ON MULTICRITERIA OPTIMIZATION

by

E. Polak and A. N. Payne

Memorandum No. ERL-M566

27 May 1975

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

ON MULTICRITERIA OPTIMIZATION

E. Polak and A. N. Payne

Department of Electrical Engineering and Computer
Sciences and the Electronics Research Laboratory
University of California, Berkeley, Calif. 94720

ABSTRACT

This paper examines the state of the art in multicriteria optimization. For this purpose, multicriteria problems are classified in terms of complexity as finite and small, finite and large, and infinite. The relative merits of typical methods for solving each of these classes are discussed and some suggestions for future work are made.

1. INTRODUCTION

Multicriteria optimization problems are almost always two or three phase problems. The first phase is common to all of them and can be stated as follows. Given (i) a set Ω of (feasible) <u>decision variables</u> (also referred to as <u>alternatives</u>), (ii) a set of <u>criteria</u> $f^i : \Omega \to \mathbb{R}$, $i = 1,2,\ldots,m$,† (also referred to as <u>attributes</u>) which produce <u>values</u> $f(x)$ (with $f = (f^1,f^2,\ldots,f^m)$), and (iii) a partial order $\leqq$ ($\leq$) in $\mathbb{R}^m$, the space of values,* construct the set of <u>noninferior decision variables</u> Ω_N and the set of <u>noninferior values</u> V_N defined as follows:

†The f^i are usually defined on a space X with $\Omega \subset X$.

*In this paper we shall assume that $v_1 \leqq v_2$ if $v_1^i \leq v_2^i$ for $i = 1,2,\ldots,m$; $v_1 \leq v_2$ if $v_1 \neq v_2$ and $v_1 \leqq v_2$. This is the most commonly used partial order.

$$V_N \triangleq \{v \in f(\Omega) | N(v) \cap f(\Omega) = \{v\}\} \tag{1}$$

where

$$N(v) \triangleq \{v' \in \mathbb{R}^m | v' \leq v\} \tag{2}$$

and

$$\Omega_N \triangleq \{x \in \Omega | f(x) \in V_N\} \tag{3}$$

The second phase consists of the following: given a set of <u>acceptable performance values</u> $V_a \subset \mathbb{R}^m$, find all the points in

$$V_{Na} \triangleq V_N \cap V_a \tag{4}$$

and in

$$\Omega_{Na} \triangleq \{x \in \Omega | f(x) \in V_{Na}\} \tag{5}$$

The third phase may consist of either selecting a point in Ω_{Na} or imposing a total order on the elements of Ω_{Na}.

In this paper we shall examine the difficulty involved in solving multicriteria problems, as well as some of the techniques suggested for their solution. For this purpose, it will be convenient to group multicriteria problems into three distinguishable classes: (i) when Ω consists of a small number of elements, (ii) when Ω consists of a large, but finite, number of elements, and (iii) when Ω is a subset of a normed space.

2. THE "SMALL" MULTICRITERIA DECISION PROBLEM

We begin with the simplest case, when Ω contains a small number of elements, say less than 20, and the number of criteria is fairly small, say less than 6. This is a common situation when one is buying a car, a radio, a saw, etc. In this case, the construction of V_N and Ω_N is quite trivial so that phase 1 poses no difficulties, while in phase 2 the number of decision variables to be considered is further restricted by V_a, the acceptable performance set. In fact, the set V_a is often used to convert criteria into constraints and thus reduce the number of criteria. The final choice (phase 3) is often facilitated by the fact that, usually, the criteria can be ordered in terms of their relative importance. Such an ordering is is called a <u>lexicographic</u> ordering of the criteria. It is used to impose, successively, more stringent acceptable performance requirements. Also, it is not uncommon to reorder the relative importance of criteria after the process of elimination of unacceptable alternatives has reached a certain stage.

Let us illustrate this progressive elimination process by showing how a hypothetical consumers analyst might arrive at a recommendation for a radial-arm saw out of a field of 8 alternatives; i.e., the cardinality of Ω is 8. The following attributes are commonly considered to be pertinent to this selection (in the initial order of decreasing importance): (a) depth of cut at 90°, (b) rip width, (c) motor type, (d) depth of cut at 45°, and (e) price. The alternatives and attributes are displayed in Table 1. Note that the attribute (c) is not a numerical value, but it could be converted to one by assigning value -1 to an induction motor and value 0 to any other type motor since induction motors are preferred. Also, since the set (1) is implicitly specified in terms of minimizations, the negatives of the depth of cut and rip width must be used to convert this problem to standard form. The acceptable performance values are specified as follows: (a) the depth of cut at 90° should be at least 3", and (b) the rip width should be at least 25".

Now, in this case, rather than first construct V_N and Ω_N, it is more expedient to first eliminate all alternatives which do not yield acceptable performance. Thus, alternatives 2, 5, 7, and 8 are immediately eliminated, leaving only alternatives 1, 3, 4, and 6 for consideration. Denoting alternative i by a_i, we see that $f(a_4) \leq f(a_1)$ and hence $a_1 \notin \Omega_N$. By inspection, the analyst now obtains that $\Omega_{Na} = \{a_3, a_4, a_6\}$. Next, the analyst reorders his priorities to let motor type take precedence over the other criteria and hence eliminates a_6. Since a depth of cut of 3 1/8" at 90° does not represent a useful improvement over 3", the analyst considers a_3 and a_4

Alternative	Attributes				
	(a) Depth of Cut at 90°	(b) Rip Width	(c) Motor Type	(d) Depth of Cut at 45°	(e) Price
1	3 in.	$25\frac{5}{8}$ in.	induction	$2\frac{1}{4}$ in.	$265
2	3	$24\frac{3}{4}$	induction	$1\frac{7}{8}$	293
3	$3\frac{1}{8}$	25	induction	2	220
4	3	$25\frac{3}{4}$	induction	$2\frac{1}{2}$	215
5	$2\frac{1}{2}$	$25\frac{1}{2}$	induction	$1\frac{7}{8}$	175
6	$3\frac{3}{4}$	$25\frac{5}{8}$	universal	$1\frac{3}{4}$	271
7	3	$19\frac{5}{8}$	universal	$1\frac{7}{8}$	123
8	$2\frac{7}{8}$	24	induction	1	300

Table 1. Data for radial-saw selection [6].

indistinguishable under this criterion. Hence the final decision is made on the basis of depth of cut at 45°, yielding our analyst's recommendation to buy a_4.

Note that the specification of acceptable performance does not always remove a criterion from the list of functions one desires to minimize. For example, in the selection of a saw, an acceptable performance requirement may be that the price should be less than $300, but one still wishes to minimize the sum to be paid for a saw.

Thus, not infrequently, in the "small" problem case, as in our example of selecting a saw, we do not have much difficulty in making a decision since the process is considerably simplified due to the reduction of the number of alternatives first by the acceptable performance requirements and subsequently by lexicographic elimination. The final sequential reduction of alternatives by the lexicographic approach appears to be "natural" in that people tend to adopt it without special training [18], [22].

3. THE LARGE, BUT FINITE, MULTICRITERIA DECISION PROBLEM

We now consider the case where Ω consists of a finite number, ν, of alternatives, but ν is too large for the construction of Ω_{Na} to be possible by inspection. This is obviously a combinatorial problem involving sorting and one would expect a sizable bibliography on the subject. However, this is not so, and we have only been able to find [14]-[16], [27]. The latest work dealing with this problem appears to be that of Kung, Luccio, and Preparata [15] who obtain bounds on $C_m(\nu)$, the number of scalar comparisons necessary for computing the sets V_N and Ω_N, where m is the number of criteria and ν is the cardinality of $f(\Omega)$. Specifically, they show that

$$C_m(\nu) \leq O(\nu \log_2\nu) \text{ for } m = 2,3 \tag{6}$$

$$C_m(\nu) \leq O(\nu(\log_2\nu)^{m-2}) \text{ for } m \geq 4 \tag{7}$$

In [26], Yao shows that

$$C_m(\nu) \geq \nu \log_2\nu + \nu - 1 \text{ for } m \geq 2 \tag{8}$$

Kung et al. exhibit algorithms which satisfy the bound (6) for m = 2,3, but there appears to be no explicit algorithm that achieves (7) for $m \geq 4$. However, as we shall shortly show, it is quite easy to specify a simple algorithm which, for any m, requires at most

$$\bar{C}_m(\nu) \triangleq m\nu(\nu+1)/2 \tag{9}$$

scalar comparisons. Now suppose that m = 10 and $\nu = 10^3$. Then the bound in (7) is

$$0(\nu(\log_2\nu)^{m-2}) = 0(10^3(\log_2 10^3)^8) = 0(10^{11}) \tag{10}$$

while

$$m\nu(\nu+1)/2 = 10(10^6+10^3) = 0(10^7) \tag{11}$$

Thus, the bound (7) is not very sharp, at least not until ν gets to be very large. Consequently, for moderate size problems the algorithm below should be quite acceptable. This algorithm requires that the vectors $v \in f(\Omega)$ be indexed as a set $\{v_i\}_{i=1}^{\nu}$, $v_i \neq v_j$. The algorithm first examines v_γ with $\gamma = \nu$. If $v_\gamma \notin V_a$, then v_γ is removed from the list. If $v_\gamma \in V_a$, then v_γ is compared with v_j, $j = \gamma - 1,\ldots,2,1$, until either (a) $v_\gamma \geq v_j$, in which case v_γ is rejected, or (b) $v_\gamma \leq v_j$, in which case v_j is rejected and the list is renumbered so that $(v_1,v_2,\ldots,v_{j-1},v_{j+1},\ldots v_\gamma) \to (v_1,v_2,\ldots,v_{\gamma-1})$, or (c) the top of the list, namely v_1, is reached, in which case $v_\gamma \in V_{Na}$ and is removed as such. The process then begins again with γ, the length of the list, decreased by one.

General Purpose Multicriteria Optimization Algorithm

Data: $\{v_i\}_{i=1}^{\nu}$, V_a.

Step 0: Set $k = 1$, $\ell = 1$, $\gamma = \nu$.

Comment: k is the index for the rejected values ($v_i = r_k$) and ℓ is the index for noninferior values ($v_i = \hat{v}_\ell$); γ is the total number of values under immediate consideration.

Step 1: If $v_\gamma \in V_a$, go to step 3; else, go to step 2.

Step 2: Set $r_k = v_\gamma$, $k = k + 1$, $\gamma = \gamma - 1$. If $\gamma = 0$, stop; else, go to step 1.

Step 3: If $\gamma = 1$, set $\hat{v}_\ell = v_1$, print the set of noninferior values $\{\hat{v}_i\}_{i=1}^{\ell}$ and corresponding alternatives $\{\hat{x}_i\}_{i=1}^{\ell}$ and stop; else, set $\tilde{v} = v_\gamma$, $\gamma = \gamma - 1$, $j = \gamma$, and go to step 4.

Step 4: If $\tilde{v} \leq v_j$ or $v_j \notin V_a$, go to step 5; else, go to step 8.

Step 5: Set $r_k = v_j$, $k = k + 1$. If $j = \gamma$, set $\gamma = \gamma - 1$ and go to step 6; else, renumber $(v_{j+1},v_{j+2},\ldots,v_\gamma) \to (v_j, v_{j+1},\ldots,v_{\gamma-1})$, set $\gamma = \gamma - 1$, and go to step 6.

Step 6: Set $j = j - 1$. If $j = 0$, set $\hat{v}_\ell = \tilde{v}$, $\ell = \ell + 1$, and go to step 7; else, go to step 4.

Step 7: If $\gamma = 0$, print the set of noninferior values $\{\hat{v}_i\}_{i=1}^{\ell}$ and corresponding alternatives $\{\hat{x}_i\}_{i=1}^{\ell}$ and stop; else, go to step 3.

Step 8: If $\tilde{v} \geq v_j^{\mu}$, set $r_k = \tilde{v}$, $k = k + 1$, and go to step 3; else, go to step 6. ¤

Since the superior algorithms for the cases $m = 2$ and $m = 3$ are easy to state, we give below our extension of the schemes proposed in [15] for these cases. Our extension accounts for V_a and for the fact that $v_i^1 = v_j^1$ is possible, whereas in [15] it was assumed $v_i^1 \neq v_j^1$ for $i \neq j$.

Bicriteria Algorithm (m=2)

Data: $\{x_i\}_{i=1}^{\nu}$, $\{v_i\}_{i=1}^{\nu}$, V_a.

Step 0: Sort the vectors v_i, $i = 1,2,\ldots,\nu$ on the basis of the first component and renumber so that $v_1^1 \leq v_2 \leq \cdots \leq v_\nu^1$. Set $i = 1$, $\gamma = 1$.

Step 1: Find the largest integer $\mu(i)$ such that $v_i^1 = v_{i+1}^1 = \ldots = v_{i+\mu(i)}^1$.

Step 2: Find an integer $\bar{\mu}(i) \in I(i) \triangleq \{i,i+1,\ldots,i+\mu(i)\}$ such that $v_{\bar{\mu}(i)}^2 = \min\{v_k^2 | k \in I(i)\}$; set $u_\gamma = v_{\bar{\mu}(i)}$, $\xi_\gamma = x_{\bar{\mu}(i)}$, and $\gamma = \gamma + 1$.

Step 3: If $i + \mu(i) < \nu$, set $i = i + \mu(i) + 1$ and go to step 1; else, go to step 4.

Comment: The set $\{u_i\}_{i=1}^{\gamma}$ consists of all the values in $f(\Omega)$ satisfying $u_1^1 < u_2^1 <\ldots< u_\gamma^1$. A value v_i not included in $\{u_i\}_{i=1}^{\gamma}$ cannot be in V_N.

Step 4: Set $i = 1$, $\ell = 1$, $b_0 = \infty$.

Step 5: If $u_i^2 \geq b_{i-1}$ or $u_i \notin V_a$ (i.e., $u_i \notin V_{Na}$), set $b_i = b_{i-1}$ and go to step 6; else, set $\hat{v}_\ell = u_i$, $\hat{x}_\ell = \xi_i$, $\ell = \ell + 1$, $b_i = u_i^2$, and go to step 6.

Step 6: If $i < \gamma$, set $i = i + 1$ and go to step 5; else, print the set of noninferior acceptable values $\{\hat{v}_j\}_{j=1}^{\ell}$ and the corresponding set of noninferior alternatives $\{\hat{x}_j\}_{j=1}^{\ell}$ and stop. ¤

The tricriteria algorithm described in [15] is more complex. It is based on the following argument. Suppose we are given a set of noninferior values $V_j = \{v_i\}_{i=1}^{j}$ in $\mathbb{R}^3$, ordered so that $v_1^1 \leq v_2^1 \leq \ldots \leq v_j^1$, and suppose that we wish to determine if the set $V_{j+1} = \{v_i\}_{i=1}^{j+1}$ where $v_{j+1}^1 \geq v_j^1$, also consists of noninferior values only. Let $w_i \triangleq (v_i^2, v_i^3) \in \mathbb{R}^2$ for $i = 1,2,\ldots,j + 1$, and let $W_j = \{w_{i_k}\}_{k=1}^{\ell}$ be the set of noninferior values in $\{w_i\}_{i=1}^{j}$. Let $\hat{w}_k = (\hat{v}_k^2, \hat{v}_k^3) = w_{i_k} = (v_{i_k}^2, v_{i_k}^3)$ and assume, without loss of generality, that

$\hat{v}_1^2 \leq \hat{v}_2^2 \leq \cdots \leq \hat{v}_\ell^2$. Then V_{j+1} consists of noninferior values if and only if the set $W_j \cup \{w_{j+1}\}$ consists of noninferior values only. Furthermore, $W_j \cup \{w_{j+1}\}$ consists of noninferior values if and only if $v_{j+1}^3 < \hat{v}_{j*}^3$, where $j*$ is the largest integer in $\{1,2,\ldots,\ell\}$ such that $\hat{v}_p^2 \leq v_{j+1}^2$ for $p = 1,2,\ldots,j*$.

<u>Tricriteria Algorithm</u> (m=3)

<u>Data</u>: $\{x_i\}_{i=1}^{\nu}$, $\{v_i\}_{i=1}^{\nu}$, V_a.

<u>Step 0</u>: Sort the set $\{v_i\}_{i=1}^{\nu}$ on the basis of the first component, and renumber $\{v_i\}_{i=1}^{\nu}$ and $\{x_i\}_{i=1}^{\nu}$ accordingly so that $v_1^1 \leq v_2^1 \leq \cdots \leq v_\nu^1$.

<u>Step 1</u>: Find the largest integer δ such that $v_\delta^1 = v_1^1$ and apply the bicriteria algorithm to $\{w_1,w_2,\ldots w_\delta\}(w_i \triangleq (v_i^2,v_i^3),\ i = 1,2,\ldots,\delta)$ to find all the noninferior values $\{w_{i_k}\}_{k=1}^{\beta}$ (with $w_{i_1}^1 < w_{i_2}^1 < \ldots < w_{i_\beta}^1$) in $\{w_i\}_{i=1}^{\delta}$. Renumber the set $(v_{i_1},v_{i_2},\ldots,v_{i_\beta},v_{\delta+1},\ldots,v_\nu) \rightarrow (v_1,v_2,\ldots,v_\gamma)$.

<u>Step 2</u>: Set $i = \beta + 1$, $k = \beta$. Set $\hat{v}_j = v_j$, $\hat{x}_j = x_j$, $w_j = (v_j^2,v_j^3)$ for $j = 1,2,\ldots,\beta$.

<u>Step 3</u>: Set $w = (v_i^2,v_i^3)$.

<u>Step 4</u>: Find the largest index $j*$ such that $w_j^1 \leq w^1$ for $j = 1,2,\ldots,j*$.

<u>Step 5</u>: If $w^2 < w_{j*}^2$ and $v_i \in V_a$, set $\hat{v}_{k+1} = v_i$, $\hat{x}_{k+1} = x_i$, renumber $(w_1,w_2,\ldots w_{j*}, w, w_{j*+1},\ldots,w_k) \rightarrow (w_1,w_2,\ldots,w_{k+1})$, set $k = k + 1$, $i = i + 1$, and go to step 6; else, set $i = i + 1$ and go to step 6.

<u>Step 6</u>: If $i \leq \gamma$, go to step 3; else, print the noninferior values $\{\hat{v}_j\}_{j=1}^{k}$ and the corresponding noninferior alternatives $\{\hat{x}\}_{j=1}^{k}$ and stop. ◻

Note that in the three algorithms which we have presented the test for $v_i \in V_a$ is carried out simultaneously with the determination of whether $v_i \in V_N$. If it is relatively easy to determine $V_a \cap \{v_i\}_{i=1}^{\nu}$, then it may be more efficient to do this first. Similarly, it may be more efficient to first determine V_N completely and then find the set $V_{Na} = V_a \cap V_N$. In fact, in many decision problems this may be the only way to proceed since the decision maker may use information provided by a knowledge of V_N to establish the set of acceptable performance values.

Thus, the construction of the noninferior, acceptable performance value set V_{Na} is a tractable task in the finite alternative case. If the third phase is to select a single point, then this task can be accomplished as follows. (i) Scan the range of values for each criterion. Those criteria in which the values vary little (e.g., $(\max_k \hat{v}_k^j - \min_k \hat{v}_k^j)/\mathrm{ave}_k(\hat{v}_k^j)$ is small) can be removed from further consideration since these criteria offer little discrimination. (ii) For the remaining criteria, establish an order of importance with appropriately narrow bands of acceptable performance. A successive application of these bands should reduce the subset of noninferior alternatives under consideration until a final choice is made.

When the final task is not to select a single alternative but to select several or to order the alternatives linearly, as is the case in university admissions or in the processing of fellowship applications [8], it is not uncommon to use a weighting method consisting of the assignment of the scalar value $\sum_{i=1}^{m} \lambda^i f^i(x_j)$, where $\lambda^i > 0$, $i = 1,2,\ldots,m$, are certain weights, to the alternative j. When there is some experience available, these weights can be computed by regression. For example, suppose a fellowship committee awards annually graduate fellowships to entering, deserving students. The attributes which the committee takes into account are grade point average (G_i), rating of letters of recommendation (L_i), GRE scores (E_i), and a rating of the school of undergraduate education (S_i). Suppose there have been N applicants for the fellowship over a number of preceding years and the committee ranking r_j, $0 \leq r_j \leq N$, of each of the N applicants has been recorded. Then, to automate their selection process, the committee chooses weights $\alpha,\beta,\gamma,\delta > 0$ which minimize

$$\sum_{i=1}^{N} (\alpha G_i+\beta L_i+\gamma E_i+\delta S_i-r_i)^2 \tag{12}$$

In the current year, the number $\alpha G_i + \beta L_i + \gamma E_i + \delta S_i$ is the i^{th} student's score to be used in the final ranking. This process is not beyond criticism since a value which is not noninferior can be given higher ranking than a noninferior one. Nevertheless, for lack of anything better, the weighting process is used quite commonly when a ranking must be produced and is usually applied to all alternatives in Ω rather than to noninferior alternatives only.

4. THE INFINITE ALTERNATIVES MULTICRITERIA DECISION PROBLEM

We shall now consider the case where Ω consists of an infinite number of points, usually defined by equality and inequality constraints. It is no longer a problem which can be solved by enumeration. We shall concentrate on the case where $\Omega \subset \mathbb{R}^n$. Let us review the most important facts about this case (see, for example, [7], [20]).

Let $\bar{f} : \mathbb{R}^n \to \mathbb{R}^{m-1}$ be defined by $\bar{f}(x) = (f^1(x),\ldots,f^{m-1}(x))$, where the f^i, $i = 1,2,\ldots,m-1$, are the first $m-1$ criterion functions. Let

$$\bar{V} = \{\bar{v} \in \mathbb{R}^{m-1} | \Omega \cap \{x | \bar{f}(x) \leqq \bar{v}\} \neq \phi\} \tag{13}$$

Let $s : \bar{V} \to \mathbb{R}$ defined by

$$s(\bar{v}) \triangleq \min\{f^m(x) | x \in \Omega,\ \bar{f}(x) \leqq \bar{v}\} \tag{14}$$

Let Γ denote the graph of $s(\cdot)$, i.e.,

$$\Gamma \triangleq \{v \in \mathbb{R}^m | v = (\bar{v},v^m),\ \bar{v} \in \bar{V},\ v^m = s(\bar{v})\} \tag{15}$$

and let

$$V \triangleq f(\Omega) \tag{16}$$

Proposition 1: The set of noninferior values, V_N, is contained in Γ, the graph of $s(\cdot)$. Furthermore,

$$V_N = \{v = (\bar{v},v^m) \in \Gamma \cap V | v^m = s(\bar{v}) < v'^m,\ \forall v' \in V \cap N(v)\} \tag{17}$$

where $N(v)$ is defined by (2). ¤

The relation (17) can be reinterpreted as follows.

Corollary: An alternative $x \in \Omega_N (f(x) \in V_N)$ if and only if x is a global minimizer of (14), for $\bar{v} = \bar{f}(x)$ and, in addition, $f^m(x) < f^m(x')$ $\forall x' \in \Omega$ satisfying $\bar{f}(x') \leq \bar{f}(x)$. ¤

Proposition 2: The sensitivity function $s(\cdot)$ is monotonically decreasing; i.e., $\bar{v}' \geq \bar{v}$ implies that $s(\bar{v}') \leq s(\bar{v})$. Furthermore, suppose $s(\cdot)$ is piecewise continuously differentiable. If $v \in \Gamma \cap V$ satisfies $\nabla s(\bar{v}) < 0$, then $v \in V_N$. ¤

Now suppose that

$$\Omega = \{x \in \mathbb{R}^n | g(x) = 0,\ h(x) \leqq 0\} \tag{18}$$

where $g : \mathbb{R}^n \to \mathbb{R}^k$, $h : \mathbb{R}^n \to \mathbb{R}^\ell$ are twice continuously differentiable. For any $\bar{v} \in \bar{V}$, let

$$\Omega_{\bar{v}} \triangleq \{x \in \Omega | \bar{f}(x) \leqq \bar{v}\} \tag{19}$$

If $\Omega_{\bar{v}}$ satisfies the Kuhn-Tucker constraint qualification for almost all $\bar{v} \in \bar{V}$ and if the corresponding minimizers of (14) satisfy second order necessary conditions of optimality (see [17, p.234]), then by interpreting the results in [24], [25], [17, p.236], we conclude that $s(\cdot)$ is piecewise continuously differentiable and that $\nabla s(\bar{v}) = -\mu(\bar{v})$, where μ is the Lagrange multiplier associated with the constraint $\bar{f}(x) \leqq \bar{v}$ in (14).

These observations lead to the following conclusions. (i) To compute a noninferior alternative solve (14) for some $\bar{v} \in \bar{V}$ to obtain a solution $x(\bar{v})$ and a corresponding multiplier $\mu(\bar{v})$. If $\bar{f}(x(\bar{v})) = \bar{v}$, then $(\bar{f}(x(\bar{v})), s(\bar{v})) \in \Gamma \cap V$, and if $\nabla s(\bar{v}) = -\mu(\bar{v}) < 0$, then $v = (\bar{v}, s(\bar{v})) \in V_N$ and $x(\bar{v}) \in \Omega_N$. If these conditions fail, try another $\bar{v}$. In principle, almost all points in Ω_N can be found in this manner. (ii) Since Ω_N consists of an infinite number of alternatives, the entire set Ω_N cannot be constructed.

As in the finite alternative case, we find that problems with two or three criteria are special because the sets V_N can be displayed as curves or parametrized families of curves when $m = 2$ or $m = 3$. This fact is utilized in the algorithms described in [20], [21] which produce an efficient, piecewise cubic approximation to V_N. Once such an approximation to V_N is obtained, one can proceed as follows. First, making use of a set of acceptable performance values V_a one can reduce the set of noninferior values V_N to be considered to $V_{Na} = V_N \cap V_a$. Then one can establish bands of equivalent performance for each criterion, which are ordered with respect to their relative importance. One can then apply these bands successively to eliminate a large number of alternatives and narrow down the final selection to a sufficiently small region so as to make the final choice fairly easy. This process is essentially the same as the one described in the finite alternative case.

When there are more than three criteria, the information display problem makes the approach described above, to say the least, quite cumbersome, if not impossible. However, since $s(\cdot)$ is likely to be piecewise continuously differentiable, one may try to impose a <u>total</u> order on Ω by means of an <u>aggregation function</u> $a : \mathbb{R}^n \to \mathbb{R}$ of the form

$$a(x) = \alpha(f^1(x), f^2(x), \ldots, f^m(x)) \tag{20}$$

and one can choose to minimize either $a(x)$ over $x \in \Omega$ or $\alpha(v)$ over $v \in \Gamma$. Since on the manifold Γ $\nabla_{\bar{v}}\alpha(v) = (I_{m-1} | \nabla s(\bar{v}))\nabla\alpha(v)$ may have discontinuities and since both $\alpha(v)$ and $\nabla_{\bar{v}}\alpha(v)$ may be difficult to compute, it is common to make a final selection by solving the aggregated problem

$$\min\{a(x) \,|\, x \in \Omega\} \tag{21}$$

In the next section we shall discuss the manner in which aggregation has been approached by several authors. However, before proceeding with this, we must point out the two serious drawbacks of aggregation. The first and most obvious drawback is that in many cases a solution of (21) will not yield an alternative in Ω_N or value in V_N. The second drawback is that aggregation is based on the assumption that a total order can actually be imposed on Ω. Now there is considerable evidence [1] that, in a multiattribute situation, humans quite commonly will prefer alternative a to alternative b, alternative b to alternative c, but they may not prefer alternative a to alternative c. This shows that the imposition of a total order is often contrary to our intuitive aims and hence is quite likely to lead to less than ideal selections. Thus, aggregation should be used with extreme caution.

5. IMPOSITION OF A TOTAL ORDER: AGGREGATION

There are basically three schemes, with endless variations, for imposing a total order on the partially ordered set of values $V = f(\Omega)$. The first and oldest (which was used even by Pareto in the last century) consists of imposing a weighting pattern on the criteria; i.e., one minimizes $\sum_{i=1}^{m} \lambda^i f^i(x)$, with $x \in \Omega$ and $\lambda^i \geq 0$. We have seen already an example of this technique in Sec. 3. As it was shown in [7], when V is directionally convex, any solution of $\min\{\sum_{i=1}^{m} \lambda^i f^i(x) \mid x \in \Omega\}$ is noninferior. Furthermore, by using all the λ in the set $\{\lambda \in \mathbb{R}^m \mid \lambda \geq 0, \sum_{i=1}^{m} \lambda^i = 1\}$, all the noninferior alternatives can be computed. Thus, starting with a weighting pattern $(\lambda^1, \ldots, \lambda^m)$, we can compute at least one noninferior alternative and then perturb the λ^i to see whether some other noninferior point in the vicinity of the first one is more desirable. This is quite common practice in linear-quadratic regulator design [2].

The second scheme is based on the hypothesis that there are indifference surfaces and that these surfaces are equi-cost surfaces for some unknown aggregate cost or utility function u. Thus, for example, Hanieski [12] assumes that we can specify trade off coefficients a^i such that given any $v = f(x)$, v lies on an indifference surface if

$$du = \sum_{i=1}^{m} a^i \frac{dv^i}{v^i} = 0 \tag{22}$$

When all $a^i = 1$, (22) describes a point at which the fractional changes in all criteria cancel each other out. Integrating (22), Hanieski then proposes to minimize

$$u(v) = \sum_{i=1}^{m} a^i \ell n v^i \tag{23}$$

subject to $v^i = f^i(x)$, $x \in \Omega$. Since this is a very simple scheme it is also easy to analyze it. Generally, in the context of a heuristic selection scheme for a final point $v \in V_N$, one tends to agree that (22) describes a point of confusion or indecision, i.e., a point v at which the decision maker finds it almost impossible to decide on a preferable small perturbation in v since any fractional gain in one criterion is offset by an approximately equal loss in the other criteria. So one may accept this as being a point on a surface of confusion or indecision, but a surface of confusion need not be a surface of indifference. For example, suppose m = 2 and $V_N = \{(v^1,v^2) | v^2 = e^{-v^1}\}$. Also suppose $a^1 = a^2 = 1$. Then on V_N

$$u(v^1,v^2) = \ell n v^2 + \ell n v^1 = -v^1 + \ell n v^1 \tag{24}$$

and we see that u has a maximum but certainly no minima ($u(v) \to -\infty$ as $v^1 \to 0$) or local minima. Thus, u does not appear to be an appropriate function to minimize. Furthermore, the smaller we make u(v), the closer v approaches a "dictatorial" solution, i.e., either $v^1 \approx \min\{f^1(x) | x \in \Omega\}$ or $v^2 \approx \min\{f^2(x) | x \in \Omega\}$, which is inconsistent with a desire for a trade off or "compromise" between criteria. Our simple example shows that one must be very cautious in translating well defined local trade off considerations into a global aggregate utility function.

Before leaving the second scheme of aggregation, we shall describe two other schemes for obtaining an aggregate utility function from local trade off considerations, so as to illustrate the range of ingenuity that has been devoted to this subject. The first is due to Geoffrion [10] and Geoffrion, Dyer and Feinberg [11]. They assume that the decision maker (DM) has a global preference function u in mind but that he can only furnish local trade off information in the form of trade off ratios. Thus, they consider the problem of solving

$$\min\{u(f(x)) | x \in \Omega\} \tag{25}$$

in a man-machine interactive process. Their method consists of two subprocedures: one for determining a usable feasible direction and one for step size determination. Given a value $\hat{v} = f(\hat{x}) \in V$, the decision maker specifies the trade off ratio $\hat{w}^i$ between the i^{th} criterion f^i and a selected reference criterion, say f^m; i.e.,

$$\hat{w}^i = \frac{\partial u(\hat{v})}{\partial v^i} / \frac{\partial u(\hat{v})}{\partial v^m}, \qquad i = 1,2,\ldots,m-1 \tag{26}$$

These m − 1 trade off ratios describe the tangent hyperplane to the indifference (equi-cost) surface of u passing through $\hat{v}$; i.e.,

$$\hat{w}^1(v^1-\hat{v}^1) + \hat{w}^2(v^2-\hat{v}^2) +\ldots+ \hat{w}^{m-1}(v^{m-1}-\hat{v}^{m-1}) + (v^m-\hat{v}^m) = 0 \tag{27}$$

To obtain the $\acute{w}^i$, the DM answers the query: "How much of a change Δv^i are you willing to permit in the value of $\hat{v}^i$ to obtain a change of Δv^i in $\acute{v}^m$, assuming that all other criteria do not change?" Then

$$\acute{w}_i = \frac{\Delta v^m}{\Delta v^i} \qquad i = 1,2,\ldots,m-1 \tag{28}$$

Thus, with $\bar{w} = (\hat{w},1)$,

$$\nabla_x u(f(\hat{x})) = \frac{\partial u(\hat{v})}{\partial v^m} \frac{\partial f(\hat{x})}{\partial x}^T \bar{w} \tag{29}$$

and even though $\partial u(\hat{v})/\partial v^m$ is not known, $\bar{w}$ is all we need to solve the Frank-Wolfe [26] direction finding problem

$$\min\{\langle \nabla_x u(f(\hat{x})),\ x - \hat{x}\rangle | x \in \Omega\} \tag{30}$$

since we can substitute $\frac{\partial f(\hat{x})}{\partial x}^T\bar{w}$ for $\nabla_x u(\hat{v})$ in (30) without affecting the resulting solution $\hat{h} = \tilde{x} - \hat{x}$ Once the direction $\hat{h}$ is computed, Geoffrion et al. require the DM to specify a step size. Thus, they have invented a well-inspired heuristic method which works well in some cases. Obviously, the method inherits all the possible pathologies that we have mentioned in conjunction with Hanieski's scheme.

To conclude our discussion of aggregation methods which extend local trade off information to a global preference function, we describe a scheme due to Briskin [3], [4], who assumes that the DM can specify all the criteria in terms of the units of a single criterion, say f^m, that the aggregate utility function is of the form

$$u(v) = v^m + \tilde{u}(v^1,\ldots,v^{m-1}) \tag{31}$$

and that $\tilde{u}$ can be specified by a set of differential equations:

$$\frac{\partial \tilde{u}(\bar{v})}{\partial v^j} = h^j(\bar{v}) \qquad j = 1,2,\ldots,m-1 \tag{32}$$

Briskin then integrates this system of equations to obtain $\tilde{u}(\cdot)$ and then minimizes $u(f(x))$ subject to $x \in \Omega$.

To see how Briskin's scheme works, we reproduce one of his examples which involves hauling freight [4]. Suppose $m = 3$, with f^1 measuring time, f^2 weight, and f^3 dollar cost for a given set of alternatives. The problem is to deliver as much weight as quickly as possible and with as little cost as possible. Briskin assumes that the criteria f^1 and f^2 can be expressed in terms of dollars. Next, he supposes that (a) the DM is willing to spend \$30 to gain one hour when the travel time is 30 hours but only \$5 when the travel time is 20 hours, (b) the rate of exchange of money for time varies exponentially with the time taken, (c) the DM is willing to spend \$20 to deliver 100 lbs. more if the delivered weight is 180 lbs., (d) the rate of exchange of money for weight varies inversely with the weight delivered, and (e) the willingness to spend money to gain time is independent of weight and vice versa. From all of this we get that $\tilde{u}$ must satisfy partial differential equations of the form

$$\frac{\partial\tilde{u}(v^1,v^2)}{\partial v^1} = -ck_1 e^{k_1 v^1} \tag{33}$$

$$\frac{\partial\tilde{u}(v^1,v^2)}{\partial v^2} = \frac{k_2}{v^2} \tag{34}$$

From the given data and (33) and (34) we get enough algebraic equations to solve for c, k_1, and k_2, which can then be substituted into the obviously assumed form of $\tilde{u}$:

$$\tilde{u}(v^1,v^2) = -ce^{k_1 v^1} + k_2 \ln v^2 \tag{35}$$

Briskin's method probably shares the faults inherent in Hanieski's method and other methods which transform local trade off preferences into a global preference function. In addition, it requires much more sophistication in specifying the functions h^j than the coefficients in the Hanieski and the Geoffrion et al. schemes.

A totally different approach to aggregation is represented by the compromise solution or goal programming methods [5], [9], [13], [23], [28]-[31]. In fact, these methods are not based on the desire to construct a global preference function but on a desire to compute an alternative x whose value f(x) is close to the ideal value $\hat{v}$ defined as follows. Let

$$\hat{v}^i \triangleq \min\{f^i(x) \mid x \in \Omega\} , \qquad i = 1,2,\ldots,m \tag{36}$$

Then $\hat{v} \triangleq (\hat{v}^1,\hat{v}^2,\ldots\hat{v}^m)$. The compromise program then is defined as

$$\min\{\|W(f(x)-\hat{v})\|_\ell \mid x \in \Omega,\ f(x) \in V_a\} , \qquad \ell \geq 1 \tag{37}$$

where W is a positive definite weighting matrix and, as before, V_a represents the set of acceptable performance values. A solution $(x^*, f(x^*))$ of (37) is called a compromise solution. Yu [28] has shown that, under mild assumptions, $x^* \in \Omega_N$ for any norm $\|\cdot\|_\ell$ with

$1 \leq \ell < \infty$. Compromise solutions are appealing, but, as pointed out by Yu [28] and Leitmann and Yu [29], compromise solutions are quite sensitive both to the units in which the criteria are expressed and to the particular norm $\|\cdot\|_\ell$ used. This is obviously a drawback and consequently Salukvadze [23] and Zeleny [30] have independently proposed that scaling be utilized to reduce this effect. Thus they suggest that the problem

$$\min\{\|WS(f(x)-\hat{v})\|_\ell \,|\, x \in \Omega,\ f(x) \in V_a\}\ , \qquad \ell \geq 1 \tag{38}$$

be solved instead of (37), with $S = \mathrm{diag}(1/\hat{v}^1, 1/\hat{v}^2, \ldots, 1/\hat{v}^m)$, provided no $\hat{v}^i = 0$. It is clear that while this eliminates the sensitivity of the solution of (38) to the units used, it makes the solution sensitive to the values of $\hat{v}^i$. Thus, while the compromise solution approach is an obviously attractive way for selecting a single noninferior alternative without constructing the entire set Ω_N, it is still not quite a perfect tool either.

The obvious conclusion is that the infinite alternative multicriteria decision problem is orders of magnitude more difficult than the finite alternative multicriteria decision problem and that methods for its solution still leave much to be desired. Thus, because of its great practical importance, the infinite alternative multicriteria problem represents an area of challenging research.

6. CONCLUSION

As we have seen, as long as the number of alternatives is finite (but not astronomically large), the multicriteria optimization problem is tractable. However, when the set of alternatives is a continuum and the number of criteria is larger than 3, the multicriteria optimization problem becomes tremendously more difficult and one's confidence in the soundness of the choice made by the various schemes proposed is nowhere near as great as in the finite case. In a way, the difficulty of the infinite alternative multicriteria optimization problem can be attributed to the fact that there are currently no entirely satisfactory methods for approximating such a problem by means of a finite discretized version of the problem. We can expect a good deal of future work to be devoted to the question of selecting a "grid" for the approximation of a continuum-type multicriteria optimization problem. Perhaps a connection of this new work with the old (such as in [20], [19]) will be based on the acceptance of points of confusion, which were discussed in Sec. 5, as a basis for selecting such a grid.

ACKNOWLEDGEMENT

Research sponsored by U.S. Army Research Office Durham Contract

DAHC04-73-C-0025, the National Science Foundation Grant ENG73-08214A01 and the Joint Services Electronics Program Contract F44620-71-C-0087.

REFERENCES

[1] Abelson, R. P., "The Choice of Choice Theories," in S. Messick and A. H. Brayfield (eds.), Decision and Choice, New York: McGraw-Hill, pp. 257-266, 1964.

[2] Athans, M., "The Role and Use of the Stochastic Linear-Quadratic-Gaussian Problem in Control System Design," IEEE Trans. Automat. Contr., vol. AC-16, pp. 529-552, Dec. 1971.

[3] Briskin, L. E., "A Method Unifying Multiple Objective Functions," Management Science, Vol. 12, pp. B406-B416, June 1966.

[4] Briskin, L. E., "Establishing a Generalized Multi-Attribute Utility Function," in J. L. Cochrane and M. Zeleny (eds.), Multiple Criteria Decision Making, Columbia, S. C.: USC Press, pp. 236-245, 1973.

[5] Charnes, A. and W. W. Cooper, Management Models and Industrial Application of Linear Programming, Vol. I, New York: Wiley, 1961.

[6] Consumer Reports, Vol. 39, pp. 164-171, Feb. 1974.

[7] Da Cunha, N. O. and E. Polak, "Constrainted Minimization Under Vector-Valued Criteria in Linear Topological Spaces," in A. V. Balakrishnan and L. W. Neustadt (eds.), Mathematical Theory of Control, New York: Academic Press, pp. 96-108, 1967.

[8] Dawes, R. M., "A Case Study of Graduate Admissions," American Psychologist, Vol. 26, pp. 180-188, Feb. 1971.

[9] Dinkelbach, W. and H. Isermann, "On Decision Making Under Multiple Criteria and Under Incomplete Information," in J. L. Cochrane and M. Zeleny (eds.), Multiple Criteria Decision Making, Columbia, S. C.: USC Press, pp. 302-312, 1973.

[10] Geoffrion, A., "Vector Maximal Decomposition Programming," Western Management Science Institute, Univ. of California, Los Angeles, Working Paper No. 164, Sept. 1970.

[11] Geoffrion, A. M., J. S. Dyer and A. Feinberg, "An Interactive Approach for Multi-Criterion Optimization with an Application to the Operation of an Academic Department," Management

Science, Vol. 19, pp. 357-368, Dec. 1972.

[12] Hanieski, J. F., "Technological Change as the Optimization of a Multidimensional Product," in J. L. Cochrane and M. Zeleny (eds.), Multiple Criteria Decision Making, Columbia, S. C.: USC Press, pp. 550-569, 1973.

[13] Huang, S. C., "Note on the Mean-Square Strategy of Vector Valued Objective Functions," JOTA, Vol. 9, pp. 364-366, May 1972.

[14] Kung, H. T., "On the Computational Complexity of Finding the Maxima of a Set of Vectors," Proc. 15th Annual IEEE Symp. on Switching and Automata Theory, pp. 117-121, Oct. 1974.

[15] Kung, H. T., F. Luccio and F. P. Preparata, "On Finding the Maxima of a Set of Vectors," J. ACM, Vol. 22, pp. 469-476, Oct. 1975.

[16] Luccio, F. and F. P. Preparata, "On Finding the Maxima of a Set of Vectors," Istituto di Scienze dell'Informazione, Università di Pisa, Pisa, Italy, Dec. 1973.

[17] Luenberger, D. G., Introduction to Linear and Nonlinear Programming, Reading, Mass.: Addison-Wesley, 1973.

[18] MacCrimmon, K. R., "An Overview of Multiple Objective Decision Making," in J. L. Cochrane and M. Zeleny (eds.), Multiple Criteria Decision Making, Columbia, S. C.: USC Press, pp. 18-44, 1973.

[19] Meisel, W. S., "Trade off Decisions in Multiple Criteria Decision Making," in J. L. Cochrane and M. Zeleny (eds.), Multiple Criteria Decision Making, Columbia, S. C.: USC Press, pp. 461-476, 1973.

[20] Payne, H. J., E. Polak, D. C. Collins and W. S. Meisel, "An Algorithm for Bicriteria Optimization Based on the Sensitivity Function," IEEE Trans. Automat. Contr., Vol. AC-20, pp. 546-548, Aug. 1975.

[21] Polak, E., "On the Approximation of Solutions to Multiple Criteria Decision Making Problems," Presented at XXII International Meeting of The Institute of Management Sciences, Kyoto, Japan, July 24-26, 1975.

[22] Russ, F., "Consumer Evaluation of Alternative Product Models," Ph.D. Dissertation, Carnegie-Mellon University, 1971.

[23] Salukvadze, M. E., "Optimization of Vector Functionals. I. The Programming of Optimal Trajectories," Automation and

Remote Control, Vol. 32, pp. 1169-1178, Aug. 1971.

[24] Smale, S., "Global Analysis and Economics I. Pareto Optimum and a Generalization of Morse Theory," in M. Peixoto (ed.), Dynamical Systems, New York: Academic Press, pp. 531-544, 1973.

[25] Wan, Y. H., "Morse Theory for Two Functions," Ph.D. Dissertation, Univ. of California, Berkeley, 1973.

[26] Wolfe, P., "Convergence Theory in Nonlinear Programming," in J. Abadie (ed.), Integer and Nonlinear Programming, New York: American Elsevier, pp. 1-36, 1970.

[27] Yao, F. F., "On Finding the Maximal Elements in a Set of Plane Vectors," Comput. Sci. Dep. Rep., University of Illinois at Urbana-Champaign, Urbana, Ill., July 1974.

[28] Yu, P. L., "A Class of Solutions for Group Decision Problems," Management Science, Vol. 19, pp. 936-946, April 1973.

[29] Yu, P. L. and G. Leitmann, "Compromise Solutions, Domination Structures and Salukvadze's Solution," JOTA, Vol. 13, pp. 362-378, Mar. 1974.

[30] Zeleny, M., "Compromise Programming," in J. L. Cochrane and M. Zeleny (eds.), Multiple Criteria Decision Making, Columbia, S. C.: USC Press, pp. 262-301, 1973.

[31] Zeleny, M., Linear Multiobjective Programming, Berlin: Springer-Verlag, 1974.