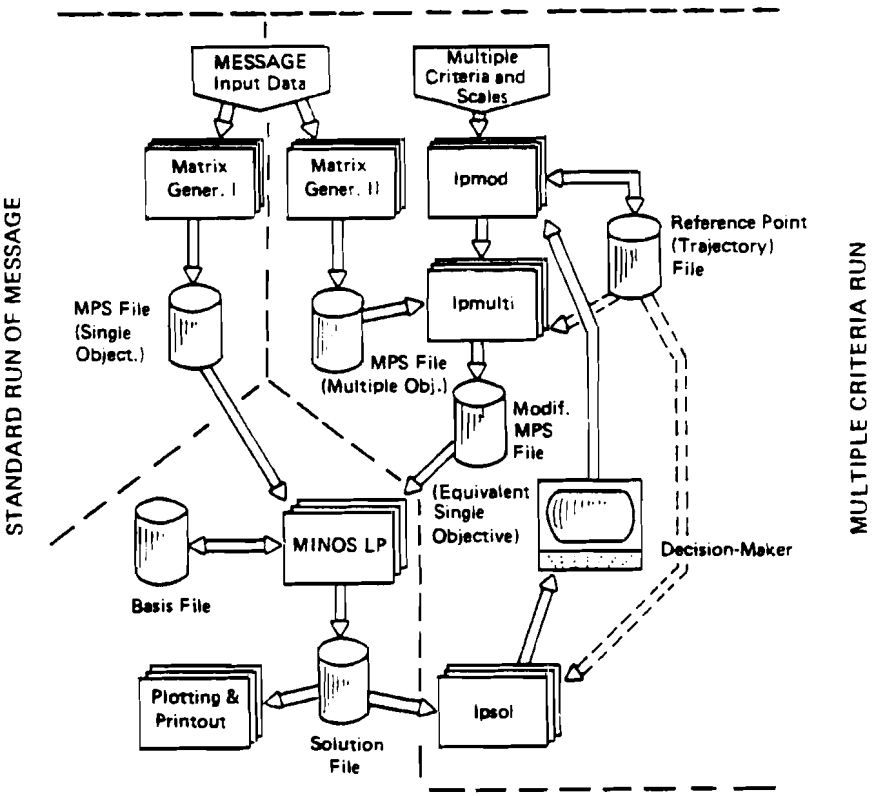


Figure 5. The combined structure of the energy model and the DIDASS package for the interactive generation of efficient energy supply strategies.

run of MESSAGE. In the combined case, however, the MPS input format file must be prepared according to the formulation of the multiple-criteria problem: for large models such as MESSAGE, the original matrix generator must be altered (Matrix Gener. II) to modify the MPS input format file in this way.

The right-hand side of Figure 5 illustrates the multiple-criteria optimization procedure. This begins with an interactive "editor" which is used to define the trajectories of various criteria and to manipulate the reference trajectories and the scaling factors (lpmod).

In the next step, the preprocessor (lpmulti in Figure 5) converts the prepared MPS input format file into a single-criterion equivalent (11). This single-criterion problem is solved with the MINOS system [9]. A postprocessor (lpsol in Figure 5) extracts selected information from the LP system output file, computes the values of the objectives and displays the information to the decision maker. The decision maker can then change the reference trajectories on the basis of this information, and possibly on the basis of experience gained in previous sessions, to generate new efficient energy supply strategies which he can analyze in the next iteration

COMPUTATIONAL EXPERIENCE

We tested the combined software by applying it to one of several scenario runs for the EEC-countries [4] for the planning period 1980-2030 under the conditions of problems M1 and M2.

The first main result was that it was necessary to scale the components of the objective vector so that the numerical values of the components are of the same order of magnitude (independent of their physical unit). If this is not done the solution of (11) is dominated by the trajectory whose components have the largest numerical values and the other trajectories are virtually insensitive to changes in their reference trajectories.

In problem M1 we experimented with different scaling factors for the cost terms because the numerical values for coal extraction and oil imports are of the same order of magnitude, while the figures for costs are greater by a factor

of 10^4. We therefore used three different sets of scaling factors for the first eleven components of vector (7)

(I) $\gamma_1 = \gamma_2 = \cdots = \gamma_{11} = 10^2$; $\gamma_{12} = \gamma_{13} = \cdots = \gamma_{33} = 1$

(II) $\gamma_1 = \gamma_2 = \cdots = \gamma_{11} = 10^3$; $\gamma_{12} = \gamma_{13} = \cdots = \gamma_{33} = 1$

(III) $\gamma_1 = \gamma_2 = \cdots = \gamma_{11} = 10^4$; $\gamma_{12} = \gamma_{13} = \cdots = \gamma_{33} = 1$

The problem M1 is solved for the three sets of scaling factors (I-III) and for given reference trajectories for costs, coal extraction, and oil imports. Figure 6 illustrates the reference trajectories and the corresponding efficient trajectories (Response) obtained in each of the three cases.

For case (I) the coal and oil trajectories (Figure 6b, 6c) are affected only very slightly by the corresponding reference trajectories, the coal response even reaching the upper bound (Figure 6b). The solution is fully dominated by the cost response and follows the cost reference trajectory. Increasing the values $\gamma_1 = \gamma_2 = \cdots = \gamma_{11}$ reduces the influence of the cost terms, and for case (III) the coal and oil responses follow the corresponding reference trajectories exactly, with a slight vertical displacement (see Figure 6b, 6c).

The trajectories s in Figure 6 indicate the solution of problem S with the scalar objective function (6) - it is interesting to compare this with the multicriteria case.

The problem described above consists of 711 rows and 761 columns. One run of the equivalent single LP problem on a VAX without an old basis from a previous session takes about 90 min CPU time, if an old basis is available the LP solution takes between 25 sec and 12 min CPU time. Using the current version of the preprocessor (lpmulti), the modification of the MPS input format file takes from 47 sec to 51 sec CPU time.

We also analyzed problem M2 using the new software. Figure 7 presents the

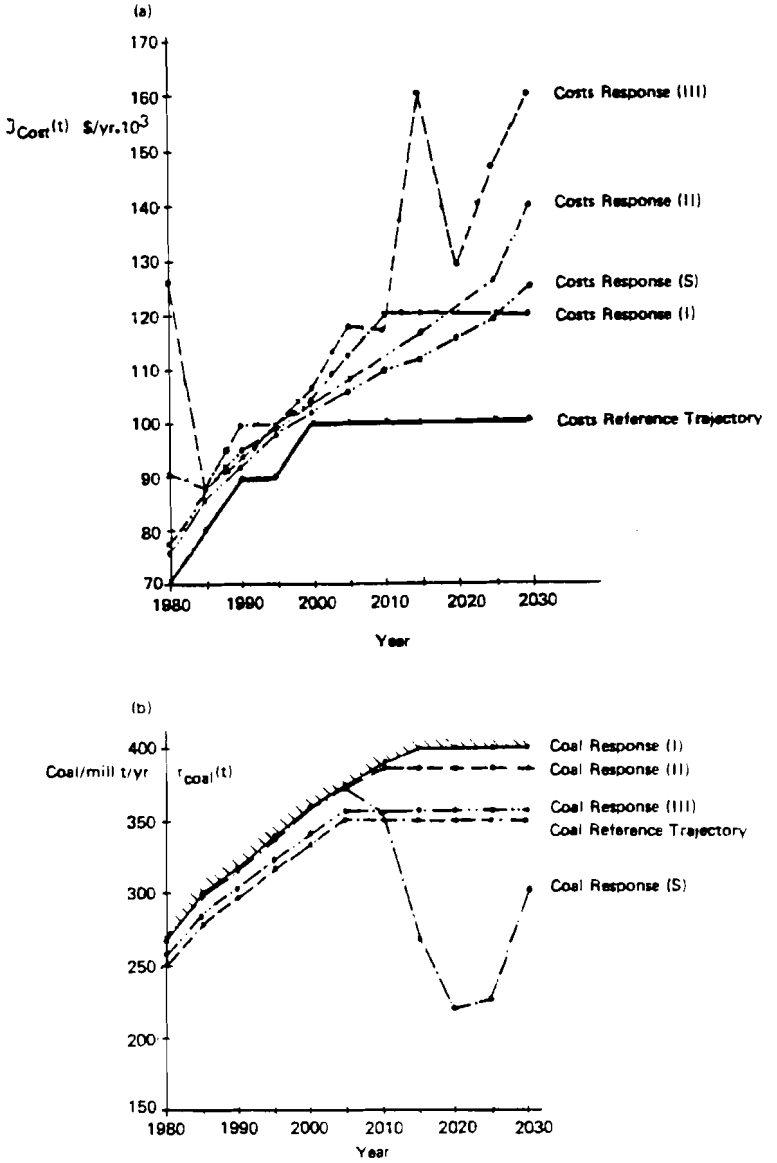


Figure 6. Experiments with different scaling factors γ_i for the cost terms (see set (I)-(III)) in problem M1 with $p = \rho = 33$, $\varepsilon = 10^{-4}$. Here s is the solution of problem S with the scalar objective function (6), given for comparison.

(a) Trajectories for the undiscounted costs $J_{cost}(t)$;

(b) Trajectories for the use of coal $r_{coal}(t)$;

(c) Trajectories of oil import policies $r_{oil}(t)$.

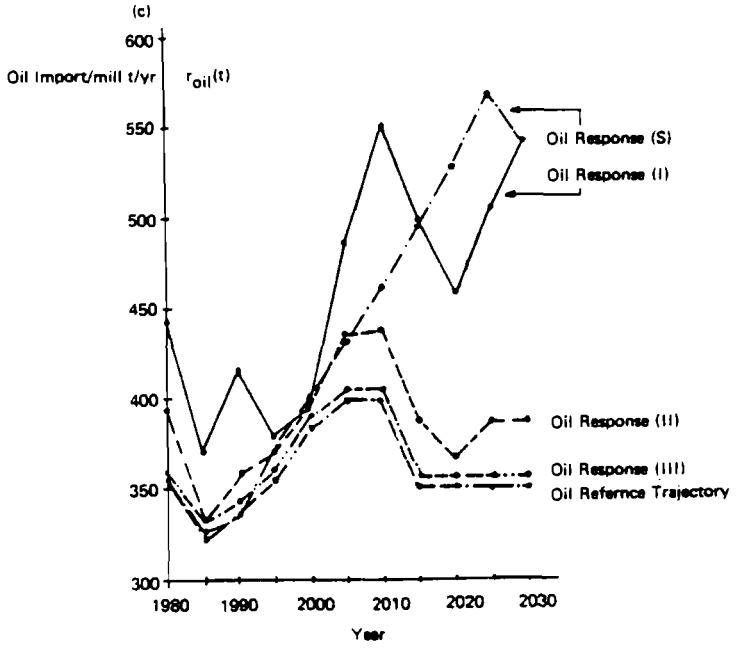


Figure 6. (continued)

results obtained for M2 assuming the same reference trajectories for coal extraction and oil imports as in problem M1 (see Figure 6b, 6c). The scaling factors corresponding to vector (8) are as follows: $\gamma_1 = 10^4$, $\gamma_2 = \gamma_3 = \cdots = \gamma_{23} = 1$. The reference point for the cost function is the scalar solution (s), which is also illustrated for the other objectives. The reference trajectories can be interpreted as follows. After a transition period ending in 2015, the decision maker wishes oil imports to a level off at 350 mill. t/year and coal extraction to remain approximately constant just below the upper bound. The reference point for the overall cost of supplying energy is assumed to be given by the scalar solution. At the scale used in Figure 7, the responses of the efficient trajectories for coal and oil appear to be identical with the reference trajectories; they actually

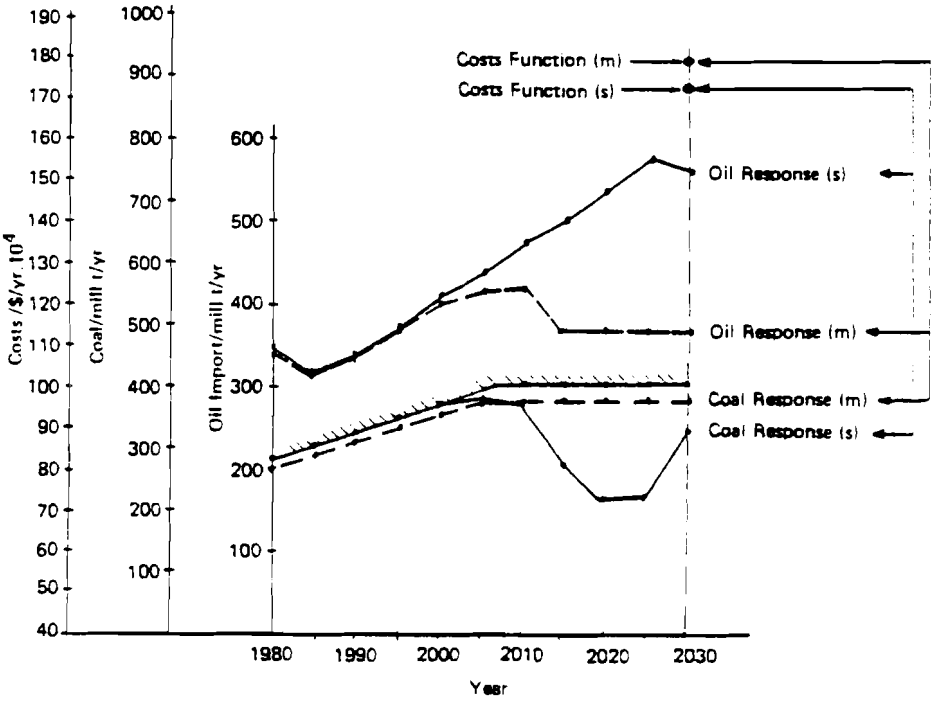


Figure 7. Efficient trajectories for the problem M2

differ by a constant value of approximately 1%. The resulting overall costs are of course higher than in the scalar case. After studying the solution on the basis of plots and printouts the decision maker will either be satisfied with this strategy or he will not; if he is not satisfied he should change the reference trajectories and/or the scaling factors before starting the next session.

SUMMARY AND CONCLUSIONS

This application has once again shown the reference level approach to be a useful tool for analyzing situations with conflicting objectives. In addition the program package DIDASS seems to be flexible enough to allow good control of the behavior of the attainable trajectories.

Further work should be done to improve the "user-friendliness" of the software. There are three ways of achieving this:

- speeding up the modification of the MPS input format file by improving the preprocessor (lpmulti)
- speeding up the solution of the equivalent LP problem
- including the history of the interactive decision-making process by displaying the sequences of references and obtained objectives visually

ACKNOWLEDGMENTS

The authors wish to thank Andrzej Wierzbicki for the initialization of this work and very helpful comments.

For the editing for grammar and consistency we want to thank Helen Gasking

References

1. A. Wierzbicki, "The use of reference objectives in multiobjective optimization. Theoretical implications and practical experiences.," WP-79-66, IIASA (1979).

2. M. Kallio, A. Lewandowski , and W. Orchard-Hays , "An implementation of the reference point approach for multiobjective optimization.," WP-80-35, IIASA (1980).

3. L. Schrattenholzer, "The energy supply model MESSAGE ," RR-81-31, IIASA (1981)

4. W. Sassin, A. Hoelzl, H.H. Rogner, and L. Schrattenholzer, "The long term energy problem: R/D strategies," SR-82-, IIASA (1982).

5. P.S. Basile, "The IIASA set of Energy Models : Its Design and Application," RR-80-31, IIASA (1980).

6. A. Wierzbicki, "A mathematical basis for satisficing decision making ," WP-80-90, IIASA (1980).

7. A. Wierzbicki, "Multiobjective trajectory optimization and semiregularization ," WP-80-179, IIASA (1980).

8. A. Lewandowski, "A program package for linear multiple criteria reference point optimization. Short User Manual ," WP-81- , IIASA (1981).

9. B.A. Murtagh and M.A. Saunders, "Minos/Augmented," Technical Report SOL-80-14, Systems Optimization Laboratory, Stanford University (1980).

APPENDIX

This appendix gives examples of equations of the type (1) and (2) taken from the energy supply model.

1. State Equation

Capacities of Technologies:

$$c(t)=c(t-1)+5z(t)-5z(t-6) \quad , \quad t=1,2,\ldots,11$$

where:

z is the vector of annual additions to capacity

$t-6$ reflects a 30-year service life

Resource Balances:

$$s(t)=s(t-1)-5r(t) \quad , \quad t=1,2,\ldots,11$$

where:

s is a vector of reserves (stocks) of primary energy carriers or man-made fuels

r is a vector of annual consumptions of primary energy

II. Constraints:

Demand/Supply Balance:

$$Dx(t) \geq d(t) + H(t) \quad , \quad t = 1,2,\ldots,11$$

where

D is a matrix describing supply/demand paths

x is a vector of annual supply activities

d is a vector of annual secondary energy demand (exogenous inputs)

H is a matrix of coefficients for secondary energy inputs to technologies

Capacity Utilization

$$B_i x(t) \leq c(t) \quad , \quad i=1,2,\ldots,n \quad t=1,2,\ldots,11$$

where

B_i are matrices defining load regions and the availability of technologies in the load regions, ?i = 1,2,...,n? (input data)

Build-Up Constraint:

$$z(t) \leq \delta z(t-1) - g \quad , \quad t=1,2,\ldots,11$$

where

δ is a diagonal matrix of growth parameters (input)

g is a vector of startup values allowing z to reach positive values from zero

Build-Up Constraint:

$$\sum_{i \in J_1} z_i(t) \leq GUB(t)$$

where

$GUB(t)$ is a vector of absolute upper limits (input data)

J_1 is a subset of the set of technologies

Resource Consumption:

$$G\,r(t) \geq Q_1 x(t) + Q_2 z(t) - Q_3 z(t-6)$$

where

G is a binary matrix which aggregates resource categories

Q_1, Q_2, Q_3 are matrices of parameters describing the specific consumption of resources by conversion technologies (input)

Resource Extraction

$$G_1 r(t) \leq p(t) \quad , \quad t = 1,2,\ldots,11$$

where

G_1 is a matrix for aggregating indigenous resource categories (input data)

p is a vector of annual production limits for each type of resource (exogenous inputs)

DYNAMIC POLYOPTIMAL CONTROL OF MULTISTAGE PROCESSES

J. Gutenbaum and A. Łodziński

Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

1. Introduction

From observations of living organisms a conclusion can be drawn that their behaviour is determined by mutability and hierarchy of objectives. At the upper levels of this hierarchy the objectives are genetically determined by mechanisms like preservation of species or "preservation of life of an individual". At lower levels the objectives determine particular modes of behaviour and specific decisions. Those objectives are subjected to continual changes.

The upper levels exert effect on lower ones by imposing constraints on the freedom of their operation. A characteristic phenomenon occurs: the lower level of the hierarchy of objectives is, the more frequent their changes are.

Concepts fundamental for the construction of adaptive algorithms for decision making, though based on imitation of activities of living organisms, do not take into account that aspect of their behaviour. Those ideas merely took into consideration the phenomenon of adaptation of decision making algorithms to environment changes /changes of a mathematical model/ while an objective /or objectives/ of performance was considered invariant. There are numerous examples of control processes, whose objectives /performance indices/ as well as descriptions /mathematical models/ vary according to conditions, at which these processes occur.

It is obvious that changes of objective functions should result in changes of control algorithms. Systems making it possible to realize such a relationship form the class of generalized adaptive systems. Their control algorithms should adapt not only to disturb-

ances having the impact upon the process description but also to the changes of objectives.

Processes of this type include cyclic ones, consisting of some distinguished stages e.g. starting up and regular operation. The process of steel production in an arc furnace, with the melting and refining stages [1] distinguished, furnishes an example of such a process. Another example is provided by the process of economic growth, in which the stages of investment, production and consumption are distinguished. Other type of processes belonging to the class considered is represented by continuous processes for which changes of situation to be followed by changes of objective functions are of random nature. These processes can be exemplified by control of water systems, in which the occurrence of situations such as flood or drought is associated with the intensity of rainfalls, having of course the random character [3], [4].

Processes, for which the phenomenon of objective function changes performed to suit variations in situation are taken into account, are called multistage ones. Particular stages differ in objective function and mathematical model used. Processes under discussion are dynamic /processes with memory/. Hence stages cannot be investigated independently. The global solution related to all the stages, has to take into account the necessity of trade-offs among local problems defined for individual stages. Hence, a poly-optimization problem is coming into play. However, problems considered involve dynamic changes of particular objective functions, as opposed to conventional polyoptimization problems, for which it is assumed that all the objective functions are applied simultaneously. It should be noted that there are cases in which not only one objective function but a set of them is to be used at every stage.

2. Formulation of the problem /the deterministic case/

Assume that a given control process can be divided into N stages $N > 1$ and the i-th stage $(i=1,2,...N)$ is described by the following state equations

$$\dot{x}_i(t) = f_i\left[x(t),\ u(t)\right];\quad t \in \left[T_{i-1}, T_i\right);\quad x_{i-1} = x(T_{i-1}) \tag{1}$$

$x(T_o) = x_o$ - the initial state

$$\left(x_i,\ T_i\right) \in S_i \tag{2}$$

S_i - the target set for the i-th stage

$u(t) \in U_i$ - the set of admissible values of the control variables at the i-th stage

It is assumed that the components of the vector u are L^2 functions

$$u \in L^2\left(\left[T_{i-1}, T_i\right),\ R^m\right);\quad i=1,2,\ldots,N \tag{3}$$

and the components of the vector x are absolutely continuous function with derivatives of the L^2 class

$$x \in W_1^2\left(\left[T_o,\ T_N\right],\ R^n\right) \tag{4}$$

For each stage $i=1,\ldots,N$ an objective function is formed

$$q_i = Q_i\left[\bar{x}_{i-1},\ \bar{x}_i,\ u(t)\right] \tag{5}$$

$$\bar{x}_i = \left(x_i,\ T_i\right) \in S_i \qquad u(t) \in U_i$$

The problem considered consists in determining the optimal /polyoptimal/ control for a given time interval $t \in \left[T_o,\ T_N\right]$. The formulation of this optimization problem is illustrated with Fig. 1.

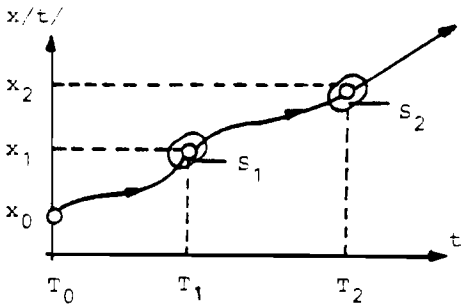


Fig. 1

3. Solution procedure

To solve the problem the two-level structure shown in Fig. 2 is used

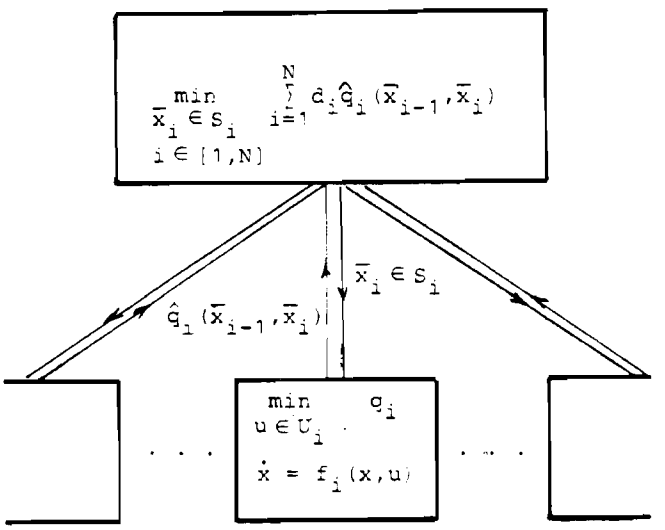


Fig. 2

At the lower level the following dynamic problem is solved:

For given $\bar{x}_{i-1} \in S_{i-1}$, $\bar{x}_i \in S_i$; $i=1,\ldots,N$ the locally optimal control is to be determined

$$\hat{u}_i(\bar{x}_{i-1}, \bar{x}_i, t) = \arg \min_{u \in U_i} q_i \qquad (6)$$

subject to the state equations given by (1) and $t \in [T_{i-1}, T_i)$

As a result the locally optimal control is derived. It is given in terms of "switching points" $\bar{x}_{i-1}$, $\bar{x}_i$. Moreover, the values of the objective function obtained for these controls are as follows

$$\hat{q}_i(\bar{x}_{i-1}, \bar{x}_i) = Q_i[\bar{x}_{i-1}, \bar{x}_i, \hat{u}_i(\bar{x}_{i-1}, \bar{x}_i, t)] \qquad (7)$$

$$t \in [T_{i-1}, T_i)$$

At the upper level the polyoptimization problem is solved. It is formulated with respect to the variables $\bar{x}_i \in S_i$; $i=1,\ldots,N$. This problem can be reduced to determining of those switching points, for which the following relation holds

$$(\hat{\bar{x}}_1, \hat{\bar{x}}_2, \ldots, \hat{\bar{x}}_N) = \arg \min_{x_i \in S_i} \sum_{i=1}^{N} d_i \hat{q}_i(\bar{x}_{i-1}, \bar{x}_i) \qquad (8)$$

$$i \in [1,N]$$

where $d_i > 0$ are the weight coefficients.

3.1. The lower-level problem /the dynamic optimization problem/

It is assumed that the switching times T_i $i=1,\ldots,N$ are known and the objective function is of the form

$$q_i = \int_{T_{i-1}}^{T_i} f_{io}(x,u,t)\,dt \qquad (9)$$

Under these assumptions the lower-level problem can be formulated as follows: for given $x_{i-1} \in S_{i-1}$ and $x_i \in S_i$ the locally optimal control is to be determined in terms of the switching points

$$\hat{u}_i(x_{i-1}, x_i\ t) = \arg \min_{u \in U_i} \int_{T_{i-1}}^{T_i} f_{i0}(x, u, t)\, dt \qquad (10)$$

subject to $\dot{x}_i = f_i(x, u)$.

It is assumed that a solution to this problem exists and is unique.

The analytic solution of such a problem can be obtained using the maximum principle.

For this purpose the Hamiltonian H is formed

$$H_i(x, u, \Psi, t) = - f_{0i}(x, u, t) + \Psi^T \cdot f_i(x, u, t) \qquad (11)$$

where Ψ are the adjoint variables.

This Hamiltonian is to be maximized with respect to $u \in U_i$. The resulting locally optimal control is a function of time, state and adjoint variables.

$$\tilde{u}(x, \Psi, t) = \arg \max_{u \in U_i} H_i(x, u, \Psi, t) \qquad (12)$$

The solution obtained is substituted into the system of canonical equations

$$\begin{aligned} \dot{x} &= f_i[x, \tilde{u}(x, \Psi, t), t] \\ \dot{\Psi} &= -\frac{\partial H_i}{\partial x}[x, \tilde{u}(x, \Psi, t), t] \end{aligned} \qquad t \in [T_{i-1}, T_i) \qquad (13)$$

This system consists of 2n equations, for which 2n conditions determining the states at the times T_{i-1} and T_i are given. Hence, from the theoretical point of riew it is possible to derive an analytic solution to the system in the form

$$\begin{aligned} \hat{x} &= g(x_{i-1}, x_i, t) \\ \hat{\Psi} &= h(x_{i-1}, x_i, t) \end{aligned} \qquad t \in [T_{i-1}, T_i) \qquad (14)$$

Substituting Equation (14) into (12), the locally optimal control is obtained in terms of time and switching parameters and the locally optimal values of the objective function in the parametric form

$$\hat{u}_i(x_{i-1}, x_i, t); \quad t \in [T_{i-1}, T_i); \quad \hat{q}_i(x_{i-1}, x_i) \qquad (15)$$

$$i = 1, \ldots, N$$

3.2. The upper-level problem /the static optimization problem/

To simplify the consideration, it is assumed that the poly-optimization problem (8) can be reduced to minimization of the global utility function. In other words

$$(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_N) = \arg \min_{\substack{x_i \in S_i \\ i \in [1,N]}} \sum_{i=1}^{N} d_i \hat{q}_i(x_{i-1}, x_i) \qquad (16)$$

are to be determined, where $d_i > 0$ are the weight coefficients.

The problem (16) can be solved using the dynamic programming approach. However, from the computational point of view the decomposition method proposed in [5] seems to be more convenient. It relies on the fact that the first term of the utility function $q(x_1, \ldots, x_N)$ depends upon the variable x_1 only, /it is assumed that the initial condition is given/, the second one upon x_1 and x_2, the third on- upon x_2 and x_3 and so on

$$q(x_1, \ldots, x_N) = d_1 \hat{q}_1(x_1) + d_2 \hat{q}_2(x_1, x_2) + \ldots + d_N \hat{q}_N(x_{N-1}, x_N) \qquad (17)$$

To accomplish the separation of individual terms of the function $q(x_1, \ldots, x_N)$, the auxiliary variables v_{i1} and v_{i2} are introduced.

$$v_{i1} = v_{i2} = x_i \qquad i=1,\ldots,N-1$$
$$v_{N1} = x_N \tag{18}$$

Substitution of the variables v_{i1} and v_{i2} for x_i in the right-hand side of Equation (17) results in

$$q(\underline{v}) = d_1 \hat{q}_1(v_{11}) + d_2 \hat{q}_2(v_{12}, v_{21}) + \ldots + d_N \hat{q}_N(v_{N-1,2}, v_{N1}) \tag{19}$$

where

$$\underline{v} = (v_{11}, v_{12}; v_{21}, v_{22}; \ldots ; v_{N-1,1}, v_{N-1,2}; v_{N1})$$

The problem of minimizing the function $q(x_1,\ldots,x_N)$, $x_i \in S_i$ is equivalent to minimization of the function $q(\underline{v})$ under the equality constraints

$$v_{i1} = v_{i2} \qquad i=1,\ldots,N-1 \tag{20}$$

and

$$v_{ij} \in S_i \qquad i=1,\ldots,N \; ; \quad j=1,2 \tag{21}$$

This new minimization problem can be solved using the Lagrange multipliers method. For this purpose, the Lagrangian L is formed

$$L(\underline{v}, \underline{\lambda}) = q(\underline{v}) + \sum_{i=1}^{N-1} \lambda_i (v_{i1} - v_{i2}) \tag{22}$$

where $\underline{\lambda} = (\lambda_1,\ldots,\lambda_{N-1})$ is the vector of Lagrange multipliers.

The Lagrangian is separable /decomposable/ with respect to the variables v_{ij}, because it can be represented in the form

$$L(\underline{v}, \underline{\lambda}) = d_1 \hat{q}_1(v_{11}) + \lambda_1 v_{11} + \sum_{i=2}^{N-1} [d_i \hat{q}_i(v_{i-1,2}, v_{i1}) - \lambda_{i-1} v_{i-1,2} + \lambda_i v_{i1}] + d_N \hat{q}_N(v_{N-1,2}, v_{N1}) \tag{23}$$

Assume that the Lagrangian (23) has a saddle point and it can be determined as a result of minimization of L with respect to $\underline{v}$ and then its maximization with respect to $\underline{\lambda}$. If this is the case, then a solution to the optimization problem is of the following form

$$\underline{v} = \arg\max \sum_{i=1}^{N} \min_{\substack{v_{i-1,2}\in S_{i-1} \\ v_{i1}\in S_i}} \left[d_i q_i\left(v_{i-1,2}, v_{i1}\right) - \lambda_{i-1} v_{i-1,2} + \lambda_i v_{i1} \right] \tag{24}$$

$$\lambda_0 = \lambda_N = 0$$

Such an approach results in the partition of the upper-level problem into local and global ones. The local problem consists in accomplishing an constrained minimization of a function of one (i=1) or two (i=2,..,N) variables. The global one involves unconstrained minimization with respect to the N-1 Lagrange multipliers. A general procedure for solving the problems mentioned is shown in Fig. 3.

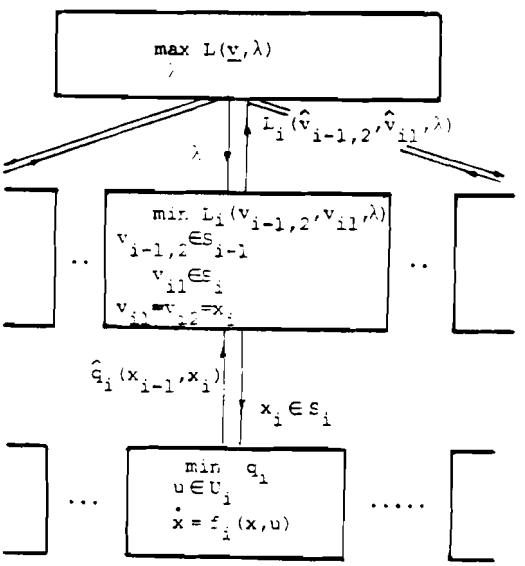


Fig. 3

Owing to the introduction of the auxiliary variables, the problem of constrained minimization of a function of N variables has been reduced to N-fold constrained minimization of a function of one or two variables /the local level/ and unconstrained minimization of a function of N-1 variables /the global level/.

For illustration purposes a simple example [5], for which an analytic solution can be obtained, is presented.

4. An example

$N = 2$

for $i = 1$:

$$q_1 = \int_{T_0}^{T_1} u^2 dt; \quad \begin{cases} \dot{x} = -x + u; \; x(0) = 0 \\ S_1 = \{ x(T_1) : x(T_1) \geq 1 \} \\ T_0 = 0, \; T_1 = 1; \; d_1 = 1 \end{cases} \tag{25}$$

for $i = 2$:

$$q_2 = \int_{T_1}^{T_2} 2u^2 dt; \quad \begin{cases} \dot{x} = u \\ S_2 = \{ x(T_2) : x(T_2) = 0 \} \\ T_2 = 2; \; d_2 = 1 \end{cases} \tag{26}$$

The sets S_1 and S_2 are shown in Fig. 4.

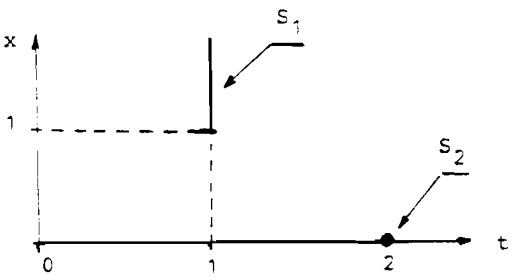


Fig. 4

The solution to the lower-level problem /the dynamic optimization problem/: i=1. The problem consists in minimization of q_1 subject to the conditions (25) and for given v_{11}

$$x(T_1) = v_{11}$$

To accomplish it, the Hamiltonian H is formed

$$H = - u^2 + \Psi(-x + u) \tag{27}$$

It reaches its maximum at

$$u = \hat{u}_1 = \Psi/2 \tag{28}$$

The canonical equations for $u = \hat{u}$ are the following

$$\begin{aligned} \dot{x} &= -x + \Psi/2 \qquad x(0) = 0 \; ; \; x(T_1) = v_{11} \\ \dot{\Psi} &= \Psi \end{aligned} \tag{29}$$

Solutions to these equations are

$$\begin{aligned} \hat{x}(t) &= v_{11}(\text{sh}1)^{-1} \text{ sh } t \\ \hat{\Psi}(t) &= 2v_{11}(\text{sh}1)^{-1} \exp t \end{aligned} \tag{30}$$

The locally optimal control can be determining using Equation (28). Performing some calculations, it is possible to derive the expression defining $\hat{q}_1$ as a function of v_{11}

$$\begin{aligned} \hat{u}_1(t) &= v_{11}(\text{sh } 1)^{-1} \exp t \\ \hat{q}_1(v_{11}) &= a\, v_{11}^2 \end{aligned} \tag{31}$$

where

$$a = \frac{e^2 - 1}{2(\text{sh } 1)^2}$$

For i = 2 and given v_{12}

$$x(T_1) = v_{12} \tag{32}$$

a similar approach is applied. It yields

$$\hat{x}(t) = 2(1 - t)v_{12} \; ; \quad \hat{u}_2(t) = - v_{12}$$
$$\hat{q}_2(v_{12}) = 2v_{12}^2 \tag{33}$$

The upper-level problem /the static optimization problem/: The Lagrangian is constructed as follows

$$L(v_{11}, v_{12}, \lambda) = \hat{q}_1(v_{11}) + \hat{q}_2(v_{12}) + \lambda(v_{11} - v_{12}) \tag{34}$$

To obtain solutions to local tasks of the upper-level problem, the Lagrangian is written in the form

$$L(v_{11}, v_{12}, \lambda) = L_1(v_{11}, \lambda) + L(v_{12}, \lambda) \tag{35}$$

where

$$L_1(v_{11}, \lambda) = a\, v_{11}^2 + \lambda\, v_{11}$$
$$L_2(v_{12}, \lambda) = 2\, v_{12}^2 - \lambda\, v_{12} \tag{36}$$

For i=1 and given λ the local task of the upper-level problem has the form

$$\hat{v}_{11} = \arg \min_{v_{11} \in S_1} L_1(v_{11}, \lambda) \; ; \quad S_1 = \{ v_{11} : 1 - v_{11} \leqslant 0 \} \tag{37}$$

It can be solved using the Kuhn-Tucker conditions. As a result one obtains

$$\hat{v}_{11}(\lambda) = \begin{cases} - \lambda / 2a \; ; & \lambda \leqslant -2a \\ 1 \quad ; & \lambda > -2a \end{cases} \tag{38}$$

$$L_1(\hat{v}_{11}, \lambda) = \begin{cases} - \lambda^2 / 4a \; ; & \lambda \leqslant -2a \\ a + \lambda \quad ; & \lambda > -2a \end{cases} \tag{39}$$

In an analogous manner the expressions for i = 2 are derived

$$\hat{v}_{12}(\lambda) = \begin{cases} 1 & ; \ \lambda \leq 4 \\ \lambda/4 & ; \ \lambda > 4 \end{cases} \tag{40}$$

$$L_2(\hat{v}_{12}, \lambda) = \begin{cases} 2 - \lambda & ; \ \lambda \leq 4 \\ \lambda^2/8 & ; \ \lambda > 4 \end{cases} \tag{41}$$

The global task of the upper-level problem consists in maximization of L with respect to λ, where

$$L(\lambda) = \begin{cases} -\lambda^2/4a + 2 - \lambda & ; \ \lambda \leq -2a \\ a + 2 & ; \ -2a < \lambda \leq 4 \\ -\lambda^2/8 + a + \lambda & ; \ \lambda > 4 \end{cases} \tag{42}$$

$L(\lambda)$ is a concave function as shown in Fig. 5 For $\lambda \in [-2a, 4]$ its maximal value is equal to $a+2$. For that interval of the values of λ we have

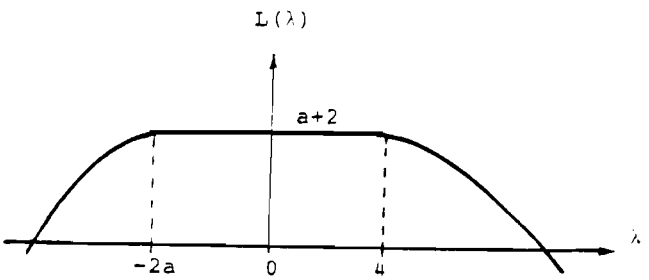


Fig. 5

$$\hat{x}(T_1) = \hat{v}_{11} = \hat{v}_{12} = 1 \tag{43}$$

Substitution of Equation (43) into (30), (31) and (33) yields the solution to the problem Fig. 6

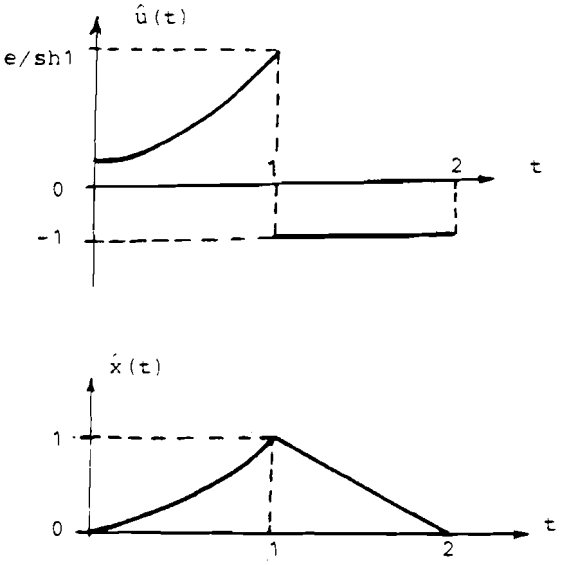


Fig. 6

$$\hat{u}(t) = \begin{cases} (\text{sh } 1)^{-1} \exp t & t \in [0,1) \\ -1 & t \in [1,2] \end{cases}$$

$$\hat{x}(t) \quad \begin{cases} (\text{sh } 1)^{-1} \text{ sh } t & t \in [0,1) \\ -t + 2 & t \in [1,2] \end{cases} \qquad (44)$$

5. Formulation of the stochastic problem

Given R objective functions, each of them corresponds to a particular situation. Moreover, it is assumed that these functions are randomly assigned to stages.

Such a formulation can be appropriate for the problem of water system control. In systems of this type the objective function depends upon situation /flood, drought, normal conditions/ and there is no a priori knowledge on situations to occur at particular stages /days, 10-day periods, or other time intervals/ in the future.

Assume that the probabilities $p(r_i)$ of the occurrence of the r-th situation at the i-th stage are known, where r_i is a realization of the random process $\Omega(i)$

$$\Omega(i) = \{ r_i \in [1,R] \} \tag{45}$$

This problem, as in the deterministic case, can be considered as a two-level one. At the lower level RN dynamic optimization tasks are to be solved /in the deterministic case only N tasks have to be solved at this level/. The solution is of the form

$$\hat{u}_{ri} = \arg \min_{u \in U_{ri}} q_{ri} \quad ; \qquad \begin{matrix} r = 1,\ldots,R \\ i = 1,\ldots,N \end{matrix} \tag{46}$$

where q_{ri} is the objective function corresponding to the occurrence of r-th situation at the i-th stage.

The upper-level problem can be reduced to minimization of the expected value of the global objective function /summed over all the stages/

$$(\hat{x}_1,\ldots,\hat{x}_N) = \arg \min_{\substack{x_i \in S_i \\ i \in [1,N]}} \sum_{i=1}^{N} \sum_{r=1}^{R} p(r_i)\hat{q}_{ri}(x_{i-1}, x_i) \tag{47}$$

where

$$\sum_{r=1}^{R} p(r_i) = 1 \; ; \quad i = 1,\ldots,N \tag{48}$$

The effectiveness of this method can be significantly increased by applying a moving optimization approach. It makes possible to take advantage of the current information on present situation and the probabilities of occurrence of situations in the future. In this case, for the current stage k $(k=1,\ldots,N)$ we have

$$\left(\hat{x}_k,\ \hat{x}_{k+1},\ldots,\hat{x}_N\right) = \arg \min_{\substack{x_i \in S_i \\ i \in [k,N]}} \Big[\hat{q}_{rk}(x_k) + \sum_{i=k+1}^{N} \sum_{r=1}^{R} p_k(r_i)\hat{q}_{ri}(x_{i-1},\ x_i)\Big] \tag{49}$$

This expression has been derived taking into account that at the current stage k the actual situation is known, $\left(p(r_k) = 1\right)$ and in the course of time the information on the probability of occurrence of the r-th situation at the i-th stage $(i > k)$ is more and more accurate. $p_k(r_i)$ denotes the probability of occurrence of the r-th situation at the i-th stage $(i > k)$, evaluated at the k-th stage.

The method presented was applied for simulation of water system control [3,4]. It was shown that a moving optimization approach makes it possible to reduce significantly the number of stages, at which flood or drought situations occur. However, it is obtained at the expense of increased computational effort, because necessary computations are to be repeated many times.

References

[1] Gutenbaum J.: Polyoptimization of systems with separate action of performance indices. Systems Science /Wrocław/ vol.3, No.3 1977

[2] Gutenbaum J.: Control of multistage processes under maximization or balanced growth of utility function. Control and Cybernetics /Warsaw/ vol. 8, No. 3 1979

[3] Gutenbaum J. et al.: Operative distribution of water resources for agriculture purposes. Proc. of the Noteć and Silistra Regional Development, Task Force Meeting, IIASA Laxenburg, May 1978

[4] Gutenbaum J., Inkielman M., Pietkiewicz-Sałdan H.: Operative control of a water system with randomly varied form of the objective function. Proc. of the 5th Polish-Italian Symposium "Applications of Systems Theory to Economics, Management and Technology", Toruń, June 1980, PWN Warsaw, pp. 295-310

[5] Łodziński A.: Optymalizacja dynamiczna procesu wieloetapowego. Doctoral Thesis, Systems Research Institute, Warsaw /under preparation/

THE QUEST FOR A CONCORDANCE BETWEEN TECHNOLOGIES AND RESOURCES AS A MULTIOBJECTIVE DECISION PROCESS

H. Górecki[1], M. Zebrowski[1], J. Kopytowski[2], and G. Dobrowolski[3]

[1] *Institute for Control and Systems Engineering in the Academy of Mining and Metallugy (ICSE-AMM), Cracow, Poland*
[2] *Institute for Industrial Chemistry (IIC), Warsaw, Poland*
[3] *Systems Research Department of the ICSE-AMM and IIC, Cracow, Poland*

INTRODUCTION

Industrial development programming is a management field which continues to stimulate the development of multiobjective decision theory. This is largely because decision problems in this area cannot be treated as single-criterion problems (such as the maximization of profit under given constraints); one reason for this is that unstable economic relations undermine the credibility of the resource estimates (in financial terms) necessary for investment. Much research has been carried out in an attempt to formulate goals in physical as well as monetary units. The resulting problems can be solved only in multidimensional space: each dimension corresponds to a resource and is treated as a criterion. This naturally leads to multiobjective decision problems. In the single-objective "classical" case, the role of the decision maker (DM) in the decision process is very important and goes far beyond problem formulation; he supervises the generation of alternatives and chooses which of the many options is to be implemented. In the multiobjective environment, however, the DM has to go much deeper into the solution of

the decision problem; he must consider the potential conflict between his preferences and the technological attainability of his goals.

There is a certain play among the resources which can be translated into constraints only after the multiobjective problem in which these relations took part as criteria has been solved. In short, this type of decision problem can be described as a quest for a concordance between technologies and resources.

Much research has been carried out in this field. One interesting way of dealing with the multiobjective decision problem is the reference point approach developed by Wierzbicki [1,2], and Kallio, Lewandowski, and Orchard-Hays [3], which has led to the creation of the computer decision support package DIDASS (Dynamic Interactive Decision Analysis and Support System).

This paper describes a method of solving multiobjective decision problems which arose from difficulties encountered in the design of production structure in the chemical industry [4]. The theoretical background was developed by Górecki [5]. We introduce the concept of the Admissible Demand Set (ADS), and a method of finding an admissible solution representing a compromise between the ADS (reflecting the expectations of the DM) and the Pareto-optimal state of the system. The result is a process-type model, described by Dobrowolski, Żebrowski, and Kopytowski in [6,7], which has been developed and used to solve problems in industrial development strategy as part of the Growth Strategy Optimization System (GSOS) [8].

One application of this model (which utilizes the reference point approach) to the generation of efficient development alternatives for the chemical industry has been described by Dobrowolski, Kopytowski, Lewandowski, and Żebrowski [9]. To illustrate the method presented in this paper we have chosen the methanol problem; the results of model runs for various admissible demand sets and a given set of methanol technologies are given in a later section.

CONCORDANCE AS A MULTIOBJECTIVE PROBLEM

Multiobjective approach

The problem of concordance mentioned above will now be described as a multiobjective decision process. The components involved in this process are as follows:

- a decision maker who has to make a final choice from the alternatives under consideration
- a set of available technologies described in terms of a model
- available resources.

In order to form a decision-making system, the components have to be related in a way that would enable the available technologies and available resources to be combined harmoniously. The pursuit of profits (potential expansion) and the availability of resources form a core from which a set of criteria may be derived and introduced into the decision-making process.

The availability of resources is determined by the availability of information from various sources as well as resulting from the knowledge and experience of experts. This may lead only to an estimated or apparent value for the availability of resources.

The model of technologies describes a technical system which transforms resources according to technological rules.

The function of the system is to help the decision maker to learn how the estimated resources and the relevant states of the model are related in criteria space. Each state of the model represents a particular subset of available technologies and also a particular level of utilization; it is assumed that each state belongs to the Pareto-optimal set in criteria space, and that the resource estimate constitutes an area in this space.

Thus, the quest for a satisfactory concordance between technologies and resources becomes an analysis of geometrical relations in criteria space.

This approach provides the decision maker with the freedom necessary to alter both the set of technologies and the available (or assumed) quantities of resources. These alterations can be prepared in the form of wishes with the help of experts; during this process, some of these wishes will be converted into goals, with the remainder being either eliminated or postponed.

In the environment described above, the decision maker has to deal with an "apparent" availability of resources and a model of technologies which also gives him estimates concerning the behavior of a hypothetical plant. However, the reliability of these two types of information is not the same. For example, a forecast concerning the availability of oil and its price is usually less reliable than technological information on a methanol plant based on a newly developed process. Consequently, it is more realistic to assume that the technology is described relatively precisely, while resources may be expressed as an availability range. Moreover, the procedure enables the decision maker to set levels of acceptability for the estimated values. This is particularly important since it reflects a subjective factor in the decision-making process. From the various estimates the decision maker can build the minimax union shown in the two-dimensional case in Figure 1.

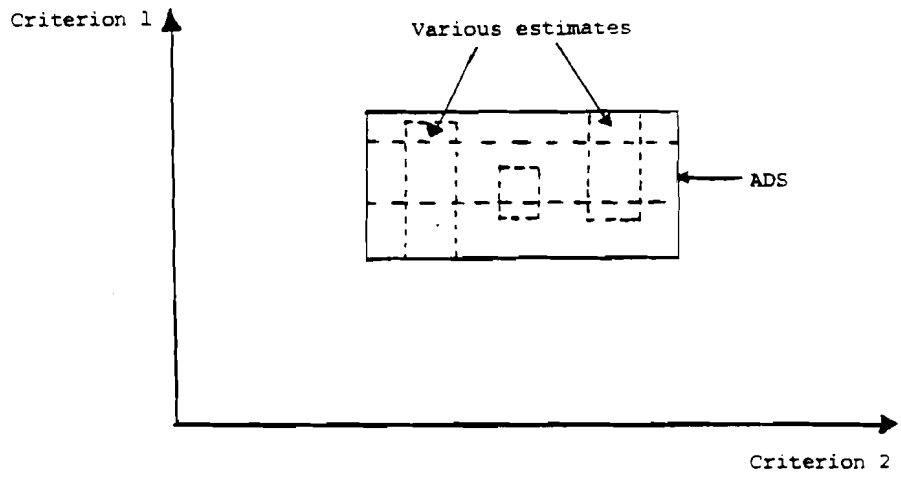


Figure 1. The construction of the admissible demand set (ADS).

We shall call this the admissible demand set (ADS), since it includes the contradiction between the actual demand (or goal) and its admissible deviations.

The existence of the ADS calls for a method of selecting a solution which will be not only desirable but also "safe". The safety requirements may be fulfilled by constructing the "skeleton" of the ADS, which is, roughly speaking, the set of lines equidistant from the boundary of the ADS (see Figure 2).

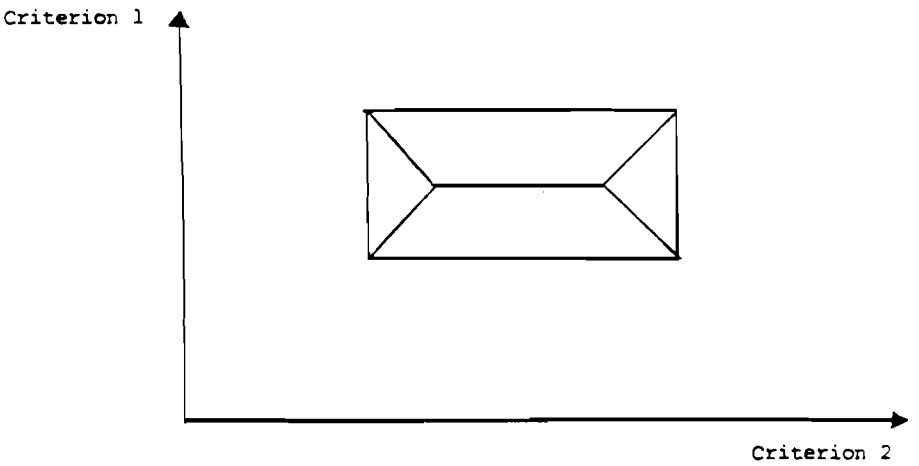


Figure 2. The skeleton of the admissible demand set.

In geometrical terms, the solution of the multiobjective problem is given by a point within the ADS which belongs simultaneously to its skeleton and to the Pareto-optimal set of the model.

Thus the decision maker has achieved a concordance between his goals (and resources) and the available technologies. Since the ADS and hence the decision-making process has to be based on unreliable information, the skeleton provides a safety margin based on the assumption that the criteria are of equal importance.

Solution procedure

We consider the situation described above, making the following additional assumptions:

1. The large-scale system under consideration is linear. This means that the model, constraints, and criteria may be described by linear equations.

2. The criteria are of equal importance.

3. The Pareto (compromise) set is not given explicitly.

4. In the example it is assumed that the admissible demand set has rectangular form in R^n, with the corresponding skeleton set.

There are three possible situations which may arise (see Figure 3):

1. The demand set is feasible. In this case it is possible to choose the compromise (Pareto) point most insensitive to change.

2. The demand set is not feasible. In this case it is possible to determine what should be done to obtain some intersection between the demand and Pareto sets.

3. The demand set is determined improperly. In this case it is necessary to make a correction in the estimate of the demand set.

The approach adopted here is based upon the use of a scalarizing function in criteria space to determine the position of any point in this space with respect to the compromise set [2].

Establishing the skeleton of the ADS makes it possible to search for the polyoptimal solution in a systematic way. As an example, we shall consider a two-criteria case, assuming that both criteria should be minimized without loss of generality.

First, we choose the best point (on the basis of our criteria) lying on the skeleton. In this case it is point 1 in Figure 3. Using the DIDASS procedure, we define the position of this point with respect to the compromise set. We now have the following possibilities:

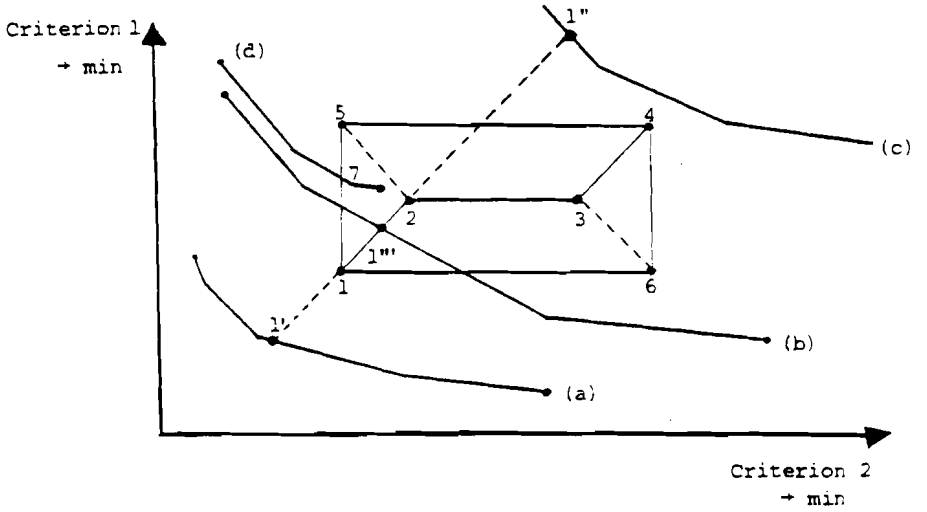


Figure 3. The relationship of the admissible demand set to various Pareto (compromise) sets (a)-(d) and the corresponding optimum points.

(a) The demand set is feasible (case a, point 1').

(b) The corresponding point in the compromise set lies on branch 1-2 of the skeleton, and is determined immediately (case b, point 1"').

(c) The demand set is not feasible (case c, point 1").

(d) The demand set may be feasible but lie in other positions, and in this case the procedure is repeated. The next point 2 is taken as the reference point, enabling us to find any point in the compromise set which lies on the branch 2-3. Point 3 is then taken as the reference point, thus determining any point in the compromise set which lies on the branch 3-4.

Branches 2-5 and 3-6 are not taken into account because they exclude the case in which the end point (case d, point 7) of the

compromise set lies inside the demand set (this case is monitored in a special way and gives valuable information). Safer solutions are found on the other branch, due to the known convexity of the compromise set.

A PRACTICAL EXAMPLE

The case of methanol

Methanol is one of the most widely used raw materials in the chemical industry -- it serves as a fuel and also as a basic raw material in both the chemical and food industries. The demand for methanol in western Europe is expected to increase rapidly, as shown by the forecast below [10]:

1979: 3259 thousand tons
1985: 4345 thousand tons
1990: 5625 thousand tons

It is estimated that in 1990 30% of this demand will arise from the use of methanol in applications that did not exist in 1975. The forecast structure of the demand for methanol in western Europe is presented in Table 1.

TABLE 1 Forecast structure of demand for methanol in western Europe: 1985 and 1990.

Application	Percentage of total methanol demand derived from each application		
	1979	1985	1990
Traditional applications			
Formaldehyde	48.8	42.3	37.0
DMT	4.9	4.2	3.6
MMA	3.4	3.1	2.6
Methyl halides	3.4	3.3	3.1
Methyl amines	4.7	4.1	3.7
Miscellaneous	24.8	22.9	20.8
New applications			
MTBE	2.1	4.2	4.3
MTBE (blending component)	0.9	1.7	1.8
Gasoline blending	6.1	4.6	3.5
Acetic acid	0.8	6.0	9.8
SCP	0.1	3.6	9.8

Seven basic methanol production technologies are considered in the problem analyzed in this paper, and are used to formulate a model. These seven technologies may be listed as follows:

1. Production of methanol from natural gas
2. Production of methanol from natural gas and carbide gas
3. Production of methanol from gasoline
4. Production of methanol from heavy residuals
5. Production of methanol from coal (Koppers-Totzek process)
6. Lurgi process - coal gasification and partial oxidation of the resulting methane
7. Lurgi process - coal gasification and subsequent methanation to obtain SNG and its gas reforming

All of these technologies have the following parameters:

Capacity: 500,000 tons per year
Design time: 2 years
Construction time: 3 years

The technologies may also be described in terms of specific parameters of three distinct types. The first reflects the amount of resources necessary to construct a particular methanol plant. The resources considered include concrete, steel, pipes, and so on. The technological parameters of the process constitute the second group, and include raw materials, products, production capacity and the relevant coefficients.

Information on resources such as energy, manpower, and the amount of waste produced corresponds to the third type of parameter. These three classes of data were originally prepared in the form defined in the WELMM data base [11] in the course of work initiated as described by Kopytowski [12].

Results of experiments

The linear model of the set of methanol technologies is of the type mentioned in the first section of this paper. The original model was extended in order to enlarge the set of resources.

For the purpose of this example we shall limit the resource analysis. The computations have been performed for two types of experiment. The first is connected with the most general resource evaluation, that of investment versus financial efficiency, while the second type is concerned with the evaluation of methanol production technologies from the point of view of their financial efficiency versus their utilization of natural gas (NG). The latter, being increasingly scarce, has to be used with special care. It is worth noting here that capital and NG simply represent resources, irrespective of their further utilization (in this case, production of methanol). The efficiency of the process reflects its ability to transform capital under given constraints. Thus, each range or estimate of the availability of resources, treated as a criterion, represents a particular strategy. The problem to be solved in each experiment is to find the relation between the strategies and the outcomes which can be expected from various methanol technologies in such circumstances.

This means that the final choice from a series of experiments is equivalent to the choice of a particular strategy and a concordant technological structure of methanol production.

Table 2 gives the results of four experiments dealing with the two-dimensional problem of the simultaneous maximization of financial efficiency and minimization of capital investment while maintaining methanol production at or above some assumed level.

Table 3 gives the results of experiments in which it was assumed that the availability of natural gas was limited. This means that we must minimize the use of natural gas while maximizing efficiency. At the same time we should take into account that capital investment is limited and a given amount of methanol must be produced.

The above example serves only to illustrate the ADS-based approach. A more detailed examination of this problem, together with a fuller description of the methodology, will be presented in a separate paper. This paper will be based on a broader set of data and will consider the structure of hydrocarbon production, of which methanol production is an element [10].

TABLE 2 The results obtained in four experiments with different admissible demand sets.

	Experiment			
	1[a]	2[a]	3[b]	4[c]
Admissible demand set				
Efficiency (billions of monetary units)	2-4	2-6	6-10	1-2
Investment (billions of monetary units)	5-10	5-10	5-10	5-10
Solution				
Efficiency (billions of monetary units)	3.13	4.0	5.84	2.28
Investment (billions of monetary units)	5.87	7.08	10.16	4.71
Technologies[d] in operation and operating level	1(100%) 2(16%)	1(100%) 2(38%)	1(100%) 2(97%)	1(92%)
Methanol production (thousands of tons)	575	690	985	460

[a] The solutions lie on the skeleton of the ADS. [b] The ADS is unattainable. [c] The levels attained are better than those proposed by the ADS. [d] The technologies are represented by the numbers assigned to them in the list on p. 9.

TABLE 3 The results obtained in four experiments, assuming that the availability of natural gas is limited.

	Experiment[a]			
	5	6[b]	7[b]	8[b]
Admissible demand set				
Efficiency (billions of monetary units)	8-10	10-12	2-6	2-6
Natural gas consumption (thousands of tons)	300-500	300-500	300-500	300-500
Solution				
Efficiency (billions of monetary units)	8.35	9.5	6.92	6.88
Natural gas consumption (thousands of tons)	400	901	0	0
Technologies[c] in operation and operating level	7(98%) 2(91%)	2(100%) 1(100%) 7(62%)	6(34%) 7(100%)	4(23%) 6(25%) 7(100%)
Methanol production (thousands of tons)	945	1320	671	740[d]

[a] In all experiments investment was limited to 200 monetary units. [b] The solutions lie at an extreme point of the Pareto-optimal set. [c] The technologies are represented by the numbers assigned to them in the list on p. 9. [d] The production of methanol was forced up to 740 thousand tons.

FUTURE EXTENSIONS

Using this methodology, it would be desirable to analyze certain special degenerate cases which arise in problems with more than two criteria.

The algorithm could be made more flexible, in particular by introducing an interactive element so that the user can change the ADS or criteria more easily. Additional subroutines may also be useful.

So far, this approach has been applied only to static systems. An obvious extension would be to develop a similar technique for analyzing dynamic systems.

REFERENCES

1. A.P. Wierzbicki (1980) The Use of Reference Objectives in Multiobjective Optimization - Theoretical Implications and Practical Experience. WP-79-66, International Institute for Applied Systems Analysis, Laxenburg, Austria.

2. A.P. Wierzbicki (1979) The Use of Reference Objective Levels in Group Assessment of Solutions of Multiobjective Optimization. WP-79-124, International Institute for Applied Systems Analysis, Laxenburg, Austria.

3. M. Kallio, A. Lewandowski, and W. Orchard-Hays (1980) An Implementation of the Reference Point Approach for Multiobjective Optimization. WP-80-35, International Institute for Applied Systems Analysis, Laxenburg, Austria.

4. H. Górecki and J. Kopytowski (1980) Polyoptimal Approach to the Problems in Design of Production Structure in the Chemical Industry. 5th Italo-Polish Seminar on the Application of Systems Theory to Economics, Management and Technology, Toruń, 1980.

5. H. Górecki (1981) Problem of Choice of an Optimal Solution in a Multicriterion Space. VIIIth IFAC Congress, Kyoto, Japan, 1981.

6. G. Dobrowolski and M. Zebrowski (1979) Model of a Balanced State of a Process Industry for Production Planning and Growth Management. Paper presented at EURO III, Amsterdam, 1979.

7. G. Dobrowolski, J. Kopytowski, and M. Żebrowski (1980) Systems Analysis in Management of Optimal Strategy of Growth. Paper presented at the 3rd IFAC/IFORS Conference, Warsaw, June 16-19, 1980.

8. A. Borek, G. Dobrowolski, and M. Żebrowski (1978) GSOS - Growth Strategy Optimization System for Chemical Industry. Paper presented at MECO-78, Athens, June 26-29, 1978.

9. G. Dobrowolski, J. Kopytowski, A. Lewandowski, and M. Żebrowski (1982) Generating Efficient Alternatives for Development of the Chemical Industry. Collaborative Paper, International Institute for Applied Systems Analysis, Laxenburg, Austria.

10. M.B. Sherwin (1981) Hydrocarbon Processing, No. 3, pp. 79-84.

11. M. Grenon and B. Lapillone (1976) The WELMM Approach to Energy Strategies and Options. RR-76-19, International Institute for Applied Systems Analysis, Laxenburg, Austria.

12. J. Kopytowski, J. Wojtania, and M. Żebrowski (1981) Fossils as Key Resources of Hydrocarbons for the Chemical Industry - the Burning Problem of Industrial Development. CP-81-20, International Institute for Applied Systems Analysis, Laxenburg, Austria.