Normal-Boundary Intersection: A New Method for Generating Pareto Optimal Points in Nonlinear Multicriteria Optimization Problems*

Indraneel Das† J. E. Dennis‡

July 1996

Abstract

This paper proposes an alternate method for finding several Pareto optimal points for a general nonlinear multicriteria optimization problem. Such points collectively capture the tradeoff among the various conflicting objectives. It is proved that this method is independent of the relative scales of the functions and is successful in producing an evenly distributed set of points in the Pareto set given an evenly distributed set of parameters, a property which the popular method of minimizing weighted combinations of objective functions lacks. Further, this method can handle more than two objectives while retaining the computational efficiency of continuation-type algorithms. This is an improvement over continuation techniques for tracing the trade-off curve since continuation strategies cannot easily be extended to handle more than two objectives.

*This research was partially supported by the Dept. of Energy, DOE Grant DE-FG03-95ER25257, the Air Force, Air Force Grant F49620-95-1-0210 and by the National Aeronautics and Space Administration under NASA Contract No. NAS1-19480 while the first author was in residence at the Institute for Computer Applications in Science and Engineering (ICASE), NASA Langley Research Center, Hampton, VA 23681-0001.

†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

A wide variety of problems arising in design optimization of engineering systems inherently involve optimizing multiple performance criteria (see, for example, Eschenauer, Koski and Osyczka [3] and Statnikov and Matusov [4]). For example, a typical bridge-construction design might involve simultaneously minimizing the total mass of the structure and maximizing its stiffness. However, it is highly improbable that these conflicting objectives would both be 'extremized' by the same design. Hence the designer makes some trade-off among the conflicting objectives in choosing the final design.

In mathematical notation a multicriteria optimization problem can be loosely posed as:

$$\text{“}\min_{x\in C}\text{”} F(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_n(x) \end{bmatrix}, \qquad n \geq 2, \qquad \ldots(MOP)$$

where

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

$F : \Re^N \mapsto \Re^n$, $h : \Re^N \mapsto \Re^{ne}$ and $g : \Re^N \mapsto \Re^{ni}$ are twice continuously differentiable mappings, and $a \in (\Re \cup \{-\infty\})^N, b \in (\Re \cup \{\infty\})^N$, 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 generally minimize every f_i simultaneously, a concept of optimality which is useful in the multiobjective framework is that of Pareto optimality, as explained below:

Definition: The vector $F(\hat{x})$ is said to *dominate* another vector $F(\bar{x})$, denoted $F(\hat{x}) \prec F(\bar{x})$, if and only if $f_i(\hat{x}) \leq f_i(\bar{x})$ for all $i \in \{1, 2, \ldots, n\}$ and $f_j(\hat{x}) < f_j(\bar{x})$ for some $j \in \{1, 2, \ldots, n\}$. A point $x^* \in C$ is said to be *globally Pareto optimal* or a *globally efficient* point for (MOP) if and only if there does not exist $x \in C$ satisfying $F(x) \prec F(x^*)$. $F(x^*)$ is then called *globally non-dominated* or *non-inferior*.

Computational methods for general nonlinear multicriteria optimization, including the one described in this paper, can at best guarantee local Pareto optimality of the obtained solution. The definition of local Pareto optimality is very similar to its global counterpart:

A point $x^* \in C$ is said to be locally Pareto optimal or a locally efficient point for (MOP) if and only if there exists an open neighborhood of x^*, $\mathcal{B}(x^*)$, such that there does not exist $x \in \mathcal{B}(x^*) \bigcap C$ satisfying $F(x) \prec F(x^*)$.

Pareto optimality will henceforth refer to local Pareto optimality unless qualified explicitly.

The shadow minimum or utopia point, F^*, is defined as the vector containing the individual global minima, f_i^*, of the objectives, i.e.,

$$F^* = \begin{bmatrix} f_1^* \\ f_2^* \\ \vdots \\ f_n^* \end{bmatrix}.$$

We assume here and henceforth the existence of a minimizer for each of our objectives. The shadow minimum could thus be attained only in the rare case when a single x minimizes all the objective functions. However, in practical situations, the best we can hope for is to get close to the shadow minimum and assure that there is an agreeable trade-off among the multiple objectives.

Very often in engineering applications the desired result helpful in facilitating design is a whole collection of Pareto optimal points, representative of the entire spectrum of efficient solutions. Thus ideally, the desired solution is the entire Pareto optimal set, which can be obtained for some small problems that allow themselves to be treated parametrically, resulting in closed-form expressions for the Pareto set (see Lin [5]). More recently, attempts have been made to approximate the entire curve of Pareto optimal solutions in bi-objective problems using techniques that trace the curve of parametrized optima (see Rakowska, Haftka and Watson [6], Rao and Papalambros [7], Lundberg and Poore [8]).

Another alternative acceptable in most applications is a discrete set of Pareto optimal points obtained by combining the multiple objectives into a single objective function and minimizing the single objective over various values of the parameters used to combine the objectives. For example, it is possible to generate a set of Pareto optimal points by minimizing a convex combination of the objectives, $w^T F(x)$, over $x \in C$, where $w \geq 0$ (component-wise) and $\sum_{i=1}^n w_i = 1$, and performing the minimization for different choices of w (see, among many others, Koski [9]). In this article, we propose a new method for generating Pareto optimal points which is at least as efficient as these methods and, unlike the techniques for tracing the curve of Pareto optimal solutions, can be applied to problems with more than two objectives.

2 Preliminaries

First let us introduce some terminology:

Convex Hull of Individual Minima (CHIM): Let x_i^* be the respective global minimizers of $f_i(x)$, $i = 1, \ldots, n$ over $x \in C$. Let $F_i^* = F(x_i^*)$, $i = 1, \ldots, n$. Let Φ be the $n \times n$ matrix whose i^{th} column is $F_i^* - F^*$ (sometimes known as the *pay-off matrix*. Then the set of points in $\Re^n$ that are convex combinations of $F_i^* - F^*$, i.e., $\{\Phi\beta : \beta \in \Re^n, \sum_{i=1}^n \beta_i = 1, \beta_i \geq 0\}$, is referred to as the Convex Hull of Individual Minima.

The set of attainable objective vectors, $\{F(x) : x \in C\}$ is denoted by $\mathcal{F}$, so $F : C \mapsto \mathcal{F}$, i.e., C is mapped by F onto $\mathcal{F}$. The space $\Re^n$ which contains $\mathcal{F}$ is usually referred to as the *objective space*. The map of C under F in the objective space is often called the *multi-loss map* (*bi-loss map*, if $n = 2$). We shall denote the boundary of $\mathcal{F}$ by $\partial\mathcal{F}$. The set of all Pareto optimal points is usually denoted by $\mathcal{P}$. The complete curve/surface of Pareto minima (continuous or not) is often referred to as the *trade-off function* (see p9, Haimes, Hall and Freedman [10]).

$CHIM_+$: Let $CHIM_\infty$ be the affine subspace of lowest dimension that contains the $CHIM$, i.e. the set $\{\Phi\beta : \beta \in \Re^n, \sum_{i=1}^n \beta_i = 1\}$. Then $CHIM_+$ is defined as the convex hull of the points in the intersection of $\mathcal{F}$ and $CHIM_\infty$. More informally, consider extending (or withdrawing) the boundary of the $CHIM$ simplex to touch $\partial\mathcal{F}$, the 'extension' of $CHIM$ thus obtained is defined as $CHIM_+$.

Henceforth, it shall be assumed that the objective functions have been defined with the shadow minimum shifted to the origin, so that all the objective functions are non-negative, i.e., $F(x)$ is redefined as:

$$F(x) \leftarrow F(x) - F^*.$$

We observe that in Fig.1, which shows the set $\mathcal{F}$ in the objective space, the point A is F_1^*, B is F_2^*, O is the shadow minimum (and the origin), the broken line segment AB is the $CHIM$, while the 'arc' ACB is the set of all Pareto minima in the objective space; alternately, the trade-off curve. In this (and any) problem with $n = 2$ (i.e., bi-objective), $CHIM = CHIM_+$. For $n > 2$ $CHIM$ may not equal $CHIM_+$ as in the case shown in fig. 3.

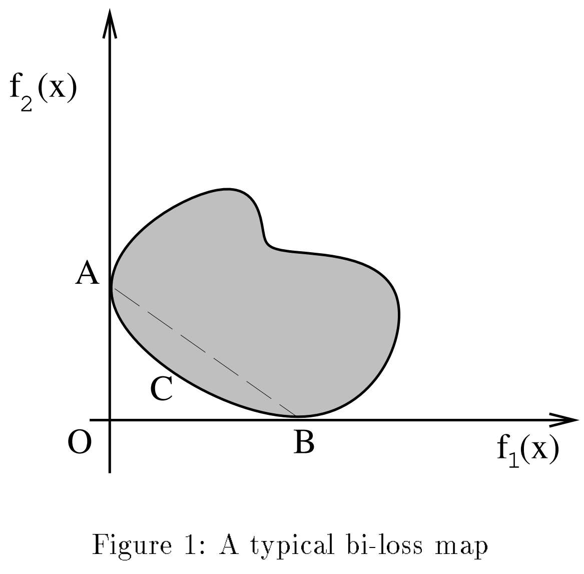


Figure 1: A typical bi-loss map

3 Central Idea

NBI is a technique intended to find the portion of $\partial\mathcal{F}$ which contains the Pareto optimal points. In order to facilitate the introduction of the preliminary idea behind NBI the discussion will assume that the vector of global minima of the objectives, F^*, is available. Later in Section 4.2 it will be argued how not having global minima usually renders very little injury to the technique.

The pivotal idea behind our approach will be introduced by means of a simple observation: the intersection point between the boundary $\partial\mathcal{F}$ and the normal pointing towards the origin emanating from any point in the $CHIM$ is a point on the portion of $\partial\mathcal{F}$ containing the efficient points. This point is also a Pareto optimal point unless it happens to lie in a 'highly concave' part of the boundary as shown in Fig. 2. It certainly is a Pareto optimal point when the trade-off surface in the objective space is convex, which happens in almost every application found in the literature. If the tradeoff surface is not convex, points in the concave part will still be obtained using NBI. If these points in the concave part are Pareto optimal this particular trait can be thought of as a merit of NBI over minimizing convex combinations of objectives which fails to obtain points in the non-convex parts of the Pareto set (see Das and Dennis[2]). If they are not Pareto optimal this might be characterized as a disadvantage. Nevertheless these points are useful even though they are not Pareto optimal, since they help in constructing a smoother approximation of the Pareto boundary.

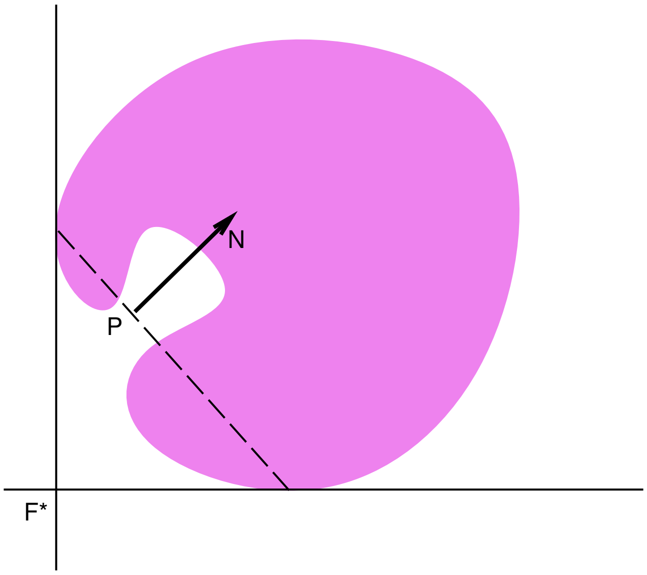


Figure 2: Boundary point obtained by NBI is not Pareto optimal

Now let us illustrate algebraically how any such boundary point can be found by solving an optimization problem. Given barycentric coordinates β, $\Phi\beta$ represents a point in the $CHIM$. Let $\hat{n}$ denote the unit normal to the $CHIM$ simplex pointing *towards* the origin; then $\Phi\beta + t\hat{n}, t \in \Re$ represents the set of points on that normal. The point of intersection of the normal and the boundary of $\mathcal{F}$ closest to the origin is the global solution of the following subproblem:

$$\begin{aligned} &\max_{x,t} \; t \\ s.t. \quad &\Phi\beta + t\hat{n} = F(x) \\ &h(x) = 0 \qquad\qquad (NBI_\beta) \\ &g(x) \leq 0 \\ &a \leq x \leq b\,. \end{aligned}$$

The vector constraint $\Phi\beta + t\hat{n} = F(x)$ ensures that the point x is actually mapped by F to a point on the normal, while the remaining constraints ensure feasibility of x with respect to the original problem (MOP). Observe that if the origin is not shifted to F^* the first set of constraints should read $\Phi\beta + t\hat{n} = F(x) - F^*$.

The subproblem above shall be referred to as the *NBI subproblem* and written as NBI_β since β is the characterizing parameter of the subproblem). Solutions of these subproblems will be referred to as *NBI points*. The idea is to solve NBI_β for various β and find several points on the boundary of $\mathcal{F}$, effectively constructing a pointwise approximation of the efficient frontier.

As indicated earlier, all NBI points are not Pareto optimal points. In biobjective problems, for every Pareto optimal point there exists a corresponding NBI subproblem of which it is the solution. The same is true for $n \geq 3$, with one difference: the coordinates of the parameter vector β for NBI_β may not be all non-negative. As a simple example, suppose $\mathcal{F}$ is a sphere in $\Re^3$ touching the coordinate axes. Then the $CHIM$ simplex is the triangle formed by joining the three points where the sphere touches the axes. Quite clearly $CHIM \neq CHIM_+$ so that there exist points in $CHIM_+ \setminus CHIM$ underneath which there are Pareto optimal points on the sphere. However since these points are not in $CHIM$, they do not satisfy $\beta_i \geq 0, \forall i$. Thus, by solving NBI_β for $\sum_{i=1}^{n} \beta_i = 1, \beta_i \geq 0, \forall i$, a portion of the Pareto set might be overlooked for problems with $n > 2$. However, these overlooked points are likely to be 'extremal' Pareto points lying near the periphery of the Pareto surface and are not interesting from the trade-off standpoint, which is our primary goal. Figure (3) illustrates a similar situation. The reader interested in how these peripheral Pareto points can be obtained can look in Das[1] for such a technique interesting at least from a theoretical standpoint.

4 Some details

4.1 Structure of Φ

The i^{th} column of Φ is described by

$$\Phi(:, i) = F(x_i^*) - F^*.$$

Since $f_i(x_i^*) = f_i^*$, clearly,

$$\Phi(i, i) = 0.$$

Moreover, since x_i^* is the minimizer of $f_i(x)$ over $x_j^*, j = 1, \ldots, n$

$$\Phi(j, i) \geq 0, j \neq i.$$

Thus a negative element in position (j, k) of Φ signifies that x_k^* is not the global minimizer of $f_k(x)$, and $f_k(x_j^*) < f_k(x_k^*)$, i.e., x_j^* improves on the current local minimum of $f_k(x)$. This fortunate occurrence provides a better starting point x_j^* for minimizing $f_k(x)$ and hence will lead to a better local minimum for f_k^* just by examining Φ.

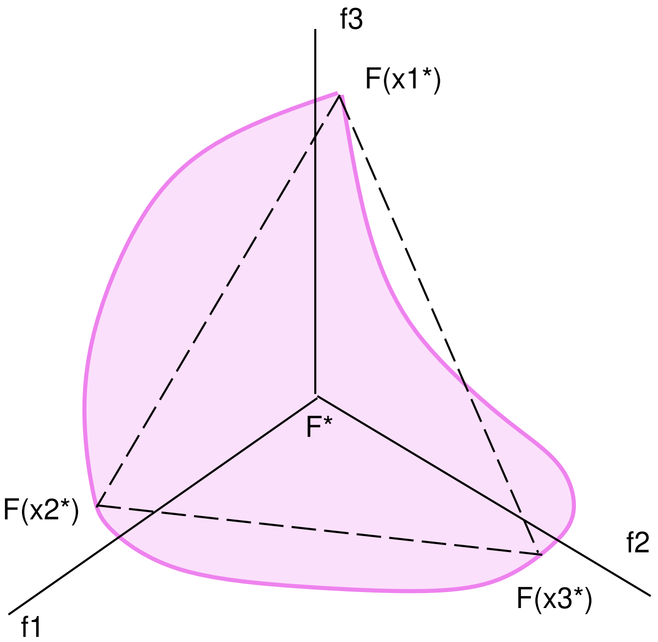


Figure 3: There exist Pareto optimal points not obtainable using NBI

4.2 Local versus global

As indicated earlier, most NBI points are guaranteed to be only locally Pareto optimal points. However, the components of the shadow minimum F^* being global minima of the objectives and the Pareto surface being convex is a sufficent, though far from necessary, condition for the NBI points to be globally Pareto optimal. In situations like the one shown in fig. 4 where the relevant part of $\partial\mathcal{F}$ is 'folded', the NBI point obtained may not be the one furthest out on the boundary along that normal because the solution of the nonlinear NBI subproblem is only guaranteed to be locally optimal. Thus the NBI point is not globally Pareto optimal.

Not being able to find globally Pareto optimal points is a drawback inherent in every method which finds a large number of efficient points of MOP. In homotopy methods, it would involve finding the global minimum of one of the two objectives in the very beginning. In methods which find efficient points by minimizing a single objective, only a global minimum of the scalarized objective would correspond to a globally efficient point.

Another important issue we had promised to deal with is the case when one or more components of the shadow minimum F^* consists of local but not global function minima. Such a case results in a different matrix Φ and more different goals $\Phi\beta$ for the NBI subproblems to improve on. These goals may be conservative or ambitious depending on the orientation of the incorrect $CHIM$ relative to the $CHIM$ formed using the true global minimizers. However having the 'incorrect $\Phi\beta$' may not preclude the NBI point from being a point on the efficient frontier, as in case of Fig. 5. Once the globally efficient point P in Fig. 5 has been found, a trivial examination of its components reveals that the current x_1^* is

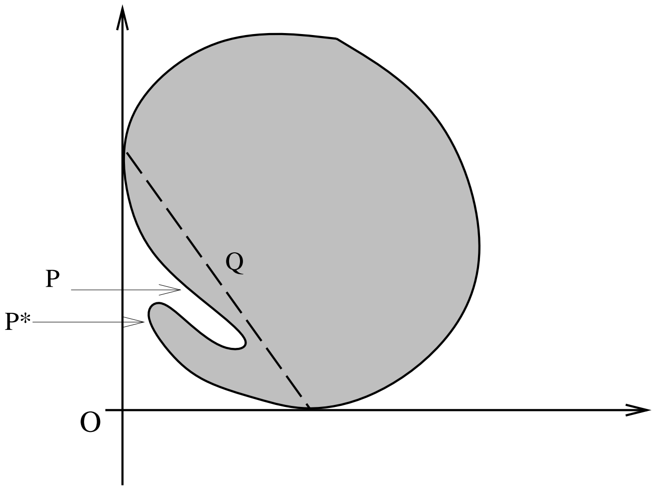


Figure 4: NBI started at Q converges to P (locally Pareto optimal), whereas the corresponding globally efficient point would have been P^*.

not the global minimizer of f_1 and provides a starting point, viz. P, for restarting the NLP to obtain a better local minimum of f_1. Then NBI can be restarted with this improved estimate of F^*. Some (if not all) globally Pareto optimal points will be obtained in most problems even if NBI is not restarted. Some points which are not Pareto optimal may be obtained if the targets $\Phi\beta$ are conservative as in Fig. 5. In cases such as the one in Fig. 6, it is possible that all globally Pareto optimal points may not be found using NBI and no indication regarding the local optimality of the function minima may be obtained.

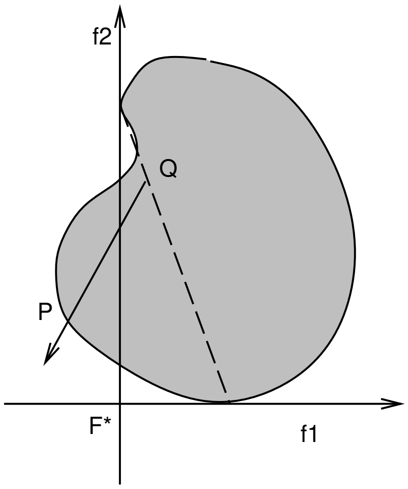


Figure 5: NBI started at Q converges to globally Pareto optimal point P even though all the function minima in the components of F^* are not global minima

However, in situations like the ones in fig. 7, owing to the fact that the individual function minima are only local, all the NBI points obtained are only locally Pareto optimal.

Computational experience (on more than just the problems mentioned here) shows that

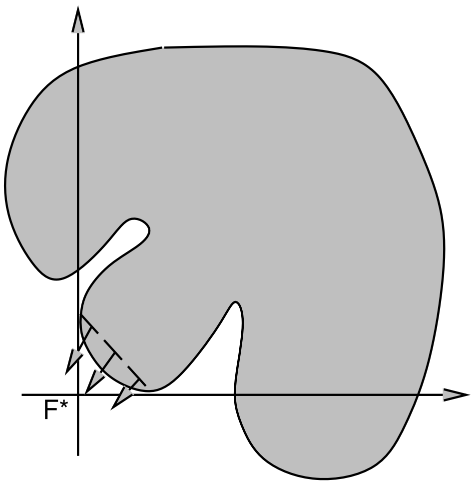


Figure 6: Local minima in the components of F^* might prevent NBI from obtaining Pareto points from every part of the efficient frontier

in cases where the global minima of the functions are not available at the onset, as NBI proceeds, either some component of Φ turns out to be negative or a function value of a particular objective is found that improves on its current local minimum value. This is not unusual given that the entire NBI procedure samples a large number of function values in the objective space.

To conclude this discussion and provide a general abstraction, it should be mentioned that whatever the components of F^* may be, NBI obtains at least the (local) boundary points dominated by F^* unless F^* is attainable, i.e., $F^* \in \mathcal{F}$. If $F^* \in \mathcal{F}$, Φ has a column of zeros and/or NBI obtains some (local) boundary point which dominates F^*, providing reason to refine F^* and start NBI all over again.

4.3 Quasi-normal instead of normal direction

The idea of a family of normals intersecting the boundary is valid even if we do not have the exact normal direction to the $CHIM$ simplex, but some *quasi-normal* direction $\hat{n}$ which has negative components, i.e. it points towards the origin. 'Shooting' a family of quasi-normal rays towards the boundary also gets us our desired boundary points. In practice we choose our quasi-normal direction to be an equally-weighted linear combination of the columns of Φ, multiplied by -1 to ensure that it points towards the origin. Explicitly,

$$\hat{n} = -\Phi e,$$

where e is the column vector of all ones.

The quasi-normal component defined as above has the property that the NBI point found for a certain β is completely independent of the scales of the objective functions. In other words, if NBI_β is re-solved with the objective functions rescaled by arbitrary factors, the NBI point found remains unchanged. This fact will be proved later.

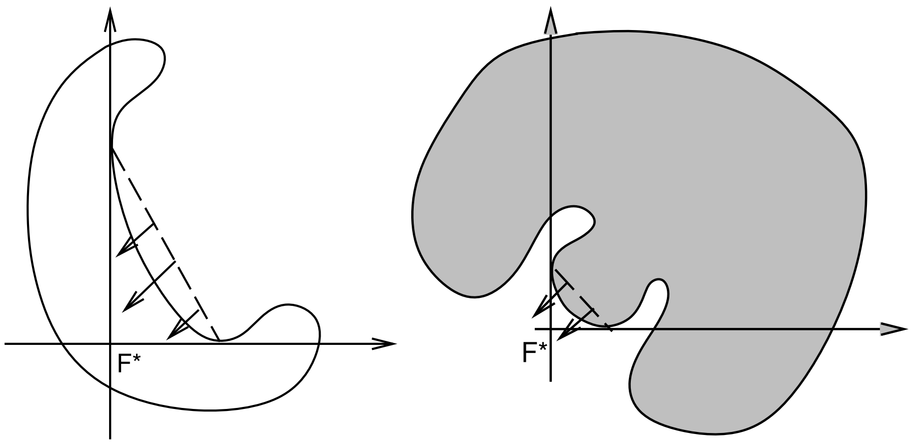


Figure 7: In the first case $\mathcal{F} = \partial\mathcal{F}$. Here, having local function minima in the components of F^* can cause NBI to find only locally Pareto optimal points.

Given that Φ has non-negative components as discussed in the previous subsection, it is clear that all components of Φe are non-negative.

Even though a quasi-normal direction will be used in our computations, we prefer to retain the name 'NBI', rather than change it to something like 'QNBI' hoping this misnomer would not be considered too harshly.

4.4 Further insight: NBI and Goal Programming

Since t is being maximized in the NBI subproblem and $\Phi\beta + t\hat{n} = F(x)$, $x \in C$, this maximization subproblem attempts to find a feasible point x as far from a 'target' point $\Phi\beta$ as possible, with $\hat{n} \leq 0$ (componentwise) guaranteeing non-increase in the components of $F(x)$ relative to the components of $\Phi\beta$ if the optimal value of t is non-negative.

This is similar to goal programming. If we take the Pareto set to be convex in the objective space, 'equality goal programming'[1] can be thought of as NBI where the direction $\hat{n}$ is one of the canonical basis vectors e_i (i.e. with 1 in the i^{th} position and 0 in the rest). To be precise, the subproblem NBI_β with $\hat{n} = -e_i$ has the same solution as the following goal programming problem:

$$\begin{aligned} &\min_x f_i(x) \\ s.t. \qquad & f_j(x) = (\Phi\beta)(j), \quad j = 1, \ldots, n, \ \ j \neq i \\ & x \in C, \end{aligned}$$

where $(\Phi\beta)(j)$ denotes the j^{th} component of the vector $\Phi\beta$.

Though posing the goals as equalities is untraditional, this kind of subproblem above for obtaining a Pareto optimal point is discussed in Lin[5] and [11].

In a future section NBI will be related to the traditional goal programming problem using Lagrange-Multiplier theory without assuming that the Pareto set is convex.

[1] Referring to goal programming where the goal constraints are equalities instead of inequalities.

4.5 Efficiently solving the subproblems

The following simple observation plays a key role in lowering the computational expense involved in solving the NBI subproblems:

Consider parameter vectors β and $\bar{\beta}$ such that β is 'close to' $\bar{\beta}$, i.e., $\|\beta - \bar{\beta}\|$ is 'small' in some norm. Then it is reasonable to expect that the solution (x^*, t^*) of NBI_β and the solution $(\bar{x}^*, \bar{t}^*)$ of $NBI_{\bar{\beta}}$ are 'close to each other'. Assume that we have solved $NBI_{\bar{\beta}}$ first and already have the point $(\bar{x}^*, \bar{t}^*)$. Then with $(\bar{x}^*, \bar{t}^*)$ as the starting point for solving NBI_β, the NBI subproblem solver can be expected to converge in relatively few iterations. It is this aspect of our algorithm that gives it the flavor of a continuation-type method.

Since we already have the individual minima of the functions, i.e., the vertices of the $CHIM$ simplex, we start at x_1^* and solve a 'nearby subproblem', and then a subproblem close to the one just solved, and so on.

Of course 'ordering the subproblems' may not be obvious for problems with more than two objective functions, but can still be achieved, as described in the next section.

5 Generating β and ordering the subproblems for more than two objectives

In this section we shall describe a (data) structure which simultaneously enables the generation of weights β and ordering the subproblems in a manner amenable not only to efficient solution but also to parallelization.

5.1 Generating β

Let us assume that for an n-objective problem, $\delta_j > 0$ is the uniform spacing between two consecutive β_j values (i.e., the 'stepsize' on the j^{th} component of β) for $j = 1, \ldots, n-1$. For simplicity, let us also assume that $\frac{1}{\delta_1}$ is an integer.

The possible values that can be assumed by β_1 are

$$[0, \delta_1, 2\delta_1, \ldots, 1].$$

Given a particular value of β_1, define $m_1 = \frac{\beta_1}{\delta_1}$. Then the possible values of β_2 corresponding to that value of β_1, (i.e. $\beta_1 = m_1\delta_1$) (all the β_i's must add up to 1) are

$$[0, \delta_2, 2\delta_2, \ldots, k_2\delta_2]\,,$$

where $k_2 = I[\frac{1-\beta_1}{\delta_2}] = I[\frac{1-m_1\delta_1}{\delta_2}]$.

Now define $m_2 = \frac{\beta_2}{\delta_2}$. Then the possible values of β_3 corresponding to $\beta_1 = m_1\delta_1$ and $\beta_2 = m_2\delta_2$ are

$$[0, \delta_3, 2\delta_3, \ldots, k_3\delta_3]\,,$$

where $k_2 = I[\frac{1-\beta_1-\beta_2}{\delta_3}] = I[\frac{1-m_1\delta_1-m_2\delta_2}{\delta_3}]$.

Thus, corresponding to $\beta_i = m_i\delta_i$, $i = 1, \ldots, j-1$, the possible values of β_j for $j = 2, \ldots, n-1$ are

$$[0, \delta_j, 2\delta_j, \ldots, k_j\delta_j],$$

where

$$k_j = I[\frac{1 - \sum_{i=1}^{j-1} m_i\delta_i}{\delta_j}].$$

Finally the last component of β is defined as

$$\beta_n = 1 - \sum_{i=1}^{n-1} \beta_i.$$

The entire data structure above can be thought of as a tree where the number of children varies with the node and generation. Each generation or level represents a component of β and each path from the root to the leaf represents a possible β vector. However, a tree structure is unnecessary for implementation; all that requires storage are the numbers δ_j. Nevertheless the tree is useful as a conceptual aid.

Of the subproblems generated by the weights in the above tree, n subproblems (with $\beta = e_i$) have already been solved in the course of finding F^*. It should also be noted that since $\frac{\delta_i}{\delta_j}$ is not necessarily an integer $\forall i < j$, the spacings between 'the last two' values of β_n may not be uniform.

5.2 Special case:

Equal stepsizes on all β_i

Let $\delta_i = \delta$, $i = 1, \ldots, n-1$

Also assume that $\frac{1}{\delta} = p$ is an integer.

As before, the possible values of β_1 are

$$[0, \delta, 2\delta, \ldots, 1]$$

Then the possible values of β_j corresponding to $\beta_i = m_i\delta_i$, $i = 1, \ldots, j-1$ for $j = 2, \ldots, n-1$ are

$$[0, \delta, 2\delta, \ldots, (p - \sum_{i=1}^{j-1} m_i)\delta]$$

As before, $\beta_n = 1 - \sum_{i=1}^{n-1} \beta_i$, and now all the β_n values are uniformly spaced.

Figure (8) shows part of the tree of β values for $n = 4$ and $\delta = 0.2$.

Number of NBI subproblems

In general, the number of NBI subproblems for a given n and a given $p = \frac{1}{\delta}$ is given by

$$\begin{pmatrix} n+p-1 \\ p \end{pmatrix}.$$

In spite of the fact that we only intend to solve 'nearby subproblems', the computational cost of solving a huge number of nonlinear programming problems can be quite daunting. This motivates the need for parallelization, as will be mentioned in the next section.

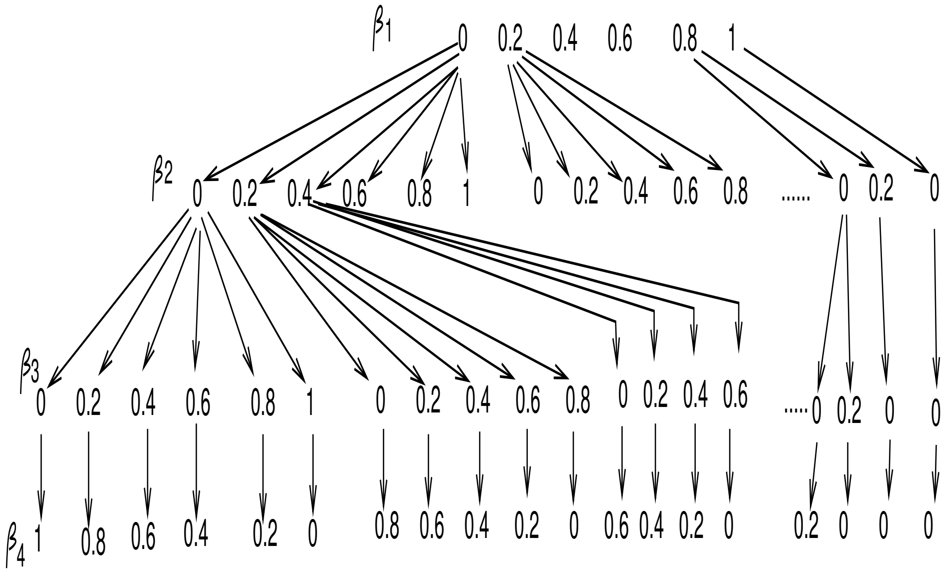


Figure 8: Generating β for $n = 4$, $\delta = 0.2$

5.3 Ordering the subproblems

Each path from the root of the tree (the topmost node) to a leaf (a member in the bottommost generation) represents a unique weight β. It should also be observed that the β vectors are already ordered on the basis of 'nearness' as one traverses the tree breadthwise. Thus a strategy for picking the order of the subproblems could be to start with the leftmost one (which has $\beta = e_n$ and is already solved) and solve the next one in the β_{n-1} generation (which is $\beta_{n-1} = \delta_{n-1}, \beta_n = 1 - \delta_{n-1}$), then the next one in the β_{n-1} generation ($\beta_{n-1} = 2\delta_{n-1}, \beta_n = 1 - 2\delta_{n-1}$), and so on until all the subproblems for $\beta_i = 0, i = 1, \ldots, n-2$ have been solved. Then we move to the next node in the β_{n-2} generation (i.e., with $\beta_i = 0, i = 1, \ldots, n-3$, $\beta_{n-2} = \delta_{n-2}$) and visit all the children of this node, with the starting points of the NBI subproblems chosen as the corresponding NBI subproblem solutions at the previous node.

This is where the scope for parallelization comes in. The solution of the first subproblem at the second node in the β_{n-2} generation did not have to wait until all the subproblems in the first node were solved. The first subproblem in the second node of the β_{n-2} generation with $\beta_{n-2} = \delta_{n-2}$, $\beta_{n-1} = \delta_{n-1}$, $\beta_n = 1 - \delta_{n-2} - \delta_{n-1}$ can be solved immediately after solving the first subproblem in the first node with $\beta_{n-2} = 0$, $\beta_{n-1} = \delta_{n-1}$, $\beta_n = 1 - \delta_{n-1}$. Thus the first subproblem in the second node can be solved in parallel with the second subproblem in the first node, ..., and the k^{th} subproblem in the second node can be solved in parallel with the $(k+1)^{th}$ subproblem of the first node. Further, the k^{th} subproblem in the third node can be solved in parallel with the $(k+1)^{th}$ subproblem of the second node, with the solution of the k^{th} subproblem of the second node as the starting point, and so on. This entire process of efficient parallelization is one of the topics of future research.

6 Relationship between the NBI subproblem and minimizing a convex combination of the objectives

In this section we illustrate how the NBI subproblem is related to the popular method of minimizing a convex combination of the objectives. This demonstrates how to go back and forth between the NBI parameter β and the convex combinations weight vector w for a particular Pareto point. The following discussion also demonstrates that corresponding to every w there exists a β such that NBI_β has the same solution as LC_w, but the converse is not true. In other words, this proves that there might be points obtainable using NBI not obtainable by minimizing convex combinations.

Given a Pareto point x^*, the problem can be thought of as being constrained only by the vector of equalities and binding inequalities and bounds at x^*. Let us denote this augmented vector of equalities by $\bar{h}(x)$. Let $w \in (\Re_+ \cup \{0\})^n$, $\sum_1^n w_i = 1$, denote a positive, convex weighting of the objectives. The weighted linear combination problem for obtaining a Pareto optimal point is then written as

$$\min_x w^T F(x)$$

$$s.t. \quad \bar{h}(x) = 0. \tag{1}$$

The solution of the problem above will be referred to as an *LC point*, and the problem denoted by LC_w. Part of the first-order necessary or KKT conditions for optimality of (x^*, λ^*) for problem (1) is

$$\nabla_x F(x^*)w + \nabla_x \bar{h}(x^*)\lambda^* = 0\,. \tag{2}$$

Similarly, if β denotes the vector of parameters in NBI_β, the NBI subproblem can be written as

$$\min_{x,t} -t$$

$$s.t. \quad F(x) - \Phi\beta - t\hat{n} = 0$$

$$\bar{h}(x) = 0\,. \tag{3}$$

Part of the KKT condition for optimality of $(x^*, t^*, \lambda^{(1)*}, \lambda^{(2)*})$ is

$$\nabla_x F(x^*)\lambda^{(1)*} + \nabla_x \bar{h}(x^*)\lambda^{(2)*} = 0 \tag{4}$$

$$-1 + \hat{n}^T \lambda^{(1)*} = 0,$$

where $\lambda^{(1)} \in \Re^n$ represents the vector of multipliers corresponding to the constraints $\Phi\beta + t\hat{n} - F(x) = 0$, and $\lambda^{(2)} \in \Re^{ne}$ denotes the multipliers of the equality constraints $\bar{h}(x) = 0$.

Claim:

Suppose $(x^*, t^*, \lambda^{(1)*}, \lambda^{(2)*})$ is the solution of NBI_β and $\sum_1^n \lambda_i^{(1)*} \neq 0$. Now define the components of the vector w as

$$w_i = \frac{\lambda_i^{(1)*}}{\sum_1^n \lambda_i^{(1)*}}.$$

Then, problem (1) with the above convex weighting vector w has the solution

$$[x^*, \lambda^* = \frac{1}{\sum_1^n \lambda_i^{(1)*}} \lambda^{(2)*}].$$

Proof:

Dividing both sides of (4) by the scalar $\sum_1^n \lambda_i^{(1)*}$ and observing that $\bar{h}(x^*) = 0$, the equivalence between (2) and (4) becomes obvious.

However, quite clearly, if for some i, the sign of $\lambda_i^{(1)*}$ is opposite to that of $\sum_1^n \lambda_i^{(1)*}$, then the vector w has a negative component and does not qualify as a weight for problem (1). In such a case, either the Pareto optimality of the NBI point $(x^*, t^*, \lambda^{(1)*}, \lambda^{(2)*})$ is questionable, or the Pareto point lies in a nonconvex part of the Pareto set (Pareto points in nonconvex parts of the Pareto set cannot be obtained by minimizing a linear combination of the objectives).

Just as the above analysis gives a method for obtaining w for problem LC_w given the corresponding solution of NBI_β, one can also obtain the NBI point corresponding to a given solution of problem LC_w with very little effort.

Claim:

Suppose (x^*, λ^*) solves problem LC_w. Let $(\bar{\beta}, t^*)$ be the solution of the $(n+1) \times (n+1)$ linear system

$$\Phi\beta + t\hat{n} = F(x^*)$$

$$\sum_{i=1}^n \beta_i = 1.$$

Then (x^*, λ^*) corresponds to the solution of NBI_β with $\beta = \bar{\beta}$, i.e., the solution of $NBI_{\bar{\beta}}$ is

$$(x^*, t^*, \lambda^{(1)*} = \frac{w}{w^T\hat{n}}, \lambda^{(2)*} = \frac{\lambda^*}{w^T\hat{n}}).$$

Proof:

Let us divide (2) on both sides by $w^T\hat{n}$.

This can always be done because, since w has nonnegative components (not all zero) and $\hat{n}$ has negative components, $w^T\hat{n} < 0$. Observing that $\lambda^{(1)*}$ defined above satisfies $\hat{n}^T\lambda^{(1)*} = 1$, it can be seen that the first part of the KKT conditions for $NBI_{\bar{\beta}}$ holds. Further observing that, $\bar{h}(x^*) = 0$ and $\Phi\beta + t\hat{n} = F(x^*)$, the required equivalence between LC_w and $NBI_{\bar{\beta}}$ follows.

7 Relationship between the NBI subproblem and goal programming using multipliers

A solution to an NBI subproblem is also a solution to a goal programming problem given that some assumptions holds. This is elaborated on below, using the same type of multiplier argument as used to relate NBI_β to LC_w.

Claim:

Suppose $(x^*, t^*, \lambda^{(1)*}, \lambda^{(2)*})$ is the solution of NBI_β. Suppose that the components of $\lambda^{(1)*}$ are all of the same sign with at least one nonzero component. If $\lambda_k^{(1)*}$ is any such nonzero component, then x^* solves the following goal programming problem:

$$\begin{aligned} &\min f_k(x) \\ \text{s.t.} \quad & f_i(x) \leq \gamma_i, \forall i \neq k \\ & \bar{h}(x) = 0 \end{aligned} \tag{5}$$

with goals γ_i given by

$$\gamma_i = \begin{cases} f_i(x^*), & \text{if } \lambda_i^{(1)} \neq 0 \\ \text{any finite number} \geq f_i(x^*), & \text{if } \lambda_i^{(1)} = 0 \end{cases}$$

$\forall i \in \{1, 2, \ldots, n\} \setminus \{k\}$.

Proof:

Since $(x^*, t^*, \lambda^{(1)*}, \lambda^{(2)*})$ solve the NBI_β subproblem, they must satisfy (4). Given that $\lambda_k^{(1)*} \neq 0$, we can divide both sides of (4) and get

$$\nabla_x f_k(x^*) + \sum_{i=1, i \neq k}^{i=n} \nabla_x f_i(x^*) \frac{\lambda_i^{(1)*}}{\lambda_k^{(1)*}} + \nabla_x \bar{h}(x) \frac{\lambda^{(2)*}}{\lambda_k^{(1)*}} = 0$$

Now $\frac{\lambda_i^{(1)*}}{\lambda_k^{(1)*}} \geq 0$ because $\lambda_i^{(1)*}$ and $\lambda_k^{(1)*}$ are of the same sign. Then with $\frac{\lambda_i^{(1)*}}{\lambda_k^{(1)*}}$ as the multipliers of the $n-1$ inequality constraints in (5), the goals γ_i satisfy complementarity by definition, since

$$\begin{aligned} &\gamma_i = f_i(x^*) \text{ whenever } \lambda_i^{(1)*} \neq 0 \\ &\Rightarrow \frac{\lambda_i^{(1)*}}{\lambda_k^{(1)*}} (f_i(x^*) - \gamma_i) = 0 \ \forall i \neq k \end{aligned}$$

Moreover, since x^* is clearly feasible for (5), $(x^*, \frac{\lambda_i^{(1)*}}{\lambda_k^{(1)*}}, \frac{\lambda^{(2)*}}{\lambda_k^{(1)*}})$ solves (satisfies first order necessary conditions for minimizer for) problem (5).

8 Proof of independence with respect to function scales using the quasi-normal

In this section we shall prove that the NBI point found using the quasi-normal $\hat{n}$ and a particular β is independent of how the individual functions are scaled.

Let the objective functions be scaled by positive scalars s_i as

$$f_i(x) \leftarrow s_i f_i(x), \quad i = 1, \ldots, n.$$

In other words, if s is the vector with components s_i and $S = diag(s)$, then

$$F(x) \leftarrow SF(x).$$

Consequently

$$\nabla_x F(x) \leftarrow \nabla_x F(x) S\,,$$
$$\Phi = S\Phi\,.$$

The quasi-normal direction $\hat{n} = -\Phi e$ after scaling becomes $= -S\Phi e$.

Claim:

If $(x^*, t^*, \lambda^{(1)*}, \lambda^{(2)*})$ solves the unscaled NBI_β (i.e. with $S = I_n$), then $(x^*, t^*, S^{-1}\lambda^{(1)*}, \lambda^{(2)*})$ solves[2] NBI_β with the i^{th} function scaled by s_i as above.

Proof:

Since $(x^*, t^*, \lambda^{(1)*}, \lambda^{(2)*})$ solves the unscaled NBI_β (still with only equality constraints as in the previous section),

$$\nabla_x F(x^*)\lambda^{(1)*} + \nabla_x \bar{h}(x^*)\lambda^{(2)*} = 0$$
$$\hat{n}^T \lambda^{(1)*} = 1$$
$$\Phi\beta + t^*\hat{n} = F(x^*)$$
$$\bar{h}(x^*) = 0.$$

The first equation can be rewritten to state that the following holds:

$$(\nabla_x F(x^*)S)(S^{-1}\lambda^{(1)*}) + \nabla_x \bar{h}(x^*)\lambda^{(2)*} = 0. \tag{6}$$

The second equation implies

$$e^T \Phi^T \lambda^{(1)*} = 1$$
$$\equiv e^T \Phi^T S S^{-1} \lambda^{(1)*} = 1.$$

Since $S = S^T$, the above is the same as

$$(e^T (S\Phi)^T)(S^{-1}\lambda^{(1)*}) = 1. \tag{7}$$

The third equation can be rewritten as

$$\Phi\beta + t^*\Phi e = F(x^*)$$
$$\equiv\; S\Phi\beta + t^* S\Phi e = SF(x^*). \tag{8}$$

Clearly, equations (6),(7) and (8) imply that $(x^*, t^*, S^{-1}\lambda^{(1)*}, \lambda^{(2)*})$ solves NBI_β with the functions scaled by S.

[2] Here 'solves' means 'finds a stationary point of the nonlinear programming problem'.

(QED)

The above result does not depend on e being the vector of all ones and consequently holds if $\hat{n}$ is scaled by a factor, say, a normalization constant.

The above result suggests that no matter how disparately the different functions might be scaled, NBI with the quasi-normal finds a set of points as if the functions were all scaled to the same order of magnitude.

9 Advantages of NBI

- **Finds a uniform spread of Pareto points:** Consider any method which parametrically combines all the objective functions into a single objective and finds efficient points by minimizing the single objective for various values of the parameters. In general the mapping from the set of parameters to the set of Pareto optimal points is *not* one-to-one. Thus it might so happen that minimizations over several different parameters produce the very same point each time, resulting in fruitless computational expense; this is never the case with NBI.

  NBI can yield a good approximation of the Pareto set by solving fewer nonlinear programming subproblems than weighted convex combinations (that is, if convex combinations can yield a good approximation at all) because the NBI points are evenly spread.

  The inter-relationship between the linear combinations subproblem and the NBI subproblem provides more insight into why the linear combinations technique fails to give a uniformly distributed set of Pareto optima. By fixing the weights w in subproblem LC_w,in effect the multipliers of the corresponding NBI subproblem get fixed, thus partly restricting the solution of the resultant subproblem. Even if the Pareto optima are uniformly distributed in the Pareto set, there is no reason why the corresponding multipliers have to be uniformly distributed. More insight into the failures of convex combinations can be obtained from Das and Dennis[2].

  However, the weights in the linear combinations approach are often very desirable because they give an idea of the relative importance of the objectives. Thus obtaining the NBI points, which are uniformly distributed, and then finding the corresponding weights w for the NBI points can be very useful.

- **Advantages over homotopy techniques:** NBI improves over homotopy/continuation techniques for tracing the curve of Pareto optimal solutions, like the one discussed in Rakowska, Haftka & Watson [6], in the following respects:

  - *Applicable for more than two objectives.* NBI is formulated to handle an arbitrary number of objectives. On the other hand, for a multiobjective problem with more than two objectives the homotopy parameter is not a scalar and the associated differential equation is a system of nonlinear partial differential equations with not readily available boundary conditions, rather than an ordinary initial value problem, as in the case of two objectives.

- *Does not require exact Hessian.* Even for a bi-objective problem, solving the homotopy initial value problem requires exact second derivative information (i.e., the Hessian of the Lagrangian), whereas the NBI subproblem solver can use any nonlinear programming technique. Even if the NLP technique for the NBI subproblem requires gradient information, secant methods for NLPs make exact Hessians unneccesary.
- *Bypasses tracking active sets.* For problems with inequality constraints or explicit bounds on variables, homotopy techniques need to keep track of the changes in active sets of the inequality constraints or bounds meticulously in course of the numerical integration, which can present difficulties if the number of inequalities or bounds is large. On the other hand, an interior point NLP solver used as the NBI subproblem solver would handle this situation quite efficiently, and will not have a problem with frequent changes in the active set.

  It must be noted though that points where the active set changes provide important information to the designer. However homotopy needs to keep track of changes in active sets even in the uninteresting parts of the Pareto set, whereas once the NBI points are found it is not difficult to trace how the active set changes along the Pareto surface by examining the binding inequalities at the Pareto points.
- *Does not assume connectedness or smoothness of the Pareto set.* The homotopy technique assumes that the Pareto curve is continuous and differentiable, and also connected, to be able to integrate along the curve. This is not the case with NBI, though it might end up reporting some subproblems as infeasible if the Pareto set is disconnected.

- NBI improves on other traditional methods like goal programming in the sense that it never requires any prior knowledge of 'feasible goals'. It improves on multilevel optimization techniques from the tradeoff standpoint, since multilevel techniques usually can only improve only a few of the 'most important' objectives, leaving no compromise for the rest.

10 A Numerical Example

Below is a brief account of employing NBI techniques on a small bi-objective problem, stated below:

$$\min_x \begin{bmatrix} f_1(x) = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 \\ f_2(x) = 3\,x_1 + 2\,x_2 - \frac{x_3}{3} + 0.01\,(x_4 - x_5)^3 \end{bmatrix}$$

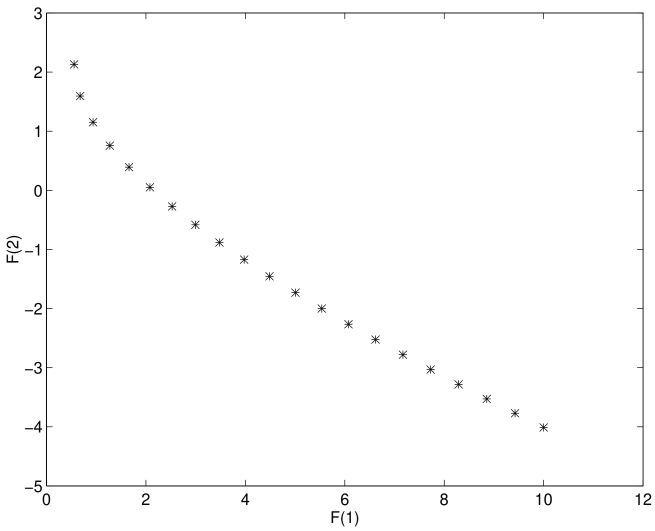
$$s.t. \quad x_1 + 2\,x_2 - x_3 - 0.5\,x_4 + x_5 = 2$$

$$4\,x_1 - 2\,x_2 + 0.8\,x_3 + 0.6\,x_4 + 0.5\,x_5^2 = 0$$

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 \leq 10.$$

NBI using the quasi-normal was run on the above problem with the evenly spread values of β with $\delta = 0.05$. The Pareto points thus obtained, independent of the scale on f_1, and are tabulated below and plotted in fig. (9):

β	Objective values
0.00, 1.00	10.0000 , -4.0111
0.05, 0.95	9.4254 , -3.7706
0.10, 0.90	8.8546 , -3.5276
0.15, 0.85	8.2882 , -3.2818
0.20, 0.80	7.7264 , -3.0329
0.25, 0.75	7.1698 , -2.7807
0.30, 0.70	6.6189 , -2.5247
0.35, 0.65	6.0743 , -2.2647
0.40, 0.60	5.5368 , -2.0000
0.45, 0.55	5.0072 , -1.7302
0.50, 0.50	4.4866 , -1.4546
0.55, 0.45	3.9764 , -1.1722
0.60, 0.40	3.4781 , -0.8820
0.65, 0.35	2.9939 , -0.5827
0.70, 0.30	2.5266 , -0.2724
0.75, 0.25	2.0801 , 0.0514
0.80, 0.20	1.6597 , 0.3922
0.85, 0.15	1.2740 , 0.7556
0.90, 0.10	0.9370 , 1.1506
0.95, 0.05	0.6754 , 1.5947
1.00, 0.00	0.5551 , 2.1306



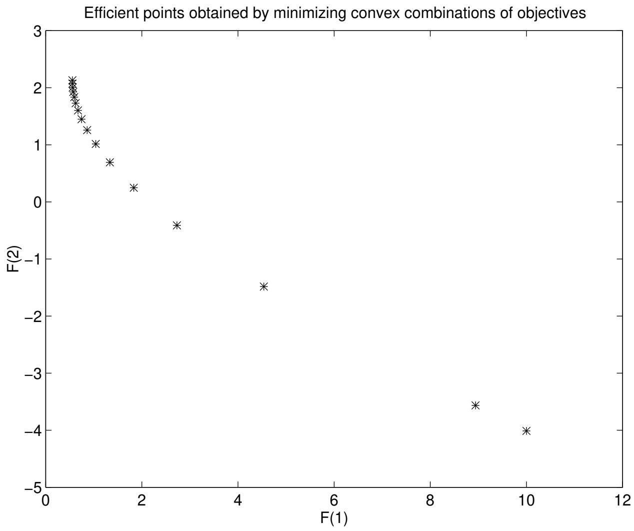
,


Figure 9: Pareto optimal vectors in the objective space using NBI and the method of convex combinations respectively

The method of convex combinations was run thrice on the same problem, with the weight vectors w assuming the same 21 uniformly spread values as the w vector above. The efficient solution scheme, i.e., starting the solution of a subproblem from the optimal point of the 'nearest subproblem', was used here as well.

When run on the original problem the minimizer of $f_2(x)$ was found six times for six different w, and there was a considerable gap 'in the middle' of the Pareto set [see fig.(9)].

With f_1 scaled by 5, the point found six times earlier was found only twice (i.e., heavily weighting the first objective made the minimizer move away from x_2^*.), but the Pareto optimal vectors obtained were concentrated at the $F(x_1^*)$ end and no 'middle ground for compromise' was captured.

With f_1 scaled by 10, the point repeated earlier was found only once, though the clustering at the $F(x_1^*)$ end increased [see fig.(10)].

The Pareto optimal vectors obtained using linear combinations are tabulated below:

Weights (w_1, w_2)	Objective values (original scale)	Objective values (f_1 scaled by 5)	Objective values (f_1 scaled by 10)
0.00 , 1.00	10.0000 , -4.0111	10.0000, -4.0111	10.0000, -4.0111
0.05 , 0.95	10.0000 , -4.0111	10.0000, -4.0111	4.8211, -1.6330
0.10 , 0.90	10.0000 , -4.0111	4.1857, -1.2896	1.1634, 0.8741
0.15 , 0.85	10.0000 , -4.0111	1.6131, 0.4330	0.7689, 1.4083
0.20 , 0.80	10.0000 , -4.0111	1.0180, 1.0451	0.6559, 1.6416
0.25 , 0.75	10.0000 , -4.0111	0.7975, 1.3592	0.6100, 1.7724
0.30 , 0.70	8.9403 , -3.5644	0.6953, 1.5506	0.5876, 1.8563
0.35 , 0.65	4.5379 , -1.4822	0.6412, 1.6796	0.5754, 1.9146
0.40 , 0.60	2.7307 , -0.4109	0.6100, 1.7725	0.5682, 1.9576
0.45 , 0.55	1.8319 , 0.2473	0.5909, 1.8425	0.5637, 1.9905
0.50 , 0.50	1.3357 , 0.6928	0.5788, 1.8973	0.5608, 2.0165
0.55 , 0.45	1.0425 , 1.0147	0.5707, 1.9413	0.5589, 2.0376
0.60 , 0.40	0.8615 , 1.2583	0.5654, 1.9773	0.5576, 2.0551
0.65 , 0.35	0.7463 , 1.4492	0.5618, 2.0075	0.5567, 2.0698
0.70 , 0.30	0.6719 , 1.6029	0.5593, 2.0331	0.5561, 2.0823
0.75 , 0.25	0.6236 , 1.7295	0.5576, 2.0551	0.5557, 2.0931
0.80 , 0.20	0.5926 , 1.8356	0.5565, 2.0741	0.5554, 2.1025
0.85 , 0.15	0.5734 , 1.9258	0.5558, 2.0909	0.5553, 2.1108
0.90 , 0.10	0.5622 , 2.0035	0.5554, 2.1057	0.5552, 2.1181
0.95 , 0.05	0.5567 , 2.0711	0.5551, 2.1188	0.5551, 2.1247
1.00 , 0.00	0.5551 , 2.1306	0.5551, 2.1306	0.5551, 2.1306

Clearly the inability of the method of convex combinations in adequately capturing the shape of the Pareto surface renders it fairly useless as a means of studying the trade-off between the conflicting objectives.

11 A truss optimization problem

Now we shall apply NBI to a truss optimization problem, a version of which has been studied in Koski [9]. The problem involves optimizing the design of a pin-jointed linear

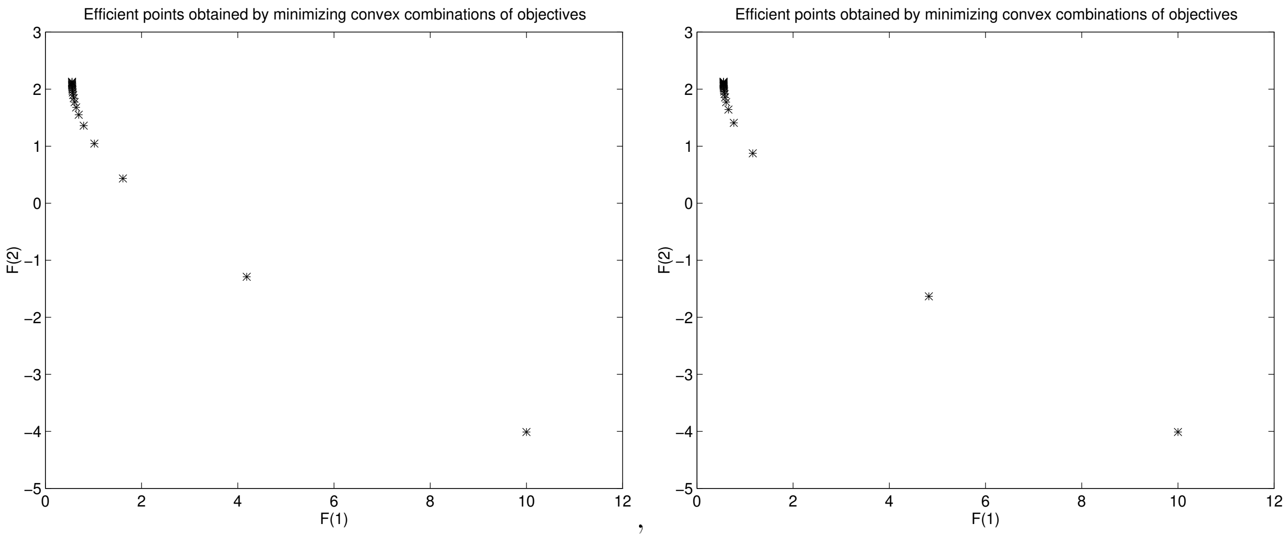


Figure 10: Pareto optimal vectors in the objective space using the method of linear combinations on the problem with f_1 scaled respectively by 5 and 10

truss structure as shown in fig. (11).

The problem is to find the optimal position of the vertical bar of fixed length L (the bars on the edge get fixed and their lengths decided accordingly) between 1/4 and 3/4 of the entire distance D and the optimal bar cross-sectional areas. The angles θ and α clearly depend on the chosen location x. Other optimization variables are the cross-sectional areas of the bars, a_1, a_2, a_3, allowed to vary between 0.8in^2 and 3.0in^2.

The objectives to be considered for minimization are the total volume of the structure, the displacement of the node and the absolute value of the stress in each of the three bars. In the structure considered by Koski in [9] the location of the middle bar was fixed, so that θ and α were also fixed. Also, his total volume was a linear function of the design variables, unlike in our formulation where total volume is expressed as $a_1 \frac{L}{\sin\theta} + a_2 L + a_3 \frac{L}{\sin\alpha}$.

Without going into further details of the problem and the data involved, which can be found in Das[1], we present some Pareto plots for subsets of the five objectives mentioned here. Fig. 12 shows the Pareto curve for minimizing the square of nodal displacement and the total volume with constraints on the absolute stresses in the three bars. The apparently unexpected gaps in the Pareto curve using NBI are points corresponding to which the NBI subproblems were infeasible owing to discontinuities in the Pareto set introduced by stringent stress inequalities.

Fig. 13 shows the Pareto curve for minimizing the stress in the right bar (the minimum value of the stress was positive and hence the absolute sign was dropped) and the total volume with constraints on the absolute stresses in the middle and left bars.

Given the individual minima and minimizers of the objectives at the outset, the number of floating point operations (flops) required in solving the subproblems for minimizing the stress in the right bar and the total volume using NBI and convex combinations for 21 parameter settings for each are shown in the table below:

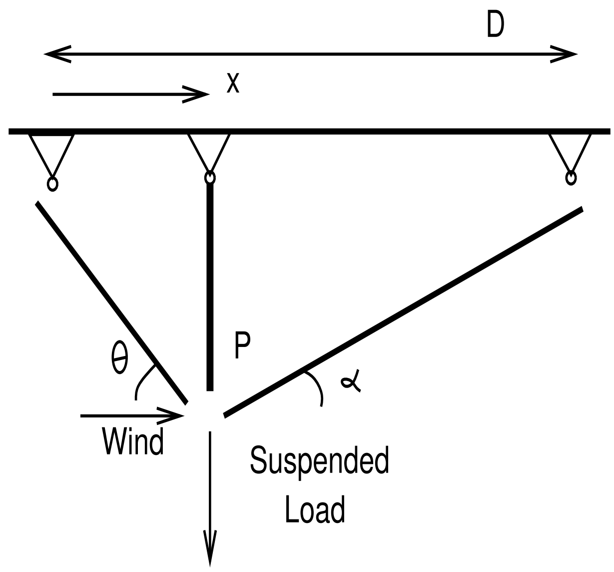


Figure 11: A truss structure under a suspended load and a wind load

	NBI	Convex Combinations
Number of distinct points	21	12
Number of flops	$2,962,068$	1,650,079
Flops per distinct point	141,050.86	137,506.58

The above table shows that NBI takes about twice as many flops but finds about twice as many distinct points, so that the number of flops per Pareto point is almost the same for the two methods (convex combinations wins marginally). But NBI yields a uniform spread of points representative of all parts of the Pareto set and hence a better model of the trade-off curve for the same effective computational cost.

Finally, fig. 14 shows the Pareto surface obtained using NBI with stress in the left bar, total volume and stress in the right bar as objectives. The uniform stepsize δ on each component of β was chosen to be 0.1 and 66 NBI subproblems were solved of which nine failed to converge owing to infeasibility. The whole process took about 11.4 million floating point operations.

A more detailed engineering-oriented treatment of this problem with tradeoff studies for more than the groups of objectives mentioned here can be found in Das [1].

12 Function scaling implicit in NBI

NBI using the quasi-normal component is unaffected by the function scales. However, as the functions get more disparately scaled, the Pareto set gets more 'stretched', and consequently the NBI points get further apart from each other. Consequently solving an NBI subproblem starting from the solution of the same nearby subproblem takes more iterations to converge. This was observed in the numerical example above and motivates

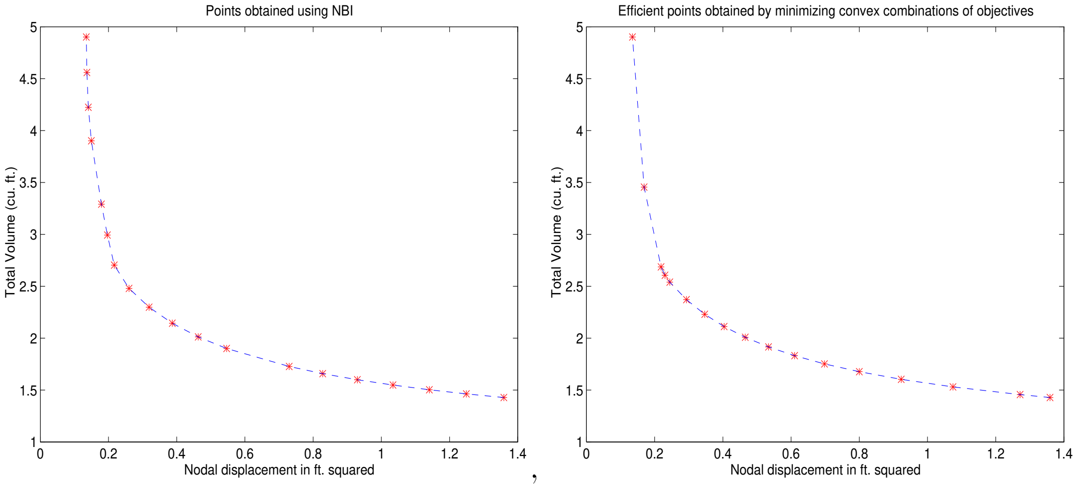


Figure 12: Pareto curves for minimizing nodal displacement and total volume (cu. ft.) using NBI and convex combinations

the need to scale the functions properly to remove this disparity in scales.

Geometrically it can be perceived that if the vertices of the $CHIM$ simplex are almost equidistant from the origin, i.e., the quantities

$$\|F(x_j^*) - F^*\|, \qquad j = 1, \ldots, n$$

are almost equal, then the quasi normal direction $\hat{n}$ is almost normal to the $CHIM$ simplex. This would achieve the 'minimally stretched' Pareto set we want and could also be a good scaling for the problem in the sense that all the functions would be about the same order of magnitude, and thus reduce possible ill-conditioning.

For the bi-objective problem, Φ is antidiagonal; thus a scaling that would achieve the above is obvious:

$$f_1 \leftarrow \frac{f_1}{f_1(x_2^*)}$$

$$f_2 \leftarrow \frac{f_2}{f_2(x_1^*)},$$

which gets each vertex of $CHIM$ to be unit distance from the origin.

However, the solution may not be so transparent for more than two objectives, and it may not be possible to get all the vertices exactly equidistant from the origin. So now we shall attempt to find function scalings $d_i > 0$ such that the functions scaled as

$$f_i \leftarrow \sqrt{d_i} f_i$$

will have the property that the *variance* among the scaled distances of the vertices from the origin, i.e.

$$\|\sqrt{D}(F(x_j^*) - F^*)\|^2, \qquad j = 1, \ldots, n$$

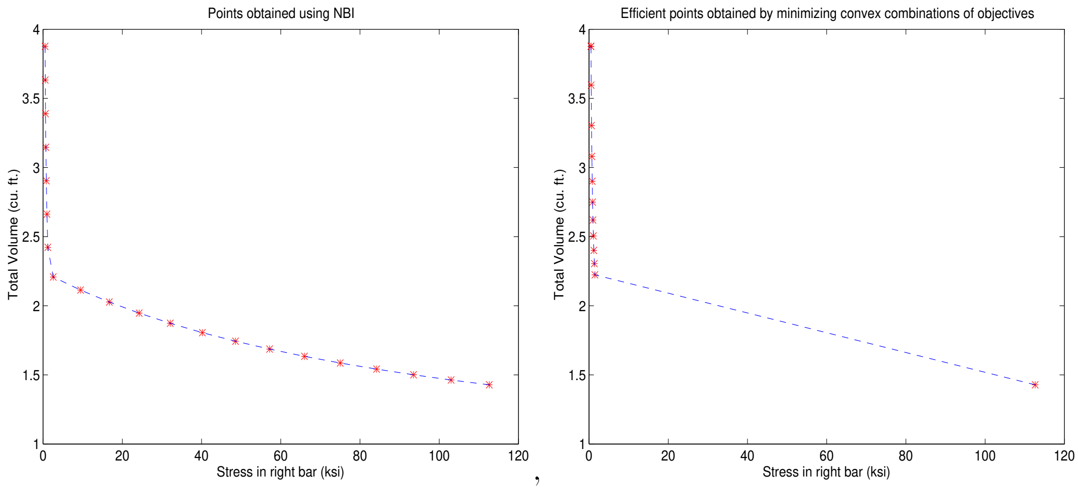


Figure 13: NBI points for minimizing stress in right bar and total volume (cu. ft.)

will be minimized ($D = diag(d)$, d represents the vector with components d_i).

Let $v_j = \|\sqrt{D}(F(x_j^*) - F^*)\|^2$, i.e.,

$$v_j = \sum_{i=1}^{n} d_i \phi_{i,j}^2,$$

where $\phi_{i,j}$ is the i^{th} row j^{th} column entry of the matrix Φ.

The mean square distance of the vertices is defined as

$$\bar{v} = \frac{1}{n}\sum_{j=1}^{n} v_j = \frac{1}{n}\sum_{i=1}^{n} d_i (\sum_{j=1}^{n} \phi_{i,j}^2).$$

The variance quantity to be minimized is given by

$$V(d) = \sum_{j=1}^{n}(v_j - \bar{v})^2;$$

i.e.,

$$V(d) = \sum_{j=1}^{n}\{\sum_{i=1}^{n} d_i\phi_{i,j}^2 - \sum_{i=1}^{n} d_i(\frac{1}{n}\sum_{j=1}^{n}\phi_{i,j}^2)\}^2.$$

Let A be the matrix with components $a_{i,j}$ given by

$$a_{i,j} = \phi_{i,j}^2 - \frac{1}{n}\sum_{k=1}^{n}\phi_{i,k}^2.$$

Then

$$V(d) = \sum_{j=1}^{n}(\sum_{i=1}^{n} d_i a_{i,j})^2;$$

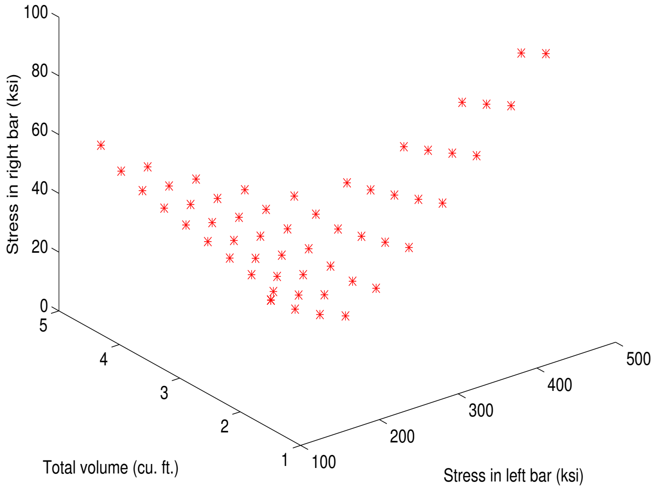


Figure 14: NBI points for minimizing stresses in left and right bars and total volume

i.e.,

$$V(d) = d^T AA^T d = \|A^T d\|^2.$$

This quadratic function is convex in d, and has an unconstrained minimizer at $d = 0$. Thus we shall demand a specific value of $\bar{v}$, which represents an average distance of the CHIM simplex from the origin and is roughly the same order of magnitude as a typical function value of any objective encountered in the computation.

Suppose we want a typical objective value to be τ, which could be something like 10. Then we would enforce

$$\bar{v} = \frac{1}{n}\sum_{i=1}^{n} d_i (\sum_{j=1}^{n} \phi_{i,j}^2) = \tau$$

along with a small lower bound on d_i. Thus the optimization problem to be solved to obtain our 'optimal' scales is

$$\min_d V(d) = d^T AA^T d$$

$$s.t. \qquad \sum_{i=1}^{n} d_i (\sum_{j=1}^{n} \phi_{i,j}^2) = n\tau$$

$$d_i >= 10^{-8} \quad , i = 1, \ldots, n.$$

Thus we can see how the matrix Φ suggests an 'improved scaling' of the objective functions, which is a bonus in the NBI approach.

It is worth observing that using the mean distance as opposed to the mean square distance in the last constraint would result in loss of convexity, hence the latter is preferred.

13 Conclusion

A technique was presented for finding Pareto optimal points of any smooth, constrained multiobjective problem with any number of objectives, perhaps restricted only by considerations of computational expense. The technique is efficient and has several useful properties, including that of obtaining an even spread of Pareto optimal points and invariance with respect to function scaling. This technique should be regarded as a tool for generating points from which the user can select the final design and not one that actually helps the user make that selection.

Further research is in progress regarding the implementational issues of parallelizing the solution of the NBI subproblem. Customized nonlinear programming techniques for solving the NBI subproblem will also be investigated.

A public domain Matlab 4.2 implementation of NBI is available free of charge at http://www.owlnet.rice.edu/~indra/NBIhomepage.html.

14 Acknowledgements

Firstly the authors would like to thank the referees for pointing out various improvements in the presentation of the material and presenting several insightful comments. The authors would also like to thank Paul Uhlig, Dept. of Mathematics, Rice University for several insightful discussions, Dr. Jagannatha Rao, Dept. of Mechanical Eng., University of Houston, for providing motivation and helpful comments, Jeffrey Hittinger, University of Michigan, Ann Arbor for a helpful discussion on data structures and Sanjeeb Dash, Rice University, for his algorithmic insights.

References

[1] I. Das. *Nonlinear Multicriteria Optimization and Robust Optimality.* Ph.D. Thesis, Dept. of Computational and Applied Mathematics, Rice University, Houston, TX 77005-1892.

[2] I. Das and J. E. Dennis. *A Closer Look at Drawbacks of Minimizing Weighted Sums of Objectives for Pareto Set Generation in Multicriteria Optimization Problems.* Dept. of Computational and Applied Mathematics Tech Report 96-36. Submitted for publication (Dec. 1996).

[3] H. Eschenauer, J. Koski and A. Osyczka. *Multicriteria Design Optimization.* Berlin, Springer-Verlag, 1990.

[4] Roman B. Statnikov and Joseph B. Matusov. *Multicriteria Optimization and Engineering.* New York, Chapman & Hall, 1995.

[5] J. G. Lin. *Three Methods for Determining Pareto-Optimal Solutions of Multiple-Objective Problems.* Directions in Large-Scale Systems, pp. 117-138. Edited by Y. C. Ho and S. K. Mitter. New York, Plenum Press, 1975.

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

[7] 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.

[8] B. N. Lundberg and A. B. Poore. *Bifurcations and Sensitivity in Parametric Programming.* Proceedings of Third Air Force/NASA Symposium on Recent Advances in Multidisciplinary Analysis and Optimization, Sept. 24-26, 1990, San Francisco, CA, pp. 50-55.

[9] J. Koski. *Multicriteria Truss Optimization.* Multicriteria Optimization in Engineering and in the Sciences. Edited by W. Stadler. New York, Plenum Press, 1988.

[10] Y. Haimes, W. Hall and H. Freedman. *Multiobjective Optimization in Water Resources Systems.* Amsterdam, Elsevier Scientific Publishing Co, 1975.

[11] 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.

[12] G. A. Katopis and J. G. Lin. *Non-inferiority of Controls under Double Performance Objectives: Minimal Time and Minimal Energy.* Proceedings of 7^{th} Hawaii Int. Conf. Syst. Sci., Honolulu, Hawaii, Jan 1974, pp. 129-131.

[13] J. G. Lin. *Circuit Design under Multiple Performance Objectives.* Proc. 1974 IEEE Int. Symp. Circuits & Systems, San Francisco, CA, pp. 549-552, Apr. 1974.