SAND2006-5024C

Multimodal Reliability Assessment for Complex Engineering Applications using Sequential Kriging Optimization

B. J. Bichon[††,*], M. S. Eldred[**,†], L. P. Swiler[**,‡], S. Mahadevan[††,§], and J. M. McFarland[††,¶]

[††] *Vanderbilt University, Nashville, TN 37235*

[**] *Sandia National Laboratories[‖], Albuquerque, NM 87185*

As engineering applications become increasingly complex, they are often characterized by implicit response functions that are both expensive to evaluate and nonlinear in their behavior. Reliability assessment given this type of response is difficult with available methods. Current reliability methods focus on the discovery of 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. This paper describes the application of sequential kriging optimization to reliability assessment through an efficient global search for multiple most probable points. By locating multiple most probable points of failure, more complex limit states can be modeled, leading to more accurate probability integration. Several possible formulations of this method will be explored and applied to a collection of example problems that currently available methods have difficulty solving either accurately or efficiently.

I. Introduction

Accurate reliability assessment is a problem of great importance to the engineering community. Poor solutions lead to designs that are either unreliable or overly expensive. However, the ability to accurately quantify the uncertainty in a design becomes increasingly difficult as the analysis of the design becomes more expensive and its behavior more nonlinear.

Current methods of reliability assessment solve an optimization problem to locate the most probable point of failure (MPP), 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 converge to sub-optimal solutions for response functions that possess multiple local optima. Convergence to sub-optimal MPPs and limit state approximations that may be inaccurate, make MPP search methods unreliable in practice. Engineers are then forced to revert to sampling methods, which are impractical when evaluation of the response function is expensive.

A reliability assessment method that is both efficient when applied to expensive response functions and accurate when the response function is highly nonlinear is needed. This paper investigates the application of a global optimization tool known as sequential kriging optimization to the search for multiple MPPs. By locating multiple points on the limit state, more accurate approximations to nonlinear limit states can be built, resulting in a more accurate assessment of the reliability.

Sequential kriging optimization[12] (SKO) is an adaptation of efficient global optimization[13] (EGO), which was developed to facilitate the optimization of expensive implicit response functions. These methods build an initial kriging model as a global surrogate for the response function, then intelligently select additional samples to be added for the next kriging model. 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 optimal solution has been found. Because SKO is a global optimization method, it is able to locate multiple local optima, which will allow for the discovery of multiple MPPs.

---

[*] Doctoral Candidate and NSF IGERT 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.

[¶] Graduate Student and NSF IGERT Fellow, Department of Mechanical Engineering.

[‖] Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed-Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

Section II describes the reliability assessment problem and traditional methods of solving it. Section III gives an overview of SKO and outlines some ideas on how it might be modified for application to MPP search. Section IV details a preliminary investigation into some of the ideas posed in Section III. Finally, Section V provides concluding remarks and some initial impressions on the promise of this method.

II. Reliability Assessment

The goal of reliability assessment is to determine the probability that an engineered device, component, system, etc. will fail in service given that its behavior is dependent in part 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 be defined by

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

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 if it is available, evaluating the multiple integral is impractical.[9] Because of these complications, methods of approximating this integral are used in practice.

A. MPP Search 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 or MPP. An approximation to the limit state is then formed at this point to facilitate the integration required to compute the probability of failure.

The MPP search is performed in uncorrelated standard normal space ("u-space") because it simplifies the probability integration; in this space, the distance from the origin to the MPP is equivalent to the number of input standard deviations from the mean response at which the limit state lies. This distance is known as the reliability index and is denoted by β. The transformation from correlated non-normal distributions (x-space) to uncorrelated standard normal distributions (u-space) is nonlinear in general, and possible approaches include the Rosenblatt,[15] Nataf,[5] and Box-Cox[1] transformations. The nonlinear transformations may also be linearized, and common approaches for this included the Rackwitz-Fiessler[14] two-parameter equivalent normal and the Chen-Lind[3] and Wu-Wirsching[18] three-parameter equivalent normals. This work employs the Nataf nonlinear transformation, which occurs in the following two steps. To transform between the original correlated x-space variables and correlated standard normals ("z-space"), the CDF matching condition is used:

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

where $F()$ is the cumulative distribution function of the original probability distribution. Then, to transform between correlated z-space variables and 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 non-normals in x-space has been modified for z-space.[5]

The forward reliability analysis algorithm for computing the probability/reliability level that corresponds to a specified response level is called the reliability index approach (RIA), and the inverse reliability analysis algorithm for computing the response level that corresponds to a specified probability/reliability level is called the performance measure approach (PMA).[17] The differences between the RIA and PMA formulations appear in the objective function and equality constraint formulations in the MPP searches. For RIA, the MPP search for achieving the specified response level $\bar{z}$ is formulated as

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

and for PMA, the MPP search for achieving the specified probability/reliability level $\bar{p}, \bar{\beta}$ is formulated as

$$\begin{aligned} \text{minimize} \quad & \pm G(\mathbf{u}) \\ \text{subject to} \quad & \mathbf{u}^T\mathbf{u} = \bar{\beta}^2 \end{aligned} \tag{5}$$

where $\mathbf{u}$ is a vector centered at the origin in u-space and $g(\mathbf{x}) \equiv G(\mathbf{u})$ by definition. In the RIA case, 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 has focused on the use of local and multipoint surrogate models to reduce the expense of the MPP search.[7,8] All of these MPP search methods employ local optimization techniques and converge to a single MPP. But the limit state of a complex engineering application may be multimodal and possess multiple significantly probable points of failure. These points are commonly referred to as multiple most probable points (multiple MPPs) despite the misnomer. The SKO method for uncertainty quantification proposed here uses a global surrogate model and global optimization methods to reduce expense and locate multiple MPPs.

B. Probability Integration

For an RIA formulation, after the MPP is found and the reliability index β is known, the next step is to integrate over the failure region to calculate the probability of failure. This can be greatly simplified from Eqn. 1 by approximating the shape of the limit state with one over which it is easier to integrate. The simplest approximation is the first-order reliability method (FORM), which approximates the limit state as a linear function. Because β represents the distance from the mean response to the MPP in standard normal space, the probability integration simplifies to:

$$p_f = \Phi(-\beta) \tag{6}$$

where $\Phi()$ is the standard normal cumulative distribution function. Another alternative is the second-order reliability method (SORM), which incorporates some curvature in the limit state approximation.[2,10,11] Breitung applies a correction based on asymptotic analysis:[2]

$$p_f = \Phi(-\beta) \prod_{i=1}^{n-1} \frac{1}{\sqrt{1+\beta\kappa_i}} \tag{7}$$

where κ_i are the principal curvatures of the limit state (the eigenvalues of an orthonormal transformation of $\nabla^2_{\mathbf{u}} G$, taken positive for a convex limit state). This method essentially uses a parabolic approximation to the limit state and is more accurate for large values of β because it collapses to first-order integration at $\beta = 0$. An alternative correction in Ref. 10 is consistent with Breitung's correction in the asymptotic regime ($\beta \to \infty$) but does not approach first-order integration as $\beta \to 0$:

$$p_f = \Phi(-\beta) \prod_{i=1}^{n-1} \frac{1}{\sqrt{1+\psi(-\beta)\kappa_i}} \tag{8}$$

where $\psi() = \frac{\phi()}{\Phi()}$ and $\phi()$ is the standard normal density function. Ref. 11 applies further corrections to Eqn. 8 based on point concentration methods.

Each of these methods makes some approximation to the shape of the limit state, making them inaccurate if the approximation is poor. If multiple MPPs are present, the true shape of the limit state cannot be linear and is likely not parabolic, so more advanced methods are needed. One possibility is to borrow concepts from system reliability where the total probability of failure is calculated as the probability of the union of multiple distinct modes of failure. If each of the n multiple MPPs on a single limit state are treated as single MPPs on n multiple limit states (with each limit state considered a distinct failure event E), then the probability of failure could be computed as:

$$p_f = P(E_1 \cup E_2 \cup \cdots \cup E_{n-1} \cup E_n) \tag{9}$$

Assuming the individual failure modes are independent, DeMorgan's Rule can be used to simplify this to:

$$p_f = 1 - \prod_{i=1}^{n} P(\overline{E_i}) = 1 - \prod_{i=1}^{n} [1 - P(E_i)] \tag{10}$$

If the reliability index for each MPP i is denoted by β_i and FORM is used for the probability integration at each MPP, the total probability of failure could be calculated by:

$$p_f = 1 - \prod_{i=1}^{n} [1 - \Phi(-\beta_i)] \tag{11}$$

and a similar formulation could be used for SORM methods.

Another possibility is to 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, so they can be more accurate than FORM and SORM, but they can also be prohibitively expensive because they generally require a large number of response function evaluations. The simplest sampling method is Basic Monte Carlo. Samples of the input variables are randomly

generated based on their distributions and the response function is then evaluated at these input values. The value of the response function at this sample point is then compared to the limit state to determine if the observed response is a success or failure. The probability of failure is then calculated as simply the ratio of observed failures to the total number of observations. One major drawback to this method is that it will generate a large number of samples in the high-probability region of the response function, but because engineers are typically concerned with high-reliability problems, this region is of little interest as the limit state most likely lies in a much less-probable region of the design space.

Importance sampling methods avoid this problem by centering the sampling density function at the MPP. This ensures the samples will lie in a more interesting region of the design space and is a much more efficient sampling method than Basic Monte Carlo. Adaptive importance sampling (AIS) further improves the efficiency by adaptively updating the sampling density function. Multimodal adaptive importance sampling[6,19] is a variation of AIS that allows for the use of multiple sampling densities making it better suited for cases where multiple sections of the limit state are highly probable.

Note that Importance Sampling methods require 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, thus making these methods inapplicable. Because SKO is a global optimization method that does not depend on the availability of accurate gradient information, convergence to the MPP should be more reliable. Moreover, SKO has the ability to locate multiple MPPs, which would provide multiple starting points and a true multimodal sampling density for the initial steps of multimodal AIS. An additional advantage of the SKO method proposed is that a byproduct of the MPP search process is a kriging model that is accurate in the vicinity of the limit state, thereby providing an inexpensive surrogate that can be used to provide response function samples. Using SKO to locate multiple MPPs, and then using the resulting kriging model to provide function evaluations in multimodal AIS for the probability integration, could result in an accurate and efficient reliability assessment tool.

III. Sequential Kriging Optimization

Sequential kriging optimization is a global optimization method put forth by Huang et al.,[12] but has its roots in the efficient global optimization (EGO) method from Jones et al.[13] The main difference between SKO and EGO is the formulation of what is known as the expected improvement function (EIF), which is the feature that sets all EGO-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 kriging model by maximizing the amount of improvement in the objective function that can be expected by adding that point. The general procedure of these EGO-type methods is:

1. Build an initial kriging model of the objective function
2. Use cross validation to ensure that the kriging model is satisfactory.
3. Find the point that maximizes the EIF. If the EIF value at this point is sufficiently small, stop.
4. Evaluate the objective function at the point where the EIF is maximized. Update the kriging model using this new point. Go to Step 3.

The following sections discuss the construction of the kriging model used, the form of the EIF, and then some ideas on how that EIF could be altered for SKO's application to MPP search.

A. Kriging Model

Kriging models are set apart from most other surrogate models by the fact that they provide not only a prediction of the modeled function value at a point, but also an indication of the uncertainty of that prediction. This uncertainty is a result of the construction of the covariance function, which 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 which 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 function being modeled $Y()$ contains some random noise (modeled as independent identically distributed normal deviates[12]), and can be described by:[4]

$$Y(\mathbf{x}) = \mathbf{h}(\mathbf{x})^T \boldsymbol{\beta} + Z(\mathbf{x}) + \epsilon \tag{12}$$

where $\mathbf{h}()$ is the trend of the model, $\boldsymbol{\beta}$ is the vector of trend coefficients, $Z()$ is a stationary Gaussian process with zero mean and covariance defined below that describes the departure of the model from its underlying trend, and ϵ is the noise

in the true function. The trend of the model can be assumed to be any function, but taking it to be a constant value is generally sufficient.[16] The covariance between outputs of the Gaussian process $Z()$ at points $\mathbf{a}$ and $\mathbf{b}$ is defined as:

$$Cov\left[Z(\mathbf{a}), Z(\mathbf{b})\right] = \sigma_Z^2 R_Z(\mathbf{a}, \mathbf{b}) \tag{13}$$

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

$$R_Z(\mathbf{a}, \mathbf{b}) = \exp\left[-\sum_{i=1}^{d} \theta_i (a_i - b_i)^2\right] \tag{14}$$

where d represents the dimensionality of the problem, and θ_i is a scale parameter that indicates the correlation between the points within dimension i. A large θ_i indicates a low correlation.

The expected value $\hat{Y}()$ and variance $s^2()$ of the kriging model prediction at point $\hat{\mathbf{x}}$ are:

$$\hat{Y}(\hat{\mathbf{x}}) = \mathbf{h}(\hat{\mathbf{x}})^T \boldsymbol{\beta} + \mathbf{v}(\hat{\mathbf{x}})^T \mathbf{V}^{-1}(\mathbf{y} - \mathbf{F}\boldsymbol{\beta}) \tag{15}$$

$$s^2(\hat{\mathbf{x}}) = \sigma_Z^2 - \begin{bmatrix}\mathbf{h}(\hat{\mathbf{x}})^T & \mathbf{v}(\hat{\mathbf{x}})^T\end{bmatrix} \begin{bmatrix}\mathbf{0} & \mathbf{F}^T \\ \mathbf{F} & \mathbf{V}\end{bmatrix}^{-1} \begin{bmatrix}\mathbf{h}(\hat{\mathbf{x}}) \\ \mathbf{v}(\hat{\mathbf{x}})\end{bmatrix} \tag{16}$$

where $\mathbf{v}()$ is a vector containing the covariance between $\hat{\mathbf{x}}$ and each of the n training points (defined by Eqn. 13), $\mathbf{V}$ is an $n \times n$ matrix containing the covariance between each pair of training points with the variance of the noise in the true function evaluations σ_ϵ^2 added to the diagonal terms, $\mathbf{y}$ is the vector of response outputs at each of the training points, and $\mathbf{F}$ is an $n \times q$ matrix with rows $\mathbf{h}(\mathbf{x}_i)^T$ (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, σ_ϵ^2 and $\boldsymbol{\theta}$) have known values.

The parameters σ_Z^2, σ_ϵ^2, and $\boldsymbol{\theta}$ are determined through maximum likelihood estimation. Denote by $\mathbf{R}$ the correlation between the outputs at the training points. $\mathbf{R}$ is then equivalent to $\mathbf{V}/(\sigma_Z^2 + \sigma_\epsilon^2)$, and the ij^{th} component of $\mathbf{R}$ is:

$$R_{ij} = \begin{cases} 1 & \text{if } i = j \\ g R_Z(\mathbf{x}_i, \mathbf{x}_j) & \text{if } i \neq j \end{cases} \tag{17}$$

where $g = \sigma_Z^2/(\sigma_Z^2 + \sigma_\epsilon^2)$. The log likelihood can then be written as:[16]

$$\log\left[p(\mathbf{y}|\mathbf{R})\right] = -\frac{1}{n}\log|\mathbf{R}| - \log(\hat{\sigma}^2) \tag{18}$$

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

$$\hat{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \mathbf{F}\hat{\boldsymbol{\beta}})^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{F}\hat{\boldsymbol{\beta}}) \tag{19}$$

where $\hat{\boldsymbol{\beta}}$ is the generalized least squares estimate of $\boldsymbol{\beta}$ from:

$$\hat{\boldsymbol{\beta}} = \left[\mathbf{F}^T \mathbf{R}^{-1} \mathbf{F}\right]^{-1} \mathbf{F}^T \mathbf{R}^{-1} \mathbf{y} \tag{20}$$

Maximizing Eqn. 18 gives estimates of $\boldsymbol{\theta}$ and g; recognizing that $\hat{\sigma}^2 = \sigma_Z^2 + \sigma_\epsilon^2$, estimates of σ_Z^2 and σ_ϵ^2 can also be obtained.

B. Expected Improvement Function

The expected improvement function is used to select the location at which a new training point should be added. It does this by calculating the probability that any point in the design space will provide a better solution than the current best solution based on the predictions and uncertainties of the kriging 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 kriging predictor $Y_P()$ can be described by a normal distribution:

$$Y_P(\mathbf{x}) \sim N\left[\hat{Y}(\mathbf{x}), s(\mathbf{x})\right] \tag{21}$$

where the expected value $\hat{Y}()$ and the variance $s^2()$ were defined in Eqns. 15 and 16, respectively. The EIF used in SKO is then defined as:[12]

$$EI(\mathbf{x}) \equiv E\left[\max\left(\hat{Y}(\mathbf{x}^*) - Y_P(\mathbf{x}), 0\right)\right]\left(1 - \frac{\sigma_\epsilon}{\sqrt{s^2(\mathbf{x}) + \sigma_\epsilon^2}}\right) \tag{22}$$

where $\mathbf{x}^*$ is the current effective best solution and is defined below. The expectation term in Eqn. 22 can then be computed by integrating over the distribution of $Y_P(\mathbf{x})$ with $\hat{Y}(\mathbf{x}^*)$ held constant, and can be expressed analytically as:[12,13]

$$E\left[\max\left(\hat{Y}(\mathbf{x}^*) - Y_P(\mathbf{x}), 0\right)\right] = \left[\hat{Y}(\mathbf{x}^*) - \hat{Y}(\mathbf{x})\right]\Phi\left[\frac{\hat{Y}(\mathbf{x}^*) - \hat{Y}(\mathbf{x})}{s(\mathbf{x})}\right] + s(\mathbf{x})\phi\left[\frac{\hat{Y}(\mathbf{x}^*) - \hat{Y}(\mathbf{x})}{s(\mathbf{x})}\right] \tag{23}$$

The effective best solution $\mathbf{x}^*$ is determined through a utility function $\upsilon()$ to account for the uncertainty in the evaluation of the true function (due to the noise ϵ). This function is defined as:[12]

$$\upsilon(\mathbf{x}) \equiv -\hat{Y}(\mathbf{x}) - \eta s(\mathbf{x}) \tag{24}$$

where η is a tunable constant that indicates the desired level of risk aversion. For instance, if $\eta = 1$, it implies a willingness to trade 1 unit of the predicted value for 1 unit of the standard deviation of prediction uncertainty.[12] Because unobserved points in the design space will likely have greater uncertainty than the observed training points, the utility function is only maximized over the n training points for the sake of computational efficiency:[12]

$$\mathbf{x}^* = \max\left[\upsilon(\mathbf{x}_1), \upsilon(\mathbf{x}_2), \ldots, \upsilon(\mathbf{x}_{n-1}), \upsilon(\mathbf{x}_n)\right] \tag{25}$$

Note that if the function being modeled is deterministic ($\sigma_\epsilon = 0$), then Eqn. 22 collapses to Eqn. 23, and the training points are directly interpolated by the kriging model (therefore $s(\mathbf{x}) = 0$ at the training points), so the utility function becomes simply $\upsilon(\mathbf{x}) = -\hat{Y}(\mathbf{x}) = -Y(\mathbf{x})$.

The point at which the EIF is maximized is selected as an additional training point. In Ref. 12 this maximization is performed using a Nelder-Mead simplex approach, which is a local optimization method. Because the EIF is quite often multimodal (particularly in the early stages of the SKO process) it is expected that Nelder-Mead may occasionally fail to converge to the true global optimum, thus global optimization techniques for finding the point that maximizes the EIF will be investigated. With the new training point added, a new kriging 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.

It is important to understand how the use of this EIF leads to optimal solutions. This particular EIF is tailored to the problem of bound-constrained minimization. Eqn. 23 indicates how much the objective function value at $\mathbf{x}$ is expected to be less than the predicted value at the current effective best solution. It contains a balance between exploitation of regions of the design space where good solutions have been discovered, and exploration of regions that have not been well explored and thus have greater uncertainty. Because the kriging model provides a random distribution at each predicted point, points with good predicted values and even a small variance will have a significant probability of producing a better solution (exploitation), but so will points that have relatively poor predicted values and greater variance (exploration). The problem of MPP search, however, is made more complicated due to its inclusion of equality constraints. For this problem, the EIF must be reformulated to give it knowledge of these constraints.

C. Incorporation of Equality Constraints for MPP Search

In this work, the equality constraints will be enforced through the use of a merit function. Possible merit functions include a penalty function P, a Lagrangian function L, and an augmented Lagrangian function La, the forms of which are:

$$P(\mathbf{x}, r_p) = f(\mathbf{x}) + r_p \mathbf{c}(\mathbf{x})^T \mathbf{c}(\mathbf{x}) \tag{26}$$

$$L(\mathbf{x}, \boldsymbol{\lambda}_c) = f(\mathbf{x}) + \boldsymbol{\lambda}_c^T \mathbf{c}(\mathbf{x}) \tag{27}$$

$$La(\mathbf{x}, \boldsymbol{\lambda}_c, r_p) = f(\mathbf{x}) + \boldsymbol{\lambda}_c^T \mathbf{c}(\mathbf{x}) + r_p \mathbf{c}(\mathbf{x})^T \mathbf{c}(\mathbf{x}) \tag{28}$$

where $f()$ is the objective function, $\mathbf{c}()$ is the vector of equality constraint functions (converted to a standard form such that for feasible point $\mathbf{c}(\mathbf{x}) = 0$), and $\boldsymbol{\lambda}_c$ is the vector of Lagrange multipliers for the equality constraints. The proper form of the penalty schedule r_p has not yet been fully investigated for this work, but one possibility may be:

$$r_p = \exp\left[\frac{k + \text{offset}}{10}\right] \tag{29}$$

where k is the current iteration number of the SKO process. Simple zeroth-order updates for the Lagrange multipliers are used:

$$\boldsymbol{\lambda}_c^{k+1} = \boldsymbol{\lambda}_c^k + 2r_p \mathbf{c}(\mathbf{x}) \tag{30}$$

The Lagrange multipliers are applied whenever a new iterate is accepted, whereas the penalty schedule is updated for any iteration that fails to reduce the

IV. Computational Experiments

Preliminary work has investigated the use of the penalty function in Eqn. 26. The application of this method to the PMA case is fairly straightforward because the objective function involves the minimization of the response function that the kriging model represents, which is the initial intent of the SKO method. Furthermore, the equality constraint for PMA is only applied to the location of the optimal solution, which is deterministic. In this case, the constraint function can be written as $c(\mathbf{u}) = \|\mathbf{u}\| - \bar{\beta}$, which is added to the expected value of the kriging model at $\mathbf{u}$, to produce the penalty function:

$$P(\mathbf{u}) = \hat{Y}(\mathbf{u}) + r_p \left(\|\mathbf{u}\| - \bar{\beta}\right)^2 \tag{31}$$

Application to the RIA case is less clear because the objective function involves the minimization of a deterministic function. Moreover, the constraint is stochastic because it involves the kriging model, so feasibility can only be determined in a probabilistic sense. The constraint function can be formulated as $c(\mathbf{u}) = \hat{Y}(\mathbf{u}) - \bar{z}$, and added to the deterministic function being minimized to produce the penalty function:

$$P(\mathbf{u}) = \|\mathbf{u}\| + r_p \left(\hat{Y}(\mathbf{u}) - \bar{z}\right)^2 \tag{32}$$

For both formulations, the resulting expected improvement and utility functions are as follows (note that these are written in terms of $\mathbf{u}$ because the MPP search is performed in u-space):

$$EI(\mathbf{u}) = \left\{ [P(\mathbf{u}^*) - P(\mathbf{u})]\, \Phi\left[\frac{P(\mathbf{u}^*) - P(\mathbf{u})}{s(\mathbf{u})}\right] + s(\mathbf{u})\phi\left[\frac{P(\mathbf{u}^*) - P(\mathbf{u})}{s(\mathbf{u})}\right] \right\} \left(1 - \frac{\sigma_\epsilon}{\sqrt{s^2(\mathbf{u}) + \sigma_\epsilon^2}}\right) \tag{33}$$

$$\upsilon(\mathbf{u}) = -P(\mathbf{u}) - \eta s(\mathbf{u}) \tag{34}$$

Preliminary results with a static r_p have shown that the penalty function provides poor scaling between the objective function and the constraint violation; one or the other simply overwhelms the EIF. The Lagrangian methods should help provide the scaling needed.

Another idea is to add the penalty (for either the RIA or PMA formulation) to the true function evaluations that are used to build the kriging model. The EIF could then be used in the form of Section III.B with no modifications. The advantage to this is that the entire kriging model is built with some knowledge of the constraints, whereas in Eqns. 31-34 only the expected value of the kriging model is modified by the penalty, while the variance remains the same. The disadvantage to this method is that the kriging model that results as a byproduct of the optimization will not be a model of the true function, but rather of the penalized function. However, after the MPPs are found, the final set of training points could be evaluated without the penalty and used to build a kriging model that could then be used in the probability integration. This adds slightly to the expense, but this method may provide a better handling of the constraints.

The final paper will include investigations of PMA and RIA methods, and formulations of both that will allow for the discovery of multiple MPPs.

V. Conclusions

The prevalence of engineering problems defined by expensive, nonlinear response functions has necessitated the development of reliability assessment methods that are both efficient and accurate. To that end, this paper has presented an idea based on the application of sequential kriging optimization to MPP search. Because this is a global optimization method, it is capable of locating multiple MPPs, which allows for more accurate probability integration. In order to apply SKO to MPP search, a method of properly incorporating the equality constraints must be developed. One possibility is through the use of a merit function, several forms of which are being investigated. The formulation of a reliability assessment tool based on sequential kriging optimization is still in its early stages, but initial results have shown promise.

VI. Acknowledgments

The authors would like to express their thanks to the Sandia Computer Science Research Institute (CSRI) for support of this collaborative work between Sandia National Laboratories and Vanderbilt University.

References

[1]Box, G.E.P. and Cox, D.R., An Analysis of Transformations, *J. Royal Stat. Soc.*, Series B, Vol. 26, 1964, pp. 211-252.

[2]Breitung, K., Asymptotic approximation for Multinormal Integrals, *J. Eng. Mech., ASCE*, Vol. 110, No. 3, 1984, pp. 357-366.

[3]Chen, X. and Lind, N.C., Fast Probability Integration by Three-Parameter Normal Tail Approximation, *Struct. Saf.*, Vol. 1, 1983, pp. 269-276.

[4]Cressie, N.A.C., *Statistics for Spatial Data*, revised edition, 1993 (Wiley: New York).

[5]Der Kiureghian, A. and Liu, P.L., Structural Reliability Under Incomplete Probability Information, *J. Eng. Mech., ASCE*, Vol. 112, No. 1, 1986, pp. 85-104.

[6]Dey, A. and Mahadevan, S., Ductile Structural System Reliability Analysis using Adaptive Importance Sampling, *Struct. Saf.*, Vol. 20, 1998, pp. 137-154.

[7]Eldred, M.S. and Bichon, B.J., New Second-Order Formulations for Reliability Analysis and Design, *AIAA J.*, in preparation.

[8]Eldred, M.S., Agarwal, H., Perez, V.M., Wojtkiewicz, S.F., Jr., and Renaud, J.E., Investigation of Reliability Method Formulations in DAKOTA/UQ, (to appear) *Structure & Infrastructure Engineering: Maintenance, Management, Life-Cycle Design & Performance*, Taylor & Francis Group.

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

[10]Hohenbichler, M. and Rackwitz, R., Improvement of Second-Order Reliability Estimates by Importance Sampling, *J. Eng. Mech., ASCE*, Vol. 114, No. 12, 1988, pp. 2195-2199.

[11]Hong, H.P., Simple Approximations for Improving Second-Order Reliability Estimates, *J. Eng. Mech., ASCE*, Vol. 125, No. 5, 1999, pp. 592-595.

[12]Huang, D., Allen, T.T., Notz, W.I., and Zeng, N., Global Optimization of Stochastic Black-Box Systems via Sequential Kriging Meta-Models, *J. Global Opt.*, Vol. 34, 2006, pp. 441-466.

[13]Jones, D., Shonlau, M., and Welch, W., Efficient Global Optimization of Expensive Black-Box Functions, *INFORMS J. of Comp.*, Vol. 12, 1998, pp. 272.283

[14]Rackwitz, R. and Fiessler, B., Structural Reliability under Combined Random Load Sequences, *Comput. Struct.*, Vol. 9, 1978, pp. 489-494.

[15]Rosenblatt, M., Remarks on a Multivariate Transformation, *Ann. Math. Stat.*, Vol. 23, No. 3, 1952, pp. 470-472.

[16]Sacks, J., Schiller, S.B., and Welch, W., Design for computer experiments, *Technometrics*, Vol. 31, 1989, pp. 41-47.

[17]Tu, J., Choi, K.K., and Park, Y.H., A New Study on Reliability-Based Design Optimization, *J. Mech. Design*, Vol. 121, 1999, pp. 557-564.

[18]Wu, Y.-T. and Wirsching, P.H., A New Algorithm for Structural Reliability Estimation, *J. Eng. Mech., ASCE*, Vol. 113, 1987, pp. 1319-1336.

[19]Zou, T., Mourelatos, Z., Mahadevan, S., and Tu, J., Reliability Analysis of Automotive Body-Door Subsystem, *Rel. Eng. and Sys. Saf.*, Vol. 78, 2002, pp. 315-324.