AIAA JOURNAL
Vol. 38, No. 6, June 2000

Ability of Objective Functions to Generate Points on Nonconvex Pareto Frontiers

Achille Messac* and Glynn J. Sundararaj†
Northeastern University, Boston, Massachusetts 02115
and
Ravindra V. Tappeta‡ and John E. Renaud§
University of Notre Dame, Notre Dame, Indiana 46556

New ground is broken in our understanding of objective functions' ability to capture Pareto solutions for multi-objective design optimization problems. It is explained why widely used objective functions fail to capture Pareto solutions when the Pareto frontier is not convex in objective space, and the means to avoid this limitation, when possible, is provided. These conditions are developed and presented in the general context of *n*-dimensional objective space, and numerical examples are provided. An important point is that most objective function structures can be made to generate nonconvex Pareto frontier solutions if the curvature of the objective function can be varied by setting one or more parameters. Because the occurrence of nonconvex efficient frontiers is common in practice, the results are of direct practical usefulness.

Nomenclature

J = aggregate objective function (a scalar)
v_i = ith entry of a generic vector $\boldsymbol{v}$
$\boldsymbol{w}$ = vector of numerical weights, used in forming the aggregate objective function
$\boldsymbol{x}$ = n-dimensional design parameter vector
$\boldsymbol{\theta}$ = m-dimensional design metric vector (also referred to as design criteria or design objectives vector)

I. Introduction

IN practical applications of optimization, particularly in the area of engineering design, the engineer is typically faced with the task of reconciling a set of disparate and conflicting objectives. The success of the optimization process is singularly dependent on the effectiveness of the objective function, which should possess the following two characteristics. First, the objective function should adequately represent the designer's preferences and objectives. Second, the objective function should have the ability to capture every point on the Pareto frontier, whether it is convex or not. The first characteristic is the subject of previous publications.[1–5] The second is the central topic of this paper.

The literature in the area of design optimization points to the inability of some commonly used objective functions to capture efficient solutions that lie in the nonconvex boundary of the feasible design space, even if the objective function is itself convex in objective space. In particular, numerous publications[6–12] address the related limitations of the weighted-sum (WS) objective function approach.

One of the major disadvantages of the WS method is that it fails to capture the points that are in the concave part of the Pareto frontier. In the work of Koski[8] and Chen et al.,[10] examples were presented to show that this easily occurs in structural optimization and robust design, respectively. In the work of Das and Dennis[7] the drawbacks of the WS method were discussed, and a geometrical explanation regarding why it fails to capture the points on the nonconvex frontier was given. In the work of Athan and Papalambros[6] and Das and Dennis[12] other methods that overcome this difficulty were discussed. The distinguishing contribution of this paper is the development of the necessary conditions for a (scalar) aggregate objective function (AOF) to capture a point in the nonconvex frontier, with a comprehensive presentation of the significant practical implications of the results.

We now explicitly define the problem we wish to address. In Fig. 1, we show a commonly occurring situation in the process of design optimization. For the sake of discussion, we consider the biobjective case. We show, however, that the important results of this paper are applicable to the general multi-objective case. We wish to minimize both objectives. As can be seen in Fig. 1, the segment a_1–a_2 of the curve defining the boundary of the feasible space is the efficient frontier. That is, for every point inside a_1–a_2, it is not possible to improve both objectives simultaneously. If one objective is improved, it must be at the expense of the other. Points inside a_1–a_2 are often referred to as Pareto points.[13] In view of their stated characteristics, Pareto points are usually the candidates of choice in the process of multi-objective optimization. Therefore, we can argue that every objective function should offer the possibility of capturing any existing Pareto point. In the absence of this possibility, a designer might be denied the ability to obtain many of the most desired design options.

At this point, two important questions are posed:

1) Is it possible to capture every Pareto point given the generic morphology of an objective function? In other words, is it possible, within the structure of a class of objective functions, to obtain a given Pareto point by altering the numerical values of the available free parameters? For example, if we are using the oft-used WS objective function, is it possible to capture a given Pareto point by changing the values of the numerical weights? This question is posed within the context of nonlinear programming. (The equally important case of linear programming brings to the fore issues that are sufficiently disparate to warrant their own distinct treatment.) Furthermore, we address the preceding question within the context of convex and nonconvex Pareto frontiers. This discussion leads us to the second question.

2) Given that the structure of an objective function allows us to capture a given Pareto point, how realistic is it to expect that we will be able to obtain the numerical values of the weights (free parameters) that indeed lead to that sought-after Pareto point? Stated differently, in a truly multi-objective setting (where the disparate objectives are typically nonlinearly coupled), can we discover the proper weights in reasonable time? The unfortunate answer to this

Received 22 June 1999; revision received 15 September 1999; accepted for publication 17 September 1999.

*Associate Professor, Mechanical Engineering Department, Multidisciplinary Design Laboratory; messac@coe.neu.edu. Associate Fellow AIAA.
†Graduate Student, Multidisciplinary Design Laboratory. Student Member AIAA.
‡Graduate Student, Aerospace and Mechanical Engineering Department, Design Automation Laboratory. Student Member AIAA.
§Associate Professor, Design Automation Laboratory. Senior Member AIAA.

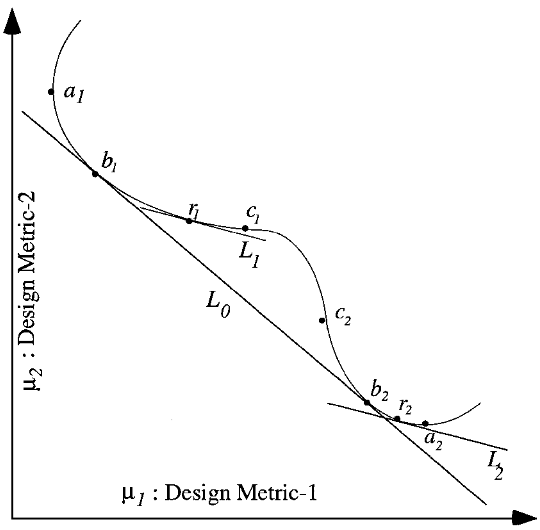

Fig. 1 Generic nonconvex Pareto frontier.

question is, usually not. The search for the proper weights usually takes place in an ad hoc environment that terminates by virtue of a time constraint, or sufficient satisfaction from the part of the designer with the current solution. The physical programming method directly addresses this question by entirely eliminating the search for proper weights from the optimization process.

These two questions are of primordial importance to the field of design optimization. In this paper, we primarily address the first. We first discuss the problem using two objectives, as this case allows for ready graphical interpretation. We then consider the general n-objective case. Several classes of objective functions are examined, for example, linear WS and general nonlinear.

It is important, in the following discussion, to carefully distinguish between the linearity of the AOF in objective space (e.g., WS method) and that of the criteria (design metrics) in decision space, that is, design parameter space. We shall always assume that the latter may be nonlinear. The designer may decide on the structure of the former, but generally not the latter. We shall also focus on the convexity of the AOF in objective space and on the convexity of the Pareto frontier, but not on that of the design metrics in design-parameter space. Again, the designer only has control over the former and can use that ability to capture the preferred design point. This paper discusses the related issues and shows how AOFs can be altered to capture points on a nonconvex segment of a Pareto frontier.

The remainder of this paper is organized as follows: Section II discusses the computational design optimization problem from a multi-objective perspective, introducing the role played by the Pareto frontier. Section III provides an analytical examination of the ability of a generic objective function to capture a Pareto point, with consideration of both two- and n-objective cases. The necessary and sufficient condition for capturability is provided. Section IV presents a set of applications of the capturability conditions of Sec. III within the context of numerical examples. Concluding remarks are provided in Sec. V.

II. Design Optimization: Multi-objective Perspective

This section discusses the computational design optimization problem from a multi-objective perspective, explores the various popular forms of AOFs used in practice, and sets the stage for Sec. III, where the capturability conditions are developed.

Mathematically, the multi-objective design optimization problem can be stated as follows:

$$\underset{x \in \Lambda}{\text{optimize}}\, \boldsymbol{\theta}(\boldsymbol{x}) = \begin{Bmatrix} \theta_1(\boldsymbol{x}) \\ \theta_2(\boldsymbol{x}) \\ \vdots \\ \theta_m(\boldsymbol{x}) \end{Bmatrix} \tag{1}$$

where, we recall, $\boldsymbol{\theta}$ and $\boldsymbol{x}$ are the design metric and design parameter vectors, respectively. For the sake of simplicity, we assume that we seek to minimize each of the design metrics. Most approaches to solving the given optimization problem entail some form of scalarization of the design metrics into a single AOF, which ultimately leads to a single answer. Unfortunately, this scalarization, which is intended to mathematically represent the designer's preferences, is extremely difficult to form correctly. Most initial attempts to form the AOF lead to a single final answer that often meets with the disapproval of the designer. Faced with this difficulty, some researchers[14, 15] have explored the possibility of identifying and representing the Pareto frontier as a way to choose from amongst a small set of good candidate solutions. An important component to addressing the difficulties associated with the formation of the AOF is the development of effective methods for visualizing the optimization process.[16]

This discussion leads us to some important concepts and questions. We will explore the notion of Pareto optimality from local and global perspectives. We will examine the means of generating the Pareto points (efficient solutions), and we note that, even when one is not interested in generating the Pareto frontier, the ability of an AOF to generate a generic efficient solution is critical. Every Pareto point is of potential interest to the designer and should, therefore, be capturable by the AOF being used during optimization. In essence, when an objective function is structurally unable to capture a region of the Pareto frontier, the designer is denied the opportunity to obtain a potential option of choice. As is discussed in the following text nearly all AOF structures are plagued with this deficiency.

The notion of Pareto optimality was introduced by Pareto in 1896.[13] A Pareto point is defined by a given value of the vector $\boldsymbol{\theta}$, for example, $\boldsymbol{\theta}_p$. The vector $\boldsymbol{\theta}_p$ represents a Pareto point if it is impossible to improve any of its entries without a simultaneous worsening of at least one other entry. From a logical point of view, it is, therefore, not possible to declare any Pareto point objectively better than any other Pareto point in objective space. Because of this realization, the concept of Pareto optimality has played an important role in the area of multi-objective optimization. As defined earlier, a Pareto point is such with respect to the entire design space. To assist us in subsequent discussions, note the distinction between the local and global minima that are generated by a generic objective function. This distinction will allow us to make two kinds of capturability properties. We call the first kind local capturability (leading to a local minimum); we call the second kind global capturability (leading to a global minimum). In this paper, when we simply say capturable, we imply local capturability.

We now use the oft-used biobjective case of Fig. 1, where both objectives are to be minimized. The curve segment a_1–a_2 represents the Pareto frontier. Every point on this segment is, therefore, of potential interest to a designer. Ideally, any AOF should have the ability to capture (i.e., yield through optimization) every point between a_1 and a_2. Unfortunately, because this Pareto frontier is not convex, it is well known that the popular WS approach fails to capture parts of the frontier. Mathematically, the WS approach is stated as

$$\underset{x \in \Lambda}{\text{optimize}}\, J(\boldsymbol{x}) = \sum_{i=1}^{m} w_i \mu_i(\boldsymbol{x}) \tag{2}$$

We now turn our attention to the following question: Which segments of the curve can the WS approach capture?

First, let us define the various points marked on the Pareto curve. The $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ components of the points a_1 and a_2, respectively, form the so-called utopia point, which is obtained by minimizing each of the design metrics independently. The points b_1 and b_2 are two equal local minima that are obtained for particular values of the weights. With small change in weight, the global optimum yields a large change. The points c_1 and c_2 represent inflection points of the Pareto frontier.

According to several publications, the answer to the preceding question is as follows: Segments a_1–b_1 and a_2–b_2 can be captured, and segment b_1–b_2 cannot be captured. Although the answer is technically correct with regard to global minimization through Eq. (2), it fails to clarify important aspects of the issue at hand.

We show in the following a simple example where the point r_1 can be captured through the WS approach, which would seem to contradict the preceding answers. Let us explain. First, we observe that, if the lines L_1 and L_2 are parallel and tangent to the Pareto curve, then point r_1 is a local minimum of Eq. (2), and point r_2 is the global minimum of the same. We assume here that the lines L_0, L_1, and L_2 represent lines of constant values of the WS objective function for particular values of the weights. It is, therefore, correct to state that the solution to the WS minimization represented by lines L_1 and L_2 is r_2 (because r_1 is a local minimum). In practice, it is highly likely that r_1, and a large part of segment b_1–b_2, will be captured by the WS approach.

The preceding observations lead us to the next question: Which part of segment b_1–b_2 can be captured at all by the WS approach? Here we shall take a pragmatic perspective and provide an answer that is of practical significance to the designer, especially one who uses a gradient-based optimization code that will generally yield local minima. We state that the segments b_1–c_1 and b_2–c_2 are capturable by the WS approach, in a departure from previous related discussions in the literature. (We prove this statement in the next section.) Emphasizing this important association of capturability property for the WS approach with the inflection point of the Pareto frontier is one contribution of this paper. In essence, the only section of the Pareto frontier that indeed cannot be captured by the WS approach is the c_1–c_2 segment, which is concave in objective space, rather than the whole segment b_1–b_2.

In the following discussion, we shall pay special attention to the popular WS approach and to other more flexible AOF structures. The WS AOF is arguably the simplest that we can examine in terms of the AOF's ability to capture the Pareto frontier. In fact, several publications in the related literature have partially addressed the issues discussed earlier. The next natural step is for us to examine the related properties of the other more flexible objective functions. Further, we will determine the general condition for capturability for any objective function, and provide the means to make practical use of the conditions provided.

In addition to the popular WS AOF approach, there are several other approaches that may at times be desirable. Note that the parameters G_i and B_i denote good and bad values of the pertaining design metric, w_i and c_i are adjustable parameters, and the subscripts min and max, respectively, denote minimum and maximum values of the particular design metric. Following are notable examples, with pertinent comments.

The absolute value method is

$$\underset{x\in\Lambda}{\text{minimize}}\, J(\boldsymbol{x}) = \sum_{i=1}^{m} \boldsymbol{w}_i |\mu_i(\boldsymbol{x}) - G_i| \tag{3}$$

The nonsmoothness of this approach is a source of numerical difficulties.

The weighted square sum (WSS) is

$$\underset{x\in\Lambda}{\text{minimize}}\, J(\boldsymbol{x}) = \sum_{i=1}^{m} \boldsymbol{w}_i \{\mu_i(\boldsymbol{x}) - G_i\}^2 \tag{4}$$

The nonzero second derivative of this approach gives it the ability to capture nonconvex Pareto points.

The weighted maximum is

$$\underset{x\in\Lambda}{\text{minimize}}\, J(\boldsymbol{x}) = \max_i \frac{\mu_i(\boldsymbol{x}) - G_i}{B_i - G_i} \tag{5}$$

Again, nonsmoothness is an issue.

The substitute objective function is

$$\underset{x\in\Lambda}{\text{minimize}}\, J(\boldsymbol{x}) = \prod_i \frac{\mu_{i,\max}(\boldsymbol{x}) - \mu_i(\boldsymbol{x})}{\mu_{i,\max}(\boldsymbol{x}) - \mu_{i,\min}(\boldsymbol{x})} \tag{6}$$

and it is relatively inflexible.

The Kreisselmeir–Steinhauser function is

$$\underset{x\in\Lambda}{\text{minimize}}\, J(\boldsymbol{x}) = \frac{1}{\rho}\,\ell n \sum_{i=1}^{m} \exp\{\rho\mu_i(\boldsymbol{x}) - \mu_{i,\max}\} \tag{7}$$

and it offers some flexibility through the parameter ρ.

The distance from utopia point is

$$\underset{x\in\Lambda}{\text{minimize}}\, J(\boldsymbol{x}) = \sum_{i=1}^{m} \boldsymbol{w}_i \{\mu_i(\boldsymbol{x}) - \mu_{i,\min}\}^2 \tag{8}$$

similar to WSS.

The exponential weighted method is

$$\underset{x\in\Lambda}{\text{minimize}}\, J(\boldsymbol{x}) = \sum_{i=1}^{m} \boldsymbol{w}_i \exp\{c_i\mu_i(\boldsymbol{x})\} \tag{9}$$

Two parameters can be manipulated to control the curvature.

The weighted compromise programming (WCP) is

$$\underset{x\in\Lambda}{\text{minimize}}\, J(\boldsymbol{x}) = \sum_{i=1}^{m} \boldsymbol{w}_i \{\mu_i(\boldsymbol{x})\}^{c_i} \tag{10}$$

As will be shown, the ability of an objective function to capture points on the Pareto frontier depends on the presence of some parameters that the designer can use to manipulate the function's curvature. By examining the given objective functions, we observe that some more readily lend themselves to increasing their curvature, thus increasing their ability to capture points in the nonconvex part of the Pareto frontier. For example, the parameter c_i in Eqs. (9) and (10) is ideally suited to effect significant changes in the objective function curvature. This parameter can, therefore, be used to help capture points that are on a highly concave Pareto frontier.

In the next section, we mathematically determine the condition for a generic objective function to capture points on the nonconvex part of the Pareto frontier.

III. Condition for Objective Function to Capture Nonconvex Pareto Frontier Points

This section provides the mathematical derivation of the condition the objective function must satisfy to capture the Pareto points on a given segment of the Pareto frontier. We also discuss the required relationship between the objective function and the Pareto frontier from a geometrical perspective. We provide an insightful geometrical interpretation of the mathematical results. Note that the following development applies to the regions of the Pareto frontier where the second derivative of the functions concerned exist.

We begin by considering the case of the WS method. This special case is then followed by the consideration of the general nonlinear objective function. We do so in the form of two propositions, which are proven to a degree that will satisfy an engineer but not a mathematician.

For the sake of the following development, we provide a concise definition for the terms locally capturable point and globally capturable point. A locally/globally capturable point is a Pareto point that is a local/global minimum of a given AOF for some setting of the parameters used in that function. In this paper, when the words locally and globally are omitted, we imply the use of the word locally. (Globally capturable points are also locally capturable.) For example, optimizing the biobjective problem of Fig. 1, the points between c_1 and c_2 are not capturable within the context of the linear WS method, but are capturable using several other methods. The capturability property of a given Pareto point is of practical implication because gradient-based optimization algorithms will generally yield that point as a solution if the starting point is in a sufficiently close neighborhood of that given Pareto point. That is, a gradient-based algorithm may yield point r_1 of Fig. 1, if the starting point is in its neighborhood. Otherwise, the point r_2 may be obtained instead. We now proceed with the development of the propositions.

A. Statement and Proof of Propositions 1 and 2

Proposition 1: A necessary and sufficient condition for a point to be capturable using a WS AOF is that the point lies on a convex part of the Pareto frontier. (We shall not herewith address endpoint effects or singularity issues.) For the sake of clarity, we first deal with the biobjective case, Proposition 1a; then we deal with the n-objective case, Proposition 1b.

Proposition 1a: The biobjective case proof is as follows:

The Pareto frontier can in general be represented by the equation

$$f[\boldsymbol{\theta}_1(\boldsymbol{x}), \boldsymbol{\theta}_2(\boldsymbol{x})] = 0 \tag{11}$$

We further assume that Eq. (11) can be rewritten in the form (with largely inconsequential loss of generality)

$$\boldsymbol{\theta}_2 \equiv \boldsymbol{\theta}_2(\boldsymbol{\theta}_1) \tag{12}$$

Proposition 1a1: We first prove that, if a generic point P is on a convex part of the frontier (e.g., point r_1 in Fig. 1), then an objective function of the form

$$J = \boldsymbol{w}_1\boldsymbol{\theta}_1 + \boldsymbol{w}_2\boldsymbol{\theta}_2 \tag{13}$$

can capture point P, where $\boldsymbol{w}_1 > 0$ and $\boldsymbol{w}_2 > 0$. The proof follows.

The value of the objective function for any point $\tilde{P}$, with coordinates $(\boldsymbol{u}_{1\tilde{p}}, \boldsymbol{u}_{2\tilde{p}})$, in the neighborhood of P can be expressed as

$$J_{\tilde{p}} \approx J_p + \boldsymbol{w}_1(\boldsymbol{u}_{1\tilde{p}} - \boldsymbol{u}_{1p}) + \boldsymbol{w}_2(\boldsymbol{u}_{2\tilde{p}} - \boldsymbol{u}_{2p}) \tag{14}$$

or

$$J_{\tilde{p}} \approx J_p + \boldsymbol{w}_1\Delta\boldsymbol{u}_1 + \boldsymbol{w}_2\Delta\boldsymbol{u}_2 \tag{15}$$

If we let

$$\frac{\boldsymbol{w}_1}{\boldsymbol{w}_2} = -\left.\frac{\mathrm{d}\boldsymbol{u}_2}{\mathrm{d}\boldsymbol{u}_1}\right|_p \tag{16}$$

Eq. (15) becomes

$$J_{\tilde{p}} \approx J_p + \boldsymbol{w}_1\left(1 - \frac{\Delta\boldsymbol{u}_2}{\Delta\boldsymbol{u}_1}\Big/\left.\frac{\mathrm{d}\boldsymbol{u}_2}{\mathrm{d}\boldsymbol{u}_1}\right|_p\right)\Delta\boldsymbol{u}_1 \tag{17}$$

Because the Pareto frontier is convex at point P,

$$\left.\frac{\mathrm{d}^2\boldsymbol{u}_2}{\mathrm{d}\boldsymbol{u}_1^2}\right|_p \geq 0 \tag{18}$$

which leads to

$$\left.\frac{\mathrm{d}\boldsymbol{u}_2}{\mathrm{d}\boldsymbol{u}_1}\right|_p \leq \frac{\Delta\boldsymbol{u}_2}{\Delta\boldsymbol{u}_1} \qquad \text{or} \qquad \left.\frac{\mathrm{d}\boldsymbol{u}_2}{\mathrm{d}\boldsymbol{u}_1}\right|_p \geq \frac{\Delta\boldsymbol{u}_2}{\Delta\boldsymbol{u}_1} \tag{19}$$

for $\tilde{P}$ on the right or left of P, respectively.

From Eqs. (17) and (19), it follows that, for $\tilde{p}$ on the right of P,

$$\Delta\boldsymbol{u}_1 > 0, \qquad J_{\tilde{p}} \geq J_p \tag{20}$$

Similarly, for $\tilde{P}$ on the left of P, we have

$$\Delta\boldsymbol{u}_1 < 0, \qquad \text{and still} \qquad J_{\tilde{p}} \geq J_p \tag{21}$$

Because we always have $J_{\tilde{p}} \geq J_p$, P is a local minimum and is, by definition, capturable by the objective function given in Eq. (13) when it lies on a convex part of the Pareto frontier, as was to be shown.

Proposition 1a2: We now prove that if an objective function of the form

$$J = \boldsymbol{w}_1\boldsymbol{\theta}_1 + \boldsymbol{w}_2\boldsymbol{\theta}_2 \tag{22}$$

can capture a generic point P, where $\boldsymbol{w}_1 > 0$ and $\boldsymbol{w}_2 > 0$, then the point P is on a convex part of the frontier, for example, point r_1 in Fig. 1.

The proof of Proposition 1a2 follows directly from that of Proposition 1a1. Essentially, we can simply follow the steps of Proposition 1a1 in reverse order.

Proposition 1b: The multi-objective case proof is as follows:

Following the preceding line of thought, the Pareto hypersurface can, in general, be represented by the equation

$$f[\boldsymbol{\theta}_1(\boldsymbol{x}), \boldsymbol{\theta}_2(\boldsymbol{x}), \ldots, \boldsymbol{\theta}_m(\boldsymbol{x})] = 0 \tag{23}$$

We assume that Eq. (23) can be rewritten in the form

$$\boldsymbol{\theta}_i \equiv \boldsymbol{\theta}_{if}(\boldsymbol{\theta}_{i-}) \tag{24}$$

where

$$\boldsymbol{\theta}_{i-} = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{i-1}, \boldsymbol{\theta}_{i+1}, \ldots, \boldsymbol{\theta}_m\}^T \tag{25}$$

Proposition 1b1: Again, we first prove that if a generic point P is on a convex part of the frontier, then an objective function of the form

$$J = \boldsymbol{w}^T\boldsymbol{\theta}, \qquad \boldsymbol{w} = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m\}^T \tag{26}$$

can capture point P, where $\boldsymbol{w}_i > 0\ \forall i$. The proof follows.

Proof:

We now rewrite Eq. (26) as

$$J = \boldsymbol{w}_{i-}^T\boldsymbol{\theta}_{i-} + \boldsymbol{w}_i\boldsymbol{\theta}_i, \qquad \boldsymbol{w}_{i-} = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{i-1}, \boldsymbol{w}_{i+1}, \boldsymbol{w}_m\}^T \tag{27}$$

or

$$\boldsymbol{\theta}_{i-\mathrm{obj}} \equiv -(1/\boldsymbol{w}_i)\boldsymbol{w}_{i-}^T\boldsymbol{\theta}_{i-} + (J/\boldsymbol{w}_i) \tag{28}$$

where $\boldsymbol{\theta}_{i-\mathrm{obj}}$ defines the AOF hyperplane.

Let us now form the function

$$\Gamma(\boldsymbol{\theta}_{i-}) \equiv \boldsymbol{\theta}_{if} - \boldsymbol{\theta}_{i-\mathrm{obj}} \tag{29}$$

An expansion of $\Gamma(\boldsymbol{\theta}_{i-})$ about the point P yields

$$\Gamma(\boldsymbol{\theta}_{i-}) \approx \Gamma(\boldsymbol{\theta}_{i-})|_p + \nabla\Gamma(\boldsymbol{\theta}_{i-})|_p^T\{\boldsymbol{\theta}_{i-} - (\boldsymbol{\theta}_{i-})|_p\} + \tfrac{1}{2}\{\boldsymbol{\theta}_{i-} - (\boldsymbol{\theta}_{i-})|_p\}^T H[\Gamma(\boldsymbol{\theta}_{i-})]|_p\{\boldsymbol{\theta}_{i-} - (\boldsymbol{\theta}_{i-})|_p\} \tag{30}$$

We now prescribe the value of J in Eq. (26) and the set of $\boldsymbol{w}_i$ such that

$$\Gamma(\boldsymbol{\theta}_{i-})|_p = 0 \tag{31}$$

$$\nabla\Gamma(\boldsymbol{\theta}_{i-})|_p = 0 \tag{32}$$

Equation (26) then forms a hyperplane that supports (is tangent to) the Pareto hypersurface at point P. Equation (30) now becomes

$$\Gamma(\boldsymbol{\theta}_{i-}) \approx \tfrac{1}{2}\{\boldsymbol{\theta}_{i-} - (\boldsymbol{\theta}_{i-})|_p\}^T H[\Gamma(\boldsymbol{\theta}_{i-})]|_p\{\boldsymbol{\theta}_{i-} - (\boldsymbol{\theta}_{i-})|_p\} \tag{33}$$

We note that

$$H[\Gamma(\boldsymbol{\theta}_{i-})]|_p = H(\boldsymbol{\theta}_{if})|_p - H(\boldsymbol{\theta}_{i-\mathrm{obj}})|_p = H(\boldsymbol{\theta}_{if})|_p - 0 \tag{34}$$

Because $H(\boldsymbol{\theta}_{if})|_p$ is positive semidefinite, by assumption (the Pareto frontier is convex), then $\Gamma(\boldsymbol{\theta}_{i-}) \geq 0$ in the neighborhood of the point P. Equations (33) and (29) then yield

$$\boldsymbol{\theta}_{if} \geq \boldsymbol{\theta}_{i-\mathrm{obj}} \tag{35}$$

which implies that point P is a local minimum, and is, therefore, capturable.

Proposition 1b2: We now prove that if an objective function of the form

$$J = \boldsymbol{w}^T\boldsymbol{\theta}, \qquad \boldsymbol{w} = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m\}^T \tag{36}$$

can capture a generic point P, where $\boldsymbol{w}_i > 0\ \forall i$, then the point P is on a convex part of the Pareto frontier as is shown in the following proof.

Proof:

As in the preceding case, the proof of Proposition 1b2 follows generally from that of Proposition 1b1. We proceed as follows.

Because an objective function of the form of Eq. (36) captures the point P, then P is a local minimum. Therefore, Eq. (35) holds. Using Eq. (29), we write $\Gamma(\boldsymbol{\theta}_{i-}) \geq 0$. Again, Eqs. (31) and (32) hold because P is a local minimum. It follows that Eq. (33) holds, and because $\Gamma(\boldsymbol{\theta}_{i-}) > 0$, then $H[\Gamma(\boldsymbol{\theta}_{i-})]|_p$ must be positive semidefinite. From Eq. (34), we conclude that $H(\boldsymbol{\theta}_{if})|_p$ must be positive semidefinite. The point P is, therefore, at a convex part of the Pareto frontier, as was to be shown.

This proof concludes the validation of Proposition 1, which also formally validates the discussion of Sec. II. That discussion addressed the ability of the linear WS objective function from a practical perspective. Next, we provide the necessary and sufficient condition for a general nonlinear AOF to capture a point on a Pareto frontier.

Proposition 2: Let $\boldsymbol{\theta}_{if} \equiv \boldsymbol{\theta}_{if}(\boldsymbol{\theta}_{i-})$ and $\boldsymbol{\theta}_{i-\mathrm{obj}} \equiv \boldsymbol{\theta}_{i-\mathrm{obj}}(\boldsymbol{\theta}_{i-})$, for a given point P, represent the hypersurfaces of the Pareto hypersurface and of the AOF, respectively. The necessary and sufficient condition for a point P to be capturable is that the Hessian of $(\boldsymbol{\theta}_{if} - \boldsymbol{\theta}_{i-\mathrm{obj}})$ be positive semidefinite and the gradient of $(\boldsymbol{\theta}_{if} - \boldsymbol{\theta}_{i-\mathrm{obj}})$ vanish.

Proof:

Define

$$\Gamma(\boldsymbol{\theta}_{i-}) \equiv \boldsymbol{\theta}_{if} - \boldsymbol{\theta}_{i-\mathrm{obj}} \tag{37}$$

and expand $\Gamma(\boldsymbol{\theta}_{i-})$ about the point P, which yields

$$\Gamma(\boldsymbol{\theta}_{i-}) \approx \Gamma(\boldsymbol{\theta}_{i-})|_p + \nabla\Gamma(\boldsymbol{\theta}_{i-})|_p^T\{\boldsymbol{\theta}_{i-} - (\boldsymbol{\theta}_{i-})|_p\} + \tfrac{1}{2}\{\boldsymbol{\theta}_{i-} - (\boldsymbol{\theta}_{i-})|_p\}^T H[\Gamma(\boldsymbol{\theta}_{i-})]|_p\{\boldsymbol{\theta}_{i-} - (\boldsymbol{\theta}_{i-})|_p\} \tag{38}$$

Note that, to define the implicit functions $\boldsymbol{\theta}_{if}$ and $\boldsymbol{\theta}_{i-\mathrm{obj}}$, we let

$$\Gamma(\boldsymbol{\theta}_{i-})|_p = 0 \tag{39}$$

Proposition 2a: We prove that if the point P is capturable, then the Hessian of $(\boldsymbol{\theta}_{if} - \boldsymbol{\theta}_{i-\mathrm{obj}})$ is positive semidefinite.

Because the point P is capturable, we must have the ability to prescribe the free parameters of the objective functions such that

$$\nabla\Gamma(\boldsymbol{\theta}_{i-})|_p = 0 \tag{40}$$

Equation (38) then becomes

$$\Gamma(\boldsymbol{\theta}_{i-}) \approx \tfrac{1}{2}\{\boldsymbol{\theta}_{i-} - (\boldsymbol{\theta}_{i-})|_p\}^T H[\Gamma(\boldsymbol{\theta}_{i-})]|_p\{\boldsymbol{\theta}_{i-} - (\boldsymbol{\theta}_{i-})|_p\} \tag{41}$$

Because the point P is capturable, P is by definition a local minimum. Then we have $\boldsymbol{\theta}_{if} \geq \boldsymbol{\theta}_{i-\mathrm{obj}}$, and, therefore, $\Gamma(\boldsymbol{\theta}_{i-}) \geq 0$. From Eq. (41), we conclude that the Hessian

$$H[\Gamma(\boldsymbol{\theta}_{i-})]|_p = H(\boldsymbol{\theta}_{if} - \boldsymbol{\theta}_{i-\mathrm{obj}})|_p \tag{42}$$

must be positive semidefinite, as was to be shown.

Proposition 2b: We now prove that if the Hessian of $(\boldsymbol{\theta}_{if} - \boldsymbol{\theta}_{i-\mathrm{obj}})$ is positive semidefinite, then the point P is capturable.

The proof can be performed by generally invoking the steps of Proposition 2a in reverse. Because the Hessian is positive semidefinite and the gradient vanishes, we have $\Gamma(\boldsymbol{\theta}_{i-}) \geq 0$, and, therefore, $\boldsymbol{\theta}_{if} \geq \boldsymbol{\theta}_{i-\mathrm{obj}}$. This implies that P is a local minimum.

This development completes the proofs of the two propositions of this section. The first addressed the special case of the WS objective function, and the second dealt with the general objective function. The former resulted in specific conditions whereas the latter yielded more general conditions. Through numerical examples, we show the practical implications of these propositions, that is, how they can be helpful in practice.

B. Geometrical Interpretation of Propositions

Next, we provide an interesting geometrical interpretation of the two propositions developed in this section. These interpretations will also be of important practical usefulness.

We use Fig. 1 to provide our geometrical interpretation of Proposition 1. We recall that Proposition 1 delineated the capabilities of the WS AOF. The curve in Fig. 1 has two inflection points, at c_1 and c_2. The curve is concave between those two points and convex elsewhere. According to Proposition 1, all of the points on the convex parts of the curve are capturable, for example, a_1, b_1, r_1. We now make the following observations: 1) The Pareto frontier and the objective function meet at the capturable point (as expected). 2) In the neighborhood of every capturable point, the objective function lies on one side and the Pareto frontier on the other side. 3) The Pareto frontier lies on the increasing-value, (that is, northeast) side of the objective function.

From these observations, we can see why the WS objective function is incapable of capturing points on the concave part of the Pareto frontier. On the convex part of the Pareto frontier, the objective function lies on the correct side. On the concave side it does not. In the case of the inflection point, the Pareto frontier and the objective function do not lie on exclusive sides. They cross each other. From the discussions of this section, we see why some points between b_1 and b_2 are capturable, for example, between b_1 and c_1 and between c_2 and b_2. For a given set of weights, the points r_1 and r_2 would be capturable. The former would be a local minimum, and the latter would be the global minimum. In the process of generating Pareto points using a gradient-based optimization algorithm, all points up to and excluding the inflection points would be capturable, for example, above c_1 and below c_2.

The preceding observations lead us to a more general geometrical condition for the case of a general nonlinear AOF. This general case is the subject of Proposition 2. Proposition 2 states that the Hessian of the function that is the difference between the Pareto frontier and the AOF is positive semidefinite at capturable points. Stated differently, the difference between the Pareto frontier and the AOF is a convex function in the neighborhood of a capturable point. A geometrical interpretation of Proposition 2 is obtained by considering the biobjective case. Proposition 2 in this case states that the second derivative of the Pareto frontier must be greater than that of the objective function at a capturable point.

The given maximum level of convexity (or curvature) of an objective function is an inherent factor regarding its ability to capture Pareto points. That is, the objective function can capture all of the points of the Pareto frontier for which the second derivative is greater than that of the objective function. In essence, the more concave the objective function in objective space, the more it is capable of capturing Pareto points.

An important practical consequence of the preceding observations is that an objective function whose curvature can be increased by using its free parameters, for example, exponents, can be made accordingly more effective at capturing Pareto solutions. Numerical examples in the next section graphically illustrate these observations.

IV. Numerical Examples

This section provides three examples that illustrate the propositions of the preceding section. The developments of this section also reveal the means of exploiting the findings of this paper in practice. The first example concerns Proposition 1, the WS method. The second and third examples address the case of a general nonlinear objective function.

Example 1a

Consider the case of minimizing the WS objective function

$$J(\boldsymbol{x}) = w_1\boldsymbol{\theta}_1(\boldsymbol{x}) + w_2\boldsymbol{\theta}_2(\boldsymbol{x}) \tag{43}$$

where

$$\boldsymbol{\theta}_1(\boldsymbol{x}) = \exp(-\boldsymbol{x}) + 1.4\exp(-\boldsymbol{x}^2) \tag{44}$$

$$\boldsymbol{\theta}_2(\boldsymbol{x}) = \exp(\boldsymbol{x}) + 1.4\exp(-\boldsymbol{x}^2) \tag{45}$$

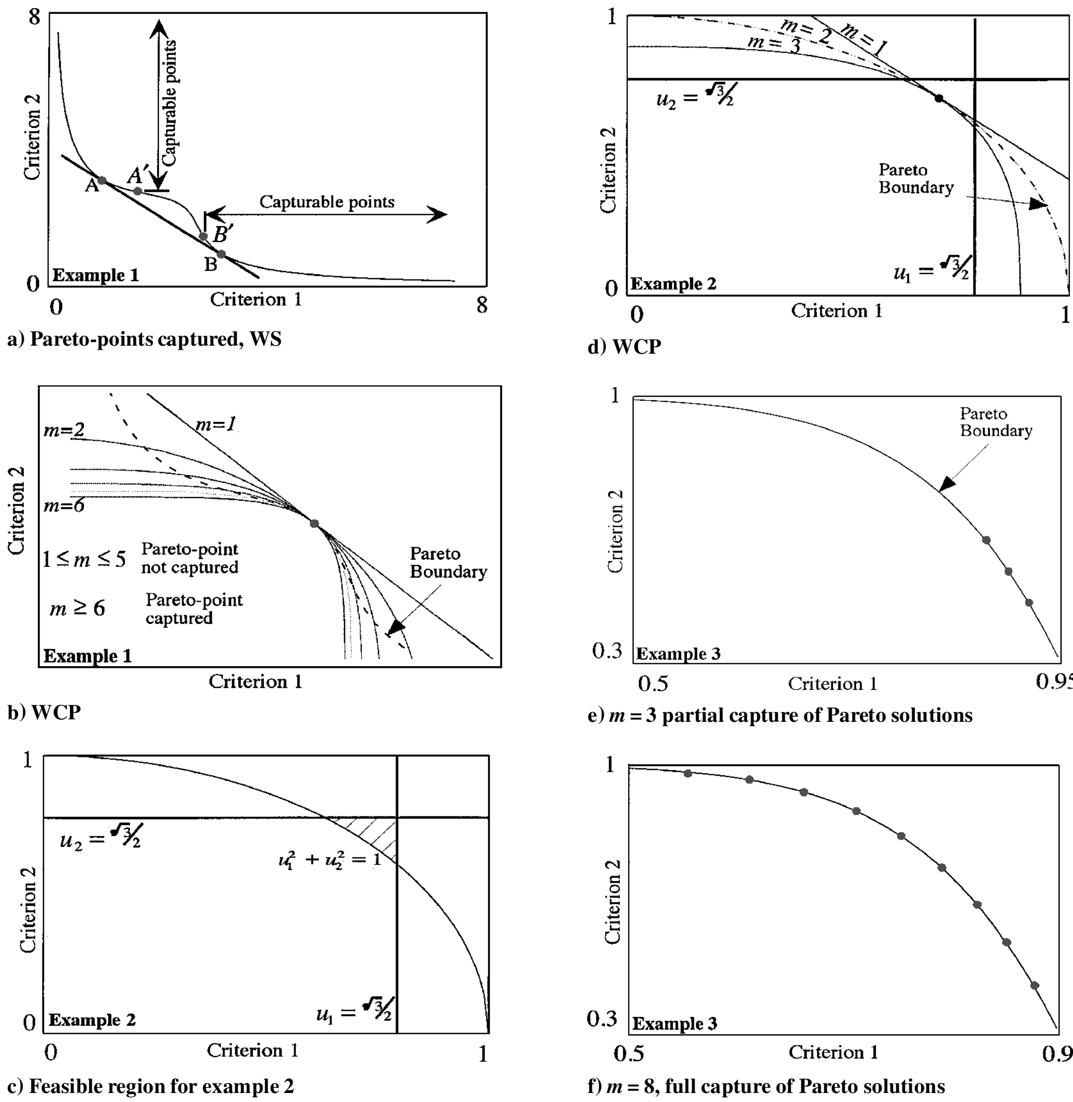

Fig. 2 Manipulating the AOF to capture Pareto solutions.

The resulting Pareto frontier is shown in Fig. 2a. By varying the weights, we capture any of the Pareto points above A' and below B'. Note that these are the inflection points of the Pareto frontier in objective space. We again note that the points between A and A', and between B and B' are indeed capturable. Because the part of the curve that is between A' and B' is concave, its second derivative is negative. Also, the second derivative of the (WS) AOF is zero. Accordingly, the inequality of Proposition 2 is violated for all points between A' and B', making these points not capturable. These results also verify Proposition 1.

To capture the points between A' and B', we must use an objective function that offers the ability to change its convexity. We explore that possibility by performing the minimization using the Weighted Compromise objective function.

Example 1b

Consider the case of minimizing the objective function

$$J(x) = w_1\theta_1^m + w_2\theta_2^m \tag{46}$$

where

$$\theta_1(x) = \exp(-x) + 1.4\exp(-x^2) \tag{47}$$

$$\theta_2(x) = \exp(x) + 1.4\exp(-x^2) \tag{48}$$

The results are depicted in Fig. 2b for values of m ranging from 1 through 6. Curves representing each of these values of m are represented. By using the values of m ranging from 1 through 5, the Pareto point shown could not be captured. For values of m of 6 and above, the Pareto point could be captured. We verified that for $m = 6$, Proposition 2 is satisfied.

Note that, in practice, it will be generally difficult to determine when Proposition 2 is or is not satisfied. However, our understanding of the inherent factors at play here will allow us to do our work more effectively. We know that increasing the curvature of the objective function will make it more effective, and increasing the parameter m will arbitrarily increase the objective function's curvature. Therefore, in practice we are able to capture any Pareto point by simply increasing m up to the needed value.

Example 2

Consider the case of minimizing the objective function

$$J(x) = w_1\theta_1^m + w_2\theta_2^m \tag{49}$$

where

$$\theta_1^2 + \theta_2^2 \geq 1 \tag{50}$$

$$\theta_1 \leq \sqrt{3}/2 \tag{51}$$

$$\theta_2 \leq \sqrt{3}/2 \tag{52}$$

In this case, because the functional forms of the objective function and of the first constraint are similar, it is relatively simple to satisfy Proposition 2. We simply need to have $m > 2$. Figure 2c depicts the feasible region in question, and Fig. 2d depicts the related curves. We let $w_1 = w_2 = 0.5$, and the results in Fig. 2d are self-explanatory. For $m = 3$, the second derivative of the objective function is -5.657 and that of the Pareto frontier is -2.828, thereby satisfying Proposition 2. To obtain these numbers, we used the expression

$$\frac{d^2\theta_2}{d\theta_1^2} = -\frac{w_1 w_2 (m-1)\theta_1^{m-2}\theta_2^m + w_1^2(m-1)\theta_1^{2(m-1)}}{w_2^2\theta_2^{2m-1}} \quad (53)$$

for the objective function and

$$\frac{d^2\theta_2}{d\theta_1^2} = -\frac{1}{\left(1-\theta_1^2\right)^{\frac{3}{2}}} \quad (54)$$

for the Pareto frontier. The values of the design metrics at the optimum are used in the preceding two expressions to verify that the inequality of Proposition 2 holds. Figures 2c and 2d illustrate this situation.

Example 3

In this example, we consider a more general situation where the Pareto frontier is concave and of changing curvature. The problem is as follows:

$$\underset{\theta}{\text{minimize}}\, J\{\theta_1(\theta), \theta_2(\theta)\} \quad (55)$$

where

$$\theta_1 = \sin\theta, \qquad \theta_2 = 1 - \sin^7\theta, \qquad 0.5236 \le \theta \le 1.2532$$

From the earlier discussions in this paper, we know that the WS approach will fail because the Pareto frontier is concave. We will instead consider two other options: the WSS and the compromise programming.

For the WSS approach, we use the objective function

$$J(x) = w_1\{\mu_1(x) - G_1\}^2 + w_2\{\mu_2(x) - G_2\}^2 \quad (56)$$

Choosing the good values to be the minimum values of the objectives, we have $G_1 = 0.5$ and $G_2 = 0.3017$. With these values, we were able to capture the full Pareto frontier. We also verified that Proposition 2 was satisfied. We also explored the possibility of using different good values, for example, greater than the minimum values used in this example. For these cases, changing the weights resulted in only partial generation of the Pareto frontier. As we know from the developments in this paper, we need the ability to change the curvature of the objective function, which is given by

$$\frac{d^2\theta_2}{d\theta_1^2} = -\frac{w_1 w_2(\theta_2 - G_2)^2 + w_1^2(\theta_1 - G_1)^2}{w_2^2(\theta_2 - G_2)^3} \quad (57)$$

The second derivative of the Pareto frontier is given by

$$\frac{d^2\theta_2}{d\theta_1^2} = -42\theta_1^5 \quad (58)$$

Although the good values do have some influence over the curvature, it is not clear how to effectively exploit that opportunity. A more direct way would be to change the exponents of the objective function. Another direct way to change the curvature also exists in the physical programming method, where a convexity parameter explicitly serves that purpose.

Our next approach to solving Example 3 is to use the WCP objective function, which is given by

$$J(x) = w_1\mu_1^m + w_2\mu_2^m \quad (59)$$

The results are as follows. For values of m less than eight, we could not capture the full Pareto frontier. In Fig. 2e, we show some of the Pareto points that could be captured. The range of capturable points is limited. The range of capturable points increases as m is increased, until m reaches the value of eight. At that point, all Pareto points are capturable. Again in this case, we verified that Proposition 2 defines the limit of capturability. That is, when Proposition 2 was violated the points were not capturable, and the converse was also true.

It is in the spirit of the preceding comments that the practical value of this paper is rooted. In general, it would be extremely impractical to have to compute the appropriate quantities to verify whether or not Propositions 1 and 2 are satisfied. We do not advise the reader to perform such a difficult task. Instead, the results of this paper explain and identify the origins of the inherent limitations of a given class of objective functions. In addition, this paper uncovers ways to circumvent these limitations, if the given class of objective functions allows. If a designer suspects that there are solutions that are available but that are not being captured by changing the objective function weights, then the designer can increase the parameter m (in the case of the compromise programming). Doing so increases the likelihood of satisfying Proposition 2. Stated differently, we can make practical use of Proposition 2, not by having to evaluate the curvatures of the Pareto frontier and of the objective function, but by varying the parameter m, which we know can significantly change curvature. What was needed is a qualitative understanding of Proposition 2, together with an understanding of the structure of the objective function (namely, that the latter's curvature will increase as m increases).

V. Conclusion

This paper expanded our understanding regarding why some objective functions fail to capture Pareto solutions. Because our ability to capture Pareto points is centrally important to the fields of design optimization and multi-objective optimization, a deep understanding of different classes of objective functions is of practical importance. This paper uncovered the role that convexity plays in the ability of objective functions to yield Pareto solutions, not simply in the case of linear WS method but also in the case of an arbitrary objective function. Quantitative conditions are derived to determine when a Pareto point is or is not capturable by a given objective function. In practice, we can increase the capability of objective functions to capture Pareto points by simply changing available parameters that increase the functions' curvatures. The results of this paper are of significant theoretical and practical value.

Acknowledgments

We gratefully acknowledge the support of the National Science Foundation through Grant DMI 97-02248 for Achille Messac and Grant DMI 98-12857 for John E. Renaud. This multidisciplinary research project is a joint effort between Northeastern University and the University of Notre Dame.

References

[1] Messac, A., "Physical Programming: Effective Optimization for Computational Design," *AIAA Journal*, Vol. 34, No. 1, 1996, pp. 149–158.

[2] Messac, A., Gupta, S., and Akbulut, B., "Linear Physical Programming: Effective Optimization for Complex Linear Systems," *Transactions on Operational Research*, Vol. 8, Oct. 1996, pp. 39–59.

[3] Messac, A., "From Dubious Construction of Objective Functions to the Application of Physical Programming," *AIAA Journal*, Vol. 38, No. 1, 2000, pp. 155–163.

[4] Messac, A., and Wilson, B., "Physical Programming for Computational Control," *AIAA Journal*, Vol. 36, No. 2, 1998, pp. 219–226.

[5] Messac, A., "Control-Structure Integrated Design with Closed-Form Design Metrics Using Physical Programming," *AIAA Journal*, Vol. 36, No. 5, 1998, pp. 855–864.

[6] Athan, T. W., and Papalambros, P. Y., "A Note on Weighted Criteria Methods for Compromise Solutions on Multiobjective Optimization," *Engineering Optimization*, No. 2, Vol. 27, 1996, pp. 155–176.

[7] Das, I., and Dennis, J. E., "A Closer Look at Drawbacks of Minimizing Weighted Sums of Objectives for Pareto Set Generation in Multicriteria Optimization Problems," *Structural Optimization*, Vol. 14, No. 1, 1997, pp. 63–69.

[8] Koski, J., "Defectiveness of Weighting Method in Multicriterion Optimization of Structures," *Communications in Applied Numerical Methods*, Vol. 1, No. 6, 1985, pp. 333–337.

[9] Osyczka, A., *Multicriterion Optimization in Engineering with Fortran Programs*, Wiley, New York, 1984, pp. 43–47.

[10] Chen, W., Wiecek, M. M., and Zhang, J., "Quality Utility—A Compromize Programming Approach to Robust Design," *Proceedings of the DETC'98, ASME Design Engineering Technical Conferences*, American Society of Mechanical Engineers, Fairfield, NJ, 1998.

[11] Rakowska, J., Haftka, R. T., and Watson, L.T., "Tracing the Efficient Curve for Multi-Objective Control-Structure Optimization," *Computing Systems in Engineering*, Vol. 2, No. 6, 1991, pp. 461–471.

[12] Das, I., and Dennis, J. E., "Normal-Boundary Intersection: A New Method for Generating the Pareto Surface in Nonlinear Multicriteria Optimization Problems," *SIAM Journal of Optimization*, Vol. 8, No. 3, 1998, pp. 631–657.

[13] Pareto, V., *Manuel d'Economie Politique*, Giard, Paris, 1896.

[14] Tappeta, R. V., and Renaud, J. E., "Interactive Multiobjective Optimization Procedure," *Proceedings of the 40th AIAA/ASME/ASCEAHS/ASC Structures, Structural Dynamics, and Materials Conference*, AIAA, Reston, VA, 1999, pp. 53–67.

[15] Tappeta, R. V., Renaud, J. E., Messac, A., and Sundararaj, G. J., "Interactive Physical Programming: Tradeoff Analysis and Decision Making in Multicriteria Optimization," *AIAA Journal*, Vol. 38, No. 5, 2000, pp. 917–926.

[16] Messac, A., and Chen, X., "Visualizing the Optimization Process in Real-Time Using Physical Programming," *Engineering Optimization Journal* (to be published).

A. Chattopadhyay
Associate Editor