INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6” x 9” black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
RICE UNIVERSITY

Nonlinear Multicriteria Optimization and

Robust Optimality

by

Indraneel Das

A Thesis Submitted

in Partial Fulfillment of the

Requirements for the Degree

Doctor of Philosophy

Approved, Thesis Committee:

John E. Dennis, Jr., Chairman
Noah Harding Professor of Computational and Applied Mathematics
Rice University

Richard A. Tapia
Noah Harding Professor of Computational and Applied Mathematics
Rice University

C. Sidney Burrus
Mark Oshman Professor of Electrical Engineering
Rice University

Andrew Meade, Jr.
Associate Professor of Mechanical Engineering and Materials Science
Rice University

Houston, Texas
April, 1997
Abstract

Nonlinear Multicriteria Optimization
and
Robust Optimality

by

Indraneel Das

This dissertation attempts to address two important problems in systems engineering, namely, multicriteria optimization and robustness optimization. In fields ranging from engineering to the social sciences designers are very often required to make decisions that attempt to optimize several criteria or objectives at once. Mathematically this amounts to finding the Pareto optimal set of points for these constrained multiple criteria optimization problems which happen to be nonlinear in many realistic situations, particularly in engineering design.

Traditional techniques for nonlinear multicriteria optimization suffer from various drawbacks. The popular method of minimizing weighted sums of the multiple objectives suffers from the deficiency that choosing an even spread of 'weights' does not yield an even spread of points on the Pareto surface and further this spread is often quite sensitive to the relative scales of the functions. A continuation/homotopy based strategy for tracing out the Pareto curve tries to make up for this deficiency, but unfortunately requires exact second derivative information and further cannot be applied to problems with more than two objectives in general. Another technique, goal programming, requires prior knowledge of feasible goals which may not be easily available for more than two objectives.

Normal-Boundary Intersection (NBI), a new technique introduced in this dissertation, overcomes all of the difficulties inherent in the existing techniques by introducing a better parametrization of the Pareto set. It is rigorously proved that NBI is completely independent of the relative scales of the functions and is quite successful in producing an evenly distributed set of points on the Pareto set given an evenly dis-
tributed set of ‘NBI parameters’ (comparable to the ‘weights’ in minimizing weighted sums of objectives). Further, this method can easily handle more than two objectives while retaining the computational efficiency of continuation-type algorithms, which is an improvement over homotopy techniques for tracing the trade-off curve. Various aspects of NBI including computational issues and its relationships with minimizing convex combinations and goal programming are discussed in this dissertation. Finally some case studies from engineering disciplines are performed using NBI.

The other facet of this dissertation deals with robustness optimization, a concept useful in quantifying the stability of an optimum in the face of random fluctuations in the design variables. This robustness optimization problem is presented as an application of multicriteria optimization since it essentially involves the simultaneous minimization of two criteria, the objective function value at a point and the dispersion in the function values in a neighborhood of the point. Moreover, a formulation of the robustness optimization problem is presented so that it fits the framework of constrained, nonlinear optimization problems, which is an improvement on existing formulations that deal with either unconstrained nonlinear formulations or constrained linear formulations.
Acknowledgments

Firstly I would like to thank my family, my mother who never let me realize the lack of a father, and my wife, without whose care, support and sacrifices this dissertation would have been a seemingly impossible task.

I am highly indebted to Dr. John Dennis who supported and encouraged the work contained in this dissertation and hardly let me realize that he was away on sabbatical in my final year. I am also grateful to the Department of Energy, the Air Force, NASA and Boeing for funding my research. I also thank my committee members, Dr. Richard Tapia, Dr. Sidney Burrus and Dr. Andrew Meade for having provided me invaluable advice and encouragement. I also wish to thank Dr. Jagannatha Rao who provided me great motivation at an early stage of my work and continued his involved encouragement all throughout. It is impossible to conclude my acknowledgments without formally thanking Dr. Michael Trosset whose insightful comments, concerns and astute advice in matters ranging from commonplace to technical were a substantial help to me. It would also have been harder for me to get my thesis ready on time had it not been for the exceptionally high-quality systems maintenance and \LaTeX{} advice offered by Michael Pearlman.

I would also like to thank my friends at Rice with who I have shared many a laugh and who I hope to remain in touch with.

I also wish to thank Dr. Evin Cramer and Dr. Greg Shubin of the Boeing Company who expressed great interest in my work and have been actively involved in applying it to problems in aircraft design.

I would like to extend my thanks to all the scientists and engineers whose work I have built upon. Finally, I would like to extend my gratitude and apologies to all those who made vital contributions in my upbringing but unfortunately did not get acknowledged by name in this limited space.
Contents

Abstract ii
Acknowledgments iv
List of Illustrations viii
List of Tables xii

1 Introduction 1
  1.1 Introduction .................................................. 1

2 Problem Formulation 3

3 Traditional Methods for Multicriteria Optimization 7
  3.1 The Method of Weighted Convex Combinations ................. 7
  3.2 Goal Programming ............................................. 8
  3.3 Multilevel Programming ...................................... 9
  3.4 Curve tracing using homotopy techniques .................... 10
  3.5 Parameter Space Investigation (PSI) ........................ 11

4 Normal-Boundary Intersection (NBI) 13
  4.1 Terminology and Notation ................................... 13
  4.2 Central Idea .................................................. 14
  4.3 Local versus global ......................................... 17
  4.4 Obtaining the entire Pareto set for more than two objectives .. 19

5 Implementing NBI 25
  5.1 Structure of the payoff matrix Φ .......................... 25
  5.2 Quasi-normal instead of normal direction .................. 25
  5.3 Efficiently solving the subproblems ......................... 26
  5.4 Generating β and ordering the subproblems for more than two objectives ................................................. 27
    5.4.1 Generating β ............................................. 27
    5.4.2 Special case: Equal stepsizes on all βi ................. 28
    5.4.3 Ordering the subproblems ............................... 30
6 Theoretical Results for NBI and its Relationship with Other Methods

6.1 NBI as a Special Case of Goal Programming

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

6.3 Relationship between the NBI subproblem and goal programming using multipliers

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

6.5 Advantages of NBI

7 A Characterization of 'Good' Pareto Optimal Points

7.1 Efficiency of order $k$

7.1.1 An Example

8 Elaborations on the Shortcomings of Minimizing Convex Combinations of Objectives

8.1 Failure in Capturing Nonconvex Parts of the Pareto Curve

8.1.1 An Equivalent Problem

8.1.2 Geometrical Interpretation of the Trigonometric Linear Combinations Problem

8.1.3 Inexistence of a $(TLC_\theta)$ subproblem for points in the nonconvex part

8.2 Nonuniform Spread of Pareto Points using Uniform Spread of $\alpha$

8.2.1 Relationship between $\alpha$ and the slope of the Pareto curve

8.2.2 Distribution of $\alpha$ for uniform spreads of Pareto points

8.2.3 Spread of $\alpha$ yielding even spread on example problem

9 A Truss Optimization Problem

9.0.4 Objectives

9.0.5 Constraints

9.0.6 Parameter settings and bounds
9.1 Numerical results ........................................... 66
  9.1.1 Minimum displacement and minimum volume ............ 66
  9.1.2 Minimum stress in right bar and minimum volume ......... 68
  9.1.3 Minimum stress in left bar and minimum volume ........... 70
  9.1.4 Minimizing stress in the left bar, total volume and stress in
        the right bar ........................................... 71
  9.1.5 Compromising among nodal displacement, total volume and
        all the stresses ........................................... 72

10A Frame Optimization Problem 82

11 The Robustness Optimization Problem: An Application 86
  11.1 Introduction ............................................. 86
  11.2 The robustness optimization problem ......................... 87
        11.2.1 Evaluating the multidimensional integral ............. 89
  11.3 Trade-off inherent in the robustness problem ............... 91
        11.3.1 Minimizing dispersion as an alternative ............. 94
        11.3.2 Environmental Variables ............................ 95
  11.4 Expectation Evaluation for Equality-Constrained Problems . 96
        11.4.1 Approximating the expectation integral for constrained problems 97
  11.5 Handling inequality constraints ............................ 100
        11.5.1 Special case: bounds on design variables ............ 106
  11.6 A numerical example ....................................... 108

12 Conclusions 113

A A Bibliography of Applications 115

B Nondominated Vectors for Truss Problem 117

Bibliography 123
Illustrations

2.1 Mapping the feasible set to the objective space ........................................ 4

4.1 A typical bi-loss map ................................................................. 14
4.2 Boundary point obtained by NBI is not Pareto optimal ......................... 15
4.3 There exist Pareto optimal points not obtainable using NBI ................. 17
4.4 NBI started at Q may converge to P (locally Pareto optimal), whereas the globally efficient point along that normal is \( P^* \). ......................... 18
4.5 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 .......................................................... 19
4.6 Local minima in the components of \( F^* \) might prevent NBI from obtaining Pareto points from every part of the efficient frontier ....... 20
4.7 In the first case \( F = \partial F \). Here, having local function minima in the components of \( F^* \) can cause NBI to find only locally Pareto optimal points. ....................................................... 21
4.8 Obtaining NBI points underneath points on \( CHIM_+ \) that are not on \( CHIM \) .............................................................................. 22

5.1 Generating \( \beta \) for \( n = 4, \delta = 0.2 \) ................................................. 29
5.2 Pareto optimal vectors in the objective space using NBI and the method of convex combinations respectively .......................... 31
5.3 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 ........................................... 33

8.1 Getting a Pareto point by solving the Trigonometric Linear Combinations problem ....................................................... 51
8.2 Getting a convex Pareto curve by solving Trigonometric Linear Combinations problems ...................................................... 52
8.3 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 $(TLC_\theta)$ objective. However points $R$ or $M$ can be obtained as a solution of a $(TLC_\theta)$ problem. 53

8.4 Failure of Trigonometric Linear Combinations in capturing nonconvex parts of the Pareto set 54

8.5 Outputs of minimizing linear combinations and NBI respectively on a biobjective optimization problem using an even spread of parameters 55

8.6 Definition of uniform spread: Projections of the arcs between consecutive Pareto points on the $f_1$ axis are constant 57

8.7 Distribution of $\alpha$ for even spread on $\psi(f_1) = \frac{1}{f_1} \text{ for } f_1 \in [0.1, 2]$ 58

8.8 Distribution of $\alpha$ for even spread on $\psi(f_1) = 2e^{-2f_1} \text{ for } f_1 \in [1, 2.5]$ 59

8.9 Distribution of $\alpha$ for even spread on $\psi(f_1) = 2e^{-\left(\frac{f_1}{2}-1\right)^2} \text{ for } f_1 \in [2, 8]$ 60

8.10 Distribution of $\alpha$ for even spread on $\psi(f_1) = \frac{1}{f_1} \text{ for } f_1 \in [2, 8]$ 60

8.11 Pareto points obtained using values of $\alpha$ predicted using nonlinear fit on NBI points 61

8.12 Distribution of values of $\alpha$ predicted using nonlinear fit on NBI points 62

9.1 A truss structure under a suspended load and a wind load 63

9.2 NBI points for minimizing square of nodal displacement and total volume (cu. ft.) 67

9.3 Pareto set for square of nodal displacement and total volume (cu. ft.) obtained by minimizing convex combinations 68

9.4 Variation in the cross-sectional area of the left bar along the Pareto set for displacement and volume 69

9.5 Variation in the cross-sectional area of the right bar along the Pareto set for displacement and volume 70

9.6 NBI points for minimizing stress in right bar and total volume (cu. ft.) 71

9.7 Variation in the nodal displacement along the Pareto set for total volume and stress in right bar 72

9.8 Pareto set for stress in right bar and total volume (cu. ft.) obtained by minimizing convex combinations 73

9.9 NBI points for minimizing stress in left bar and total volume (cu. ft.) 74
9.10 Pareto points for minimizing stress in left bar and total volume (cu. ft.) using the method of convex combinations ................................................. 75
9.11 Variation in the position of the middle bar along the Pareto set for total volume and stress in left bar ................................................................. 76
9.12 NBI points for minimizing stresses in left and right bars and total volume .................................................................................................................. 76
9.13 Stress in left bar versus stress in right bar subplot for Pareto set for volume and stresses in left and right bars ......................................................... 77
9.14 Stress in left bar versus total volume subplot for Pareto set for volume and stresses in left and right bars ............................................................... 77
9.15 Total volume versus stress in right bar subplot for Pareto set for volume and stresses in left and right bars ............................................................... 78
9.16 NBI points for minimizing square of nodal displacement and total volume (cu. ft.) ................................................................................................. 78
9.17 Pareto points for minimizing square of nodal displacement and total volume using convex combinations ............................................................... 79
9.18 Nodal displacement plotted for various points on the volume vs maximum bar stress Pareto set versus maximum bar stress ........................................... 79
9.19 Nodal displacement plotted along the Pareto set for total volume and maximum bar stress .......................................................................................... 80
9.20 Triobjective subplot for the three bar stresses for the Pareto set for the three stresses and total volume ................................................................. 80
9.21 Biobjective subplot for each the three bar stresses vs volume for the four-objective Pareto set for three stresses and total volume .............................. 81

10.1 2-D frame under wind load .................................................................... 82
10.2 Pareto curves for minimizing the top storey displacement and total volume of frame with 3 supports and 4 storeys using NBI and convex combinations ................................................................. 85

11.1 The 'flatter' side of the minimum is more robust w.r.t. optimality .......... 91
11.2 Drawbacks of choosing minimizer of expectation as final design ........ 92
11.3 Drawback of minimizing dispersion in function values ......................... 93
11.4 Pareto set for minimizing exact expectation and original objective ....... 95
11.5 Pareto set for minimizing approximate expectation and original objective .................................................. 96
11.6 Expectation of \( f \) can be lower in a neighborhood of A than in the same for B, but solely because infeasible \( x \) values map to smaller \( f \) values and contribute to a smaller value of the expectation. .......................... 103
11.7 Trade-off curve for minimizing \( f(r) \) and its expectation using NBI ........................................ 110
11.8 Trade-off surface for weighted sum \( f(r) \), its expectation and \( \phi \) ..................................... 111
11.9 Pareto curve for weighted sum and its expectation subject to \( \phi \leq 0.01 \) .................. 111
11.10 Variation in \( \phi \) along Pareto curve for weighted sum and its expectation subject to \( \phi \leq 0.01 \) .......................................................... 112
## Tables

5.1  Number of NBI subproblems for different $n$ and $\delta$ .......................... 30
5.2  Nondominated vectors obtained using NBI ............................................. 32
5.3  Table of nondominated vectors obtained by minimizing convex sums of objectives ................................................................. 34

B.1  Table of nondominated objective vectors for truss problem ................. 117
Chapter 1

Introduction

1.1 Introduction

In recent years techniques for mathematical optimization have carved a niche for themselves in various aspects of design and decision-making in fields ranging from chemical and mechanical engineering to finance and political science. As more and more practical applications of optimization in design come to light, a realization that follows almost pari passu is that most of the final decisions are in fact guided by not one, but several criteria. For example, a structural engineer might think of designing a bridge aiming to minimize the total weight of the structure, but also realizes that maximizing the stiffness of the structure (i.e. minimizing the displacements of crucial node points on the structure) should be taken into account as a criterion. However, it is highly improbable that these conflicting objectives would both be ‘extremized’ by the same design, hence some trade-off between the objectives functions is desired to ensure an efficient design.

Problems like the one cited above are abundant in the literature and the author provides instances of at least a part of the gamut of applications of multicriteria optimization in Appendix A.

A comprehensive developmental history of multicriteria optimization is available in W. Studler’s [45], [44], [43]. The birth of multicriteria optimization can be traced back to as early as the late nineteenth century in consumer economics. In a book entitled Mathematical Psychics (1881), F. Edgeworth first coined the notion of multicriteria optimality which is today known as Pareto optimality, after V. Pareto, a welfare economist who continued and expanded Edgeworth’s work. The chief motivation for economists was the need to satisfy more than one customer simultaneously. Further contributions followed from game theory, in particular min-max theorems by Borel and von Neumann. Finally researchers in various fields of engineering and the social sciences started using these concepts in dealing with problems within their disciplines. Markowitz coined the term ‘efficient frontier’ to describe the curve of Pareto optimal solutions in the 1950’s in the context of portfolio optimization ([27], [28]). It wasn’t until the late 1970’s and early 1980’s that sophisticated techniques for mathemati-
cal programming made their way into multicriteria optimization. Until recently, the larger share of work in algorithm development for multicriteria optimization had been in the context of linear programming. Even though theory for nonlinear multicriteria optimization was available (Cohon [3] (1978), Sawaragi, Nakayama and Tanino [37] (1985)), algorithmic work for obtaining computational solutions of multicriteria problems was mostly done for problems with linear objectives and linear constraints as in Hartley [15] (1983), Thanassoulis [52] (1983), Yu [57] (1985), Steuer [48] (1986), Szidarovsky, Gershon and Duckstein [51] (1986), to name a few.

Currently nonlinear single criteria optimization is in a fair state of development so that designers have resources available to help them pose and solve single criteria nonlinear programming problems. Nonlinear multicriteria optimization on the other hand is in a more primitive state; difficulties exist in currently available techniques that greatly hinder multicriteria optimization from being used in as widespread a fashion as it otherwise would be. This dissertation is an attempt to help bridge this gap.

This disseration begins by formulating the multicriteria optimization problem and introducing relevant concepts in Chapter 2, including the concept of 'optimality' in the multiobjective framework. Chapter 3 surveys the leading methods that have been used so far in solving multicriteria optimization problems. Chapter 4 introduces a new technique, Normal-Boundary Intersection (NBI) for finding 'evenly distributed', 'optimal' points for the multicriteria optimization problem. Chapter 5 elaborates on the issues involved in implementing NBI. Chapter 6 establishes theoretical results related to NBI and its relationship with some of the traditional methods. Chapter 7 includes a brief description of a new concept that can be very useful in selecting the final point given several points that are 'optimal'. Chapter 8 investigates the shortcomings of minimizing weighted sums of the multiple objective functions, one of the most popular methods for multicriteria optimization. Chapter 9 and 10 present case studies of two multicriteria structural systems. Chapter 11 deals with robustness optimization, an important problem related to multicriteria optimization from a basic philosophical standpoint, and presents formulations for incorporating uncertainties in parameters in nonlinear design optimization problems.
Chapter 2

Problem Formulation

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

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

where

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

\[F : \mathbb{R}^N \rightarrow \mathbb{R}^n, \quad h : \mathbb{R}^N \rightarrow \mathbb{R}^{ne} \quad \text{and} \quad g : \mathbb{R}^N \rightarrow \mathbb{R}^{ni}\]

are twice continuously differentiable mappings, \(a \in (\mathbb{R} \cup \{-\infty\})^N, b \in (\mathbb{R} \cup \{\infty\})^N\), where \(N\) is 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 \(F(\dot{x})\) is said to dominate another vector \(F(\ddot{x})\), denoted \(F(\dot{x}) \prec F(\ddot{x})\), if and only if \(f_i(\dot{x}) \leq f_i(\ddot{x})\) for all \(i \in \{1, 2, \ldots, n\}\) and \(f_j(\dot{x}) < f_j(\ddot{x})\) 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 nondominated or non-inferior.

Computational methods for general nonlinear multicriteria optimization, including the one described here, 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 $x^*$, $B(x^*)$, such that there does not exist $x \in B(x^*) \cap C$ satisfying $F(x) \prec F(x^*)$.

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

The space in which $x$ belongs is referred to as the *decision space*, while the image of $C$ under the mapping $F$ is called the *criterion space* or *objective space* (see fig. 2.1). This set of attainable objective vectors, $\{F(x) : x \in C\}$, will henceforth be denoted by $\mathcal{F}$, so $F : C \mapsto \mathcal{F}$. The map of $C$ under $F$ in the objective space is 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 denoted usually by $\mathcal{P}$. Some disciplines like portfolio optimization refer to the boundary of nondominated solutions in the objective space as the *efficient frontier*.

Strictly speaking, as in Steuer [48], the term 'efficient' refers to the Pareto optimal points in the decision space while 'nondominated' refers to their images in the objective space.

![Figure 2.1 Mapping the feasible set to the objective space](image-url)
The complete surface/curve of Pareto minima is often referred to as the trade-off surface/curve. Let us suppose that this trade-off surface is described by

$$\mathcal{F}_p(f_1, f_2, \ldots, f_n) = 0.$$ 

Under some theoretical assumptions, the above define $n$ mappings describing each coordinate $f_i$ in terms of the others for points on the Pareto surface. Let these mapping functions be denoted by

$$f_i^p(f_1, \ldots, f_{i-1}, f_{i+1}, \ldots, f_n), \ i = 1, \ldots, n.$$ 

Haines, Hall and Freedman on page 9 of [14] refer to the above functions as trade-off functions.

A Pareto surface will be called convex if all the trade-off functions are convex, i.e. the Hessian of each $f_i^p$ is positive definite on its respective domain. By nonconvex parts of the Pareto set we shall refer to parts of the Pareto set where at least one of the trade-off functions has negative curvature, i.e. a negative definite Hessian.

A nonconvex Pareto surface can also be characterized geometrically in a way that will be useful in Chapter 8. A Pareto surface is nonconvex if there exists a tangent plane in the objective space that is tangent to more than one, countable number of distinct Pareto points.

A simple but useful concept is that of the shadow minimum or utopia point, $F^*$, defined as the vector with the individual (global) minimum of each objective function. $f^*_i$, as its components, i.e.,

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

We assume here and henceforth the existence of a minimum 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 acceptable compromise 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 [23]). More recently, attempts have been made to approximate the entire curve of Pareto optimal solutions in biobjective problems using techniques that trace the curve of parametrized optima (see Rakowska, Haftka and Watson [31], Rao and Papalambros [32], Lundberg and Poore [25]).

Another alternative acceptable in most applications is a discrete set of Pareto optimal points obtained by parametrically combining the multiple objectives into a single objective function and minimizing the single objective over various values of the parameters. The forthcoming chapter offers descriptions of the leading methods for multicriteria optimization that have been used so far.
Chapter 3

Traditional Methods for Multicriteria Optimization

3.1 The Method of Weighted Convex Combinations


In this technique $n$ weights $w_i$ are chosen such that $w_i \geq 0$, $i = 1, \ldots, n$ and $\sum_{i=1}^{n} w_i = 1$ and the following problem is solved:

$$\min_{x} \sum_{i=1}^{n} w_i f_i(x) = w^T F(x)$$

$$s.t. \quad x \in C. \quad \ldots (LC)$$

Claim:

The global minimizer $x^*$ of the above problem is a globally Pareto optimal point for $(MOP)$.

Proof:

Suppose $x^*$ is not globally Pareto optimal for $(MOP)$. Then there exists $\bar{x} \in C$ such that $F(\bar{x}) < F(x^*)$. Then there exists an index $k$ such that $f_k(\bar{x}) < f_k(x^*)$. Assuming $w_k > 0$, $w^T F(\bar{x}) < w^T F(x^*)$, contradicting that $x^*$ is the global minimizer of $w^T F(x)$ over $C$. If $w_k = 0$ for every $k$ such that $f_k(\bar{x}) < f_k(x^*)$, then an additional assumption of uniqueness of the global minimizer $x^*$ is required to establish this result.
The same argument can be repeated to prove that a local minimizer of the weighted convex combinations problem is locally Pareto optimal.

It should be noted that $\sum_{i=1}^{n} w_i = 1$ is only introduced as a normalization. Solving the weighted convex combinations problem for various parameter settings of $w$ thus yields several Pareto optimal points, with repetitions avoided by the use of the normalization.

The two major difficulties with the above idea are as follows:

- 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. What is even worse is that very often the same point is found for different convex combinations, i.e., the mapping between the parameters $w$ and the Pareto points is not one-to-one. As Lin [24] puts it, in the absence of convexity "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 middle 'ground' for any compromise, although such 'ground' may actually exist". Examples can be found in Lin [24], Katopis and Lin [17] and Lin [22].

The above shortcomings are investigated in detail in Chapter 8.

3.2 Goal Programming

Another popular method called goal programming involves fixing some desired target values for all but one of the objectives. Several books on goal programming are available including Ignizio [16], Romero [34], Schniederjans [38] and Lee [21].

If $\gamma_j$ is the goal specified on the $j^{th}$ objective then the corresponding goal programming problem is:

$$\min_{x} f_i(x)$$

s.t.  
$$f_j(x) \leq \gamma_j, \ j = 1, \ldots, n, \ j \neq i$$

$$x \in C.$$  

...(GP)

Claim:

If there exists a unique global minimizer $x^*$ of the goal programming problem (GP), then it is a globally Pareto optimal point for (MOP).
Proof:

Let us assume that the claim is false. Then, as before, the non-Pareto optimality of $x^*$ implies the existence of $\tilde{x} \in C$ and an index $k$ such that $f_k(\tilde{x}) < f_k(x^*)$. If this index $k = i$, this implies that $x^*$ is not a global minimizer of (GP). If $k \neq i$, then $x^*$ is not the unique global minimizer of (GP), hence the contradiction.

 Needless to say, the proof for local Pareto optimality of a local minimizer of (GP) proceeds similarly.

Solving (GP) for different goal settings can yield several Pareto optimal points. However, a prior determination of the goals may not always be possible. For biobjective problems it is straightforward to pick goals $\gamma_1$ on $f_1$ while minimizing $f_2$. This is because the minimum of $f_1$ and the value of $f_1$ at the minimizer of $f_2$ provide sharp lower and upper bounds on $\gamma_1$ for which (GP) will be feasible, provided the Pareto curve is continuous. However, for $n > 2$, goals $\gamma_1, \ldots, \gamma_{n-1}$ need to be set on more than one objective and knowing lower and upper bounds on every $\gamma_i$ is no longer sufficient to determine what combinations of $\gamma_i$ will result in (GP) being feasible. Thus solving (GP) for various goal settings without prior knowledge of the Pareto surface is likely to yield many infeasible goal programming problems even if the Pareto surface is continuous.

3.3 Multilevel Programming

Multilevel programming involves ordering the objectives based on importance and optimizing the individual objectives one by one starting with the most important one, taking care not to disturb the optimality of the more important objectives while optimizing a less important one. This is often referred to as the lexicographic method, as in Chankong and Haimes [2].

More explicitly let us assume that without loss of generality $f_1, f_2, \ldots, f_n$ is such an ordering of the objectives in decreasing order of importance. Then the multilevel programming problem is

$$\min_{x \in C} f_1(x)$$

↓ get solution $x^{(1)}$

$$\min_{x \in C} f_2(x) \text{ s.t. } f_1(x) \leq f_1(x^{(1)})$$

↓ get solution $x^{(2)}$
\[
\min_{x \in C} f_3(x) \quad \text{s.t.} \quad f_i(x) \leq f_i(x_*^{(i)}), \quad i = 1, 2 \\
\vdots \\
\min_{x \in C} f_n(x) \quad \text{s.t.} \quad f_i(x) \leq f_i(x_*^{(i)}), \quad i = 1, \ldots, n
\]

One drawback of multilevel programming is that the feasible sets for the subproblems lower down in the hierarchy may be empty. So the obtained point, though Pareto optimal, will have a high emphasis on minimizing the objectives higher up in the hierarchy with no trade-off or compromise for the rest. Moreover this technique can be used only to obtain a few Pareto optimal points and not to generate the entire Pareto set. However, if the hierarchical ordering is of great relevance in a particular decision-making problem this can certainly be the most appropriate method of choice.

### 3.4 Curve tracing using homotopy techniques

A more recent method as in Rao and Papalambros [32] and Rakowska, Haftka and Watson [31] applicable only to problems with two objectives is based on continuation/homotopy techniques. Consider a biobjective optimization problem with only equality constraints, converted to the weighted linear combinations form as below:

\[
\min_{x} \quad tf_1(x) + (1 - t)f_2(x) \\
\text{s.t.} \quad h(x) = 0.
\]

First-order necessary conditions for optimality of \((x, \lambda)\) in the above problem are:

\[
t \nabla_x f_1(x) + (1 - t) \nabla_x f_2(x) + \nabla_x h(x) \lambda = 0
\]

\[
h(x) = 0,
\]

which can be written equivalently as the system of nonlinear equations

\[
H(x(t), \lambda(t), t) = 0.
\]

The simplest homotopy technique, continuation,* involves differentiating \(H\) w.r.t. \(t\) to obtain an initial value problem (IVP), and then integrating it to obtain the curve of Pareto optimal points. Differentiation of \(H\) w.r.t. \(t\) gives

\[
\nabla_x H(x, \lambda, t)^T \nabla_t x + \nabla_\lambda H(x, \lambda, t)^T \nabla_t \lambda + \nabla_t H(x, \lambda, t) = 0.
\]

*Improvements on continuation can be made by using arc-length continuation, but our treatment suffices to illustrate the idea.
Let $J(x, \lambda, t) = \begin{bmatrix} \nabla_x H(x, \lambda, t)^T & \nabla_\lambda H(x, \lambda, t)^T \end{bmatrix}$.

Then

$$J(x, \lambda, t) \begin{bmatrix} \nabla_t x \\ \nabla_t \lambda \end{bmatrix} + \nabla_t H(x, \lambda, t) = 0,$$

so

$$\begin{bmatrix} \nabla_t x \\ \nabla_t \lambda \end{bmatrix} = -J^{-1}(x, \lambda, t) \nabla_t H(x, \lambda, t).$$

This ODE is solved with the initial conditions

$$x(t = 0) = x^*_2, \quad \lambda(t = 0) = \lambda^*_2$$

and is integrated up until $t = 1$, $(x^*_2, \lambda^*_2)$ being the solution of

$$\min_x f_2(x) \text{ s.t. } h(x) = 0.$$  

Though often efficient, these homotopy methods are deficient in the sense that they require exact second derivative information, and meticulous tracking of the active set for inequality constrained problems. More importantly, they cannot be extended easily to handle more than 2 objectives. These points are elaborated on in Chapter 6.

### 3.5 Parameter Space Investigation (PSI)

Developed by Sobol, Artobolevskii, Genkin, Kreinin, Sergeev, Statnikov and Matusov from 1967 onwards, this technique has been used profusely by Statnikov and Matusov [47] in tackling actual problems from engineering. A later article written by Steuer and Sun [49] as a computational investigation also tracks down its developmental history. The authors of this article observe that, though known in the republics of the former Soviet Union, this method is still relatively unknown in the West.

The PSI method is aimed at solving low-dimensional design problems in engineering that are highly nonlinear and even nonsmooth. It is essentially a random sampling procedure which involves the following steps:

- Sampling points on a uniform grid in the space of the design points constrained only by finite bounds.
- Evaluating the objectives and constraints (nonlinear inequalities) at each of the points.
- Discarding the points not satisfying the inequality constraints.
• Ordering the objective vectors and retaining the nondominated ones.

After one such run the designer sets goals on the objectives and tries to find objective vectors that satisfy them. If no such objective vector is found, either the goals are relaxed or more points are sampled and the process repeated. As observed by Steuer and Sun [49], it is practically impossible to apply the PSI method to problems with more than ten variables since the computational task is daunting. Furthermore there is no guarantee that the point chosen finally is actually Pareto optimal. However, this may be the only way to handle some of the complicated, nonsmooth problems in engineering at which the method is aimed (see Statnikov and Matusov [47]). It reportedly also has the additional property that the number of simulations or function evaluations required is insensitive to the number of objectives.
Chapter 4

Normal-Boundary Intersection (NBI)

Normal-Boundary Intersection (NBI) is a new technique motivated by a geometrical intuition to provide a better parametrization of the Pareto set than that provided by the weights in the weighted convex combinations problem. This parametrization is better in the sense that the points obtained using NBI provide a more even coverage of the Pareto surface, and this coverage does not miss the interesting 'middle part' of the Pareto surface. This 'middle part' of the Pareto surface is interesting because this is the part of the Pareto set furthest from all the individual function minima, thus providing in most cases the best compromise among the conflicting objectives.

In order to discuss NBI further, some terminology needs to be introduced to describe the geometric ideas underlying NBI.

4.1 Terminology and Notation

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 \( \Phi \) be the \( n \times n \) matrix whose \( i^{th} \) column is \( F_i^* - F^* \) sometimes known as the payoff matrix. Then the set of points in \( \mathbb{R}^n \) that are convex combinations of \( F_i^* - F^* \), i.e., \( \{ \Phi \beta : \beta \in \mathbb{R}^n, \sum_{i=1}^{n} \beta_i = 1, \beta_i \geq 0 \} \), is referred to as the Convex Hull of Individual Minima.

\( CHIM_+ \): Let \( CHIM_\infty \) be the affine subspace of lowest dimension that contains the \( CHIM \), i.e., the set \( \{ \Phi \beta : \beta \in \mathbb{R}^n, \sum_{i=1}^{n} \beta_i = 1 \} \). 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 nonnegative, i.e., \( F(x) \) is redefined as:

\[
F(x) \leftarrow F(x) - F^*.
\]
We observe that in Fig. 4.1, 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., biobjective), $CHIM = CHIM_+$. For $n > 2$ CHIM may not equal $CHIM_+$ as in the case shown in Fig. 4.3.

![A typical bi-loss map](image)

**Figure 4.1** A typical bi-loss map

### 4.2 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 4.3 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. 4.2. 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 trade-off 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 nonconvex parts of the Pareto set, as discussed in Chapter 8. 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.

![Diagram](image)

**Figure 4.2** 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 $\beta$, $\Phi \beta$ represents a point in the $CHIM$. Let $\hat{n}$ denote the unit normal to the $CHIM$ simplex pointing *towards* the origin; then $\Phi \beta + t\hat{n}, t \in \mathbb{R}$ 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:
\[
\max_{x,t} \quad t \\
\text{s.t.} \quad \Phi \beta + t \hat{n} = F(x) \\
h(x) = 0 \quad (NBI_\beta) \\
g(x) \leq 0 \\
a \leq x \leq b.
\]

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 \textit{NBI subproblem} and written as \( NBI_\beta \) since \( \beta \) is the characterizing parameter of the subproblem). Solutions of these subproblems will be referred to as \textit{NBI points}. The idea is to solve \( NBI_\beta \) for various \( \beta \) 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 \( \beta \) for \( NBI_\beta \) may not be all nonnegative. As a simple example, suppose \( \mathcal{F} \) is a sphere in \( \mathbb{R}^3 \) touching the coordinate axes. Then the \textit{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 \( NBI_\beta \) 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. Fig. 4.3 illustrates a similar situation. A method, interesting at least from a theoretical standpoint, for obtaining these peripheral Pareto points is described in Section 4.4.
4.3 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 sufficient, though far from necessary, condition for the NBI points to be globally Pareto optimal. In situations like the one shown in Fig. 4.4 where the relevant part of $\partial F$ 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 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 $\Phi$ and 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. 4.5. Once the globally efficient point $P$ in Fig. 4.5 has been found, a trivial examination of its components reveals that the current $x_1^*$ is 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. 4.5. In cases such as the one in Fig. 4.6, 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.

However, in situations like the ones in Fig. 4.7, owing to the fact that the individual function minima are only local, all the NBI points obtained are only locally Pareto optimal.
Figure 4.5  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

Computational experience (on more than just the problems mentioned here) shows that in cases where the global minima of the functions are not available at the onset, as NBI proceeds, either some component of $\Phi$ 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 $F^* \in \mathcal{F}$, $\Phi$ 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.

4.4 Obtaining the entire Pareto set for more than two objectives

It has been described earlier and in Fig. 4.3 why for a problem with more than 2 objectives it is possible to miss parts of the Pareto set 'near the edge', i.e., peripheral Pareto points, by running $NBI_\beta$ only for $\beta$ values corresponding to points on the CHIM. It is likely that for most engineering applications this will not be a shortcom-
Figure 4.6 Local minima in the components of $F^*$ might prevent NBI from obtaining Pareto points from every part of the efficient frontier.

ing, since the engineer probably will only want points in the interesting 'middle' part of the Pareto set. However pathological examples can be contructed where the only points that are Pareto optimal are solutions of $NBI_\beta$, where the $\beta$ values correspond to points lying on $CHIM_+$ but not on $CHIM$. Even otherwise, devising a strategy to find the Pareto points on the 'edge' is required because it renders completeness to NBI as a technique even though it may not be necessary in practice, hence the section follows. The reader interested only in practical applications may choose to omit the details of this section.

The central idea will be motivated by means of Fig. 4.8, which can be thought of as being a 'top view' of Fig. 4.3. Solving $NBI_\beta$ for $\beta$ corresponding to points on the $CHIM$ ABC in Fig. 4.8 fails to obtain the Pareto points lying underneath the shaded part of $CHIM_+$ (because the points in the shaded part do not lie on $CHIM$). We shall first develop how the Pareto points underneath the shaded part can be obtained just by modifying slightly the NBI subproblem.

Imagine a family of rays emanating from edge AB in the plane of $CHIM_\infty$ (i.e., the plane containing simplex ABC) directed towards the boundary as shown in the Fig. 4.8. The idea is to obtain points $P'$, $Q'$, $R'$, etc. on the boundary of $\mathcal{F}$ which lie on $CHIM_+$, and then solve the regular NBI subproblem with the quantity $\Phi \beta$
Figure 4.7 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.

substituted by points on the segments PP', QQ', RR', etc., so that now Pareto points lying underneath the points on the segments are obtained. In order to do this, a parametrization is needed for the points on those segments (i.e., our selected set of points on $CHIM$ but not on $CHIM_+$), so the points P', Q', R' where the ray intersects the boundary are required.

Let $\hat{n}_2$ be the direction of this family of rays. An obvious choice here is a linear combination of the two remaining edges AC and BC, say

$$\hat{n}_2 = (F(x^*_3) - F(x^*_2)) + (F(x^*_1) - F(x^*_2)) = M_2 e,$$

where the columns of the $3 \times 2$ matrix $M_2$ are $F(x^*_3) - F(x^*_2)$ and $F(x^*_1) - F(x^*_2)$.

Roughly speaking, the aim is to 'shoot' a family of rays along $\hat{n}_2$ from points on the segment AB towards the part of the boundary of $\mathcal{F}$ joining A and B. The points on the edge AB can be written as $\Phi \beta^{(2)}$ where $\beta^{(2)}$ is a positive convex combination vector as before with its second component $= 0$. Given the point $Q \equiv \Phi \beta^{(2)}$ for a particular $\beta^{(2)}$, the point $Q'$ can be found by solving the following:

$$\max_{t_2} t_2$$

$$s.t. \quad \Phi \beta^{(2)} + t_2 \hat{n}_2 = F(x).$$
Given the optimal $t_2^*$ that solves the above, the points on the segment $QQ'$ can be parametrized as

$$\Phi \beta^{(2)} + t_2 \tilde{n},$$

where $t_2 \in [0, t_2^*]$.

Now to obtain the Pareto points lying underneath points on the segment $QQ'$, the following $NBI_{\beta^{(2)}, t_2}$ subproblem is solved for a discrete set of values of $t_2 \in [0, t_2^*]$, if $t_2^* > 0$:

$$\max_{x,t} t$$

s.t. $\Phi \beta^{(2)} + t_2 \tilde{n}_2 + t \tilde{n} = F(x)$

$$x \in C.$$

As is evident, first finding $t_2^*$ for every setting of $\beta^{(2)}$ and then solving $NBI_{\beta^{(2)}, t_2}$ for all admissible settings of $\beta^{(2)}$ and $t_2$ yield a mesh of Pareto points on $\partial F$ underneath the shaded part of $CHIM$ shown in Fig. 4.8.

The above process should then be repeated for every edge in the $CHIM$ simplex. Below is a general description of this entire process.
Let $B_i$ be the set of possible values of $\beta$ with its $i^{th}$ component 0, i.e.

$$B_i = \{ \beta : \sum_{j=1}^{n} \beta_j = 1, \beta_i = 0, \beta_j \geq 0 \ \forall j \neq i \}.$$ 

Let $B_i^\delta$ be a finite subset of the above set such that $\beta$ is uniformly spread with stepsize $\delta$ on each component, as discussed earlier, i.e.,

$$B_i^\delta = \{ \beta : \sum_{j=1}^{n} \beta_j = 1, \beta_j = k\delta \text{ for some } k \in \{0,1,2,\ldots\} \text{ and } \beta_j \geq 0 \ \forall j \neq i \}$$

or

$$B_i^\delta = \{ \beta^{(i)} : \sum_{j=1}^{n} \beta_j = 1, \beta_j = k\delta \text{ for some } k \in \{0,1,2,\ldots\} \text{ and } \beta_j \geq 0 \ \forall j \neq i \},$$

where $\beta^{(i)}$ denotes that the $i^{th}$ component of $\beta$ is 0.

Now the following loops are executed:
For $i = 1 : n$

\[
\begin{align*}
\{ \\
\text{Find the boundary of } CHIM_+ \text{ near the } i^{th} \text{ edge of } CHIM \text{ by solving} \\
\max_{x,t_i} t_i \\
\text{s.t. } \Phi \beta + t_i \hat{n}_i = F(x) \quad \cdots \text{(NBI bdry)} \beta \\
x \in C \\
\text{for all } \beta \in B_i^\delta, \text{ where } \hat{n}_i = M_i e. \\
\text{and the } j^{th} \text{ column of the } n \times n - 1 \text{ matrix } M_i \text{ is given by } F(x_j) - F(x_i). \\
\text{Let } t_i(\beta) \text{ be the solution of (NBI bdry). Then,}
\end{align*}
\]

For each $\beta \in B_i^\delta$

\[
\begin{align*}
\{ \\
\text{if } t_i(\beta) > 0 \\
\text{Let } \tau_i(\beta) = \frac{t_i(\beta)}{p} \text{ be the uniform stepsize on } [0, t_i(\beta)], p \text{ is an integer.}
\end{align*}
\]

For $j = 2 : p - 1$

\[
\begin{align*}
\{ \\
\text{Set } t_i = j\tau_i(\beta) \\
\text{Solve NBI}_{\beta,t_i}, \text{i.e. solve} \\
\max_{x,t} t \\
\text{s.t. } \Phi \beta + t_i \hat{n}_i + t \hat{n} = F(x) \\
x \in C
\end{align*}
\]

\}
\}
Clearly the computational cost involved in obtaining these peripheral Pareto points is likely to be high, and we are fortunate that in practical situations it is usually sufficient to run NBI subproblems corresponding to points on the *CHIM*, since this usually captures the interesting 'middle' part of the Pareto set.
Chapter 5

Implementing NBI

This chapter discusses several details of implementing NBI and reports computational results on a small biobjective problem using a Matlab 4.2 implementation. Computational results for applying NBI on multicriteria structural optimization problems are discussed in details in Chapters 9 and 10.

5.1 Structure of the payoff matrix $\Phi$

The $i^{th}$ column of $\Phi$ is described by

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

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 $\Phi$ signifies not only that $x_k^*$ is not the global minimizer of $f_k(x)$, but $f_k(x_j^*) < f_k(x_k^*)$, and so $x_j^*$ improves on the current local minimum of $f_k(x)$. This provides a better starting point $x_j^*$ for minimizing $f_k(x)$ and hence will lead to a better local minimum for $f_k^*$.

5.2 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 instead 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 the sum of the columns of $\Phi$, 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 \( \beta \) is completely independent of the scales of the objective functions. In other words, if \( NBI_\beta \) is re-solved with the objective functions rescaled by arbitrary factors, the NBI point found remains unchanged. This comforting fact will be proved later.

Given that \( \Phi \) has nonnegative components as discussed in the previous subsection, it is clear that all components of \( \Phi e \) are nonnegative.

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.

### 5.3 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 \( \beta \) and \( \bar{\beta} \) such that \( \beta \) 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 \( NBI_\beta \) and the solution \( (\bar{x}^*, \bar{t}^*) \) of \( NBI_{\bar{\beta}} \) are 'close to each other'. Assume that we have solved \( NBI_\beta \) first and already have the point \( (\bar{x}^*, \bar{t}^*) \). Then with \( (\bar{x}^*, \bar{t}^*) \) as the starting point for solving \( NBI_{\bar{\beta}} \), 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.

Let us illustrate the above strategy for a biobjective problem. The weights \( \beta \) for only two objectives can be expressed as \( [\beta, 1 - \beta], \beta \in [0, 1] \). We can take \( \beta \) to assume the values:

\[ [0, \delta, 2\delta, \ldots, k\delta], \]
where $\delta < 1$ is the (uniform) spacing between two consecutive $\beta_1$ values and $k = I[\frac{1}{\delta}]$. i.e. the greatest integer $\leq \frac{1}{\delta}$. Then the set of 'uniformly distributed' weights is given by $[\beta, 1 - \beta]$, where $\beta$ ranges over the values as above.

Now assuming $\delta \ll 1$ (say $\delta = 0.05$), the minimizer of $f_2(x)$, i.e., $x^*_2$, is expected to be a small perturbation of the solution to the NBI subproblem with $\beta = [\delta, 1 - \delta]$. Thus the NBI subproblem with this $\beta$ is solved starting from $x^*_2$, and its solution is used as the starting point for solving the NBI subproblem with $\beta = [2\delta, 1 - 2\delta]$, and so on, until the last weight is reached.

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.

### 5.4 Generating $\beta$ 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 $\beta$ and ordering the subproblems in a manner amenable not only to efficient solution but also to parallelization.

#### 5.4.1 Generating $\beta$

Let us assume that for an $n$-objective problem, $\delta_j > 0$ is the uniform spacing between two consecutive $\beta_j$ values (i.e., the 'stepsize' on the $j^{th}$ component of $\beta$) 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 $\beta_1$ are

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

Given a particular value of $\beta_1$, define $m_1 = \frac{\beta_1}{\delta_1}$. Then the possible values of $\beta_2$ corresponding to that value of $\beta_1$, (i.e. $\beta_1 = m_1 \delta_1$) (all the $\beta_i$'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 \( \beta_3 \) 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 \( k_2 = I[1 - \frac{\delta_1}{\delta_2} - \frac{\beta_2}{\delta_2}] = I[1 - m_1 \delta_1 - m_2 \delta_2] \).

Thus, corresponding to \( \beta_i = m_i \delta_i, \; i = 1, \ldots, j - 1 \), the possible values of \( \beta_j \) for \( j = 2, \ldots, n - 1 \) are

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

where

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

Finally the last component of \( \beta \) 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 \( \beta \) and each path from the root to the leaf represents a possible \( \beta \) vector. However, a tree structure is unnecessary for implementation; all that requires storage are the numbers \( \delta_j \). 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 \( \beta_n \) may not be uniform.

### 5.4.2 Special case: Equal stepsizes on all \( \beta_i \)

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 \( \beta_1 \) are

\[
[0, \delta, 2\delta, \ldots, 1].
\]
Then the possible values of $\beta_j$ 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 $\beta_n$ values are uniformly spaced.

Fig. 5.1 shows part of the tree of $\beta$ values for $n = 4$ and $\delta = 0.2$.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{beta_tree.png}
\caption{Generating $\beta$ for $n = 4$, $\delta = 0.2$}
\end{figure}

**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

$$\binom{n + p - 1}{p}.$$ 

Table 5.1 tabulates this quantity for some values of $n$ and $\delta$.

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.
<table>
<thead>
<tr>
<th>Spacing(δ):</th>
<th>1/5</th>
<th>1/8</th>
<th>1/10</th>
<th>1/16</th>
<th>1/20</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of obj(n)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>9</td>
<td>11</td>
<td>17</td>
<td>21</td>
</tr>
<tr>
<td>3</td>
<td>21</td>
<td>45</td>
<td>66</td>
<td>153</td>
<td>231</td>
</tr>
<tr>
<td>4</td>
<td>56</td>
<td>165</td>
<td>286</td>
<td>969</td>
<td>1,771</td>
</tr>
<tr>
<td>5</td>
<td>126</td>
<td>495</td>
<td>1,001</td>
<td>4,845</td>
<td>10,626</td>
</tr>
<tr>
<td>6</td>
<td>252</td>
<td>1,287</td>
<td>3,003</td>
<td>20,349</td>
<td>53,130</td>
</tr>
<tr>
<td>7</td>
<td>462</td>
<td>3,003</td>
<td>8,008</td>
<td>74,613</td>
<td>230,230</td>
</tr>
<tr>
<td>8</td>
<td>792</td>
<td>6,435</td>
<td>19,448</td>
<td>245,157</td>
<td>888,030</td>
</tr>
</tbody>
</table>

Table 5.1 Number of NBI subproblems for different n and δ

5.4.3 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 β = e_n and is already solved) and solve the next one in the β_{n-1} generation (which is β_{n-1} = δ_{n-1}, β_n = 1 - δ_{n-1}), then the next one in the β_{n-1} generation (β_{n-1} = 2δ_{n-1}, β_n = 1 - 2δ_{n-1}), and so on until all the subproblems for β_i = 0, i = 1, ..., n - 2 have been solved. Then we move to the next node in the β_{n-2} generation (i.e., with β_i = 0, i = 1, ..., n - 3, β_{n-2} = δ_{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 β_{n-2} = δ_{n-2}, β_{n-1} = δ_{n-1}, β_n = 1 - δ_{n-2} - δ_{n-1} can be solved immediately after solving the first subproblem in the first node with β_{n-2} = 0, β_{n-1} = δ_{n-1}, β_n = 1 - δ_{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.

### 5.5 A Numerical Example

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

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

NBI using the quasi-normal was run on the above problem with the evenly spread values of \( \beta \) with \( \delta = 0.05 \). The Pareto points thus obtained are listed in Table 5.2 and plotted in Fig. 5.2.

![Figure 5.2 Pareto optimal vectors in the objective space using NBI and the method of convex combinations respectively](image)

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

**Table 5.2** Nondominated vectors obtained using NBI
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. 5.2].

With $f_1$ scaled by 5, the point found earlier six times 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. 5.3].

The nondominated vectors obtained using linear combinations are shown in Table 5.3.

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.

---

**Figure 5.3** 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
Table 5.3 Table of non-dominated vectors obtained by minimizing convex sums of objectives

5.6 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 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

\[ \| F(x_j^*) - F^* \|, \quad j = 1, \ldots, n \]
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 biobjective problem, \( \Phi \) is antidiagonal; thus a scaling that would achieve the above is obvious:

\[
\begin{align*}
f_1 & \leftarrow \frac{f_1}{f_1(x_2^*)} \\
f_2 & \leftarrow \frac{f_2}{f_2(x_1^*)}
\end{align*}
\]

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.

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

will be minimized (\( D = \text{diag}(d) \), \( d \) 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 \( \Phi \).

The mean square distance of the vertices is defined as

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

The variance quantity to be minimized is given by

\[
V(d) = \sum_{j=1}^{n} (v_j - \bar{v})^2;
\]
i.e.,

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

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

\[ V(d) = \sum_{j=1}^{n} \left( \sum_{i=1}^{n} d_i a_{i,j} \right)^2; \]

i.e.,

\[ V(d) = d^T A A^T d = \| A^T d \|^2. \]

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 \( \tau \) (say, 10). Then we would enforce

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

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

\[ \min_{d} V(d) = d^T A A^T d \]

\[ s.t. \quad \sum_{i=1}^{n} d_i \left( \sum_{j=1}^{n} \phi_{i,j}^2 \right) = n \tau \]

\[ d_i \geq 10^{-8}, \quad i = 1, \ldots, n. \]

Thus we can see how the matrix \( \Phi \) 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.
Chapter 6

Theoretical Results for NBI and its Relationship with Other Methods

This chapter focuses on how the NBI parametrization is related to other standard parametrizations of the Pareto set introduced by goal programming and the method of convex combinations. It is also proved here that NBI points are invariant with respect to the relative scales of the objective functions, which is one of the biggest strengths of NBI.

6.1 NBI as a Special Case of 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 nonnegative.

This is similar to goal programming. If we take the Pareto surface to be convex in the objective space, ‘equality goal programming'\(^4\) can be thought of as NBI where the direction \( \hat{n} \) is the negative of one of the canonical basis vectors \( e_i \), i.e. a vector with 1 in the \( f^{th} \) position and 0 in the rest). In other words, the subproblem \( NBI_\beta \) with \( \hat{n} = -e_i \) has the same solution as the following goal programming problem:

\[
\begin{align*}
\min \limits_x & \quad f_i(x) \\
\text{s.t.} & \quad f_j(x) = (\Phi \beta)(j), \quad j = 1, \ldots, n, \ j \neq i \\
& \quad x \in C,
\end{align*}
\]

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

The proof of the above is trivial. Since only the \( i^{th} \) component of \( \hat{n} \) is nonzero, the NBI constraints \( \Phi \beta + t \hat{n} = F(x) \) allow a decrease only in the \( i^{th} \) component of

\(^4\)Referring to goal programming where the goal constraints are equalities instead of inequalities.
\( F(x), \) i.e. in \( f_i(x), \) while the rest of the objectives are constrained to be equal to the corresponding components of \( \Phi \beta. \)

Though posing the goals as equalities is untraditional, a related procedure for obtaining a Pareto optimal point is discussed in Lin[23] and [24].

In Section 6.3 NBI will be related to the traditional goal programming problem using Lagrange-Multiplier theory without assuming that the Pareto surface is convex.

6.2 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 \( \beta \) 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 \( \beta \) such that \( NBI_{\beta} \) 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 \( w \in (\mathbb{R}_+ \cup \{0\})^n \), \( \sum_i w_i = 1 \), denote a positive, convex weighting of the objectives. The weighted linear combination problem for obtaining a Pareto optimal point is then written as

\[
\min_x w^T F(x) \\
\text{s.t.} \quad \bar{h}(x) = 0. \tag{6.1}
\]

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 \( (x^*, \lambda^*) \) for problem (6.1) is

\[
\nabla_x F(x^*) w + \nabla_x \bar{h}(x^*) \lambda^* = 0. \tag{6.2}
\]

Similarly, if \( \beta \) denotes the vector of parameters in \( NBI_{\beta} \), the NBI subproblem can be written as

\[
\min_{x, t} -t
\]
\[ \begin{align*}
\text{s.t.} \quad F(x) - \Phi \beta - t\hat{n} &= 0 \\
\hat{h}(x) &= 0.
\end{align*} \tag{6.3} \]

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

\[
\begin{align*}
\nabla_x F(x^*)\lambda^{(1)*} + \nabla_x \hat{h}(x^*)\lambda^{(2)*} &= 0 \\
-1 + \hat{n}^T \lambda^{(1)*} &= 0,
\end{align*} \tag{6.4}
\]

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

**Theorem 6.1**

Let \((x^*, t^*, \lambda^{(1)*}, \lambda^{(2)*})\) be the solution of \(NBI_\beta\) and suppose that \(\sum_1^n \lambda^{(1)*}_i \neq 0\).

Now define the components of the vector \(w\) as

\[
w_i = \frac{\lambda^{(1)*}_i}{\sum_1^n \lambda^{(1)*}_i}.
\]

Then, problem (6.1) with the above convex weighting vector \(w\) has the solution

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

**Proof** Dividing both sides of (6.4) by the scalar \(\sum_1^n \lambda^{(1)*}_i\) and observing that \(\hat{h}(x^*) = 0\), the equivalence between (6.2) and (6.4) becomes obvious.

However, quite clearly, if for some \(i\), the sign of \(\lambda^{(1)*}_i\) is opposite to that of \(\sum_1^n \lambda^{(1)*}_i\), then the vector \(w\) has a negative component and does not qualify as a weight for problem (6.1). 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 \(NBI_\beta\), one can also obtain the NBI point corresponding to a given solution of problem \(LC_w\) with very little effort.
Theorem 6.2

Assume \( (x^*, \lambda^*) \) 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 \( (x^*, \lambda^*) \) corresponds to the solution of \( NBI_{\beta} \) 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 (6.2) 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, \( \bar{\lambda}(x^*) = 0 \) and \( \Phi \beta + t \hat{n} = F(x^*) \), the required equivalence between \( LC_w \) and \( NBI_{\bar{\beta}} \) follows.

\( \square \)

6.3 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 kind of multiplier-based approach used to relate \( NBI_{\beta} \) to \( LC_w \).

Theorem 6.3

Assume \( (x^*, t^*, \lambda^{(1)*}, \lambda^{(2)*}) \) is the solution of \( NBI_{\beta} \). Assume 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:
\[
\min f_k(x)
\]
\[
s.t. \quad f_i(x) \leq \gamma_i, \forall i \neq k
\]
\[
\bar{h}(x) = 0
\]

with goals \(\gamma_i\) given by
\[
\gamma_i = \begin{cases} 
  f_i(x^*), & \text{if } \lambda_i^{(1)\ast} \neq 0 \\
  \text{any finite number } \geq f_i(x^*), & \text{if } \lambda_i^{(1)\ast} = 0
\end{cases}
\]

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

**Proof**

Since \((x^*, t^*, \lambda^{(1)\ast}, \lambda^{(2)\ast})\) solve the \(NBI_{\beta}\) subproblem, they must satisfy (6.4). Given that \(\lambda_k^{(1)\ast} \neq 0\), we can divide both sides of (6.4) and get

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

Now \(\frac{\lambda_i^{(1)\ast}}{\lambda_k^{(1)\ast}} \geq 0\) because \(\lambda_i^{(1)\ast}\) and \(\lambda_k^{(1)\ast}\) are of the same sign. Then with \(\frac{\lambda_i^{(1)\ast}}{\lambda_k^{(1)\ast}}\) as the multipliers of the \(n - 1\) inequality constraints in (6.5), the goals \(\gamma_i\) satisfy complementarity by definition, since

\[
\gamma_i = f_i(x^*) \text{ whenever } \lambda_i^{(1)\ast} \neq 0
\]

\[
\Rightarrow \frac{\lambda_i^{(1)\ast}}{\lambda_k^{(1)\ast}}(f_i(x^*) - \gamma_i) = 0 \quad \forall i \neq k.
\]

Moreover, since \(x^*\) is clearly feasible for (6.5), \((x^*, \lambda_i^{(1)\ast}, \lambda_k^{(1)\ast}, \lambda_i^{(2)\ast})\) solves (satisfies first order necessary conditions for minimizer for) problem (6.5).

\[
\square
\]

### 6.4 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 \(\beta\) 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 $S = \text{diag}(s)$, then

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

Consequently

$$\nabla_x F(x) \leftarrow \nabla_x F(x) S,$$

$$\Phi = SS\Phi.$$

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

**Theorem 6.4**

If $(x^*, t^*, \lambda^{(1)*}, \lambda^{(2)*})$ is a stationary point of the unscaled $NBI_\beta$ (i.e. with $S = I_n$), then $(x^*, t^*, S^{-1}\lambda^{(1)*}, \lambda^{(2)*})$ is a stationary point of $NBI_\beta$ with the $i^{th}$ function scaled by $s_i$ as described above.

**Proof** Since $(x^*, t^*, \lambda^{(1)*}, \lambda^{(2)*})$ solves the unscaled $NBI_\beta$ (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:

$$(\nabla_x F(x^*)S)(S^{-1}\lambda^{(1)*}) + \nabla_x \bar{h}(x^*)\lambda^{(2)*} = 0.$$ \hspace{1cm} (6.6)

The second equation implies

$$e^T\Phi^T\lambda^{(1)*} = 1$$

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

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

$$(e^T(S\Phi)^T)(S^{-1}\lambda^{(1)*}) = 1.$$ \hspace{1cm} (6.7)
The third equation can be rewritten as
\[
\Phi \beta + \Phi e = F(x^*)
\]
\[
\equiv S\Phi \beta + \Phi S e = SF(x^*).
\]
(6.8)

Clearly, equations (6.6),(6.7) and (6.8) imply that \((x^*, t^*, S^{-1} \lambda^{(1)*}, \lambda^{(2)*})\) solves \(NBI_3\) with the functions scaled by \(S\).

The above result does not depend on \(e\) being the vector of all ones. The only requirement is that \(e\) should have positive components since the quasi-normal needs negative components for NBI to be meaningful. In particular the result holds if \(\hat{n}\) is scaled by a factor, say, a normalization constant. In our implementation \(\hat{n}\) is always normalized to have unit \(l_2\) norm.

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.

### 6.5 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 is provided in Chapter 8.

However, the weights in the linear combinations approach are often 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 [31], 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 biobjective 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 unnecessary.

  - *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 methods must 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 trade-off standpoint, since multilevel techniques usually can only improve only a few of the ‘most important’ objectives, leaving no compromise for the rest.
Chapter 7

A Characterization of 'Good' Pareto Optimal Points

This chapter is aimed at suggesting a technique for choosing the final point given a finite set of design points that are all Pareto optimal. The need for such a technique is motivated by the fact that easy visualization of the entire Pareto surface at once is not possible if the dimension of the objective space exceeds 3 in which case the designer may have no choice other than go through the list of Pareto optimal points in order to make his/her final choice. In such a case lower dimensional subplots for 2 or 3 element subsets of the n objectives fail to be of any help as will be shown in Chapter 9.

7.1 Efficiency of order k

The chief instrument for this technique will be a new concept which we have named efficiency of order k and defined as below:

Definition:
Let \( \mathcal{P} \) be the discrete set of Pareto optimal points at hand and \( \mathcal{F}_D = \{ F(x^*) : x^* \in \mathcal{P} \} \) the corresponding set of nondominated objective vectors.

Consider all possible \( k \)-element subsets of the \( n \) given criteria \( 1 \leq k \leq n \). A point \( x^* \in \mathcal{P} \) is defined as efficient of order \( k \) if \( F(x^*) \) is not dominated by any member of \( \mathcal{F}_D \) for any of the \( k \)-element subsets of objectives.

Theorem 7.1
If \( x^* \) is efficient of order \( k \), then it is efficient of order \( k + 1 \).

Proof
The proof is by contrapositive. If \( x^* \) is not efficient of order \( k + 1 \), then there exists a \( k + 1 \) dimensional vector objective with components \( f_{i_1}(x), f_{i_2}(x), \ldots, f_{i_{k+1}}(x) \) such that \( \exists \) a feasible \( \bar{x} \) with the property that \( f_{i_p}(\bar{x}) \leq f_{i_p}(x^*) \), for all \( 1 \leq p \leq k + 1 \), strict inequality holding for one of the components. Suppose, with no loss of generality, that this component is \( f_{i_1} \), i.e. \( f_{i_1}(x^*) < f_{i_1}(\bar{x}) \). Then \( x^* \) is not Pareto optimal for the \( k \) dimensional objective with components \( f_{i_1}(x), f_{i_2}(x), \ldots, f_{i_k}(x) \), since there exists
\[ \bar{x} \text{ such that } f_{i_p}(\bar{x}) \leq f_{i_p}(x^*), \text{ for all } 1 \leq p \leq k \text{ and } f_{i_1}(x^*) < f_{i_1}(\bar{x}). \text{ But then a } k \text{ element subset of the } n \text{ objectives has been found for which } x^* \text{ is not Pareto optimal. hence } x^* \text{ is not efficient of order } k. \]

\[ \square \]

Corollary:

It obviously follows by induction that if \( x^* \) is efficient of order \( k \), then it is efficient of order \( j \) for every \( j \in \{k + 1, \ldots, n\} \).

Significance:

It is clear that efficiency of order \( n \) is the ordinary concept of Pareto optimality. On the other extreme, a point that is efficient of order 1 must be the shadow minimum. So if the shadow minimum is not attainable, which is most often the case, there does not exist any point that is efficient of order 1. On the other hand, there usually are many points that are Pareto optimal, and it is in general unclear how to choose among them. The concept of efficiency of order \( k \) provides something intermediate, something stronger than Pareto optimality but weaker than the shadow minimum. Thus one way to choose the final Pareto optimal point would be to choose the final point as one that has the lowest order efficiency. This provides a way of retaining Pareto optimal points satisfying the strongest requirements and eliminating ones which are inferior. Efficiency of order \( k \) is our suggestion to provide a rigidly-defined means of branding some Pareto optimal points as being perhaps superior or more desirable than others, applicable irrespective of the particulars of the problem at hand.

7.1.1 An Example

As an example consider the set of nondominated vectors in \( \mathbb{R}^3 \) shown in the table below:
One can begin by eliminating points that are not efficient of order 2. Points 1, 2, 4 and 5 can be ruled out based on the following observations:

- Point 1 is dominated by point 2 for the \((f_2, f_3)\) subset
- Point 2 is dominated by point 3 for the \((f_1, f_2)\) subset
- Point 4 is dominated by points 2 and 3 for the \((f_2, f_3)\) subset
- Point 5 is dominated by point 3 for the \((f_1, f_3)\) subset

This leaves point 3 as the only possible candidate satisfying efficiency of order 2. A methodical comparison shows that point 3 is in fact nondominated for each of the three 2-element subsets and hence is efficient of order 2. However it is not efficient of order 1 since its components do not contain the minimum values for all the objective functions.

Another example using NBI points from a four-objective problem arising in truss design is given at the end of Chapter 9.
Chapter 8

Elaborations on the Shortcomings of Minimizing Convex Combinations of Objectives

It has been mentioned earlier that the popular method for parametrically minimizing convex combinations of the objectives suffers from two major drawbacks: its failure in obtaining nonconvex parts of the Pareto surface and its even more undesirable incapability in yielding an even spread of Pareto points given an even spread of parameter settings. This chapter attempts to shed light from a geometrical standpoint on why these happen.

8.1 Failure in Capturing Nonconvex Parts of the Pareto Curve

8.1.1 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 \( 1 - \alpha \) and \( \alpha \) where \( \alpha \) is a scalar \( \in [0,1] \), problem \((LC)\) becomes

\[
\min_x (1 - \alpha)f_1(x) + \alpha f_2(x) \\
s.t. \quad x \in C \quad \ldots(LC_2).
\]

An equivalent formulation of the above is

\[
\min \quad \frac{\cos \theta}{\cos \theta + \sin \theta} f_1(x) + \frac{\sin \theta}{\cos \theta + \sin \theta} f_2(x) \\
s.t. \quad x \in C \quad \ldots(TLC_\theta),
\]

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

For a given value of \( \theta \), problem \((LC_2)\) with

\[
\alpha = \frac{\sin \theta}{\cos \theta + \sin \theta}
\]
yields exactly the same solution as problem \((TLC_\theta)\). Thus \((LC_2)\) and \((TLC_\theta)\) 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 \((TLC_\theta)\) for some \(\theta \in [0, \frac{\pi}{2}]\). It follows that if a Pareto point is not a solution to problem \((TLC_\theta)\) for all \(\theta \in [0, \frac{\pi}{2}]\), then it cannot be obtained by minimizing any convex combination of the two objectives.

### 8.1.2 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. 8.1). Let us label the rotated axes as \(\tilde{f}_1, \tilde{f}_2\). Then an elementary coordinate transformation gives

\[
\begin{bmatrix}
\tilde{f}_1 \\
\tilde{f}_2
\end{bmatrix} = \begin{bmatrix}
\cos \theta & \sin \theta \\
-\sin \theta & \cos \theta
\end{bmatrix} \begin{bmatrix}
f_1 \\
f_2
\end{bmatrix}
\]

so that

\[
\tilde{f}_1 = f_1 \cos \theta + f_2 \sin \theta.
\]

Thus \(\frac{\tilde{f}_1(x)}{\cos(\theta) + \sin(\theta)}\) is what problem \((TLC_\theta)\) aims to minimize over \(x \in C\), which has the same solution as \(\min_{x \in C} \tilde{f}_1(x)\). Geometrically it can be seen in Fig. 8.1 how minimizing \(\tilde{f}_1\) gets us the Pareto point \(P\) (i.e., we need a feasible point corresponding to which the value of the \(\tilde{f}_1\) coordinate is minimum). One can imagine this minimization process as translating the \(\tilde{f}_2\) axis parallel to itself until it hits the curve. The point where it hits the curve is a Pareto point.

Solving problem \((TLC_\theta)\) 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. 8.2 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 \((TLC_\theta)\) for some \(\theta\) 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.
Figure 8.1 Getting a Pareto point by solving the Trigonometric Linear Combinations problem

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, $\theta$ and hence the $(TLC_\theta)$ problem get uniquely defined (since $\theta = \tan^{-1}(\text{slope}) - \frac{\pi}{2}$, as shown in Fig. 8.1). 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 $(TLC_\theta)$, in which case the point in question is not a (global) minimum of problem $(TLC_\theta)$. This is illustrated for point P in Fig. 8.3. This point cannot solve $(TLC_\theta)$ for any other $\theta$ since, given the slope at the point, $\theta$ has been uniquely defined. Hence the justification follows.
8.1.3 **Inexistence of a \((TLC_\theta)\) 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 \((TLC_\theta)\) subproblem which it solves.

The first uses the aid of Fig. 8.4. Let us recall that solving \((TLC_\theta)\) 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. 8.4 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 of the Pareto curve. Thus there does not exist a
Figure 8.3 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 \((TLC_\theta)\) objective. However points R or M can be obtained as a solution of a \((TLC_\theta)\) problem.

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 least at one point where it does not match the slope of the boundary curve. Thus it cannot solve a \((TLC_\theta)\) subproblem for any \(\theta \in [0, \frac{\pi}{2}]\).

Finally, given the equivalence between problem \((TLC_\theta)\) 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.
8.2 Nonuniform Spread of Pareto Points using Uniform Spread of $\alpha$

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 gives an analytic explanation of the observation that even if the Pareto curve is convex, a uniform spread of $\alpha$ does not guarantee a uniform spread of 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 $\alpha$ are actually clumped in certain regions of the Pareto set, providing the user no information about the nature of trade-off between the two objectives elsewhere.

As an example consider the bicriteria problem in Chapter 5, restated below for convenience:

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

As in Chapter 5, the set of Pareto points obtained by minimizing convex combinations of the two objectives for 21 different weights with \( \alpha \) varying from 0 to 1 in even increments of 0.05 are shown in Fig. 8.5, followed by the corresponding set of Pareto points produced by Normal-Boundary Intersection (NBI) using an even spread of NBI parameters (also available in Das and Dennis [5]).

![Efficient points obtained by minimizing convex combinations of objectives](image1)
![Efficient points obtained by minimizing convex combinations of objectives](image2)

**Figure 8.5** Outputs of minimizing linear combinations and NBI respectively on a biobjective optimization problem using an even spread of parameters

8.2.1 Relationship between \( \alpha \) and the slope of the Pareto curve

An explicit relationship between the slope of the Pareto curve at a Pareto point and the \( \alpha \) for which this point solves \( (LC_2) \) can be derived based on the development in earlier sections. As shown in Fig. 8.1, the slope of the curve at the Pareto point, denoted by \( \frac{\partial f_2}{\partial f_1} \), is related to the parameter \( \theta \) in the \( (TLC_\theta) \) 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 \( \alpha \) and \( \theta \),
\[ \alpha = \frac{\sin \theta}{\sin \theta + \cos \theta} \]
\[ = \frac{1}{1 + \cot \theta} \]
so,
\[ \alpha = \frac{1}{1 - \frac{\partial f_2}{\partial f_1}}. \] (8.1)

Thus an even spread of \( \alpha \) 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. 8.6). If the minimum and maximum of \( f_1 \) values on the Pareto set are \( f_1^* \) and \( \bar{f}_1 \) respectively, then the same for \( f_2 \) are \( \psi(\bar{f}_1) \) and \( \psi(f_1^*) \). As shown in Fig. 8.6, 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 \( \alpha \). Then using the fact that the projections of the arcs between two consecutive Pareto points on the \( f_1 \) axis are all equal, \( \alpha \) 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 (8.1) in the above yields
\[ \frac{1}{1 - \frac{\partial f_2}{\partial f_1}} = \frac{f_1 - f_1^*}{\bar{f}_1 - f_1^*}. \]
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 $\alpha$.

8.2.2 Distribution of $\alpha$ for uniform spreads of Pareto points

Given that an even spread of Pareto points would correspond to an even spread of $\alpha$ for only specific Pareto curves, we shall try to find out the distribution of $\alpha$ 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 requires
the arcs joining consecutive Pareto points to have projections of equal lengths on
the line segment AB joining the two extreme points of the Pareto curve. 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. 8.6). Histogram
plots of the $\alpha$ values corresponding to an even spread of points on the Pareto curve
are shown for some assumed shapes of the Pareto curve in figs. 8.7, 8.8, 8.9 and 8.10.

![Distribution of $\alpha$ for even spread on $\psi(f_1) = \frac{1}{f_1}$ for $f_1 \in [0.1, 2]$](image)

**Figure 8.7** Distribution of $\alpha$ for even spread on $\psi(f_1) = \frac{1}{f_1}$ for $f_1 \in [0.1, 2]$

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

### 8.2.3 Spread of $\alpha$ 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 $\alpha$ 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. 8.5. The model fit to the set of points $(f_1, f_2)$
was the following sum of three exponentials

$$
\hat{f}_2 = \hat{\psi}(f_1) = p_1 + p_2 e^{-p_3 f_1} + p_4 e^{-p_5 f_1} + p_6 e^{-p_7 f_1}.
$$
Figure 8.8  Distribution of $\alpha$ for even spread on $\psi(f_1) = 2e^{-2f_1}$ for $f_1 \in [1, 2.5]$

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

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

Estimating $\frac{\partial f_2}{\partial f_1}$ by $\hat{\psi}'(f_1)$ at the Pareto points obtained using NBI, a set of predicted $\alpha$ values were obtained using (8.1). Minimizing convex combinations of the two objectives for these values of $\alpha$ now yields the spread shown in Fig. 8.11, which is a big improvement on the convex combinations spread in Fig. 8.5.

The distribution for these $\alpha$ values is shown in a histogram in Fig. 8.12.

Thus the above clarifies in detail why the method of convex combinations suffers from its drawbacks and is best avoided as a means of generating the Pareto set.
Figure 8.9  Distribution of $\alpha$ for even spread
on $\psi(f_1) = 2e^{-\frac{(f_1-1)^2}{8}}$ for $f_1 \in [2,8]$

Figure 8.10  Distribution of $\alpha$ for even spread on $\psi(f_1) = \frac{1}{f_1}$ for $f_1 \in [2,3]$
Figure 8.11 Pareto points obtained using values of $\alpha$ predicted using nonlinear fit on NBI points
Figure 8.12 Distribution of values of $\alpha$ predicted using nonlinear fit on NBI points
Chapter 9

A Truss Optimization Problem

Now we shall describe an application which, though small in terms of variable size, illustrates all of the aspects of NBI. A larger dimensional problem involving a frame structure will be considered in the next chapter. The problem described here arises in structural optimization and a version of it has been studied in Koski [20]. The problem involves optimizing a pin-jointed truss structure (so that no flexural deformations need to be considered) as shown in Fig. 9.1.

![Truss Structure Diagram](image)

**Figure 9.1** A truss structure under a suspended load and a wind load

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 $\theta$ and $\alpha$ clearly depend on the chosen location $x$. Other optimization variables are the cross-sectional areas of the bars, denoted respectively by $a_1, a_2, a_3$, which are allowed to vary between 0.8in$^2$ and 3.0in$^2$. Let $u_1$ and $u_2$ denote respectively the horizontal and vertical displacements of the node $P$, $\delta_1, \delta_2$ and $\delta_3$ the elongations of the three bars respectively, and $E$ the modulus of elasticity of the material of the bars.
9.0.4 Objectives

The possible objectives to be minimized are given below, two or more element subsets of these objectives will be considered at a time for trade-off studies:

- Total displacement of node $P$ (squared for differentiability everywhere), i.e.
  \[ d = u_1^2 + u_2^2 \]

- Total volume of the structure, i.e.
  \[ V = a_1 \frac{L}{\sin \theta} + a_2 L + a_3 \frac{L}{\sin \alpha} \]

- Absolute value of stress in the left bar,
  \[ |s_1| = \left| E \frac{\delta_1}{L \sin \theta} \right| = \frac{E}{L} |\delta_1| \sin \theta \]

- Absolute stress in the middle bar,
  \[ |s_2| = \frac{E}{L} |\delta_2| \]

- Absolute stress in the right bar,
  \[ |s_3| = \frac{E}{L} |\delta_3| \sin \alpha \]

A similar structure was considered by Koski in [20] but the location of the middle bar was fixed, so that $\theta$ and $\alpha$ were also fixed and thus minimizing $|s_i|$ was equivalent to minimizing $\delta_i$. Moreover the volume in his case was a linear function of his design variables, unlike in our formulation.

9.0.5 Constraints

The first set of constraints will express equilibrium under the applied load (forces). Let the wind load be $W_1$ and the suspended load be $W_2$ and let the respective bar forces be $y_1, y_2, y_3$ (assume all the bars are under tension). Then equilibrium can be expressed as

\[ y_1 \cos \theta - y_3 \cos \alpha = W_1 \]
\[ y_1 \sin \theta + y_2 + y_3 \sin \alpha = W_2 \]
or alternately,

\[ A^T y = W \tag{9.1} \]

where

\[ A = \begin{bmatrix} \cos \theta & \sin \theta \\ 0 & 1 \\ - \cos \alpha & \sin \alpha \end{bmatrix} \]

Assuming the bars are linearly elastic, the vector of bar forces \( y \) can then be related to the vector of displacements \( \delta \) as

\[ y = C \delta , \tag{9.2} \]

where \( C \) is a diagonal matrix with diagonal elements \( \frac{E}{L} a_1 \sin \theta, \frac{E}{L} a_2, \frac{E}{L} a_3 \sin \alpha \) respectively.

Neglecting geometric nonlinearities (i.e. assuming the change in length of each bar is ‘small’), standard linear analysis relates the changes in bar lengths to the displacement of the node \( P \) in the global coordinates as

\[ Au = \delta . \tag{9.3} \]

Substituting equation 9.3 in equation 9.2 we get

\[ y = C Au . \tag{9.4} \]

Substitution of equation 9.4 in equation 9.1 yields

\[ A^T CAu = W \]

or

\[ Ku = W , \tag{9.5} \]

where \( K = A^T CA \) is the structure stiffness matrix. Multiplying out \( A^T CA \), \( K \) can be written explicitly in terms of the design variables as

\[ K = \frac{E}{L} \begin{bmatrix} a_1 \sin \theta \cos^2 \theta + a_3 \sin \alpha \cos^2 \alpha & a_1 \sin^2 \theta \cos \theta - a_3 \sin^2 \alpha \cos \alpha \\ a_1 \sin^2 \theta \cos \theta - a_3 \sin^2 \alpha \cos \alpha & a_1 \sin^3 \theta + a_2 + a_3 \sin^2 \alpha \end{bmatrix} . \]

The first set of equality constraints are imposed using the equilibrium conditions in the form of equation 9.5. The second set of equations relate \( x, \theta \) and \( \alpha \) so as to ensure that the bars are actually connected at node \( P \):

\[ x = L \cot \theta \]
Further, an inequality constraint should be imposed on every stress component that is not included as an objective so that the absolute value of those stress components do not exceed a maximum stress \( s_{\text{max}} \).

The stress components \( s_i \) were earlier expressed as functions of \( \delta_i \). One can either use equation 9.3 to substitute \( \delta \) and reduce the number of variables in the problem for the price of added nonlinearity in the formulation, or can leave \( s_i \) in terms of \( \delta_i \) and impose equation 9.3 as a constraint and settle for more variables and constraints with less nonlinearity. Our choice will be the first alternative.

9.0.6 Parameter settings and bounds

The values of the various parameters in the problems were set as below:

- \( E = 29 \times 10^3 \) ksi (modulus of elasticity of a variety of steel)
- \( L = 60 \) ft
- \( D = 120 \) ft
- \( W_1 \) (wind load) = 100 kips
- \( W_2 \) (suspended load) = 1000 kips
- \( s_{\text{max}} = 550 \) ksi

As stated before, lower and upper bounds were imposed on the cross-sections (0.8in\(^2\) and 3.0in\(^2\)) and the location of the middle bar (0.25\(D \leq x \leq 0.75D\)).

9.1 Numerical results

Multiobjective trade-off studies were performed for various combinations of the five objectives mentioned earlier, associated numerical results are reported in the following sections:

9.1.1 Minimum displacement and minimum volume

The biobjective problem of minimizing the square of the displacement of \( P \) and minimizing the total volume of the structure was solved using NBI for 21 evenly
spaced values of $\beta$. Equations 9.5 and 9.6 were imposed as equality constraints. Inequality constraints were of the form

$$|s_i| \leq s_{\text{max}} \text{ or equivalently } s_i \leq s_{\text{max}} \text{ and } s_i \geq -s_{\text{max}}$$

each $s_i$ being a function of $\delta_i$ and hence $u$. Thus the optimization variables were $a_1, a_2, a_3, \theta, \alpha, x, u_1, u_2$, with bounds on $a_1, a_2, a_3, x$, which we shall call our design variables and the remaining $\theta, \alpha, u_1, u_2$ response or state variables.

Fig. 9.2 shows a plot of the Pareto optimal points obtained using NBI (finite difference gradients used all throughout). The NBI parameter $\beta$ was varied uniformly with a stepsize of 0.05 on each component. The two noticeable ‘gaps’ in the otherwise uniform spread of Pareto points correspond to NBI subproblems which did not converge in the stipulated maximum number of iterations, perhaps because the bounds on the cross-sections or the bar stresses were too stringent.

![Points obtained using NBI](image)

**Figure 9.2** NBI points for minimizing square of nodal displacement and total volume (cu. ft.)

The same problem was solved using the method of convex combinations, the convex combination vectors $w$ being assigned the same 21 uniformly spread values as that for the NBI parameter $\beta$. The obtained spread of Pareto points is shown in Fig. 9.3, which is much less uniform than that for NBI.
Figure 9.3  Pareto set for square of nodal displacement and total volume (cu. ft.) obtained by minimizing convex combinations

It is also interesting to study the variation in the optimization variables along the Pareto set. Two such figures (9.4 and 9.5) show the variation in the cross-sectional area of the left bar and that of the right bar respectively along the Pareto set. As expected, these cross-sections get smaller near the end of the Pareto curve where the total volume reaches a minimum. One interesting observation here is that $x$ and hence $\theta$ and $\alpha$ remain fixed along the Pareto set.

9.1.2 Minimum stress in right bar and minimum volume

In this version of the problem, the stress in the right bar was included as an objective and the constraint $|s_3| \leq s_{\text{max}}$ was dropped from the problem. (Normally it would be necessary to minimize the absolute value of the stress or the square of the stress, but the minimum value of the stress turned out to be positive, thus minimizing the stress was equivalent to minimizing the absolute stress). The Pareto set for minimizing the stress in the right bar obtained using NBI is shown in Fig. 9.6. Even though the sharp drop in values of the total volume in the left part of the Pareto set looks vertical, it
in fact has a very small slant and all the points thereon are Pareto optimal. However, the interesting region is clearly elsewhere.

One objective neglected in this biobjective problem is the displacement of the node. So the displacement of $P$ was plotted at each of the Pareto points to get an idea of which part of the Pareto set gives an acceptable nodal displacement. This plot is shown in Fig. 9.7 and shows that in the 'interesting' part of the Pareto set (i.e. where the total volume does not fall rapidly), the displacement itself rises fairly rapidly so that it might be unwise to move too far beyond the 'kink' of the Pareto curve in favor of minimizing total volume.

In contrast to NBI, minimizing convex combinations for 21 evenly spread values of $w$ yields Fig. 9.8 with points only in the 'uninteresting part'. The number of distinct Pareto points obtained here are 12.

Given the individual minima and minimizers of the objectives at the outset, the number of floating point operations (flops) required in solving the subproblems using NBI and convex combinations are shown in the table below:
Figure 9.5  Variation in the cross-sectional area of the right bar along the Pareto set for displacement and volume

<table>
<thead>
<tr>
<th></th>
<th>NBI</th>
<th>Convex Combinations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of distinct points</td>
<td>21</td>
<td>12</td>
</tr>
<tr>
<td>Number of flops</td>
<td>2,962,068</td>
<td>1,650,079</td>
</tr>
<tr>
<td>Flops per distinct point</td>
<td>141,050.86</td>
<td>137,506.58</td>
</tr>
</tbody>
</table>

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.

9.1.3 Minimum stress in left bar and minimum volume

A similar trade-off study between the stress in the left bar (with constraints on the absolute values of the stresses in the middle and right bars) and the total volume was performed. The set of Pareto points generated using 21 values of NBI is shown in Fig. 9.9, while that for convex combinations with 21 evenly spread values of $w$ is
shown in Fig. 9.10. The position of the middle bar varies from 90 ft. down to 30 ft. along the Pareto set in the manner shown in Fig. 9.11.

9.1.4 Minimizing stress in the left bar, total volume and stress in the right bar

The Pareto surface for minimizing the total volume and the stresses in the left and right bars was obtained using NBI with 66 evenly spread values of $\beta$ (stepsize on each component = 0.1). The only inequality constraint was now on the stress of the middle bar, i.e. $|s_2| \leq s_{max}$. Of the 66 subproblems, nine failed to converge owing to infeasibility, while the remaining 57 did and yielded the plot shown in Fig. 9.12. The whole process took about 11.4 million floating point operations. Relative to the discrete set of 57 vectors obtained every point was nondominated.

One interesting object in multiobjective problems with more than two objectives is a lower-dimensional subplot for a proper subset of the objectives. Figs. 9.13, 9.14 and 9.15 show biobjective subplots for the three biobjective subsets of this triobjective problem.
Figure 9.7 Variation in the nodal displacement along the Pareto set for total volume and stress in right bar

9.1.5 Compromising among nodal displacement, total volume and all the stresses

None of the Pareto surfaces obtained thus far attempt to compromise among all the five objectives, so this will be our focus from now on. Since it is beneficial for the purposes of plotting and interpretation to keep the number of objectives as small as possible, we shall first try to combine all the stresses into a single minimizable quantity. However, rather than form a linear combination of the three stresses and pose it as an objective, we shall choose our objective to be the maximum of the absolute value of the three stresses. It is quite difficult in interpret a linear combination of the three stresses in physical terms, whereas minimizing the maximum stress over the three bars is more tangible. After all, what the designer usually wants is that all the stresses lie below a certain maximum stress (which is how the problem had been posed at the onset), and it is this maximum stress which will now be introduced as a variable and minimized as an objective.

However, it was discovered that the minimizer of the total displacement is also the minimizer of the maximum absolute bar stress, hence we decided to drop displacement
as an objective and simply look at the trade-off curve for total volume and the maximum stress. This problem solved using NBI yielded the plot shown in Fig. 9.16, while that using convex combinations yielded Fig. 9.17. The Pareto curve obtained using convex combinations is a noticeable improvement over other Pareto curves obtained using convex combinations and 20 distinct and fairly uniformly spread Pareto points have been obtained given 21 weights, however the total number of flops required by convex combinations is about 44 million while that for NBI is only 9.6 million.

In order to observe the variation in nodal displacement, the nodal displacement for the Pareto points was plotted versus the maximum bar stress in Fig. 9.18 and versus the Pareto curve in Fig. 9.19. Clearly the nodal displacement increases monotonically with the maximum bar stress as is expected since a greater nodal displacement results in a larger maximum bar strain and hence a larger maximum stress. This justifies the dropping of nodal displacement as an objective all the more.

Finally, the four-objective problem of minimizing the total volume and the stresses in the three bars was solved using NBI for 165 evenly spread settings of \( \beta \) (stepsize on each component = 0.125). Of these 13 subproblems failed to converge. The discrete set of 152 objective vectors obtained from the remaining subproblems were
Figure 9.9  NBI points for minimizing stress in left bar and total volume (cu. ft.)

all nondominated and are shown in the table in Appendix B. The entire process took 20.7 million floating point operations.

Figs. 9.20 and 9.21 show the subplots for the three bar stresses and the total volume versus each of the stresses. It should be admitted without any pretense that it is quite difficult to actually use such plots as an aid in the design process, but such is the curse of an increased number of objectives in which case the Pareto set cannot be visualized easily. Hence, the concept of efficiency of order k was applied in dealing with this situation. A program developed in C++ for finding the efficiency order of vectors in a discrete set identified only 3 out of the 152 nondominated vectors as being efficient of order 3. None of the vectors were efficient of order 2. These three vectors, which are ‘best choices’ since they have the lowest order of efficiency, are tabulated below. The number in front of each vector refers to the cardinal number of the vector indicating the order in which it appears in a table in Appendix B.
Figure 9.10 Pareto points for minimizing stress in left bar and total volume (cu. ft.) using the method of convex combinations

<table>
<thead>
<tr>
<th>Net Vol. (cu. ft.)</th>
<th>Left bar stress(ksi)</th>
<th>Middle bar stress(ksi)</th>
<th>Right bar stress(ksi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>68.</td>
<td>2.9024</td>
<td>2.0583</td>
<td>3.2611</td>
</tr>
<tr>
<td>73.</td>
<td>3.941</td>
<td>1.591</td>
<td>3.0369</td>
</tr>
<tr>
<td>96.</td>
<td>3.246</td>
<td>1.7772</td>
<td>4.2073</td>
</tr>
</tbody>
</table>
Figure 9.11  Variation in the position of the middle bar along the Pareto set for total volume and stress in left bar

Figure 9.12  NBI points for minimizing stresses in left and right bars and total volume
Figure 9.13  Stress in left bar versus stress in right bar subplot for Pareto set for volume and stresses in left and right bars

Figure 9.14  Stress in left bar versus total volume subplot for Pareto set for volume and stresses in left and right bars
Figure 9.15  Total volume versus stress in right bar subplot for Pareto set for volume and stresses in left and right bars.

Figure 9.16  NBI points for minimizing square of nodal displacement and total volume (cu. ft.)
Figure 9.17  Pareto points for minimizing square of nodal displacement and total volume using convex combinations

Figure 9.18  Nodal displacement plotted for various points on the volume vs maximum bar stress Pareto set versus maximum bar stress
Figure 9.19 Nodal displacement plotted along the Pareto set for total volume and maximum bar stress

Figure 9.20 Triobjective subplot for the three bar stresses for the Pareto set for the three stresses and total volume
Figure 9.21 Biobjective subplot for each the three bar stresses vs volume for the four-objective Pareto set for three stresses and total volume
Chapter 10

A Frame Optimization Problem

In this chapter we will consider another structural optimization problem. The structure involved is a 2-D frame composed of vertical column elements and horizontal beam elements, analyzed in the framework of linear elasticity. The basic difference from the truss structure is that the beams and columns can offer resistance to bending moments and hence flexural deformations need to be considered while analyzing the structure.

![2-D frame under wind load](image)

**Figure 10.1** 2-D frame under wind load

Fig. 10.1 shows a frame with three supports and four storeys, which is the structure we are optimizing. The wind load imposed on the structure is also shown in Fig. 10.1. In general, such a frame structure with $p$ supports and $k$ storeys has $pk$ columns and $(p - 1)k$ beams. All the beams and columns are assumed to have rectangular cross-sections. Let $I_i$ denote the moment of inertia of the $i^{th}$ beam or column element about axis along the direction of the bending moment through one end. If $h_i$ denotes the length of the side of the rectangular cross-section parallel to this direction and $b_i$
denotes the other length of the other side, then

\[ I_i = \frac{b_i h_i^3}{12}. \]

Given \( b_i \) and \( h_i \) for every beam and column, the displacement at every node of the structure under a given loading can be computed. However, instead of using the dimensions of the cross-sections as the variables, the quantities \( a_i \) denoting the cross-sectional areas of the elements are introduced. Substituting \( a_i = b_i h_i \) reduces the nonlinearity of the problem (\( a_i \) appears linearly as a design variable). The design variables now are \( a_i \) and \( h_i \) for each beam and column.

Thus there are two design variables for each beam or column, with a total of \( 2(pk + (p - 1)k) = 2k(2p - 1) \) design variables for the problem. The state variables are the displacements for each of the three degrees of freedom for each node (horizontal and vertical displacements and rotation of the node), yielding a total of \( 3pk \) state variables. The state equations are square, i.e. they are of size \( 3pk \times 3pk \).

**Objective functions**

The two objectives considered in this problem are:

- The horizontal displacement of the rightmost node on the top storey. This will be referred to as the top storey displacement for the sake of brevity.

- The total volume of the structure.

**Additional constraints**

Lower and upper bounds were imposed on every \( a_i \) and \( h_i \). In addition, constraints were imposed to guarantee that the optimization does not produce beams that are too slender. These constraints enforce that the cross-section 'did not deviate much' from being a square. In particular, the constraints ensured that one side of the rectangular cross-section of any beam or column was no longer than \( 3/2 \) and no shorter than \( 2/3 \) of the length of the other side, i.e.

\[
\frac{2}{3} h_i \leq b_i \leq \frac{3}{2} h_i \\
\equiv \frac{2}{3} h_i \leq \frac{A_i}{h_i} \leq \frac{3}{2} h_i \\
\equiv \frac{2}{3} h_i^2 - A_i \leq 0 \quad \text{and} \quad A_i - \frac{3}{2} h_i^2 \leq 0.
\]
Data
Other data used in the problem were as below:

- Modulus of elasticity of material: $29 \times 10^3$ ksi.
- Length of each beam: 30 ft.
- Length of each column: 40 ft.
- Bounds on $a_i$: [50, 80] sq. in.
- Bounds on $h_i$: [10, 20] in.

Pareto curves
The top storey displacement was computed using a black-box approach by solving the state equations at each iteration and not by posing the state equations as equality constraints as in the all-at-once approach used in the truss problem. The structure stiffness matrix was set up using the direct stiffness method. It was assumed that the support settlements are all zero. The problem was solved using NBI and the method of convex combinations for 16 uniformly distributed parameter settings. The Pareto curves thus obtained depicting the trade-off between top storey displacement and total volume are shown in Fig. 10.2.
Figure 10.2  Pareto curves for minimizing the top storey displacement and total volume of frame with 3 supports and 4 storeys using NBI and convex combinations
Chapter 11

The Robustness Optimization Problem: An Application

This chapter will focus on another very important problem in engineering design, the robustness optimization problem. This is inherently a multicriteria optimization problem. This chapter aims to devise a formulation and computational technique for robustness optimization for constrained nonlinear problems and relate it to multicriteria optimization.

11.1 Introduction

Optimization problems arising in engineering design often involve variables and parameters that are inherently stochastic. Manufacturing processes usually produce units which have random variations. Environmental factors under which systems operate are very often random quantities. Traditional deterministic optimization assumes that given an optimal design, the manufacturer will manufacture it accurately, thus failing to take into account the effect of these stochastic aberrations in the variables on the function value.

Let us consider a particular example. Consider a design that achieves minimum displacement of the top storey in a particular multi-storey frame design problem. When the manufacturer starts out to build beams and columns with the specified optimal geometric properties, it is quite likely that there will be random variations in the manufacturing process and the specifications will not be met precisely. This will cause a departure of the actual design from the intended design and hence a change in the top-storey displacement.

In addition to the manufacturing errors, the loads (wind, snow or otherwise) actually imposed on the structure may be different from the ones taken into account in the 'optimal design' owing to the stochastic nature of such loads.

Based on such considerations, it is more prudent to choose a design which accounts for the possible aberrations in the relevant parameters in addition to achieving a low value of the objective function. This motivates the need for a technique which incorporates a measure of robustness during the optimization process. This is different
from 'sensitivity analysis' which only attempts to present a measure of how robust
the design is after the optimization process is over. This paper presents what Mulvey,
Vanderbei and Zenios [30] call a 'proactive' approach, i.e. one where the optimization
process is designed to yield a robust solution.

The following sections begin by introducing the expectation minimization problem
as a robust version of the standard unconstrained optimization problem and formulating
an approximation to the expectation which can be posed as an objective function
to drive a numerical optimization procedure. A discussion follows of why the problem
of finding a robust optimum is inherently a biobjective problem where optimality is
traded off for gain in robustness. Next, this formulation is extended to optimization
problems constrained by state equations and a computationally efficient and 'optimizable'
approximation to the expectation of the objective is derived. Finally, inequality
constraints are introduced in the problem and the effects of random aberrations on
these inequalities are discussed. The chapter concludes with a numerical example
arising from the truss design problem discussed in Chapter 9.

11.2 The robustness optimization problem

Consider the unconstrained optimization problem

$$\min_x f(x)$$

where $f : \mathbb{R}^N \rightarrow \mathbb{R}$ is a twice continuously differentiable mapping. Here $x$ denotes the
vector of design variables. We will defer the introduction of environmental variables
$z$ in the formulation until Section 11.3.2 for ease of presentation.

If $x^*$ is the chosen optimal point for the above problem, the robust designer would
want the objective function value to be 'near optimal' for all nearby values of $x^*$. So
instead of posing the problem as the standard unconstrained optimization problem
as above, a more acceptable formulation would be the following min-max problem:

$$\min_x \max_{\xi \in \mathcal{D}(x)} f(x + \xi)$$

where $\mathcal{D}(x) = \{ \xi : \xi \in \mathbb{R}^N, |x_i + \xi_i| \leq \alpha_i, i = 1, 2, \ldots, n \}$ and the aberration $\xi_i$
fluctuates randomly in the interval $[x_i - \alpha_i, x_i + \alpha_i]$. For simplicity of notation, $\mathcal{D}(x)$
will henceforth be denoted simply by $\mathcal{D}$. 
As posed above, the problem is nonsmooth and hence not tractable by quasi-Newton methods. A more tractable and quite reasonable reformulation of the problem is obtained by converting the problem to the one below:

\[
\min_x \bar{R}(x) = \frac{\int_D f(x + \xi) d\xi}{\int_D d\xi}.
\]

Here \( \bar{R} \) is a smooth measure of instability of the function \( f \) in a neighborhood of \( x \). To explain the notation, \( R \) stands for robustness, and \( \bar{R} \) for robustness complement or instability. This formulation though not equivalent to the min-max formulation has the same flavor and addresses the same issue of taking into consideration the 'activity in a neighborhood' of the point as opposed to at only that one point. The above formulation is more desirable in situations where safeguarding against the worst-case behavior is too conservative.

It can be seen that \( \bar{R}(x) \) is nothing other than \( E(f(x + \xi)) \), the expectation of \( f(x + \xi) \) given that \( \xi \) is distributed uniformly over the hypercuboid \( D \). Since there is no reason to restrict ourselves to the uniform distribution, we generalize the expectation minimization problem given any smooth distribution as below:

\[
\min_x \bar{R}(x) = \int_D f(x + \xi) w(\xi) d\xi,
\]

where \( w(\xi) \) is the multidimensional probability density function of \( \xi \) over the domain \( D \) (so \( \int_D w(\xi) d\xi = 1 \)). Observe that the domain \( D \) does not need to be bounded, e.g. if \( w(\xi) \) is the multinormal distribution, the domain is all of \( \mathbb{R}^n \). Moreover, we assume throughout the paper that the p.d.f. is independent of \( x \), i.e. the distribution of the aberrations is the same regardless of the target design point.

This robustness optimization problem has been studied usually in the setting of a linear program, as in Mulvey, Vanderbei and Zenios [30] and Mayer [29]. Several applications that have been studied include the diet problem with a stochastic treatment of the nutrition contents of some foods, the power capacity expansion problem (demand for electric power being uncertain), airline allocation for the Air Force and minimum weight structural design, optimized over a random distribution of wind loads. Robustness optimization has also been incorporated in multiobjective space vehicle design by Erikstad, et. al. [9]. The expectation minimization approach has been referred to in Slowinski and Teghem [42], Stancu-Minasian [46] and Ermoliev and Wets [10] among many others. We shall endeavor to handle nonlinear problems in the robust setting as well.
11.2.1 Evaluating the multidimensional integral

Since the integral for evaluating the expectation may not be available analytically in closed form, especially if $\text{dim}(x)$ is large, we need alternate procedures for evaluating the integral. A numerical quadrature scheme can require too many evaluations of $f(x)$, and can be quite difficult mathematically when $\text{dim}(x)$ gets large. In order to bypass these difficulties, we shall resort to building a one-point quadratic model of $f(x)$ and then use this approximation to evaluate the integral. Thus,

$$
\tilde{R}(x) = \int_D f(x + \xi)w(\xi)d\xi
$$

$$
\approx \int_D (f(x) + \nabla_x f(x)^T \xi + \frac{1}{2} \xi^T \nabla^2_x f(x) \xi)w(\xi)d\xi.
$$

Clearly,

$$
\int_D f(x)w(\xi)d\xi = f(x)\int_D w(\xi)d\xi = f(x).
$$

For ease of notation, let $\gamma_i$ be the $i^{th}$ component of $\nabla_x f(x)$ and $h_{ij}$ the $i^{th}$ row $j^{th}$ column entry of $\nabla^2_x f(x)$, the Hessian of $f(x)$ w.r.t. $x$. Then

$$
\int_D \nabla_x f(x)^T \xi w(\xi)d\xi
$$

$$
= \int_D \sum_{i=1}^N \gamma_i \xi_i w(\xi)d\xi
$$

$$
= \sum_{i=1}^N \gamma_i \int_D \xi_i w(\xi)d\xi = \sum_{i=1}^N \gamma_i \mu_i = \nabla_x f(x)^T \mu
$$

where $\mu$ is the vector of expectations of the aberrations $\xi_i$.

Further

$$
\int_D \frac{1}{2} \xi^T \nabla^2_x f(x) \xi w(\xi)d\xi
$$

$$
= \int_D (\sum_{1 \leq i < j \leq n} \xi_i \xi_j h_{ij} + \frac{1}{2} \sum_{i=1}^n \xi_i^2 h_{ii})w(\xi)d\xi
$$

where the symmetricity of the Hessian of $f$ was used to get rid of the factor of $\frac{1}{2}$ in the summation over the off-diagonal terms. Elementary statistics tells us that

$$
\int_D \xi_i \xi_j w(\xi)d\xi = E(\xi_i \xi_j) = \rho_{ij} \sigma_i \sigma_j + \mu_i \mu_j
$$

and

$$
\int_D \xi_i^2 w(\xi)d\xi = E(\xi_i^2) = \sigma_i^2 + \mu_i^2
$$
where \( \rho_{ij} \) is the correlation coefficient between \( \xi_i \) and \( \xi_j \) and \( \sigma_i \) is the standard deviation of \( \xi_i \). Now the integral of the second order term can be expressed in terms of moments (upto second order) of the joint distribution as

\[
\int_\mathcal{D} \frac{1}{2} \xi^T \nabla^2_x f(x) \xi \, w(\xi) \, d\xi = \sum_{1 \leq i < j \leq n} h_{ij}(\rho_{ij} \sigma_i \sigma_j + \mu_i \mu_j) + \frac{1}{2} \sum_{i=1}^{n} h_{ii}(\sigma_i^2 + \mu_i^2).
\]

Observe that the above is a special case of the following general result in matrix form for any symmetric matrix \( H \) with \( H = \nabla^2_x f(x) \)

\[
\int_\mathcal{D} \frac{1}{2} \xi^T H \xi \, w(\xi) \, d\xi = \frac{1}{2} s^T (H \ast R) s + \mu^T H \mu
\]  

(11.1)

where \( R \) is the correlation matrix of the distribution function \( w(\xi) \) with ones on the diagonal. The operation \( \ast \) denotes the Hadamard product or more plainly, term-by-term multiplication. Here \( s \) is the vector of standard deviations with \( \sigma_i \) as its \( i^{th} \) entry.

Thus the integral can be approximated as

\[
\hat{R}(x) = \int_\mathcal{D} f(x + \xi) w(\xi) d\xi
\]

\[
\approx f(x) + \nabla_x f(x)^T \mu + \sum_{1 \leq i < j \leq n} h_{ij}(\rho_{ij} \sigma_i \sigma_j + \mu_i \mu_j) + \frac{1}{2} \sum_{i=1}^{n} h_{ii}(\sigma_i^2 + \mu_i^2).
\]

If the aberrations are assumed to be stochastically independent, then \( \rho_{ij} = 0 \forall i \neq j \), so

\[
\hat{R}(x) \approx f(x) + \nabla_x f(x)^T \mu + \sum_{1 \leq i < j \leq n} h_{ij} \mu_i \mu_j + \frac{1}{2} \sum_{i=1}^{n} h_{ii}(\sigma_i^2 + \mu_i^2).
\]

If the distributions are symmetric about \( x \) (e.g. uniform or multinormal distributions centered at \( x \)) then \( \mu_i = 0, \forall i \), i.e. each aberration component has mean zero, so

\[
\hat{R}(x) \approx f(x) + \sum_{1 \leq i < j \leq n} h_{ij} \rho_{ij} \sigma_i \sigma_j + \frac{1}{2} \sum_{i=1}^{n} h_{ii} \sigma_i^2.
\]

Note all throughout the dependence of \( h_{ij} \) on \( x \).

As a small numerical example, let us consider the one-dimensional problem

\[
\min_x \frac{1}{x} + 2.5 \sqrt{x}.
\]

The graph of this function is shown in Fig. 11.1; clearly it is desirable to stay away from the ‘steep side’ of the minimum. Thus, even though the minimizer of this function occurs at \( x^* = 0.8617 \), the point that minimizes the expectation assuming a
uniform distribution of the aberration over an interval of width 1.4 centered at $x$ is $x^* = 1.139$. This was obtained using the exact integral; using the approximation derived above minimizing the expectation yields $x^* = 1.059$. Both of these robust minima migrate considerably to the flatter side of the curve.

Figure 11.1 The 'flatter' side of the minimum is more robust w.r.t. optimality

11.3 Trade-off inherent in the robustness problem

Simply minimizing the expectation has its drawbacks, as shown in the examples in Fig. 11.2. Often what the user aims to achieve by minimizing expectation is to 'move away from the sharp minimum'. However, poor knowledge of the distribution $w(\xi)$ on the aberrations might result in the minimum of the expectation ending up very near the sharp minimum of the function. For example, this would happen if the uniform distribution on $\xi$ in the example problem earlier had been replaced by a normal distribution with mean 0 and variance 0.01. This would also happen if the minimum value is very low compared to other function values that lie just outside the domain of the aberrations, so that the comparatively low contributions of the function values
around the function minimum still make the function minimizer a minimizer of the expectation. This is the situation which we attempted to illustrate in the first of the two figures in Fig. 11.2. The second figure shows a perhaps pathological case where the function values oscillate wildly in a region, but since the expectation 'smooths out' all the fluctuations, the highly unstable oscillatory region is reported as being robust.

The problems pointed out above arise because minimizing the expectation pays no heed to the dispersion in the objective function values near the point. However, even though suggested as an alternative by many, including Slowinski and Teghem [42], solely minimizing the dispersion in the function values is insufficient since it pays no attention to the function values. Thus a minimizer of dispersion could conceivably end up at a flat maximum of a function as in case of Fig. 11.3.

We feel that robustness optimization is an attempt to compromise between two objectives: minimizing the original objective function value at a point and minimizing an estimate of the behavior of the function in a neighborhood of the point. This observation directs us towards a biobjective formulation of the robustness problem. The following biobjective formulations are possible for trading off between 'optimality' and 'robustness':

- Minimizing expectation and minimizing dispersion
Figure 11.3 Drawback of minimizing dispersion in function values

- Minimizing expectation and minimizing the original objective
- Minimizing the original objective and minimizing dispersion

Though often implicit in its appearance, this biobjective approach is no stranger in the robustness optimization literature. Su and Renaud [50] penalize their original objective by a term which they call the sensitivity index. This is nothing but a very rough estimate of the expectation, amounting overall to minimizing a linear combination of the function and its expectation. Mulvey, Vanderbei and Zenios [30] plot the 'efficient frontier' for a robust version of a problem called AFIRO from NETLIB linear programs using a similar penalty parameter approach. Box and Jones [1] go a step further and actually minimize a convex combination of the expectation and variance for different parameter settings to find the Pareto set. However, their objective had a particular quadratic form; such assumptions cannot be made about our objective functions. Also, their system output was allowed to have a random component not captured by the physical model which will not be the case in our target applications arising chiefly from structural design. Furthermore, minimizing dispersion in our non-linear setting can be stifled in the absence of higher order moment information on the distribution of \( \xi \), as shown in the next subsection. And finally, our approach will employ NBI and not convex combinations as the tool for biobjective optimization.
11.3.1 Minimizing dispersion as an alternative

An obvious measure of dispersion in the objective values in a neighborhood about the target point is

\[ V(x) = \int_D (f(x + \xi) - f(x))^2 w(\xi) \, d\xi \]

If \( f(x + \xi) \) is considered as an estimator of \( f(x) \), then \( V(x) \) is the mean-squared error (MSE) of estimation.

Let us consider approximating this integral using the same second-order Taylor expansion technique used for the expectation. Then we get

\[ V(x) \approx \int_D (\nabla_x f(x)^T \xi + \frac{1}{2} \xi^T \nabla^2_x f(x) \xi)^2 w(\xi) \, d\xi. \]

When expanded, the integrand will clearly involve up to fourth order terms in \( \xi \), which when integrated will give rise to terms involving up to fourth order moments of the distribution \( w(\xi) \). This causes no difficulty if the exact joint p.d.f. of \( \xi \) is known in which case the higher order moments can be computed, but this will rarely be the case in a practical situation. At best the engineers are likely to have some estimates of the aberration means (possibly 0) and standard deviations and no information about higher order moments or even correlations.

If a first-order approximation for \( f(x + \xi) \) is used as in Su and Renaud [50], then the approximation for the integrand becomes

\[ V(x) \approx \int_D (\nabla_x f(x)^T \xi)^2 w(\xi) \, d\xi. \]

But now \( V(x) \) suffers from a serious drawback: it vanishes at any point where \( \nabla_x f(x) = 0 \). In other words, this approximation suggests that the most robust points are the unconstrained minimizers and maximizers of the function, which completely defeats our purpose.

Given the above concerns, we shall strive to avoid formulations that involve minimizing dispersion. This leaves us with one approach, that of minimizing expectation and minimizing the original objective.

The Pareto curves for minimizing expectation and minimizing the original function value using NBI for our earlier single variable problem are shown in Fig. 11.4 (using exact integral for expectation) and Fig. 11.5 (using the approximation).

With the trade-off curves in hand one can see that it is not worth giving up optimality by increasing it beyond 3.5, because then on very little gain in robustness
can be achieved by giving up optimality. However, without the trade-off curve one might have chosen a minimizer of some arbitrarily chosen weighted sum of the function and the expectation with no knowledge of what part of the Pareto set it fell in. Without the biobjective formulation in the first place, the user would have been very likely to choose the minimizer of expectation as the final point, which the trade-off curve suggests as being not the best choice.

11.3.2 Environmental Variables

Following Box and Jones [1] and many others in the statistics literature, we introduce the concept of environmental variables. Essentially these are the variables which have random aberrations about their ‘reported’ values, but unlike the design variables, the ‘reported’ values are fixed by the ‘environment’ and is not up to the designer to choose.

For example, the wind load in a structural problem falls in this category. The specified wind load may be distributed uniformly over the interval [50, 70] kips. Equivalently, the wind load can be thought of as being 60 kips with an aberration distributed uniformly over [−10, 10] kips. The best the designer can do is minimize the expectation over the random wind load provided, but (s)he cannot specify a ‘target’
wind load about which (s)he wants the wind load to be distributed with the specified aberration.

We shall introduce additional environmental variables in our problem and denote them by $z$. The expectation minimization problem now is written as

$$\min_x \mathcal{R}_z(x) = \int_{\mathcal{P}} f(x + \xi_x, z + \xi_z) w(\xi_x, \xi_z) d\xi_x d\xi_z,$$

where $\xi_x$ denotes the vector of aberrations from the target design variables $x$ and $\xi_z$ denotes the aberrations from the fixed environmental variables $z$. Note that $\mathcal{R}_z(x)$ is minimized only over $x$; $z$ as a subscript on $\mathcal{R}$ denotes that $z$ is only a fixed parameter and not an optimization variable.

### 11.4 Expectation Evaluation for Equality-Constrained Problems

In this section we shall attempt to extend the formulation for the robustness optimization problem for nonlinearly constrained problems. Firstly we shall try to obtain a formulation for equality-constrained robustness optimization where the equality constraints can be thought of as being the state equations of the system, i.e., a system of nonlinear equations, perhaps discretized differential equations, derived from physics or chemistry that accurately describes the state of the system. In other words, the
state equations cannot be violated without something unphysical happening. Thus the equality constraints here are more special than in applications like image restoration where the (linear) equality constraints can be violated owing to noisy coefficients, even though there exists a feasible state of the system (i.e., there is an image even if the equality constraints do not hold), as in Mulvey, et. al. [30]. Our state variables will be denoted by \( y \), the control or design variables by \( x \) and the environmental variables by \( z \). We will assume that given a particular setting of \( x \) and \( z \), the state equations

\[
C(x, y, z) = 0
\]

can be solved to obtain a unique \( y \), so that implicitly the state equations define a function \( \psi : (x, z) \mapsto y \), i.e. \( y = \psi(x, z) \). Thus given the aberrations \( \xi_x \) and \( \xi_z \) in \( x \) and \( z \), the state equation \( C(x + \xi_x, y, z + \xi_z) = 0 \) can be solved to obtain \( y = \psi(x + \xi_x, z + \xi_z) \).

If our traditional equality-constrained optimization problem is

\[
\min_{x, y} f(x, y, z) \\
\text{s.t. } C(x, y, z) = 0,
\]

or equivalently

\[
\min_{x} f(x, \psi(x, z), z),
\]

then the expected cost minimization version of the above is

\[
\min_{x} \bar{R}_z(x) = \int_{\mathcal{D}} f(x + \xi_x, \psi(x + \xi_x, z + \xi_z), z + \xi_z) w(\xi_x, \xi_z) \, d\xi_x \, d\xi_z.
\]

### 11.4.1 Approximating the expectation integral for constrained problems

Since it is unlikely that the function \( \psi \) will be available in closed form in most practical situations, the need for approximating the integral is more pressing here than in the unconstrained case. As in the unconstrained case, a Taylor series approximation technique will be used here. A quadrature-based numerical integration scheme would involve solving the state equations many times and is thus computationally more expensive. We shall begin by replacing \( \psi(x + \xi_x, z + \xi_z) \) with its first-order approximation, i.e.,

\[
\psi(x + \xi_x, z + \xi_z) \approx \psi(x, z) + \nabla_x \psi^T \xi_x + \nabla_z \psi^T \xi_z.
\]
where $\nabla_x \psi^T = \nabla_x \psi^T(x,z)$ is the $\dim(y) \times \dim(x)$ Jacobian matrix (transpose of the gradient) with $\frac{\partial \psi}{\partial x_j}$ as its $i^{th}$ row $j^{th}$ column element; $\nabla_z \psi$ is defined similarly. Now the integrand can be approximated using a Taylor series expansion as below:

$$
\begin{align*}
&f(x + \xi_x, \psi(x + \xi_x, z + \xi_z), z + \xi_z) \\
&\approx f(x + \xi_x, \psi(x, z) + \nabla_x \psi^T \xi_x + \nabla_z \psi^T \xi_z, z + \xi_z).
\end{align*}
$$

The above expanded in a Taylor series about $(x, \psi(x,z), z)$ gives

$$
\begin{align*}
f(x, \psi(x,z), z) + \nabla_x f^T \xi_x + \nabla_y f^T (\nabla_x \psi^T \xi_x + \nabla_z \psi^T \xi_z) + \nabla_z f^T \xi_z \\
+ \text{second order terms} + \text{higher order terms}.
\end{align*}
$$

Here the arguments $(x, z)$ have often been dropped to make the algebra less cumbersome.

The first-order approximation to the integrand can be integrated as below:

$$
\begin{align*}
\int_D \left( f(x, \psi(x,z), z) + \nabla_x f^T \xi_x + \nabla_y f^T (\nabla_x \psi^T \xi_x + \nabla_z \psi^T \xi_z) + \nabla_z f^T \xi_z \right) w(\xi_x, \xi_z) \, d\xi_x \, d\xi_z \\
= f(x, \psi(x,z), z) + \nabla_x f^T \mu_x + \nabla_z f^T \mu_z + \nabla_y f^T \nabla_x \psi \mu_x + \nabla_y f^T \nabla_z \psi \mu_z,
\end{align*}
$$

where $\mu_x$ and $\mu_z$ are the mean vectors of $\xi_x$ and $\xi_z$ respectively.

The second-order terms in the expansion can be written as

$$
\frac{1}{2} \begin{bmatrix}
\xi_x^T \\
\xi_y^T \\
\xi_z^T
\end{bmatrix}
\begin{bmatrix}
\nabla_x^2 f & H_{xy} & H_{xz} \\
H_{yx}^T & \nabla_y^2 f & H_{yz} \\
H_{zx}^T & H_{zy}^T & \nabla_z^2 f
\end{bmatrix}
\begin{bmatrix}
\xi_x \\
\xi_y \\
\xi_z
\end{bmatrix},
$$

where $\xi_y = \nabla_x \psi^T \xi_x + \nabla_z \psi^T \xi_z$, $H_{xy}$ is the $\dim(x) \times \dim(y)$ matrix whose $i^{th}$ row $j^{th}$ column entry is $\frac{\partial^2 f}{\partial x_i \partial y_j}$, $H_{yz}$ and $H_{xz}$ are defined similarly.

Expanding the above, substituting $\nabla_x \psi^T \xi_x + \nabla_z \psi^T \xi_z$ for $\xi_y$ and regrouping terms yields the following useful form for the second-order quantities

$$
\frac{1}{2} \xi_x^T \mathcal{H}_1 \xi_x + \frac{1}{2} \xi_z^T \mathcal{H}_2 \xi_z + \xi_x^T \mathcal{H}_3 \xi_z,
$$

where

$$
\begin{align*}
\mathcal{H}_1 &= \nabla_x^2 f + H_{xy} \nabla_x \psi^T + \nabla_x \psi H_{xy}^T + \nabla_x \psi \nabla_y^2 f \nabla_x \psi^T, \\
\mathcal{H}_2 &= \nabla_z^2 f + H_{yz} \nabla_z \psi^T + \nabla_z \psi H_{yz}^T + \nabla_z \psi \nabla_y^2 f \nabla_z \psi^T, \\
\mathcal{H}_3 &= H_{xz} + H_{xy} \nabla_z \psi^T + \nabla_z \psi H_{yz} + \nabla_z \psi \nabla_y^2 f \nabla_z \psi^T.
\end{align*}
$$
Observe that $\mathcal{H}_1$ and $\mathcal{H}_2$ are symmetric; thus in order to integrate the first two terms in the second-order expansion the result derived in equation (11.1) can be applied directly, yielding

$$
\int_D \frac{1}{2} \xi_x^T \mathcal{H}_1 \xi_x = \frac{1}{2} s_x^T (\mathcal{H}_1 \ast R_x) s_x + \mu_x^T \mathcal{H}_1 \mu_x
$$

and

$$
\int_D \frac{1}{2} \xi_z^T \mathcal{H}_2 \xi_z = \frac{1}{2} s_z^T (\mathcal{H}_2 \ast R_z) s_z + \mu_z^T \mathcal{H}_2 \mu_z.
$$

The third term involving the asymmetric matrix $\mathcal{H}_3$ contains the aberrations in both the design and the environmental variables. Let us assume for simplicity that the aberrations in the environmental variables and the aberrations in the design variables are uncorrelated. This is certainly the case with our example from structural optimization because the fluctuation in the wind and suspended loads and the manufacturing errors in the beam cross-sections can be assumed to be independent of each other. In the absence of correlation between $\xi_x$ and $\xi_z$ the integral of the third term clearly vanishes.

It should be emphasized that it is not any harder to actually integrate this cross-term, it just happens to be unnecessary for our purposes.

Assembling all of the above quantities, the approximate expectation with second-order terms becomes

$$
\bar{R}(x) \approx f(x, \psi(x, z), z) + \nabla_x f^T \mu_x + \nabla_z f^T \mu_z + \nabla_y f^T \nabla_x \psi^T \mu_x + \nabla_y f^T \nabla_z \psi^T \mu_z
$$

$$+
\frac{1}{2} s_x^T (\mathcal{H}_1 \ast R_x) s_x + \mu_x^T \mathcal{H}_1 \mu_x + \frac{1}{2} s_z^T (\mathcal{H}_2 \ast R_z) s_z + \mu_z^T \mathcal{H}_2 \mu_z.
$$

(11.2)

The only remaining issue is that of obtaining the quantities $\nabla_x \psi$ and $\nabla_z \psi$. This can be done by differentiating the state equations as described below.

Let $r$ denote any variable in the set of design or environmental variables. Then differentiating $C(x, \psi(x, z), z) = 0$ with respect to $r$ gives

$$
\nabla_r C(x, \psi(x, z), z) + \nabla_y C(x, \psi(x, z), z)^T \nabla_r \psi(x, z) = 0
$$

$$
\Rightarrow \nabla_y C(x, \psi(x, z), z)^T \nabla_r \psi(x, z) = -\nabla_r C(x, \psi(x, z), z).
$$

Assembling the derivatives for every $r \in \{z_i : i = 1, \ldots, \dim(x)\}$ in order gives

$$
\nabla_y C(x, \psi(x, z), z)^T \nabla_z \psi(x, z)^T = -\nabla_x C(x, \psi(x, z), z)^T.
$$
In order to obtain $\nabla_x \psi$ we will assume that $\nabla_y C(x, \psi(x, z), z)^T$ is square and invertible, i.e. $\dim(y) = \dim(C)$, so that there are as many state variables as state equations. Though this assumption may sound restrictive, a wide spectrum of interesting problems in engineering fall in this category, as does our structural example. Note that if the number of state variables is less than the number of state equations there would be fewer deterministic variables than equations and it would be impossible to make the state equations hold for every setting of $x$ and $z$, contradicting one of our basic assumptions.

Assuming thus that $\nabla_y C(x, \psi(x, z), z)^{-T}$ exists, we have the following required equations

$$\nabla_x \psi^T(x, z) = -\nabla_y C^{-T}(x, y, z) \nabla_x C^T(x, y, z)$$
$$\nabla_z \psi^T(x, z) = -\nabla_y C^{-T}(x, y, z) \nabla_z C^T(x, y, z)$$

or, omitting the arguments,

$$\nabla_x \psi^T = -\nabla_y C^{-T} \nabla_x C^T$$
$$\nabla_z \psi^T = -\nabla_y C^{-T} \nabla_z C^T,$$

which need to be substituted into our previously-derived expression for $\bar{R}_z(x)$ (equation 11.2) to obtain $\bar{R}_z(x, y)$, the expectation as a function of $x$ and $y$ (since the derivatives of $C$ have $y = \psi(x, z)$ as an argument).

This completes our treatment on how, for robustness optimization problems constrained by square state equations, the expectation can be computed approximately using derivatives of the state equations and the objective function. Now we solve

$$\min_{x,y} \bar{R}_z(x, y)$$
$$\text{s.t.} \quad C(x, y, z) = 0.$$
\[ \text{s.t. } C(x, y, z) = 0 \]
\[ g(x, y, z) \leq 0. \]

The inequality constraints \( g(x, y, z) \) will be classified into two groups, hard inequality constraints \( g^h(x, y, z) \) and soft inequality constraints \( g^s(x, y, z) \). This classification is very important for our formulation of inequality-constrained robustness optimization.

**Hard inequalities** are those that simply cannot be violated because of physical considerations or otherwise, no matter what the random aberrations in the design or environmental variables are. A constraint demanding the length of a rod to be nonnegative is an example. There does not exist any set of random aberrations in the design variables which can cause length, mass or time to be negative. Thus our robustness optimization must take into account the fact that these hard inequalities can never be violated, i.e.

\[ g^h(x + \xi_x, \psi(x + \xi_x, z + \xi_z), z + \xi_z) \leq 0 \]

for all settings of \( \xi_x \) and \( \xi_z \). This is achieved by an old trick often used in handling inequality constraints in nonlinear programming known as 'squared slacks', i.e. the inequalities \( g(x, y, z) \leq 0 \) are replaced by

\[ g^h(x, y, z) = -\frac{1}{2} v \ast v \quad (11.3) \]

where \( v \) is a vector of unconstrained real numbers and has the same dimension as \( g^h \). Here \( v \ast v \), the Hadamard product of \( v \) with itself, is simply a vector with each component of \( v \) squared. The permissible aberrations \( (\xi_x, \xi_z) \) are thus defined as those that satisfy

\[ g^h(x + \xi_x, \psi(x + \xi_x, z + \xi_z), z + \xi_z) = -\frac{1}{2} v \ast v, \]

i.e., \( \mathcal{D} = \{(\xi_x, \xi_z) : g^h(x + \xi_x, \psi(x + \xi_x, z + \xi_z), z + \xi_z) \leq 0 \} \).

The idea of squared-slacks has been discarded in nonlinear programming because it often results in loss of convexity of the problem and destroys good semi-local convergence of NLP algorithms. But it proves to be very useful for our formulation since it provides a way of defining the permissible aberrations in terms of nonlinear equalities only.
Just as we assumed for the equality-constrained case that there exists a function \( \psi : (x, z) \mapsto y \), we assume here that there exists a function \( \psi_g(x, z) \mapsto (y, v) \) which is implicitly defined by the set of equalities

\[
C(x, y, z) = 0
\]

\[
g^h(x, y, z) + \frac{1}{2} v \cdot v = 0.
\]

Observe that since the second set of equalities introduces \( \dim(g^h) \) new equations and \( \dim(g^h) \) new variables \( v_i \),

\[
\dim(C) = \dim(y) \Rightarrow \dim(C \cup g^h) = \dim(y \cup v).
\]

Thus the problem constrained by equalities and only hard inequalities is reduced so that it fits the robustness optimization formulation for equality-constrained optimization derived in the previous section.

**Soft inequalities** are those which can be violated because of untoward aberrations in the design or environmental variables. For example, requiring the cross-section of a beam to be \( \geq 0.8 \text{ in}^2 \) is a soft inequality constraint, because there is no guarantee that the manufacturing process will not erroneously produce a beam with a smaller cross-section. If there was a guarantee, this should be classified as a hard inequality. Note that this classification into hard and soft inequalities is not based on whether the designer would like the inequality to hold rigorously, but rather on whether it actually always is satisfied in physical terms in spite of aberrations from targets. Thus, the designer might absolutely want the beam cross-section to not fall below 0.8 \text{ in}^2 because it may make the beam very susceptible to fracture otherwise, but still cannot specify it as a hard inequality unless the manufacturing process guarantees that it will not be violated.

One undesired phenomenon that may occur here is that the expected value of the objective function might be low at a point because the aberred values of the design variables map to low objective function values, even though those aberred values actually violate the soft inequalities. This is illustrated in Fig. 11.6.

This is one reason why our model should also impose a penalty for violating the soft inequalities owing to aberrations in \( x \). Another obvious reason is that it may be undesirable to have a violation in the soft inequalities because of fluctuations in \( x \) and \( z \), so that the designer may want what Mulvey, et. al. in [30] call *model robustness*, i.e. the (soft inequality) constraints staying ‘almost feasible’ for all aberred scenarios.
Figure 11.6 Expectation of $f$ can be lower in a neighborhood of A than in the same for B, but solely because infeasible $x$ values map to smaller $f$ values and contribute to a smaller value of the expectation.

of the variables. Thus the 'degree of infeasibility' should also be minimized or forced to be under a certain permissible level. In order to quantify the 'degree of infeasibility' the following measure, denoted by $\phi$, is proposed:

$$
\phi(x) = \int_D \tau_p(g^s(x + \xi_x, \psi_g(x + \xi_x, z + \xi_z), z + \xi_z)) w(\xi_x, \xi_z) d\xi_x d\xi_z
$$

where

$$
\tau_p(g^s(x, \psi_g(x, z), z)) = \frac{1}{p} \sum_{i=1}^{i=\text{dim}(g^s)} \left( \max(0, g_i^s(x, \psi_g(x, z), z)) \right)^p
$$

where $p$ is a positive integer of choice.

It is easy to see that $\tau_p(g^s)$ becomes positive only when some component of $g^s$ becomes positive and thus measures the degree to which the soft constraints are violated at a point, while $\phi$ measures this violation over a region.

Let us consider evaluating or approximating the integral for $\phi$. If $p = 1$, which is along the lines of Mulvey, et. al. [30], then $\tau_p$ is continuous but not differentiable at any point where $g_i^s$ vanishes for some $i$. The optimization process thus requires nonsmooth optimization techniques, which we wish to avoid.
For $p = 2$, $\tau_p$ is both continuous and differentiable everywhere and

$$\nabla x\tau_2 = \sum_i \max\{0, g_i^t\} \nabla x g_i^t.$$  

However, second derivatives of $\tau_2$ clearly do not exist. If we are to use a Taylor approximation of the integrand as before, a second-order expansion of the integrand is desirable (the first-order term vanishes and plays no role if the means of the aberrations are 0, which is very often the case). Moreover, these derivatives must themselves be at least once and preferably twice differentiable for the approximation to the integral to be tractable by gradient-based optimization. With this in mind, let us choose $p = 4$, which is thrice differentiable. It can be shown easily that

$$\nabla x\tau_4 = \sum_i (\max\{0, g_i^t\})^3 \nabla x g_i^t$$

and

$$\nabla^2 x\tau_4 = \sum_i (\max\{0, g_i^t\})^3 \nabla^2 x g_i^t + 3 \sum_i (\max\{0, g_i^t\})^2 \nabla x g_i^t (\nabla x g_i^t)^T.$$  

Approximating the integral by integrating a Taylor approximation of $\tau_4$ suffers from a major drawback. Since the first and second derivatives of $\tau_4$ vanish whenever the quantities $\max\{0, g_i^t\}$ vanish $\forall i$, and $\max\{0, g_i^t\}$ vanishes $\forall i$ at any point that satisfies all the soft inequalities, the approximation $\phi$ to the integral of $\tau_4$ vanishes at any point that satisfies $g_i^t \leq 0 \forall i$. This completely defeats our purpose since this implies that the 'amount of infeasibility' in the neighborhood of any feasible point is 0, no matter how badly the inequalities are violated because of aberrations from this feasible target point.

**An alternate approximation**  Given the aforementioned drawback, we shall use a more traditional discretized approximation to the integral for $\phi$. In other words, a finite set of points $\hat{D}$ is chosen from the infinite set $D$ and the integral replaced by a sum of the integrand over all points in $\hat{D}$, i.e.

$$\phi_\varepsilon(x) = \int_D \tau_p(g^\varepsilon(x + \xi_x, \psi_g(x + \xi_x, z + \xi_z), z + \xi_z)) w(\xi_x, \xi_z) d\xi_x d\xi_z$$

$$\approx \sum_{(\xi_x, \xi_z) \in \hat{D}} \tau_p(g^\varepsilon(x + \xi_x, \psi_g(x + \xi_x, z + \xi_z), z + \xi_z)) w(\xi_x, \xi_z) \Delta(\xi_x, \xi_z)$$

where $\Delta(\xi_x, \xi_z)$ is the weight associated with the point $(\xi_x, \xi_z)$ in the quadrature rule.

The above still requires us to have $\psi_g(x + \xi_x, z + \xi_z) \forall (\xi_x, \xi_z) \in \hat{D}$, which involves solving the state equations $|\hat{D}|$ times and can be very expensive. Thus in course of
the actual computation, $\psi_g(x + \xi_x, z + \xi_z)$ can be approximated by a first-order Taylor approximation as before, i.e.

$$\psi_g(x + \xi_x, z + \xi_z) \approx \psi_g(x, z) + \nabla_x \psi^T_g \xi_x + \nabla_z \psi^T_g \xi_z.$$ 

Substituting the above, $\tau_p \circ g^*$ can be evaluated at an approximation of its argument as

$$\tau_p(g^*(x + \xi_x, \psi_g(x + \xi_x, z + \xi_z), z + \xi_z)) \approx \tau_p(g^*(x + \xi_x, \psi_g(x, z) + \nabla_x \psi^T_g \xi_x + \nabla_z \psi^T_g \xi_z, z + \xi_z)).$$

This discretized integral still involves evaluating the soft constraints $g^*$ $|\mathcal{D}|$ times. This may not be a problem if the computation of $g^*$ is not too expensive (e.g. if the soft inequality constraints consist mostly of bounds on variables), but if it is expensive it can in turn be approximated using a first-order Taylor approximation, i.e.

$$g^*(x + \xi_x, \psi_g(x, z) + \nabla_x \psi^T_g \xi_x + \nabla_z \psi^T_g \xi_z) \approx g^*(x, \psi_g(x, z), z) + (\nabla_x g^*)^T \xi_x + (\nabla_y g^*)^T (\nabla_x \psi^T_g \xi_x + \nabla_z \psi^T_g \xi_z) + (\nabla_z g^*)^T \xi_z.$$ 

Now $\tau_p$ can be evaluated at the above approximation of $g^*$. The resulting approximation to the discretized integral for $\phi_z$ is the seemingly clumsy-looking expression below.

$$\phi_z(x) \approx \sum_{(\xi_x, \xi_z) \in \mathcal{D}} \tau_p(g^*(x, \psi_g(x, z), z) + (\nabla_x g^*)^T \xi_x + (\nabla_y g^*)^T (\nabla_x \psi^T_g \xi_x + \nabla_z \psi^T_g \xi_z) + (\nabla_z g^*)^T \xi_z) w(\xi_x, \xi_z).$$

Observe that it suffices to choose the value of $p$ to be 2, because now a gradient-based optimizer can minimize the discretized integral using the first derivatives of $\tau_2$ which happen to exist. Also observe that the existence of second derivatives of $g^*$ is required in order to claim differentiability of $\tau_2$ with the first-order approximation of $g^*$ as its argument.

Brief comments on $\nabla_x \psi^T_g$ and $\nabla_z \psi^T_g$ are in order, since something special can be said about their structure. The united system of state equations and hard inequalities can now be differentiated obtain the gradients of $\psi_g$. Since $\nabla_x C = 0$ and $\nabla_y g^h = \text{diag}(v)$, the required gradients turn out to be

$$\nabla_x \psi^T_g = \begin{bmatrix} \nabla_y C & \nabla_y g^h \\ 0 & \text{diag}(v) \end{bmatrix} - \begin{bmatrix} (\nabla_x C)^T \\ (\nabla_x g^h)^T \end{bmatrix}$$

\[
\n\nabla_z \psi_g^T = \begin{bmatrix}
\nabla_y C & \nabla_y g^h \\
0 & \text{diag}(v)
\end{bmatrix}^{-T} \begin{bmatrix}
(\nabla_z C)^T \\
(\nabla_z g^h)^T
\end{bmatrix}.
\]

Substituting the above along with \((y, v)\) for \(\psi_g(x, z)\) in our approximations for \(\bar{R}_z(x)\) and \(\phi_z(x)\), \(\bar{R}_z\) and \(\phi_z\) are obtained as functions of \((x, y, v)\) and can now be minimized subject to satisfying the state equations, the squared-slack equality constraints (11.3) and the soft inequality constraints.

One might argue that the variables \(v\) can be eliminated from the above by replacing \(v_i\) with \(\sqrt{2g_i^h}\) for every \(i\), but then we need to ensure that the hard inequalities hold at every iteration of the optimization, or else the quantities \(\sqrt{2g_i^h}\) which appear in \(\nabla_z \psi_g\) and \(\nabla_z \psi_g\) (and hence \(\bar{R}_z\) and \(\phi_z\)) can become imaginary.

In the presence of soft inequality constraints, the earlier expectation minimization problem is replaced by the biobjective problem of minimizing \(\bar{R}_z(x, y)\) and \(\phi_z(x, y)\). We shall differ from other researchers in that instead of minimizing \(\bar{R}_z(x, y)\) penalized by a multiple of \(\phi_z(x, y)\), we will recommend using the NBI parametrization to obtain the Pareto curve for the problem.

The earlier biobjective formulation for robustness optimization involving minimization of the expectation and the actual objective function now becomes a triobjective problem where the objectives to be minimized are \(f\), \(\bar{R}_z\) and \(\phi_z\).

11.5.1 Special case: bounds on design variables

In many interesting problems the only soft inequalities in the problem are bounds on the design variables \(x\). In such a case it may be possible to evaluate exactly the integral for \(\phi\) depending on the joint p.d.f. \(w(\xi)\). We shall find the integral assuming that the aberrations on the design variables are independent and uniformly distributed, as is the case in our numerical example.

Let the aberration in \(x_i\), denoted by \(\xi_i\) for simplicity, be distributed uniformly over the interval \([\xi_i^-, \xi_i^+]\). If all the soft inequalities are explicit bounds, say \(a_i \leq x_i \leq b_i\), \(i = 1, \ldots, \text{dim}(x)\), \(\phi_z(x)\) can be expressed as

\[
\phi(x) = \frac{1}{p} \int_{\mathcal{D}} \sum_{i=1}^{\text{dim}(x)} w(\xi) d\xi \left(\max\{0, a_i - (x_i + \xi_i)\}\right)^p + \left(\max\{0, x_i + \xi_i - b_i\}\right)^p.
\]
\[ \frac{1}{p} \sum_{i=1}^{\text{dim}(x)} \int_{\mathcal{D}} (\max\{0, a_i - (x_i + \xi_i)\})^p w(\xi) d\xi \]
\[ = \frac{1}{\xi_i^+ - \xi_i^-} \int_{\mathcal{D}} (\max\{0, a_i - (x_i + \xi_i)\})^p d\xi_i \]
\[ = \frac{1}{\xi_i^+ - \xi_i^-} \int_{\xi_i^- a_i - x_i} (a_i - x_i - \xi_i)^p d\xi \]
\[ = \begin{cases} 
\frac{1}{\xi_i^+ - \xi_i^-} \int_{\xi_i^-}^{\min(\xi_i^+, a_i - x_i)} (a_i - x_i - \xi_i)^p d\xi_i, & \text{if } x_i < a_i - \xi_i^- \\
0, & \text{if } x_i \geq a_i - \xi_i^-.
\end{cases} \]

Thus, whenever \( x_i < a_i - \xi_i^- \), the above equals
\[ \frac{1}{(p+1)(\xi_i^+ - \xi_i^-)} \left[ (a_i - x_i - \xi_i^-)^{p+1} - (a_i - x_i - \min\{\xi_i^+, a_i - x_i\})^{p+1} \right]. \] (11.5)

Similarly it can be shown that
\[ \frac{1}{(p+1)(\xi_i^+ - \xi_i^-)} \left[ (x_i + \xi_i^+ - b_i)^{p+1} - (x_i + \max\{b_i - x_i, \xi_i^-\} - b_i)^{p+1} \right], \] if \( x_i > b_i - \xi_i^+ \)
\[ = \begin{cases} 
\frac{1}{(p+1)(\xi_i^+ - \xi_i^-)} \left[ (x_i + \xi_i^+ - b_i)^{p+1} - (x_i + \max\{b_i - x_i, \xi_i^-\} - b_i)^{p+1} \right], & \text{if } x_i > b_i - \xi_i^+ \\
0, & \text{if } x_i \leq b_i - \xi_i^+.
\end{cases} \] (11.6)

Some simplifications occur if \( \xi_i^+ \geq 0 \) and \( \xi_i^- \leq 0 \). An example of such a situation is the case when \( \xi_i \) is distributed symmetrically about zero. Given that \( x_i \geq a_i \) which any feasible point satisfies,
\[ \xi_i^+ \geq 0 \Rightarrow x_i \geq a_i - \xi_i^+ \Rightarrow \min\{a_i - x_i, \xi_i^+\} = a_i - x_i. \]

Similarly, since any feasible point satisfies \( x_i \leq b_i \),
\[ \xi_i^- \leq 0 \Rightarrow x_i \leq b_i - \xi_i^- \Rightarrow \max\{b_i - x_i, \xi_i^-\} = b_i - x_i. \]
Substituting the above, 11.5 and 11.6 become

$$\int_{D} (\max\{0, a_i - (x_i + \xi_i)\})^p w(\xi) d\xi = \begin{cases} \frac{(a_i - x_i - \xi_i^-)^{p+1}}{(p+1)(\xi_i^+ - \xi_i^-)} & \text{if } x_i < a_i - \xi_i^- \\ 0 & \text{if } x_i \geq a_i - \xi_i^- \end{cases}$$

and

$$\int_{D} (\max\{0, a_i - (x_i + \xi_i)\})^p w(\xi) d\xi = \begin{cases} \frac{(x_i + \xi_i^+ - b_i)^{p+1}}{(p+1)(\xi_i^+ - \xi_i^-)} & \text{if } x_i > b_i - \xi_i^+ \\ 0 & \text{if } x_i \leq b_i - \xi_i^+ \end{cases}$$

Now the above (or their more general versions in 11.5 and 11.6) can be substituted in 11.4 to evaluate $\phi$. The value $p = 2$ was chosen in our computation.

### 11.6 A numerical example

The technique for robustness optimization of functions on constrained domains discussed in the preceding sections was applied to one of the problems arising in the design of the truss structure discussed in detail in Chapter 9. Our favorite point was picked from the trade-off curve for minimum displacement and minimum volume. This point corresponded to the point obtained by minimizing the following weighted sum of the two objectives

$$f(r) = 0.7254 \text{ square of nodal displacement } + 0.2746 \text{ total volume},$$

where

$$r = [a_1, a_2, a_3, xpos, \theta, \alpha, u_1, u_2].$$

The variable $xpos$ denotes the position of the middle bar and the rest of the symbols have the meanings described in Chapter 9.

The function minimizer of the above was found to be

$$r^* = [0.9210, 1.4511, 0.8000, 30.0000, 1.1071, 0.5880, 0.1979, 0.9099].$$

Here we are interested in the robustness minimizer of the function $f(r)$. Aberrations in $r^*$ could result either in loss of Pareto optimality of the chosen design or have the less severe effect of shifting the position of the design point to a less desirable part of the trade-off curve.

The random distribution data for the problem were as below:
• Design variable aberrations

- $\xi_{ai}$, the aberration in each cross-section: Uniform[-0.4, 0.4] sq. in.

- $\xi_{xpos}$, the aberration in the position of the middle bar: Uniform[-7, 7] ft.
  
  The aberration in the position of the middle bar is taken to arise not from fabrication or assembly errors errors, but from decisions based on aesthetics and other considerations. During the actual assembly, the designer may decide that the structure would be visually more pleasing if the middle bar was shifted by 5 ft. to one side.

• Environmental variable distributions

- Horizontal wind load: Normal(100, 15²) kips

- Suspended load: Normal(1000, 25²) kips

All the random variables were assumed to be uncorrelated.

Expression 11.2 was used to evaluate the expectation of $f(r)$ over the domain of aberrations. The minimizer of the expectation turned out to be

$$r^{R*} = [2.4749, 0.8000, 3.0000, 61.3408, 0.7743, 0.7967, 0.1895, 0.7686].$$

The trade-off curve for minimizing $f(r)$ and its expectation obtained using NBI is shown in Fig. 11.7.

It is worth observing that the effect of the aberrations in the environmental variables dominates that for the design variables. If the distributions on the environmental variables are assumed to be 0, the minimizer of the expectation turns out to be

$$[0.8665, 1.5331, 0.8000, 35.6510, 1.0347, 0.6183, 0.2153, 0.9135],$$

which deviates less from $r^*$ than $r^{R*}$. The aberrations in the loads cause the ‘displacement component’ of $f(r)$ to have large fluctuations. Hence the robust minimizer shifts to a configuration that allows for a stiffer design and suffers less nodal displacement.

The violation in soft inequalities, $\phi$, was then taken into account as the third objective. Since the only soft inequalities are the bounds on the design variables, the quantity $\phi$ was evaluated using 11.4. This triobjective problem was solved using NBI; a plot of the points obtained is shown in Fig. 11.8. This Pareto surface appears to be almost degenerate; furthermore an examination of the values of the three objectives
reveals that $\phi$ is of the order $10^{-1}$ at most of the points, which is a little too high for our purposes.

Thus we attempted to generate the Pareto set for minimizing the weighted sum $f(r)$ and its expectation subject to $\phi \leq 10^{-2}$. The trade-off curve obtained using NBI is shown in Fig. 11.9. The variation in $\phi$ along the Pareto curve is shown in Fig. 11.10. It can be observed that $\phi$ is low in the desirable ‘middle part’ of the trade-off curve, hence designating these points as good choices.
Figure 11.8  Trade-off surface for weighted sum $f(r)$, its expectation and $\phi$

Figure 11.9  Pareto curve for weighted sum and its expectation subject to $\phi \leq 0.01$
Figure 11.10  Variation in $\phi$ along Pareto curve for weighted sum and its expectation subject to $\phi \leq 0.01$
Chapter 12

Conclusions

In this dissertation, the problem of multiobjective optimization was described and analyzed and a new technique was proposed for finding Pareto sets. Finally, the robustness optimization problem was discussed from a multiobjective point of view. It can be hoped that with increases in computational power, designers and planners will begin to think more and more in terms of meeting multiple criteria rather than shirk from the idea due to lack of computational resources and techniques. As Ralph E. Steuer put it in [48]:

'We can now openly admit that a problem has multiple objectives when it possesses multiple conflicting criteria. There is no need to ignore or gloss over the fact. We can deal with multiple objectives head on, because now we have the tools to solve large-scale multiple-criteria optimization problems.'

He was of course referring to linear multicriteria optimization problems. Now, several years later, we are perhaps in a position to make a similar claim about nonlinear multicriteria optimization problems.

The work in this dissertation points to several research directions. The most important of these involves the issues arising in solving real-world, large-scale optimization problems where the functions and constraints are very expensive to evaluate. Such situations arise in aircraft design and other multi-disciplinary optimization problems. Given the computational costs involved, it may be far too expensive to solve a large number of NBI subproblems, or even to find all the function minima at the onset. The author and Dr. John E. Dennis, Jr. are involved in a collaborative project with The Boeing Company to formulate a technique which bypasses finding all the individual function minima and solving all the NBI subproblems and yields points only in the ‘interesting’ part of the Pareto set. No computation is then wasted on solving NBI subproblems which converge to the less useful peripheral part of the Pareto set.

Related to the above is the yet untested idea of using NBI in conjunction with statistical techniques like DACE (Design and Analysis of Computer Experiments, see Sacks, Welch, Mitchell and Wynn [36]) which produce approximate models of ex-
pensive multidimensional functions. DACE modeling techniques have been successfully used to model and optimize expensive functions on constrained domains (see Serafini [41], Trosset and Torczon [53]), hence such techniques seem to bear prospect for solving the NBI subproblems.

Moreover, a parallel implementation of the entire process of solving the NBI subproblems based on the parallelization strategy discussed in Section 4.3 of Chapter 5 could make the entire process much faster and applicable to larger problems.

Exploiting the special nature of the NBI subproblem and developing customized nonlinear programming techniques for solving it is also considered among future work. In particular, it is important to focus on the structure of the NBI subproblem arising in important applications like structural optimization.

Lastly, though not the least importantly, a large share of the author's immediate future work will be devoted to applying multicriteria and robustness optimization techniques developed in this dissertation to practical problems arising in industry.
Appendix A

A Bibliography of Applications

This appendix provides a list of various multicriteria optimization problems arising from practical applications in various disciplines. This is by no means exhaustive and is only meant to provide the reader an idea of the wide range of applicability of this topic.

- Design of machines and machine parts
  - Robot arm design (Eschenauer, Koski and Osyczka [11])
  - Helicopter rotor design (Ganguli and Chopra [12])
  - Truck frame design, gear unit design, metal-cutting tools design (Statnikov and Matusov [47])

- Aircraft design
  - Designing aircraft control systems (Schy and Giesy [39], Schy, Giesy and Johnson [40])
  - Optimal layout in aircraft and spacecraft structures (G. Knepe, H. Baier in Eschenauer, et. al. [11])

- Structural design
  - Reflector design for a parabolic antenna (Eschenauer, et. al. [11])
  - Truss design (Koski [20], Rakowska, Haftka and Watson [31])

- Water resources management
  - Water quality management, aimed at optimizing sheet erosion, phosphorus loading, biological oxygen demand and pollution cost (Das and Haimes [6]).
  - Stream resource allocation, aimed at minimizing total cost of waste treatment, maximizing the amount of water in the reservoir for supply and minimizing the pollution (Haimes, Hall and Freedman [14]).
- Reservoir storage capacity determination to achieve minimum cost, minimum water loss due to evaporation and maximum reservoir capacity (Reid and Vemuri [33]).

A nice survey of research work in resources planning is available in Cohon, Scavone and Solanki [4].

- Natural gas pipeline routing (Engberg, Cohon and ReVelle [8])
- Health care financial resources allocation (Urlt and Beaudry [54])
- Chemical process optimization, aimed at minimizing the total investment and the net operationg costs (Kitagawa, et. al. [19])
- Telescope Design (K. H. Stenvers in Eschenauer, et. al. [11]), multicriteria formulation involved maximizing the coefficient of thermal expansion of the material and the stiffness of the structure and minimizing shape deviations due to humidity and the total weight of the structure.
- Design of ceramic components (J. Koski and R. Silvennionen in Eschenauer, et. al. [11])
- Minimum transfer time and minimum energy expenditure in control systems (Katopis and Lin [17])
- Electrical circuit design, minimizing total power consumption and error in biasing voltage (Lin [22])
- Design of electromagnetic devices (Russenchuck [35])
- Sausage Blending (Steuer [48])
- Managerial Compensation Planning (Steuer [48])
- Cogeneration plant planning (Yokoyama and Ito [56]), minimizing total annual cost and total annual energy expenditure.
- Thermal power dispatch (Dhillon, Parti and Kothari [7])
- River basin development (Vedula and Rogers [55] for the Cauvery basin in India, Major and Lenton [26] for the Rio Colorado in Argentina)
- Bauxite mining under water supply, environmental and economic requirements (Szidarovsky, Gershon and Duckstein [51])
Appendix B

Nondominated Vectors for Truss Problem

This is a list of the non-dominated vectors obtained by using NBI on the problem of minimizing the stresses in the three bars (in ksi, i.e. kips/sq.in.) and the total volume (in cu. ft.) of the structure described in Chapter 9. Of the following, only vectors 68, 73 and 96 are efficient of order 3. This table is provided to let the reader judge the superiority of these points in comparison to the rest.

<table>
<thead>
<tr>
<th></th>
<th>Net Vol.</th>
<th>Left bar stress</th>
<th>Middle bar stress</th>
<th>Right bar stress</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>3.8761</td>
<td>2.8288</td>
<td>7.9678</td>
<td>0.0048</td>
</tr>
<tr>
<td>2.</td>
<td>1.4229</td>
<td>3.9116</td>
<td>5.7338</td>
<td>2.5321</td>
</tr>
<tr>
<td>3.</td>
<td>1.5973</td>
<td>3.3015</td>
<td>5.6587</td>
<td>2.3106</td>
</tr>
<tr>
<td>4.</td>
<td>1.8108</td>
<td>2.7247</td>
<td>5.653</td>
<td>2.1222</td>
</tr>
<tr>
<td>5.</td>
<td>2.0759</td>
<td>2.1919</td>
<td>5.7389</td>
<td>1.9779</td>
</tr>
<tr>
<td>6.</td>
<td>2.4046</td>
<td>1.7134</td>
<td>5.9378</td>
<td>1.8877</td>
</tr>
<tr>
<td>7.</td>
<td>2.8076</td>
<td>1.2981</td>
<td>6.2685</td>
<td>1.8606</td>
</tr>
<tr>
<td>8.</td>
<td>3.3143</td>
<td>0.9714</td>
<td>6.7835</td>
<td>1.922</td>
</tr>
<tr>
<td>9.</td>
<td>3.8761</td>
<td>2.8288</td>
<td>7.9678</td>
<td>0.0048</td>
</tr>
<tr>
<td>10.</td>
<td>1.4502</td>
<td>4.0173</td>
<td>5.1329</td>
<td>2.5326</td>
</tr>
<tr>
<td>11.</td>
<td>1.6207</td>
<td>3.4039</td>
<td>5.0507</td>
<td>2.3076</td>
</tr>
<tr>
<td>12.</td>
<td>1.8308</td>
<td>2.8242</td>
<td>5.039</td>
<td>2.1164</td>
</tr>
<tr>
<td>13.</td>
<td>2.0926</td>
<td>2.2886</td>
<td>5.1191</td>
<td>1.9693</td>
</tr>
<tr>
<td>14.</td>
<td>2.4177</td>
<td>1.807</td>
<td>5.3116</td>
<td>1.876</td>
</tr>
<tr>
<td>15.</td>
<td>2.8147</td>
<td>1.3867</td>
<td>5.6317</td>
<td>1.8439</td>
</tr>
<tr>
<td>16.</td>
<td>3.3544</td>
<td>1.088</td>
<td>6.2052</td>
<td>1.9333</td>
</tr>
<tr>
<td>17.</td>
<td>1.639</td>
<td>3.5019</td>
<td>4.4338</td>
<td>2.3004</td>
</tr>
<tr>
<td>18.</td>
<td>1.8459</td>
<td>2.9195</td>
<td>4.4164</td>
<td>2.1064</td>
</tr>
<tr>
<td>19.</td>
<td>2.1069</td>
<td>2.3833</td>
<td>4.4951</td>
<td>1.9586</td>
</tr>
<tr>
<td>20.</td>
<td>2.4279</td>
<td>1.8981</td>
<td>4.6803</td>
<td>1.8618</td>
</tr>
</tbody>
</table>

Table B.1  Table of nondominated objective vectors for truss problem

... continued
<table>
<thead>
<tr>
<th></th>
<th>Net Vol.</th>
<th>Left bar stress</th>
<th>Middle bar stress</th>
<th>Right bar stress</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>2.8691</td>
<td>1.5155</td>
<td>5.0789</td>
<td>1.8674</td>
</tr>
<tr>
<td>22</td>
<td>3.3928</td>
<td>1.2032</td>
<td>5.6241</td>
<td>1.9432</td>
</tr>
<tr>
<td>23</td>
<td>1.8901</td>
<td>3.0396</td>
<td>3.8454</td>
<td>2.1212</td>
</tr>
<tr>
<td>24</td>
<td>2.1575</td>
<td>2.5088</td>
<td>3.9355</td>
<td>1.9789</td>
</tr>
<tr>
<td>25</td>
<td>2.4721</td>
<td>2.0182</td>
<td>4.1093</td>
<td>1.8766</td>
</tr>
<tr>
<td>26</td>
<td>2.9275</td>
<td>1.6477</td>
<td>4.5331</td>
<td>1.8943</td>
</tr>
<tr>
<td>27</td>
<td>3.4306</td>
<td>1.3178</td>
<td>5.0417</td>
<td>1.9525</td>
</tr>
<tr>
<td>28</td>
<td>2.2079</td>
<td>2.6342</td>
<td>3.3756</td>
<td>1.999</td>
</tr>
<tr>
<td>29</td>
<td>2.5166</td>
<td>2.1386</td>
<td>3.5389</td>
<td>1.8917</td>
</tr>
<tr>
<td>30</td>
<td>2.9852</td>
<td>1.7793</td>
<td>3.9862</td>
<td>1.9206</td>
</tr>
<tr>
<td>31</td>
<td>3.4683</td>
<td>1.4324</td>
<td>4.4592</td>
<td>1.9618</td>
</tr>
<tr>
<td>32</td>
<td>2.5614</td>
<td>2.2592</td>
<td>2.969</td>
<td>1.907</td>
</tr>
<tr>
<td>33</td>
<td>3.0423</td>
<td>1.9104</td>
<td>3.4381</td>
<td>1.9463</td>
</tr>
<tr>
<td>34</td>
<td>3.5063</td>
<td>1.5472</td>
<td>3.8773</td>
<td>1.9713</td>
</tr>
<tr>
<td>35</td>
<td>3.4959</td>
<td>1.6208</td>
<td>3.2094</td>
<td>1.9396</td>
</tr>
<tr>
<td>36</td>
<td>3.5263</td>
<td>1.7291</td>
<td>2.6139</td>
<td>1.9427</td>
</tr>
<tr>
<td>37</td>
<td>3.8761</td>
<td>2.8288</td>
<td>7.9678</td>
<td>0.0048</td>
</tr>
<tr>
<td>38</td>
<td>1.4045</td>
<td>4.1631</td>
<td>5.7725</td>
<td>2.2633</td>
</tr>
<tr>
<td>39</td>
<td>1.5706</td>
<td>3.5459</td>
<td>5.6827</td>
<td>2.0347</td>
</tr>
<tr>
<td>40</td>
<td>1.7731</td>
<td>2.9597</td>
<td>5.6575</td>
<td>1.837</td>
</tr>
<tr>
<td>41</td>
<td>2.0235</td>
<td>2.4144</td>
<td>5.7173</td>
<td>1.6801</td>
</tr>
<tr>
<td>42</td>
<td>2.3329</td>
<td>1.9194</td>
<td>5.8819</td>
<td>1.5734</td>
</tr>
<tr>
<td>43</td>
<td>2.7095</td>
<td>1.4817</td>
<td>6.1658</td>
<td>1.524</td>
</tr>
<tr>
<td>44</td>
<td>3.1591</td>
<td>1.1062</td>
<td>6.5794</td>
<td>1.5367</td>
</tr>
<tr>
<td>45</td>
<td>3.7258</td>
<td>0.8305</td>
<td>7.2008</td>
<td>1.6491</td>
</tr>
<tr>
<td>46</td>
<td>1.5923</td>
<td>3.6468</td>
<td>5.0718</td>
<td>2.0303</td>
</tr>
<tr>
<td>47</td>
<td>1.7911</td>
<td>3.0575</td>
<td>5.0399</td>
<td>1.8294</td>
</tr>
<tr>
<td>48</td>
<td>2.0382</td>
<td>2.5094</td>
<td>5.0939</td>
<td>1.6697</td>
</tr>
<tr>
<td>49</td>
<td>2.3447</td>
<td>2.0119</td>
<td>5.2534</td>
<td>1.5607</td>
</tr>
<tr>
<td>50</td>
<td>2.7186</td>
<td>1.5718</td>
<td>5.5325</td>
<td>1.5089</td>
</tr>
</tbody>
</table>

... continued
<table>
<thead>
<tr>
<th></th>
<th>Net Vol.</th>
<th>Left bar stress</th>
<th>Middle bar stress</th>
<th>Right bar stress</th>
</tr>
</thead>
<tbody>
<tr>
<td>51.</td>
<td>3.213</td>
<td>1.2346</td>
<td>6.0256</td>
<td>1.5597</td>
</tr>
<tr>
<td>52.</td>
<td>3.7559</td>
<td>0.9387</td>
<td>6.6049</td>
<td>1.6519</td>
</tr>
<tr>
<td>53.</td>
<td>1.8307</td>
<td>3.1737</td>
<td>4.4608</td>
<td>1.8403</td>
</tr>
<tr>
<td>54.</td>
<td>2.0915</td>
<td>2.6373</td>
<td>4.5391</td>
<td>1.6923</td>
</tr>
<tr>
<td>55.</td>
<td>2.3949</td>
<td>2.1371</td>
<td>4.6931</td>
<td>1.5806</td>
</tr>
<tr>
<td>56.</td>
<td>2.7499</td>
<td>1.681</td>
<td>4.9386</td>
<td>1.5127</td>
</tr>
<tr>
<td>57.</td>
<td>3.2679</td>
<td>1.3639</td>
<td>5.4737</td>
<td>1.5837</td>
</tr>
<tr>
<td>58.</td>
<td>3.7852</td>
<td>1.6461</td>
<td>6.0075</td>
<td>1.654</td>
</tr>
<tr>
<td>59.</td>
<td>2.1486</td>
<td>2.7683</td>
<td>3.991</td>
<td>1.7181</td>
</tr>
<tr>
<td>60.</td>
<td>2.4468</td>
<td>2.2638</td>
<td>4.1358</td>
<td>1.6019</td>
</tr>
<tr>
<td>61.</td>
<td>2.7935</td>
<td>1.8006</td>
<td>4.3666</td>
<td>1.527</td>
</tr>
<tr>
<td>62.</td>
<td>3.3221</td>
<td>1.4924</td>
<td>4.9204</td>
<td>1.6069</td>
</tr>
<tr>
<td>63.</td>
<td>3.8024</td>
<td>1.1432</td>
<td>5.3886</td>
<td>1.6458</td>
</tr>
<tr>
<td>64.</td>
<td>2.4989</td>
<td>2.3907</td>
<td>3.5789</td>
<td>1.6234</td>
</tr>
<tr>
<td>65.</td>
<td>2.838</td>
<td>1.9209</td>
<td>3.7962</td>
<td>1.542</td>
</tr>
<tr>
<td>66.</td>
<td>3.3757</td>
<td>1.6206</td>
<td>4.3662</td>
<td>1.6298</td>
</tr>
<tr>
<td>67.</td>
<td>3.8256</td>
<td>1.2454</td>
<td>4.7804</td>
<td>1.6428</td>
</tr>
<tr>
<td>68.</td>
<td>2.9024</td>
<td>2.0583</td>
<td>3.2611</td>
<td>1.574</td>
</tr>
<tr>
<td>69.</td>
<td>3.4291</td>
<td>1.7486</td>
<td>3.8116</td>
<td>1.6524</td>
</tr>
<tr>
<td>70.</td>
<td>3.8508</td>
<td>1.3493</td>
<td>4.1757</td>
<td>1.6414</td>
</tr>
<tr>
<td>71.</td>
<td>3.4825</td>
<td>1.8765</td>
<td>3.2569</td>
<td>1.675</td>
</tr>
<tr>
<td>72.</td>
<td>3.878</td>
<td>1.4549</td>
<td>3.5745</td>
<td>1.6417</td>
</tr>
<tr>
<td>73.</td>
<td>3.941</td>
<td>1.591</td>
<td>3.0369</td>
<td>1.6725</td>
</tr>
<tr>
<td>74.</td>
<td>1.55</td>
<td>3.7955</td>
<td>5.7175</td>
<td>1.7639</td>
</tr>
<tr>
<td>75.</td>
<td>1.7433</td>
<td>3.2015</td>
<td>5.676</td>
<td>1.5584</td>
</tr>
<tr>
<td>76.</td>
<td>1.9818</td>
<td>2.646</td>
<td>5.7147</td>
<td>1.3914</td>
</tr>
<tr>
<td>77.</td>
<td>2.2763</td>
<td>2.1383</td>
<td>5.8529</td>
<td>1.2721</td>
</tr>
<tr>
<td>78.</td>
<td>2.6345</td>
<td>1.6849</td>
<td>6.104</td>
<td>1.2069</td>
</tr>
<tr>
<td>79.</td>
<td>3.0585</td>
<td>1.2876</td>
<td>6.4722</td>
<td>1.1979</td>
</tr>
<tr>
<td>80.</td>
<td>1.775</td>
<td>3.3109</td>
<td>5.0828</td>
<td>1.5625</td>
</tr>
<tr>
<td></td>
<td>Net Vol.</td>
<td>Left bar stress</td>
<td>Middle bar stress</td>
<td>Right bar stress</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>----------------</td>
<td>-------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>81.</td>
<td>2.0325</td>
<td>2.7716</td>
<td>5.1552</td>
<td>1.4117</td>
</tr>
<tr>
<td>82.</td>
<td>2.3286</td>
<td>2.2653</td>
<td>5.2962</td>
<td>1.2938</td>
</tr>
<tr>
<td>83.</td>
<td>2.6705</td>
<td>1.798</td>
<td>5.5184</td>
<td>1.2147</td>
</tr>
<tr>
<td>84.</td>
<td>3.0675</td>
<td>1.3777</td>
<td>5.8387</td>
<td>1.1827</td>
</tr>
<tr>
<td>85.</td>
<td>3.6262</td>
<td>1.0953</td>
<td>6.446</td>
<td>1.2883</td>
</tr>
<tr>
<td>86.</td>
<td>2.0974</td>
<td>2.9094</td>
<td>4.621</td>
<td>1.4441</td>
</tr>
<tr>
<td>87.</td>
<td>2.3892</td>
<td>2.3994</td>
<td>4.7545</td>
<td>1.3226</td>
</tr>
<tr>
<td>88.</td>
<td>2.7236</td>
<td>1.9257</td>
<td>4.9634</td>
<td>1.2372</td>
</tr>
<tr>
<td>89.</td>
<td>3.1169</td>
<td>1.5023</td>
<td>5.277</td>
<td>1.202</td>
</tr>
<tr>
<td>90.</td>
<td>3.6767</td>
<td>1.2207</td>
<td>5.8861</td>
<td>1.3084</td>
</tr>
<tr>
<td>91.</td>
<td>2.4496</td>
<td>2.5334</td>
<td>4.2123</td>
<td>1.3511</td>
</tr>
<tr>
<td>92.</td>
<td>2.7774</td>
<td>2.054</td>
<td>4.4095</td>
<td>1.2601</td>
</tr>
<tr>
<td>93.</td>
<td>3.1815</td>
<td>1.6397</td>
<td>4.7422</td>
<td>1.2341</td>
</tr>
<tr>
<td>94.</td>
<td>3.7267</td>
<td>1.3458</td>
<td>5.3256</td>
<td>1.3282</td>
</tr>
<tr>
<td>95.</td>
<td>2.8317</td>
<td>2.1828</td>
<td>3.8565</td>
<td>1.2835</td>
</tr>
<tr>
<td>96.</td>
<td>3.246</td>
<td>1.7772</td>
<td>4.2073</td>
<td>1.2662</td>
</tr>
<tr>
<td>97.</td>
<td>3.777</td>
<td>1.4711</td>
<td>4.7654</td>
<td>1.3482</td>
</tr>
<tr>
<td>98.</td>
<td>3.3104</td>
<td>1.9145</td>
<td>3.6722</td>
<td>1.2982</td>
</tr>
<tr>
<td>99.</td>
<td>3.8278</td>
<td>1.5968</td>
<td>4.2061</td>
<td>1.3686</td>
</tr>
<tr>
<td>100.</td>
<td>3.8792</td>
<td>1.723</td>
<td>3.6479</td>
<td>1.3895</td>
</tr>
<tr>
<td>101.</td>
<td>1.7208</td>
<td>3.4495</td>
<td>5.7075</td>
<td>1.2861</td>
</tr>
<tr>
<td>102.</td>
<td>1.9781</td>
<td>2.91</td>
<td>5.7794</td>
<td>1.135</td>
</tr>
<tr>
<td>103.</td>
<td>2.2708</td>
<td>2.4008</td>
<td>5.9145</td>
<td>1.0142</td>
</tr>
<tr>
<td>104.</td>
<td>2.604</td>
<td>1.926</td>
<td>6.1212</td>
<td>0.9278</td>
</tr>
<tr>
<td>105.</td>
<td>2.9927</td>
<td>1.4987</td>
<td>6.4268</td>
<td>0.8887</td>
</tr>
<tr>
<td>106.</td>
<td>2.0528</td>
<td>3.0561</td>
<td>5.2627</td>
<td>1.1758</td>
</tr>
<tr>
<td>107.</td>
<td>2.3423</td>
<td>2.5442</td>
<td>5.392</td>
<td>1.0523</td>
</tr>
<tr>
<td>108.</td>
<td>2.6691</td>
<td>2.064</td>
<td>5.5875</td>
<td>0.9604</td>
</tr>
<tr>
<td>109.</td>
<td>3.0398</td>
<td>1.6212</td>
<td>5.8609</td>
<td>0.9059</td>
</tr>
<tr>
<td>110.</td>
<td>3.5257</td>
<td>1.2767</td>
<td>6.339</td>
<td>0.9496</td>
</tr>
</tbody>
</table>

... continued
<table>
<thead>
<tr>
<th></th>
<th>Net Vol.</th>
<th>Left bar stress</th>
<th>Middle bar stress</th>
<th>Right bar stress</th>
</tr>
</thead>
<tbody>
<tr>
<td>111</td>
<td>2.4126</td>
<td>2.6865</td>
<td>4.8674</td>
<td>1.0393</td>
</tr>
<tr>
<td>112</td>
<td>2.734</td>
<td>2.2018</td>
<td>5.0533</td>
<td>0.9929</td>
</tr>
<tr>
<td>113</td>
<td>3.0957</td>
<td>1.7513</td>
<td>5.3107</td>
<td>0.9307</td>
</tr>
<tr>
<td>114</td>
<td>3.5376</td>
<td>1.3693</td>
<td>5.7107</td>
<td>0.9369</td>
</tr>
<tr>
<td>115</td>
<td>2.7987</td>
<td>2.3393</td>
<td>4.5186</td>
<td>1.025</td>
</tr>
<tr>
<td>116</td>
<td>3.1525</td>
<td>1.8822</td>
<td>4.7622</td>
<td>0.9562</td>
</tr>
<tr>
<td>117</td>
<td>3.6024</td>
<td>1.507</td>
<td>5.1763</td>
<td>0.9692</td>
</tr>
<tr>
<td>118</td>
<td>3.21</td>
<td>2.0136</td>
<td>4.2148</td>
<td>0.9823</td>
</tr>
<tr>
<td>119</td>
<td>3.6672</td>
<td>1.6446</td>
<td>4.6418</td>
<td>1.0015</td>
</tr>
<tr>
<td>120</td>
<td>3.7318</td>
<td>1.7821</td>
<td>4.1071</td>
<td>1.0337</td>
</tr>
<tr>
<td>121</td>
<td>2.0125</td>
<td>3.2065</td>
<td>5.912</td>
<td>0.9112</td>
</tr>
<tr>
<td>122</td>
<td>2.3037</td>
<td>2.696</td>
<td>6.0443</td>
<td>0.789</td>
</tr>
<tr>
<td>123</td>
<td>2.6274</td>
<td>2.2131</td>
<td>6.2342</td>
<td>0.6945</td>
</tr>
<tr>
<td>124</td>
<td>2.9873</td>
<td>1.7612</td>
<td>6.4885</td>
<td>0.6309</td>
</tr>
<tr>
<td>125</td>
<td>2.3865</td>
<td>2.849</td>
<td>5.5418</td>
<td>0.8367</td>
</tr>
<tr>
<td>126</td>
<td>2.7064</td>
<td>2.3629</td>
<td>5.7251</td>
<td>0.739</td>
</tr>
<tr>
<td>127</td>
<td>3.0593</td>
<td>1.905</td>
<td>5.9669</td>
<td>0.6693</td>
</tr>
<tr>
<td>128</td>
<td>3.5293</td>
<td>1.5469</td>
<td>6.4167</td>
<td>0.6994</td>
</tr>
<tr>
<td>129</td>
<td>2.7831</td>
<td>2.5108</td>
<td>5.2119</td>
<td>0.7815</td>
</tr>
<tr>
<td>130</td>
<td>3.1302</td>
<td>2.0478</td>
<td>5.4434</td>
<td>0.7069</td>
</tr>
<tr>
<td>131</td>
<td>3.5127</td>
<td>1.6152</td>
<td>5.7379</td>
<td>0.6625</td>
</tr>
<tr>
<td>132</td>
<td>3.2</td>
<td>2.1898</td>
<td>4.918</td>
<td>0.7435</td>
</tr>
<tr>
<td>133</td>
<td>3.5743</td>
<td>1.7502</td>
<td>5.1978</td>
<td>0.6921</td>
</tr>
<tr>
<td>134</td>
<td>3.6363</td>
<td>1.8854</td>
<td>4.6583</td>
<td>0.722</td>
</tr>
<tr>
<td>135</td>
<td>2.3693</td>
<td>3.0191</td>
<td>6.2323</td>
<td>0.5917</td>
</tr>
<tr>
<td>136</td>
<td>2.6924</td>
<td>2.5357</td>
<td>6.4211</td>
<td>0.4967</td>
</tr>
<tr>
<td>137</td>
<td>3.0412</td>
<td>2.0743</td>
<td>6.6557</td>
<td>0.4236</td>
</tr>
<tr>
<td>138</td>
<td>2.7839</td>
<td>2.6962</td>
<td>5.9342</td>
<td>0.5518</td>
</tr>
<tr>
<td>139</td>
<td>3.1297</td>
<td>2.2322</td>
<td>6.1634</td>
<td>0.4761</td>
</tr>
<tr>
<td>140</td>
<td>3.5641</td>
<td>1.8437</td>
<td>6.5499</td>
<td>0.4759</td>
</tr>
</tbody>
</table>

... continued
<table>
<thead>
<tr>
<th>Net Vol.</th>
<th>Left bar stress</th>
<th>Middle bar stress</th>
<th>Right bar stress</th>
</tr>
</thead>
<tbody>
<tr>
<td>141.</td>
<td>3.2135</td>
<td>2.386</td>
<td>5.6627</td>
</tr>
<tr>
<td>142.</td>
<td>3.5802</td>
<td>1.9399</td>
<td>5.9291</td>
</tr>
<tr>
<td>143.</td>
<td>3.6561</td>
<td>2.0871</td>
<td>5.4145</td>
</tr>
<tr>
<td>144.</td>
<td>2.8</td>
<td>2.8946</td>
<td>6.6836</td>
</tr>
<tr>
<td>145.</td>
<td>3.1495</td>
<td>2.4338</td>
<td>6.9195</td>
</tr>
<tr>
<td>146.</td>
<td>3.2489</td>
<td>2.601</td>
<td>6.4465</td>
</tr>
<tr>
<td>147.</td>
<td>3.6156</td>
<td>2.1548</td>
<td>6.7128</td>
</tr>
<tr>
<td>148.</td>
<td>3.7044</td>
<td>2.3129</td>
<td>6.2211</td>
</tr>
<tr>
<td>149.</td>
<td>3.3054</td>
<td>2.8338</td>
<td>7.2676</td>
</tr>
<tr>
<td>150.</td>
<td>3.7777</td>
<td>2.4777</td>
<td>7.7216</td>
</tr>
<tr>
<td>151.</td>
<td>3.7785</td>
<td>2.5608</td>
<td>7.0735</td>
</tr>
<tr>
<td>152.</td>
<td>3.8761</td>
<td>2.8288</td>
<td>7.9678</td>
</tr>
</tbody>
</table>
Bibliography


