50th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference
17th
4 - 7 May 2009, Palm Springs, California

AIAA 2009-2261

Reliability-Based Design Optimization using Efficient Global Reliability Analysis

Barron J. Bichon[††‖]*, Sankaran Mahadevan[‖†], and Michael S. Eldred[**‡]

[††] *Southwest Research Institute, San Antonio, TX 78238*

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

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

Finding the optimal (lightest, least expensive, etc.) design for an engineered component that meets or exceeds a specified level of reliability is a problem of obvious interest across a wide spectrum of engineering fields. Various methods for this reliability-based design optimization problem have been proposed. Unfortunately, this problem is rarely solved in practice because, regardless of the method used, solving the problem is too expensive or the final solution is too inaccurate to ensure that the reliability constraint is actually satisfied. This is especially true for engineering applications involving expensive, implicit, and possibly nonlinear performance functions (such as large finite element models). The Efficient Global Reliability Analysis method was recently introduced to improve both the accuracy and efficiency of reliability analysis for this type of performance function. This paper explores how this new reliability analysis method can be used in a design optimization context to create a method of sufficient accuracy and efficiency to enable the use of reliability-based design optimization as a practical design tool.

I. Introduction

Reliability-based design optimization (RBDO) is used to perform design optimization (such as minimizing the weight of a component) while accounting for reliability constraints. A general RBDO problem is typically of the form:

$$\begin{aligned} &\text{minimize} && f(\mathbf{d}) \\ &\text{subject to} && P[g(\mathbf{d},\mathbf{x}) \geq \bar{z}] \leq \bar{p}_f \end{aligned} \tag{1}$$

where the objective function f is a function of only the deterministic design variables $\mathbf{d}$, but the response function in the reliability constraint g is a function of $\mathbf{d}$ and $\mathbf{x}$, a vector of random variables defined by known probability distributions. The probability of failure p_f is defined as the probability that the response function $g(\mathbf{d},\mathbf{x})$ exceeds some threshold value $\bar{z}$ and is calculated as:

$$p_f = \int_{g>\bar{z}} \cdots \int f_{\mathbf{x}}(\mathbf{x})\, d\mathbf{x} \tag{2}$$

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,

*Research Engineer, Materials Engineering Department; Doctoral Candidate and NSF IGERT Fellow, Department of Civil and Environmental Engineering; Member AIAA.

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

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

§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.

evaluating the multiple integral is impractical.[18] Because of these complications, methods of approximating this integral are used in practice.

Common approximations include sampling methods and methods based on the concept of the most probable point (MPP). MPP-based methods, such as the First-Order Reliability Method (FORM), require solving an equality-constrained optimization problem to locate the MPP, which may fail. They then rely on a low-order approximation to the shape of the limit state (the contour on the response function where $g = \bar{z}$) to calculate p_f and will therefore be inaccurate if this approximation is poor. Sampling methods, such as Monte Carlo sampling, do not rely on an approximation to the shape of the limit state and are therefore generally more accurate, but they are usually too computationally expensive to be practical. The authors previously introduced Efficient Global Reliability Analysis (EGRA) to provide a method that is both computationally efficient and accurate for any arbitrarily shaped limit state.[3] This paper investigates how EGRA might be used in the various formulations of the RBDO problem.

A. Nested RBDO

The simplest and most direct RBDO approach is the nested approach in which a full reliability analysis is performed for every optimization function evaluation. This involves a nesting of two distinct levels of optimization within each other, one at the design level and one at the reliability analysis level.

An important performance enhancement for nested methods when MPP-based reliability analysis methods are used is the use of sensitivity analysis to analytically compute the design gradients of the probability constraint. When design variables are separate from the uncertain variables (i.e., they are not distribution parameters), then the following first-order expression may be used:[1,19,21]

$$\nabla_{\mathbf{d}} p_f = -\phi(-\beta)\frac{1}{\|\nabla_{\mathbf{u}} G\|}\nabla_{\mathbf{d}} g \tag{3}$$

where β is the reliability index (the distance from the median response to the MPP), ϕ is the standard normal probability density function, $\mathbf{u}$ are the random variables transformed into standard normal space (where the MPP search is performed), and $G(\mathbf{u}) \equiv g(\mathbf{x})$ by definition. The authors provide similar sensitivities for second-order reliability analyses in Ref. 13.

The ability to provide this kind of sensitivity information is an advantage of MPP-based methods, but because of the possible inaccuracy in this type of reliability analysis, they can lead to optimal solutions that do not satisfy the constraint on the probability of failure. Approximate sensitivities can be derived for Monte Carlo Sampling, but in a nested formulation, the probability of failure must be calculated a large number of times, making this type of reliability analysis rarely feasible due to the computational expense involved.

B. Single-Loop RBDO

This formulation of the RBDO problem simultaneously optimizes the objective function and searches for the MPP, satisfying the probabilistic constraint only at the optimal solution. The Karush-Kuhn-Tucker (KKT) conditions are used to reformulate the first-order reliability constraint into an equivalent deterministic constraint. In this way, the need to locate the MPP for the constraints is completely eliminated, but at convergence, the MPPs of the active constraints are still found. The new optimization problem is stated as:

$$\begin{array}{ll} \text{minimize} & f(\mathbf{d}) \\ \text{subject to} & \beta \geq \bar{\beta} \\ & G(\mathbf{d}, \mathbf{u}) = 0 \\ & \nabla_{\mathbf{u}} G^T \mathbf{u} + \beta \|\nabla_{\mathbf{u}} G\| = 0 \end{array} \tag{4}$$

The first constraint states that the final reliability index must be greater than a minimum reliability index $\bar{\beta}$, which is derived from the original probability constraint assuming a first order limit state, $\bar{\beta} = -\Phi^{-1}(\bar{p}_f)$, where Φ is the standard normal cumulative distribution function. The second constraint ensures that the MPP lies on the limit state. The third constraint is the KKT optimality constraint, the derivation of which also relies on the first-order limit state assumption. Note that the equations shown assume that $\bar{z} = 0$, but the response function can always be rearranged to facilitate this assumption.

Investigations into this formulation of the RBDO problem in Ref. 8, 22, 23 have shown that it is far more efficient than the nested formulation, but the accuracy of the final solution relies on the accuracy of the underlying first-order assumption.

C. Sequential RBDO

An alternative RBDO approach is the sequential approach, in which additional efficiency is sought through breaking the nested relationship of the MPP and design searches. The general concept is to iterate between optimization and uncertainty quantification, updating the optimization goals based on the most recent probabilistic assessment results. This update may be based on safety factors[35] or other approximations.[12]

A particularly effective approach for updating the optimization goals is to use the p_f sensitivity analysis in combination with local surrogate models.[37] In Ref. 14, first-order Taylor series approximations were explored, and in Ref. 13, second-order Taylor series approximations were investigated. In both cases, a trust-region model management framework[17] is used to adaptively manage the extent of the approximations and ensure convergence of the RBDO process. Surrogate models can be used for both the objective function and the constraints, although the use of constraint surrogates alone is sufficient to remove the nesting.

In particular, trust-region surrogate-based RBDO employs surrogate models of f and p_f within a trust region Δ^k centered at $\mathbf{d}_c$:

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{d}_c) + \nabla_{\mathbf{d}} f(\mathbf{d}_c)^T(\mathbf{d} - \mathbf{d}_c) \\ \text{subject to} \quad & p_f(\mathbf{d}_c) + \nabla_{\mathbf{d}} p_f(\mathbf{d}_c)^T(\mathbf{d} - \mathbf{d}_c) \\ & \|\mathbf{d} - \mathbf{d}_c\|_\infty \leq \Delta^k \end{aligned} \tag{5}$$

where first-order Taylor series are shown, but as mentioned, second-order expansions could also be used.

Sequential RBDO is more efficient than the nested formulation, but generally more expensive than the single-loop formulation. It can be more accurate than single-loop results due to the possible inclusion of second-order approximations to the limit state, but will still be inaccurate if that approximation is poor.

II. Efficient Global Optimization and Reliability Analysis

Efficient global optimization (EGO) was originally proposed by Jones et al.[20] to solve unconstrained optimization problems. The main features of this method are its use of Gaussian Process models and the Expected Improvement Function (EIF). 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 this method is:

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 modifications to the EGO method to make it applicable to constrained optimization problems and 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 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 response function being modeled $G(\mathbf{u})$ can be described by:[9]

$$G(\mathbf{u}) = \mathbf{h}(\mathbf{u})^T\boldsymbol{\beta} + Z(\mathbf{u}) \tag{6}$$

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 below) 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.[29] 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:

$$Cov\left[Z(\mathbf{a}), Z(\mathbf{b})\right] = \sigma_Z^2 R(\mathbf{a}, \mathbf{b}) \tag{7}$$

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,[29] and is used here for $R()$:

$$R(\mathbf{a}, \mathbf{b}) = \exp\left[-\sum_{i=1}^{d}\theta_i(a_i - b_i)^2\right] \tag{8}$$

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:

$$\mu_G(\mathbf{u}) = \mathbf{h}(\mathbf{u})^T\boldsymbol{\beta} + \mathbf{r}(\mathbf{u})^T\mathbf{R}^{-1}(\mathbf{g} - \mathbf{F}\boldsymbol{\beta}) \tag{9}$$

$$\sigma_G^2(\mathbf{u}) = \sigma_Z^2 - \begin{bmatrix}\mathbf{h}(\mathbf{u})^T & \mathbf{r}(\mathbf{u})^T\end{bmatrix}\begin{bmatrix}\mathbf{0} & \mathbf{F}^T\\ \mathbf{F} & \mathbf{R}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{h}(\mathbf{u})\\ \mathbf{r}(\mathbf{u})\end{bmatrix} \tag{10}$$

where $\mathbf{r}(\mathbf{u})$ is a vector containing the covariance between $\mathbf{u}$ and each of the n training points (defined by Eq. 7), $\mathbf{R}$ 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:[29]

$$\log\left[p(\mathbf{g}|\mathbf{R})\right] = -\frac{1}{n}\log|\mathbf{R}| - \log(\hat{\sigma}_Z^2) \tag{11}$$

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:

$$\hat{\sigma}_Z^2 = \frac{1}{n}(\mathbf{g} - \mathbf{F}\boldsymbol{\beta})^T\mathbf{R}^{-1}(\mathbf{g} - \mathbf{F}\boldsymbol{\beta}) \tag{12}$$

Maximizing Eq. 11 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:

$$\hat{G}(\mathbf{u}) \sim N\left[\mu_G(\mathbf{u}), \sigma_G(\mathbf{u})\right] \tag{13}$$

where the mean $\mu_G()$ and the variance $\sigma_G^2()$ were defined in Eqs. 9 and 10, respectively. The EIF is defined as:[20]

$$EI\big(\hat{G}(\mathbf{u})\big) \equiv E\left[\max\Big(G(\mathbf{u}^*) - \hat{G}(\mathbf{u}), 0\Big)\right] \tag{14}$$

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:

$$EI\big(\hat{G}(\mathbf{u})\big) = \int_{-\infty}^{G^*} (G^* - G)\; f_{\hat{G}}\; dG \tag{15}$$

where G is a realization of $\hat{G}$. This integral can be expressed analytically as:[20]

$$EI\big(\hat{G}(\mathbf{u})\big) = (G^* - \mu_G)\; \Phi\left(\frac{G^* - \mu_G}{\sigma_G}\right) + \sigma_G\, \phi\left(\frac{G^* - \mu_G}{\sigma_G}\right) \tag{16}$$

where Φ and ϕ are the standard normal cumulative distribution and probability density functions, respectively, and 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. Because the EIF is often highly multimodal[20] an implementation of the DIRECT global optimization algorithm is used[16] to locate the point at which the EIF is maximized.

It is important to understand how the use of this EIF leads to optimal solutions. Eq. 16 indicates how much the objective function value at $\mathbf{x}$ 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).

C. Constrained Formulation of EGO

EGO was originally created as an unconstrained minimizer. This paper uses a new formulation to enable the solution of constrained problems using EGO, but alternative methods have been previously introduced in Refs. 2, 4, 30, 31. A brief introduction to this constrained EGO formulation is presented here, but a detailed comparison to previous methods is outside the scope of this paper and will be published separately.

Methods based on merit functions are commonly used to solve constrained optimization problems. This type of method adds penalty terms to the objective function based on the constraint violations, thus making them less desirable to the optimizer. One very successful type of merit function is the augmented Lagrangian formulation. Using this formulation, the merit function for the RBDO problem in Eq. 1 could be written as:

$$\mathcal{M} = f + \lambda\,[p_f - \bar{p}_f] + r_p\,[p_f - \bar{p}_f]^2 \tag{17}$$

where λ is a Lagrange multiplier and r_p is the penalty coefficient. Note that this equation assumes the presence of only one constraint, but λ, p_f, and $\bar{g}$ could easily be replaced with vectors to form a more general formulation. For use with EGO, Gaussian process models stand in for f and p_f:

$$\mathcal{M} = \hat{f} + \lambda\,[\hat{p}_f - \bar{p}_f] + r_p\,[\hat{p}_f - \bar{p}_f]^2 \tag{18}$$

To drive the EGO algorithm, the expected improvement for this merit function is needed, but its calculation is not straightforward.

Recall that at any point, the terms $\hat{f}$ and $\hat{p}_f$ do not represent values, but Gaussian distributions. A linear combination of Gaussian distributions results in another Gaussian distribution, but the square term in Eq. 18 means that the resulting distribution of $\mathcal{M}$ is not Gaussian. To accomodate the use of EGO in an augmented Lagrangian formulation, the Expected Violation Function (EVF) is used.

1. Expected Violation Function

The EVF was originally introduced in Ref. 2, but was used differently in that work. Here it is used to replace the constraint violation terms $(\hat{p}_f - \bar{p}_f)$ in Eq. 18 with the *expected violation*. For an inequality constraint with an upper bound, the expected violation is calculated in a similar fashion as the previously derived expected improvement function:

$$EV = (\mu_p - \bar{p}_f)\left[1 - \Phi\left(\frac{\bar{p}_f - \mu_p}{\sigma_p}\right)\right] + \sigma_p\,\phi\left(\frac{\bar{p}_f - \mu_p}{\sigma_p}\right) \tag{19}$$

Similar equations can be derived for inequality constraints with lower bounds and also for equality constraints, but they are not of interest here. Using the expected violation, the augmented Lagrangian merit function becomes:

$$\mathcal{M} = \hat{f} + \lambda EV + r_p EV^2 \tag{20}$$

Because this formulation adds deterministic values to the Gaussian distribution $\hat{f}$, the merit function is now truly Gaussian with parameters defined by:

$$\mu_{\mathcal{M}} = \mu_f + \lambda EV + r_p EV^2 \tag{21}$$

$$\sigma^2_{\mathcal{M}} = \sigma^2_f \tag{22}$$

These values can then be used in Eq. 16 to calculate the expected improvement on the merit function.

When EGO converges, the GP models of the objective function and any constraints are theoretically accurate in the vicinity of the optimum solution due to the sample density in this region. Recognizing this, a gradient-based optimizer can be used to refine the EGO solution using these GPs rather than the true functions so that no additional true function evaluations are necessary. In this study, the Nonlinear Interior Point Solver (NIPS) in OPT++,[24] an optimization package from Sandia National Laboratories, is used to perform this solution refinement.

D. Efficient Global Reliability Analysis

The goal of EGRA is to create an accurate Gaussian process model of the response function g using only a very small number of samples. This inexpensive surrogate model can then be sampled to give a highly accurate estimate on the probability of failure at minimal cost. The computational expense is kept low by focusing the training samples in the areas where accuracy is important - near the limit state. To determine whether a given point is expected to lie on the limit state, the Expected Feasibility Function has been introduced by the authors.[3]

1. Expected Feasibility Function

The expected violation function provides an indication of how much the true value of the response at a point can be expected to violate the given constraint. In EGRA, the concern is on how much the true value of the response at a point be expected to be equal to the value that satisfies the equality constraint $G(\mathbf{u}) = \bar{z}$. Inspired by the contour estimation work in Ref. 27, this expectation can be calculated in a similar fashion as Eq. 15 by integrating over a region in the immediate vicinity of the threshold value $\bar{z} \pm \epsilon$:

$$EF\big(\hat{G}(\mathbf{u})\big) = \int_{\bar{z}-\epsilon}^{\bar{z}+\epsilon} \big[\epsilon - |\bar{z} - G|\big]\, f_{\hat{G}}\, dG \tag{23}$$

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:

$$\begin{aligned} EF\big(\hat{G}(\mathbf{u})\big) = (\mu_G - \bar{z})&\left[2\,\Phi\left(\frac{\bar{z} - \mu_G}{\sigma_G}\right) - \Phi\left(\frac{z^- - \mu_G}{\sigma_G}\right) - \Phi\left(\frac{z^+ - \mu_G}{\sigma_G}\right)\right] \\ - \sigma_G&\left[2\,\phi\left(\frac{\bar{z} - \mu_G}{\sigma_G}\right) - \phi\left(\frac{z^- - \mu_G}{\sigma_G}\right) - \phi\left(\frac{z^+ - \mu_G}{\sigma_G}\right)\right] \\ + \quad \epsilon&\left[\Phi\left(\frac{z^+ - \mu_G}{\sigma_G}\right) - \Phi\left(\frac{z^- - \mu_G}{\sigma_G}\right)\right] \end{aligned} \tag{24}$$

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 $\mathbf{u}$, while $\bar{z}$ 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.

2. 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 $\frac{(n+1)(n+2)}{2}$ 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 vicnity of the limit state.
7. This surrogate model is then used to calculate the probability of failure using 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.

III. Reliability-Based Design Optimization with EGO and EGRA

A variety of ways to incorporate EGO and EGRA into each of the traditional RBDO formulations (nested, single-loop, and sequential) are investigated.

A. Nested RBDO

In this method, EGO is used as the optimizer and for each candidate design point a full reliability analysis is performed using EGRA. For each EGRA analysis, the design point $\mathbf{d}$ is set; EGRA operates only over the random variables, $\mathbf{u}$. Because this must be performed at every design point and there is no sharing of information to either guide the optimizer or assist subsequent EGRA runs, this is expected to be the most expensive option.

B. Quasi-Single-Loop RBDO

In this method, EGRA is first used to locate the limit state and build a GP of g over the entire $\mathbf{d}+\mathbf{u}$ space in one step. With this GP built, any optimizer could be used at the design level; to calculate the probability of failure corresponding to any design point, one needs only to sample the GP previously built by EGRA. This optimization requires no additional evaluations of the response function in the probabilistic constraints. Because EGRA is allowed to search over the full expanse of $\mathbf{d}+\mathbf{u}$ at all times instead of being restricted to adding points within a given $\mathbf{d}$ hyperplane, this method is expected to be more efficient than the nested method.

At convergence, a full EGRA analysis is run to verify the reliability of the design. The function evaluations used in this step are not counted in the method's cost, but the result is used in determining whether the method converged to a feasible solution.

C. Sequential RBDO

This formulation uses a GP of the response function in the probabilistic constraint that spans $\mathbf{d}+\mathbf{u}$, but rather than update that model for every iteration of the design optimizer, it is only updated at its convergence. The algorithm follows this outline:

1. Build an initial GP for the objective function. At each design point, use EGRA to solve for the p_f at that point using a GP for g that spans the design and uncertain spaces.
2. Use EGO to fully solve the RBDO problem. At each design point, to calculate the corresponding p_f, sample the current GP of g.
3. When EGO converges, calculate the true p_f using EGRA at $\mathbf{d}^*$. The points used by EGRA to resolve the limit state are then added to the GP of g.
4. Re-solve the RBDO problem using this new GP. Iterate until the method converges.

This method has one major advantage over the quasi-single-loop method: because it only adds points to the GP at candidate optimal points, it does not waste time increasing the accuracy of the GP in regions of the design space that are far from optimal. Assessing convergence of this method is difficult. Due to the stochastic nature of EGO, the "converged to the same point" technique typically used for sequential problems cannot be applied. Instead, this study uses a metric on the accuracy of the underlying constraint function GP models. At the optimal point from an EGO solve, if the value of all of the constraint functions' GP models are within 1% of the actual constraint function values after verification, the method is said to have converged. However, while the latest analysis may have been necessary to ensure this accuracy in the models, it is not necessarily the optimal solution found by the algorithm. All of the candidate optimal designs are post-processed to find the best solution.

IV. Computational Experiments

Three tests problems are investigated to compare the results of these various formulations of RBDO with EGRA to one another and to other RBDO algorithms. For each problem, two "baseline" tests were run using previously available state-of-the-art methods in DAKOTA/UQ,[34] which is the uncertainty quantification component of DAKOTA,[15] an open-source software framework for design and performance analysis of computational models on high performance computers developed at Sandia National Laboratories. The first test is a nested formulation using the Nonlinear Interior Point Solver (NIPS) from the OPT++[24] library for the design optimizer and the second-order iterated Advanced Mean Value method[13] (AMV^2+) to perform the reliability analysis. This nested formulation is used to find the baseline solution, but is expected to be relatively expensive. To provide a more fair comparison on the expense of the methods, a second more efficient baseline is run using a sequential, trust-region based method, again using NIPS and AMV^2+. These are both gradient-based methods; to simulate their use on implicit performance functions, numerical differentiation is used to compute the gradients and, where needed, Hessian information is generated using quasi-Hessians with Symmetric Rank 1 updating.

All other tests that use either EGO or EGRA (or both) are performed using NESSUS-PA, a probabilistic algorithm library currently under development at Southwest Research Institute. The NIPS method uses the same OPT++ implementation, but only the first-order AMV+ method is currently available in NESSUS-PA.

A. Short column

The first application problem involves the plastic analysis and design of a short column with rectangular cross section (width b and depth h) having uncertain material properties (yield stress Y) and subject to uncertain loads (bending moment M and axial force P).[13, 22] The response function, designed to represent the stress in the column at which the yield stress is exceeded, is defined as:

$$g = 1 - \frac{4M}{bh^2Y} - \frac{P^2}{b^2h^2Y^2} \tag{25}$$

The distributions for P, M, and Y are Normal($\mu = 500$, $\sigma = 100$), Normal($\mu = 2000$, $\sigma = 400$), and Lognormal($\mu = 5$, $\sigma = 0.5$), respectively, with a correlation coefficient of 0.5 between P and M (uncorrelated otherwise). Failure for this response function is defined by $g \leq 0$.

An objective function of cross-sectional area and a target probability of failure are used in the design problem:

$$\begin{array}{ll} \min & bh \\ \text{s.t.} & \beta^* \geq 2.5 \\ & 5.0 \leq b \leq 15.0 \\ & 15.0 \leq h \leq 25.0 \end{array} \tag{26}$$

Table 1 gives a summary of the results from all investigated methods. Because both EGO and EGRA are stochastic methods, average solutions are given with some additional information on the best solutions found for 10 runs of these methods.

Table 1. Results for the short column example.

Design/Reliability Methods	Avg. Obj. Fn (Best Feasible)	Avg. Violation (# violations)	Avg. Fn Evals (Best Feasible)
Nested NIPS/AMV^2+	216.7	0.0	4123
Sequential NIPS/AMV^2+	216.7	0.0	2434
Nested NIPS/EGRA	215.9 (–)	0.024 (10)	2179.6 (–)
Nested EGO/AMV+	219.1 (219.7)	0.017 (3)	283.0 (280)
Nested EGO/EGRA	216.3 (216.1)	0.015 (3)	327.8 (341)
Quasi-Single-Loop NIPS	216.4 (216.7)	0.027 (5)	140.3 (129)
Quasi-Single-Loop EGO	217.6 (217.3)	0.026 (5)	139.0 (133)
Sequential EGO/EGRA	217.8 (216.5)	0.0 (0)	146.1 (111)

The Nested NIPS/EGRA test failed to find a single feasible solution in 10 runs of the algorithm. This is likely due to the inherent problems with using a stochastic reliability analysis method with a gradient-based design optimizer. If the variance in the reliability estimates is on the order of (or worse, greater than) the real change in reliability given a small change in design (for calculating the gradient via finite differences), it will greatly hinder the optimizer's progress. This leads to both an expensive analysis due to the greater number of steps needed, and poor results.

The remaining Nested tests performed reasonably well. They used less than 15% of the function evaluations required by the Sequential NIPS/AMV^2+ baseline test. The optimal objective function values are competitive with the baselines, but 3 of the runs in each test produced an infeasible result. The Quasi-Single-Loop tests were even less expensive requiring an average of less than 6% of the function evaluations required by the Sequential NIPS/AMV^2+ baseline test, but half of the optimal solutions were infeasible.

The Sequential EGO/EGRA test is clearly the superior method amongst the new methods introduced in this study. It required, on average, just 6% of the function evaluations needed by the Sequential NIPS/AMV^2+ baseline test. The average objective function value is within 0.5% of the baseline tests' results. It produced the least expensive feasible solution amongst all tests at just 111 function evaluations, and this method produced no infeasible solutions in 10 runs of the algorithm.

B. Cantilever Beam

This test problem investigates is the cantilever beam problem that has been previously investigated in Ref. 13, 14, 25, 32, 35 and others to test both reliability analysis and design methods. Figure 1 shows a schematic of this problem.

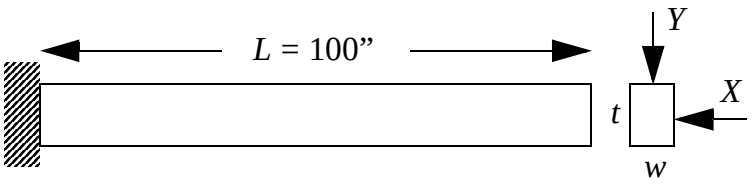

Figure 1. Schematic of the cantilever beam example problem.

This problem has two response functions of interest; one concerning the stress, g_S and one concerning the displacement, g_D:

$$g_S = R - \frac{600}{wt}\left(\frac{Y}{t} + \frac{X}{w}\right) \tag{27}$$

$$g_D = D - \frac{4L^3}{Ewt}\sqrt{\frac{Y^2}{t^4} + \frac{X^2}{w^4}} \tag{28}$$

Failure for these response functions are defined by $g_S \geq 0$ and $g_D \geq 0$. All variables are described in Table 2.

Table 2. Variable detail for the cantilever beam example.

Variable	Mean	COV	Distribution
Horizonal Load, X	500	0.2	Normal
Vertical Load, Y	1,000	0.1	Normal
Yield Strength, R	40,000	0.05	Normal
Modulus of Elasticity, E	29e6	0.05	Normal
Length, L	100	—	Deterministic
Max Displacement, D	2.25	—	Deterministic

The design problem is to minimize the weight (or, equivalently, the cross-sectional area) of the beam subject to probabilistic constraints on the displacement and stress. If the random variables are fixed at their means, the resulting deterministic design problem (with constraints $g_S \leq 0$ and $g_D \leq 0$) has the solution $(w,\, t) = (2.352,\, 3.326)$ with an objective function of 7.824. When seeking specified reliability levels of 3.0 for these constraints, the design problem becomes:

$$\begin{aligned} \min \quad & wt \\ \text{s.t.} \quad & \beta_S^* \geq 3.0 \\ & \beta_D^* \geq 3.0 \\ & 1.0 \leq w \leq 4.0 \\ & 1.0 \leq t \leq 4.0 \end{aligned} \tag{29}$$

Table 3 gives a summary of the results from all investigated methods. Because both EGO and EGRA are stochastic methods, average solutions are given with some additional information on the best solutions found for 10 runs of these methods.

The results for this test problem are similar to those seen in the previous tests. Note that constraint 1 refers to the stress constraint, and constraint 2 to the displacement. Once again, the sequential test produces extremely competitive results with an average solution within 1% of the Nested NIPS/AMV^2+ baseline result, an average cost of 8-15% of that used by the Sequential NIPS/AMV^2+ baseline (while producing an optimal objective function value that is 13% lower), and no infeasible solutions.

Table 3. Results for the cantilever beam example.

Design/Reliability Methods	Avg. Obj. Fn (Best Feas.)	Avg. Viol. 1 (# viols)	Avg. Viol. 2 (# viols)	Avg. Evals 1 (Best Feas.)	Avg. Evals 2 (Best Feas.)
Nested NIPS/AMV^2+	9.520	0.0	0.0	4176	4599
Sequential NIPS/AMV^2+	11.076	0.0	0.0	1611	1593
Nested EGO/AMV+	9.732 (9.526)	0.00004 (2)	0.0 (0)	110.5 (100)	356 (360)
Nested EGO/EGRA	9.772 (9.531)	0.0005 (1)	0.0 (0)	111.1 (102)	194.6 (169)
Quasi-Single-Loop NIPS	9.830 (9.701)	0.0427 (5)	0.0264 (1)	82.8 (108)	84.9 (68)
Quasi-Single-Loop EGO	9.836 (9.646)	0.0194 (5)	0.0291 (2)	85.1 (76)	127.9 (101)
Sequential EGO/EGRA	9.602 (9.524)	0.0 (0)	0.0 (0)	129.0 (134)	242.4 (272)

V. Conclusions

This paper explores the applicability of EGRA and EGO to the field of reliability-based design optimization. Several possibilities for applying these methods to RBDO were outlined and evaluated, the most promising of which is a sequential formulation. This method only performs a full EGRA analysis at the candidate optimal solution from a run of EGO on the current set of Gaussian process models. Additionally, the Gaussian process model of the response function in the probabilistic constraint spans both the design and uncertain spaces, thus allowing the information found at one design point to influence the search for the limit state at subsequent design points. These two characteristics, combined with the relatively low cost of performing an EGRA analysis, lead to a highly efficient RBDO method. This Sequential EGO/EGRA method was shown to produce near-optimal results at a small fraction of the cost of previously available methods.

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]Allen, M. and Maute, K., Reliability-Based Design Optimization of Aeroelastic Structures, *Struct. Multidiscip. Opt.*, Vol. 27, 2004, pp. 228-242.

[2]Audet, C., Dennis, J.E., Moore, D.W., Booker, A., and Frank, P.D., A Surrogate-Model-Based Method for Constrained Optimization, *Proc. 8th AIAA/NASA/USAF/ISSMO Symposium on Multidisciplinary Anal. and Opt.* paper AIAA-2000-4891, 2000.

[3]Bichon, B.J., Eldred, M.S., Swiler, L.P., Mahadevan, S., and McFarland, J.M., Efficient Global Reliability Analysis for Nonlinear Implicit Performance Functions, *AIAA J.*, Vol. 46, No. 10, 2008, pp. 2459-2468.

[4]Björkman, M., and Holström, K., Global Optimization of Costly Nonconvex Functions using Radial Basis Functions, *Opt. and Eng.*, Vol. 1, 2000, pp. 373-397.

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

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

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

[8]Chiralaksanakul, A., and Mahadevan, S., First-Order Approximation Methods in Reliability-Based Design Optimization, *J. of Mech. Design*, Vol. 127, 2005.

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

[10]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.

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

[12]Du, X. and Chen, W., Sequential Optimization and Reliability Assessment Method for Efficient Probabilistic Design, *J. Mech. Design*, Vol. 126, 2004, pp. 225-233.

[13]Eldred, M.S. and Bichon, B.J., Second-Order Reliability Formulations in DAKOTA/UQ, paper AIAA-2006-1828 in *Proc. 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, Newport, RI, May 2006.

[14]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.

[15]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. Version 4.0 User's Manual. Sandia Technical Report SAND2006-6337, Revised October 2006, Sandia National Laboratories, Albuquerque, NM.

[16]Gablonsky, J.M., An Implementation of the DIRECT Algorithm, Technical Report CRSC-TR98-29, Center for Research in Scientific Computation, North Carolina State University, August, 1998.

[17]Giunta, A.A. and Eldred, M.S., Implementation of a Trust Region Model Management Strategy in the DAKOTA Optimization Toolkit, *Proc. 8th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Anal. and Opt.*, paper AIAA-2000-4935, Long Beach, CA, September 6-8, 2000.

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

[19]Hohenbichler, M. and Rackwitz, R., Sensitivity and Importance Measures in Structural Reliability, *Civil Eng. Syst.*, Vol. 3, 1986, pp. 203-209.

[20]Jones, D., Schonlau, M., and Welch, W., Efficient Global Optimization of Expensive Black-Box Functions, *J. of Global Opt.*, Vol. 13, 1998, pp. 455-492.

[21]Karamchandani, A. and Cornell, C.A., Sensitivity Estimation Within First and Second Order Reliability Methods, *Struct. Saf.*, Vol. 11, 1992, pp. 95-107.

[22]Kuschel, N., and Rackwitz, R., Two Basic Problems in Reliability-Based Structural Optimization, *Math. Meth. of Oper. Research*, Vol. 46, 1997.

[23]Liang, J., Mourelatos, Z.P., and Tu, J., A Single-Loop Method for Reliability-Based Design Optimization, *Proc. 2004 ASME Design Eng. Tech. Conf.*, Paper DETC2004/DAC-57255.

[24]Meza, J.C., OPT++: An Object-Oriented Class Library for Nonlinear Optimization, Technical Report SAND94-8225, Sandia National Laboratories, Livermore, CA, 1994.

[25]Qu, X. and Haftka, R.T., "Reliability-Based Design Optimization Using Probabilistic Sufficiency Factor," paper AIAA-2003-1657 in *Proc. 44th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, Norfolk, VA, April 2003.

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

[27]Ranjan, P., Bingham, D., and Michailidis, G., Sequential Experiment Design for Contour Estimation from Complex Computer Codes, *Technometrics*, to appear, 2007.

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

[29]Sacks, J., Schiller, S.B., and Welch, W., Design for Computer Experiments, *Technometrics*, Vol. 31, 1989, pp. 41-47.

[30]Sasena, M.J., *Flexibility and Efficiency Enhancements for Constrained Global Design Optimization with Kriging Approximations*, PhD Thesis, Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI, 2002.

[31]Schonlau, M. *Computer Experiments and Global Optimization*, PhD Thesis, University of Waterloo, Waterloo, Canada, 1997.

[32]Sues, R., Aminpour, M. and Shin, Y., Reliability-Based Multidisciplinary Optimization for Aerospace Systems, paper AIAA-2001-1521 in *Proc. of the 42nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Seattle, WA, April 16-19, 2001.

[33]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.

[34]Wojtkiewicz, S.F., Jr., Eldred, M.S., Field, R.V., Jr., Urbina, A., and Red-Horse, J.R., A Toolkit For Uncertainty Quantification In Large Computational Engineering Models, *Proc. 42nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynam., and Mat. Conf.*, paper AIAA-2001-1455, Seattle, WA, April 16-19, 2001.

[35]Wu, Y.-T., Shin, Y., Sues, R., and Cesare, M., Safety-Factor Based Approach for Probability-Based Design Optimization, *Proc. 42nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynam., and Mat. Conf.*, paper AIAA-2001-1522, Seattle, WA, April 16-19, 2001.

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

[37]Zou, T., Mahadevan, S., and Rebba, R., Computational Efficiency in Reliability-Based Optimization, *Proc. 9th ASCE Specialty Conf. on Prob. Mech. and Struct. Rel.*, Albuquerque, NM, July 26-28, 2004.

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