Struct Multidisc Optim (2010) 41:853–862
DOI 10.1007/s00158-009-0460-7

RESEARCH PAPER

# The weighted sum method for multi-objective optimization: new insights

**R. Timothy Marler · Jasbir S. Arora**

Received: 17 April 2009 / Revised: 17 September 2009 / Accepted: 17 November 2009 / Published online: 12 December 2009


**Abstract** As a common concept in multi-objective optimization, minimizing a weighted sum constitutes an independent method as well as a component of other methods. Consequently, insight into characteristics of the weighted sum method has far reaching implications. However, despite the many published applications for this method and the literature addressing its pitfalls with respect to depicting the Pareto optimal set, there is little comprehensive discussion concerning the conceptual significance of the weights and techniques for maximizing the effectiveness of the method with respect to a priori articulation of preferences. Thus, in this paper, we investigate the fundamental significance of the weights in terms of preferences, the Pareto optimal set, and objective-function values. We determine the factors that dictate which solution point results from a particular set of weights. Fundamental deficiencies are identified in terms of a priori articulation of preferences, and guidelines are provided to help avoid blind use of the method.

**Keywords** Multi-objective optimization · Weighted sum · Preferences

R. T. Marler (✉)
Center for Computer Aided Design, College of Engineering;
111 ERF, The University of Iowa, Iowa City, IA, 52242, USA
e-mail: tmarler@engineering.uiowa.edu

J. S. Arora
Center for Computer Aided Design, College of Engineering;
238 ERF, The University of Iowa, Iowa City, IA, 52242, USA
e-mail: jsarora@engineering.uiowa.edu

## 1 Introduction

Despite deficiencies with respect to depicting the Pareto optimal set, the weighted sum method for multi-objective optimization (MOO) continues to be used extensively not only to provide multiple solution points by varying the weights consistently, but also to provide a single solution point that reflects preferences presumably incorporated in the selection of a single set of weights. However, the method's effectiveness in this latter capacity has not been studied thoroughly. To be sure, there are many different approaches for determining the weights (Marler and Arora 2004), but ultimately, however, these are all just different processes for organizing one's preferences and priorities. Instead of proposing yet another algorithm for translating preferences into weights, we focus on the mathematical characteristics of the solution and on the fundamental conceptual meaning of the weights. Furthermore, we determine the factors that dictate which Pareto optimal solution point results from a particular set of weights. Through this analysis, we augment the currently available literature and provide new insight into the nature of the method and new ideas for maximizing its effectiveness with respect to articulating preferences.

New understanding of the fundamentals presented in this paper, is useful in setting the weights directly (without the use of an additional algorithm). In addition, understanding inadequacies that are inherent in the method is instructive, and in this vein, we present deficiencies that cannot be avoided and that have not been discussed in the current literature.

Setting weights is just one approach to articulating preferences, and it is applicable to many different methods. Consequently, understanding how the weights affect the solution to the weighted sum method has implications concerning other approaches that involve similar method-parameters.

## 1.1 Overview of the formulation

The general multi-objective optimization (MOO) problem is posed as follows:

$$\begin{aligned}&\underset{\mathbf{x}}{\text{Minimize}}: \mathbf{F}(\mathbf{x}) = [F_1(\mathbf{x}), F_2(\mathbf{x}), \cdots, F_k(\mathbf{x})]^T \\ &\text{subject to}: \; g_j(\mathbf{x}) \le 0; \; j = 1, 2, \ldots, m\end{aligned} \tag{1}$$

where $k$ is the number of objective functions and $m$ is the number of inequality constraints. $\mathbf{x} \in E^n$ is a vector of design variables, and $\mathbf{F}(\mathbf{x}) \in E^k$ is a vector of objective functions $F_i(\mathbf{x}) : E^n \to E^1$. The *feasible design space* is defined as $\mathbf{X} = \{\mathbf{x} \,|\, g_j(\mathbf{x}) \le 0, \; j = 1, 2, ..., m\}$. The *feasible criterion space* is defined as $\mathbf{Z} = \{\mathbf{F}(\mathbf{x}) \,|\, \mathbf{x} \in \mathbf{X}\}$.

The idea of a solution for (1) can be unclear, because a single point that minimizes all objectives simultaneously usually does not exist. Consequently, the idea of *Pareto optimality* is used to describe solutions for MOO problems. A solution point is Pareto optimal if it is not possible to move from that point and improve at least one objective function without detriment to any other objective function. Alternatively, a point is *weakly Pareto optimal* if it is not possible to move from that point and improve all objective functions simultaneously.

Solving a particular MOO formulation can serve as a necessary and/or sufficient condition for Pareto optimality, and a method's performance in this capacity is important. However, these characterizations veer slightly from their meaning in terms of single-objective optimization. If a formulation provides a necessary condition, then for a point to be Pareto optimal, it must be a solution to that formulation. Consequently, the method can yield every Pareto optimal point as a solution. However, some solutions to a formulation that provides a necessary condition may not be Pareto optimal. Alternatively, if a formulation provides a sufficient condition, then its solution is always Pareto optimal, although such a formulation may not be able to capture all of the Pareto optimal points.

Typically, there are infinitely many Pareto optimal solutions for a multi-objective problem. Thus, it is often necessary to incorporate user preferences for various objectives in order to determine a single suitable solution. With methods that incorporate *a priori articulation of preferences*, the user indicates preferences before running the optimization algorithm and subsequently allows the algorithm to determine a single solution that presumably reflects such preferences. Alternatively, with *a posteriori articulation of preferences*, one manually selects a single solution from a representation of the Pareto optimal set. This paper focuses on the use of the weighted sum method for a priori articulation, while its use for a posteriori articulation has been addressed thoroughly in the literature.

Using the weighted sum method to solve the problem in (1) entails selecting scalar weights $w_i$ and minimizing the following composite objective function:

$$U = \sum_{i=1}^{k} w_i F_i(\mathbf{x}) \tag{2}$$

If all of the weights are positive, as assumed in this study, then minimizing (2) provides a sufficient condition for Pareto optimality, which means the minimum of (2) is always Pareto optimal (Zadeh 1963; Goicoechea et al. 1982). Although some literature indicates that $\sum_{i=1}^{k} w_i = 1$ and $\mathbf{w} \ge \mathbf{0}$, if any one of the weights is zero, there is a potential for the solution to be only weakly Pareto optimal.

With regards to the design space, minimizing the weighted sum provides a necessary condition if the multi-objective problem is convex, which means the feasible design space is convex (each constraint is convex) and all of the objective functions are convex (Geoffrion 1968; Koski 1985; Miettinen 1999). A function defined on a convex set is convex if and only if the Hessian matrix of the function is positive semidefinite or positive definite at all points in the set. Thus, if the Hessian matrix for each constraint and for each objective function is positive semidefinite or positive definite, then the weighted sum method can provide all Pareto optimal points (the complete Pareto optimal curve).

With regards to the criterion space, because the Pareto optimal set is always on the boundary of the feasible criterion space $\mathbf{Z}$, some studies suggest that $\mathbf{Z}$ must be convex for the weighted sum method to yield all of the Pareto optimal points (Stadler 1995; Athan and Papalambros 1996). However, Lin (1975) demonstrates that, in fact, if $\mathbf{Z}$ is only *p-directionally convex* (Holtzman and Halkin 1966) for any definitely negative vector $\mathbf{p}$, then the weighted sum approach provides a necessary condition. The term *definitely negative* implies $\mathbf{p} < \mathbf{0}$. p-directionally convex is defined as follows:

**Definition 1** *p-Directionally Convex*: Given a nonzero vector $\mathbf{p} \in E^k$, $\mathbf{Z} \subset E^k$ is said to be p-directionally convex if given any two different points in $\mathbf{Z}$, $\mathbf{F}_1$ and $\mathbf{F}_2$, and any two positive scalars, $w_1$ and $w_2$, with $w_1 + w_2 = 1$, there is a positive number $\beta$ such that $w_1\mathbf{F}_1 + w_2\mathbf{F}_2 + \beta\mathbf{p} \in \mathbf{Z}$.

This means that only the portion of the feasible criterion space that is visible when looking in the direction of $\mathbf{p}$ from within the feasible space needs to be convex. Therefore, only the Pareto optimal hypersurface needs to be convex for the weighted sum method to provide a necessary condition for Pareto optimality. Currently, the authors do not know of any method for predicting whether or not the Pareto optimal hypersurface is convex.

### 1.2 Review of the literature and motivation

Following the introduction of the weighted sum method by Zadeh (1963), the method has been mentioned prominently in the literature. However, most general treatments of MOO simply outline the weighted sum approach and indicate that it provides a Pareto optimal solution. Intricacies of the method and of the solutions that it yields are typically not discussed. In particular, the significance of the weights is not thoroughly explored, and thus, despite the presence of many algorithms for determining weight values, no fundamental guidelines have been presented for selecting weights for accurate a priori articulation of preferences.

The weighted sum method is often presented strictly as a tool, especially over the past few years, and literature regarding examples of applications is extensive. However, the focus is on the application, and the problems tend to be limited to those with just two objective functions. For instance, as development for a new approach, Koski and Silvennoinen (1987) provide an early application and use the weighted sum method to obtain multiple Pareto optimal solutions with a systematic change in the weights, while minimizing the volume and the nodal displacement of a four-bar space truss. Kassaimah et al. (1995) use the method for the two-objective optimization of laminated plates, where critical buckling shear stress is maximized and deflection is minimized. A few different somewhat arbitrary sets of weights are considered, and the corresponding solutions are compared, but the method itself is not studied and the articulation of preferences is not considered. Proos et al. (2001) apply the method to topology optimization, minimizing compliance and maximizing the first mode of the natural frequency, for two-dimensional plane stress problems. Again, the weights are altered to yield a representation of the Pareto optimal set. Saramago and Steffen (1998) use a weighted sum to combine two objective functions in an optimization-based approach to predicting robotic motion, but the weights have the same value and are thus irrelevant.

Marler and Arora (2005) study a three-objective problem, but they use the weighted sum method as a platform for studying various function-transformation methods and their affect on the depiction of the Pareto optimal set. It is found that a convex combination of functions is advantageous for depicting the Pareto optimal set, as apposed to determining a single solution. In addition, it is shown that some transformation methods can actually be detrimental to the process of obtaining a diverse and uniform spread of Pareto optimal points, and criteria are presented for determining when certain transformation methods (in conjunction with the weighted sum method) may not perform well.

Some literature provides a more substantial look at the method, focusing on its two primary deficiencies when representing the Pareto optimal set for a posteriori articulation of preferences. First, many authors demonstrate the method's inability to capture Pareto optimal points that lie on non-convex portions of the Pareto optimal curve (Koski 1985; Stadler and Dauer 1992; Stadler 1995; Chen et al. 1999; Athan and Papalambros 1996; Huang et al. 2007). Secondly, it is well known that the method does not provide an even distribution of points in the Pareto optimal set, with a consistent change in weights, where *even* implies a consistent Euclidean distance between consecutive solution points. Das and Dennis (1997) evaluate why this deficiency occurs, and for a two-objective problem, they derive the explicit form of the unique Pareto optimal curve for which an even distribution of weights yields an even distribution of Pareto solutions. In general, however, the points that result from a systematic selection of weights may cluster together. These two deficiencies are often answered with alternative methods for representing a Pareto optimal set (Das and Dennis 1998; Tappeta et al. 2000; Messac and Mattson 2002; Messac et al. 2003; Huang et al. 2007; Zhang et al. 2008). Such alternatives can be effective and valuable, but they are usually independent of the weighted sum method.

Although the difficulties mentioned above are well documented, they concern the method's use for yielding a complete Pareto optimal set rather than a single solution (a priori articulation of preferences). Criteria for maximizing the effectiveness of the method in the latter respect are not available. In fact, there are few, if any numerical studies of the weighted sum method that focus on a priori articulation of preferences. Essentially, there is no fundamental unifying analysis of the method.

Many researchers have, however, developed systematic approaches to selecting weights, surveys of which are provided by Eckenrode (1965), Hobbs (1980), Hwang and Yoon (1981), and Voogd (1983). We identify two broad classes of approaches, but with these approaches, weights may still be set ineffectively in terms of accurately articulating preferences. With *rating methods*, the decision-maker assigns independent values of relative importance to each objective function. With *ranking methods* (Yoon and Hwang 1995), which can be viewed as a subset of rating methods, the different objective functions are ordered by importance. Then, the least important objective receives a weight of one, and integer weights with consistent increments are assigned to objectives that are more important. The same approach is used with *categorization methods*, in which different objectives are grouped in broad categories such as highly important and moderately important. *Ratio questioning* or *paired comparison methods* are also common and provide systematic means to rate objective functions by comparing two objectives at a time. In this vein, Saaty (1977, 2003), and Saaty and Hu (1998) provide an eigenvalue method for determining weights, which involves $k(k-1)/2$ pair-wise comparisons between objective functions. This yields a comparison matrix, and the eigenvalues of the matrix are used as the weights. Rao and Roy (1989) discuss a method for determining weights based on fuzzy set theory. While disregarding preferences, Gennert and Yuille (1988) suggest that the weights should be selected to maximize the final minimum value of the weighted sum function. For the most part, algorithms for determining weights that reflect user preferences either fall into the category of rating methods or paired comparison methods.

Although the above-mentioned algorithms provide valid processes for determining the weights, there is little rigorous investigation as to what the weights actually represent. Thus, this paper studies the significance of the weights themselves. By considering the fundamental significance of the weights from different viewpoints, we uncover intrinsic flaws in the rating methods and paired comparison methods for a priori articulation of preferences. In addition, we identify which factors govern which solution is determined from a specific set of weights, and we explain why even with an effective algorithm that helps one decide on values for the weights, a weighted sum is not necessarily capable of incorporating preferences completely.

## 2 Interpretation of the weights

One often views the weights as general gauges of relative importance for each objective function. However, selecting a set of weights that reflects preference towards one objective or another can be difficult, because preferences tend to be indistinct. In addition, even with full knowledge of the objectives and satisfactory selection of weights, the final solution may not necessarily reflect the intended preferences that are supposedly incorporated in the weights. Nonetheless, the specific Pareto optimal point that is provided as the solution depends on which weights are used, so it is important to determine how the weights relate to preferences, to the Pareto optimal set, and to the individual objective functions. These relationships are explored in this section.

### 2.1 Preferences

First, the relationship between preferences and weights is examined. Concurrently, the theoretical foundation of rating methods is studied. We discover that such methods allow one to inadvertently set the weights according to the relative magnitude of the objective functions rather than the relative importance of the objectives. In this way, we demonstrate how function-transformation can be advantageous with a priori articulation of preferences. In addition, we reveal inherent deficiencies in the weighted sum method with respect to its capacity for incorporating preferences.

A *preference function* is an abstract function (of points in the criterion space) in the mind of the decision-maker, which perfectly incorporates his/her preferences. Most MOO methods that involve minimizing a single aggregated objective function, attempt to approximate the preference function with some mathematical representation, called a *utility function*. The gradients of the preference function $P[\mathbf{F}(\mathbf{x})]$ and the utility function in (2) are given respectively as follows:

$$\nabla_{\mathbf{x}} P\left[\mathbf{F}(\mathbf{x})\right] = \sum_{i=1}^{k} \frac{\partial P}{\partial F_i} \nabla_{\mathbf{x}} F_i\left(\mathbf{x}\right) \tag{3}$$

$$\nabla_{\mathbf{x}} U = \sum_{i=1}^{k} w_i \nabla_{\mathbf{x}} F_i\left(\mathbf{x}\right) \tag{4}$$

Each component of the gradient $\nabla_{\mathbf{x}} P$ qualitatively represents how the decision-maker's satisfaction changes with a change in the design point and a consequent change in function values. Comparing (3) and (4) suggests that *if the weights are selected properly, then the utility function can have a gradient that is parallel to the gradient of the preference function*. This is significant because the purpose of any utility function is to approximate the preference function, so imposing similarities between these two functions such as parallel gradients, can result in a more accurate representation of one's preferences. If the utility function should ideally be the same as the preference function, then certainly the gradients of the two functions should be the same as well.

The above relationships indicate that $w_i$ represents $\partial P/\partial F_i$. Conceptually, $\partial P/\partial F_i$ is the approximate change in the preference-function value (change in a decision-maker's satisfaction) that results from a change in the objective-function value for $F_i$. In this way, $\partial P/\partial F_i$ provides a mathematical definition or metric for the importance of $F_i$. However, it only makes sense to consider the importance of an objective or change in preference-function value, in relative terms. Thus, the value of a weight is significant relative to the values of other weights; the independent absolute magnitude of a weight is irrelevant in terms of preferences.

#### 2.1.1 Deficiencies in the method

Although (3) and (4) provide a general, conceptual idea of what the weights represent in terms of preferences, using them to set precise values for weights can be difficult. In fact, a fundamental deficiency in the weighted sum method is that *it can be difficult to discern between setting weights to compensate for differences in objective-function magnitudes* *and setting weights to indicate the relative importance of an objective as is done with the rating methods*. For practical purposes, $\partial P/\partial F_i$ can be approximated as $\Delta P/\Delta F_i$, and when objective functions have different ranges and orders of magnitude, an appropriate value for $\Delta F_i$ may not be apparent. For example, consider two objective functions to be minimized, where function-one is stress, which for the sake of argument ranges between 10,000 and 40,000 psi, and function-two is displacement, which ranges between 1 and 2 in. If displacement were to decrease by one unit, the consequence would be a significant improvement; $\Delta P/\Delta F_2$ would be relatively high. Concurrently, if stress were to decrease by a unit of one, the consequence would be negligible; $\Delta P/\Delta F_1$ would be insignificant. The result would be a relatively large value for $w_2$, compared to $w_1$. However, this result does not necessarily reflect the relative importance of the objectives; the values for the weights have been dictated purely by the relative magnitudes of the objective functions.

Alternatively, one could consider percentage changes in the objective functions rather than an absolute change of unity as discussed above. However, then the question arises as to what one should consider when evaluating percentages. Should $\Delta F_i$ be a percentage of the average value for $F_i$, the maximum value for $F_i$, the minimum value for $F_i$, or some factor of the range of values for $F_i$? The choice is arbitrary.

Consequently, one may set the weights to reflect objective-importance directly, but this approach can also lead to difficulties. With the previous example, suppose displacement is twice as important as stress. Considering the relative value of the weights, if $w_1 = 1$, then $w_2 = 2$. However, given the relative magnitudes of the objective functions, stress would still dominate the weighted sum, and the use of $w_2 = 2$ would be irrelevant. Again, the magnitudes of the objective functions affect how the weights are set, and the articulation of preferences becomes blurred at best.

This difficulty arises because what dictate the solution to the weighted sum are the relative magnitudes of the terms $w_i F_i$, in (2), not just the magnitudes of the functions or weights alone. *The value of a weight is significant not only relative to other weights but also relative to its corresponding objective function*. This is a critical idea, although it is often overlooked. With many weighted methods, including the weighted sum method, preferences are articulated a priori by using weights to induce the dominance of a particular term in a formulation (and the objective function associated with the term). However, the process of selecting weights and thus indicating preferences is complicated if this dominance is instead, induced by relative function-magnitude. Thus, *when using weights to represent the relative importance of the objectives, transforming the functions so that they all have similar magnitudes and do not naturally dominate the aggregate objective function can help one set the weights to reflect preferences more accurately*.

As an example demonstrating the significance of the relative values of the weights, consider the following MOO problem:

$$\begin{aligned} &\text{Find}: \ \mathbf{x} = [x_1, x_2]^T \\ &\underset{\mathbf{x}}{\text{to minimize}}: \ \begin{array}{l} F_1 = 20\,(x_1 - 0.75)^2 + (2x_2 - 2)^2 \\ F_2 = (x_1 - 2.5)^2 + (x_2 - 1.5)^2 \end{array} \end{aligned} \tag{5}$$

The design space for this problem is shown in Fig. 1, and the criterion space is shown in Fig. 2.

When $F_1$ is minimized independently, it has a minimum value of zero, and $F_2$ has a value of 3.3125. When $F_2$ is minimized independently, it has a minimum value of zero, and $F_1$ has a value of 62.25. $F_1$ has a range between zero and 62.25, and $F_2$ has a range between zero and 3.3125. Thus, the range of values for $F_1$ is much larger than that of $F_2$. If this problem is solved using the weighted sum method, and weights are selected with relative values to suggest equal importance of eth two objectives (i.e. $w_1 = w_2 = 1$), then the solution is $\mathbf{x} = (0.833, 1.100)$ and $F = (0.179, 2.938)$. Yet, even when considered to be equally important, $F_1$ has a value close to its absolute minimum, and $F_2$ has a value close to its maximum. If the weights are set to suggest that $F_1$ is twice as important as $F_1$ (i.e. $w_1 = 1$, $w_2 = 2$), then the solution is $\mathbf{x} = (0.909, 1.167)$ and $F = (0.179, 2.938)$. Even when one weight is twice as large as the other, the solution does not change a significant amount. However, when the objective functions are divided by their respective maxima and thus normalized, then it becomes easy to set the weights such that they are significant relative to each other and relative to the objective functions values. Then, with the objective functions normalized such that they both a have a range between zero and one, if the weights are again set

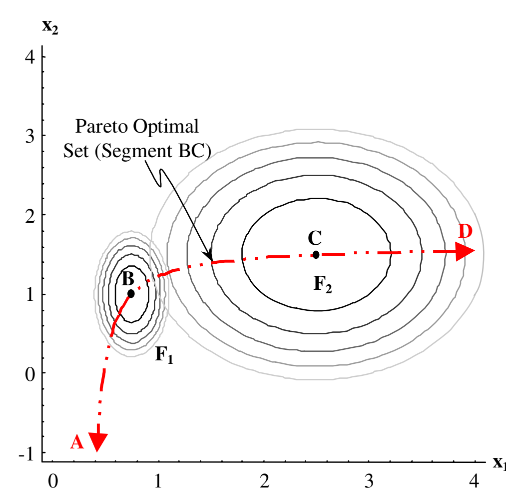


**Fig. 1** Example problem design space

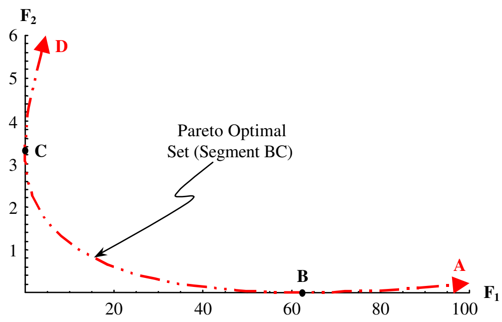


**Fig. 2** Example problem criteria space

such that $w_1 = w_2 = 1$, the solution is $\mathbf{x} = (1.598, 1.412)$ and $F = (15.054, 0.822)$.

Further study of (2), (3), and (4) reveals another fundamental deficiency in the weighted sum method: *the weighted sum utility function is only a linear approximation (in the criterion space) of the preference function*. Equation (2) is a linear function of the objective functions; it is the most basic approximation of one's preference function. The consequences of this characteristic are explained as follows. $\partial P/\partial F_i$ may actually be a nonlinear function of the objective functions, changing from point to point in the criterion space. However, using a weighted sum as a utility function to approximate the preference function, inherently assumes that this term is constant, represented by a scalar weight. Thus, values that are selected for the weights are only locally relevant. That is, they are not necessarily appropriate for the complete criterion space, only for the point at which they are determined, presumably the starting point (initial guess). As the design point changes with an iterative optimization process, the values of the objective functions change, and consequently, preferences (between various points in the criterion space) may change in a nonlinear fashion. For instance, in the context of the stress-displacement example described above, minimizing displacement may be more important than minimizing stress only until stress reaches a relatively high value, suggesting potential failure. Then, minimizing stress becomes more critical, and its associated weighting value should be increased. Anticipating such a change and altering weighting values while an optimization problem is running is impractical. Thus, decision-makers implicitly assume that $\partial P/\partial F_i$ is constant, when they determine a set of weights to reflect the relative importance of the objectives. This means that one assumes that the gradient of the preference function (in the criterion space) is constant, which is not necessarily the case. This is why even with a process that enables one to determine acceptable values for the weights, the final solution to the weighted sum problem may not accurately reflect initial preferences that were supposedly incorporated in the weights.

It is interesting to note that from a computational perspective, nonlinear composite objective functions often help to find Pareto optimal solutions on non-convex Pareto optimal sets more effectively (Messac et al. 2000a, b; Marler 2009). Such functions typically involve various types of parameters, not just weighting factors. Thus, they provide a more complex and versatile utility function for altering preferences.

## 2.2 Pareto optimal set

In this section and the next, factors that affect the solution to the weighted sum method are determined. In addition, fundamental characteristics of the weights are presented with regard to the Pareto optimal set, whereas the previous section concerned fundamental characteristics with respect to preferences. The characteristics discussed here provide a theoretical foundation for the paired comparison methods. We point out a fundamental deficiency with such methods for articulating preferences, and we discover that function transformations can actually be detrimental with these methods.

Figure 3 depicts a general illustrative model of a Pareto optimal set in the criterion space, for two objective functions.

The weights represent the gradient of $U$ in (2) with respect to the vector function $\mathbf{F}(\mathbf{x})$, shown as follows:

$$\nabla_{\mathbf{F}} U = \left\{ \begin{array}{c} \dfrac{\partial U}{\partial F_1} \\ \dfrac{\partial U}{\partial F_2} \end{array} \right\} = \left\{ \begin{array}{c} w_1 \\ w_2 \end{array} \right\} \tag{6}$$

The Pareto optimal solution (for a given set of weights) is found by determining where the $U$-contour with the lowest

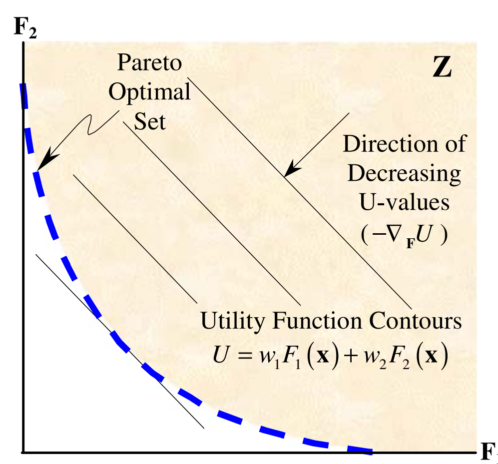


**Fig. 3** Pareto optimal set in the criterion space

possible value intersects the boundary of the feasible criterion space. The $U$-contours are linear and are thus tangent to the Pareto optimal curve at the solution point. Changes in the relative values of the weights result in changes in the orientation of the $U$-contours. *The subsequent change in the solution depends on the shape of the Pareto optimal curve in the criterion space*. This has significant repercussions when trying to provide an even spread of Pareto optimal points by solving a series of weighted sum problems, for a posteriori articulation of preferences. In fact, some more complicated methods have been developed so that the shape of the Pareto optimal hypersurface has a minimal affect the accuracy with which it is depicted (Marler and Arora 2004).

In terms of higher dimensions (more than two objective functions), the minimum value of (2) for a fixed set of weights determines a supporting hyperplane to $\mathbf{Z}$ (i.e. $U(\mathbf{F}) = \text{constant}$) (Zadeh 1963; Gembicki 1974), and the normal of the hyperplane is the vector of weights. The Pareto optimal hypersurface is always on the boundary of $\mathbf{Z}$, so the supporting hyperplane is tangent to the hypersurface at the point that minimizes the weighted sum.

Steuer (1989) uses such a hyperplane to provide an interpretation of the weights, which we suggest forms the basis for paired comparison methods. This interpretation is explained as follows. $w_1 = 1$ is considered a reference weight, and $F_1$ is the reference function (theoretically, any function can be used as a reference). Then, $\Delta_i$ represents the amount by which $F_i$ must increase in order to compensate for a decrease (improvement) of $\Delta_1$ in $F_1$, while remaining on the hyperplane. Each weight is then defined as

$$w_i = \lim_{\Delta_i \to 0} \frac{\Delta_1}{\Delta_i} \tag{7}$$

If one assumes $\Delta_1 = 1$, then each weight is approximated as

$$w_i \approx \frac{1}{\Delta_i} \tag{8}$$

Thus, $w_i$ represents the trade-off between $F_i$ and the reference function, at the solution point to the weighted sum problem. This is the idea behind paired comparison methods for setting weights.

To clarify this idea, note that one can draw the same conclusion using Fig. 3 for the case of two objective functions. Considering that the solution to the weighted sum problem is always Pareto optimal, the slope of the Pareto optimal curve in Fig. 3 is determined as follows:

$$\frac{dF_2}{dF_1} = -\frac{w_1}{w_2} \tag{9}$$

The left side of (9) can be approximated as $\Delta F_2/\Delta F_1$. Then, assuming $w_1 = 1$ is again a reference weight and $\Delta F_1 = 1$ (comparable to using $\Delta_1 = 1$ as above), (9) is reduced to

$$w_2 \approx -\frac{1}{\Delta F_2} \tag{10}$$

There is no negative sign in (8), because the numerator is assumed to represent a *decrease* of one.

When using the concepts described above, the objective functions should not be transformed. This is because using (7) involves comparing the different objective functions and evaluating trade-offs (how much of a loss in an objective is one willing to sacrifice for a gain in another objective). In contrast to the idea of evaluating relative importance, which was discussed with regard to the rating methods, assigning values to the weights based on trade-offs can be hampered by function transformations. Trade-off comparisons between functions are easier when the functions retain their original physical significance and original units. However, most transformation methods yield unitless functions, the values of which have little or no physical significance.

When using the stress-displacement example, one may be willing to accept an increase in displacement of 0.002 in. if stress were to reduce by 5,000 psi. If stress is the reference function, then a decrease (improvement) of 1 psi must correspond to a displacement increase of $5{,}000 \times 0.002 = 10$. Therefore, if the reference weight is $w_1 = 1$, then $w_2 = 10$. However, if the objectives are transformed or normalized and thus become unitless, it becomes difficult to conduct any type of trade-off analysis; it is difficult to conceptualize what level of increase in normalized displacement one might accept in exchange for a reduction in normalized stress of one.

Although (7) is derived based on the mathematical characteristics of (2), it provides another general method for articulating preferences, and here we point out an inherent deficiency in this method. Note that (7) is a relationship that necessarily exists at the solution point for the weighted sum. However, a user's decisions concerning trade-offs can change depending on the point at which the different objective functions are evaluated. One may be forced to determine trade-offs based on values for the different functions when the functions are evaluated at a point other than the solution point to which (7) applies. Therefore, as with (3) and (4), (7) only provides an approximate means for incorporating preferences in the weighted sum utility function a priori.

Figure 3 and (6) provide reinforcement of the conclusions drawn in Section 2.1 concerning the significance of the weights. Note that the direction of the gradient of $U$ in the criterion space, and thus the orientation of the U-contours and corresponding solution point, depends on the relative values (not the absolute values) of the weights. In addition, the orientation of the U-contours relative to the Pareto optimal set also depends on the range of objective-function values for points in the set. That is, for a given set of weights (a given gradient vector for the weighted sum), scaling an objective function changes the Pareto optimal solution point that the weighted sum method provides. Thus, as suggested earlier, the value of a weight is significant relative to the values of other weights *and* relative to the value of its corresponding objective function.

### 2.3 Objective functions

Whereas the discussion above considers the significance of weights in terms of the Pareto optimal set, this section focuses on the relationship between a set of weights and the objective function values. First, consider a problem with two unconstrained objective functions:

$$U = w_1 F_1 (\mathbf{x}) + w_2 F_2 (\mathbf{x}) \tag{11}$$

For $U$ to have a minimum, it is necessary that the gradient of $U$ be equal to zero, as follows:

$$\nabla_{\mathbf{x}} U = w_1 \nabla_{\mathbf{x}} F_1 + w_2 \nabla_{\mathbf{x}} F_2 = \mathbf{0} \tag{12}$$

Assuming the weights are positive and noting that minimizing a weighted sum always provides a Pareto optimal solution, (12) indicates that at a Pareto optimal point, the gradients of the two objective functions are co-linear and point in opposite directions. Essentially, the linear combination of the gradients equals zero. In compliance with the definition of Pareto optimality, this suggests that moving from a solution that satisfies (12), in order to improve a function, is detrimental to at least one other function.

General constrained problems are considered using Karush–Kuhn–Tucker necessary conditions, which are simplified as follows:

$$\sum_{i=1}^{k} w_i \nabla_{\mathbf{x}} F_i = -\sum_{i=1}^{m_a} \mu_i \nabla_{\mathbf{x}} g_i \tag{13}$$

where $m_a$ is the number of active constraints and $\mu_i$ is the Lagrange multiplier for gi($\mathbf{x}$). Equation (13) indicates that a linear combination of the objective function gradients is equal and opposite to a linear combination of the constraint gradients (for active constraints), at a Pareto optimal point. If a constraint is active at the solution point for a weighted sum, which necessarily provides a Pareto optimal point, then that constraint forms part of the Pareto optimal set. Equations (12) and (13) suggest that *the Pareto optimal point resulting from the use of a specific set of weights depends on active constraints that form part of the Pareto optimal set and on the relationship between the gradients of the different objective functions*.

## 3 Discussion and conclusions

This paper has provided insight into how the weighted sum method works and has explored the significance of the weights with respect to preferences, the Pareto optimal set, and the objective-function values. We have revealed fundamental deficiencies with respect to preferences, and we have shown that although the weighted sum method is easy to use, it provides only a linear approximation of the preference function. Thus, the solution may not preserve one's initial preferences no matter how the weights are set, a crucial idea that is often overlooked. The solution depends on multiple factors, one of which is the relative magnitude of the objective functions. However, when setting the weights, only the relative importance of the objectives should be considered, not the relative magnitudes of the function values. In this way, we argue that function-transformation can be helpful with a priori articulation of preferences, when the weights are used to represent the relative importance of the objectives (as with rating methods). This finding provides an important guideline that should always be considered when setting weights and/or transforming objective functions. Alternatively, we argue that if the weights are viewed as representing trade-offs between objective functions (as with paired comparison methods), then retaining the original units of the objectives, without transformation, is advantageous.

Much of the discussion in this paper pertains to a priori articulation of preferences, and although some literature suggests that the weights be set such that $\sum_{i=1}^{k} w_i = 1$ and $\mathbf{w} \geq \mathbf{0}$, which results in a convex combination of objective functions, when using the weighted sum for a priori articulation of preferences, there is no need to place any restriction on the values of the weights other than $\mathbf{w} \geq \mathbf{0}$, which ensures Pareto optimality. In fact, using unrestricted weights can be beneficial, making the determination of appropriate weight values easier. However, when systematically altering the weights for a posteriori articulation of preferences, using a convex combination of objective function can help avoid repeating weighting vectors in terms of relative values (Marler and Arora 2005). Thus, weights should be treated differently when used for a priori articulation than they are when used for a posteriori articulation.

Conclusions based on this work and guidelines for setting weights are presented as follows:

1) When setting weights directly for a priori articulation of preferences, the value of a weight must be significant relative to other weights and relative to its corresponding objective function.
2) When using a ranking method to set weights, all objective functions should be transformed such that they have similar ranges.

3) When using a paired comparison method to set weights or when viewing the weights as trade-offs, objective functions should not be transformed/normalized.
4) When articulating well-understood preferences with the paired comparison methods, unrestricted positive weights should be used.
5) The weighted sum provides only a basic approximation of one's preference function. It is fundamentally incapable of incorporating complex preference information. Thus, even if one determines acceptable values for the weights a priori, the final solution may not accurately reflect initial preferences.
6) The Pareto optimal solution that results from a specific set of weights depends on constraints that form part of the Pareto optimal set, the relationships between the gradients of the different objective functions, the relative magnitudes of the objective functions, and the shape of the Pareto optimal hypersurface.
7) The weighted sum method provides a basic and easy-to-use approach for multi-objective optimization and is useful as such, but alternate methods should be considered for the following cases:

    a. Non-convex or potentially non-convex problems where it is critical to depict the complete Pareto optimal set thoroughly
    b. When one wishes to depict the Pareto optimal set for even a convex problem thoroughly, but needs to minimize computational expenses (minimize the number of times (1) must be solved)
    c. When it is critical to articulate complex preferences accurately

## References


Athan TW, Papalambros PY (1996) A note on weighted criteria methods for compromise solutions in multi-objective optimization. Eng Optim 27:155–176

Chen W, Wiecek MM, Zhang J (1999) Quality utility—a compromise programming approach to robust design. J Mech Des 121:179–187

Das I, Dennis JE (1997) A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems. Struct Optim 14:63–69

Das I, Dennis JE (1998) Normal-boundary intersection: a new method for generating the Pareto surface in nonlinear multicriteria optimization problems. SIAM J Optim 8:631–657

Eckenrode RT (1965) Weighting multiple criteria. Manage Sci 12:180–192

Gembicki FW (1974) Performance and sensitivity optimization: a vector index approach. PhD dissertation, Case Western Reserve University, Cleveland, OH

Gennert MA, Yuille AL (1988) Determining the optimal weights in multiple objective function optimization. In: Second international conference on computer vision (held in Los Alamos, CA), Institute of Electrical and Electronics Engineers, Piscataway, NJ, pp 87–89

Geoffrion AM (1968) Proper efficiency and the theory of vector maximization. J Math Anal Appl 22:618–630

Goicoechea A, Hansen DR, Duckstein L (1982) Multiobjective decision analysis with engineering and business applications. Wiley, New York

Hobbs BF (1980) A comparison of weighting methods in power plant siting. Decis Sci 11:725–737

Holtzman JM, Halkin H (1966) Directional convexity and the maximum principle for discrete systems. SIAM J Control 4:263–275

Huang C-H, Galuski J, Bloebaum CL (2007) Multi-objective Pareto concurrent subspace optimization for multidisciplinary design. AIAA J 45:1894–1906

Hwang C-L, Yoon K (1981) Multiple attribute decision making, methods and applications: a state-of-the-art survey. In: Beckmann M, Kunzi HP (eds) Lecture notes in economics and mathematical systems, no 186. Springer, Berlin

Kassaimah SA, Mohamed AM, Kolkailah FA (1995) Bi-criteria optimum design of laminated plates under uniform load and shear. In: Proceedings of the 27th international SAMPLE technical conference (held in Albuquerque, NM), 27, pp 731–737

Koski J (1985) Defectiveness of weighting method in multicriterion optimization of structures. Commun Appl Numer Methods 1:333–337

Koski J, Silvennoinen R (1987) Norm methods and partial weighting in multicriterion optimization of structures. Int J Numer Methods Eng 24:1101–1121

Lin JG (1975) Three methods for determining Pareto-optimal solutions of multiple-objective problems. In: Ho YC, Mitter SK (eds) Directions in large-scale systems. Plenum, New York

Marler RT (2009) A study of multi-objective optimization methods for engineering applications. VDM, Saarbrucken

Marler RT, Arora JS (2004) Survey of multi-objective optimization methods for engineering. Struct Multidiscipl Optim 26:369–395

Marler RT, Arora JS (2005) Transformation methods for multi-objective optimization. Eng Optim 37:551–569

Messac A, Mattson CA (2002) Generating well-distributed sets of Pareto points for engineering design using physical programming. Eng Optim 3:431–450

Messac A, Sukam CP, Melachrinoudis E (2000a) Aggregate objective functions and Pareto frontiers: required relationships and practical implications. Optim Eng 1:171–188

Messac A, Sundararaj GJ, Tappeta RV, Renaud JE (2000b) Ability of objective functions to generate points on nonconvex Pareto frontiers. AIAA J 38(6):1084–1091

Messac A, Ismail-Yahaya A, Mattson CA (2003) The normalized normal constraint method for generating the Pareto frontier. Struct Multidiscipl Optim 25:86–98

Miettinen K (1999) Nonlinear multiobjective optimization. Kluwer Academic, Boston

Proos KA, Steven GP, Querin OM, Xie YM (2001) Multicriterion evolutionary structural optimization using the weighted and the global criterion methods. AIAA J 39:2006–2012

Rao JR, Roy N (1989) Fuzzy set theoretic approach of assigning weights to objectives in multicriteria decision making. Int J Syst Sci 20:1381–1386

Saaty TL (1977) A scaling method for priorities in hierarchies, multiple objectives and fuzzy sets. J Math Psychol 15:234–281

Saaty TL (2003) Decision-making with the AHP: why is the principal eigenvalue necessary. Eur J Oper Res 145:85–91

Saaty TL, Hu G (1998) Ranking by eigenvector versus other methods in the analytic hierarchy process. Appl Math Lett 11:121–125

Saramago SFP, Steffen V Jr (1998) Optimization of the trajectory planning of robot manipulators taking into account the dynamics of the system. Mech Mach Theory 33:883–894

Stadler W (1995) Caveats and boons of multicriteria optimization. Microcomput Civ Eng 10:291–299

Stadler W, Dauer JP (1992) Multicriteria optimization in engineering: a tutorial and survey. In: Kamat MP (ed) Structural optimization: status and promise. American Institute of Aeronautics and Astronautics, Washington, DC

Steuer RE (1989) Multiple criteria optimization: theory, computation, and application. Krieger, Malabar

Tappeta RV, Renaud JE, Messac E, Sundararaj GJ (2000) Interactive physical programming: tradeoff analysis and decision making in multicriteria optimization. AIAA J 38:917–926

Voogd H (1983) Multicriteria evaluation for urban and regional planning. Pion, London

Yoon KP, Hwang C-L (1995) Multiple attribute decision making, an introduction. Sage, London

Zadeh LA (1963) Optimality and non-scalar-valued performance criteria. IEEE Trans Automat Contr AC-8:59–60

Zhang K-S, Han Z-H, Li W-J, Song W-P (2008) Bilevel adaptive weighted sum method for multidisciplinary multi-objective optimization. AIAA J 46:2611–2622