A Fusion-Based Multi-Information Source Optimization Approach using Knowledge Gradient Policies

S. F. Ghoreishi * and D. L. Allaire†

Department of Mechanical Engineering, Texas A&M University, College Station, TX, 77843

Optimization of complex systems often involves evaluation of a quantity several times, which is potentially computationally prohibitive. This can be alleviated by considering information sources representing the original model with lower fidelity and cost. This paper describes an optimization method for the case where the objective function is represented by different information sources with varying fidelities and computational costs. The proposed methodology creates a multi-information source value-of-information framework that defines optimal strategies for querying of information sources. The surrogate Gaussian process model is used for fusion of information sources. Then, the knowledge gradient policy is incorporated by considering the probability of violation of constraints for sequential decision making to identify the next design and information source to evaluate. The high performance of the developed methodology is demonstrated in terms of making balance between cost and information gain of various information sources of a one-dimensional example test problem and an aerodynamic design example.

I. Introduction

In computational science and engineering, in order to optimize a complex system, quantities of interest need to be evaluated several times, which is potentially expensive. These quantities of interest can be often described by multiple computational models, called information sources, which can be used to evaluate the quantities of interest to alleviate the computational burden of optimization process. These information sources can be different experiments, numerical methods, or previous data. Each of these information sources has an associated level of accuracy to estimate the quantities of interest. This level of accuracy to represent the real world is referred to as fidelity which leads to different computational cost. This paper is concerned with the development of an approach for incorporating different available information sources, with potentially differing levels of fidelity and cost, to enable accurate and efficient constrained optimization of a real world quantity of interest.

One common approach in multifidelity optimization is to treat the models as a hierarchy and replace or calibrate low-fidelity information with high-fidelity results [1–5]. Trust region methods are used to approximate the high-fidelity model for optimizing an objective function when the derivatives of the objective are available. In Refs. 6 and 7, a trust region based model-management method is employed in which the gradients of the low-fidelity objective function and constraints are scaled or shifted to match those of a high-fidelity model. Ref. 8 presents a convergent multifidelity optimization algorithm using the trust region method which does not require the high-fidelity derivatives. Calibration techniques such as efficient global optimization [2] and the surrogate management framework [3] are often employed in cases where gradients are not available. In these cases, Kriging surrogate models, also referred to as Gaussian process regression [9], are often used as a method of interpolation, which is based on random spatial processes [10]. In Ref. 3, a framework is presented for generating a sequence of approximations to the objective function and using these approximations as surrogates for optimization without requiring the derivatives of the objective.

*Ph.D. Candidate, AIAA Student Member

†Assistant Professor, AIAA Member

In Ref. 11, a multifidelity sequential kriging optimization method is proposed which uses the expected improvement as a measure to determine the next design point and the level of fidelity model to evaluate. In Ref. 12, a method is proposed based on statistical techniques to adaptively build a multifidelity surrogate for multifidelity optimization. In Ref. 13, a multifidelity multidisciplinary design optimization approach is presented which synthesizes information from various models and manages model fidelity throughout the design process through the application of estimation theory and global sensitivity analysis.

In order to handle constraints in the optimization, surrogate-based methods have been proposed to approximate the constraints which are expensive to evaluate. In Ref. 14, the probability of feasibility is calculated using a Kriging surrogate approximating each constraint and multiplied by the expected improvement at the sampled point. In Ref. 5, efficient global optimization is used to perform optimization and the penalty method is used to consider constraints. Ref. 15 maximizes the expected improvement with samples constrained to lie in the feasible space defined based on the mean values of the Kriging models for the constraints. In many practical applications, the constraint function returns values only for the feasible space which makes the constraint function discontinuous, or only feasibility or infeasibility of a sample point is available instead of the value of the constraint at that sample, which makes the constraint function binary. In these cases, instead of predicting the constraint response values over the entire design space using surrogates, the goal is to predict whether or not the response is greater than a threshold. For considering these constraints, classification approaches have been used to construct a boundary separating the feasible and infeasible regions. The boundaries are constructed using methods such as convex hulls [16], support vector machines [17], etc. In Ref. 18, probabilistic support vector machine (PSVM) [19, 20] is used to quantify the probability of constraint violation, and is implemented within a constrained efficient global optimization formulation to select samples for locating the optimal solution.

In this work, we propose a method for optimizing a function which is expensive to evaluate, where a variety of information sources with different levels of computational cost and model fidelity are available to approximate the objective. Here, we relate the fidelity to uncertainty due to model inadequacy which is uncertainty due to the omission of some aspects of reality, improper modeling, or unrealistic assumptions [21–23]. This fidelity is characterized by assigning a probability distribution to the output of each individual model on the basis of the model inadequacy associated with that particular model. Surrogate models are constructed for information sources using Gaussian processes, and a single fused Gaussian process is constructed by fusing the information obtained from evaluating the design and the associated information source. There are several techniques used in practice for combining information from multiple information sources. Among them are the adjustment factors approach [24–26], Bayesian model averaging [27–31], and fusion under known and unknown correlation [32, 33]. Here, the fused Gaussian process is constructed based on the data and fidelity variances of the information sources which are queried. The next design to evaluate and the model to query are determined based on the cost of querying and the expected improvement criterion, which we characterize here using the knowledge gradient policy. The knowledge gradient policy was introduced in Ref. 34 as a look-ahead policy which takes an information-economic approach to maximize (or minimize) an objective using a single information source with noisy observations from that information source [35–38]. In this work, our contribution is incorporating the available information from different information sources to the fused model and using this fused model for decision making. Unlike the previous methods in which the expected improvement of each information source and the associated cost were used for decision making, in our proposed methodology, the expected improvement in the fused model by considering the fidelity variance and cost of each information source are used to select the next design and model. In other words, instead of performing the knowledge gradient for each individual model constructed from observed experiments, our proposed method performs the knowledge gradient policy for computation of expected improvement using the fused model constructed from the whole observed data. A key feature of the proposed method is the adaptive selection process which arises from consideration of both cost and information fusion. Our proposed multi-information source optimization approach is applied to optimize a one-dimensional example test problem and an aerodynamic design example, and the results are compared with the method presented in Ref. 12. It is shown that our proposed method identifies better solutions, especially in the case of limited budget available for querying.

The rest of the paper is organized as follows. Section II presents the approach proposed here. In Section III, the approach is applied to a one-dimensional test function and an aerodynamic example problem and the results are presented, and conclusions are drawn in Section IV.

II. Approach

In this section, our proposed approach is introduced for an optimization problem subject to m inequality constraints. A mathematical statement of the problem is formulated as:

$$\begin{aligned} \mathbf{x}^* = & \underset{\mathbf{x}\in\chi}{\operatorname{argmax}} f(\mathbf{x}) \\ & \text{s.t.} \quad g_j(\mathbf{x}) \leq 0, \qquad j = 1, \ldots, m. \end{aligned} \tag{1}$$

where f is the objective function, $\{g_j$, where $j = 1, 2, \ldots, m\}$ are the m real-world constraints, and $\mathbf{x}$ is a set of design variables in the input design space χ. In the optimization process, the real-world objective function and the real-world constraints must be estimated at each iteration. For this estimation task, often several information sources, such as numerical simulation models, experiments, expert opinions, etc., are available. These sources have varying fidelities, or approximation accuracies over the design space, and varying computational costs. The approach we describe here enables the exploitation of all available sources of information, in an information-economic sense, that balances the cost and associated fidelity of each information source when choosing how to estimate the quantity of interest and the constraints. In our formulation, the core issue is how to manage dynamic querying to choose what new information sources to sample and with what input, at each step of the overall decision process. The choice of next design point to query must be based on how much a sample tells us immediately about the design goal at hand and if that sample is expected to satisfy the current constraints. This choice must also trade off the costs associated with a particular information source query and the expected improvement or information gained by executing the query.

A. Fused Gaussian Process

Let assume we have M information sources available. Each information source describes the quantity of interest, $f(\mathbf{x})$, at design point $\mathbf{x}$ with the associated fidelity represented by additive independent identically distributed Gaussian noise as [39–41]:

$$y_i(\mathbf{x}) = f(\mathbf{x}) + \mathcal{N}(0, \lambda_i). \tag{2}$$

where $y_i(\mathbf{x})$ denotes the output distribution of i^{th} information source at design $\mathbf{x}$ and λ_i is the fidelity variance of the corresponding information source. Let $\mathbf{X}_i = [\mathbf{x}_{1,i}, \ldots, \mathbf{x}_{N_i,i}]$ be N_i evaluated design points with the observed objective values $\mathbf{y}_i = [y_{1,i}, \ldots, y_{N_i,i}]$ for the i^{th} information source. Using the Gaussian process for modeling the objective over the design space for this information source will lead to the following covariance function:

$$cov(\mathbf{y}_i) = K(\mathbf{X}_i, \mathbf{X}_i) + \lambda_i I, \tag{3}$$

where $K(\mathbf{X}_i, \mathbf{X}_i)$ is a $N_i \times N_i$ matrix specifying the covariance between the outputs which is written as a function of the inputs. We consider the commonly used squared exponential covariance function as:

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\sum_{h=1}^{d} \frac{(x_h - x'_h)^2}{2l_h^2}\right), \tag{4}$$

where d is the dimension of the input space, σ_f^2 is the signal variance, and l_h is the characteristic length-scale that indicates the correlation between the points within dimension h.

The first step in our approach is to construct a fused Gaussian process. We first assume that N previous queries of information sources are available that take the form of tuples $(i_{1:N}, \mathbf{x}_{1:N}, y_{1:N})$, in which $i_{1:N} \in [1, \ldots, M]$ represent the information sources that data belong to, $\mathbf{x}_{1:N}$ represent the design decisions corresponding to these information sources denoted by $\mathbf{X}_N$, and $y_{1:N}$ represent the observations obtained which are denoted by $\mathbf{y}_N$. Constructing a Gaussian process over data, we model the covariance between these data from various information sources using the following covariance function:

$$cov(\mathbf{y}_N) = K(\mathbf{X}_N, \mathbf{X}_N) + \Lambda, \tag{5}$$

where Λ is a diagonal matrix of size $N \times N$ in which each diagonal element is the fidelity variance associated with the information source that data belongs to. This matrix is given as:

$$\Lambda = \begin{bmatrix} \lambda_{i_1} & 0 & \cdots & 0 \\ 0 & \lambda_{i_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{i_N} \end{bmatrix}. \tag{6}$$

The joint distribution of the observed objective functions, $\mathbf{y}_N$, and unobserved function values, $\mathbf{y}_f$, at any set of P samples $\mathbf{x}_{1:P}$ denoted by $\mathbf{X}$ in the design space can be written as:

$$\begin{bmatrix} \mathbf{y}_N \\ \mathbf{y}_f \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(\mathbf{X}_N, \mathbf{X}_N) + \Lambda & K(\mathbf{X}_N, \mathbf{X}) \\ K(\mathbf{X}, \mathbf{X}_N) & K(\mathbf{X}, \mathbf{X}) \end{bmatrix}\right), \tag{7}$$

where $K(\mathbf{X}_N, \mathbf{X}_N)$ is a $N \times N$ matrix with mn^{th} entry as $k(\mathbf{x}_m, \mathbf{x}_n)$, and $K(\mathbf{X}_N, \mathbf{X})$ is a $N \times P$ matrix with mp^{th} entry as $k(\mathbf{x}_m, \mathbf{x}_p)$. Deriving the conditional distribution, we construct the fused Gaussian process with mean function μ and covariance matrix Σ as:

$$\mathbf{y}_f \sim GP(\mu, \Sigma), \tag{8}$$

where

$$\mu = K(\mathbf{X}_N, \mathbf{X})^T [K(\mathbf{X}_N, \mathbf{X}_N) + \Lambda]^{-1} \mathbf{y}_N, \tag{9}$$

$$\Sigma = K(\mathbf{X}, \mathbf{X}) - K(\mathbf{X}_N, \mathbf{X})^T [K(\mathbf{X}_N, \mathbf{X}_N) + \Lambda]^{-1} K(\mathbf{X}_N, \mathbf{X}). \tag{10}$$

The parameters σ_f^2, $\{l_j, j = 1, \ldots, d\}$ as well as the fidelity variance associated with each information source in Λ which are all represented as Ψ, can be obtained by performing the maximum likelihood method. This involves taking the log of the marginal likelihood, which can be written as:

$$\log p(\mathbf{y}_N | \mathbf{X}_N, \Psi) = -\frac{1}{2} \mathbf{y}_N^T K^{-1} \mathbf{y}_N - \frac{1}{2} \log |K| - \frac{N}{2} \log 2\pi, \tag{11}$$

where $K = K(\mathbf{X}_N, \mathbf{X}_N) + \Lambda$. The maximum likelihood estimates of parameters are obtained by computing the partial derivatives of the log marginal likelihood with respect to the parameters as:

$$\begin{aligned} \frac{\partial}{\partial \psi} \log p(\mathbf{y}_N | \mathbf{X}_N, \Psi) &= \frac{1}{2} \mathbf{y}_N^T K^{-1} \frac{\partial K}{\partial \psi} K^{-1} \mathbf{y}_N - \frac{1}{2} tr(K^{-1} \frac{\partial K}{\partial \psi}) \\ &= \frac{1}{2} tr\left(((K^{-1}\mathbf{y}_N)(K^{-1}\mathbf{y}_N)^T - K^{-1}) \frac{\partial K}{\partial \psi}\right), \end{aligned} \tag{12}$$

for $\psi \in \Psi$.

In our approach, we construct the fused Gaussian process by fusing the data available as training sets for all the information sources and those data queried from the information sources. After querying from each information source, the fused Gaussian process is updated. This fused Gaussian process then contains information synthesized from knowledge gained from each information source that has been queried, so it has the most comprehensive knowledge among all the information sources.

B. Gaussian Process for Constraints

For determining the next design point to evaluate, instead of computing the real-world constraints which are typically expensive, surrogate Gaussian process is constructed for each constraint. Assuming that the j^{th} constraint is evaluated at N_j design points denoted by $\mathbf{X}_j$ and the corresponding constraint values are represented by $\mathbf{g}_j$, the posterior mean μ_j of its Gaussian process can be computed in closed form for any unevaluated design $\mathbf{x}$ as:

$$\mu_j(\mathbf{x}) = K_j(\mathbf{X}_j, \mathbf{x})^T [K_j(\mathbf{X}_j, \mathbf{X}_j) + s_j^2 I]^{-1} \mathbf{g}_j, \tag{13}$$

where $K_j(\mathbf{X}_j, \mathbf{X}_j)$ is a $N_j \times N_j$ matrix with mn^{th} entry as $k_j(\mathbf{x}_{m,j}, \mathbf{x}_{n,j})$, $K_j(\mathbf{X}_j, \mathbf{x})$ is a $N_j \times 1$ vector with m^{th} entry as $k_j(\mathbf{x}_{m,j}, \mathbf{x})$, and s_j^2 is the noise variance of the Gaussian process. The posterior variance for design $\mathbf{x}$ can be computed as:

$$\sigma_j^2(\mathbf{x}) = k_j(\mathbf{x}, \mathbf{x}) - K_j(\mathbf{X}_j, \mathbf{x})^T [K_j(\mathbf{X}_j, \mathbf{X}_j) + s_j^2 I]^{-1} K_j(\mathbf{X}_j, \mathbf{x}). \tag{14}$$

These Gaussian processes for constraints are then used, as discussed in the following subsection, to find the design point with desired certainty regarding the probability of violation of constraints.

C. Choosing the Next Design and Information Source

The next step of our approach, which is determining the next design point and the information source to query from, is formulated in a decision-theoretic manner. Based on the available data, which have been used to construct the fused Gaussian process, we have a posterior distribution on $F(y|\mathbf{x})$, and a given design decision $\mathbf{x}$ has expected value under the posterior, $\mathbb{E}[F(y|\mathbf{x})|i_{1:N}, \mathbf{x}_{1:N}, y_{1:N}]$. If we were to make no additional queries, the best expected objective value would be:

$$y_N^* = \max_{\mathbf{x}\in\chi} \mathbb{E}[F(y|\mathbf{x})|i_{1:N}, \mathbf{x}_{1:N}, y_{1:N}]. \tag{15}$$

Similarly, if we were to make one additional query, we would obtain a value:

$$y_{N+1}^* = \max_{\mathbf{x}\in\chi} \mathbb{E}[F(y|\mathbf{x})|i_{1:N+1}, \mathbf{x}_{1:N+1}, y_{1:N+1}]. \tag{16}$$

The difference $y_{N+1}^* - y_N^*$ is the improvement in value that results from the additional query. The essential idea in such a value of information analysis is to choose the query or set of queries that maximizes this improvement. This improvement depends on y_{N+1}, and so it is unknown when choosing i_{N+1} and $\mathbf{x}_{N+1}$, but its posterior predictive distribution can be determined, and thus the expected value of $y_{N+1}^* - y_N^*$ can be evaluated under the posterior at time N. This value is referred to as the expected improvement (EI) defined as:

$$\begin{aligned} EI(i,\mathbf{x}) &= \mathbb{E}\left[\max_{\mathbf{x}\in\chi} \mathbb{E}[F(y|\mathbf{x})|i_{1:N}, i_{N+1}=i, \mathbf{x}_{1:N}, \mathbf{x}_{N+1}=\mathbf{x}, y_{1:N}] - \max_{\mathbf{x}\in\chi} \mathbb{E}[F(y|\mathbf{x})|i_{1:N}, \mathbf{x}_{1:N}, y_{1:N}]\right] \\ &= \mathbb{E}\left[\max_{\mathbf{x}\in\chi} \mathbb{E}[F(y|\mathbf{x})|i_{1:N}, i_{N+1}=i, \mathbf{x}_{1:N}, \mathbf{x}_{N+1}=\mathbf{x}, y_{1:N}]\right] - \max_{\mathbf{x}\in\chi} \mathbb{E}[F(y|\mathbf{x})|i_{1:N}, \mathbf{x}_{1:N}, y_{1:N}]. \end{aligned} \tag{17}$$

We propose to use a Knowledge Gradient (KG) approach as a measure of expected improvement in order to determine the query $N+1$ given N prior queries and the information source to query from, and thus build off the work of Refs. 36–38. KG takes an information-economic approach to maximizing (or minimizing) an objective using a single information source with noisy observations from that information source. Specifically, let assume we have a set of designs with normally distributed beliefs about their quality. After N queries, we have a vector of means $\theta^N = \mathbb{E}[F(y|\mathbf{x})|i_{1:N}, \mathbf{x}_{1:N}, y_{1:N}]$ which is a discretization of a continuous function and a vector of precisions, $\beta^N = 1/\mathrm{var}(F(y|\mathbf{x})|i_{1:N}, \mathbf{x}_{1:N}, y_{1:N})$. The knowledge state is thus $S^N = (\theta^N, \beta^N)$. If the design process stops now, we would choose the best design based on the current knowledge, which is given by:

$$\mathbf{x}^N = \operatorname*{argmax}_{\mathbf{x}\in\chi} \theta_{\mathbf{x}}^N. \tag{18}$$

The value of being in state S^N is then given as $V^N(S^N) = \theta_{\mathbf{x}^N}^N$. If one additional query can be made, then the value of being in the next knowledge state is given as $V^{N+1}(S^{N+1}(\mathbf{x})) = \max_{\mathbf{x}'\in\chi} \theta_{\mathbf{x}'}^{N+1}$. The knowledge gradient is then defined as:

$$\nu_{\mathbf{x}}^{KG,N} = \mathbb{E}[V^{N+1}(S^{N+1}(\mathbf{x})) - V^N(S^N)|S^N], \tag{19}$$

which can be viewed as the gradient of $V^N(S^N)$ with respect to querying the single information source at $\mathbf{x}$ [37]. The knowledge gradient policy for sequentially choosing the next query is then given as:

$$\mathbf{x}^{KG,N} = \operatorname*{argmax}_{\mathbf{x}\in\chi} \nu_{\mathbf{x}}^{KG,N}. \tag{20}$$

Here, we need to compute the knowledge gradient for the fused Gaussian process in order to determine the next design point and the information source to query. The knowledge gradient at design point $\mathbf{x}$ for the i^{th} information source given N available data can be written as [37]:

$$\nu_{\mathbf{x},i}^{KG,N} = \mathbb{E}[\max_{\mathbf{x}'\in\chi} \theta_{\mathbf{x}'}^N + \tilde{\sigma}_{\mathbf{x}',i}(\Sigma^N, \mathbf{x}) Z|S^N] - \max_{\mathbf{x}'\in\chi} \theta_{\mathbf{x}'}^N, \tag{21}$$

where Z is a standard normal random variable, $\theta_{\mathbf{x}'}^N$ and the column vector $\tilde{\sigma}_{\mathbf{x}',i}(\Sigma^N, \mathbf{x})$ are the predicted expected value and uncertainty of θ at design point $\mathbf{x}'$ if design $\mathbf{x}$ is evaluated at step N. The column vector $\tilde{\sigma}_{\mathbf{x}',i}(\Sigma^N, \mathbf{x})$ is given by:

$$\tilde{\sigma}_{\mathbf{x}',i}(\Sigma^N, \mathbf{x}) = \frac{\Sigma^N e_{\mathbf{x}}}{\sqrt{\lambda_i + \Sigma_{\mathbf{xx}}^N}}, \tag{22}$$

where $e_{\mathbf{x}}$ is a column vector of 0's with a 1 in the position corresponding to $\mathbf{x}$, Σ^N is the covariance between the measurements $\mathbf{x}'$ in the fused Gaussian process, and λ_i is the fidelity variance of information source i. Calculation of the knowledge gradient, which is based on a piecewise linear function, is discussed in detail in two algorithms presented in Ref. 38.

By denoting $C_i(\mathbf{x})$ as the cost of querying information source i at design $\mathbf{x}$, without considering constraints, we find the query $(i_{N+1}, \mathbf{x}_{N+1})$ that minimizes the expected cost per unit expected improvement, given by:

$$\begin{aligned}(i_{N+1}, \mathbf{x}_{N+1}) &= \underset{i\in[1,\ldots,M],\mathbf{x}\in\chi}{\operatorname{argmin}} \frac{C_i(\mathbf{x})}{EI(i,\mathbf{x})} \\ &= \underset{i\in[1,\ldots,M],\mathbf{x}\in\chi}{\operatorname{argmin}} \frac{C_i(\mathbf{x})}{\nu^{KG,N}_{\mathbf{x},i}}.\end{aligned} \tag{23}$$

Equation (23) selects the best design sample and the information source to query. However, not all the evaluated points in the design space satisfy all the constraints. We need to minimize Equation (23), to obtain the maximum expected improvement with minimum possible cost over all feasible design points.

In order to consider constraints, penalty functions have been widely used due to their simplicity. These functions include both static and dynamic strategies. In the static case, a constant penalty is applied to those solutions that violate feasibility in any way. In some problems, the optimal solution lies on the boundary of the feasible region. In addition, in some highly constrained problems, finding feasible samples might be difficult, and the search has to begin with infeasible samples. Therefore, restricting the search to only feasible solutions or imposing very severe penalties makes it difficult to find the optimum solution. On the other hand, if the penalty is not severe enough, then too large a region is searched and much of the search time will be used to explore regions far from the feasible region. In the dynamic case, these issues are handled by increasing the severity of the penalty as the search progresses. This has the property of allowing highly infeasible solutions early in the search, while continually increasing the penalty imposed to eventually move the final solution to the feasible region [42]. The Gaussian processes constructed for constraints in Subsection B are Bayesian representation of constraints. The uncertainty in these GPs may come from different sources of uncertainty, i.e. model discrepancy, parametric uncertainty, code uncertainty, etc. Therefore, we need to develop a methodology which is able to handle the stochasticity in constraints. To perform the penalty method in our approach, the penalty function multiplied by the probability of violation of constraints is added to the cost of querying of information source i as:

$$C_{p_i}(\mathbf{x}) = C_i(\mathbf{x}) + f_p \times P_v(\mathbf{x}), \tag{24}$$

where $C_{p_i}(\mathbf{x})$ is the penalized cost function at design $\mathbf{x}$, $C_i(\mathbf{x})$ is the cost of querying from information source i, f_p is the penalty function, and $P_v(\mathbf{x})$ is the probability of violation of constraints at design $\mathbf{x}$. In the first steps of the process, we aim to involve the infeasible samples in the search process to explore the information they might carry. Our approach achieves this idea by using a dynamic penalty function strategy. In this strategy, the value of penalty function is small in the first iterations, which helps the methodology to explore the entire search space, and as the process goes on, its value increases which decreases the probability of an infeasible sample to be selected. Thus, we consider the penalty function which varies with the number of iterations, defined as:

$$f_p = C_v \times \left(1 - \exp(-\frac{t}{\tau})\right), \tag{25}$$

where C_v is a positive penalty constant imposed as the cost of violation of constraints, t is the iteration number, and τ is a user-defined value which determines the rate of increase of penalty as the number of iterations increases. This value is specified based on the available budget and risk attitude of the user.

The probability of violation of constraint j at a design sample is computed based on the normal distribution coming from the corresponding Gaussian process. According to Equations (13 - 14), the value of constraint g_j at a design point $\mathbf{x}$ is distributed normally with mean $\mu_j(\mathbf{x})$ and variance $\sigma_j^2(\mathbf{x})$ as:

$$g_j(\mathbf{x}) \sim \mathcal{N}(\mu_j(\mathbf{x}), \sigma_j^2(\mathbf{x})). \tag{26}$$

We use the method of cumulative distribution function to compute the probability of violating the constraint g_j by design point $\mathbf{x}$ as:

$$P_j(\mathbf{x}) = P\left(g_j(\mathbf{x}) > 0\right) = 1 - P\left(g_j(\mathbf{x}) < 0\right). \tag{27}$$

Assuming that we have m independent constraints, the probability of violation of constraints at design $\mathbf{x}$, $P_v(\mathbf{x})$, can be computed as:

$$\begin{aligned}
P_v &= P\left(g_1(\mathbf{x}) > 0 \cup g_2(\mathbf{x}) > 0 \cup \dots \cup g_m(\mathbf{x}) > 0\right) \\
&= 1 - P\left(g_1(\mathbf{x}) < 0 \cap g_2(\mathbf{x}) < 0 \cap \dots \cap g_m(\mathbf{x}) < 0\right) \\
&= 1 - \prod_{j=1}^{m} P\left(g_j(\mathbf{x}) < 0\right) \\
&= 1 - \prod_{j=1}^{m} \left(1 - P\left(g_j(\mathbf{x}) > 0\right)\right) \\
&= 1 - \prod_{j=1}^{m} \left(1 - P_j(\mathbf{x})\right).
\end{aligned} \tag{28}$$

Therefore, to select the next information source, i_{N+1}, and the next design sample, $\mathbf{x}_{N+1}$, Equation (23) is modified as:

$$(i_{N+1}, \mathbf{x}_{N+1}) = \underset{i \in [1,\dots,M], \mathbf{x} \in \chi}{\operatorname{argmin}} \left(\frac{C_i(\mathbf{x}) + f_p \times P_v(\mathbf{x})}{\nu_{\mathbf{x},i}^{KG,N}} \right). \tag{29}$$

Latin Hypercube sampling is used to generate samples in the input design space χ for maximization of the knowledge gradient for problems with multidimensional continuous design variables. Assuming that S Latin Hypercube samples are generated as alternatives, and M information sources are available, $S \times M$ values are computed according to Equation (29). As we seek low evaluation cost for a large performance gain by considering the constraints, the sample and the information source that obtain the minimum value are selected to be evaluated. After querying from the selected information source, the fused Gaussian process of the objective function gets updated according to Equations (8-10). Furthermore, the constraints are evaluated at the selected design sample and the corresponding Gaussian processes get updated. This process repeats until a termination criterion, such as exhaustion of the querying budget, is met. In order to find the optimum solution of Equation (1), we discretize the design space evenly. The grids must be sufficiently dense to get a good approximation to the continuous design space. Then, the value of the objective function and the probability of violating the constraints are computed at each point according to Equations (9) and (28) respectively. Among all the points with probability of violation less than a specified value, γ, the point that has the maximum function value is selected as the optimum solution.

Our proposed fusion-based multi-information source optimization approach using a knowledge gradient policy is presented in Algorithm 1.

Algorithm 1: Multi-Information Source KG Policy

1: Construct the fused Gaussian process for the objective function by fusing the data of all information sources according to Equations (8 - 10).

2: Construct Gaussian process for the constraints.

repeat

3: Generate Latin Hypercube samples in input design space χ.

4: Select the sample and the information source according to Equation (29).

5: Update the fused GP of the objective function and GP of the constraints based on the observation obtained from the chosen information source at selected design sample.

until termination

6: Return the point with the largest estimated value according to the fused GP among the points with the probability of violating the constraints less than the specified value γ.

III. Application and Results

In this section, we present the key features of our fusion-based multi-information source optimization approach using a knowledge gradient policy on two demonstrations. The first case is an analytical problem with one-dimensional input and output, and the second demonstration is the minimization of the drag coefficient of a NACA 0012 airfoil subject to a constraint on the lift coefficient.

A. One-Dimensional Function

The first example is maximization of a one-dimensional constrained function shown in Figure 1 and defined as:

$$\begin{aligned} x^* = &\underset{\mathbf{x}\in[0\ 1.2]}{\text{argmax}}\ -(1.4-3x)\sin(18x) \\ &\text{s.t.}\quad x^2 - 1.2 \leq 0. \end{aligned} \tag{30}$$

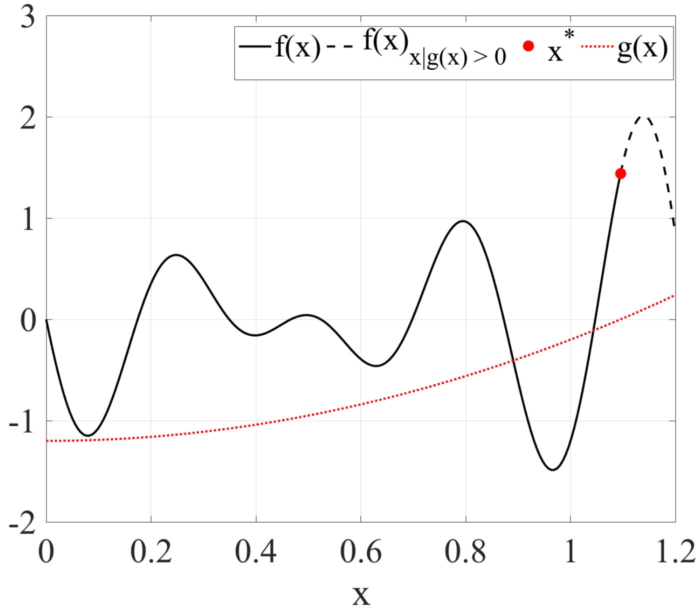

Figure 1: One-dimensional optimization problem of Equation (30).

We consider two information sources f_1 and f_2 with added model discrepancy with fidelity variances of $\lambda_1 = 10^{-5}$ and $\lambda_2 = 10^{-2}$, and constraint with noise variance of $\sigma^2 = 10^{-3}$, given as:

$$\begin{aligned} f_1(x) &= -(1.4-3x)\sin(18x) + \mathcal{N}(0,\, \lambda_1), \\ f_2(x) &= -(1.4-3x)\sin(18x) + \mathcal{N}(0,\, \lambda_2), \\ g_1(x) &= x^2 - 1.2 + \mathcal{N}(0,\, \sigma^2). \end{aligned} \tag{31}$$

The evaluation costs for information sources are set to $C_1 = 20$ and $C_2 = 15$, and we assume $C_v = 5$.

Figure 2 shows the objective function and the constraint with $(1-2\gamma)$ confidence interval. The solid lines show the feasible region and dashed lines show the objective function where the probability of violation is greater than γ. In the right plot, it can be seen that the optimum point is in the right side of the plot, however due to the infeasibility of the tallest hip of the function, the selected point is in valley of the function. On the other hand, for $\gamma = 10\%$ the optimum solution is shifted to the left side, due to the expansion of infeasible region in the right hand side of the objective function. This shows the effect of risk attitude on the optimal solution and decision making task.

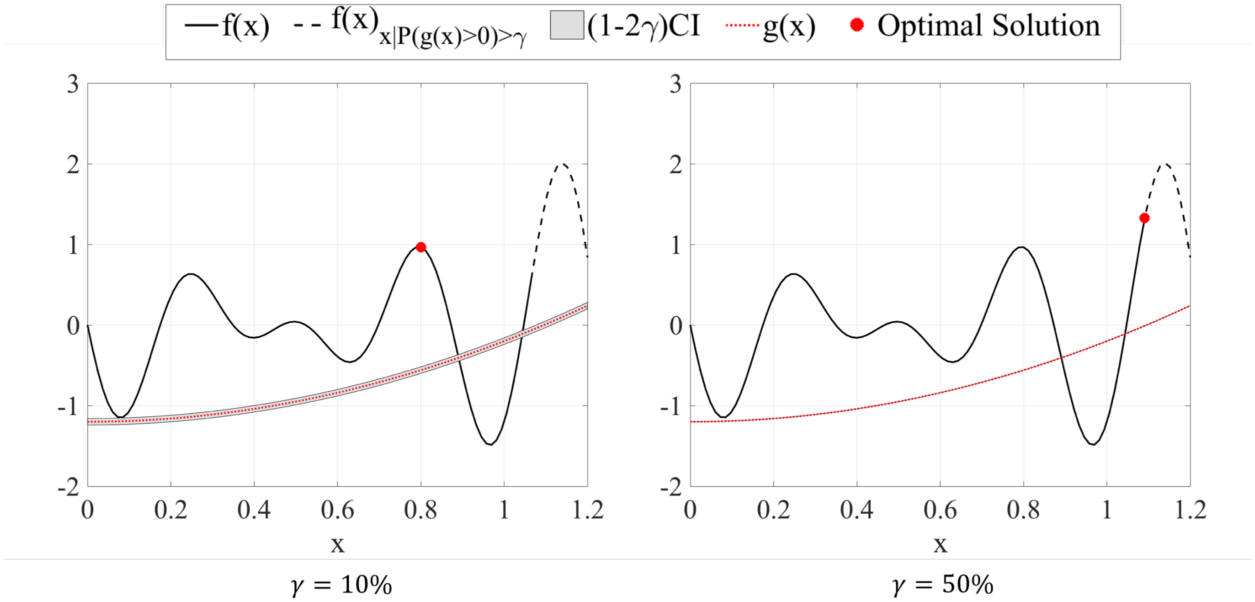

Figure 2: The objective function, constraint and optimal solution of optimization problem in Equation (30) for $\gamma = 10\%$ and $\gamma = 50\%$.

In order to asses the performance of our approach, we compare our methodology with the method presented in Ref. 12 which is called MF algorithm. In Ref. 12, each information source has a separate Gaussian process, and predictions are obtained by fusing the information via the method presented in Ref. 43. The next design to evaluate is selected by applying the expected improvement function on these surrogates, and the next information source to query is chosen based on a heuristic that aims to balance information gain and cost of query. Table 1 represents the expected results obtained by our approach and MF algorithm for $\gamma = 10\%$ and $\gamma = 50\%$. The expected results are averaged over 500 replications of the simulations. The total cost of querying as the stopping criterion for these simulations is set to $C_{tot} = 150$. As it can be seen, our approach improves on the work of Ref. 12. It is due to the fact that our methodology

Table 1: Expected results obtained by our proposed approach and MF Algorithm [12] over 500 replications of the simulations of optimization problem in Equation (30) for $\gamma = 10\%$ and $\gamma = 50\%$.

		Optimal Solution	Proposed Approach	MF Algorithm [12]
$\gamma = 10\%$	x^*	0.8000	0.8010	0.6999
	f^*	0.9657	0.9766	0.8537
$\gamma = 50\%$	x^*	1.0909	1.0769	0.9556
	f^*	1.3261	1.3449	1.2309

allows rigorously exploiting correlations across the design space which reduces the uncertainty by querying one new design sample, even if it is queried from a information source with lower fidelity. Thus, we obtain a more accurate estimate of the true objective function from each sample. Note that this feature is also accomplished by constructing our fused Gaussian process which has the most comprehensive knowledge among all the information sources by fusing all the queried samples and fidelity variances of information sources.

Figure 3 shows the effect of value of τ on sampling from the infeasible region. As it has been shown in the left plot of Figure 2, the right hand side of function is infeasible for $\gamma = 10\%$. However, the infeasibility of the right hand side should be learnt by sequential sampling of both objective and constraint. This requires exploration of our design space during sequential decision making process. This fact can be controlled by the use of parameter τ as part of our penalty function. As it can be seen in Figure 3, as the value of τ increases, the rate of increase of penalty decreases and in the first iterations, samples can be selected in the infeasible region. Therefore, the value of τ is selected based on the available budget and risk attitude of user.

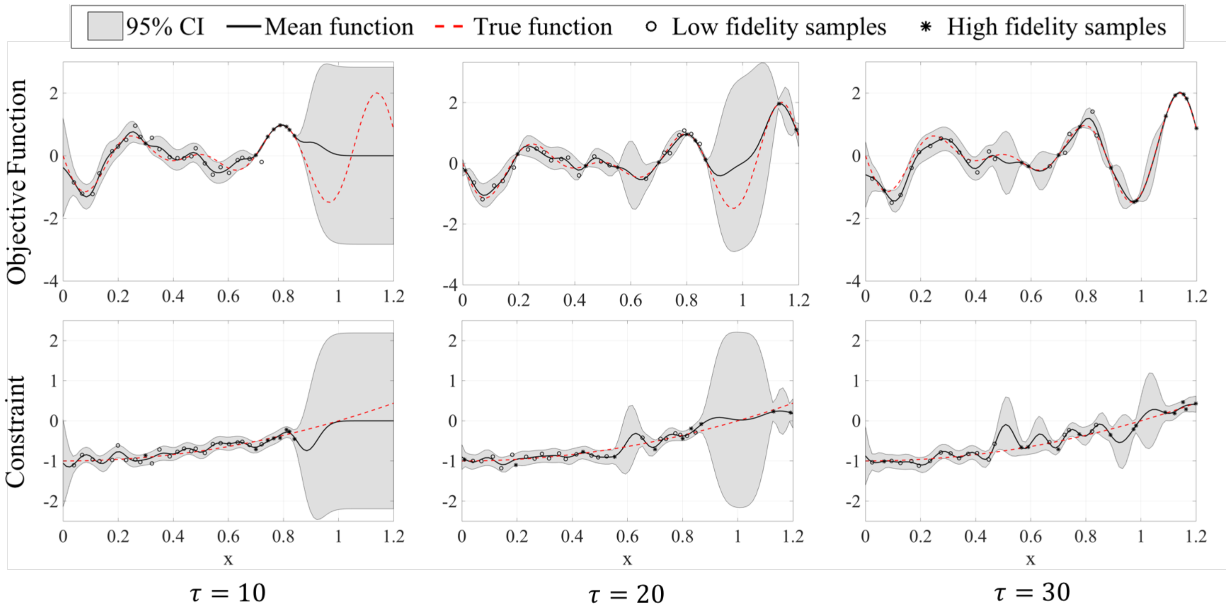

Figure 3: Effect of τ in querying infeasible samples.

B. NACA 0012 Drag Coefficient with Lift Coefficient Constraint

The second demonstration to apply our fusion-based multi-information source optimization approach is a two-dimensional constrained aerodynamic design example. The airfoil of interest is the NACA 0012, a common validation airfoil [44, 45]. Since Navier-Stokes equations used to solve the fluid dynamics are expensive to solve, simplified equations have been developed to solve this problem. These simplified equations consider some assumptions which lead to variable fidelities. Here, the computational fluid dynamics programs XFOIL [46] and SU2 [47] are used as the two information sources. XFOIL is a solver for the design and analysis of airfoils in the subsonic regime. It combines a panel method with the Karman-Tsien compressibility correction for the potential flow with a two-equation boundary layer model. This causes XFOIL to overestimate lift and underestimate drag [48]. SU2 uses a finite volume scheme and Reynolds-averaged Navier-Stokes (RANS) method with the Spalart-Allmaras turbulence model, which allows SU2 to be significantly more accurate than XFOIL in the more turbulent flow regimes at higher values of Mach number and angle of attack.

In this problem, we are particularly concerned with finding the Mach number M and angle of attack α that minimize the coefficient of drag C_D of a NACA 0012 airfoil subject to maintaining a minimum coefficient of lift C_L:

$$\begin{aligned} \mathbf{x}^* = & \underset{\mathbf{x}\in\chi}{\operatorname{argmax}} \; -C_D \\ & \text{s.t.} \quad 0.4 - C_L \leq 0. \end{aligned} \tag{32}$$

where $\mathbf{x}^* = [M^*, \alpha^*]$. The design space is $\chi = I_M \times I_\alpha$ with $I_M = [0.15 \; 0.75]$ and $I_\alpha = [-2.2 \; 13.3]$. The objective and constraint, as well as contour plot of objective and constraint are shown in Figure 4.

We set the evaluation costs to be $C_1 = 500$ and $C_2 = 300$, and the fidelity variances are set to be $\lambda_1 = 10^{-5}$ and $\lambda_2 = 10^{-3}$ for information sources and $\sigma^2 = 10^{-5}$ for fidelity variance of the constraint. The maximum probability of violating the constraint is $\gamma = 10\%$, and the computational budget is limited to 15000. The penalty term is set to $C_v = 400$.

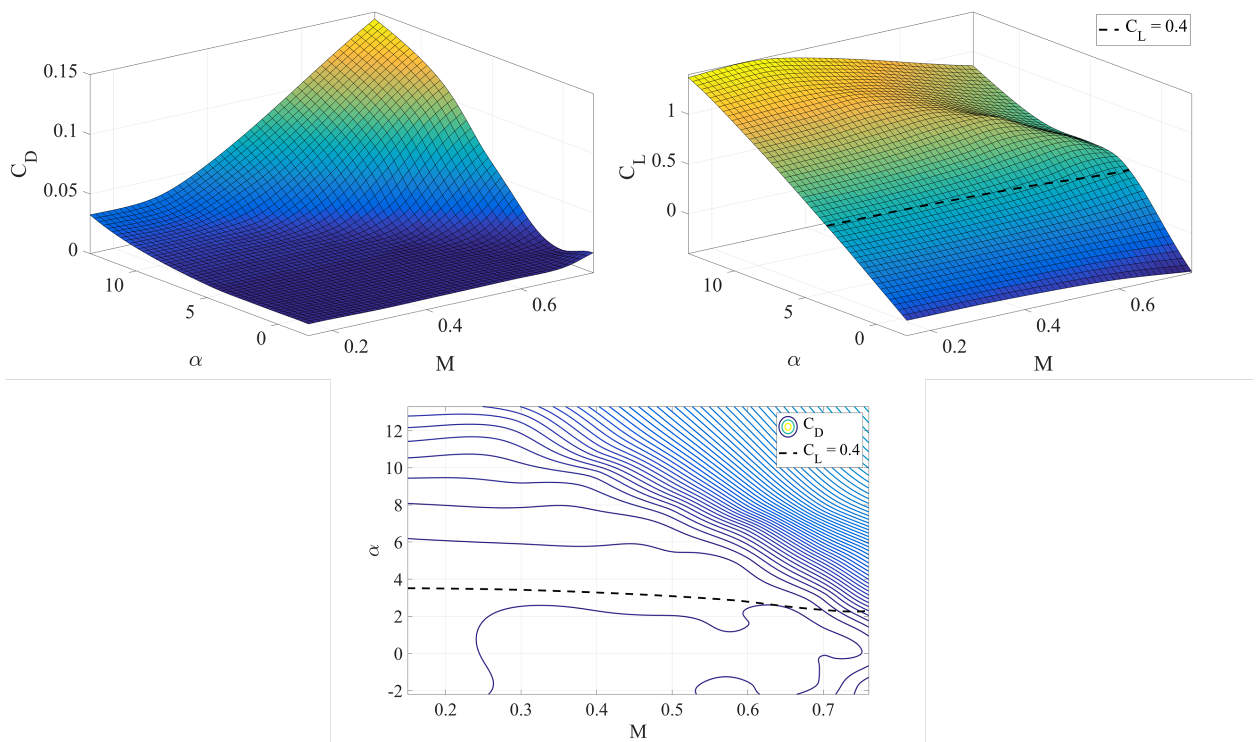

Figure 4: Optimization problem of Equation (32).

Table 2 represents the expected results obtained by our approach and MF algorithm [12] for $\gamma = 10\%$. The expected results are averaged over 500 replications of the simulations. The Monte Carlo solution is obtained by discretization of the design space into 2500 grids and finding the point with minimum coefficient of drag which does not violate the constraint on coefficient of lift. It is clear that our proposed methodology outperforms the MF Algorithm by obtaining the lowest expected objective value which is closer to the Monte Carlo result.

Table 2: Expected results obtained by our proposed approach and MF Algorithm [12] over 500 replications of the simulations of optimization problem in Equation (32) for $\gamma = 10\%$.

	Monte Carlo Solution	Proposed Approach	MF Algorithm [12]
f^*	0.0091	0.0098	0.0382

IV. Conclusion

This paper has presented an approach to perform constrained optimization of expensive to evaluate functions when different information sources with varying fidelities and evaluation costs are available. The approach considers the trade off between cost and performance gain for querying from information sources of varying fidelity to find the design decision which optimizes the objective function. This is achieved by fusing the information obtained from information sources to construct the fused Gaussian process. Then, the knowledge gradient policy is incorporated as a measure of expected improvement to identify the next design to evaluate, as well as to select the information source with which to perform the evaluation. This is performed based on the evaluation cost and fidelity of the information sources. The proposed strategy samples the design space by balancing exploration and exploitation tasks both between and within the available information sources. We demonstrated our approach on the optimization of a one-dimensional example test problem and an aerodynamic design example. It has been shown that the proposed approach finds the feasible optimum value of objective in a high performance decision-theoretic manner.

Acknowledgments

This work was supported by the AFOSR MURI on multi-information sources of multi-physics systems under Award Number FA9550-15-1-0038, program manager, Dr. Jean-Luc Cambier.

References

[1]Alexandrov, N., Lewis, R., Gumbert, C., Green, L., and Newman, P., "Optimization with variable-fidelity models applied to wing design," *38th Aerospace Sciences Meeting and Exhibit*, 2000, p. 841.

[2]Jones, D. R., Schonlau, M., and Welch, W. J., "Efficient global optimization of expensive black-box functions," *Journal of Global optimization*, Vol. 13, No. 4, 1998, pp. 455–492.

[3]Booker, A. J., Dennis Jr, J., Frank, P. D., Serafini, D. B., Torczon, V., and Trosset, M. W., "A rigorous framework for optimization of expensive functions by surrogates," *Structural optimization*, Vol. 17, No. 1, 1999, pp. 1–13.

[4]Jones, D. R., "A taxonomy of global optimization methods based on response surfaces," *Journal of global optimization*, Vol. 21, No. 4, 2001, pp. 345–383.

[5]Sasena, M. J., Papalambros, P., and Goovaerts, P., "Exploration of metamodeling sampling criteria for constrained global optimization," *Engineering optimization*, Vol. 34, No. 3, 2002, pp. 263–278.

[6]Alexandrov, N. M., Dennis Jr, J. E., Lewis, R. M., and Torczon, V., "A trust-region framework for managing the use of approximation models in optimization," *Structural optimization*, Vol. 15, No. 1, 1998, pp. 16–23.

[7]Alexandrov, N. M., Lewis, R. M., Gumbert, C. R., Green, L. L., and Newman, P. A., "Approximation and model management in aerodynamic optimization with variable-fidelity models," *Journal of Aircraft*, Vol. 38, No. 6, 2001, pp. 1093–1101.

[8]March, A. and Willcox, K., "Provably convergent multifidelity optimization algorithm not requiring high-fidelity derivatives," *AIAA journal*, Vol. 50, No. 5, 2012, pp. 1079–1089.

[9]Rasmussen, C. E., "Gaussian processes for machine learning," 2006.

[10]Kleijnen, J. P., "Kriging metamodeling in simulation: A review," *European journal of operational research*, Vol. 192, No. 3, 2009, pp. 707–716.

[11]Huang, D., Allen, T., Notz, W., and Miller, R., "Sequential kriging optimization using multiple-fidelity evaluations," *Structural and Multidisciplinary Optimization*, Vol. 32, No. 5, 2006, pp. 369–382.

[12]Lam, R., Allaire, D. L., and Willcox, K. E., "Multifidelity optimization using statistical surrogate modeling for non-hierarchical information sources," *56th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, 2015, p. 0143.

[13]Allaire, D., Willcox, K., and Toupet, O., "A bayesian-based approach to multifidelity multidisciplinary design optimization," *13th AIAA/ISSMO Multidisciplinary Analysis Optimization Conference*, 2010, p. 9183.

[14]Schonlau, M., *Computer experiments and global optimization*, University of Waterloo Waterloo, 1998.

[15]Audet, C., Denni, J., Moore, D., Booker, A., and Frank, P., "A surrogate-model-based method for constrained optimization," *8th Symposium on Multidisciplinary Analysis and Optimization*, 2000, p. 4891.

[16]Missoum, S., Ramu, P., and Haftka, R. T., "A convex hull approach for the reliability-based design optimization of nonlinear transient dynamic problems," *Computer methods in applied mechanics and engineering*, Vol. 196, No. 29, 2007, pp. 2895–2906.

[17]Basudhar, A., Missoum, S., and Sanchez, A. H., "Limit state function identification using support vector machines for discontinuous responses and disjoint failure domains," *Probabilistic Engineering Mechanics*, Vol. 23, No. 1, 2008, pp. 1–11.

[18]Basudhar, A. and Missoum, S., "Classification-based Quantification of Constraint Violation for Efficient Global Optimization and Failure Probability Confidence Bound Calculation," .

[19]Platt, J. et al., "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Advances in large margin classifiers*, Vol. 10, No. 3, 1999, pp. 61–74.

[20]Basudhar, A. and Missoum, S., "Reliability assessment using probabilistic support vector machines," *International Journal of Reliability and Safety*, Vol. 7, No. 2, 2013, pp. 156–173.

[21]Friedman, S., Ghoreishi, S. F., and Allaire, D. L., "Quantifying the Impact of Different Model Discrepancy Formulations in Coupled Multidisciplinary Systems," *19th AIAA Non-Deterministic Approaches Conference*, 2017, p. 1950.

[22]Imani, M. and Braga-Neto, U. M., "Maximum-Likelihood Adaptive Filter for Partially Observed Boolean Dynamical Systems," *IEEE Transactions on Signal Processing*, Vol. 65, No. 2, pp. 359–371.

[23]Ghoreishi, S. and Allaire, D., "Adaptive uncertainty propagation for coupled multidisciplinary systems," *AIAA Journal*, 2017, pp. 1–11.

[24]Mosleh, A. and Apostolakis, G., "The assessment of probability distributions from expert opinions with an application to seismic fragility curves," *Risk Analysis*, Vol. 6, No. 4, 1986, pp. 447–461.

[25]Reinert, J. M. and Apostolakis, G. E., "Including model uncertainty in risk-informed decision making," *Annals of nuclear energy*, Vol. 33, No. 4, 2006, pp. 354–369.

[26]Riley, M. E., Grandhi, R. V., and Kolonay, R., "Quantification of modeling uncertainty in aeroelastic analyses," *Journal of Aircraft*, Vol. 48, No. 3, 2011, pp. 866–873.

[27]Leamer, E. E., *Specification searches: Ad hoc inference with nonexperimental data*, Vol. 53, John Wiley & Sons Incorporated, 1978.

[28]Madigan, D. and Raftery, A. E., "Model selection and accounting for model uncertainty in graphical models using Occam's window," *Journal of the American Statistical Association*, Vol. 89, No. 428, 1994, pp. 1535–1546.

[29]Draper, D., "Assessment and propagation of model uncertainty," *Journal of the Royal Statistical Society. Series B (Methodological)*, 1995, pp. 45–97.

[30]Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T., "Bayesian model averaging: a tutorial," *Statistical science*, 1999, pp. 382–401.

[31]Imani, M. and Braga-Neto, U., "Particle filters for partially-observed Boolean dynamical systems," *Automatica*, 2017.

[32]Allaire, D. and Willcox, K., "Fusing information from multifidelity computer models of physical systems," *Information Fusion (FUSION), 2012 15th International Conference on*, IEEE, 2012, pp. 2458–2465.

[33]Thomison, W. D. and Allaire, D. L., "A Model Reification Approach to Fusing Information from Multifidelity Information Sources," *19th AIAA Non-Deterministic Approaches Conference*, 2017, p. 1949.

[34]Gupta, S. S. and Miescke, K. J., "Bayesian look ahead one-stage sampling allocations for selection of the best population," *Journal of statistical planning and inference*, Vol. 54, No. 2, 1996, pp. 229–244.

[35]Scott, W., Frazier, P., and Powell, W., "The correlated knowledge gradient for simulation optimization of continuous parameters using gaussian process regression," *SIAM Journal on Optimization*, Vol. 21, No. 3, 2011, pp. 996–1026.

[36]Frazier, P. I., Powell, W. B., and Dayanik, S., "A knowledge-gradient policy for sequential information collection," *SIAM Journal on Control and Optimization*, Vol. 47, No. 5, 2008, pp. 2410–2439.

[37]Powell, W. B. and Ryzhov, I. O., *Optimal learning*, Vol. 841, John Wiley & Sons, 2012.

[38]Frazier, P., Powell, W., and Dayanik, S., "The knowledge-gradient policy for correlated normal beliefs," *INFORMS journal on Computing*, Vol. 21, No. 4, 2009, pp. 599–613.

[39]Kennedy, M. C. and O'Hagan, A., "Bayesian calibration of computer models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 63, No. 3, 2001, pp. 425–464.

[40]Ghoreishi, S. F. and Allaire, D. L., "Compositional uncertainty analysis via importance weighted gibbs sampling for coupled multidisciplinary systems," *18th AIAA Non-Deterministic Approaches Conference*, 2016, p. 1443.

[41]Ghoreishi, S. F., *Uncertainty analysis for coupled multidisciplinary systems using sequential importance resampling*, Master's thesis, 2016.

[42]Smith, A., Smith, A. E., Coit, D. W., Baeck, T., Fogel, D., and Michalewicz, Z., "Penalty functions," 1997.

[43]Winkler, R. L., "Combining probability distributions from dependent information sources," *Management Science*, Vol. 27, No. 4, 1981, pp. 479–488.

[44]Thibert, J., Grandjacques, M., Ohman, L., et al., "NACA 0012 airfoil," *AGARD Advisory Report*, Vol. 138, 1979.

[45]Rumsey, C., "2D NACA 0012 Airfoil Validation Case," *Turbulence Modeling Resource, NASA Langley Research Center*, 2014, pp. 33.

[46]Drela, M., "XFOIL: An analysis and design system for low Reynolds number airfoils," *Low Reynolds number aerodynamics*, Springer, 1989, pp. 1–12.

[47]Palacios, F., Alonso, J., Duraisamy, K., Colonno, M., Hicken, J., Aranake, A., Campos, A., Copeland, S., Economon, T., Lonkar, A., et al., "Stanford University Unstructured (SU 2): an open-source integrated computational environment for multi-physics simulation and design," *51st AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition*, 2013, p. 287.

[48]Barrett, R. and Ning, A., "Comparison of Airfoil Precomputational Analysis Methods for Optimization of Wind Turbine Blades," *IEEE Transactions on Sustainable Energy*, Vol. 7, No. 3, 2016, pp. 1081–1088.