AIAA JOURNAL
Vol. 46, No. 10, October 2008

Efficient Global Reliability Analysis for Nonlinear Implicit Performance Functions

B. J. Bichon*
Vanderbilt University, Nashville, Tennessee 37235
M. S. Eldred† and L. P. Swiler‡
Sandia National Laboratories, Albuquerque, New Mexico 87185
and
S. Mahadevan§ and J. M. McFarland¶
Vanderbilt University, Nashville, Tennessee 37235

DOI: 10.2514/1.34321

Many engineering applications are characterized by implicit response functions that are expensive to evaluate and sometimes nonlinear in their behavior, making reliability analysis difficult. This paper develops an efficient reliability analysis method that accurately characterizes the limit state throughout the random variable space. The method begins with a Gaussian process model built from a very small number of samples, and then adaptively chooses where to generate subsequent samples to ensure that the model is accurate in the vicinity of the limit state. The resulting Gaussian process model is then sampled using multimodal adaptive importance sampling to calculate the probability of exceeding (or failing to exceed) the response level of interest. By locating multiple points on or near the limit state, more complex and nonlinear limit states can be modeled, leading to more accurate probability integration. By concentrating the samples in the area where accuracy is important (i.e., in the vicinity of the limit state), only a small number of true function evaluations are required to build a quality surrogate model. The resulting method is both accurate for any arbitrarily shaped limit state and computationally efficient even for expensive response functions. This new method is applied to a collection of example problems including one that analyzes the reliability of a microelectromechanical system device that current available methods have difficulty solving either accurately or efficiently.

I. Introduction

AS ENGINEERING applications become increasingly complex, they are often characterized by implicit response functions that are expensive to evaluate and perhaps nonlinear in their behavior. Reliability analysis, given this type of response, is difficult with available methods. Current reliability methods focus on the search for a single most probable point of failure, and then build a low-order approximation to the limit state at this point. This creates inaccuracies when applied to engineering applications for which the limit state has a higher degree of nonlinearity or is multimodal. Sampling methods, on the other hand, do not rely on an approximation to the shape of the limit state and are therefore generally more accurate when applied to problems with nonlinear limit states. However, sampling methods typically require a large number of response function evaluations, which can make their application infeasible for computationally expensive problems. This expense has been partially alleviated through the development of various efficient sampling methods.

Analytical methods of reliability analysis solve a local optimization problem to locate the most probable point (MPP) of failure, and then quantify the reliability based on its location and an approximation to the shape of the limit state at this point. Typically, gradient-based solvers are used to solve this optimization problem, which may fail to converge for nonsmooth response functions with unreliable gradients or may converge to only one of several solutions for response functions that possess multiple local optima. In addition to these MPP convergence issues, the evaluated probabilities can be adversely affected by limit state approximations that may be inaccurate. Engineers are then forced to revert to sampling methods, which do not rely on MPP convergence or simplifying approximations to the true shape of the limit state. However, as mentioned earlier, employing such methods is impractical when evaluation of the response function is expensive. Thacker et al. [1] provide a good overview of the errors in current methods when applied to nonlinear problems from structural dynamics and material fatigue, motivating the need for new methods with greater accuracy.

A reliability analysis method that is both efficient when applied to expensive response functions and accurate for a response function of any arbitrary shape is needed. This paper develops a method based on efficient global optimization [2] (EGO) to adaptively search for multiple points on or near the limit state throughout the random variable space. By locating multiple points near the limit state, more complicated and nonlinear limit states can be accurately modeled, resulting in an accurate assessment of the reliability.

EGO was developed to facilitate the unconstrained minimization of expensive implicit response functions. The method builds an initial Gaussian process model as a global surrogate for the response function, then adaptively selects additional samples to be added for inclusion in a new Gaussian process model in subsequent iterations. The new samples are selected based on how much they are expected to improve the current best solution to the optimization problem. When this expected improvement is acceptably small, the globally

Received 29 August 2007; revision received 22 April 2008; accepted for publication 5 May 2008.

*Doctoral Candidate and National Science Foundation Integrative Graduate Education and Research Traineeship Program Fellow, Department of Civil and Environmental Engineering. Student Member AIAA.

†Principal Member of Technical Staff, Optimization and Uncertainty Estimation Department. Associate Fellow AIAA.

‡Principal Member of Technical Staff, Optimization and Uncertainty Estimation Department. Member AIAA.

§Professor, Department of Civil and Environmental Engineering and Mechanical Engineering. Member AIAA.

¶Doctoral Candidate and National Science Foundation Integrative Graduate Education and Research Traineeship Fellow, Department of Mechanical Engineering. Student Member AIAA.

optimal solution has been found. The application of this methodology to equality-constrained reliability analysis is the primary contribution of this work. Combining this EGO-based search for the limit state contour with an efficient sampling method to calculate the probability of failure results in what is referred to as efficient global reliability analysis (EGRA) in this paper.

The use of Gaussian process models in reliability analysis was previously investigated in [3,4]. However, there are key differences in the previous work and the EGRA method introduced here. The earlier methods used a number of randomly selected samples to construct the model and global accuracy of that model was sought. This results in either a lack of accuracy if too few samples are used, or wasted expense creating models that are accurate in areas where they need not be. EGRA avoids these problems by not requiring the surrogate model to have high accuracy throughout the random variable domain, but only in the vicinity of the limit state. This is accomplished by focusing the training data around the limit state and greatly reduces the number of samples required. Additionally, the search for the limit state is performed using an iterative process with a rigorous convergence criteria, ensuring that the final model provides an accurate depiction of the limit state.

Section II describes the reliability analysis problem and traditional methods of solving it. Section III gives an overview of the EGO algorithm, and outlines how it is adapted for application to reliability analysis. Section IV describes a collection of example problems and compares the performance of EGRA to that of other available methods. Finally, Sec. V provides concluding remarks on this new method.

II. Reliability Analysis

The goal of reliability analysis is to determine the probability that an engineered device, component, system, etc., will fail in service, given that its behavior is dependent on random inputs. This behavior is defined by a response function $g(\mathbf{x})$, where $\mathbf{x}$ represents the vector of random variables defined by known probability distributions. Failure is then defined by that response function exceeding (or failing to exceed) some threshold value $\bar{z}$. The probability of failure p_f is then defined by

$$p_f = \int_{g>\bar{z}} \cdots \int f_{\mathbf{x}}(\mathbf{x}) \mathrm{d}\mathbf{x} \tag{1}$$

where $f_{\mathbf{x}}$ is the joint probability density function of the random variables $\mathbf{x}$, and the integration is performed over the failure region where $g > \bar{z}$. In general, $f_{\mathbf{x}}$ is impossible to obtain and, even when it is available, evaluating the multiple integral is impractical [5]. Because of these complications, methods of approximating this integral are used in practice.

A. Most-Probable-Point-Based Methods

These methods involve solving a nonlinear optimization problem to locate the point on the limit state (the contour on the response function where $g = \bar{z}$) that has the greatest probability of occurring. This point is known as the most probable point. A first- or second-order approximation to the limit state is then constructed at this point to facilitate the integration required to compute the probability of failure.

The MPP search is performed in uncorrelated standard normal space because it simplifies the probability integration; in this space, the distance from the origin to the MPP is equivalent to the reliability index and is denoted by β. The transformation from correlated nonnormal distributions (x space) to uncorrelated standard normal distributions (u space) is nonlinear in general, and possible approaches include the Rosenblatt [6], Nataf [7], and Box–Cox [8] transformations. The nonlinear transformations may also be linearized, and common approaches for this include the Rackwitz–Fiessler [9] two-parameter equivalent normal and the Chen–Lind [10] and Wu–Wirsching [11] three-parameter equivalent normals. This paper employs the Nataf nonlinear transformation [7], which occurs in the following two steps. To transform between the original correlated x-space variables and correlated standard normals (z space), the cumulative distribution function matching condition is used:

$$\Phi(z_i) = F(x_i) \tag{2}$$

where $F(\,)$ is the cumulative distribution function of the original probability distribution and $\Phi(\,)$ is the standard normal cumulative distribution function. Then, to transform from correlated z-space variables to uncorrelated u-space variables, the Cholesky factor $\mathbf{L}$ of a modified correlation matrix is used:

$$\mathbf{z} = \mathbf{L}\mathbf{u} \tag{3}$$

where the original correlation matrix for nonnormals in x space has been modified for z space [7].

The forward-reliability analysis algorithm for computing the probability/reliability level that corresponds to a specified response level $\bar{z}$ is formulated as

$$\text{minimize} \quad \mathbf{u}^T\mathbf{u} \qquad \text{subject to } G(\mathbf{u}) = \bar{z} \tag{4}$$

where $\mathbf{u}$ is a vector centered at the origin in u space and $G(\mathbf{u}) \equiv g(\mathbf{x})$ by definition. The optimal MPP solution $\mathbf{u}^*$ defines the reliability index from $\beta = \pm\|\mathbf{u}^*\|_2$, which in turn defines the probability of failure through the probability integration.

Recent research in [12–14] has focused on the use of local and multipoint surrogate models to reduce the expense of the MPP search. All of these MPP search methods employ local optimization techniques and converge to a single MPP. However, the limit state of a complex engineering application may be multimodal and possess multiple significantly probable points of failure. Mahadevan and Shi [15] deal with these types of functions by searching for all of these MPPs and forming a local approximation at each, whereas Zou et al. [16] construct an approximation to the indicator response function. The method for reliability analysis proposed here uses a global surrogate model and global optimization methods to reduce expense and locate multiple points along the limit state.

The simplest approximation to compute the failure probability is the first-order reliability method (FORM), which approximates the limit state as a linear function and computes the failure probability as $p_f = \Phi(-\beta)$. Several second-order approximations are also available [17–20] using the curvature of the limit state at the MPP. All of these methods use an approximation to the limit state based only on information in the immediate vicinity of the MPP, making their results inaccurate if the approximation is poor or if the MPP is improperly located.

B. Sampling Methods

An alternative to MPP search methods is to directly perform the probability integration numerically by sampling the response function. Sampling methods do not rely on a simplifying approximation to the shape of the limit state, and so they can be more accurate than FORM and the second-order reliability method (SORM), but they can also be prohibitively expensive because they generally require a large number of response function evaluations. Importance sampling methods reduce this expense by focusing the samples in the important regions of the random variable space. One option is to center the sampling density function at the MPP rather than at the mean [21]. This ensures a large number of failure samples, thus increasing the efficiency of the sampling method. Adaptive importance sampling (AIS) further improves the efficiency by adaptively updating the sampling density function. Curvature-based [22] and multimodal [14,23] options are available to refine the sampling density.

Note that importance sampling methods require that the location of at least one MPP be known because it is used to center the initial sampling density. However, current gradient-based, local search methods used in MPP search may fail to converge or may converge to poor solutions for highly nonlinear problems, possibly making these

methods inapplicable. As the next section describes, EGO is a global optimization method that does not depend on the availability of accurate gradient information, making convergence more reliable for nonsmooth response functions. Moreover, EGO has the ability to locate multiple failure points, which would provide multiple starting points and thus a good multimodal sampling density for the initial steps of multimodal AIS. The resulting Gaussian process model is accurate in the vicinity of the limit state, thereby providing an inexpensive surrogate that can be used to provide response function samples. As will be seen, using EGO to locate multiple points along the limit state, and then using the resulting Gaussian process model to provide function evaluations in multimodal AIS for the probability integration, results in an accurate and efficient reliability analysis tool. Because of its roots in efficient global optimization, this method of reliability analysis is called efficient global reliability analysis in this paper.

III. Efficient Global Optimization

Efficient global optimization was originally proposed by Jones et al. [2] and has been adapted into similar methods such as sequential kriging optimization (SKO) [24]. The main difference between SKO and EGO lies within the specific formulation of what is known as the expected improvement function (EIF), which is the feature that sets all EGO/SKO-type methods apart from other global optimization methods. The EIF is used to select the location at which a new training point should be added to the Gaussian process model by maximizing the amount of improvement in the objective function that can be expected by adding that point. A point could be expected to produce an improvement in the objective function if its predicted value is better than the current best solution, or if the uncertainty in its prediction is such that the probability of it producing a better solution is high. Because the uncertainty is higher in regions of the design space with fewer observations, this provides a balance between exploiting areas of the design space that predict good solutions and exploring areas where more information is needed. The general procedure of these EGO-type methods is as follows:

1) Build an initial Gaussian process model of the objective function.

2) Find the point that maximizes the EIF. If the EIF value at this point is sufficiently small, stop.

3) Evaluate the objective function at the point where the EIF is maximized. Update the Gaussian process model using this new point. Go to step 2.

The following sections discuss the construction of the Gaussian process model used, the form of the EIF, and then a description of how that EIF is modified for application to reliability analysis.

A. Gaussian Process Model

Gaussian process (GP) models are set apart from other surrogate models because they provide not just a predicted value at an unsampled point, but also an estimate of the prediction variance. This variance gives an indication of the uncertainty in the GP model, which results from the construction of the covariance function. This function is based on the idea that when input points are near one another, the correlation between their corresponding outputs will be high. As a result, the uncertainty associated with the model's predictions will be small for input points that are near the points used to train the model and will increase as one moves further from the training points.

It is assumed that the true response function being modeled $G(\mathbf{u})$ can be described by [25]

$$G(\mathbf{u}) = \mathbf{h}(\mathbf{u})^T\boldsymbol{\beta} + Z(\mathbf{u}) \tag{5}$$

where $\mathbf{h}(\,)$ is the trend of the model, $\boldsymbol{\beta}$ is the vector of trend coefficients, and $Z(\,)$ is a stationary Gaussian process with zero mean (and covariance defined next) that describes the departure of the model from its underlying trend. The trend of the model can be assumed to be any function, but taking it to be a constant value has been reported to be generally sufficient [26]. For the work presented here, the trend is assumed constant and $\boldsymbol{\beta}$ is taken as simply the mean of the responses at the training points. The covariance between outputs of the Gaussian process $Z(\,)$ at points $\mathbf{a}$ and $\mathbf{b}$ is defined as

$$\mathrm{Cov}\,[Z(\mathbf{a}), Z(\mathbf{b})] = \sigma_Z^2 R(\mathbf{a}, \mathbf{b}) \tag{6}$$

where σ_Z^2 is the process variance and $R(\,)$ is the correlation function. There are several options for the correlation function, but the squared-exponential function is common [26], and is used here for $R(\,)$:

$$R(\mathbf{a}, \mathbf{b}) = \exp\left[-\sum_{i=1}^{d}\theta_i(a_i - b_i)^2\right] \tag{7}$$

where d represents the dimensionality of the problem (the number of random variables), and θ_i is a scale parameter that indicates the correlation between the points within dimension i. A large θ_i is representative of a short correlation length.

The expected value $\mu_G(\,)$ and variance $\sigma_G^2(\,)$ of the GP model prediction at point $\mathbf{u}$ are

$$\mu_G(\mathbf{u}) = \mathbf{h}(\mathbf{u})^T\boldsymbol{\beta} + \mathbf{r}(\mathbf{u})^T\mathbf{R}^{-1}(\mathbf{g} - \mathbf{F}\boldsymbol{\beta}) \tag{8}$$

$$\sigma_G^2(\mathbf{u}) = \sigma_Z^2 - [\,\mathbf{h}(\mathbf{u})^T \quad \mathbf{r}(\mathbf{u})^T\,]\begin{bmatrix}\mathbf{0} & \mathbf{F}^T\\ \mathbf{F} & \mathbf{R}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{h}(\mathbf{u})\\ \mathbf{r}(\mathbf{u})\end{bmatrix} \tag{9}$$

where $\mathbf{r}(\mathbf{u})$ is a vector containing the covariance between $\mathbf{u}$ and each of the n training points [defined by Eq. (6)], $\mathbf{R}$ is an $n \times n$ matrix containing the correlation between each pair of training points, $\mathbf{g}$ is the vector of response outputs at each of the training points, and $\mathbf{F}$ is an $n \times q$ matrix with rows $\mathbf{h}(\mathbf{u}_i)^T$ (the trend function for training point i containing q terms; for a constant trend, $q = 1$). This form of the variance accounts for the uncertainty in the trend coefficients $\boldsymbol{\beta}$, but assumes that the parameters governing the covariance function (σ_Z^2 and $\boldsymbol{\theta}$) have known values.

The parameters σ_Z^2 and $\boldsymbol{\theta}$ are determined through maximum likelihood estimation. This involves taking the log of the probability of observing the response values $\mathbf{g}$ given the covariance matrix $\mathbf{R}$, which can be written as [26]

$$\log[p(\mathbf{g}|\mathbf{R})] = -\frac{1}{n}\log|\mathbf{R}| - \log(\hat{\sigma}_Z^2) \tag{10}$$

where $|\mathbf{R}|$ indicates the determinant of $\mathbf{R}$, and $\hat{\sigma}_Z^2$ is the optimal value of the variance given an estimate of $\boldsymbol{\theta}$ and is defined by

$$\hat{\sigma}_Z^2 = \frac{1}{n}(\mathbf{g} - \mathbf{F}\boldsymbol{\beta})^T\mathbf{R}^{-1}(\mathbf{g} - \mathbf{F}\boldsymbol{\beta}) \tag{11}$$

Maximizing Eq. (10) gives the maximum likelihood estimate of $\boldsymbol{\theta}$, which in turn defines σ_Z^2.

B. Expected Improvement Function

The expected improvement function is used to select the location at which a new training point should be added. The EIF is defined as the expectation that any point in the search space will provide a better solution than the current best solution based on the expected values and variances predicted by the GP model. An important feature of the EIF is that it provides a balance between exploiting areas of the design space where good solutions have been found and exploring areas of the design space where the uncertainty is high. First, recognize that at any point in the design space, the GP prediction $\hat{G}(\,)$ follows a Gaussian distribution:

$$\hat{G}(\mathbf{u}) \sim N[\mu_G(\mathbf{u}), \sigma_G(\mathbf{u})] \tag{12}$$

where the mean $\mu_G(\,)$ and the variance $\sigma_G^2(\,)$ were defined in Eqs. (8) and (9), respectively. The EIF is defined as [2]

$$EI[\hat{G}(\mathbf{u})] \equiv E[\max(G(\mathbf{u}^*) - \hat{G}(\mathbf{u}), 0)] \tag{13}$$

where $G(\mathbf{u}^*)$ is the current best solution chosen from among the true function values at the training points (henceforth referred to as simply G^*). This expectation can then be computed by integrating over the distribution of $\hat{G}(\mathbf{u})$ with G^* held constant:

$$EI[\hat{G}(\mathbf{u})] = \int_{-\infty}^{G^*} (G^* - G) f_{\hat{G}} \, \mathrm{d}G \tag{14}$$

where G is a realization of $\hat{G}$. This integral can be expressed analytically as [2]

$$EI[\hat{G}(\mathbf{u})] = (G^* - \mu_G)\Phi\left(\frac{G^* - \mu_G}{\sigma_G}\right) + \sigma_G \phi\left(\frac{G^* - \mu_G}{\sigma_G}\right) \tag{15}$$

where it is understood that μ_G and σ_G are functions of $\mathbf{u}$.

The point at which the EIF is maximized is selected as an additional training point. With the new training point added, a new GP model is built and then used to construct another EIF, which is then used to choose another new training point, and so on, until the value of the EIF at its maximized point is below some specified tolerance. In [24], this maximization is performed using a Nelder–Mead simplex approach, which is a local optimization method. Because the EIF is often highly multimodal [2], it is expected that Nelder–Mead may fail to converge to the true global optimum. In [2], a branch-and-bound technique for maximizing the EIF is used, but was found to often be too expensive to run to convergence. In this paper, an implementation of the DIRECT global optimization algorithm is used [27].

It is important to understand how the use of this EIF leads to optimal solutions. Equation (15) indicates how much the objective function value at $\mathbf{u}$ is expected to be less than the predicted value at the current best solution. Because the GP model provides a Gaussian distribution at each predicted point, expectations can be calculated. Points with good expected values and even a small variance will have a significant expectation of producing a better solution (exploitation), but so will points that have relatively poor expected values and greater variance (exploration).

The application of EGO to reliability analysis, however, is made more complicated because the response function appears in the equality constraint rather than the objective function [see Eq. (4)]. This makes the maximization of the EIF inappropriate because feasibility is the main concern. This application is therefore a significant departure from the original objective of EGO and requires a new formulation. For this problem, the expected feasibility function is introduced.

C. Expected Feasibility Function

The expected improvement function provides an indication of how much the true value of the response at a point can be expected to be less than the current best solution. It therefore makes little sense to apply this to the forward-reliability problem where the goal is not to minimize the response, but rather to find where it is equal to a specified threshold value. The expected feasibility function (EFF) is introduced here to provide an indication of how well the true value of the response is expected to satisfy the equality constraint $G(\mathbf{u}) = \bar{z}$. Inspired by the contour estimation work in [28], this expectation can be calculated in a similar fashion as Eq. (14) by integrating over a region in the immediate vicinity of the threshold value $\bar{z} \pm \epsilon$:

$$EF[\hat{G}(\mathbf{u})] = \int_{\bar{z}-\epsilon}^{\bar{z}+\epsilon} [\epsilon - |\bar{z} - G|] f_{\hat{G}} \, \mathrm{d}G \tag{16}$$

where G denotes a realization of the distribution $\hat{G}$, as before. Allowing z^+ and z^- to denote $\bar{z} \pm \epsilon$, respectively, this integral can be expressed analytically as

$$\begin{aligned} EF[\hat{G}(\mathbf{u})] = (\mu_G - \bar{z})\left[2\Phi\left(\frac{\bar{z} - \mu_G}{\sigma_G}\right) - \Phi\left(\frac{z^- - \mu_G}{\sigma_G}\right) - \Phi\left(\frac{z^+ - \mu_G}{\sigma_G}\right)\right] - \sigma_G\left[2\phi\left(\frac{\bar{z} - \mu_G}{\sigma_G}\right) - \phi\left(\frac{z^- - \mu_G}{\sigma_G}\right) - \phi\left(\frac{z^+ - \mu_G}{\sigma_G}\right)\right] + \epsilon\left[\Phi\left(\frac{z^+ - \mu_G}{\sigma_G}\right) - \Phi\left(\frac{z^- - \mu_G}{\sigma_G}\right)\right] \end{aligned} \tag{17}$$

where ϵ is proportional to the standard deviation of the GP predictor ($\epsilon \propto \sigma_G$). In this case, z^-, z^+, μ_G, σ_G, and ϵ are all functions of the location $\mathbf{u}$, whereas $\bar{z}$ is a constant. Note that the EFF provides the same balance between exploration and exploitation as is captured in the EIF. Points where the expected value is close to the threshold ($\mu_G \approx \bar{z}$) and points with a large uncertainty in the prediction will have large expected feasibility values.

D. Efficient Global Reliability Analysis Algorithm

The following process makes up the EGRA algorithm:

1) Generate a small number of samples from the true response function.
   a) Only $[(n+1)(n+2)]/2$ samples are used (n is the number of uncertain variables). This initial selection is arbitrary, but the number of samples required to define a quadratic polynomial is used as a convenient rule of thumb.
   b) The samples uniformly span the uncertain space over the bounds $\pm 5\sigma$.
   c) Latin hypercube sampling (LHS) is used to generate the samples.
2) Construct an initial Gaussian process model from these samples.
3) Find the point with maximum expected feasibility.
   a) The expected feasibility function is built with $\epsilon = 2\sigma_G$.
   b) To ensure the global optimum of this multimodal function is found, DIRECT is used.
   c) If the maximum expected feasibility is less than 0.001, go to step 7.
4) Evaluate the true response function at this point.
5) Add this new sample to the previous set and build a new GP model. Go to step 3.
6) When the maximum expected feasibility is small, the Gaussian process model is accurate in the vicinity of the limit state.
   a) The final point may not be near the limit state. This means that no point near the limit state has enough remaining variance to have a larger EFF than the maximum point found. Therefore, the GP model has converged in the vicinity of the limit state.
7) This surrogate model is then used to calculate the probability of failure using multimodal adaptive importance sampling.

Multimodal adaptive importance sampling (MAIS) is typically used to reduce the sampling cost, but it can also be more accurate than using even a large number of Monte Carlo or LHS samples if enough evaluations can be afforded to allow the method to converge. Because all of the MAIS samples are evaluated using the GP model, they can be provided at little cost. Additionally, MAIS would not be easy with other methods that use a random selection of true samples with which to construct the GP model because several iterations would be necessary just to locate the limit state. With EGRA, there are already a large number of samples near the limit state available with which to construct the multimodal sampling density.

IV. Computational Experiments

The EGRA algorithm has been implemented by the authors in the DAKOTA/UQ [29] software, which is the uncertainty quantification component of DAKOTA [30], an open-source software framework for design and performance analysis of computational models on high-performance computers developed at Sandia National Laboratories. Computational results for all methods shown here are either from published results or new analyses performed using the DAKOTA/UQ software with solvers provided by an implementation of the DIRECT algorithm developed at North Carolina State University [27].

A. Numerical Example 1 (Multimodal Function)

The first problem has a highly nonlinear and multimodal response defined by

$$g(\mathbf{x}) = \frac{(x_1^2 + 4)(x_2 - 1)}{20} - \sin\frac{5x_1}{2} - 2 \tag{18}$$

The distribution of x_1 is normal ($\mu = 1.5,\ \sigma = 1$) and x_2 is normal ($\mu = 2.5,\ \sigma = 1$); the variables are uncorrelated. The response level of interest for this study is $\bar{z} = 0$ with failure defined by $g > \bar{z}$. This problem can be formulated as

$$p_f = P[g(\mathbf{x}) > 0] \tag{19}$$

Figure 1 shows a contour plot in x space of this response function throughout the ± 5 standard deviation search space. This function clearly has several local optima to the forward-reliability MPP search problem [see Eq. (4)].

To illustrate how the EGRA method works, this problem is explored in detail. A GP model was built with 10 randomly selected samples. Note that this is larger than the recommended number of initial samples; this is done only to enhance visualization in the early stages of the process. Figure 2 shows contours of the mean value, the variance, and the expected feasibility function for this set of samples.

Comparing the mean value plot to the true contours in Fig. 1 shows that this is a poor approximation. Note that the variance is very low near the training data, but is large throughout the rest of the space. The ⋆ [near $(-3.5, 2.4)$] on the expected feasibility function contour shows the point that maximizes the function and is the new sample that will be added at the next iteration. Selecting this point is a combination of the mean value and the variance, but, at this stage, the large variance dominates and the new point is chosen to explore the search space. Figure 3 shows the same contours after this new point has been added to the training data.

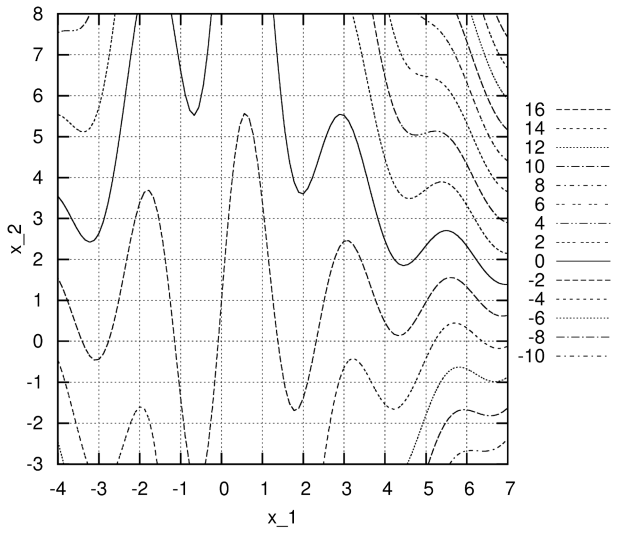


Fig. 1 Contour plot of the multimodal function. The solid line is $g = \bar{z} = 0$.

This is clearly still a poor approximation to the multimodal function. Again, the uncertainty in the model dominates the expected feasibility and the maximum point [at (6.5, 7.5)] is chosen to explore. The next few samples follow in a similar fashion, all chosen to explore due to the high uncertainty from building a GP model with such a small amount of data. Figure 4 shows the mean, variance, and expected feasibility contours after five points have been added (15 total samples in the training data).

The new point selected at this iteration (indicated by the ⋆ on the expected feasibility contour plot) is at a point with low variance, but with an expected value very near the limit state (the solid line on the mean value contour plot). The uncertainty in the model has dropped to a point where the exploitive terms of the EFF have a considerable effect. Note also that the value of the expected feasibility is much smaller than it was with just five fewer samples, showing that the method is converging. Figure 5 shows the contours when there are 30 total samples.

At this point, the mean value plot is very close to the true contour and the variance is very low. The expected feasibility function is shown without the samples so as not to hide the very tight contours. The Gaussian process model "knows" where the limit state is, and so all subsequent samples are "exploitation" samples. The samples chosen are selected by the little uncertainty that still remains. Because the variance is larger as one moves away from the sample data, this ensures that samples are not chosen too close to one another, protecting the structure of the GP. Figure 6 shows the final contour plot of the expected values of the GP model. Note that all of the samples added since Fig. 5 lie almost exactly on the limit state and, in general, the much larger density of samples near the limit state than elsewhere in the search space.

This problem was also solved using reliability methods available in DAKOTA/UQ that reduce the cost of the MPP search through the use of local surrogate models. Two response function approximation methods were investigated [12]: second-order iterated advanced mean value (AMV^2+) and two-point adaptive nonlinear approximation (TANA) [31].

A case using no response function approximation was also investigated. Combining this with first- or second-order integration results in the traditional FORM/SORM formulations. To produce results consistent with an implicit response function, numerical gradients and quasi-Newton-Hessians from Symmetric Rank 1 updates were used. For each method, at the converged MPP, both first- and second-order integration were used to calculate the probability.

For each method, the only algorithmic variation explored here is that the surrogate models can be built in either x space or u space.

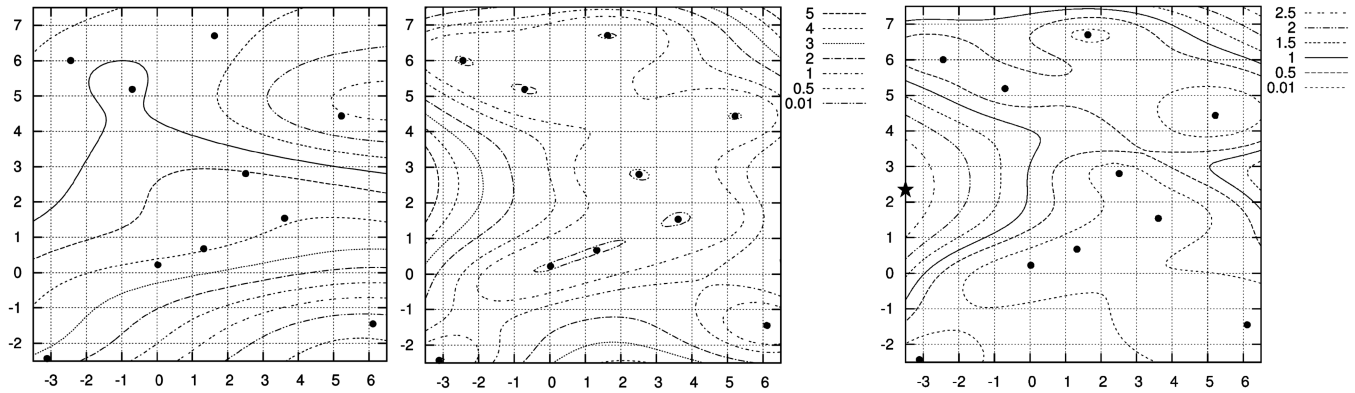


Fig. 2 Contours of the mean value (solid line is the limit state), variance, and expected feasibility function for 10 initial samples. Dots represent the samples used to create the GP; ⋆ is the maximum (EFF) point.

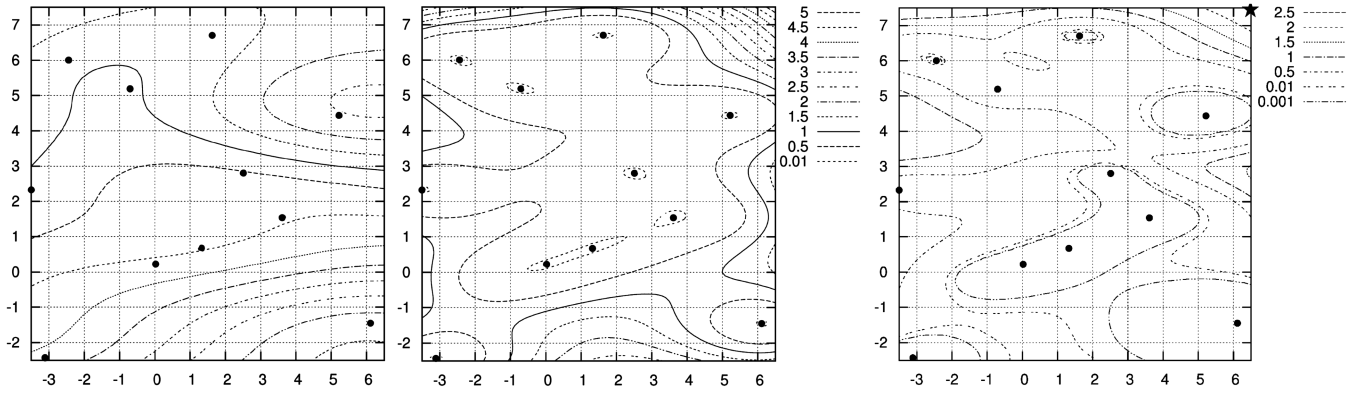


Fig. 3 Contours of the mean value (solid line is the limit state), variance, and expected feasibility function for 11 samples. Dots represent the samples used to create the GP; ⋆ is the maximum (EFF) point.

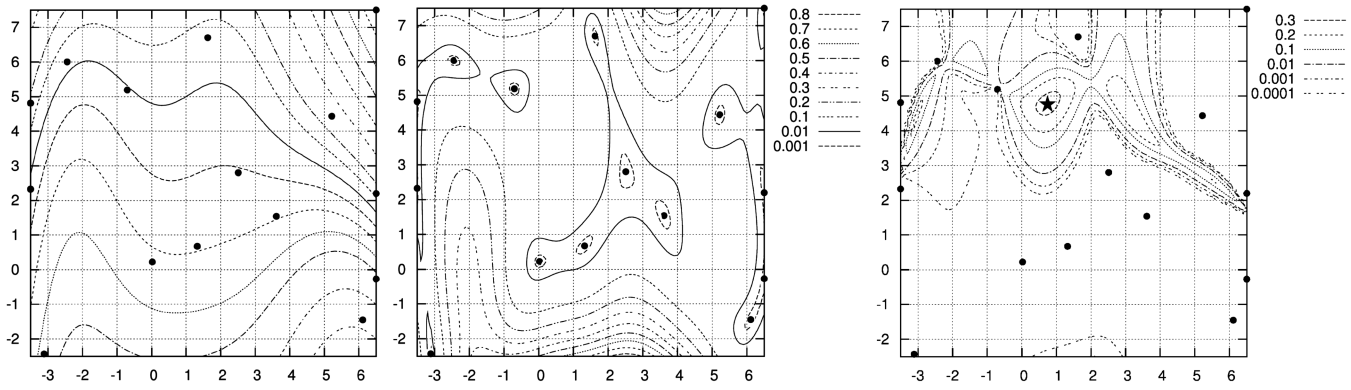


Fig. 4 Contours of the mean value (solid line is the limit state), variance, and expected feasibility function for 15 samples. Dots represent the samples used to create the GP; ⋆ is the maximum (EFF) point.

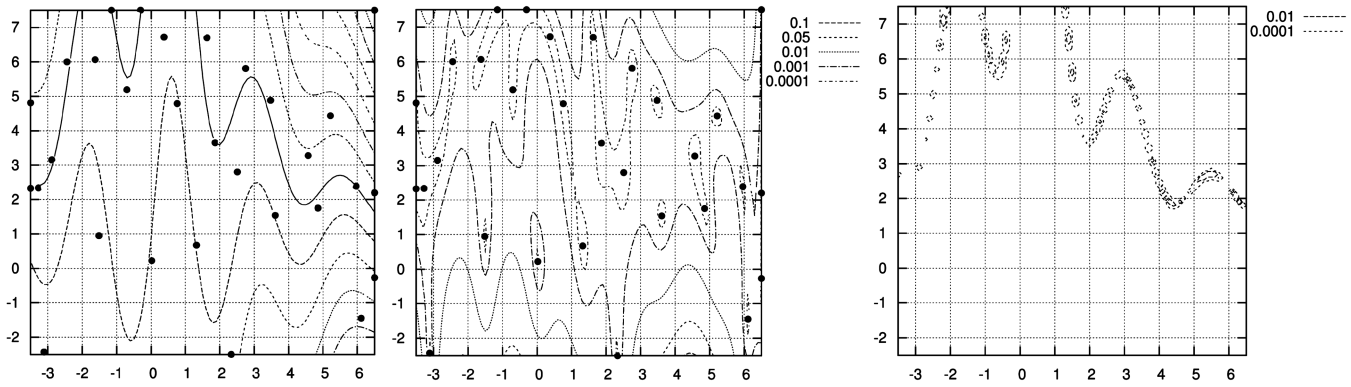


Fig. 5 Contours of the mean value (solid line is the limit state), variance, and expected feasibility function for 30 samples. Dots represent the samples used to create the GP.

Determining which space is appropriate depends upon the form of the response and the space transformation. This choice can have significant effects on both accuracy and efficiency for methods that use low-order approximations to the response function. For instance, if a linear approximation is used for a response that is linear in x space but nonlinear in u space, then building the approximation in x space will yield better results. Gaussian process models are not greatly affected by this choice because they do not rely on curve-fitting or any assumptions on the shape of the response. However, they can smooth out finer details in the true function, meaning they will generally provide better approximations with fewer samples if the true function itself is smooth. If it is known that the response function is smoother in one space than the other, the user should take advantage of that information and build the GP in that space.

Table 1 gives a summary of the results from all methods. To establish an accurate estimate of the true solution, 20 independent simulations were performed using 1 million Latin hypercube samples per simulation. The average probability from these simulations is reported as the "true" solution. Because EGRA is stochastic, it was also run 20 times, and the average probabilities are reported. To measure the accuracy of the method, two errors are reported for the EGRA results: the error in the average probability and the average of the absolute errors from 20 independent simulations. For comparison, the same errors are given for 20 runs of LHS studies composed of 10,000 and 100,000 samples.

Most of the MPP search methods converge to the same MPP and thus report the same probability. Note, however, that the x-space TANA results are not included because the method failed to

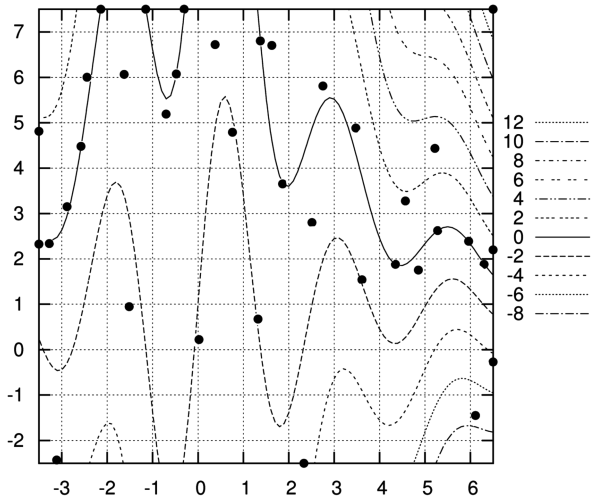


Fig. 6 Final contour of the mean value (solid line is the limit state) with 37 samples. Dots represent the samples used to create the GP.

converge. The probabilities are more accurate when second-order integration is used, but still have significant errors (20%). For this multimodal problem, EGRA is more expensive than AMV^2+, but cheaper than all the other methods, and provides much more accurate results with an average absolute error less than 0.5%. Another important factor to note is the consistency of the method. The average error from 20 EGRA studies is comparable to that from 20 runs of 1 million LHS samples, and is superior to a similar study using 100,000 samples.

B. Numerical Example 2 (Cubic Function)

The second example is a two-dimensional nonlinear function taken from the literature.

$$g(\mathbf{x}) = x_1^3 + x_2^3 - 18 \tag{20}$$

The distribution of x_1 is normal ($\mu = 10,\ \sigma = 5$) and x_2 is normal ($\mu = 9.9,\ \sigma = 5$); the variables are uncorrelated. The response level of interest for this study is $\bar{z} = 0$ with failure defined by $g < \bar{z}$. The problem formulation is then

$$p_f = P[g(\mathbf{x}) < 0] \tag{21}$$

This problem was introduced by Zou et al. [14] to test a method that used a trust-region managed adaptive response surface method to locate the MPP and then used first-order integration, second-order integration, and multimodal adaptive importance sampling to calculate the probability of failure.

Table 2 gives a summary of the results from the same methods investigated in the previous example, plus the published results from [14]. To establish an accurate estimate of the true solution, 20 independent simulations were performed using 1 million Latin hypercube samples per simulation. The average probability from these simulations is reported as the true solution. Again, two errors are reported for EGRA and LHS: the error in the average probability and the average of the absolute errors from 20 independent simulations.

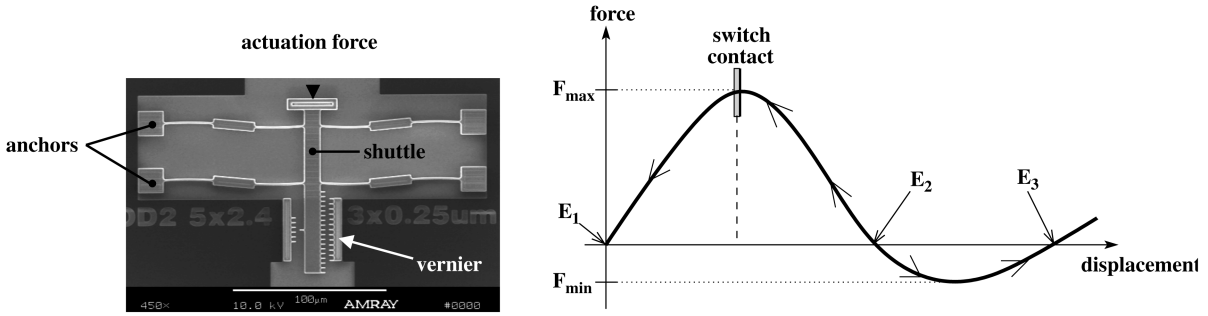
This problem only has one significant MPP, and so the large disparity in some of the local search methods clearly indicates convergence to different points. It is interesting to note that for this problem, x-space TANA provides the most efficient solution, whereas it failed to converge for the previous problem. Once again, second-order integration provides better results, but is still not an accurate approximation to the true shape of the limit state, and so there are still large errors. Because this test problem is not multimodal, performing MAIS with only the MPP as a starting point (as is done by Zou et al. [14]) is sufficient to capture the higher level of nonlinearity in the limit state and generate an excellent result. However, if this method were applied to the previous test problem, it would likely be either much less accurate or require a substantial increase in cost to adequately locate and sample the other significantly probable regions of the space. It should also be pointed out that despite MAIS being a stochastic method, only the error for a single result is reported by Zou et al. [14] and not an average absolute error, as is included for the other sampling methods. For this nonlinear problem, EGRA is more expensive than x-space TANA, but cheaper than all the other methods, and provides much more accurate results. This problem also shows the EGRA results to be equally as consistent as the exhaustive LHS studies considered to be the "truth."

Table 1 Results for numerical example 1

Reliability method	Function evaluations	First-order p_f (% error)	Second-order p_f (% error)	Sampling p_f (% error, avg. error)
No approximation	70	0.11797 (277.0%)	0.02516 (−19.6%)	——
x space AMV^2+	26	0.11797 (277.0%)	0.02516 (−19.6%)	——
u space AMV^2+	26	0.11777 (277.0%)	0.02516 (−19.6%)	——
u space TANA	131	0.11797 (277.0%)	0.02516 (−19.6%)	——
LHS solution	10 k	——	——	0.03117 (0.385%, 2.847%)
LHS solution	100 k	——	——	0.03126 (0.085%, 1.397%)
LHS solution	1 M	——	——	0.03129 (truth, 0.339%)
x space EGRA	35.1	——	——	0.03134 (0.155%, 0.433%)
u space EGRA	35.2	——	——	0.03133 (0.136%, 0.296%)

Table 2 Results for numerical example 2

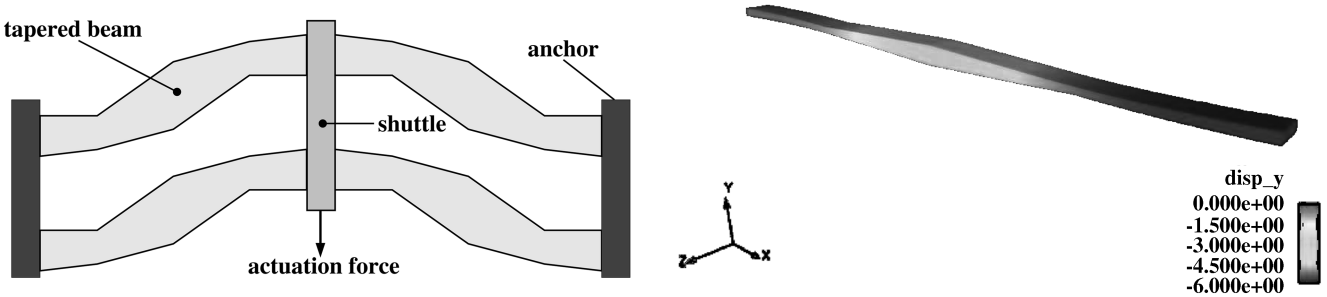
Reliability method	Function evaluations	First-order p_f (% error)	Second-order p_f (% error)	Sampling p_f (% error, avg. error)
No approximation	125	0.01301 (127.1%)	0.004164 (−27.3%)	——
x space AMV^2+	66	0.01301 (127.1%)	0.004165 (−27.3%)	——
u space AMV^2+	66	0.01301 (127.1%)	0.004165 (−27.3%)	——
x space TANA	21	0.01301 (127.1%)	0.004157 (−27.4%)	——
u space TANA	36	0.01301 (127.1%)	0.004165 (−27.3%)	——
Zou et al. [14] MAIS	97	0.02560 (346.9%)	0.016200 (182.8%)	——
LHS solution	10 k	——	——	0.005590 (2.413%, 7.279%)
LHS solution	100 k	——	——	0.005686 (0.746%, 3.637%)
LHS solution	1 M	——	——	0.005728 (truth, 0.757%)
Zou et al. [14] MAIS	560	——	——	0.005750 (0.380%, no data)
x space EGRA	28.7	——	——	0.005694 (0.590%, 0.725%)
u space EGRA	30.1	——	——	0.005703 (0.446%, 0.737%)



a) Scanning electron micrograph of a MEMS bistable mechanism in its second stable position. The attached vernier provides position measurements

b) Schematic of force-displacement curve for bistable MEMS mechanism. The arrows indicate stability of equilibria E_1 and E_3 and instability of E_2

Fig. 7 Bistable MEMS mechanism.



a) Schematic of a tapered beam bistable mechanism in as-fabricated position (not to scale)

b) Scale rendering of tapered beam leg for bistable mechanism

Fig. 8 Tapered beams for bistable MEMS mechanism.

C. Bistable Microelectromechanical System Device

This final application problem involves analyzing the reliability of a microelectromechanical system (MEMS) device presented in [32,33]. The device is a compliant bistable mechanism, where, instead of mechanical joints, material elasticity enables the bistability of the mechanism [34–36]. Figure 7a contains an electron micrograph of a MEMS compliant bistable mechanism in its second stable position. The first stable position is the as-fabricated position. One achieves transfer between stable states by applying force to the center shuttle via a thermal actuator, electrostatic actuator, or other means to move the shuttle past an unstable equilibrium, thus making it useful as a microswitch, relay, or nonvolatile memory.

Bistable switch actuation characteristics depend on the relationship between actuation force and shuttle displacement for the manufactured switch. Figure 7b contains a schematic of a typical force-displacement curve for a bistable mechanism. The switch characterized by this curve has three equilibria: E_1 and E_3 are stable equilibria, whereas E_2 is an unstable equilibrium (arrows indicate stability). A device with such a force-displacement curve could be used as a switch or actuator by setting the shuttle to position E_3, as shown in Fig. 7a (requiring large actuator force $F_{\max}$) and then actuating by applying the comparably small force $F_{\min}$ in the opposite direction to transfer back through E_2 toward the equilibrium E_1. One could use this force profile to complete a circuit by placing a switch contact near the displaced position corresponding to maximum (closure) force, as illustrated. Repeated actuation of the switch relies on being able to reset it with actuation force $F_{\max}$.

The device design considered in [32,33] is similar to that in the electron micrograph in Fig. 7a. The primary structural difference in the present design is the tapering of the legs, shown schematically in Fig. 8a. Figure 8b shows a scale drawing of one tapered beam leg (one-quarter of the full switch system). A single leg of the device is approximately 100 μm wide and 5–10 μm tall. This topology is a cross between the fully compliant bistable mechanism reported in [35] and the thickness-modulated curved beam in [37].

Because of manufacturing processes, fabricated geometry can deviate significantly from design-specified beam geometry. As a consequence of photo lithography and etching processes, fabricated in-plane geometry edges (contributing to widths and lengths) can be 0.1 ± 0.08 μm less than specified. This variability in the manufactured geometry leads to substantial variability in the positions of the stable equilibria and in the maximum and minimum force on the force-displacement curve. The manufactured thickness

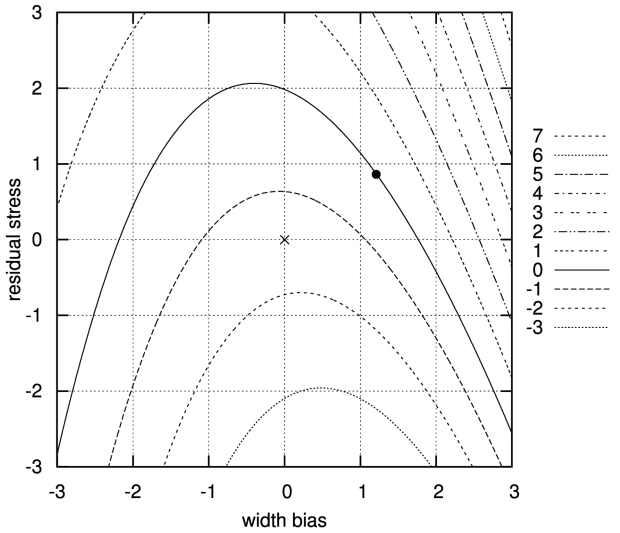


Fig. 9 Contour plot of $F_{\min}(\mathbf{d}, \mathbf{x})$ as a function of uncertain variables x. Solid line is the limit state $g(\mathbf{x}) = 0.0$; x is the mean; circle is the MPP.

Table 3 Results for the bistable MEMS device problem

Reliability method	Function evaluations	First-order p_f	Sampling p_f (avg. error)
No approximation	396	0.06815	——
x space AMV^2+	50	0.06815	——
u space AMV^2+	50	0.06815	——
x space TANA	60	0.06816	——
u space TANA	81	0.06815	——
LHS solution	1 k	——	0.1130 (3.953%)
x space EGRA	15.3	——	0.1099 (0.153%)
u space EGRA	15.2	——	0.1099 (0.138%)

of the device is also random, though this does not contribute as much to variability in the force-displacement behavior. Stochastic material properties such as Young's modulus and residual stress also influence the characteristics of the fabricated beam. For this application, two key random variables are considered: ΔW (edge bias on beam widths, which yields effective manufactured widths of $W_i + \Delta W$, $i = 0, \ldots, 4$) and S_r (residual stress in the manufactured device). The distribution of ΔW is normal ($\mu = -0.2$, $\sigma = 0.08$) measured in micrometers, and the distribution of S_r is normal ($\mu = -11$, $\sigma = 4.13$) measured in megapascal.

Given the geometric design variables $\mathbf{d}$ and the specified random variables $\mathbf{x} = [\Delta W, S_r]$, the limit state is the minimum actuation force $F_{\min}(\mathbf{d}, \mathbf{x})$ and failure is defined to be an actuation force with magnitude less than 5.0. The problem formulation is then

$$p_f = P[F_{\min}(\mathbf{d}, \mathbf{x}) + 5.0 > 0] \tag{22}$$

Figure 9 displays the results of a parameter study of the response function $g(\mathbf{x}) = F_{\min}(\mathbf{d}, \mathbf{x}) + 5.0$ as a function of the uncertain variables $\mathbf{x}$ for the given design. The contour plot is scaled to a ± 3 standard deviation range in the transformed u space. The limit state $g(\mathbf{x}) = 0$ (equivalent to $F_{\min}(\mathbf{d}, \mathbf{x}) = -5.0$) is indicated by the solid line. For some design variable sets $\mathbf{d}$ (not depicted), the limit state is relatively well behaved in the range of interest and first-order probability integrations would be sufficiently accurate. For the design variable set used to generate Fig. 9, the limit state has significant nonlinearity, and thus demands more sophisticated probability integrations. The most probable point converged to by the MPP search methods is denoted in Fig. 9 by the circle.

Table 3 gives a summary of the results from the same methods investigated in the previous examples. Second-order integration results are not provided because the quasi Hessians that are computed during the MPP search, although helpful for informing the MPP search algorithm, do not provide adequate curvature information to generate accurate probability estimates. For this problem, the probabilities calculated by second-order integration were found to be less than those calculated by first-order integration. Inspection of Fig. 9 shows that this is clearly not the case. Moreover, even when numerical Hessians are used, second-order integration is not possible for this problem because the large, negative principal curvatures create numerical difficulties. For the MPP search methods, the first-order probabilities are the most accurate available. Finite element analysis of this MEMS device is too expensive for the exhaustive LHS studies used in the previous examples, and so 20 simulations with only 1000 samples each were performed. Because the true solution to this problem is not known, the errors in the various solutions cannot be provided. However, for the stochastic methods, the average absolute error from among the 20 runs is reported. This error is calculated by comparing the individual solutions of a method to the mean solution of that method, and so it is similar to the coefficient of variation of the solutions.

Because the true solution is unknown, no definitive argument can be made on the accuracy of the EGRA solution. However, given the similarity to the LHS solution and the fact that the MPP search methods rely on low-order approximations to the shape of a limit state that is known to be multimodal (see Fig. 9), it is reasonable to say that EGRA provides a more accurate estimate to the probability of failure than the MPP search methods. A remarkable feature of these results is that EGRA requires less than one-third the number of function evaluations than even the cheapest of the MPP search methods, and the average error in the EGRA solutions is considerably smaller than for the LHS solutions. Given the larger variance in the LHS solutions, the accuracy of EGRA relative to the LHS results for the preceding examples, and the fact that the average EGRA solution here lies easily within the LHS error bounds, there is the implication that the average EGRA solution may actually be more accurate than the average LHS solution. Moreover, that EGRA solution comes at a cost of only about 15 true function evaluations versus the 1000 required for LHS.

V. Conclusions

Current reliability methods require engineers to sacrifice either accuracy or efficiency. This paper has presented a new method called efficient global reliability analysis that uses Gaussian process models and is both accurate for any arbitrarily shaped limit state and efficient for even computationally expensive response functions. These are desirable features for application to expensive implicit performance functions, where the shape of the limit state may be unknown and impossible to discern.

In the example problems used to test this new method, EGRA produced results that are far more accurate than MPP search methods, while requiring far fewer true function evaluations than sampling methods. The authors realize that these problems are too few to provide definitive judgment on the performance of the new method; also, all problems investigated involve only two uncorrelated normal variables. However, it is expected that EGRA will perform well regardless of distribution type and correlation because the construction of the surrogate model is completely unconcerned with the fact that it is operating in a probabilistic space. The concept of one point in the space being more or less probable than another is not introduced until the sampling stage, which is easily handled for any type of distribution and/or correlation. Investigation is ongoing to determine how well EGRA will scale to problems with larger dimensions.

Also being investigated is how EGRA might be used in reliability-based design optimization (RBDO). Inspection of Fig. 9 shows that, although there is only one truly *most* probable point for the MEMS device problem, there is clearly an additional local solution for $\mathbf{u}$ satisfying Eq. (4) at approximately $(-1.7, 1)$. This is a common occurrence when using RBDO methods where the optimizer has a tendency to push the design into a corner where the mean response is encircled by the failure domain. Because EGRA can easily provide accurate probability estimates in these situations, this motivates the use of EGRA in RBDO. Research aimed at developing such a formulation is ongoing.

Acknowledgments

This study was partially supported by funds from the National Science Foundation, through the Integrative Graduate Education and Research Traineeship multidisciplinary doctoral program in Risk and Reliability Engineering at Vanderbilt University (Grant No. 013429). The authors would also like to express their gratitude to the Sandia Computer Science Research Institute for support of this collaborative work with Sandia National Laboratories. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of

Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.

References

[1] Thacker, B. H., Riha, D. S., Millwater, H. R., and Enright, M. P., "Errors and Uncertainties in Probabilistic Engineering Analysis," *Proceedings of the 42nd Structures, Structural Dynamics, and Materials Conference*, AIAA Paper 2001-1239, April 2001.

[2] Jones, D., Schonlau, M., and Welch, W., "Efficient Global Optimization of Expensive Black-Box Functions," *Journal of Global Optimization*, Vol. 13, No. 4, 1998, pp. 455–492. doi:10.1023/A:1008306431147

[3] Giunta, A. A., McFarland, J. M., Swiler, L. P., and Eldred, M. S., "Promise and Peril of Uncertainty Quantification via Response Surface Approximations," *Structure & Infrastructure Engineering: Maintenance, Management, Life-Cycle Design & Performance*, Vol. 2, No. 3, 2006, pp. 175–189. doi:10.1080/15732470600590507

[4] Wang, L., Beeson, D., Akkaram, S., and Wiggs, G., "Gaussian Process Meta-Models for Efficient Probabilistic Design in Complex Engineering Design Spaces," *Proceedings of ASME International Design Engineering Technical Conference & Computers and Information in Engineering Conference*, American Society of Mechanical Engineers Paper DETC2005-85406, Sept. 2005.

[5] Haldar, A., and Mahadevan, S., *Probability, Reliability, and Statistical Methods in Engineering Design*, Wiley, New York, 2000.

[6] Rosenblatt, M., "Remarks on a Multivariate Transformation," *Annals of Mathematical Statistics*, Vol. 23, No. 3, 1952, pp. 470–472. doi:10.1214/aoms/1177729394

[7] Der Kiureghian, A., and Liu, P. L., "Structural Reliability Under Incomplete Probability Information," *Journal of Engineering Mechanics, American Society of Civil Engineers*, Vol. 112, No. 1, 1986, pp. 85–104.

[8] Box, G. E. P., and Cox, D. R., "Analysis of Transformations," *Journal of the Royal Statistical Society, Series B*, Vol. 26, No. 2, 1964, pp. 211–252.

[9] Rackwitz, R., and Fiessler, B., "Structural Reliability Under Combined Random Load Sequences," *Computers and Structures*, Vol. 9, No. 5, 1978, pp. 489–494. doi:10.1016/0045-7949(78)90046-9

[10] Chen, X., and Lind, N. C., "Fast Probability Integration by Three-Parameter Normal Tail Approximation," *Structural Safety*, Vol. 1, No. 4, 1983, pp. 269–276. doi:10.1016/0167-4730(82)90003-0

[11] Wu, Y.-T., and Wirsching, P. H., "New Algorithm for Structural Reliability Estimation," *Journal of Engineering Mechanics, American Society of Civil Engineers*, Vol. 113, No. 9, 1987, pp. 1319–1336.

[12] Eldred, M. S., and Bichon, B. J., "Second-Order Formulations in DAKOTA/UQ," *Proceedings of the 47th AIAA/ASME/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, AIAA Paper 2006-1828, 2006.

[13] Eldred, M. S., Agarwal, H., Perez, V. M., Wojtkiewicz, S. F., Jr., and Renaud, J. E., "Investigation of Reliability Method Formulations in DAKOTA/UQ," *Structure & Infrastructure Engineering: Maintenance, Management, Life-Cycle Design & Performance*, Vol. 3, No. 3, 2007, pp. 199–213. doi:10.1080/15732470500254618

[14] Zou, T., Mourelatos, Z., Mahadevan, S., and Meernik, P., "Reliability Analysis of Automotive Body-Door Subsystem," *Reliability Engineering and System Safety*, Vol. 78, No. 3, 2002, pp. 315–324. doi:10.1016/S0951-8320(02)00178-3

[15] Mahadevan, S., and Shi, P., "Multiple Linearization Method for Nonlinear Reliability Analysis," *Journal of Engineering Mechanics*, Vol. 127, No. 11, 2001, pp. 1165–1173. doi:10.1061/(ASCE)0733-9399(2001)127:11(1165)

[16] Zou, T., Mourelatos, Z., Mahadevan, S., and Tu, J., "Indicator Response Surface Method for Simulation-Based Reliability Analysis," *Journal of Mechanical Design*, Vol 130, No. 7, 2008, p. 071401. doi:10.1115/1.2918901

[17] Breitung, K., "Asymptotic Approximation for Multinormal Integrals," *Journal of Engineering Mechanics, American Society of Civil Engineers*, Vol. 110, No. 3, 1984, pp. 357–366.

[18] Hohenbichler, M., and Rackwitz, R., "Improvement of Second-Order Reliability Estimates by Importance Sampling," *Journal of Engineering Mechanics, American Society of Civil Engineers*, Vol. 114, No. 12, 1988, pp. 2195–2199.

[19] Hong, H. P., "Simple Approximations for Improving Second-Order Reliability Estimates," *Journal of Engineering Mechanics*, Vol. 125, No. 5, 1999, pp. 592–595. doi:10.1061/(ASCE)0733-9399(1999)125:5(592)

[20] Tvedt, L., "Distribution of Quadratic Forms in Normal Space-Application to Structural Reliability," *Journal of Engineering Mechanics*, Vol. 116, No. 6, 1990, pp. 1183–1197. doi:10.1061/(ASCE)0733-9399(1990)116:6(1183)

[21] Harbitz, A., "Efficient Sampling Method for Probability of Failure Calculation," *Structural Safety*, Vol. 3, No. 2, 1986, pp. 109–115. doi:10.1016/0167-4730(86)90012-3

[22] Wu, Y.-T., "Adaptive Importance Sampling Method for Structural System Reliability Analysis," *Reliability Technology 1992*, Vol. AD-28, edited by T. A. Cruse, ASME Winter Annual Meeting, 1992, pp. 217–231.

[23] Dey, A., and Mahadevan, S., "Ductile Structural System Reliability Analysis Using Adaptive Importance Sampling," *Structural Safety*, Vol. 20, No. 2, 1998, pp. 137–154. doi:10.1016/S0167-4730(97)00033-7

[24] Huang, D., Allen, T. T., Notz, W. I., and Zeng, N., "Global Optimization of Stochastic Black-Box Systems via Sequential Kriging Meta-Models," *Journal of Global Optimization*, Vol. 34, No. 3, 2006, pp. 441–466. doi:10.1007/s10898-005-2454-3

[25] Cressie, N. A. C., *Statistics for Spatial Data*, revised edition, Wiley, New York, 1993.

[26] Sacks, J., Schiller, S. B., and Welch, W., "Design for Computer Experiments," *Technometrics*, Vol. 31, No. 1, 1989, pp. 41–47. doi:10.2307/1270363

[27] Gablonsky, J. M., "Implementation of the DIRECT Algorithm," Center for Research in Scientific Computation, Technical Rept. CRSC-TR98-29, North Carolina State Univ., Raleigh, NC, Aug. 1998.

[28] Ranjan, P., Bingham, D., and Michailidis, G., "Sequential Experiment Design for Contour Estimation from Complex Computer Codes," *Technometrics*, 2008 (to be published).

[29] Wojtkiewicz, S. F., Jr., Eldred, M. S., Field, R. V., Jr., Urbina, A., and Red-Horse, J. R., "Toolkit For Uncertainty Quantification in Large Computational Engineering Models," *Proceedings of the 42nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, AIAA Paper 2001-1455, April 2001.

[30] Eldred, M. S., Giunta, A. A., Brown, S. L., Adams, B. M., Dunlavy, D. M., Eddy, J. P., Gay, D. M., Griffin, J. D., Hart, W. E., Hough, P. D., Kolda, T. G., Martinez-Canales, M. L., Swiler, L. P., Watson, J.-P., and Williams, P. J., "DAKOTA, a Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis," Ver. 4.0 User's Manual, Sandia National Lab. Technical Rept. SAND 2006-6337, Oct. 2006 (Revised).

[31] Xu, S., and Grandhi, R. V., "Effective Two-Point Function Approximation for Design Optimization," *AIAA Journal*, Vol. 36, No. 12, 1998, pp. 2269–2275.

[32] Adams, B. M., Eldred, M. S., Wittwer, J., and Massad, J., "Reliability-Based Design Optimization for Shape Design of Compliant Micro-Electro-Mechanical Systems," *Proceedings of the 11th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, AIAA Paper 2006-7000, Sept. 2006.

[33] Adams, B. M., Bichon, B. J., Carnes, B., Copps, K. D., Eldred, M. S., Hopkins, M. H., Neckels, D. C., Notz, P. K., Subia, S. R., and Wittwer, J. W., "Solution-Verified Reliability Analysis and Design of Bistable MEMS Using Error Estimation and Adaptivity," Sandia Technical Rept. SAND 2006-6286, Oct. 2006.

[34] Ananthasuresh, G. K., Kota, S., and Gianchandani, Y., "Methodical Approach to the Design of Compliant Micromechanisms," *Proceedings of IEEE Solid-State Sensor and Actuator Workshop*, Inst. of Electrical and Electronics Engineers, New York, 1994, pp. 189–192.

[35] Jensen, B. D., Parkinson, M. B., Kurabayashi, K., Howell, L. L., and Baker, M. S., "Design Optimization of a Fully-Compliant Bistable Micromechanism," *Proceedings 2001 ASME International Mechanical Engineering Congress and Exposition*, American Society of Mechanical Engineers IMECE2001/MEMS-23852, Nov. 2001.

[36] Kemeny, D. C., Howell, L. L., and Magleby, S. P., "Using Compliant Mechanisms to Improve Manufacturability in MEMS," *Proceedings of the 2002 ASME DETC*, American Society of Mechanical Engineers, New York, No. DETC2002/DFM-34178, 2002.

[37] Qiu, J., and Slocum, A. H., "Curved-Beam Bistable Mechanism," *Journal of Microelectromechanical Systems*, Vol. 13, No. 2, 2004, pp. 137–146. doi:10.1109/JMEMS.2004.825308

A. Messac
Associate Editor