Robustness Optimization for Constrained, Nonlinear Programming Problems*

Indraneel Das†

March 1997

Abstract

In realistic situations engineering designs should take into consideration random aberrations from the stipulated design variables arising from manufacturing variability. Moreover, many environmental parameters are often stochastic in nature. Traditional nonlinear optimization attempts to find a deterministic optimum of a cost function and does not take into account the effect of these random variations on the objective. This paper attempts to device a technique for finding optima of constrained nonlinear functions that are robust with respect to such variations.

The expectation of the function over a domain of aberrations in the parameters is taken as a measure of 'robustness' of the function value at a point. It is pointed out that robustness optimization is ideally an attempt to trade off between 'optimality' and 'robustness'. A newly-developed multicriteria optimization technique known as Normal-Boundary Intersection is used here to find evenly-spaced points on the Pareto curve for the 'optimality' and 'robustness' criteria. This Pareto curve enables the user to make the trade-off decision explicitly, free of arbitrary 'weighting' parameters.

This paper also formulates a derivative-based approximation for evaluating the expected value of the objective function on the nonlinear manifold defined by the state equations for the system. Existing procedures for evaluating the expectation usually involve numerical integration techniques requiring many solutions of the state equations for one evaluation of the expectation. The procedure presented here bypasses the need for multiple solutions of the state equations and hence provides a cheaper and more easily optimizable approximation to the expectation. Finally, this paper discusses how nonlinear inequality constraints should be treated in the presence of random parameters in the design. Computational results are presented for finding a robust optimum of a structural optimization problem.

*This research was partially supported by the Dept. of Energy, DOE Grant DE-FG03-95ER25257 and the Air Force, Air Force Grant F49620-95-1-0210

†Graduate Student, Dept. of Computational & Applied Mathematics, Rice University, Houston, TX 77005-1892. indra@rice.edu.

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 and environmental conditions will hold as specified, 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 story in a particular multi-story 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-story 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 random variations in such environmental factors.

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 [7] call a 'proactive' approach, i.e. one where the optimization process is designed to yield a robust solution.

2 The robustness optimization problem

Consider the unconstrained optimization problem

$$\min_x f(x)$$

where $f : \Re^N \mapsto \Re$ 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 3.3 for ease of presentation.

If x^* is the chosen optimal design 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 $\mathcal{D}(x) = \{\xi : \xi \in \Re^N, |x_i + \xi_i| \leq \alpha_i, i = 1, 2, \ldots, n\}$ and the aberration ξ_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:

$$\min_x \bar{R}(x) = \frac{\int_{\mathcal{D}} f(x+\xi)d\xi}{\int_{\mathcal{D}} d\xi}.$$

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 ξ is distributed uniformly over the hypercuboid $\mathcal{D}$. Since there is no reason to restrict ourselves to the uniform distribution, we generalize the expectation minimization problem given any smooth distribution as below:

$$\min_x \bar{R}(x) = \int_{\mathcal{D}} f(x+\xi)w(\xi)d\xi\,,$$

where $w(\xi)$ is the multidimensional probability density function of ξ over the domain $\mathcal{D}$ (so $\int_{\mathcal{D}} w(\xi)d\xi = 1$). Observe that the domain $\mathcal{D}$ does not need to be bounded, e.g. if $w(\xi)$ is the multinormal distribution, the domain is all of $\Re^n$. 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 classically been studied in the setting of a linear program, as in Mulvey, Vanderbei and Zenios [7] and Mayer [6]. 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. [3]. The expectation minimization approach has been referred to in Slowinski and Teghem [8], Stancu-Minasian [9] and Ermoliev and Wets [4] among many others. We shall endeavor to handle nonlinear problems in the robust setting as well.

2.1 Evaluating the multidimensional integral

Since the integral for evaluating the expectation may not be available analytically in closed form, especially if $\dim(x)$ 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 $\dim(x)$ 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,

$$\bar{R}(x) = \int_{\mathcal{D}} f(x+\xi) w(\xi) d\xi$$

$$\approx \int_{\mathcal{D}} (f(x) + \nabla_x f(x)^T \xi + \frac{1}{2} \xi^T \nabla_x^2 f(x) \xi)\, w(\xi) d\xi\,.$$

Clearly,

$$\int_{\mathcal{D}} f(x)\, w(\xi) d\xi = f(x) \int_{\mathcal{D}} w(\xi) d\xi = f(x)\,.$$

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

$$\int_{\mathcal{D}} \nabla_x f(x)^T \xi\, w(\xi) d\xi$$

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

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

where μ is the vector of expectations of the aberrations ξ_i.
Further

$$\int_{\mathcal{D}} \frac{1}{2} \xi^T \nabla_x^2 f(x) \xi\, w(\xi)\, d\xi$$

$$= \int_{\mathcal{D}} (\sum_{1 \le i < j \le n} \xi_i \xi_j h_{ij} + \frac{1}{2} \sum_{i=1}^{n} \xi_i^2 h_{ii})\, w(\xi)\, d\xi$$

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

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

$$\text{and } \int_{\mathcal{D}} \xi_i^2\, w(\xi)\, d\xi = E(\xi_i^2) = \sigma_i^2 + \mu_i^2$$

where ρ_{ij} is the correlation coefficient between ξ_i and ξ_j and σ_i is the standard deviation of ξ_i. Now the integral of the second order term can be expressed in terms of moments (upto second order) of the joint distribution as

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

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

$$\int_{\mathcal{D}} \frac{1}{2} \xi^T H \xi\, w(\xi)\, d\xi = \frac{1}{2} s^T (H * R) s + \mu^T H \mu \qquad (1)$$

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

Thus the integral can be approximated as

$$\bar{R}(x) = \int_{\mathcal{D}} f(x+\xi) w(\xi) d\xi$$

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

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

$$\bar{R}(x) \approx f(x) + \nabla_x f(x)^T \mu + \sum_{1 \le i < j \le n} h_{ij}\mu_i\mu_j + \frac{1}{2}\sum_{i=1}^{n} h_{ii}(\sigma_i^2 + \mu_i^2)\,.$$

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

$$\bar{R}(x) \approx f(x) + \sum_{1 \le i < j \le n} h_{ij}\rho_{ij}\sigma_i\sigma_j + \frac{1}{2}\sum_{i=1}^{n} h_{ii}\sigma_i^2\,.$$

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. 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.

3 Trade-off inherent in the robustness problem

Simply minimizing the expectation has its drawbacks, as shown in the examples in Fig. 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 ξ in the example problem earlier had been replaced by a normal distribution with mean 0 and variance

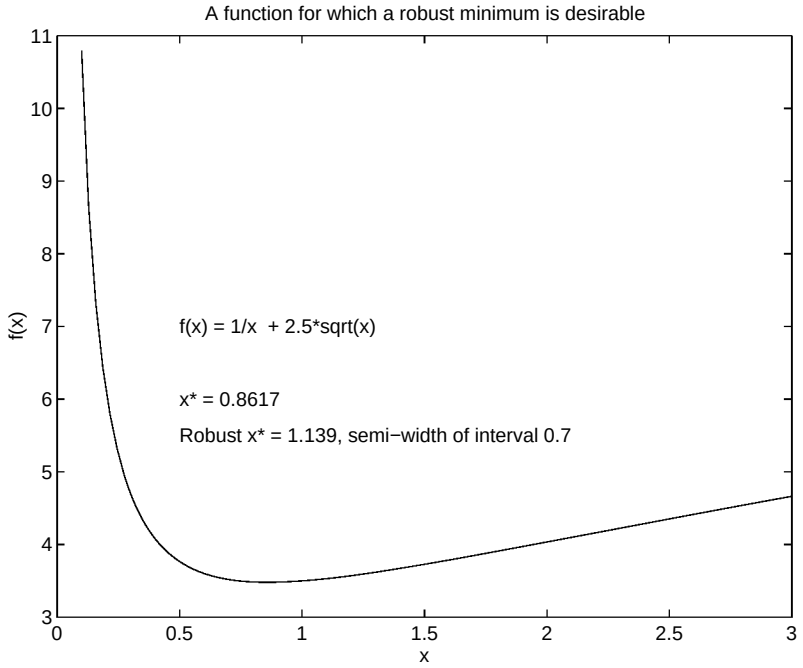

Figure 1: The 'flatter' side of the minimum is more robust w.r.t. optimality

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. 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 [8], 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. 3.

It can thus be seen that robustness optimization is ideally 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

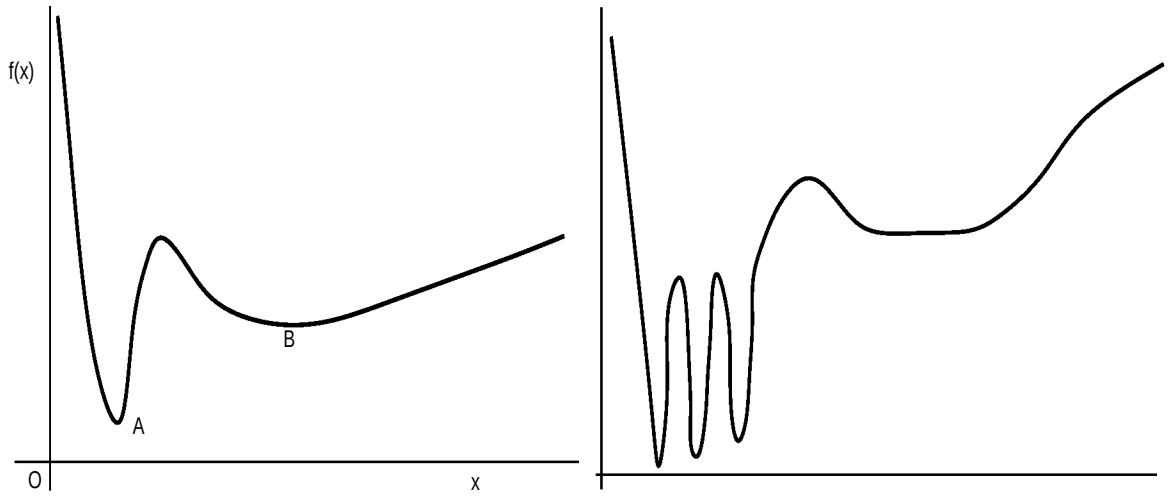

Figure 2: Drawbacks of choosing minimizer of expectation as final design

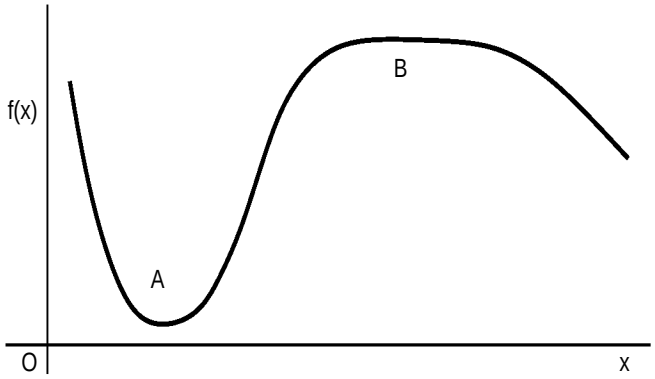

Figure 3: Drawback of minimizing dispersion in function values

- 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 [10] penalize their original objective by a term which they call the sensitivity index. This is nothing but a crude estimate of the expectation, amounting overall to minimizing a linear combination of the function and its expectation. Though this approach yields a Pareto optimal point[1] for the bicriteria optimization problem of minimizing the function and its expectation, it gives the user no indication as to whether the point falls in a 'desirable' part of the curve of Pareto optimal solutions. To remedy this, Mulvey, Vanderbei and Zenios [7] plot the 'efficient frontier', i.e., the Pareto curve, for a robust version of a problem called AFIRO from NETLIB linear programs using a similar penalty parameter approach. Box and Jones [1] minimize a convex

[1] Pareto optimality is a concept of optimality used in multicriteria optimization defined in Appendix A.

combination of the expectation and variance for different parameter settings to find several Pareto optimal points for the biobjective optimization problem. 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 aimed applications arising chiefly from structural design. Approximating a measure of dispersion in our nonlinear setting can be stifled in the absence of higher order moment information on the distribution of ξ, as shown in the next subsection.

A problem exists with the idea of minimizing weighted sums of objectives over various parameter settings for generating a point-wise approximation to the Pareto curve; it usually fails to generate an even spread of points on the Pareto curve given an even spread of parameter settings. Often 'highly biased' clusters of points are generated which give the user a very poor model of the trade-off curve. A detailed investigation of this phenomenon can be found in Das [2]. Our approach employs Normal-Boundary Intersection (NBI) as the tool for biobjective optimization. NBI introduces a sophisticated parametrization of the Pareto set which generates evenly-spaced points on the trade-off surface and has various other appealing properties. A brief description of NBI is included in Appendix A for the convenience of the reader; a full description can be found in Das [2]. More information can also be found on the NBI homepage at http://www.owlnet.rice.edu/~indra.

3.1 Minimizing dispersion as an alternative

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

$$V(x) = \int_{\mathcal{D}} (f(x+\xi) - f(x))^2 \, w(\xi) \, d\xi$$

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

$$V(x) \approx \int_{\mathcal{D}} (\nabla_x f(x)^T \xi + \frac{1}{2} \xi^T \nabla_x^2 f(x) \xi)^2 \, w(\xi) \, d\xi \, .$$

When expanded, the integrand will clearly involve upto fourth order terms in ξ, 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., 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 [10], then the approximation for the integrand becomes

$$V(x) \approx \int_{\mathcal{D}} (\nabla_x f(x)^T \xi)^2 \, w(\xi) \, d\xi \, .$$

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 which involve minimizing dispersion. This leaves us with one alternative, that of minimizing expectation and minimizing the original objective.

3.2 Robustness versus optimality trade-off curves

The Pareto curves for minimizing expectation and minimizing the original function value using NBI for our earlier single variable problem are shown in Fig. 4 (using exact integral for expectation) and Fig. 5 (using the approximation).

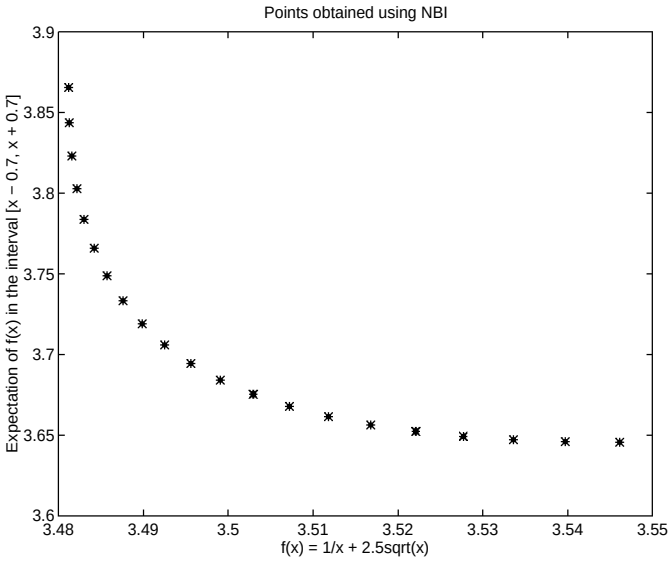

Figure 4: Pareto set for minimizing exact expectation and original objective

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.

3.3 Environmental Variables

Following Box and Jones [1] and many others in the statistical literature, we introduce the concept of environmental variables. These are the variables which have random aberrations

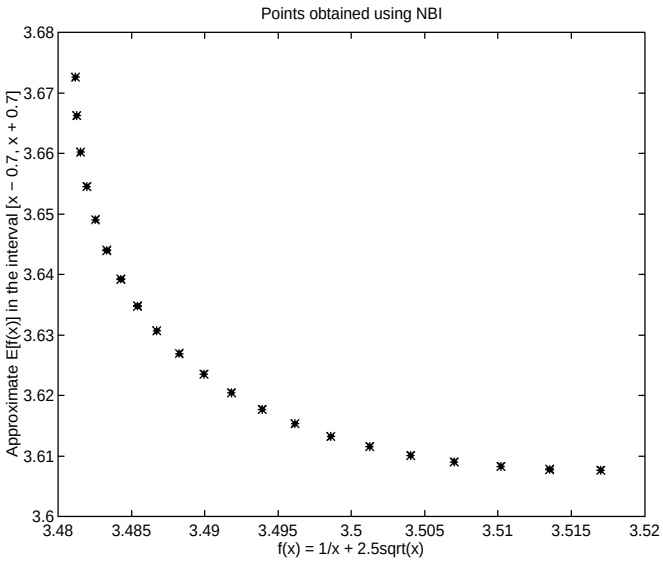

Figure 5: Pareto set for minimizing approximate expectation and original objective

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 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

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

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

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

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. [7]. 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. In other words, the state equations implicitly define a function $\psi : (x, z) \mapsto y$, i.e. $y = \psi(x, z)$. Thus given the aberrations ξ_x and ξ_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

$$\min_{x,y} \; f(x, y, z)$$

$$\text{s.t.} \qquad C(x, y, z) = 0\,,$$

or equivalently

$$\min_{x} f(x, \psi(x, z), z)\,,$$

then the expected cost minimization version of the above is

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

4.1 Approximating the expectation integral for constrained problems

Since it is unlikely that the function ψ 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 $\frac{\partial \psi_i}{\partial x_j}$ 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:

$$f(x + \xi_x, \psi(x + \xi_x, z + \xi_z), z + \xi_z)$$

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

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

$$f(x,\psi(x,z),z) + \nabla_x f^T \xi_x + \nabla_y f^T (\nabla_x \psi^T \xi_x + \nabla_z \psi^T \xi_z) + \nabla_z f^T \xi_z$$

$$+ \text{ second order terms } + \text{ higher order terms } .$$

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:

$$\int_{\mathcal{D}} (f(x,\psi(x,z),z) + \nabla_x f^T \xi_x + \nabla_y f^T (\nabla_x \psi^T \xi_x + \nabla_z \psi^T \xi_z) + \nabla_z f^T \xi_z \,) \, w(\xi_x, \xi_z) \, d\xi_x \, d\xi_z$$

$$= f(x,\psi(x,z),z) + \nabla_x f^T \mu_x + \nabla_z f^T \mu_z + \nabla_y f^T \nabla_x \psi^T \mu_x + \nabla_y f^T \nabla_z \psi^T \mu_z \, ,$$

where μ_x and μ_z are the mean vectors of ξ_x and ξ_z respectively.

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

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

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 ξ_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

$$\begin{aligned} \mathcal{H}_1 &= \nabla_x^2 f + H_{xy} \nabla_x \psi^T + \nabla_x \psi H_{xy}^T + \nabla_x \psi \nabla_y^2 f \nabla_x \psi^T \, , \\ \mathcal{H}_2 &= \nabla_z^2 f + H_{yz}^T \nabla_z \psi^T + \nabla_z \psi H_{yz} + \nabla_z \psi \nabla_y^2 f \nabla_z \psi^T \, , \\ \mathcal{H}_3 &= H_{xz} + H_{xy} \nabla_z \psi^T + \nabla_x \psi H_{yz} + \nabla_x \psi \nabla_y^2 f \nabla_z \psi^T \, . \end{aligned}$$

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 (1) can be applied directly, yielding

$$\int_{\mathcal{D}} \frac{1}{2} \xi_x^T \mathcal{H}_1 \xi_x \; = \; \frac{1}{2} s_x^T (\mathcal{H}_1 * R_x) s_x + \mu_x^T \mathcal{H}_1 \mu_x$$

and

$$\int_{\mathcal{D}} \frac{1}{2} \xi_z^T \mathcal{H}_2 \xi_z \; = \; \frac{1}{2} s_z^T (\mathcal{H}_2 * R_z) s_z + \mu_z^T \mathcal{H}_2 \mu_z \, .$$

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 ξ_x and ξ_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

$$\bar{R}_z(x) \approx f(x,\psi(x,z),z) + \nabla_x f^T \mu_x + \nabla_z f^T \mu_z + \nabla_y f^T \nabla_x \psi^T \mu_x + \nabla_y f^T \nabla_z \psi^T \mu_z$$
$$+ \frac{1}{2} s_x^T (\mathcal{H}_1 * R_x) s_x + \mu_x^T \mathcal{H}_1 \mu_x + \frac{1}{2} s_z^T (\mathcal{H}_2 * R_z) s_z + \mu_z^T \mathcal{H}_2 \mu_z \,. \qquad (2)$$

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 $r \in \{x_i : i = 1,\ldots,\dim(x)\}$ in order gives

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

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 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

$$\min_{x,y} \ \bar{R}_z(x,y)$$

$$\text{s.t.} \quad C(x,y,z) = 0 \, .$$

5 Handling inequality constraints

Our previous formulation will now be extended to handle inequality constraints $g(x,y,z) \leq 0$ (which includes explicit bounds on variables). The standard optimization problem constrained by equalities and inequalities now is

$$\min_{x,y} \ f(x,y,z)$$

$$\text{s.t.} \quad C(x,y,z) = 0$$

$$g(x,y,z) \leq 0 \, .$$

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 ξ_x and ξ_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

$$g^h(x,y,z) = -\frac{1}{2} v * v \tag{3}$$

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

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

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

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 proves to be very useful for our formulation.

Just as it was assumed for the equality-constrained case that there exists a function $\psi : (x, z) \mapsto y$, it will be assumed here that there exists a function $\psi_g(x, z) \mapsto (y, v)$ which is implicitly defined by the set of equalities

$$C(x, y, z) = 0$$

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

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 ≥ 0.8 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 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. (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 [7] call *model robustness*, i.e. the (soft inequality) constraints staying 'almost feasible' for all aberred scenarios 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 ϕ, is proposed:

$$\phi_z(x) = \int_{\mathcal{D}} \tau_p(g^s(x + \xi_x, \psi_g(x + \xi_x, z + \xi_z), z + \xi_z)) \, w(\xi_x, \xi_z) \, d\xi_x \, d\xi_z$$

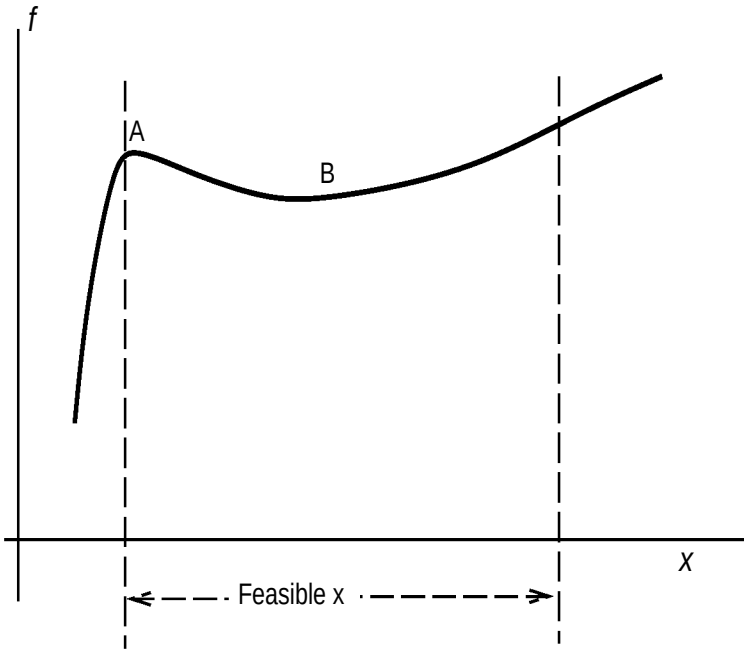

Figure 6: Expectation of f can be lower in a neighborhood of A than in the same for B, but solely because infeasible x values map to smaller f values and contribute to a smaller value of the expectation.

where

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

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 ϕ measures this violation over a region.

Let us consider evaluating or approximating the integral for ϕ. If $p = 1$, which is along the lines of Mulvey, et. al. [7] τ_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 have chosen to avoid from the start.

For $p = 2$, τ_p is both continuous and differentiable everywhere and

$$\nabla_x \tau_2 = \sum_i \max\{0, g_i^s\} \nabla_x g_i^s \, .$$

However, second derivatives of τ_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 likely to be the case most often). 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 easily shown that

$$\nabla_x \tau_4 = \sum_i (\max\{0,\, g_i^s\})^3 \nabla_x g_i^s$$

and

$$\nabla_x^2 \tau_4 = \sum_i (\max\{0,\, g_i^s\})^3 \nabla_x^2 g_i^s + 3 \sum_i (\max\{0,\, g_i^s\})^2 \nabla_x g_i^s (\nabla_x g_i^s)^T \,.$$

Approximating the integral by integrating a Taylor approximation of τ_4 suffers from a major drawback. Since the first and second derivatives of τ_4 vanish whenever the quantities $\max\{0, g_i^s\}$ vanish $\forall i$, and $\max\{0, g_i^s\}$ vanishes $\forall i$ at any point that satisfies all the soft inequalities, the approximation ϕ to the integral of τ_4 vanishes *at any point that satisfies* $g_i^s \leq 0 \,\forall i$. This completely defeats our purpose since this implies that the 'amount of infeasibility' in the neighorhood 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 ϕ. In other words, a finite set of points $\hat{\mathcal{D}}$ is chosen from the infinite set $\mathcal{D}$ and the integral replaced by a sum of the integrand over all points in $\hat{\mathcal{D}}$, i.e.

$$\begin{aligned}\phi_z(x) &= \int_{\mathcal{D}} \tau_p(g^s(x+\xi_x, \psi_g(x+\xi_x, z+\xi_z), z+\xi_z))\, w(\xi_x,\xi_z)\, d\xi_x\, d\xi_z \\ &\approx \sum_{(\xi_x,\xi_z)\in\hat{\mathcal{D}}} \tau_p(g^s(x+\xi_x, \psi_g(x+\xi_x, z+\xi_z), z+\xi_z))\, w(\xi_x,\xi_z) \triangle(\xi_x,\xi_z)\end{aligned}$$

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

The above still requires us to have $\psi_g(x+\xi_x, z+\xi_z)\ \forall(\xi_x,\xi_z) \in \hat{\mathcal{D}}$, which involves solving the state equations $|\hat{\mathcal{D}}|$ 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.

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

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

$$\tau_p(g^s(x+\xi_x, \psi_g(x+\xi_x, z+\xi_z), z+\xi_z)) \;\approx\; \tau_p(g^s(x+\xi_x, \psi_g(x,z) + \nabla_x \psi_g^T \xi_x + \nabla_z \psi_g^T \xi_z, z+\xi_z)) \,.$$

This discretized integral still involves evaluating the soft constraints g^s $|\hat{\mathcal{D}}|$ times. This may not be a problem if the computation of g^s 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.

$$\begin{aligned}&g^s(x+\xi_x, \psi_g(x,z) + \nabla_x \psi_g^T \xi_x + \nabla_z \psi_g^T \xi_z, z+\xi_z) \\ \approx\; &g^s(x, \psi_g(x,z), z) + (\nabla_x g^s)^T \xi_x + (\nabla_y g^s)^T (\nabla_x \psi_g^T \xi_x + \nabla_z \psi_g^T \xi_z) + (\nabla_z g^s)^T \xi_z \,.\end{aligned}$$

Now τ_p can be evaluated at the above approximation of g^s. The resulting approximation to the discretized integral for ϕ_z is the seemingly clumsy-looking expression below:
$\phi_z(x)$

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

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 τ_2 which happen to exist. Also observe that the existence of second derivatives of g^s is required in order to claim differentiability of τ_2 with the first-order approximation of g^s as its argument.

Brief comments on $\nabla_x\psi_g^T$ and $\nabla_z\psi_g^T$ 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 to obtain the gradients of ψ_g. Since $\nabla_v C = 0$ and $\nabla_v g^h = \mathrm{diag}(v)$, the required gradients turn out to be

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

$$\nabla_z\psi_g^T = \begin{bmatrix} \nabla_y C & \nabla_y g^h \\ 0 & \mathrm{diag}(v) \end{bmatrix}^{-T} \begin{bmatrix} (\nabla_z C)^T \\ (\nabla_z g^h)^T \end{bmatrix}.$$

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 ϕ_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 (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 $\nabla_x\psi_g$ and $\nabla_z\psi_g$ (and hence $\bar{R}_z$ and ϕ_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 $\phi_z(x,y)$ added to it 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 tri-objective problem where the objectives to be minimized are f, $\bar{R}_z$ and ϕ_z.

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 the integral for ϕ exactly

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 ξ_i for simplicity, be distributed uniformly over the interval $[\xi_i^-, \xi_i^+]$. If all the soft inequalities are explicit bounds, say $a_i \le x_i \le b_i$, $i = 1, \ldots, dim(x)$, $\phi_z(x)$ can be expressed as

$$\phi(x) = \frac{1}{p}\int_{\mathcal{D}} w(\xi)d\xi \sum_{i=1}^{dim(x)} (\max\{0, a_i - (x_i + \xi_i)\})^p + (\max\{0, x_i + \xi_i - b_i\})^p$$

$$= \frac{1}{p}\sum_{i=1}^{dim(x)} \int_{\mathcal{D}} (\max\{0, a_i - (x_i + \xi_i)\})^p w(\xi)d\xi + \frac{1}{p}\sum_{i=1}^{dim(x)} \int_{\mathcal{D}} (\max\{0, x_i + \xi_i - b_i\})^p w(\xi)d\xi \quad (4)$$

where $w(\xi) = \Pi_{i=1}^{dim(x)}(\xi_i^+ - \xi_i^-)^{-1}$ is a constant, and $\mathcal{D} = \{\xi : \xi_i^- \le \xi_i \le \xi_i^+ \ \ \forall i\}$. Observe that the subscript z has been dropped from ϕ because ϕ does not depend on the environmental variables in this special situation.

Now,

$$\int_{\mathcal{D}} (\max\{0, a_i - (x_i + \xi_i)\})^p w(\xi)d\xi$$

$$= \frac{1}{\xi_i^+ - \xi_i^-}\int_{\mathcal{D}} (\max\{0, a_i - (x_i + \xi_i)\})^p d\xi_i$$

$$= \frac{1}{\xi_i^+ - \xi_i^-}\int_{\xi_i \le a_i - x_i} (a_i - x_i - \xi_i)^p d\xi$$

$$= \begin{cases} \frac{1}{\xi_i^+ - \xi_i^-}\int_{\xi_i^-}^{\min\{\xi_i^+,\, a_i - x_i\}} (a_i - x_i - \xi_i)^p \, d\xi_i, & \text{if } x_i < a_i - \xi_i^- \\ 0, & \text{if } x_i \ge a_i - \xi_i^- \, . \end{cases}$$

Thus, whenever $x_i < a_i - \xi_i^-$, the above equals

$$\frac{1}{(p+1)(\xi_i^+ - \xi_i^-)}[\,(a_i - x_i - \xi_i^-)^{p+1} - (a_i - x_i - \min\{\xi_i^+,\, a_i - x_i\})^{p+1}\,]\,. \quad (5)$$

Similarly it can be shown that

$$\int_{\mathcal{D}} (\max\{0, a_i - (x_i + \xi_i)\})^p w(\xi)d\xi$$

$$= \begin{cases} \frac{1}{(p+1)(\xi_i^+ - \xi_i^-)}[\,(x_i + \xi_i^+ - b_i)^{p+1} - (x_i + \max\{b_i - x_i,\, \xi_i^-\} - b_i)^{p+1}\,] & \text{if } x_i > b_i - \xi_i^+ \\ 0, \text{ if } x_i \le b_i - \xi_i^+ \, . \end{cases} \quad (6)$$

Some simplifications occur if $\xi_i^+ \ge 0$ and $\xi_i^- \le 0$. An example of such a situation is the case when ξ_i is distributed symmetrically about zero. Given that $x_i \ge a_i$ which any feasible point satisfies,

$$\xi_i^+ \ge 0 \Rightarrow x_i \ge a_i - \xi_i^+ \Rightarrow \min\{a_i - x_i,\, \xi_i^+\} = a_i - x_i \, .$$

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, 5 and 6 become

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

and

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

Now the above (or their more general versions in 5 and 6) can be substituted in 4 to evaluate ϕ. The value of the integer p was chosen to be 2 in our computation.

6 A numerical example

The technique for robustness optimization of functions on constrained domains discussed in the preceding sections was applied to a truss design problem described below. The problem involves optimizing the design of a pin-jointed truss structure under a wind load and a suspended load as shown in Fig. (7).

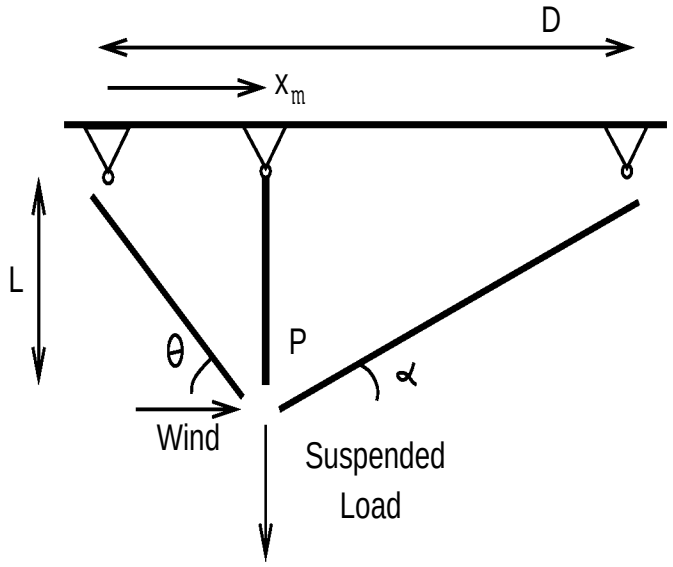

Figure 7: A truss structure under a suspended load and a wind load

The problem is to find the optimal position x_m of the vertical bar of fixed length $L = 60$ ft. between 1/4 and 3/4 of the fixed distance $D = 120$ ft. and the optimal bar cross-sectional areas. The bars on the edge get fixed and their lengths decided accordingly. The angles θ and α clearly depend on the chosen location x. The cross-sectional areas of

the three bars in sq. in., a_1, a_2, a_3, were allowed to vary between 0.8in^2 and 3.0in^2. The horizontal and vertical displacement of the node are denoted by u_1 and u_2 respectively and measured in feet.

Observe that allowing the position of the middle bar to vary introduces more nonlinearity in the problem than is found in the version of this problem studied by Koski [5].

The vector of variables in the problem is

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

The first four variables are design variables while the remining four are state variables. The equality constraints arise from expressing equilibrium of the truss under the given loads and from imposing that the three bars are connected at P.

The aim of the designer is to find a robust Pareto optimal point for minimizing the square of the displacement of the node P and the total volume of the truss. Let us denote the first objective by $f_1(r)$ and the second by $f_2(r)$, so

$$f_1(r) = u_1^2 + u_2^2\,,$$

$$f_2(r) = a_1 \frac{L}{\sin\theta} + a_2 L + a_3 \frac{L}{\sin\alpha}\,.$$

This biobjective problem (with deterministic data) was solved using NBI and a point was picked from the Pareto curve based on 'engineering judgment'. This Pareto point is also the minimizer of the following weighted sum of the two objectives[2]

$$f(r) = 0.7254\, f_1(r) + 0.2746\, f_2(r)\,.$$

The minimizer of $f(r)$ was

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

A robust minimizer of the function $f(r)$ is what we desire. Aberrations in r^* could result in either 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
  - ξ_{a_i}, the aberration in each cross-section: Uniform[-0.4, 0.4] sq. in.
  - ξ_{x_m}, 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.

[2]The weights can be obtained by looking at the optimal multipliers in the NBI subproblem as shown in Das [2].

- Environmental variable distributions
  - Horizontal wind load: Normal$(100, 15^2)$ kips
  - Suspended load: Normal$(1000, 25^2)$ kips

All the random variables were assumed to be uncorrelated.

Expression 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. 8.

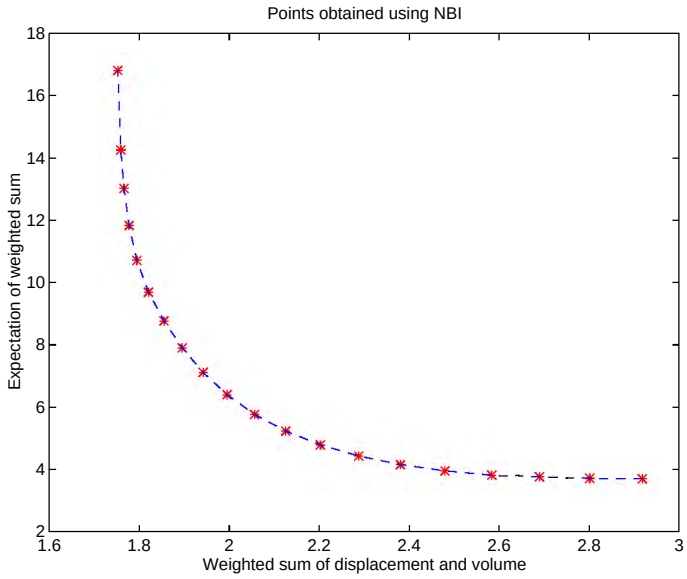

Figure 8: Trade-off curve for minimizing $f(r)$ and its expectation using NBI

It is worth observing that here 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' f_1 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, ϕ, was then taken into account as the third objective. Since the only soft inequalities are the bounds on the design variables, the quantity ϕ was evaluated using 4. This triobjective problem was solved using NBI; a plot of the points obtained is shown in Fig. 9. This Pareto surface appears to be almost degenerate; further

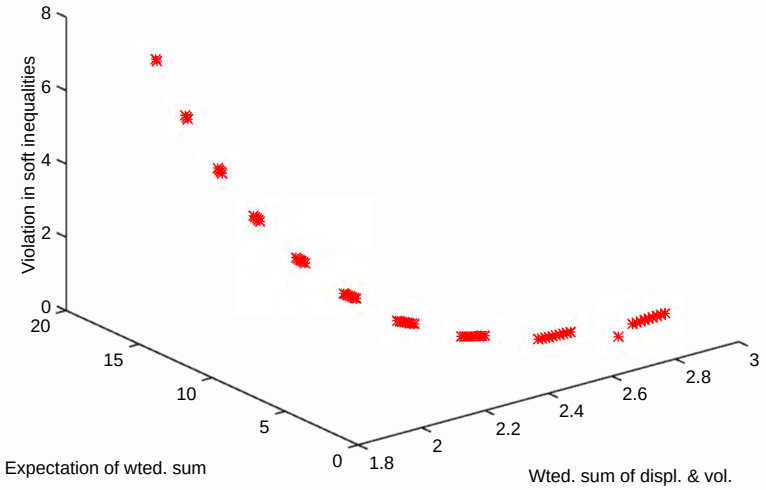

Figure 9: Trade-off surface for weighted sum $f(r)$, its expectation and ϕ

an examination of the values of the three objectives reveals that ϕ 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 thus obtained using NBI is shown in Fig. 10. The variation in ϕ along the Pareto curve is shown in Fig. 11. It can be observed that ϕ is low in the desirable 'middle part' of the trade-off curve, hence designating these points as good choices.

7 Conclusion

The robustness optimization problem was discussed in the context of nonlinear optimization. A systematic multicriteria procedure was formulated to allow the designer to explicitly trade off robustness for optimality and vice-versa. This renders unnecessary the use of ad hoc penalty parameters and weighting strategies widely used thus far for combining the 'robustness' and 'optimality' criteria into one scalar objective function.

An algorithmic strategy was also developed for evaluating the expectation of the cost function over a neighborhood of aberrations. For problems constrained by state equations, existing strategies for evaluating the expectation have usually been based on Monte-Carlo integration techniques or more sophisticated variants of such methods. Such techniques treat the state equation solver as a black box and require solving the state equations multiple times to obtain a reasonably accurate estimate of the integral. Since optimizing the expectation involves many evaluations of the expectation, the computational task reaches formidable proportions. The derivative-based approximation suggested in this paper makes the computational task more tractable, but restricts the applicability of the method to problems where the derivatives are available. The application of automatic differentiation

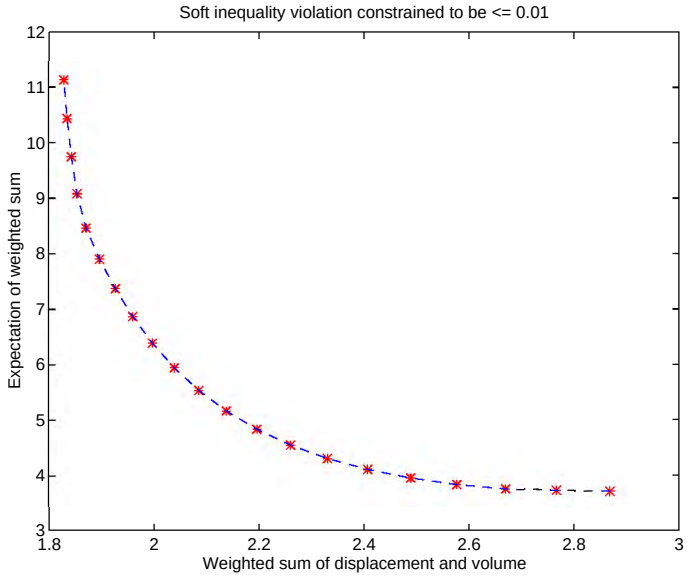

Figure 10: Pareto curve for weighted sum and its expectation subject to $\phi \leq 0.01$

techniques for finding the derivatives of the state equations is one of the topics of future research.

8 Acknowledgments

The author would like to thank Dr. John Dennis and Dr. Michael Trosset, Rice University, and Dr. Andrew Booker and Dr. Stephen Jones, The Boeing Company, for their helpful comments and suggestions.

9 Appendix A: Normal-Boundary Intersection

Normal-Boundary Intersection (NBI) is a new technique for solving multicriteria optimization problems. To describe NBI, let us first describe the multicriteria optimization problem and other necessary terminology. In mathematical notation a multicriteria optimization problem can be loosely posed as:

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

where

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

$F : \Re^N \mapsto \Re^n$, $h : \Re^N \mapsto \Re^{ne}$ and $g : \Re^N \mapsto \Re^{ni}$ are twice continuously differentiable mappings, $a \in (\Re \cup \{-\infty\})^N, b \in (\Re \cup \{\infty\})^N$, where N is the number of variables, n the

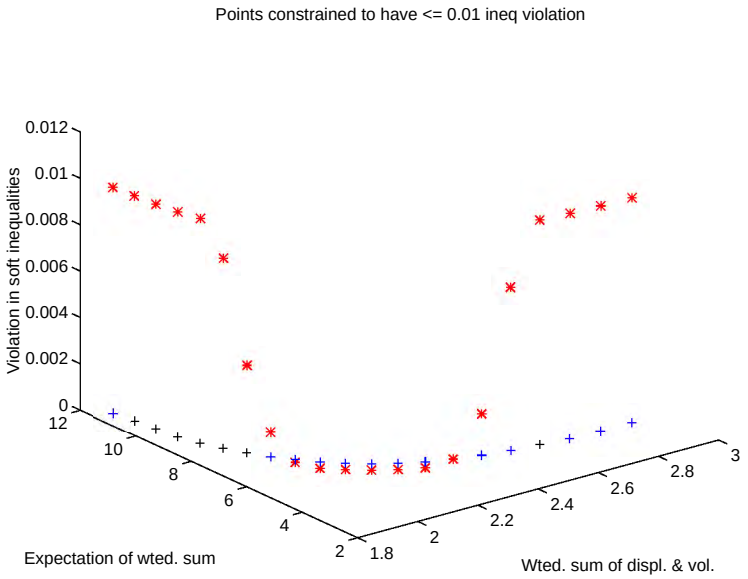

Figure 11: Variation in ϕ along Pareto curve for weighted sum and its expectation subject to $\phi \leq 0.01$.

number of objectives, ne and ni the number of equality and inequality constraints.
The domain of F is called the *decision space* while its range is referred to as the *objective space* (see fig. 12). The set of attainable objective vectors, $\{F(x) : x \in C\}$, is denoted by $\mathcal{F}$, so $F : C \mapsto \mathcal{F}$.

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

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 $x \in C$ such that $f_i(x) \leq f_i(x^*)\ \forall i \in \{1, 2, \ldots, n\}$ with at least one strict inequality.

It is possible to generate a set of Pareto optimal points by minimizing a convex combination of the objectives, $w^T F(x)$, over $x \in C$, where $w_i \geq 0\ \forall i \in \{1, 2, \ldots, n\}$ and $\sum_{i=1}^{n} w_i = 1$, and performing the minimization for different choices of w (see, among many others, Koski [5]). This very popular method suffers from two major drawbacks in that it fails to obtain nonconvex parts of the Pareto set and more importantly, does not yield a uniform spread of points on the Pareto set given a uniform spread of w.

Some other necessary definitions are given below.
The shadow minimum or utopia point, F^*, is defined as the vector containing the

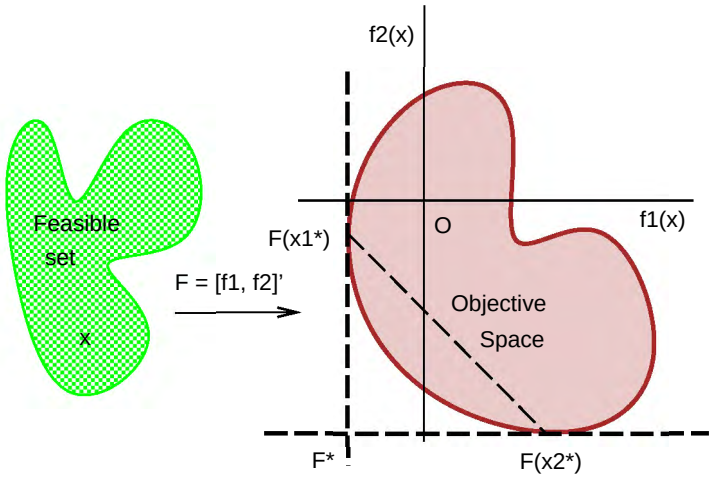

Figure 12: Mapping the feasible set to the objective space

individual global minima, f_i^*, of the objectives, i.e.,

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

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

Of course, global minima are not available in practice. However, this does not pose much of a problem in most cases as discussed extensively in Das [2].

Given a convex combination vector β, $\Phi\beta$ represents a point on the $CHIM$. Let $\hat{n}$ denote the unit normal to the $CHIM$ simplex pointing *towards* the origin; then $\Phi\beta + t\hat{n}, t \in \Re$ represents the set of points on that normal. Then the point of intersection between the normal and the boundary of $\mathcal{F}$ closest to the origin is the solution of the following nonlinear programming problem:

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

The vector constraint $\Phi\beta + t\hat{n} = F(x) - F^*$ 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 constrained set in the original problem (MOP).

The subproblem above is referred to as the *NBI subproblem* and written as NBI_β. The idea is to solve NBI_β for various β and find several points on the boundary of $\mathcal{F}$, effectively constructing a pointwise approximation to the part of the boundary of $\mathcal{F}$ containing the Pareto minimal set.

Solving NBI_β for various β with the components of β evenly-spaced is equivalent to 'shooting a family of normals' from a uniform grid of points on the $CHIM$ towards the Pareto boundary. Since the projection of the Pareto points hence obtained on the $CHIM$ are evenly spread, the points themselves are. Numerous instances of such even spreads of points are provided throughout the paper. It has also been rigorously proved that the even spread of Pareto points obtained using NBI is invariant with respect to the relative scales of the multiple objective functions.

For more details, the reader is referred to Das [2].

References

[1] G. E. P. Box and S. Jones, *Designing products that are robust to the environment*, Total Quality Management, 3 (1992), pp. 265–282.

[2] I. Das, *Nonlinear Multicriteria Optimization and Robust Optimality*, PhD thesis, Rice University, Houston, TX 77251, April 1997.

[3] S. Erikstad, U. Lautenschlager, B. Bras, J. Allen, and F. Mistree, *Integrating robustness into multiobjective space-vehicle design process*, Journal of Guidance Control and Dynamics, 18 (1995), pp. 1163–1168.

[4] Y. Ermoliev and R. J-B Wets (Eds.), *Numerical Techniques for Stochastic Optimization*, Springer Series in Computational Mathematics, Springer-Verlag, 1988.

[5] J. Koski, *Multicriteria truss optimization*, in Multicriteria Optimization in Engineering and in the Sciences, W. Stadler, ed., Plenum Press, New York, 1988.

[6] J. Mayer, *Computational techniques for probabilistic constrained optimization problems*, in Stochastic Programming, K. Marti, ed., Lecture Notes in Economics and Mathematical Systems, Berlin, 1990, Springer-Verlag, pp. 141–164. Proceedings of GAMM/IFIP-Workshop, Neubiberg, FRG, May 29 - 31, 1990.

[7] J. Mulvey, R. Vanderbei, and S. Zenios, *Robust optimization of large-scale systems*, Operations Research, 43 (1995), pp. 264–281.

[8] R. Slowinski and J. Teghem (Eds.), *Stochastic Versus Fuzzy Approaches to Multiobjective Mathematical Programming under Uncertainty*, Kluwer Academic Publishers, Dordrecht,The Netherlands, 1990.

[9] I. M. Stancu-Minasian, *Stochastic Programming with Multiple Objective Functions*, D. Reidel Publishing Co., Dordrecht, 1984.

[10] J. Su and J. E. Renaud, *Automatic differentiation in robust optimization*, AIAA-96-4005-CP, (1996), pp. 201–215.