



## Efficient Pareto Frontier Exploration using Surrogate Approximations

BENJAMIN WILSON, DAVID CAPPELLERI, TIMOTHY W. SIMPSON, MARY FRECKER  
*The Pennsylvania State University, Department of Mechanical & Nuclear Engineering, University Park,  
PA 16802, USA*  
*email: tws8@psu.edu*

*Received September 8, 2000; Revised May 2, 2001*

**Abstract.** In this paper we present an efficient and effective method of using surrogate approximations to explore the design space and capture the Pareto frontier during multiobjective optimization. The method employs design of experiments and metamodeling techniques (e.g., response surfaces and kriging models) to sample the design space, construct global approximations from the sample data, and quickly explore the design space to obtain the Pareto frontier without specifying weights for the objectives or using any optimization. To demonstrate the method, two mathematical example problems are presented. The results indicate that the proposed method is effective at capturing convex and concave Pareto frontiers even when discontinuities are present. After validating the method on the two mathematical examples, a design application involving the multiobjective optimization of a piezoelectric bimorph grasper is presented. The method facilitates multiobjective optimization by enabling us to efficiently and effectively obtain the Pareto frontier and identify candidate designs for the given design requirements.

**Keywords:** Pareto frontier, multiobjective optimization, approximation models, response surface, kriging

### 1. Introduction

Engineering design by its very nature is multiobjective, often requiring tradeoffs between disparate and conflicting objectives. Designing the cross-section of a cantilever beam is a classic example of the tradeoffs embodied in design—minimizing the weight and deflection of the beam requires a tradeoff between both objectives since improving one worsens the other. The pervasiveness of these tradeoffs in engineering design has given rise to a rich and vast array of approaches for multiobjective and multicriteria optimization. Examples include the weighted sum and compromise programming approaches (Osyczka, 1985; Stadler and Dauer, 1993; Steuer, 1986), genetic algorithm-based approaches (Azarm et al., 1999; Balling et al., 1999; Cheng and Li, 1997; Osyczka and Kundu, 1995; Schaumann et al., 1998), Pareto point approximations (Kasprzak and Lewis, 1999; Li et al., 1998; Zhang et al., 1999), and “brute force” approaches like Parameter Space Investigation (Lieberman, 1991; Sobol, 1992).

Many researchers have studied the limitations of weighted sum approaches to capture the Pareto set in non-convex problems (Athans and Papalambros, 1996; Koski, 1985; Messac et al., 1999). Messac et al. (1999) derive quantitative conditions for determining whether or not a Pareto point is capable of being captured with a given objective function formulation.

Das and Dennis (1997) examine the drawbacks of using weighted sums to find the Pareto set during multicriteria optimization, noting that an evenly distributed set of weights fails to produce an even distribution of points in the Pareto set. Their observations led to the creation of the Normal-Boundary Intersection (NBI) approach to parameterize the Pareto set and generate an evenly distributed set of points along the Pareto frontier using an evenly distributed set of parameters (Das, 1998, 1999; Das and Dennis, 1998).

Balling (1999) likens multiobjective optimization to “shopping” in that our goal as designers should be to produce a “rich set of good designs” from which the consumer can pick the best design. He advocates the need for research in two areas: (1) efficient methods for obtaining rich Pareto sets and (2) interactive graphical computer tools to assist decision makers in the “shopping” process. Balling and his colleagues (1999) have developed an approach that combines genetic algorithms to find Pareto optimal designs with an interactive GUI that has slider bars to vary the importance of the objectives to determine their impact on the design solution. Tappeta and Renaud (1999a, 1999b) are also developing an interactive multiobjective optimization procedure to explore design solutions around a Pareto point using second-order Pareto surface approximations derived from sensitivity information at the Pareto point. Implementation of their approach using the Physical Programming methodology is described in Tappeta et al. (1999).

In this paper, *we seek to reduce the computational expense of interactive approaches to multiobjective optimization by focusing on efficient methods for obtaining rich Pareto sets*. Consequently, we propose a method that employs design of experiments (e.g., central composite designs, Latin hypercubes) and surrogate approximations (e.g., response surfaces, kriging models) to rapidly explore and capture the Pareto frontier. Our method for efficient Pareto frontier exploration is introduced in the next section. This is followed in Sections 3 and 4 with two mathematical example problems to demonstrate the capability of the proposed Pareto frontier exploration method to function effectively for convex and non-convex multicriteria optimization problems even when discontinuities are present. The method is then implemented in Section 5 to design a piezoelectric bimorph actuator for use in minimally invasive surgery. Closing remarks are given in Section 6.

## 2. Technology base

In this work, we do not seek to approximate the Pareto frontier directly as previous researchers have done, nor do we require that any weights for the objectives be specified *a priori* for a weighted sum or compromise programming approach. Instead, we propose a method to explore the entire design space rapidly by combining design of experiments (e.g., central composite designs, Latin hypercubes) and metamodeling techniques (e.g., response surfaces, kriging models) to construct inexpensive-to-run approximations of computationally expensive engineering analyses and simulations (Simpson et al., 2001). These approximations are then used in lieu of the computationally expensive analyses, providing “surrogates” to help explore the multiobjective design space and identify a rich set of points along the Pareto frontier. Candidate points can then be used to obtain the actual (or near actual) Pareto frontier from the original analysis codes after good designs are identified for the multiple competing objectives.

### 2.1. The Pareto frontier exploration method

Our method for Pareto frontier exploration is shown in figure 1. As seen in the figure, the proposed method employs design of experiments and surrogate approximations to facilitate exploring and capturing the Pareto frontier. As shown in figure 1, the first step is to identify the design space. This is typically a multi-dimensional hypercube defined by the upper and lower bounds of each design variable over some region of interest. Once the design space has been identified, an experimental design is selected to sample the design space. A variety of different types of experimental designs exist, including central composite designs (Myers and Montgomery, 1995), Latin hypercubes (McKay et al., 1979), and orthogonal arrays (Owen, 1992) to name a few. After choosing an experimental design, the design space is sampled to obtain data to construct surrogate approximations of each objective and constraint.

Once the sample data has been obtained, the next step is to construct surrogate approximations. In this paper, we demonstrate the method using two types of surrogates: (1) second-order polynomial response surfaces (Myers and Montgomery, 1995), and (2) kriging models which employ an underlying constant term and a Gaussian correlation

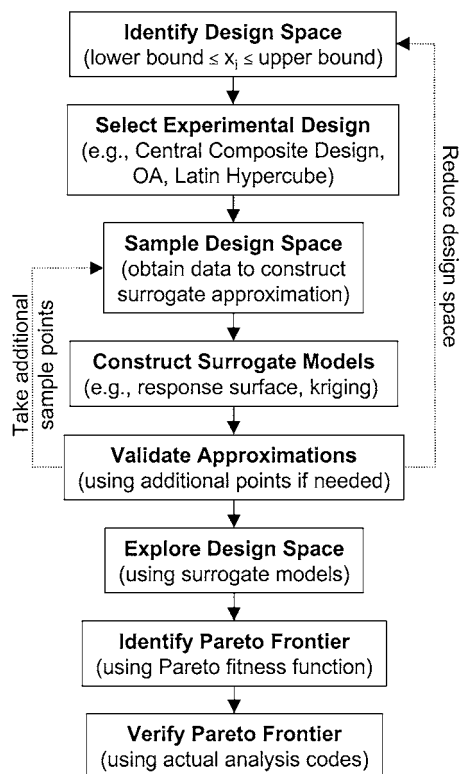


Figure 1. Pareto frontier exploration method.

function (see Booker, 1998; Giunta et al., 1998; Koehler and Owen, 1996; Sacks et al., 1989; Simpson et al., 1998; for a detailed description of kriging and example applications). A variety of surrogate approximations exist, however, and choosing the appropriate approximation is an open research question receiving considerable attention. Recent reviews of surrogate modeling applications in mechanical and aerospace engineering are in (Simpson et al., 2001), structural optimization are in (Barthelemy and Haftka, 1993), and multidisciplinary optimization are in (Sobieszczanski-Sobieski and Haftka, 1997).

After the approximations have been constructed, they must be validated to ensure sufficient accuracy. Validation can be achieved through a variety of means, including residual error analysis (Myers and Montgomery, 1995) and cross-validation (Meckesheimer et al., 2001). In the case of interpolative approximations such as kriging models, additional validation points are often required to assess metamodel accuracy. The additional validation points are used to compute error measures—where error is defined as the difference between the predicted and actual values—such as maximum absolute error (MAE), average absolute error (AAE), and  $R^2$  over the additional validation data as follows (Myers and Montgomery, 1995):

$$\text{MAE} = \max. \{|y_i - \hat{y}_i|\}_{i=1, \dots, n_{\text{error}}} \quad (1)$$

$$\text{AAE} = \frac{1}{n_{\text{error}}} \sum_{i=1}^{n_{\text{error}}} |y_i - \hat{y}_i| \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n_{\text{error}}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_{\text{error}}} (y_i - \bar{y})^2} \quad (3)$$

where  $y_i$  is the actual value of the response,  $\hat{y}_i$  is the predicted value of the response,  $\bar{y}$  is the average of the actual response values, and  $n_{\text{error}}$  is the number of additional validation points. For MAE and AAE, a smaller value indicates a more accurate fit; for  $R^2$ , the closer it is to 1, the better the fit. These error measures can also be computed as percentage (relative) errors by dividing the absolute error by the actual value of the response. If the approximation errors are too large, the Pareto frontier obtained using the surrogate approximations will not be a good approximation of the actual Pareto frontier. To improve the approximation error, additional sample points may be taken or the design space may be reduced in an effort to improve the accuracy of the approximation. If the number of additional sample points outweighs the estimated number of analyses required for optimization or the design space becomes too small to include all regions of interest for the multiple objectives, then traditional multiobjective optimization approaches should be employed to obtain the Pareto frontier.

Once the surrogate models have been validated, they can be used to rapidly explore the design space. Since the approximations are simple, they are extremely fast to execute; therefore, an exhaustive grid search over the design space is simple and computationally inexpensive compared to using the actual analyses. For continuous variables, the number of points sampled in the grid search is dictated only by the available computation time, which is influenced by the dimensionality of the design space and the desired resolution of the search grid. For discrete variables, the search grid resolution can be chosen to reflect available discrete choices.

After the design space has been explored, the Pareto frontier can be obtained using a Pareto fitness function such as that proposed by Schaumann et al. (1998):

$$F_i = [1 - \max_{j \neq i} (\min(f_{1i} - f_{1j}, f_{2i} - f_{2j}, \dots))]^p \quad (4)$$

where:

- $F_i$  = Pareto fitness value of  $i$ th design
- $f_{1i}$  = first objective function value of the  $i$ th design
- $f_{2i}$  = second objective function value of the  $i$ th design
- $p$  = Pareto exponent ( $p = 1$  in this study)

Using Eq. (4), Pareto optimal designs have a fitness function value greater than or equal to one; non-Pareto designs have fitness values between 0 and 1. This occurs because the objectives  $f_1$ ,  $f_2$ , etc. in Eq. (4) are scaled to range between zero and one using the following equation (Schaumann et al., 1998):

$$f_{1i} = \frac{\text{raw } f_{1i} - \text{raw } f_{1\min}}{\text{raw } f_{1\max} - \text{raw } f_{1\min}}, i = 1, \dots, n \quad (5)$$

where:

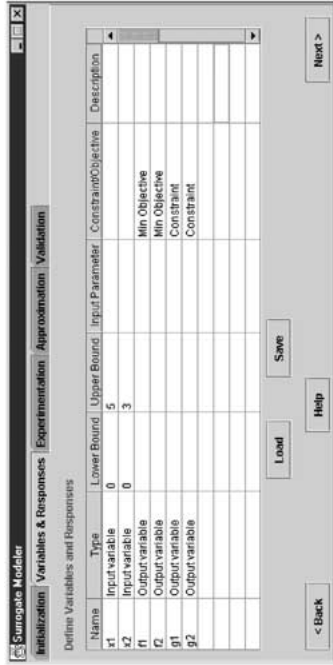
- $f_{1i}$  = scaled first objective value
- $\text{raw } f_{1i}$  = raw (un-scaled) value of first objective for  $i$ th design
- $\text{raw } f_{1\max}$  = maximum raw (un-scaled) value of first objective over all designs
- $\text{raw } f_{1\min}$  = minimum raw (un-scaled) value of first objective over all designs
- $n$  = number of objectives being considered

Equation 5 assumes that all objectives are to be minimized; however, the scaling can be easily reversed to maximize objectives if needed. Furthermore, the Pareto fitness function can be used to capture the Pareto frontier and corresponding design solutions for any number of objectives although we only demonstrate its use for problems with two objectives in order to make it easy to visualize and compare the actual and predicted Pareto frontiers.

The proposed method enables us to capture the entire Pareto frontier all at once without having to specify weights on the objectives or utilize any optimization algorithm. The predicted Pareto frontier can then be explored graphically to determine suitable design solutions that yield the best compromise between the multiple competing objectives. The corresponding design variables can be stored along with the Pareto points, enabling the actual Pareto frontier to be easily constructed from the original analysis code by substituting these design variables into the original analyses rather than the surrogate approximation.

## 2.2. Surrogate modeling software

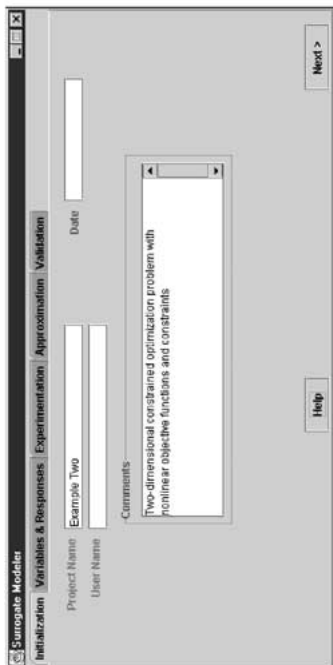
To facilitate construction and validation of the surrogate approximations, a platform-independent Java-based software application has been developed. After initialization (see



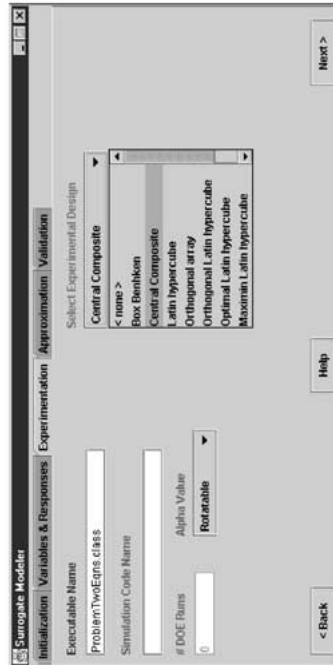
(b) Variables & Responses



(d) Approximation



(a) Initialization



(c) Experimentation

Figure 2. Screen shots of surrogate modeling software.

figure 2(a)), the surrogate modeling application queries the user for the design variables and their ranges of interest and the output responses (see figure 2(b)). The user then selects an experimental design, and a set of data points are generated and sent to the analysis code or simulation (see figure 2(c)). The application then constructs the selected surrogate approximation using the resulting sample data (see figure 2(d)) and outputs a separate Java class file, containing the corresponding surrogate model. This Java class can then be compiled and queried as needed in place of the original analysis code. The validation tab, which is not shown, is still under development.

To demonstrate the utility of surrogate approximations and the effectiveness of the proposed method for capturing the Pareto frontier, two mathematical example problems are presented next.

### 3. Example problem 1

The first example is a convex, bi-criteria mathematical function with linear boundary constraints from Li et al. (1998). The optimization problem for this example is formulated as follows:

$$\begin{aligned}
 \text{Minimize: } & f_1(x_1, x_2) = (x_1 - 2)^2 + (x_2 - 1)^2 \\
 & f_2(x_1, x_2) = x_1^2 + (x_2 - 6)^2 \\
 \text{subject to: } & g_1(x_1, x_2) = x_1 - 1.6 \leq 0 \\
 & g_2(x_1, x_2) = 0.4 - x_1 \leq 0 \\
 & g_3(x_1, x_2) = x_2 - 5 \leq 0 \\
 & g_4(x_1, x_2) = 2 - x_2 \leq 0
 \end{aligned} \tag{6}$$

Since the four linear constraints,  $g_1$ – $g_4$ , specify the region of interest (i.e.,  $x_1 \in [0.4, 1.6]$  and  $x_2 \in [2, 5]$ ), we choose to sample only within this region to avoid infeasible solutions. Therefore, the only surrogate approximations that need to be constructed in this example are for  $f_1$  and  $f_2$ , not the constraints.

Two experimental designs are used to sample the design space: (1) a central-composite design (CCD) and (2) a Latin hypercube (LH); scaled versions of each design are shown in figure 3. Both designs contain nine points to ensure a fair comparison between the resulting approximations.

The actual values of  $f_1$  and  $f_2$  at each sample point are recorded, and the resulting set of sample data is used to construct second-order response surface models and kriging models for each objective for each sample set. These surrogate models are then used to predict the responses at a set of new design points; we sample over 10,000 points from a  $101 \times 101$  grid to generate the Pareto frontier using the Pareto fitness function in Eq. (4). The resulting Pareto frontiers obtained using the response surface models and the kriging models are shown in figures 4(a) and (b), respectively. Since this example consists of two convex, second-order polynomial functions, the second-order polynomial response surface models predict the Pareto frontier exactly for both experimental designs.

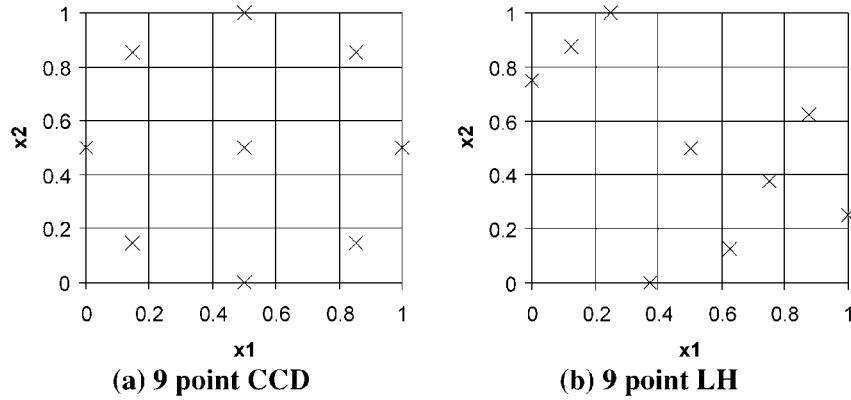


Figure 3. Experimental designs for examples 1 and 2.

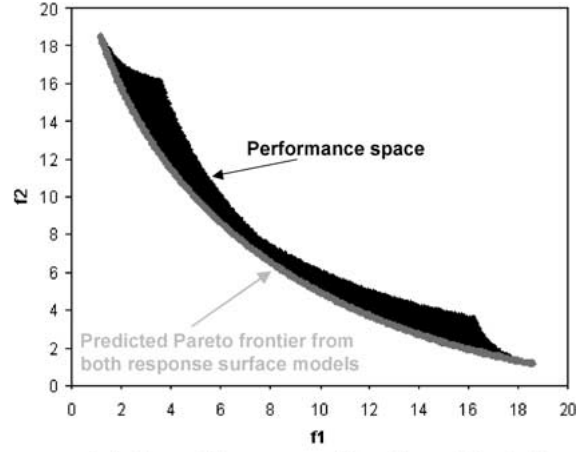
While the polynomial response surface models approximate  $f_1$  and  $f_2$  exactly, error analysis using maximum percent error, average percent error, and  $R^2$ , is performed on the kriging model for validation of the Pareto frontier predictions. Rather than use an additional set of validation points to verify the kriging models, the predicted Pareto frontier is compared directly with the actual Pareto frontier since the actual frontier can be easily obtained for this example over the search grid. The results of this analysis are listed in Table 1 and are obtained by comparing the values of  $f_1$  and  $f_2$  along the actual Pareto frontier with the corresponding predicted values of  $f_1$  and  $f_2$  based on the kriging models.

With the exception of the maximum percent error for  $f_1$ , the errors listed in Table 1 are very small, indicating that the kriging models—using only a constant term and a Gaussian correlation function—are nearly as accurate as second-order polynomial response surfaces. Also note that the kriging models based on the Latin hypercubes are much more accurate than those based on the central composite design, particularly in terms of maximum percent error. However, the  $R^2$  values for all four approximations are almost 1, indicating accurate predictions along the Pareto frontier as seen in figure 4(b).

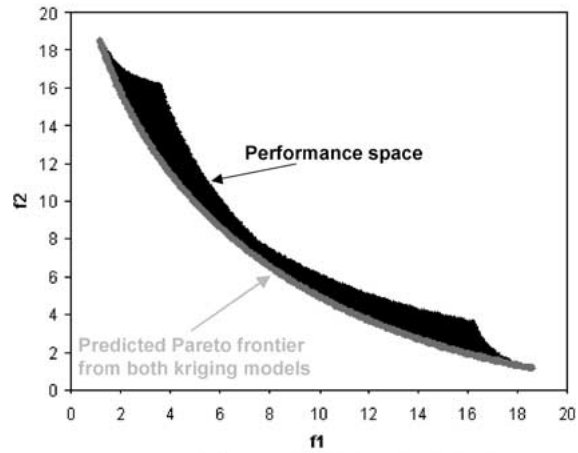
Table 1. Validation of kriging Pareto frontiers.

Error measure	Kriging with central composite design		Kriging with Latin hypercube	
	$f_1$	$f_2$	$f_1$	$f_2$
MAE (%)	22.024	6.531	0.552	0.128
AAE (%)	5.414	1.611	0.030	0.015
$R^2$	0.9972	0.9997	0.9999	0.9999





(a) from Response Surface Models



(b) from Kriging Models

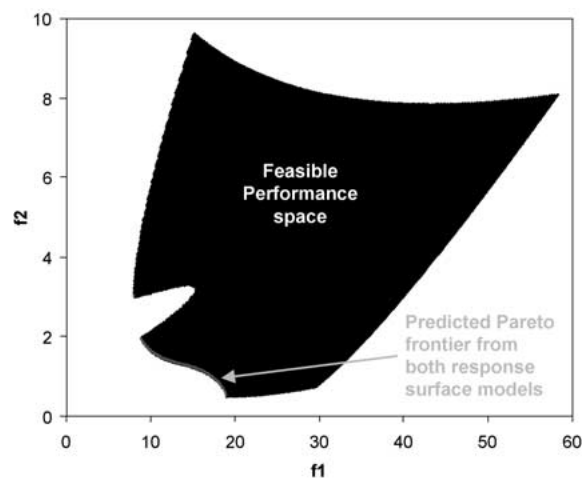
Figure 4. Predicted Pareto frontiers for example 1.

#### 4. Example problem 2

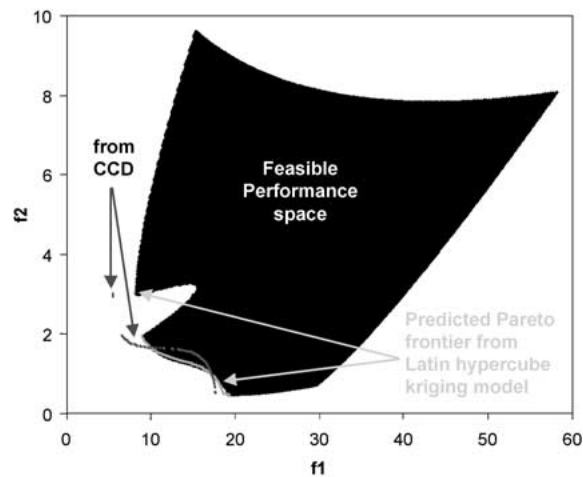
Our second example is a two-dimensional problem with non-linear objectives and constraints from Tappeta and Renaud (1999a). The problem definition is as follows.

$$\begin{aligned}
 \text{Minimize: } & f_1(x_1, x_2) = (x_1 + x_2 - 7.5)^2 + (x_2 - x_1 + 3)^2/4 \\
 & f_2(x_1, x_2) = (x_1 - 1)^2/4 + (x_2 - 4)^2/2 \\
 \text{subject to: } & g_1(x_1, x_2) = 2.5 - (x_1 - 2)^3/2 - x_2 \geq 0 \\
 & g_2(x_1, x_2) = 3.85 + 8(x_2 - x_1 + 0.65)^2 - x_2 - x_1 \geq 0
 \end{aligned} \tag{7}$$

The bounds  $0 \leq x_1 \leq 5$  and  $0 \leq x_2 \leq 3$  are used to ensure that the entire feasible performance space is utilized. Response surface and kriging models are created for the second example problem, using the same nine point Latin hypercube and central-composite designs shown in figure 3 scaled to the new design space. We again sample over 10,000 points from a  $101 \times 101$  grid to generate the Pareto frontier using the Pareto fitness function given in Eq. (4) and find that the response surface models, for both experimental designs, predict the Pareto frontier exactly as shown in figure 5(a). The Pareto frontiers obtained from the kriging models from both experimental designs are shown in figure 5(b).



(a) from Response Surface Models



(b) from Kriging Models

Figure 5. Predicted Pareto frontiers for example 2.

Table 2. Validation of kriging Pareto frontiers.

Error measure	Kriging with central composite design		Kriging with Latin hypercube	
	$f_1$	$f_2$	$f_1$	$f_2$
MAE (%)	31.741	44.421	0.074	40.135
AAE (%)	16.080	19.568	0.057	21.545
$R^2$	0.7449	0.8070	0.9999	0.8926

As seen in figure 5(b), the kriging models created using the Latin hypercube design predict the Pareto frontier quite accurately, but the kriging models created using the central composite design do not. Error measurements for  $f_1$  and  $f_2$  for the points along the Pareto frontier for both sets of kriging models are listed in Table 2. As in Section 3, these error measures are computed by comparing points along the actual Pareto frontier with points along the predicted Pareto frontier since it is easy to obtain the actual frontier for this example. These error measures are percentage errors expressed relative to the actual values of  $f_1$  and  $f_2$  along the true Pareto frontier. Error estimates for  $g_1$  and  $g_2$  are not computed since relative errors measured along the Pareto frontier will be very large because  $g_1$  and  $g_2$  are zero along the frontier.

Based on this data, we note that neither kriging model predicts  $f_2$  well along the frontier; however, the kriging model based on the Latin hypercube does accurately predict  $f_1$ . The  $R^2$  values for  $f_1$  and  $f_2$  for the kriging models based on the central composite design are poor, and the maximum percent error and average percent error are also quite large. Despite the large maximum percent error and average percent error for  $f_2$  for the kriging models based on the Latin hypercube, the  $R^2$  values for  $f_1$  and  $f_2$  are acceptable for an approximation. So while none of the kriging models predict as well as they did in the first example, the kriging models based on the Latin hypercube are more accurate than those based on the central composite design.

One explanation for the discrepancy between the performance of the two designs results from an interaction between the kriging model and the experimental design type. Due to the location and spacing of the sample points in the design space in a central composite design, the correlation matrix based on the Gaussian correlation function in the fitted kriging model tends to be singular or near singular when maximum likelihood estimation is performed to obtain the theta parameters used to fit the model. To confirm this, the condition numbers of the correlation matrices for each response for each design are listed in Table 3. We have found that a condition number smaller than  $10^{-12}$  indicates that significant round-off error can occur during prediction because the correlation matrix is close to singular. Such is the case for the kriging models for  $f_1$  and  $f_2$  based on the central composite design.

Despite the slight approximation error in the kriging models, the Pareto points themselves (i.e., the design variables  $x_1$  and  $x_2$  corresponding to each point on the Pareto frontier) are nearly identical to the actual Pareto points obtained from the original set of equations as seen in figure 6. Figure 6(a) shows the Pareto points obtained from the response surface

Table 3. Condition numbers of kriging model correlation matrices.

Condition number	Kriging with central composite design	Kriging with Latin hypercube
$f_1$	2.8821D-15	1.2165D-09
$f_2$	1.3950D-14	8.5201D-10
$g_1$	1.0345D-08	2.2178D-09
$g_2$	6.4121D-13	2.2982D-06

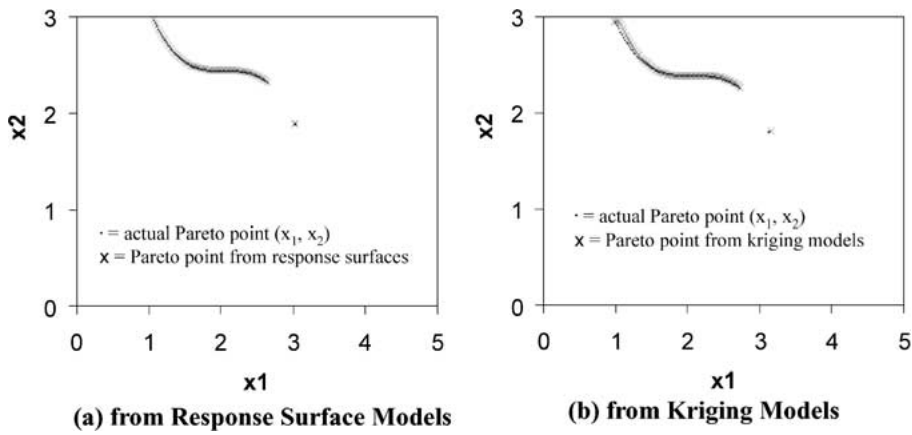


Figure 6. Mapping the design space of the Pareto frontier for example 2.

models overlaying the actual Pareto points; those obtained from the kriging models based on the Latin hypercube design are shown in figure 6(b).

While considerable error exists in the Pareto frontier obtained using the kriging models based on the central composite design, the kriging models based on the Latin hypercube design and both sets of response surface models are able to capture the Pareto frontier accurately. In this case, *the Pareto frontier is successfully captured without the use of optimization even though it is non-convex and discontinuous*. This example also demonstrates the importance of validating the surrogate approximations to ensure that they are sufficiently accurate before attempting to capture the Pareto frontier; otherwise, the predicted and actual frontiers may be vastly different.

## 5. Design of a piezoelectric bimorph actuator

Our final example comes from current research work in which we are trying to simultaneously optimize the maximum deflection and blocked force of a piezoelectric bimorph actuator for minimally invasive surgery (Cappelleri and Frecker, 1999; Cappelleri et al., 1999, 2001). A piezoelectric bimorph actuator is created by laminating layers of piezoelectric

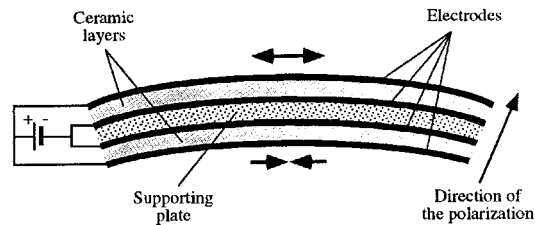


Figure 7. Bimorph actuator (Fatikow and Rembold, 1997).

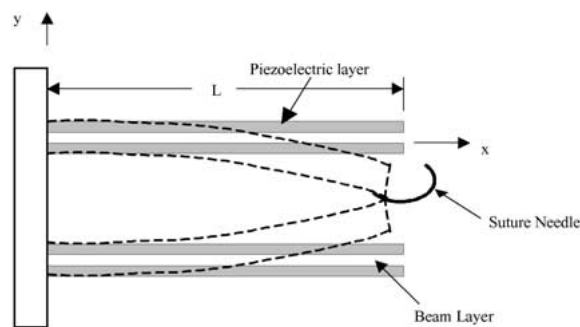


Figure 8. A piezoelectric bimorph grasper.

ceramic material (PZT) onto a thin sandwich beam or plate. When opposing voltages are applied to the two ceramic layers, a bending moment is induced in the beam, see figure 7. A pair of cantilevered piezoelectric bimorph actuators can be used as a simple grasping device, where the bimorph actuators are used as active “fingers” as shown in figure 8.

The objective in this example is to design a PZT bimorph grasper for application in minimally invasive surgical procedures. The performance of the PZT bimorph actuator is evaluated in terms of the tip force and deflection: a large tip deflection is required so that the jaws of the grasper can close completely, and a large tip force is required to securely grasp a suture needle and prevent it from rolling in the jaws. These performance criteria are determined by the thickness and width of each PZT layer, the length of the actuator, the material properties, and the applied voltage. The force available at the tip is modeled as the blocked force (i.e., the force exerted with no deflection), and the deflection is modeled as the free deflection. The results from preliminary analysis indicate that a standard, commercially available bimorph of the size required for MIS is infeasible since there is insufficient grasping force and tip deflection (Cappelleri et al., 2001). Consequently, a variable thickness design, where the thickness of the layers is varied along the length, is proposed to improve the deflection and force performance of the PZT bimorph actuator.

The piezoelectric bimorph actuator is modeled as a composite beam with a thin steel sandwich layer and PZT5H (IEEE, 1978) top and bottom layers. Rather than allowing the thickness of the PZT layers to vary continuously along the length, the layers are discretized into five sections to allow for simple modeling, where the thickness of each section,

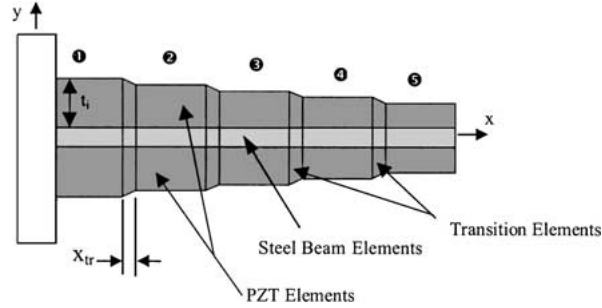


Figure 9. Variable thickness actuator.



Figure 10. Finite element model of actuator.

$t_i$  ( $i = 1, \dots, 5$ ), are the design variables (see figure 9), varying from 1 to 3 mm. A finite element model of the composite variable thickness design is created with standard cantilever supports at the base nodes. The finite element model consists of 1458 eight-node three-dimensional (brick) elements as shown in figure 10. The entire model is 5 mm long and 3 mm wide. Each PZT section has three elements across its height and width, and 10 elements along the length. As the thickness of the PZT sections is varied in the optimization procedure, the number of elements remains constant. Node compatibility at the section boundaries is ensured through the use of transition elements as shown in figure 9. These elements are formed by joining the end nodes from one piezoelectric section to the start nodes of the next piezoelectric section. The length of the transition elements,  $x_{tr}$ , is constant. The steel beam is modeled using 486 brick elements, with each section having three elements along the height and width and ten across the length.

The finite element (FE) analysis is performed using ABAQUS, a commercial non-linear finite element analysis software (ABAQUS, 1997). Two types of FE analysis are conducted to predict (1) the free deflection of the bimorph and (2) the blocked force, each while the actuator is subjected to a prescribed input voltage. The blocked force condition is simulated by constraining the nodes at the tip of the bimorph in the  $y$ -direction. Since the operating frequency is low (on the order of 1–2 Hz) in MIS applications, only quasi-static response is considered. The optimization problem for the variable thickness bimorph actuator can be stated as follows (note that we are maximizing the objectives in this example rather than minimizing as in the previous examples).

$$\begin{aligned}
 &\text{Maximize:} && \text{Blocked force} = f(t_1, t_2, t_3, t_4, t_5) \\
 &&& \text{Deflection} = f(t_1, t_2, t_3, t_4, t_5) \\
 &\text{subject to:} && 1 \text{ mm} \leq t_i \leq 3 \text{ mm}, i = 1, \dots, 5
 \end{aligned} \tag{8}$$

Table 4. Validation of surrogate approximations.

Error measure	Response surfaces		Kriging models	
	Deflection	Force	Deflection	Force
MAE (%)	13.78	18.99	18.06	8.02
AAE (%)	6.23	4.33	7.09	2.83
$R^2$	0.953	0.766	0.999	0.999

While Latin hypercubes were found to be more effective in the previous examples, only a face-centered central composite design (CCD) is used to sample the design space since it only requires three levels of the design variables to be analyzed ( $t_i = 1$  mm, 2 mm, and 3 mm). The face-centered CCD consists of 27 points: 16 points from a half-fraction Resolution V factorial design, 10 “star” points, and 1 center point. Data from these 27 sample points is used to construct response surface models and kriging approximations for both blocked force and deflection. Both sets of approximations are then used in lieu of the ABAQUS finite element analysis to search the design space to predict the Pareto frontier.

After constructing both sets of approximations, a set of twenty-five random points from a Latin hypercube are used to validate each set of approximations since it is too computationally expensive to determine the actual Pareto frontier in ABAQUS for comparison as we did in Sections 3 and 4. The maximum percent error, average percent error, and  $R^2$  values for the response surface and kriging models for the deflection and blocked force are listed in Table 4. The error in the predicted deflection is comparable for both approximations, with the response surface models yielding slightly lower maximum percent error and average percent error; however, in the case of the blocked force response, the kriging model fits the data much more accurately. Comparison of  $R^2$  values across approximations further reveals that the kriging models are much more accurate than the response surface models since  $R^2$  is nearly 1 for both kriging models.

To understand the tradeoff between the deflection and blocked force objectives, the approximations are used to search the design space and find the Pareto frontier. The design space is explored by predicting the blocked force and deflection of 3125 design points based on a  $5^5$  grid that searches over all section thicknesses in 0.5 mm increments based on readily available material sizes using the response surface models and the kriging models. The Pareto frontier is then obtained by selecting points on the boundary of the design space as predicted by the response surface models and the kriging models as shown in figure 11. The points on the Pareto frontier, as predicted by the response surface models and kriging models, are compared to one another, and the points with common thickness settings are evaluated in ABAQUS to verify the actual response. The results of this analysis are also shown in figure 11.

As seen in the figure, the kriging models are good predictors of the points along the Pareto frontier, while the response surface models are not for points that have a large deflection and small blocked force. This result is consistent with the data in Table 4 as the response surface model is generally less accurate than the kriging model based on the set of validation points. It is also evident from figure 11 that by varying the thickness of the piezoelectric

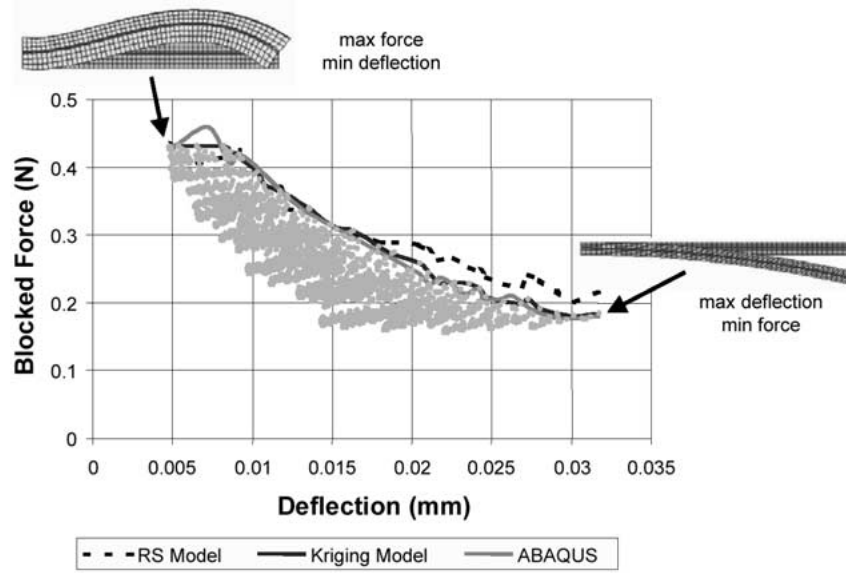


Figure 11. Predicted Pareto frontiers from response surface and kriging models.

layers a substantial blocked force can be generated, but the tip deflection is still quite low.

As a final comparison, we employ the weighted sum approach typically used in multiobjective optimization to demonstrate the effectiveness of our proposed method at capturing the entire Pareto frontier. Toward this end, we vary the weighting factor  $w$  to identify candidate designs based on a weighted sum of both objectives (i.e., maximize deflection and blocked force):

$$f = (1 - w) \frac{\delta_{\text{free}}^* - \delta_{\text{free}}}{\delta_{\text{free}}^*} + (w) \frac{F_{\text{blocked}}^* - F_{\text{blocked}}}{F_{\text{blocked}}^*} \quad (9)$$

where  $\delta_{\text{free}}^*$  and  $F_{\text{blocked}}^*$  are the individual optimum values of free deflection and blocked force, respectively. The weighting factor,  $w$ , is varied from 0.0 to 1.0, using increments of 0.01. Tables 5 and 6 show the effect of the weighting factor on the optimal solutions for the response surface and kriging models, respectively.

Although we expect to obtain several intermediate solutions by varying the weighting factor, only the two extreme solutions and one intermediate solution are obtained using the response surface models. As the weight on the deflection is increased, the maximum deflection solution is obtained until the weighting factor reaches 0.59, after which an intermediate solution is obtained until  $w = 0.88$ . The maximum force solution is obtained when the weighting factor is in the range 0.89–1.00. Using the kriging models, three intermediate solutions are obtained during optimization. This confirms the inadequacy of



Table 5. Effect of weighting factor on solution using response surface models.









Weighting factor, $w$	Solution	Predicted		ABAQUS	
		$\delta$ (mm)	$F_{\text{blocked}} (N)$	$\delta$ (mm)	$F_{\text{blocked}} (N)$
0.00–0.59		0.0318	0.2171	0.0317	0.1808
0.60–0.88		0.0096	0.4245	0.0092	0.4180
0.89–1.00		0.0052	0.4321	0.0048	0.4286

Table 6. Effect of weighting factor on solution using kriging models.

Weighting factor, $w$	Solution	Predicted		ABAQUS	
		$\delta$ (mm)	$F_{\text{blocked}} (N)$	$\delta$ (mm)	$F_{\text{blocked}} (N)$
0.0–0.78		0.0317	0.1837	0.0317	0.1808
0.79		0.0143	0.2990	0.0148	0.3060
0.80–0.92		0.0092	0.4180	0.0092	0.4180
0.93–0.95		0.0082	0.4308	0.0082	0.3950
0.95–1.0		0.0048	0.4286	0.0048	0.4286

using the weighted sum approach to capture intermediate solutions or the Pareto frontier for non-convex problems. The proposed method is much more efficient and effective at capturing the entire Pareto frontier without having to specify weights for the objectives or use an optimization algorithm.

## 6. Closing remarks

Preliminary results indicate that the proposed method provides a useful means for quickly and effectively capturing the Pareto frontier and identifying candidate designs when working with multiple competing objectives. The method employs surrogate approximations to reduce the computational expense of design analyses, enabling rapid design space search in multiple dimensions. Meanwhile, the Pareto fitness function given by Eq. (4) allows us to identify the predicted Pareto frontier without the need to specify weights for each objective

function or use any optimization. The predicted Pareto frontier can then be verified using the original analyses to estimate the actual Pareto frontier. The method works well for both convex and non-convex Pareto frontiers, even when discontinuities are present, as demonstrated in Section 3 and 4. We note that it is important that the surrogate approximations are validated and sufficiently accurate; otherwise, the predicted Pareto frontier will be inaccurate. The example problem in Section 4 demonstrated how inaccurate approximations can yield poor predictions along the Pareto frontier. If achieving accurate surrogate approximations is difficult, alternative metamodel types or experimental design strategies could be employed in an effort to improve modeling accuracy. This has been demonstrated by comparing results from two types of experimental designs (i.e., central composite designs and Latin hypercubes) and two types of surrogate approximations (i.e., response surfaces and kriging models) in each example problem. The method is very flexible and can easily accommodate a wide variety of experimental designs and approximations.

While three examples do not completely validate the method itself, the results to date are promising. While the examples contained herein are for multicriteria problems with two objectives, we do not anticipate any problems adapting our method to multicriteria problems with more than two objectives. The Pareto fitness function in Eq. (4) easily scales to higher dimensions; however, visualizing more than two objectives at once remains a challenge. The proposed method also does not restrict the use of an optimization algorithm to facilitate the search for the Pareto frontier using the surrogate models; however, we do wish to avoid having to specify weights *a priori* for each objective. We envision that a genetic algorithm-based approach such as that proposed in (Azarm et al., 1999; Balling et al., 1999; Schaumann et al., 1998) could greatly facilitate the design space search; however, we have not yet implemented such an approach.

### Acknowledgments

We appreciate the helpful comments and suggestions from our reviewers. The work by Benjamin Wilson and Dr. Simpson was supported by Dr. Kam Ng, ONR 333, through the Naval Sea Systems Command under Contract No. N00039-97-D-0042 and N00014-00-G-0058. The work by David Cappelleri and Dr. Frecker was supported by the Charles E. Culpeper Foundation Biomedical Pilot Initiative.

### References

- ABAQUS Version 5.7-1, Hibbitt, Karlsson and Sorensen, Inc., 1080 Main Street, Pawtucket, RI, (URL: <http://www.hks.com/>), 1997.
- T. W. Athan and P. Y. Papalambros, "A note on weighted criteria methods for compromise solutions in multi-objective optimization," *Engineering Optimization* vol. 27, no. 2, pp. 155–176, 1996.
- S. Azarm, B. J. Reynolds, and S. Narayanan, "Comparison of two multiobjective optimization techniques with and within genetic algorithms," *Advances in Design Automation*, Las Vegas, NV, ASME, Sept. 12–15, 1999, Paper No. DETC99/DAC-8584.
- R. Balling, "Design by shopping: A new paradigm?," *Proc. Third World Congress of Structural and Multidisciplinary Optimization*, C. L. Bloebaum and K. E. Lewis et al., eds. Buffalo, NY, University at Buffalo vol. 1, May 17–21, 1999, pp. 295–297.

- R. J. Balling, J. T. Taber, M. R. Brown, and K. Day, "Multiobjective urban planning using genetic algorithm," *Journal of Urban Planning and Development* vol. 125, no. 2, pp. 86–99, 1999.
- J.-F. M. Barthelemy and R. T. Haftka, "Approximation concepts for optimum structural design—A review," *Structural Optimization* vol. 5, pp. 129–144, 1993.
- A. J. Booker, "Design and analysis of computer experiments," *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization*, St. Louis, MO, AIAA vol. 1, Sept. 2–4, 1998, pp. 118–128. AIAA-98-4757.
- D. J. Cappelleri and M. I. Frecker, "Optimal design of smart tools for minimally invasive surgery," *Optimization in Industry II*, F. Mistree and A. D. Belegundu, eds. Banff, Alberta, Canada, ASME, July 6–11, 1999.
- D. J. Cappelleri, M. I. Frecker, and T. W. Simpson, "Optimal design of a PZT bimorph actuator for minimally invasive surgery," *7th International Symposium on Smart Structures and Materials*, Newport Beach, CA, SPIE, Mar. 5–9, 1999.
- D. J. Cappelleri, M. I. Frecker, T. W. Simpson, and A. Snyder, "A metamodel-based approach for optimal design of a PZT bimorph actuator for minimally invasive surgery," *Journal of Mechanical Design* 2001, to appear.
- F. Y. Cheng and D. Li, "Multiobjective optimization design with Pareto genetic algorithm," *Journal of Structural Engineering* vol. 123, no. 9, pp. 1252–1261, 1997.
- I. Das, "An improved technique for choosing parameters for Pareto surface generation using normal-boundary intersection," *Proc. Third World Congress of Structural and Multidisciplinary Optimization (WCSMO-3)* C. L. Bloebaum and K. E. Lewis et al., eds., Buffalo, NY, University at Buffalo, vol. 2, May 17–21, 1999, pp. 411–413.
- I. Das, "Optimization large systems via optimal multicriteria component assembly," *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization*, St. Louis, MO, AIAA, vol. 1, Sept. 2–4, 1998, pp. 661–669, AIAA-98-4791.
- I. Das and J. E. Dennis, "A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems," *Structural Optimization* vol. 14, no. 1, pp. 63–69, 1997.
- I. Das, and J. E. Dennis, "Normal-boundary intersection: A new method for generating the Pareto surface in nonlinear multicriteria optimization problems," *SIAM Journal on Optimization* vol. 8, no. 3, pp. 631–657, 1998.
- W. Fatikow and Rembold, U., *Microsystem Technology and Microrobotics*, Springer-Verlag: New York, 1997.
- A. Giunta, L. T. Watson, and J. Koehler, "A comparison of approximation modeling techniques: Polynomial versus interpolating models," *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization*, St. Louis, MO, AIAA vol. 1, Sept. 2–4, 1998, pp. 392–404. AIAA-98-4758.
- IEEE Group on Sonics and Ultrasonics, Transducers and Resonators Committee, *IEEE Standard on Piezoelectricity (176-1978)*, ANSI/IEEE: New York, 1978.
- E. M. Kasprzak and K. E. Lewis, "A method to determine optimal relative weights for Pareto solution sets," *Proc. Third World Congress of Structural and Multidisciplinary Optimization (WCSMO-3)*, C. L. Bloebaum and K. E. Lewis et al., eds., Buffalo, NY, University at Buffalo vol. 2, May 17–21, 1999, pp. 408–410.
- J. R. Koehler and A. B. Owen, "Computer experiments," *Handbook of Statistics*, S. Ghosh and C. R. Rao, eds. Elsevier Science: New York, 1996, pp. 261–308.
- J. Koski, "Defectiveness of weighting method in multicriterion optimization of structures," *Communications in Applied Numerical Methods* vol. 1, pp. 333–337, 1985.
- Y. Li, G. M. Fadel, and M. M. Wiecek, "Approximating Pareto curves using the hyper-ellipse," *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization*, St. Louis, MO, AIAA vol. 3, Sept. 2–4, 1998, pp. 1990–2002.
- E. R. Liberman, "Soviet multi-objective mathematical programming methods: An overview," *Management Science* vol. 37, no. 9, pp. 1147–1165, 1991.
- M. D. McKay, R. J. Beckman, and W. J. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics* vol. 21, no. 2, pp. 239–245, 1979.
- M. Meckesheimer, R. R. Barton, T. W. Simpson, and A. Booker, "Computationally inexpensive metamodel assessment strategies," *ASME Design Technical Conferences—Design Automation Conference* A. Diaz, ed. Pittsburgh, PA, ASME, Sept. 9–12, 2001, DETC2001/DAC-21028.
- A. Messac, J. G. Sundararaj, R. V. Tappeta, and J. E. Renaud, "The ability of objective functions to generate non-convex Pareto frontiers," *40th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and*

- Materials Conference and Exhibit*, St. Louis, MO, AIAA vol. 1, Apr. 12–15, 1999, pp. 78–87. AIAA-99-1211.
- R. H. Myers, and D. C. Montgomery, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, John Wiley & Sons: New York, 1995.
- A. Osyczka, "Multicriteria optimization for engineering design," in *Design Optimization*, J. S. Gero, ed. Academic Press: New York, 1985, pp. 193–227.
- A. Osyczka and S. Kundu, "A new method to solve generalized multicriteria optimization problems using the simple genetic algorithm," *Structural Optimization* vol. 10, no. 2, pp. 94–99, 1995.
- A. B. Owen, "Orthogonal arrays for computer experiments, integration and visualization," *Statistica Sinica* vol. 2, pp. 439–452, 1992.
- J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, "Design and analysis of computer experiments," *Statistical Science* vol. 4, no. 4, pp. 409–435, 1989.
- E. J. Schaumann, R. J. Balling, and K. Day, "Genetic algorithms with multiple objectives," *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization*, St. Louis, MO, AIAA vol. 3, Sept. 2–4, 1998, pp. 2114–2123.
- T. W. Simpson, T. M. Mauery, J. J. Korte, and F. Mistree, "Comparison of response surface and kriging models for multidisciplinary design optimization," *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization*, St. Louis, MO, AIAA vol. 1, Sept. 2–4, 1998, pp. 381–391, AIAA-98-4755.
- T. W. Simpson, T. M. Mauery, J. J. Korte, and F. Mistree, "Kriging metamodels for global approximation in simulation-based multidisciplinary design optimization," *AIAA Journal* 2001, to appear.
- T. W. Simpson, J. Peplinski, P. N. Koch, and J. K. Allen, "Metamodels for computer-based engineering design: Survey and recommendations," *Engineering with Computers* vol. 17, pp. 129–150, 2001.
- J. Sobieszczanski-Sobieski and R. T. Haftka, "Multidisciplinary aerospace design optimization: Survey of recent developments," *Structural Optimization* vol. 14, pp. 1–23, 1997.
- I. M. Sobol, "An efficient approach to multicriteria optimum design problems," *Surveys on Mathematics for Industry* vol. 1, pp. 259–281, 1992.
- W. Stadler and J. Dauer, "Multicriteria optimization in engineering: A tutorial and survey," *Structural Optimization: Status and Promise*, M. P. Kamat, ed., AIAA: Washington, D.C., 1993, pp. 209–249.
- R. E. Steuer, *Multiple Criteria Optimization: Theory, Computation, and Application*, Wiley: New York, 1986.
- R. V. Tappeta and J. E. Renaud, "Interactive multiobjective optimization design strategy for decision based design," *Advances in Design Automation*, Las Vegas, NV, ASME, Sept. 12–15, 1999a, Paper No. DETC99/DAC-8581.
- R. V. Tappeta, and J. E. Renaud, "Interactive multiobjective optimization procedure," *40th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference and Exhibit*, St. Louis, MO, AIAA vol. 1, Apr. 12–15, 1999b, pp. 27–41, AIAA-99-1207.
- R. V. Tappeta, J. E. Renaud, A. Messac, and J. G. Sundararaj, "Interactive physical programming: Tradeoff analysis and decision making in multicriteria optimization," *40th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference and Exhibit*, St. Louis, MO, AIAA vol. 1, Apr. 12–15, 1999, pp. 53–67, AIAA-99-1209.
- J. Zhang, M. M. Wiecek, and W. Chen, "Local approximation of the efficient frontier in robust design," *Advances in Design Automation*, Las Vegas, NV, ASME, Sept. 12–15, 1999, Paper No. DETC99/DAC-8566.