AIAA JOURNAL
Vol. 39, No. 11, November 2001

Required Relationship Between Objective Function and Pareto Frontier Orders: Practical Implications

Achille Messac* and Amir Ismail-Yahaya†
Rensselaer Polytechnic Institute, Troy, New York 12180-3590

It is well known that the weighted sum aggregate objective function fails to capture Pareto points that are located on a concave region of the Pareto frontier. Conditions are established for a general aggregate objective function to capture points on a concave Pareto frontier region. Specifically, these conditions are developed in terms of the relative orders of the Pareto frontier and aggregate objective function. These critical conditions can be used in practice to allow an aggregate objective function to capture any potentially desirable Pareto point. Conversely, failure to satisfy these critical conditions can represent a pivotal failure point of the general application of design optimization because we may fail to obtain the design we seek.

I. Introduction

NUMEROUS recent publications have focused on the ability of aggregate objective functions (AOF) to capture Pareto[1] points on concave regions of Pareto frontiers. The drawbacks of the weighted sum (WS) AOF are examined by Das and Dennis,[2] Koski,[3] and Messac et al.[4] Among the undesirable properties of the WS are that 1) an even distribution of scalar weights in the AOF does not generally yield an even distribution of optimum points and that 2) the spacing of the optimum points is largely dependent on the relative scaling of the design metrics. Note that in this paper we will use the term design metric rather than the term design objective or design criteria. In addition, we will use the term design parameter rather than the term design variable or decision variables. We also use AOF to differentiate succinctly between a single design metric and the aggregation of a collection thereof, which is the quantity ultimately minimized. (For an interesting discussion of lexicon-related issues in design, sec Messac and Chen.[5]) It is shown via examples by Das and Dennis[6] that the normal boundary intersection (NBI) method generates evenly spaced Pareto optimal points and that the spacing of the points is independent of the relative scaling of the design metrics. NBI is not limited to biobjective problems and is primarily used for tracing the Pareto frontier. In a similar vein, the physical programming method[7–14] is shown to generate a well-distributed set of Pareto points for an evenly varying set of input parameters.[14]

In other works, different AOF approaches are shown to overcome the drawbacks of the WS. These include the weighted Tchebycheff method (also called compromise programming[15]) and the exponential weighted criteria method.[16] These works are, in general, closely associated with the notion of generation of the Pareto frontier. In particular, Rao and Papalambros[17] and Rakowska et al.[18] develop a homotopy-based strategy for tracing the Pareto frontier. Related work can also be found by Lin[19] and Chen et al.[20]

Closely related to the present paper is a recent work by Messac et al.,[21] where required relationships are developed for AOFs to capture generic points on a Pareto frontier. These conditions are theoretical in nature and only point to general guidance in the formation of AOFs. One of the key results of that work is that the Hessian of the difference between the Pareto frontier and AOF at an optimum point must be positive semidefinite, when this difference is implicitly expressed in terms of a generic design metric. Although useful, the results of Ref. 21 cannot readily and explicitly be used in practice. The present paper, in contrast, develops explicit conditions that can be used in practice in a direct way. Other related work addressing Pareto frontier issues includes that of Fadel and Li.[22]

It may be helpful at this point to comment on the practical implications of this paper. As we know, one of the critical phases of the design optimization process is the formation of the AOF. The other important phases include the development of the design metrics in terms of the design variables. This latter phase might include using such tools as commercial finite element codes or user-written software modules. The important point here is that, even if this challenging task is performed correctly and efficiently, unless the appropriate AOF is used, the designer will fail to obtain designs that might be desirable. Specifically, if the structure of the AOF inherently does not allow for obtaining a given desirable solution, then a different structure should be used. Examples of these inappropriate AOFs include 1) using the WS AOF for a concave Pareto frontier or 2) using the weighted square sum (WSS) AOF in the case of a Pareto frontier that is highly concave,[12] one with a strong curvature. These drawbacks have been discussed in the literature largely within the context of the WS method.[2] This paper extends this discussion to the case of AOFs of order n and develops conditions that can be used in practice directly to ensure that the AOF can capture any point on the Pareto frontier. Interestingly, in the process of this development, we shall answer the following question: What is the required relationship between the power of a design metric in the AOF and that of the same design metric in the Pareto frontier? This paper provides guidance of pragmatic consequence to the optimization practitioner.

This paper is organized as follows. Section II presents the main analytical development, where the general conditions are derived for an AOF to capture points of a Pareto frontier. Associated practical consequences are examined from the perspective of providing practical guidance to the design optimization practitioner. Numerical results and discussions are provided in Sec. III and concluding remarks in Sec. IV.

II. Analytical Development

In this section, we develop conditions that govern the relationship between the order of the AOF and that of the Pareto frontier, for a given class of assumed functional forms. These conditions yield useful practical findings. We present these developments in two progressive steps. In the first, we assume that the Pareto frontier takes the form of a polynomial where each design metric appears in the form of a single term with a given exponent. This analysis provides important preliminary findings. This analytical development is followed by a more general case where the Pareto frontier is a separable polynomial composed of an additive form of polynomials of individual design metrics. The latter development allows for a more comprehensive set of findings that have important practical application.

Received 12 January 2001; revision received 30 April 2001; accepted for publication 1 May 2001.

*Associate Professor, Department of Mechanical Engineering, Aeronautical Engineering and Mechanics, Multidisciplinary Design and Optimization Laboratory, 110 8th Street; messac@rpi.edu. Associate Fellow AIAA.

†Graduate Student, Department of Mechanical Engineering, Aeronautical Engineering and Mechanics, Multidisciplinary Design and Optimization Laboratory, 110 8th Street.

We begin by defining the general multiobjective optimization problem (P1):

$$\min_{x \in D} \mu(\boldsymbol{x}) = \begin{pmatrix} \mu_1(\boldsymbol{x}) \\ \mu_2(\boldsymbol{x}) \\ \vdots \\ \mu_m(\boldsymbol{x}) \end{pmatrix} \tag{1}$$

where $D = \{\boldsymbol{x} \in R^n \mid \boldsymbol{h}(\boldsymbol{x}) = 0,\ \boldsymbol{g}(\boldsymbol{x}) \leq 0,\ \boldsymbol{x}_{\min} \leq \boldsymbol{x} \leq \boldsymbol{x}_{\max}\}$; with $\boldsymbol{h}: R^n \rightarrow R^{n_e}$; $\boldsymbol{g}: R^n \rightarrow R^{n_i}$; $\boldsymbol{x}_{\min} \in (R \cup \{-\infty\})^n$; $\boldsymbol{x}_{\max} \in (R \cup \{+\infty\})^n$; m is the number of design metrics; $m \geq 2$, n_e and n_i are the numbers of equality and inequality constraints, respectively; and μ_i is at least twice differentiable for $i \in \{1, \ldots, m\}$. For any design parameter vector $\boldsymbol{x} = (x_1, \ldots, x_n)$, a design metric vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)$ is defined according to the function $\boldsymbol{\mu}: R^n \rightarrow R^m$. We now define relevant feasible regions. Let $Z = \{z \in R^m \mid z = \mu(x),\ \boldsymbol{x} \in D\}$ be the set of images of all points in D, where D is the feasible region in design parameter space and Z the feasible region in design metric space; $\mu_1(\boldsymbol{x}), \ldots, \mu_m(\boldsymbol{x})$ are the coordinates of the image of $\boldsymbol{x}$ in the design metric space.

In practice, the preceding problem is often solved through a scalarization process whereby the design metrics are aggregated to yield an AOF, subject to the constraints. For the purpose of the following analytical development, we employ the relevant and equivalent form of the preceding optimization problem.

Problem P2:

$$\text{minimize} \quad J(\mu_1, \ldots, \mu_m)$$

$$\text{subject to: } G(\mu_1, \ldots, \mu_m) \geq 0 \tag{2}$$

where on the Pareto frontier we have $G(\boldsymbol{\mu}) = 0$. Without material loss of generality in the foregoing analysis, we shall assume that $\mu_k > 0$, $1 \leq k \leq m$. Because we are interested only in Pareto optimal points, we assume that the AOF used is capturing only points that are Pareto optimal. A Pareto optimal point is said to be capturable if it is a solution of Problem P2 for some settings of the parameters used in the AOF $J(\mu_1, \ldots, \mu_m)$.

Optimality Conditions for Simpler Pareto Frontier

As stated earlier, in the first component of the analytical development, we assume a particular form of the functions [Eq. (2)] where each design metric is raised to a given exponent. We shall comment on the limitations of the resulting findings and on how this case leads to a more complex Pareto frontier form that has more significant practical implications. The simpler form of the Pareto frontier follows.

Problem P3:

$$\text{minimize} \quad J(\boldsymbol{\mu}) = w_1 \mu_1^{s_1} + \cdots + w_m \mu_m^{s_m} \tag{3a}$$

$$\text{subject to} \quad G(\boldsymbol{\mu}) = \alpha_1 \mu_1^{t_1} + \cdots + \alpha_m \mu_m^{t_m} - c \geq 0 \tag{3b}$$

We restrict our analysis to the nondegenerate cases where $w_k \neq 0$ and $s_k \neq 0$. Furthermore, we require that the constants w_k and s_k be positive to ensure that the points captured be Pareto optimal and also that c be positive. A detailed analysis of related issues is presented in Ref. 21, where AOFs that are guaranteed to yield Pareto points are said to be admissible and their characteristics are defined and developed. In the foregoing analytical development, we will discover the required conditions on both the exponents and the scalar coefficients in both the AOF and the Pareto frontier. Next, we address three issues in sequence: 1) optimality conditions, 2) convexity/concavity of Pareto frontier, and 3) examination of optimality conditions.

Optimality Conditions

At this point we proceed to establish the conditions for optimality by invoking the Karush–Kuhn–Tucker conditions. Specifically, we employ the second-order necessary conditions for optimality.[23] For a point $\boldsymbol{\mu}^*$ that is a minimum, the following conditions apply:

$$\nabla J(\boldsymbol{\mu}^*) + \lambda \nabla G(\boldsymbol{\mu}^*) = 0 \tag{4a}$$

$$G(\boldsymbol{\mu}^*) \geq 0 \tag{4b}$$

$$\lambda \leq 0 \tag{4c}$$

$$\lambda G(\boldsymbol{\mu}^*) = 0 \tag{4d}$$

together with

$$\boldsymbol{y}^T \nabla^2 [L(\boldsymbol{\mu}^*)] \boldsymbol{y} \geq 0 \tag{4e}$$

$$\nabla [G(\boldsymbol{\mu}^*)]^T \boldsymbol{y} = 0 \tag{4f}$$

where

$$L(\boldsymbol{\mu}) = J(\boldsymbol{\mu}) + \lambda G(\boldsymbol{\mu}) \tag{4g}$$

Substituting Eq. (3) into Eq. (4) yields the following conditions:

$$w_k s_k \mu_k^{s_k - 1} + \lambda \alpha_k t_k \mu_k^{t_k - 1} = 0, \qquad 1 \leq k \leq m \tag{5a}$$

$$\sum_{k=1}^{m} \alpha_k \mu_k^{t_k} - c \geq 0 \tag{5b}$$

$$\lambda \leq 0 \tag{5c}$$

$$\lambda G(\boldsymbol{\mu}^*) = 0 \tag{5d}$$

Before we proceed further, let us consider the possibility that the Lagrange multiplier vanishes. In this case, from Eq. (5a), we would have $\mu_k = 0$ for $1 \leq k \leq m$, which is a trivial case, which also requires $c = 0$ [see Eq. (5b)]. Therefore, we must have

$$w_k s_k \mu_k^{s_k - 1} + \lambda \alpha_k t_k \mu_k^{t_k - 1} = 0, \qquad 1 \leq k \leq m \tag{6a}$$

$$\sum_{k=1}^{m} \alpha_k \mu_k^{t_k} - c = 0 \tag{6b}$$

$$\lambda < 0 \tag{6c}$$

in addition to

$$\boldsymbol{y}^T \nabla^2 [L(\boldsymbol{\mu}^*)] \boldsymbol{y} \geq 0 \tag{6d}$$

$$\nabla [G(\boldsymbol{\mu}^*)]^T \boldsymbol{y} = 0 \tag{6e}$$

where

$$\nabla^2 [L(\boldsymbol{\mu}^*)] = H = \operatorname{diag}(h_{kk}), \qquad 1 \leq k \leq m \tag{6f}$$

$$h_{kk} = w_k s_k (s_k - 1) \mu_k^{s_k - 2} + \lambda \alpha_k t_k (t_k - 1) \mu_k^{t_k - 2} \tag{6g}$$

$$\nabla G(\boldsymbol{\mu})^T = \left\{ t_1 \alpha_1 \mu_1^{t_1 - 1}, t_2 \alpha_2 \mu_2^{t_2 - 1}, \ldots, t_m \alpha_m \mu_m^{t_m - 1} \right\} \tag{6h}$$

Convexity/Concavity of Pareto Frontier

Before we proceed to solve Eq. (6), we shall comment on the concavity/convexity of the Pareto frontier. Because this paper examines Pareto frontiers that are concave, at least in the neighborhood of the desired solution, we need to ensure that the Pareto frontier functions that we use are indeed concave. First, we make a special note regarding the convex case. Because the inequality constraint defined in Eq. (4b) is with a greater-than-or-equal-to sign, the Pareto frontier will be part of the boundary of a convex feasible space when its Hessian is negative semidefinite. Conversely, the Pareto frontier will be a part of a concave feasible space when its Hessian is positive definite. One way to see this is to think of $G(\boldsymbol{\mu})$ as the equation of a bowl and of $G(\boldsymbol{\mu}) = 0$ as the equation of the feasible space boundary. If the bowl faces down (negative semidefinite Hessian), we will have $G(\boldsymbol{\mu}) \geq 0$ in the feasible space (center of the bowl-convex space). Next, we note that the Hessian of the Lagrangian will always be positive semidefinite for the convex case. The preceding assertion is proved by examining Eq. (4g): 1) λ is negative, 2) the Hessian of $G(\boldsymbol{\mu})$ is negative semidefinite, and 3) the AOF is itself convex and its Hessian positive semidefinite. In this case, we actually have a convex optimization problem. Therefore, the convex Pareto frontier case is trivial and of no particular interest in this investigation. We shall restrict our attention to the case of concave Pareto frontiers, which is the chief subject of this paper.

We now establish that the feasible space defined by Eq. (5b) is indeed concave and thereby in conformity with the objectives of the present study. To do so, we take two generic extreme axes points that are on the boundary of the feasible design metric space and show that the points on the segment that interpolates these extreme points are infeasible. For the ith generic point P_i, we have $\mu_k = 0$ for $k \neq i$, yielding [Eq. (5b)]

$$\alpha_i \mu_i^{t_i} - c \geq 0 \Rightarrow \bar{\mu}_i = (c/\alpha_i)^{1/t_i} \tag{7a}$$

We have

$$P_i = \{0, \ldots, 0, \bar{\mu}_i, 0, \ldots, 0\}^T \tag{7b}$$

The interpolating segment points take the form

$$P_0 = \theta P_i + (1 - \theta) P_j, \qquad \{0 < \theta < 1\} \tag{7c}$$

We now show that the point P_0 violates the constraint of Eq. (4b) and is, therefore, not part of the feasible space, which must then be concave. The stated violation is shown as follows:

$$G(P_0) = \alpha_i (\theta \bar{\mu}_i)^{t_i} + \alpha_j [(\theta - 1)\bar{\mu}_j]^{t_j} - c \tag{7d}$$

$$G(P_0) = c\{\theta^{t_i} + (1 - \theta)^{t_j} - 1\} \tag{7e}$$

$$c\{\theta^{t_i} + (1 - \theta)^{t_j} - 1\} < c\{\theta + (1 - \theta) - 1\} \tag{7f}$$

$$c\{\theta + (1 - \theta) - 1\} = 0 \tag{7g}$$

Therefore we have

$$G(P_0) < 0 \tag{7h}$$

The inequality (7f) holds when $t_i > 1$, with $\{0 < \theta < 1\}$. Assuming that $t_i > 1$, we have just shown that the Pareto frontier defined by Eq. (5b) is part of a concave feasible space.

Examination of Optimality Conditions [Equation (6)]

We begin by using Eqs. (6a) and (6g) to obtain

$$\lambda = -(w_k s_k / \alpha_k t_k) \mu_k^{s_k - t_k} \tag{8a}$$

$$h_{kk} = w_k s_k (s_k - t_k) \mu_k^{s_k - 2} \tag{8b}$$

Since $w_k > 0$ and $s_k > 0$, for λ to be negative we require that α_k and t_k have the same sign and be nonzero.

We now turn our attention to Eq. (6e), which implies that the elements of the vector $\boldsymbol{y}$ cannot be chosen independently in Eq. (6d). We can employ Eqs. (6e) and (6h) to express this dependence as

$$\boldsymbol{y} = \begin{bmatrix} V^T \\ I^{m-1} \end{bmatrix} \begin{Bmatrix} y_2 \\ \vdots \\ y_m \end{Bmatrix} \equiv \tilde{V}\tilde{\boldsymbol{y}} \tag{9}$$

where

$$V^T = \left\{ -\frac{t_2 \alpha_2}{t_1 \alpha_1} \mu_2^{t_2 - t_1}, \ldots, -\frac{t_m \alpha_m}{t_1 \alpha_1} \mu_m^{t_m - t_1} \right\} \tag{10a}$$

$$\tilde{\boldsymbol{y}}^T = \{y_2, \ldots, y_m\} \tag{10b}$$

and I^{m-1} is an identity matrix of order $m - 1$. Note that, to obtain Eqs. (10a), we chose to eliminate $\boldsymbol{y}_1$, and in the process, divided by $t_1 \alpha_1 \mu_1^{t_1 - 1}$, which we know to be nonzero.

Using Eq. (9), we can express Eqs. (4e) and (6f) as

$$\boldsymbol{y}^T H \boldsymbol{y} = \tilde{\boldsymbol{y}}^T \tilde{V}^T H \tilde{V} \tilde{\boldsymbol{y}} \geq 0 \tag{11}$$

or

$$\tilde{\boldsymbol{y}}^T \tilde{H} \tilde{\boldsymbol{y}} \geq 0 \tag{12}$$

$$\tilde{H} = \tilde{V}^T H \tilde{V} \tag{13}$$

We now express Eq. (13) as [see Eq. (8b)]

$$\tilde{H} = H_a + H_b \tag{14}$$

where

$$H_a = V h_{11} V^T \tag{15}$$

$$H_b = \mathrm{diag}(h_{22}, \ldots, h_{mm}) \tag{16}$$

We note that Eq. (4e) states that $\tilde{H}$ shall be positive semidefinite. Inspection of Eq. (8b) reveals that if $s_k \geq t_k$, $2 \leq k \leq m$, then H_b is positive semidefinite. In addition, if $s_1 \geq t_1$, then H_a is also positive semidefinite. The latter is so because [Eq. (15)] a matrix formed by the product of a vector times its transpose is positive definite. From these observations, we conclude that, if $s_k \geq t_k$, $1 \leq k \leq m$, then $\tilde{H}$ is positive semidefinite. This is because the sum of two positive semidefinite matrices is positive semidefinite [Eq. (14)]. However, for the matrix $\tilde{H}$ to be positive semidefinite, we need not require $s_k \geq t_k$ for all $1 \leq k \leq m$. Let us explore this latter case, that is, the case where the relation $s_k \geq t_k$ is not satisfied for at least one value of k. We express Eq. (12) explicitly as

$$(\tilde{H})_{ij} = w_1 s_1 \mu_1^{s_1 - 2}(s_1 - t_1) \times \left\{ \frac{t_{i+1}\alpha_{i+1}\mu_{i+1}^{t_{i+1}-1}}{t_1 \alpha_1 \mu_1^{t_1 - 1}} \right\} \left\{ \frac{t_{j+1}\alpha_{j+1}\mu_{j+1}^{t_{j+1}-1}}{t_1 \alpha_1 \mu_1^{t_1 - 1}} \right\} + w_{i+1} s_{i+1}(s_{i+1} - t_{i+1}) \mu_{i+1}^{s_{i+1}-2} \delta_{ij} \tag{17}$$

where

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \tag{18}$$

The first important observation that we make is that the generic entries of the matrix $\tilde{H}$ depend on the particular values of the design metrics and are, therefore, problem dependent. Because these entries are not dependent solely on s_k and t_k, it is not possible to make direct conclusions regarding the positive definiteness of $\tilde{H}$ in the conventional sense. However, it will still be possible to make important findings.

Recall that our objective here is to determine whether it is possible for $\tilde{H}$ to be positive semidefinite, even when the relation $s_k \geq t_k$ is not satisfied. Let us assume that, for some value of i, for example, $i = k$, we have $s_{k+1} < t_{k+1}$. In this case the second term in Eq. (17) is negative. Because the matrix $\tilde{H}$ must be positive semidefinite for all values of the design metrics, let us examine the case where μ_{k+1} and μ_1 are significantly nonzero and the other design metrics are negligibly small. In such a case, $\tilde{H}$ degenerates into a zero matrix that has a single significantly nonzero diagonal element given by

$$(\tilde{H})_{kk} = w_1 s_1 \mu_1^{s_1 - 2}(s_1 - t_1) \left\{ \frac{t_{k+1}\alpha_{k+1}\mu_{k+1}^{t_{k+1}-1}}{t_1 \alpha_1 \mu_1^{t_1 - 1}} \right\}^2 + w_{k+1} s_{k+1}(s_{k+1} - t_{k+1}) \mu_{k+1}^{s_{k+1}-2} \tag{19}$$

We observe that the right-hand side of Eq. (19) is the only nonzero eigenvalue of the matrix $\tilde{H}$. We need to investigate if it can be made positive. We first note that the second term is negative. If the first term can overcome this negative component, then the eigenvalue can become positive. As can be seen, if s_1 is sufficiently greater than t_1, then the eigenvalue will be positive.

We are now able to articulate three important findings, where we recall that the positive semidefiniteness of $\tilde{H}$ is a necessary condition for optimality. First, if $s_k \geq t_k$ for all values of k, $\tilde{H}$ is positive semidefinite. Second, even if for a generic kth design metric $s_k < t_k$, $\tilde{H}$ could become positive semidefinite if, at least for one other generic rth design metric, s_r is sufficiently greater than t_r. Third, if $s_k < t_k$ for all k, then $\tilde{H}$ will not be positive semidefinite. These three statements have significant practical implications. In practice, because we have control over the values of s_i, we can simply make them sufficiently large to make $\tilde{H}$ positive definite.

The analysis thus far has allowed us to make a direct link between the power of a given design metric in the AOF and that of the same design metric in the Pareto frontier. However, the form assumed for the Pareto frontier is restrictive. Note that, even in this restrictive case, the findings are of practical value because for more general cases the conclusions will be accordingly more onerous. In particular, the findings thus far point to the weaknesses of the WS objective function, the WSS objective function, and of any other form of objective functions that limit the highest power of the AOF with respect to each design metric.

Optimality Conditions for Separable Pareto Frontier

Next, we examine a more general case that will provide guidance of a more general nature. In particular, we assume that the Pareto frontier is composed of separable polynomials in terms of the each design metric. That is, each design metric μ_i now appears as a polynomial of degree t_i, rather than as a single term with exponent t_i. As we proceed with this case, we will not repeat several equations that again hold.

Problem P4 takes the form

$$\text{minimize} \quad J(\boldsymbol{\mu}) = \sum_{i=1}^{m} w_i \mu_i^{s_i} \tag{20a}$$

$$\text{subject to} \quad G(\boldsymbol{\mu}) = \sum_{k=1}^{m} P_k(\mu_k) \geq 0 \tag{20b}$$

where

$$P_k(\mu_k) \equiv P_k^{t_k}(\mu_k) = \sum_{j=0}^{t_k} \alpha_{kj} \mu_k^j \tag{20c}$$

We note that $P_k^{t_k}$ is a polynomial of μ_k of order t_k. From the form of Eq. (20b), we note that there are no cross terms, and the polynomial is, therefore, separable in terms of the design metrics. We will see that this form of the Pareto frontier is sufficiently general to allow us to draw conclusions of general practical applicability in the process of engineering design optimization. We shall comment later on the case of more general Pareto frontier forms.

To solve P4, we note that Eqs. (4a–4g) again apply. Substituting Eqs. (20) into Eq. (4) leads to

$$w_k s_k \mu_k^{s_k - 1} + \lambda P_k'(\mu_k) = 0, \qquad 1 \leq k \leq m \tag{21a}$$

$$P_k'(\mu_k) \equiv \frac{\partial P_k}{\partial \mu_k} = \sum_{j=1}^{t_k} j \cdot \alpha_{kj} \mu_k^{j-1} \tag{21b}$$

$$G(\boldsymbol{\mu}) = \sum_{k=1}^{m} P_k(\mu_k) = 0 \tag{21c}$$

$$\lambda < 0 \tag{21d}$$

in addition to

$$\boldsymbol{y}^T \nabla^2 [L(\mu^*)] \boldsymbol{y} \geq 0 \tag{21e}$$

$$\nabla [G(\mu^*)]^T \boldsymbol{y} = 0 \tag{21f}$$

where

$$\nabla^2 [L(\boldsymbol{\mu}^*)] = H = \operatorname{diag}(h_{kk}), \qquad 1 \leq k \leq m \tag{21g}$$

$$h_{kk} = w_k s_k (s_k - 1) \mu_k^{s_k - 2} + \lambda P_k''(\mu_k) \tag{21i}$$

$$P_k''(\mu_k) \equiv \frac{\partial^2 P_k}{\partial \mu_k^2} = \sum_{j=1}^{t_k} j \cdot (j-1) \cdot \alpha_{kj} \mu_k^{j-2} \tag{21j}$$

$$\nabla G(\mu)^T = \{P_1', P_2', \ldots, P_m'\} \tag{21h}$$

Following the same reasoning as that of P3 yields the conclusion that λ is strictly negative. From Eq. (21a), we obtain

$$\lambda = -\frac{w_k s_k}{P_k'(\mu_k)} \mu_k^{s_k - 1} \tag{22}$$

For λ to be a finite strictly negative number, we need

$$P_k'(\mu_k) > 0 \tag{23}$$

Note the geometrical interpretation of Eq. (23). Because $P_k'(\mu_k)$, $1 \leq k \leq m$, are the components of the gradient vector $\nabla \boldsymbol{G}(\mu_k)$, Eq. (23) simply states that the gradient shall have all positive entries. Geometrically, this means that the gradient vector of the Pareto frontier will point north–east in the objective (design metric) space. This is exactly what we expected. If the perpendicular to the Pareto frontier were not to point north–east at a given point, then that point would be dominated by its neighbor. (Recall that $\mu_k > 0$.)

Substituting Eq. (22) into Eq. (21i) yields

$$h_{kk} = w_k s_k \mu_k^{s_k - 2} \{s_k - 1 - \mu_k [P_k''(\mu_k) / P_k'(\mu_k)]\} \tag{24}$$

Following a reasoning that is similar to the earlier case of a simpler form of the Pareto frontier function, we again restrict our attention to concave Pareto frontiers. The Hessian of the Pareto frontier function is given by

$$\nabla^2 G(\mu) \equiv H_G = \operatorname{diag}\{P_1''(\mu_1), \ldots, P_m''(\mu_m)\} \tag{25}$$

As already discussed, the Pareto frontier of Eq. (20b) defines a concave feasible space when H_G is positive definite, which occurs when

$$P_k''(\mu_k) > 0, \qquad 1 \leq k \leq m \tag{26}$$

Again, following a line of development similar to that of the earlier case, we write an equation analog to Eq. (17) as

$$(\tilde{H})_{ij} = (P_{i+1}'/P_1')(P_{j+1}'/P_1') h_{11} + h_{i+1,j+1} \delta_{ij} \qquad 1 \leq (i, j) \leq m - 1 \tag{27}$$

Determining detailed conditions for the positive semidefiniteness of $\tilde{H}$ would involve an unwieldy set of analyses and conclusions that would be of little, if any, practical value. Rather than doing so, we seek instead to uncover findings that can be used in practice. We begin by noting that $\tilde{H}$ will be positive semidefinite [recall Eq. (12)] when

$$h_{kk}(\mu_k) \geq 0, \qquad 1 \leq k \leq m \tag{28}$$

Using some elementary algebra, Eq. (28) yields the condition

$$P_k'(\mu_k)(s_k - 1) - \mu_k P_k''(\mu_k) \geq 0, \qquad 1 \leq k \leq m \tag{29}$$

We make the critical observation that, because $P_k'(\mu_k)$ and $P_k''(\mu_k)$ are positive, s_k must be sufficiently large if we wish to satisfy Eq. (28). Explicitly, in terms of the separable Pareto frontier polynomial function, Eq. (29) becomes

$$\sum_{i=1}^{t_k} \left\{ i \cdot \alpha_{ki} (s_k - i) \mu_k^{i-1} \right\} \geq 0, \qquad 1 \leq k \leq m \tag{30}$$

We observe that, when $s_k \geq t_k$, the $(s_k - i)$ factors are all positive, which helps. However, $s_k \geq t_k$ alone does not guarantee that Eq. (30) is satisfied. This is the case because the α_{ki} constants may be positive or negative. These issues will be explored in the following examples.

A useful way to contrast the two classes of Pareto frontiers studies in this section is to view the first as a generalized hyperellipsoid centered at the origin of the design metric space and to view the second as a more general case that is, nevertheless, separable. We use the term generalized because each design metric has a different exponent and the prefix hyper because we are in an m-dimensional space. The following two examples will amplify this discussion.

III. Examples and Practical Implications

In this section, we present two examples that illustrate the concepts presented in this paper. This is followed by a discussion of practical implications. In the first example, we use the simple form of the concave Pareto frontier, which involves three ellipses centered at the origin. In the second, we show the consequences of a more general concave Pareto frontier. Note that these two examples are purposely chosen to satisfy our objective of transparency in the demonstration of our findings and to convey maximum insight into the issues raised by this paper.

Example 1: Generalized Hyperellipsoid Frontier

In this example, the Pareto frontier is composed of three ellipses that are centered at the origin. The problem is posed as follows:

$$\text{minimize} \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \tag{31a}$$

$$\text{subject to} \quad \mu_1^2 + \left(\mu_2^2/9\right) \geq 1 \tag{31b}$$

$$\mu_1^4 + \mu_2^4 \geq 16 \tag{31c}$$

$$\left(\mu_1^3/27\right) + \mu_2^3 \geq 1 \tag{31d}$$

$$0 \leq \mu_1 \leq 2.9 \tag{31e}$$

$$0 \leq \mu_2 \leq 2.9 \tag{31f}$$

The feasible region in the objective space is shown in Fig. 1. We use the AOF on the form of Eq. (3a) as

$$J = \mu_1^s + b\mu_2^s \tag{32}$$

where b is a constant that can be used to yield different points of the Pareto frontier. Examining Fig. 1, we note that the arcs containing the points B, D, and C, respectively, denote the parts of the Pareto frontier represented by Eqs. (31b), (31c), and (31d). First, we note that the corner points do not satisfy the differentiability conditions and are, therefore, not constrained by the findings of this paper. Accordingly, point A can be obtained by simply letting s be equal to 1, the WS approach. For the arc containing B, because the Pareto frontier is of order two for both design metrics and is an ellipse of centered at the origin, all of the points of the subject arc can be obtained with $s \geq 2$. In fact, for $s = 2$, the AOF and the Pareto frontier arc overlap, and all of the Pareto points are simultaneously obtained. If we are interested in a particular point on the arc, a higher value of s and a particular value of the constant b are needed. Similarly, we need $s \geq 4$ and $s \geq 3$ for the arcs containing points D and C, respectively. This example showed that, in the best of circumstances, the exponent of the AOF should generally be higher than that of the Pareto frontier, for a given design metric. The next example shows that, in the more usual case where the Pareto frontier is not an ellipsoid centered at the origin, the requirement on the AOF becomes even more stringent, thereby causing some concern regarding our pervasive use of the WS and WSS AOF approaches, or even of others.

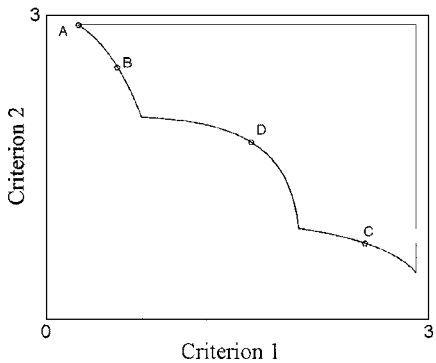

Fig. 1 Increasing the curvature of the AOF to capture Pareto optimal points.

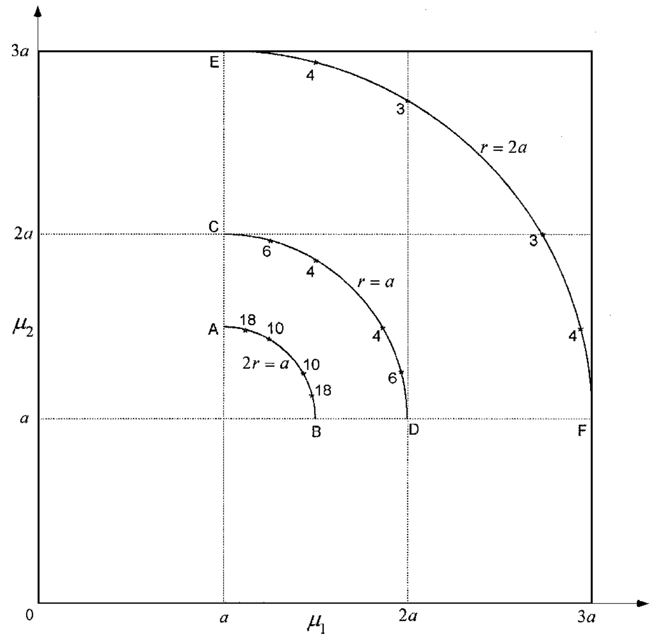

Fig. 2 Second-order Pareto frontier requires high-order AOF.

Example 2: Separable Pareto Frontier

In this example, different Pareto frontier cases are examined. They are circles of different radii, but which have a common center that is not at the origin. The common center is at the point $\mu_1 = \mu_2 = a$. The problem is posed as follows:

$$\text{minimize} \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \tag{33a}$$

$$\text{subject to} \quad (\mu_1 - a)^2 + (\mu_2 - a)^2 \geq r^2 \tag{33b}$$

$$a \leq \mu_1 \leq r + a \tag{33c}$$

$$a \leq \mu_2 \leq r + a \tag{33d}$$

Figure 2 shows three cases of Pareto frontiers denoted by three arcs of different radii r: arc A–B with $2r = a$, arc C–D with $r = a$, and arc E–F with $r = 2a$. We will discuss shortly the meaning of the numbers in Fig. 2. The AOF takes again the form

$$J = \mu_1^s + b\mu_2^s \tag{34}$$

We now proceed to examine this problem. Reconciling Eq. (33b) with the form of Eq. (20c), we have

$$\alpha_{k1} = -2a, \qquad \alpha_{k2} = 1, \qquad k = 1, 2 \tag{35}$$

Further, the condition of Eq. (30) takes the form

$$-2a(s-1) + 2(s-2)\mu_k \geq 0, \qquad k = 1, 2 \tag{36}$$

Let us consider the case of $s = 2$. From Eq. (36), we obtain the condition $-2a \geq 0$. Because we have $a > 0$, we make the interesting finding that when the Pareto frontier circle is not centered at the origin, a second-order AOF (or the WSS) is not sufficient to capture any optimal point at all. In other words, to capture any optimal point, we must use an AOF that is of order higher than two, even though the Pareto frontier is of order two.

Next, we further examine our ability to capture Pareto optimal points in terms of a and r. From Eqs. (36), (33c) and (33d), we obtain the condition

$$a[(s-1)/(s-2)] \leq \mu_k \leq a + r, \qquad k = 1, 2 \tag{37}$$

which yields

$$s \geq a/r + 2 \tag{38}$$

Equation (38) states that the higher the ratio a/r is the higher the required power to obtain a Pareto point and states that we need to satisfy Eq. (38) to capture any Pareto optimal point at all. Let us now uncover the required conditions to capture, not at least one, but all Pareto optimal points on the arc in question. To do so, we let

$$\mu_k = a + \beta_k r, \qquad 0 \leq \beta_k \leq 1, \qquad k = 1, 2 \tag{39}$$

Substituting Eq. (39) into (36) yields the condition

$$\beta_k \geq (a/r)[1/(s-2)], \qquad k = 1, 2 \tag{40}$$

Using Eq. (38), we are able to determine which region of the Pareto frontier we can capture for a given set of parameters. Figure 2 shows these regions. Let us consider the case of arc E–F, where we have $r = 2a$. When $s = 3$, we can capture all of the points that are between the two threes on the arc. For $s = 4$, we capture the larger region of the arc that is between the two fours. Similar statements can be made for the arcs C–D, and A–B. In the latter case, we observe that the requirements are quite stringent. We need a power of 18 in the AOF to capture the points in the subject region. In practice, we rarely use such AOFs, which implies that we would routinely fail to capture these designs even if they were the ones we seek. In other words, we would miss these points, and we might not even know we did.

We finally consider the conditions for capturing solutions that are near the points A or B. To do so, we let

$$\mu_k = a + \varepsilon, \qquad k = 1, 2 \tag{41}$$

where ε is an infinitesimally small number. Substituting Eq. (41) into Eq. (36) yields

$$s = a/\varepsilon + 2 \tag{42}$$

From Eq. (42), we note that, as $\varepsilon \to 0$, we have $s \to \infty$. This finding states that the power of the AOF must approach infinity as we seek the point approaching A. This means that, in practice, we would generally fail to capture such points on concave Pareto frontiers. We conclude our discussion of this example by noting that similar corresponding conclusions are obtained in the case of Pareto frontiers of order higher than two.

Practical Implications

It is important at this point to take a pragmatic perspective of the analytical developments and results obtained. In other words, how can these findings help us in practice, in as much as we usually do not have access to the analytical expression of the Pareto frontier? Interestingly, this lack of the analytical expression of the Pareto frontier is of little consequence. First, the well-known result that the WS AOF method fails in nonconvex Pareto frontier cases is simply a special case of this paper's analytical development. This paper goes further by explicitly addressing what happens in concave Pareto frontier cases for other than WS AOF. From this paper's findings, we know what can be done to ensure the capture of all potentially desirable solutions, that is, the complete Pareto frontier. We simply have to increase the order of the suspect design metrics, in the AOF. For example, assume that we change the weights in an AOF in the hope of obtaining a more desirable optimal solution, and in the process we observe a finite jump between the old and the new solution. This is generally due to the presence of a nonconvex region of the Pareto frontier. We note that this can occur with the WS AOF or any other continuous AOF. When this situation occurs, we can be sure to eliminate this jump in the optimum solution (that results from changing the weights) simply by increasing the powers of the corresponding design metrics. In other words, we can, in theory, obtain the complete Pareto frontier by having the proper form of the AOF, that is, one that offers the ability to increase the orders of the design metrics.

The next practical question that comes to mind regards the numerical consequence of increasing the order of the Pareto frontier. In practice, unfortunately, having too high an order for the AOF results in numerical conditioning deficiencies. The higher the order of the AOF, the more likely numerical conditioning difficulties will occur. An alternative is to employ the Tchebycheff (infinite) norm. Unfortunately, again, this method is not ideally suited to practical applications. For a method to be practical, while allowing the user to increase the AOF order (implicitly or explicitly), we must have a more complex form of the AOF than the WS or most of those popularly used in practice. The physical programming optimization approach[4,7–14] is one such method, where piecewise continuous functions composed of high-order splines form preference functions, each associated with one design metric. In the case of physical programming, the potentially deleterious numerical conditioning consequences are entirely circumvented.

IV. Conclusions

This paper investigated the capturability of optimal solutions that are on concave regions of Pareto frontiers (PF). This investigation went significantly beyond the common knowledge that the WS AOF approach fails to capture points on PF. We presented relationships that the AOF must satisfy for it to capture points of concave PF. We assumed two functional forms of the PF. The first takes the form of a hyperellipsoid that is centered at the origin. In this case, we presented relationships between the power of the AOF and the PF that must be satisfied for the AOF to be effective. In the second PF functional form, we let the PF be separable in terms of the design metrics. In this case, too, practical findings were developed.

We note that the usual objection/concern with the use of simplistic examples does not apply here. This paper does not present a method that it claims will work in complex cases, although it only presents simplistic examples. Rather, this paper examines difficulties associated with nonconvex PF. It is evident that, in this case, the difficulties discussed will only exacerbate in more complex cases, thereby reinforcing the message of the paper. This paper also offers a positive note, because it presents a way of overcoming the inability of objective functions to capture concave PF points.

One general finding is that it is desirable to make the AOF be of order higher than that of the PF, for each design metric. The specific details are discussed in the paper. Because we can have control over the power of the AOF, we can put the findings of this paper to good practical use. More important, this paper puts in explicit terms the limitations of the WS AOF, the WSS AOF, and all other AOFs that inherently limit the highest powers of the design metrics.

Acknowledgments

This research was partially supported by National Science Foundation CAREER Grant DMI-0196243. Assistance with example 1 by Cyriaque Puemi-Sukam is hereby acknowledged.

References

[1] Pareto, V., *Manuale di Economica Politica*, Societa Editrice Libraria, Milano, Italy, 1906; Schwier, A. S., *Manual of Political Economy*, Macmillan, New York, 1971 (English translation).

[2] 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.

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

[4] Messac, A., Sundararaj, G. J., Tappeta, R. V., and Renaud, J. E., "Ability of Objective Functions to Generate Non-Convex Pareto Frontiers," *AIAA Journal*, Vol. 38, No. 3, 2000, pp. 1084–1191.

[5] Messac, A., and Chen, W., "The Engineering Design Discipline: Is Its Confounding Lexicon Hindering Its Evolution?" *Journal of Engineering Evaluation and Cost Analysis*, Vol. 3, 2000, pp. 67–83.

[6] 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.

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

[8] Messac, A., and Hattis, P., "Physical Programming Design Optimization for High Speed Civil Transport (HSCT)," *Journal of Aircraft*, Vol. 33, No. 2, 1996, pp. 446–449.

[9] 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.

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

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

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

[13]Messac, A., Melachrinoudis, E., Sukam, C. P., "Mathematical and Pragmatic Perspectives of Physical Programming," *AIAA Journal*, Vol. 39, No. 5, 2001, pp. 885–893.

[14]Martinez, M., Messac, A., and Rais-Rohani, M., "Manufacturability-Based Optimization of Aircraft Structures Using Physical Programming," *AIAA Journal*, Vol. 39, No. 3, 2001, pp. 517–542.

[15]Steuer, R. E., *Multiple Criteria Optimization*: *Theory, Computation and Application*, Wiley, New York, 1986, pp. 440–450.

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

[17]Rao, J. R., and Papamlabros, P. Y., "A Non-Linear Programming Continuation Strategy for One Parameter Design Optimization Problems," *Proceedings of ASME Design Automation Conference*, American Society of Mechanical Engineers, Fairfield, NJ, 1989, pp. 77–89.

[18]Rakowska, J., Haftka, R. T., and Watson, L. T., "Tracing the Efficient Curve for Multi-Objective Control-Structure Optimization," *Computer Systems Engineering*, Vol. 2, 1991, pp. 461–471.

[19]Lin, J. G., "Three Methods for Determining Pareto-Optimal Solutions of Multiple-Objective Problems," *Directions in Large-Scale Systems*, edited by Y. C. Ho and S. K. Mitter, Plenum, New York, 1975, pp. 117, 118.

[20]Chen, W., Wiecek, M. M., and Zhang, J., "Quality Utility—A Compromise Programming Approach to Robust Design," American Society of Mechanical Engineers, ASME Paper DAC5601, Sept. 1998.

[21]Messac, A., Melachrinoudis, E., and Sukam, C. P., "Required Relationships Between Aggregate Objective Functions and Pareto Frontiers, with Practical Implications," *Optimization and Engineering Journal*, Vol. 1, No. 2, 2000, pp. 171–188.

[22]Fadel, G., and Li, Y., "Approximating the Pareto Curve to Help Solve Bi-Objective Design Problems," *Structural Optimization Journal* (to be published).

[23]Bazaraa, M. S., Sherali, H. D., and Shetty, C. M., *Nonlinear Programming and Algorithms*, 2nd ed., Wiley, New York, 1993, pp. 171, 172.

E. Livne
Associate Editor