

A Closer Look at Drawbacks of Minimizing Weighted Sums of Objectives for Pareto Set Generation in Multicriteria Optimization Problems*

Indraneel Das[†]

John Dennis[‡]

Abstract

A standard technique for generating the Pareto set in multicriteria optimization problems is to minimize (convex) weighted sums of the different objectives for various different settings of the weights. However, it is well-known that this method succeeds in getting points from all parts of the Pareto set only when the Pareto curve is convex. This article provides a geometrical argument as to why this is the case.

Secondly, it is a frequent observation that even for convex Pareto curves, an evenly distributed set of weights fails to produce an even distribution of points from all parts of the Pareto set. This article aims to identify the mechanism behind this observation. Roughly, the weight is related to the slope of the Pareto curve in the objective space in a way such that an even spread of Pareto points actually corresponds to often very uneven distributions of weights. Several examples are provided showing assumed shapes of Pareto curves and the distribution of weights corresponding to an even spread of points on those Pareto curves.

*This research was supported by the Dept. of Energy, DOE Grant DE-FG03-95ER25257

[†]Graduate Student, Dept. of Computational & Applied Mathematics, Rice University, Houston, TX 77251-1892. indra@rice.edu.

[‡]Noah Harding Professor Of Computational & Applied Mathematics, Rice University, Houston, TX 77251-1892. dennis@rice.edu

1 Introduction

Many problems in a wide variety of engineering disciplines are characterized by the need to minimize several nonlinear functions of the variables simultaneously. For example, a typical bridge-construction design might involve simultaneously minimizing the total mass of the structure and maximizing its stiffness. An airplane design problem might require maximizing fuel efficiency, payload, and minimizing the weight of the structure. Such multicriteria problems can be mathematically expressed as

$$\min_{x \in C} F(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_n(x) \end{bmatrix}, \quad n \geq 2 \quad \dots(MOP)$$

where

$$C = \{x : h(x) = 0, g(x) \leq 0, a \leq x \leq b\}$$

$F : \mathfrak{R}^N \mapsto \mathfrak{R}^n$, $h : \mathfrak{R}^N \mapsto \mathfrak{R}^{ne}$ and $g : \mathfrak{R}^N \mapsto \mathfrak{R}^{ni}$ are twice continuously differentiable mappings and $a \in (\mathfrak{R} \cup \{-\infty\})^N$, $b \in (\mathfrak{R} \cup \{\infty\})^N$, N being the number of variables, n the number of objectives, ne and ni the number of equality and inequality constraints.

Since no single x^* would in general minimize every f_i simultaneously, a concept of optimality which is useful in the multiobjective framework is that of Pareto optimality. To acquaint readers not familiar with the concept, it is defined below:

Definition: A point $x^* \in C$ is said to be (globally) Pareto optimal or a (globally) efficient point or a non-dominated or a non-inferior point for (MOP) if and only if there does not exist $x \in C$ such that $F(x) \leq F(x^*)$ with at least one strict inequality (the \leq implies term-by-term inequality).

A very popular approach for converting this multicriteria problem into a scalar optimization problem is to minimize a convex combination of the different objectives (see, for example, Koski [1], Jahn, et al [2]). In other words, n weights α_i are chosen such that $w_i \geq 0$, $i = 1, \dots, n$ and $\sum_{i=1}^n w_i = 1$ and the following problem is solved:

$$\begin{aligned} \min_x \sum_{i=1}^n w_i f_i(x) &= w^T F(x) \\ \text{s.t.} \quad x &\in C \quad \dots(LC) \end{aligned}$$

It follows immediately that the global minimizer x^* of the above problem is a Pareto optimal point for (MOP), since if not, then there must exist a feasible x which improves on at least one of the (positively weighted) objectives without increasing the others and hence produces a smaller value of the weighted sum¹.

A common approach then is to perform the above minimization for an even spread of α in order to generate several points in the Pareto set (which for a two objective problem produces points on the *Pareto curve* or *tradeoff curve*). The two major difficulties with this idea are as follows:

¹a unicity assumption on the global minimizer may be required if some of the components of w are zero

- If the Pareto curve is not convex, there does not exist any w for which the solution to problem (LC) lies in the nonconvex part.
- Even if the Pareto curve is convex, an even spread of weights w does not produce an even spread of points on the Pareto curve.

The following sections attempt to explain geometrically why these happen.

2 Failure in Capturing Nonconvex Parts of the Pareto Curve

2.1 An Equivalent Problem

In order to simplify the analysis it shall henceforth be assumed that there are only two objectives $f_1(x)$ and $f_2(x)$ (i.e. $n = 2$). Then if we let the weights on the two objectives be represented by α and $1 - \alpha$ where α is a scalar $\in [0, 1]$, problem (LC) becomes

$$\begin{aligned} \min_x \quad & (1 - \alpha)f_1(x) + \alpha f_2(x) \\ \text{s.t.} \quad & x \in C \quad \dots(LC_2). \end{aligned}$$

An equivalent formulation of the above is

$$\begin{aligned} \min \quad & \frac{\cos \theta}{\cos \theta + \sin \theta} f_1(x) + \frac{\sin \theta}{\cos \theta + \sin \theta} f_2(x) \\ \text{s.t.} \quad & x \in C \quad \dots(TLC_\theta) \end{aligned}$$

where the scalar θ varies from 0 to $\frac{\pi}{2}$. We shall call this problem the Trigonometric Linear Combinations problem (TLC_θ) .

For a given value of θ , problem (LC_2) with

$$\alpha = \frac{\sin \theta}{\cos \theta + \sin \theta}$$

yields exactly the same solution as problem (TLC_θ) . Thus (LC_2) and (TLC_θ) are equivalent in the sense that a Pareto point is a solution of problem (LC_2) for some $\alpha \in [0, 1]$ if and only if it is also the solution of problem (TLC_θ) for some $\theta \in [0, \frac{\pi}{2}]$. It follows that *if a Pareto point is not a solution to problem (TLC_θ) for all $\theta \in [0, \frac{\pi}{2}]$, then it cannot be obtained by minimizing any convex combination of the two objectives.*

2.2 Geometrical Interpretation of the Trigonometric Linear Combinations Problem

Consider rotating the f_1 - f_2 axes (anticlockwise) in the objective space by an angle $\theta \in [0, \frac{\pi}{2}]$ (as in fig. 1). Let us label the rotated axes as \bar{f}_1, \bar{f}_2 . Then, an elementary coordinate transformation gives

$$\begin{bmatrix} \bar{f}_1 \\ \bar{f}_2 \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$$

so that

$$\bar{f}_1 = f_1 \cos \theta + f_2 \sin \theta$$

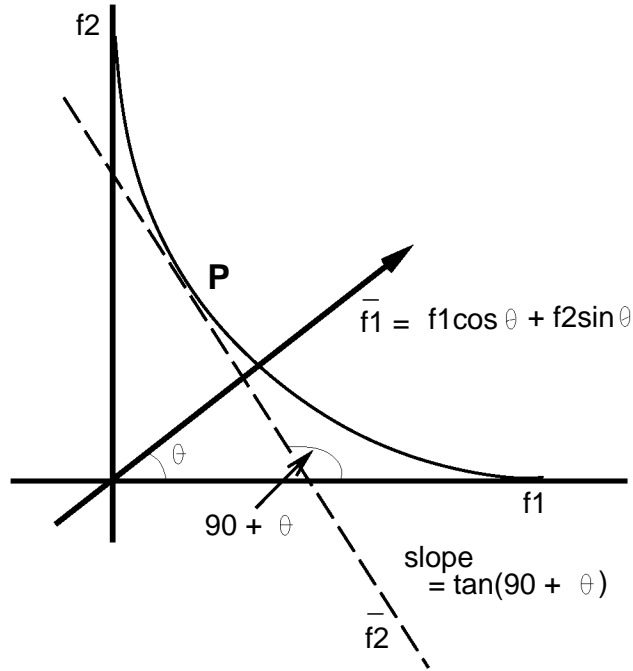


Figure 1: Getting a Pareto point by solving the Trigonometric Linear Combinations problem

Thus $\frac{\bar{f}_1(x)}{\cos(\theta)+\sin(\theta)}$ is what problem (TLC_θ) aims to minimize over $x \in C$, which has the same solution as $\min_{x \in C} \bar{f}_1(x)$. Geometrically, it can be seen in fig. 1 how minimizing \bar{f}_1 gets us the Pareto point P (i.e., we need a feasible point corresponding to which the value of the \bar{f}_1 coordinate is minimum). One can imagine this minimization process as translating the \bar{f}_2 axis parallel to itself until it hits the curve. The point where it hits the curve is a Pareto point.

Solving problem (TLC_θ) for all $\theta \in [0, \frac{\pi}{2}]$ is equivalent to repeating the above process for all axis rotations $\in [0, \frac{\pi}{2}]$. This can be thought of as varying the slope of the tangent from 0 to $-\infty$ while maintaining contact with the Pareto boundary and picking out the points of contact. This is illustrated in fig. 2 for a convex Pareto curve and shows how this process can yield all the Pareto points.

An alternative characterization:

A given Pareto point is a solution to problem (TLC_θ) for some θ if and only if the tangent to the Pareto curve at that point does not intersect the boundary of the set of attained vectors at any point where it is not a tangent to the boundary curve. This characterization relies on the boundary of the set of attained vectors being continuous and differentiable (i.e. $\frac{\partial f_2}{\partial f_1}$ exists at every point on the boundary of the set of attained vectors), which will be assumed throughout the remainder of this treatise.

Justification:

Given the slope of the Pareto curve at the point in question, θ and hence the (TLC_θ)

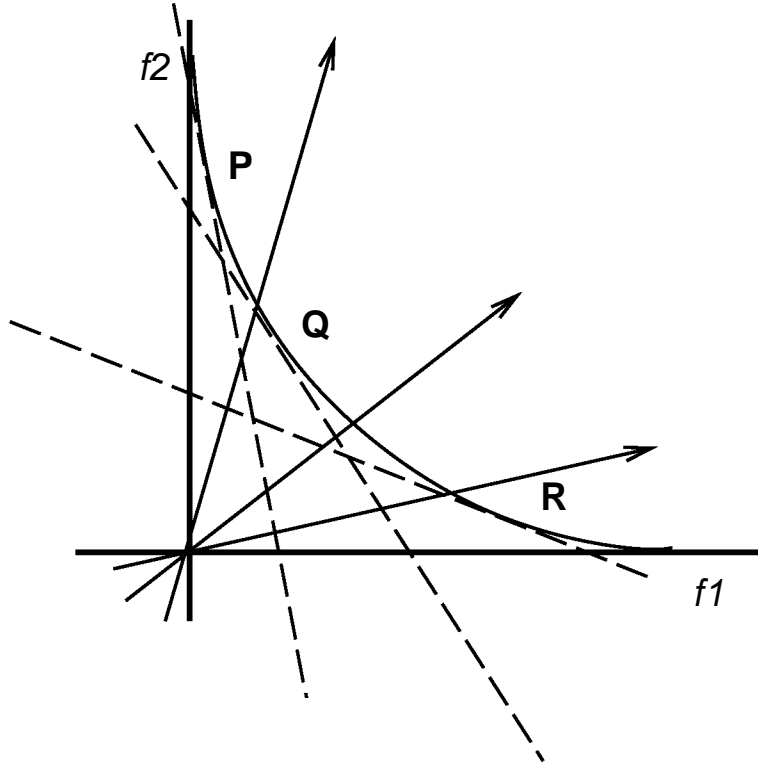


Figure 2: Getting a convex Pareto curve by solving Trigonometric Linear Combinations problems

problem gets uniquely defined (since $\theta = \tan^{-1}(\text{slope}) - \frac{\pi}{2}$, as shown in fig. 1). However if the tangent at this point intersects another point on the Pareto curve where the slope is different from the slope of the tangent, then the continuity and differentiability of the boundary curve imply that it is possible to slide the tangent down further (perpendicular to itself) to get a lower objective function value in problem (TLC_θ) , in which case the point in question is not a (global) minimum of problem (TLC_θ) . This is illustrated for point P in fig. 3. This point cannot solve (TLC_θ) for any other θ since, given the slope at the point, θ has been uniquely defined. Hence the justification follows.

2.3 Inexistence of a (TLC_θ) subproblem for points in the nonconvex part

It can be argued in two ways that given a point in the nonconvex part of the Pareto set, there does not exist a (TLC_θ) subproblem which it solves.

The first uses the aid of fig. 4. Let us recall that solving (TLC_θ) for all $\theta \in [0, \frac{\pi}{2}]$ is equivalent to varying the slope of the tangent from 0 to $-\infty$ while maintaining contact with the Pareto boundary. In fig. 4 it is shown that if the slope of the tangent matches that of the segment AB, it touches the Pareto curve at two distinct points (this is one characterization of a smooth, nonconvex Pareto curve, i.e. there exists a slope for which the tangent touches at least two distinct points on the curve). If the slope of the tangent is greater (i.e. less negative) than the slope of the segment AB, then it touches points in the MQ arc of the Pareto curve. If the slope is any less then it touches points in the PR arc

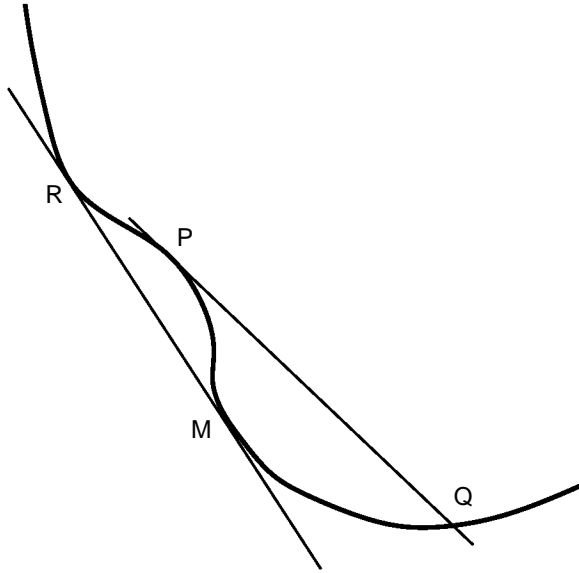


Figure 3: Pareto point P cannot be a solution of any trigonometric linear combinations problem, since the tangent at P intersects the boundary at Q but does not match the slope at Q, so the tangent can slide further down and decrease the (TLC_θ) objective. However points R or M can be obtained as a solution of a (TLC_θ) problem.

of the Pareto curve. Thus there does not exist a value of the slope for which the tangent touches the Pareto points in the RM arc of the Pareto curve, thus missing the nonconvex part.

The second argument follows from the alternative characterization given earlier. Given that the boundary of the set of attained vectors is continuous and differentiable, a tangent to any point in the nonconvex part of the Pareto set must intersect the boundary at at least one point where it does not match the slope of the boundary curve. Thus it cannot solve a (TLC_θ) subproblem for any $\theta \in [0, \frac{\pi}{2}]$.

Finally, given the equivalence between problem (TLC_θ) and (LC_2) , it can be concluded that Pareto points in the nonconvex parts of the Pareto set cannot be obtained by minimizing a convex combination of the objectives.

3 Nonuniform Spread of Pareto Points using Uniform Spread of α

It was shown in the previous section that problem (LC_2) fails to find Pareto points in the nonconvex part of the Pareto set. This section argues that even if the Pareto set is convex, a uniform spread of α does not guarantee a uniform spread of Pareto points on the Pareto curve. In many cases it has in fact been observed that the points obtained using a uniformly spread set of values of α are actually clumped in certain regions of the Pareto set, providing the user no information about the nature of tradeoff between the two objectives elsewhere. J. G. Lin in [5] observes that “Pareto-optimal solutions obtained by this method are often found to be so few, or the corresponding indexes so extreme, that there seems to be no

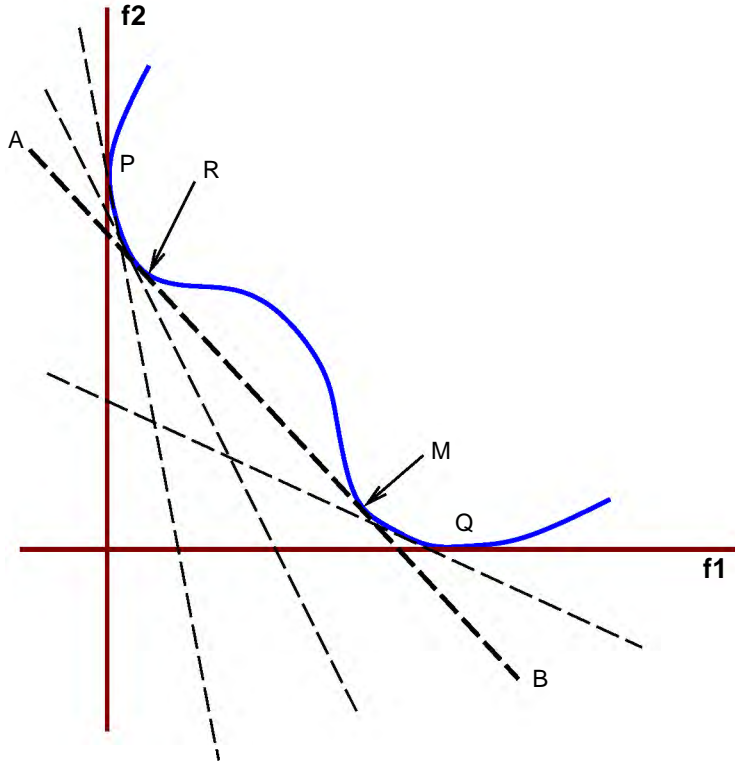


Figure 4: Failure of Trigonometric Linear Combinations in capturing nonconvex parts of the Pareto set

middle ‘ground’ for any compromise, although such ‘ground’ may actually exist”.

As an example consider the biobjective problem given below:

$$\begin{aligned} \min_x \quad & \left[\begin{array}{l} f_1(x) = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 \\ f_2(x) = 3x_1 + 2x_2 - \frac{x_3}{3} + 0.01(x_4 - x_5)^3 \end{array} \right] \\ \text{s.t.} \quad & x_1 + 2x_2 - x_3 - 0.5x_4 + x_5 = 2 \\ & 4x_1 - 2x_2 + 0.8x_3 + 0.6x_4 + 0.5x_5^2 = 0 \\ & x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 \leq 10 \end{aligned}$$

The set of Pareto points obtained by minimizing convex combinations of the two objectives for 21 different weights with α varying from 0 to 1 in even increments of 0.05 are shown in fig. 5, followed by the corresponding set of Pareto points produced by Normal-Boundary Intersection (NBI) using an even spread of NBI parameters (see Das and Dennis [6] for details).

3.1 Relationship between α and the slope of the Pareto curve

An explicit relationship between the slope of the Pareto curve at a Pareto point and the α for which this point solves (LC_2) can be derived based on the development in earlier

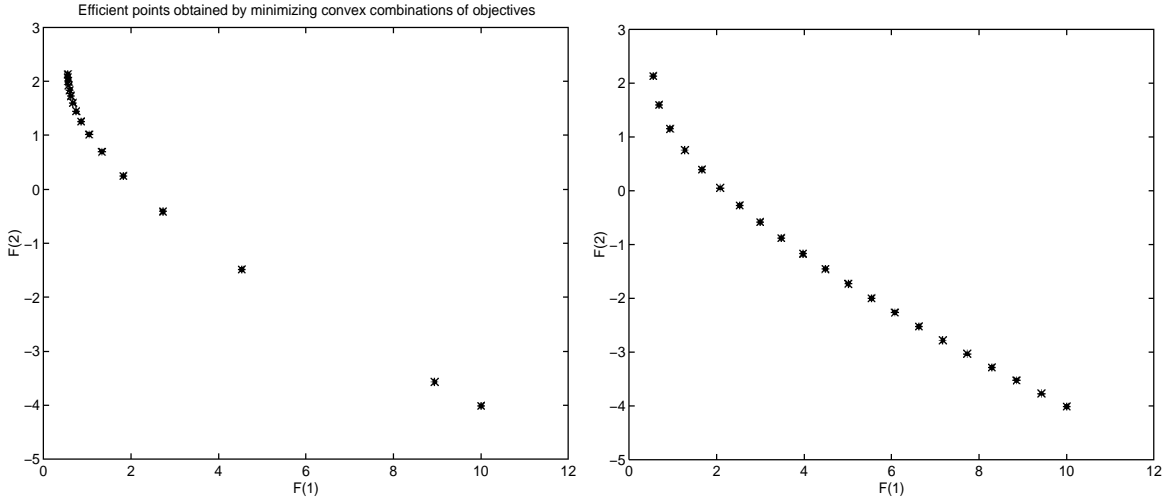


Figure 5: Outputs of minimizing linear combinations and NBI respectively on a bi-objective optimization problem using an even spread of parameters

sections. As shown in fig. 1, the slope of the curve at the Pareto point, denoted by $\frac{\partial f_2}{\partial f_1}$, is related to the parameter θ in the (TLC_θ) subproblem by

$$\frac{\partial f_2}{\partial f_1} = \tan\left(\frac{\pi}{2} + \theta\right)$$

i.e.,

$$\frac{\partial f_2}{\partial f_1} = -\cot \theta$$

Then, using the relationship between α and θ ,

$$\begin{aligned} \alpha &= \frac{\sin \theta}{\sin \theta + \cos \theta} \\ &= \frac{1}{1 + \cot \theta} \end{aligned}$$

so,

$$\alpha = \frac{1}{1 - \frac{\partial f_2}{\partial f_1}} \quad (1)$$

Thus an even spread of α would correspond to an even spread of points on the Pareto curve only if the shape of the Pareto curve is such that the quantity $\frac{1}{1 - \frac{\partial f_2}{\partial f_1}}$ is evenly spread for an even spread of Pareto points.

In order to make this notion concrete let us suppose that the functional form of the Pareto curve in the objective space is given by the mapping $\psi : f_1 \mapsto f_2$ (this is denoted simply by $f_2(f_1)$ in fig. 6). If the minimum and maximum of f_1 values on the Pareto set are \bar{f}_1^* and \bar{f}_1 respectively, then the same for f_2 are $\psi(\bar{f}_1)$ and $\psi(\bar{f}_1^*)$. As shown in fig. 6, we define a discrete set of Pareto points as being uniformly spread if the projections of the arcs between two consecutive Pareto points on the f_1 axis are all equal. This definition may not correspond to the most uniform spread possible (which would be to say that the distances

between two consecutive Pareto points are all identical), but provides an adequate sense of ‘uniformity’ in most practical cases and enables us to make our point without distracting the reader by meticulous details.

Let us assume that the Pareto points are uniformly spread as defined above corresponding to an even spread of α . Then using the fact that the projections of the arcs between two consecutive Pareto points on the f_1 axis are all equal, α can be written as a function of f_1 (at the Pareto points) as

$$\alpha = \frac{(f_1 - f_1^*)}{(\bar{f}_1 - f_1^*)}$$

If this above relationship is extended for all $f_1 \in [f_1^*, \bar{f}_1]$, then substitution of (1) in the above yields

$$\frac{1}{1 - \frac{\partial f_2}{\partial f_1}} = \frac{f_1 - f_1^*}{\bar{f}_1 - f_1^*}$$

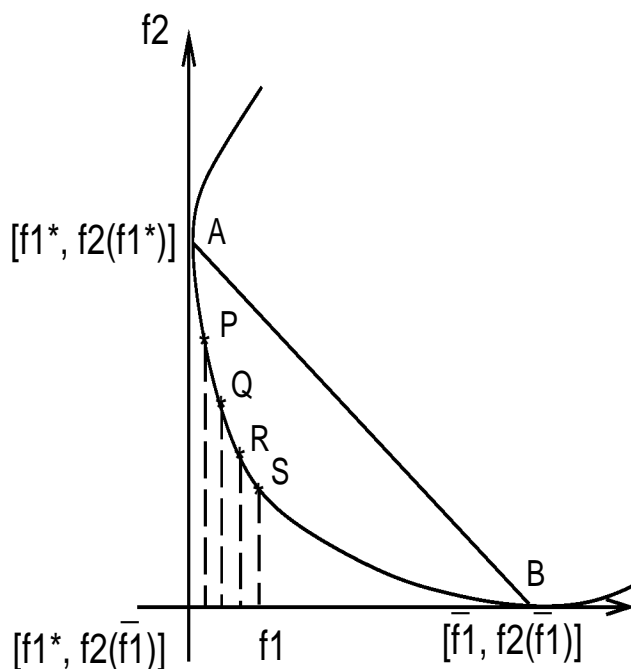


Figure 6: Definition of uniform spread: Projections of the arcs between consecutive Pareto points on the f_1 axis are constant

Rewriting $\frac{\partial f_2}{\partial f_1}$ as $\psi'(f_1)$ and rearranging the above equation results in the following differential equation:

$$\psi'(f_1) = 1 - \frac{\bar{f}_1 - f_1^*}{f_1 - f_1^*}$$

which has the general solution

$$\psi(f_1) = f_1 - (\bar{f}_1 - f_1^*) \log(f_1 - f_1^*) + K(\text{constant})$$

The above illustrates that only for very specific shapes of Pareto curves is it possible to get an even spread of Pareto points using an even spread of α .

3.2 Distribution of α for uniform spreads of Pareto points

Given that an even spread of Pareto points would correspond to an even spread of α for only specific Pareto curves, we shall try to find out the distribution of α values corresponding to an even spread of Pareto points given some typical Pareto curves. The spread of Pareto points selected here are even in the sense of NBI, which is a little less naive than the characterization of even spread in the previous section. Instead of the projections of arcs between two consecutive Pareto points on the f_1 axis being constant, the NBI even spread requires that those projections on the segment joining the extreme points of the Pareto set be constant (segment AB in fig. 6). Histogram plots of the α values corresponding to an even spread of points on the Pareto curve are shown for some assumed shapes of the Pareto curve in figs. 7, 8, 9 and 10.

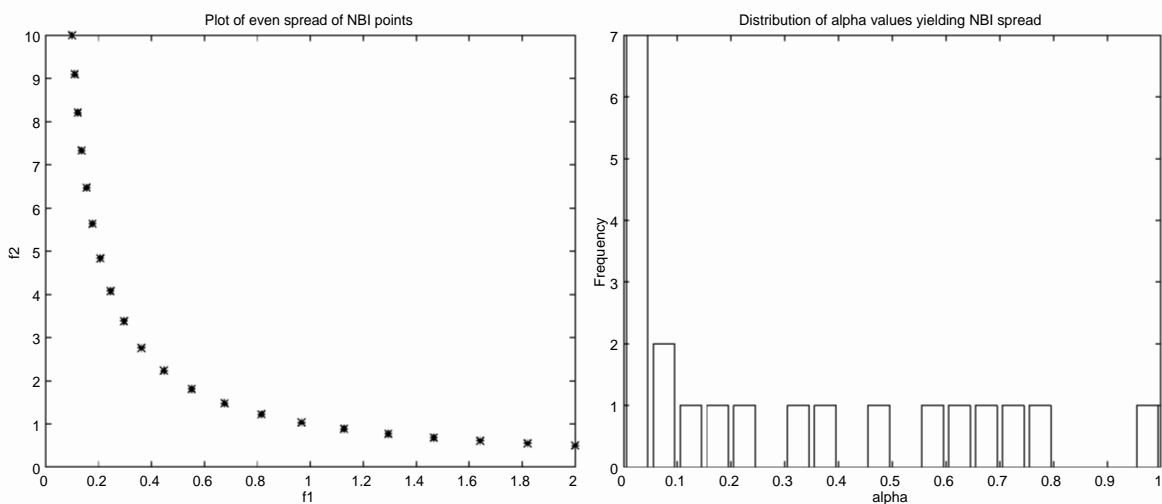


Figure 7: Distribution of α for even spread on $\psi(f_1) = \frac{1}{f_1}$ for $f_1 \in [0.1, 2]$

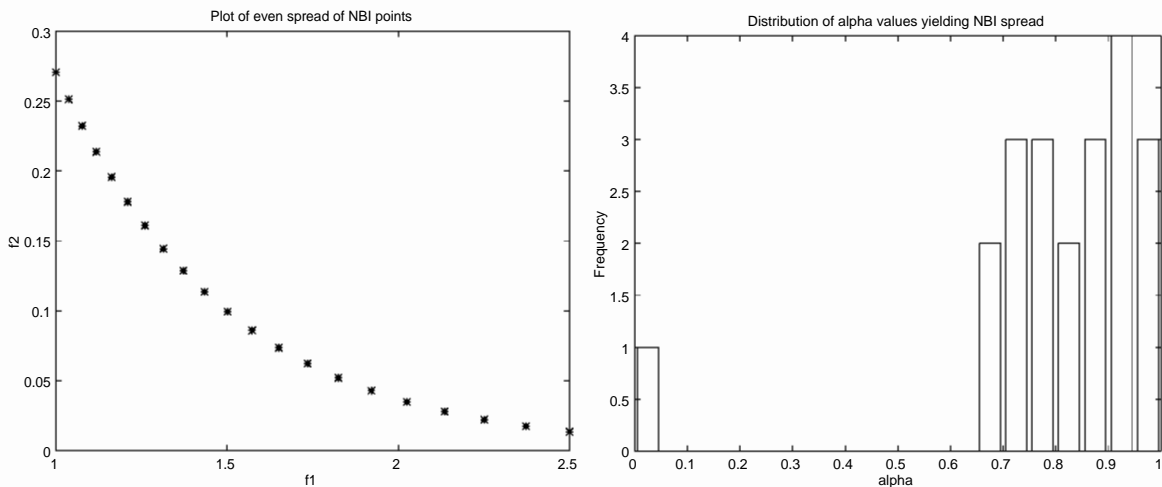


Figure 8: Distribution of α for even spread on $\psi(f_1) = 2e^{-2f_1}$ for $f_1 \in [1, 2.5]$

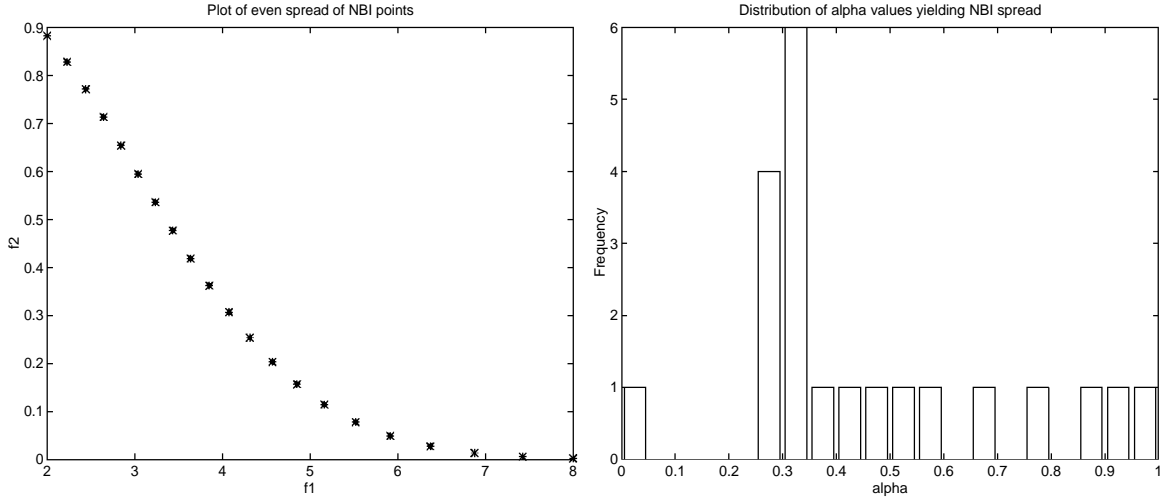


Figure 9: Distribution of α for even spread on $\psi(f_1) = 2e^{-\frac{(f_1-1)^2}{8}}$ for $f_1 \in [2, 8]$

Clearly, without prior knowledge of the shape of the Pareto curve it is impossible to try to find values of α that map out even a fairly uniform spread of points on the Pareto curve.

3.3 Spread of α yielding even spread on example problem

If the shape of the Pareto curve for the earlier two objective example were known, it would be possible to determine the values of α which would yield a uniform spread of Pareto points. However, since $\psi(f_1)$ is not known for this Pareto curve, we tried to estimate it roughly by fitting a nonlinear model of f_1 to the set of points on the curve obtained using NBI (as in fig. 5). The model fit to the set of points (f_1, f_2) was the following sum of three exponentials

$$\hat{f}_2 = \hat{\psi}(f_1) = p_1 + p_2 e^{-p_3 f_1} + p_4 e^{-p_5 f_1} + p_6 e^{-p_7 f_1}$$

The quantity $(f_2 - \hat{f}_2)^2$ summed over all the Pareto points was minimized over the parameters p_1, p_2, \dots, p_7 yielding the following nonlinear fit to the Pareto curve with an acceptable residual of 0.0161

$$\hat{f}_2 = -12.8735 + 314.2471e^{-11.6855 f_1} + 13.9954e^{-0.0456 f_1} + 1.524e^{-0.9896 f_1}$$

Estimating $\frac{\partial f_2}{\partial f_1}$ by $\hat{\psi}'(f_1)$ at the Pareto points obtained using NBI, a set of predicted α values were obtained using (1). Minimizing convex combinations of the two objectives for these values of α now yields the spread shown in fig. 11, which is a big improvement on the convex combinations spread in fig. 5.

The distribution for these α values is shown in a histogram in fig. 12.

4 Conclusion

The drawbacks of minimizing weighted convex combinations are obvious. It is quite impossible to know the correct weights needed to generate points evenly spread on the Pareto curve without actually knowing the shape of the Pareto curve. The scene also looks

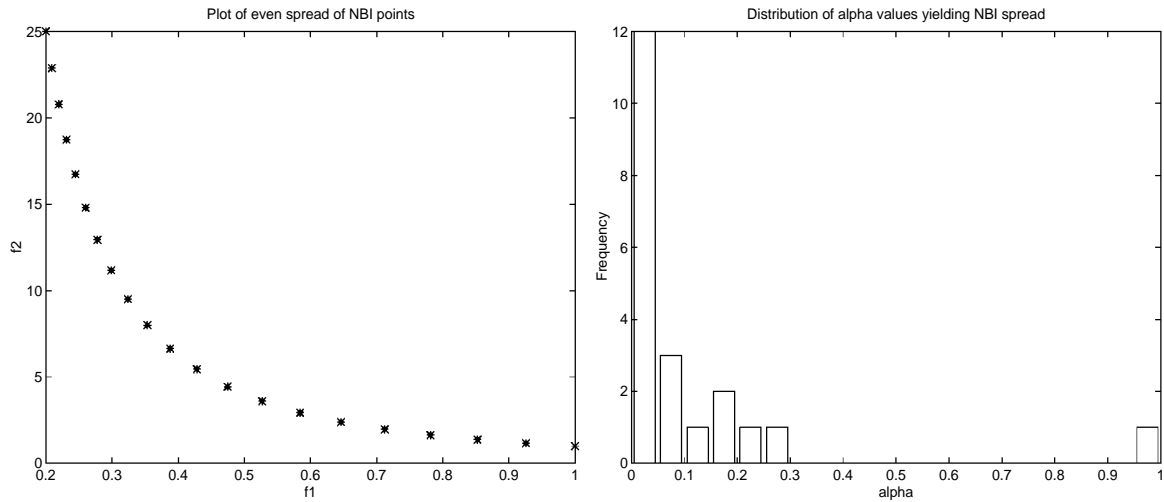


Figure 10: Distribution of α for even spread on $\psi(f_1) = \frac{1}{f_1^2}$ for $f_1 \in [2, 8]$

bleak for designers who hope to minimize just one weighted sum of objectives and expect to get a point in the middle region of the Pareto set. An alternative to this can be Normal-Boundary Intersection (see Das & Dennis [6]), a new technique developed recently which possesses nice properties including that of obtaining an even spread of points (visit <http://www.owl.net.rice.edu/indra/NBIhomepage.html> for software and papers). The chief role of this paper is to caution people who blindly lump several objectives into one using a weighted combination and the authors hope to have accomplished that.

References

- [1] J. Koski. *Multicriteria Truss Optimization*. Multicriteria Optimization in Engineering and in the Sciences. Edited by W. Stadler. New York, Plenum Press, 1988.
- [2] J. Jahn, J. Klose and A. Merkel. *On the Application of a Method of Reference Point Approximation to Bicriterial Optimization Problems in Chemical Engineering*. ADVANCES IN OPTIMIZATION. Ed. by W. Oettli, D. Pallaschke. Proceedings of the sixth French-German Colloquium on Optimization, Lambrecht, FRG, June 2-8, 1991. Springer-Verlag, Berlin.
- [3] H. Eschenauer, J. Koski and A. Osyczka. *Multicriteria Design Optimization*. Berlin, Springer-Verlag, 1990.
- [4] J. R. Rao and P. Y. Papalambros. *A Non-linear Programming Continuation Strategy for One Parameter Design Optimization Problems*. Proceedings of ASME Design Automation Conference, Montreal, Quebec, Canada, Sept. 17-20, 1989, pp. 77-89.
- [5] J. G. Lin. *Multiple-Objective Problems: Pareto-Optimal Solutions by Method of Proper Equality Constraints*. IEEE Transactions on Automatic Control, vol. AC-21, no.5, October 1976, pp. 641-650.

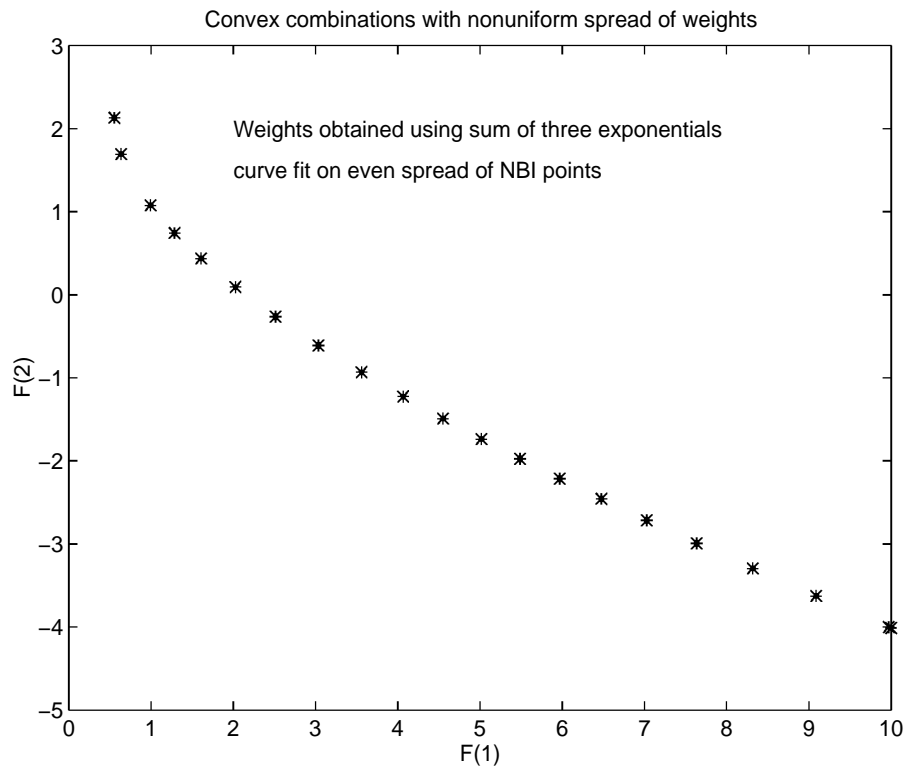


Figure 11: Pareto points obtained using values of α predicted using nonlinear fit on NBI points

- [6] I. Das and J. E. Dennis. *Normal-Boundary Intersection: An Alternate Approach for Generating Pareto-optimal Points in Multicriteria Optimization Problems*. Submitted to SIAM J. on Optimization, July 1996.

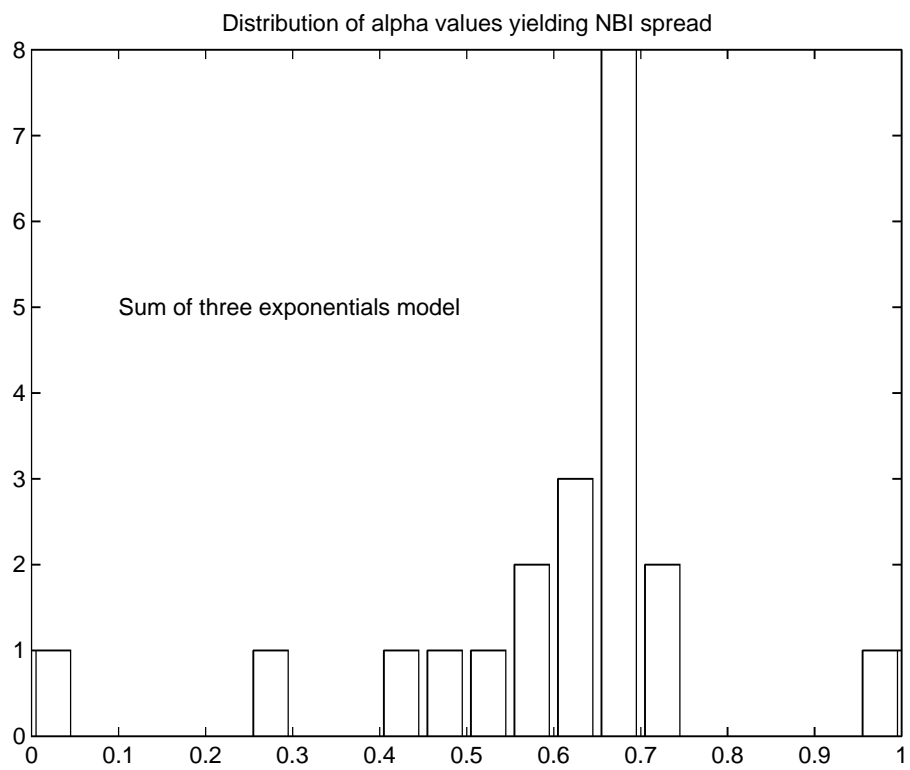


Figure 12: Distribution of values of α predicted using nonlinear fit on NBI points