10th World Congress on Structural and Multidisciplinary Optimization
May 19 -24, 2013, Orlando, Florida, USA

Inclusion of Aleatory and Epistemic Uncertainty in Design Optimization

Sirisha Rangavajhala[1], Chen Liang[2] , Sankaran Mahadevan[3]

[1] Philips Research North America, New York City, New York, USA, sirisha.rangavajhala@gmail.com
[2] Vanderbilt University, Nashville, Tennessee, USA, chen.liang.1@vanderbilt.edu

1. Abstract
This paper presents a design optimization methodology under three sources of uncertainty: physical variability (aleatory); data uncertainty (epistemic) due to sparse or imprecise data; and model uncertainty (epistemic) due to modeling errors/approximations. A likelihood-based method is use to fuse multiple formats of information, and a non-parametric probability density function (PDF) is constructed. Two types of model errors are considered: model form error and numerical solution error, each of which is a function of the design variables that changes at every iteration of the optimization. Gaussian process (GP) surrogate models are constructed for efficient computation of model errors in the optimization. The treatment in this paper yields a distribution of the output that accounts for various sources of uncertainty. The use of a probabilistic approach to include both aleatory and epistemic uncertainties allows for their efficient integration into the optimization framework. The proposed methods are illustrated using a three-dimensional wing design problem involving fluid-structure interaction analysis.
2. Keywords: Optimization, Epistemic, Surrogate model

3. Introduction
Optimization methods under uncertainty within engineering design have generally been pursued in two directions: (1) reliability-based design optimization (RBDO), where the focus has been on achieving desired reliability levels for constraints, and (2) robust design optimization (RDO), where the primary focus has been on minimizing objective function variations, typically quantified by their respective variances. In this paper, we consider three sources of uncertainty in design optimization and engineering analysis: (1) Variability (aleatory) due to the inherent physical variability in all processes. (2) Data uncertainty (epistemic) due to lack of sufficient information and (3) Model uncertainty (epistemic) caused due to model form assumptions as well as numerical approximations at several stages of the analysis.
Design optimization under aleatory uncertainty is a well-researched topic in the literature. However, when epistemic uncertainty is considered, the existing methods address only parts of the entire problem scope. Methods to quantify model errors usually consider a single set of inputs, and this process can be computationally expensive in itself. The design optimization setting however poses additional challenges, because each iterate of the optimization represents a different set of inputs, and the model errors are therefore potentially different at different regions of the design space. In this paper, we consider the aleatory uncertainty, data uncertainty and model errors within an optimization framework. A likelihood-based approach is used to probabilistically quantify these uncertainties. The constraint and objective functions are assumed to be available only through computationally expensive simulation models. For each function, the model form error and numerical solution error are quantified, as a function of the design variables. The computationally expensive objective/constraint calculation model is replaced by a Gaussian process surrogate model for the sake of computational efficiency [1]. The resulting surrogate model error is also quantified as a function of the design variable inputs. The contributions of is paper can be summarized as follows:

1. Development of a probabilistic framework to include natural variability, data uncertainty and model uncertainty within the design optimization problem
2. Quantification of various types of model errors as functions of design variable values including model form error, discretization error, and surrogate model error
3. Uncertainty quantification in model output, through probability distributions, due to variability, data uncertainty, and model errors.
4. Illustration of the proposed methodology with a wing design problem.

4. RBDO formulation

An RBDO formulation of the above problem is given in [2] as following:

$$\min_{\mu_X, d} [\mu_f(X, d, P, p_d)]$$
$$s.t. \quad Prob(g_i(X, d, P, p_d) \le 0) \ge p_t^i \; i = \{1, \dots, n_q\} \quad (1)$$
$$Prob(X \ge lb_X) \ge p_{lb}^t$$
$$Prob(X \ge ub_X) \ge p_{ub}^t$$
$$lb_d \le d \le ub_d$$

where X is the vector of random design variables with bounds lb_X and ub_X, respectively; μ_f and μ_X are the mean of f and X, respectively; d is the vector of deterministic design variables with bounds lb_d and ub_d; P is the vector of random parameters; P_d is the vector of deterministic parameters; and p_t^i is the target reliability required for the i^{th} inequality constraint; p_{lb}^t and p_{ub}^t are the target reliabilities for the design variable bounds.

5. Sources of errors and uncertainties

5.1. Data uncertainty (epistemic): design and non-design variables

A likelihood-based approach is used to represent the data uncertainty due to sparse point data and interval data [3]. The methodology can handle mixed data, e.g., both point data and interval data for the same variable, and fit a non-parametric PDF thereby avoiding the assumption of a distribution type. This PDF can be easily used within uncertainty propagation either using sampling methods such as Monte Carlo simulations or analytical methods such as the First Order Reliability Method [4]. Note that this approach can also handle correlated variables by using a non-parametric joint PDF and the corresponding joint likelihood, given sparse or interval data for multiple variables.

5.2. Model Errors

Two types of model errors are considered: numerical solution error and model form error [5]. Let G(x) represent the true function value that we are interested in estimating, where x represents a set of point-valued inputs. The following equation can be written as following:

$$G = g_{raw} + \epsilon_{num} + \epsilon_{mf} \quad (2)$$

where g_{raw} is the function value as evaluated by a computational code, ϵ_{num} is the numerical solution error, and ϵ_{mf} is the model form error. Two types of numerical solution errors are considered in this paper: discretization error and error due to surrogate model use.

5.2.1. Discretization Error

Discretization error is a critical component of numerical solution errors; this arises due to the solution of a continuum problem through finite element or finite difference solution methods that discretize the continuous domain. In this paper, we assume multidimensional mesh refinement, where the mesh can be refined in more than one direction, and possible cases where meshes do not match topologically [6]. A GP model is used for discretization error estimation, as an enhancement of the traditional Richardson extrapolation to alleviate its drawbacks such as requirement of monotonic convergence, and being in the asymptotic convergence region of the mesh size [7].

5.2.2. Surrogate Model Error

Use of the GP surrogate model introduces surrogate model error. In order to estimate this error as a function of the design variables within the optimization, we construct a GP model for the surrogate model error, denoted as ϵ_{se} , which is also a Gaussian variable for each x. The constraint function can then be written as following:

$$G(x) = g_h(x) + \epsilon_{se}(x) + \epsilon_{mf}(x) \quad (3)$$

Note that at each x, the numerical solution errors of the constraint function, is a Gaussian variable, thereby resulting in a distribution of the model prediction, instead of a point value.

5.2.3. Model Form Error

We assume that the model form error also varies as the inputs to the model change. Assume that a set of observation data from physical experiments is available at a set of q design inputs, O_{set} = $\{X^T_{obs1}, ... X^T_{obsq}\}$. The following equation can be written as following [5]:

$$G(x) = g_{obs}(x) + \epsilon_{exp} = g_h(x) + \epsilon_{se}(x) \quad (4)$$

where g_{obs} is the observed experimental value, and ϵ_{exp} is the experimental error. We assume that the experimental error is the same at all design inputs, although this assumption can be changed easily if needed. The model form error can be evaluated as following:

$$\epsilon_{mf}(x) = g_{obs}(x) - [g_h(x) + \epsilon_{se}(x)] + \epsilon_{exp} \quad (5)$$

It is evaluated at each of the values O_{set}, where at each x, ϵ_{mf} (x) is a Gaussian variable, assuming ϵexp is Gaussian. A GP model, $g_{mf}(x)$ is fit with the training data $\{O_{set}; \epsilon_{mf}(O_{set})\}$ to yield the model form error as a function of the input. The expression for G(x) can finally be written as following:

$$g_{true}(x) = g_h(x) + g_{se}(x) + g_{mf}(x) \quad (6)$$

Fig. 6 summarizes the discussion on quantifying various model errors.

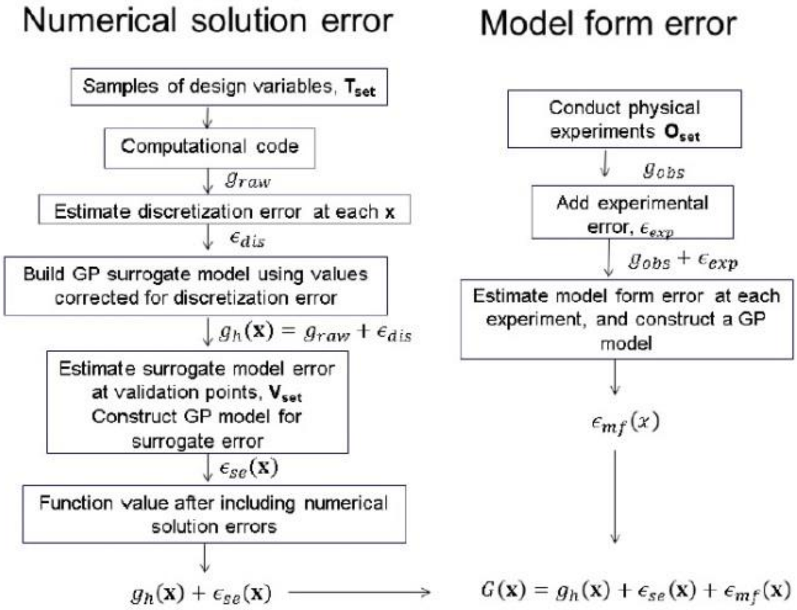


Figure 1: Summary of the model error quantification method

5.3. Stochastic Model Prediction in Optimization

The model errors discussed in this section are stochastic quantities, thereby resulting in a stochastic model prediction for constraints and objectives at a given design input. These stochastic model predictions are now used in the optimization formulation, and must be appropriately used to yield constraint failure probabilities and the objective function. Figures 2 and 3 illustrate the methods used in this paper. As shown in Figure 2, in the traditional method, each random sample is propagated through the objective function model, resulting in a single point value of f; these propagated output samples are then used to compute the mean and standard deviation of f. In other words, the only source of uncertainty in such a case arises because of the input. In the proposed model error treatment on the other hand, each random sample of the input results in a Gaussian distribution of the corresponding prediction. If the optimization requires minimization of mean objective value, μ_f = mean(μ_{fi}), i = {1, ..., n}, or the worst case scenario, μ_f= max(μ_{fi}), i = {1, ..., n} can be used.

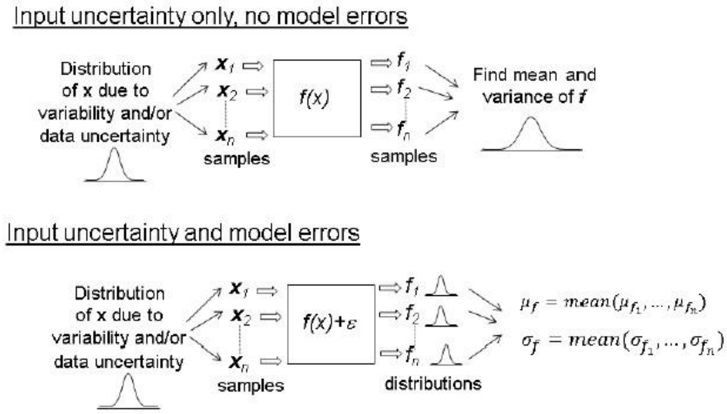
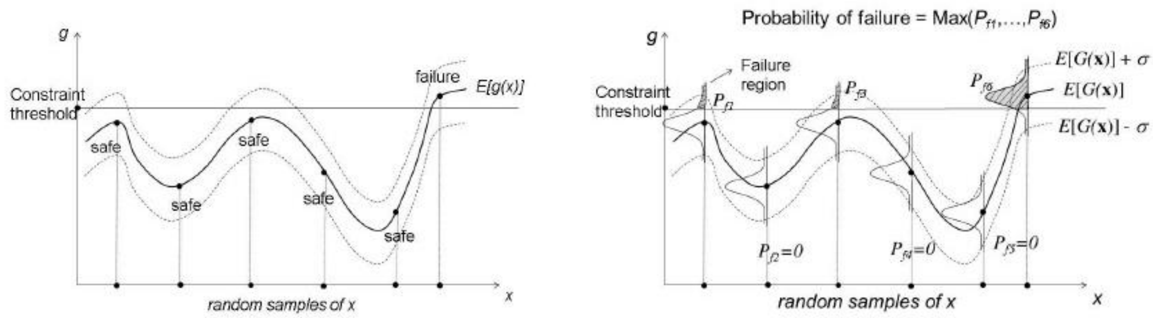


Figure 2: Objective function evaluation with stochastic model predictions

For constraint failure evaluation, we exploit the fact that the corrected model value after correction for model form and discretization errors is a Gaussian variable. At each random sample of the input x_i, we find the corresponding probability of failure given that the true model prediction, $g_{true}(x)$ is a Gaussian distribution, as shown in Figure 3(b), as P_{fi} = Prob($G(x_i) \leq 0$), which is a simple evaluation of a Gaussian cumulative density function value. The probability of failure used in the optimization can be computed as the worst case scenario, P_{fi} = max[Prob($G(x_i) \leq 0$)], $i = \{1, \ldots, n\}$, where n is the number of input samples. An alternative simpler approach is to only consider the mean values of the GP error models. The corrected function value can then be written as follows.

$$G(\mathrm{x}) = E[g_h(\mathrm{x})] + E[g_{se}(\mathrm{x})] + E[g_{mfe}(\mathrm{x})] \tag{7}$$

By ignoring the variance of the model errors, we underestimate the uncertainties in the model predictions used in the optimization. This may lead to erroneously small probabilities of failure. However, consideration of only the mean model errors can be computationally faster.



(a) Using only mean model predictions (b) Using stochastic model predictions
Figure 3: Constraint failure probability evaluation with considering model errors

6. Design Optimization Formulation under Uncertainty

The optimization problem is presented and the steps involved are outlined as follows:

1. Choose the training data, T_{set}, verification data, V_{set}, and observation data, O_{set} as using a suitable design of experiments.
2. Construct the GP models for the function value after correction for discretization error, surrogate model error, and model form error as functions of the design inputs.
3. Perform uncertainty quantification for all epistemic variables using the likelihood approach. The likelihood based PDF is used to determine the variance of the epistemic variable, which is then used in the optimization problem, and the mean of the epistemic variable is designed. The aleatory variables are described through probability distributions.
4. Solve the following optimization problem as formulation 1 shows, where F and G are the models corrected for numerical solution error and model form error.
5. The bounds on the means of the epistemic variables can also be posed as constraints

$$lb_X + k\sigma_X \leq \mu_X \leq ub_X - k\sigma_X$$

where σ_X for the epistemic variable can be estimated from the likelihood-based PDF constructed.
The above probabilistic formulation integrates both aleatory and epistemic uncertainties within the optimization framework.

7. Numerical Example

A cantilevered wing with a NACA 0012 airfoil is adopted [8]. Figure 4 shows the top view of the wing configuration. We use ANSYS to perform the fluid-structure interaction analysis of the wing. Two variations of a wing design problem are considered with different assumptions. The first problem has a single design variable, and is used to explain all the elements of the proposed approach in detail. The second problem has five design variables, and is used to examine the performance of the method as the problem size increases. Only the first problem will be illustrated here. The second problem will be addressed in the paper.

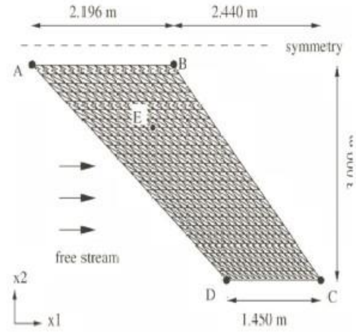


Figure 4: Two-dimensional view of the wing configuration

The design optimization problem is to maximize the lift generated by the wing subject to a stress constraint, and the design variable is the backsweep angle. The variability of the manufacturing process is characterized as epistemic described by sparse and interval data. The constraint and objective function values are evaluated by ANSYS, and surrogate models are built to be used in the optimization, thereby introducing discretization error and surrogate model error. In addition, model form error is also considered. All of the above errors are considered as varying functions of the backsweep angle. Five sub-problems are considered within this example to illustrate the proposed developments: (1) Case 1: deterministic optimization problem, (2) Case 2: Optimization under uncertainty with variability alone, and no data or model error considerations, (3) Case 3: Optimization under uncertainty with variability and data uncertainty alone, but no model error considerations, and (4) Case 4: optimization with variability, data uncertainty and mean values of model errors, and (5) Case 5: optimization with variability, data uncertainty and model errors, using the entire distributions of the model errors.

Table 1: Optimization results summary

Problem	μ_{bw}^*	μ_L^*
Case 1	0.2794	1927.56
Case 2	0.4097	1796.52
Case 3	0.3643	1814.05
Case 4	0.3326	1919.87
Case 5	0.3603	1929.16

Table 2: Optimization results summary

	Probability of failure		Mean objective	
Design Case	ignoring model errors	including distribution of model errors	ignoring model errors	including distribution of model errors
Case 1	0	0.0422	1889.8	1908.8
Case 2	0	0.0487	1796.51	1907.98
Case 3	0	$4.67 \times 10-5$	1912.15	1929.16
Case 4	-	$5.65 \times 10-4$	-	1908.83
Case 5	-	$7.8 \times 10-4$	-	1929.16

Tables 1 and 2 summarize the results of the above problems, and the following observations are made.

1. Case 1 versus Cases 2 and 3: When data uncertainty is considered in the problem, the design objective worsened (Table 1), which is generally expected when uncertainty in the design variables is considered in the problem.
2. Cases 1, 2 and 3 versus Case 4: When model errors are corrected, the worsening of the objective due to data uncertainty is seemingly canceled (in this problem) by the model error corrections applied.
3. Case 4 versus Case 5: The probability of failure when only mean model errors (Case 4) was observed to be zero, while that when the entire distribution of model error is used (Case 5) was observed to be 7.835×10^{-4}. This is to be expected, as shown in Figure 3.
4. In Table 2, we conduct a scenario analysis to understand the impact of model errors. In row 1, we use the optimum solution of the deterministic problem, and assume data uncertainty at this solution. Say the system was designed at the deterministic solution of backsweep angle of 0.2794 rad, and there was variability in the actual wing manufactured. How does this impact the probability of failure and the objective function? Furthermore, how does the model error consideration impact this analysis? For row 1, it is seen that the even when mean model errors are considered, the probability of failure is zero. However, if the entire distribution of the model errors is considered, there is non-zero probability of failure. This means that a complete consideration of model errors and the uncertainties in their estimation can increase the probability of failure.

For the objective function, we do a similar study – consider data uncertainty about the deterministic optimum solution, and study how the objective function changes. If no model errors are considered, the objective function performance deteriorates when compared to the one predicted by the deterministic optimization. In this problem, the consideration of distribution of model errors does not cause a profound difference compared to using mean values, since the objective function involves the mean value as against probability of failure.

8. Conclusion

In this paper, we proposed approaches to systematically include aleatory and epistemic uncertainties in the design variables, non-design variables, and model errors into the optimization problem. We employ a likelihood-based approach to handle variables described by sparse and interval data. Numerical solution errors and model form errors are determined prior to the optimization as a function of design variables, and are used in the optimization formulation. The proposed method is illustrated for a 3-D wing design problem which involves a fluid-structure interaction analysis. Two design problems involving wing design are used to illustrate the proposed methods with cases that gradually increase complexity of the problem. It is observed that model errors and the associated uncertainties can have a significant impact on the optimum solution and the constraint reliabilities obtained. In this work, we have used GP models to represent the behavior of model errors with changing design inputs. The GP framework offers unique advantages in the context of the proposed work; it can not only be integrated into the probabilistic framework easily, but also can provide quantitative estimates of the uncertainties due to various sources in the error models. We assumed an additive error model, and the errors are sequentially corrected. In this paper, we constructed GP models for the errors, thereby resulting in Gaussian distributions for the error estimates. Alternate representations of the error models that do not result in a Gaussian distribution might be of interest for some problems.

9. Acknowledgments

This study was supported by funds from NASA Langley Research Center under Cooperative Agreement No. NNX08AF56A1 (Technical Monitor: Lawrence Green). The support is gratefully acknowledged. The computational resources of Vanderbilt University's ACCRE have been used in this paper.

10. References

[1] Rasmussen, C. E. and Williams, C. K. I., *Gaussian processes for machine learning*, Springer, 2006.

[2] Valdebenito, M. and Schuller, G., "A survey on approaches for reliability-based optimization," *Structural and Multidisciplinary Optimization*, 42, 645–663, 2010.

[3] Sankararaman, S. and Mahadevan, S., "Likelihood-based representation of epistemic uncertainty due to sparse point data and/or interval data," *Reliability Engineering and System Safety*, 96(7), 814 – 824, 2011.

[4] Haldar, A. and Mahadevan, S., *Probability, Reliability, and Statistical Methods in Engineering Design*, John Wiley and Sons, Inc, 2000.

[5] Rebba, R., Mahadevan, S., and Huang, S., "Validation and error estimation of computational models," *Reliability Engineering and Systems Safety*, 91, 1390–1397, 2006.

[6] Rangavajhala, S., Sura, V., Hombal, V. K., and Mahadevan, S., "Discretization error estimation in multidisciplinary simulations," *AIAA Journal*, 49(12), 2673–2683, 2011.

[7] Richardson, L. F., “The approximate arithmetical solution by finite differences of physical problems involving differential equations, with an application of the stresses in amasonry dam,” *Philosophical transactions of the Royal society of London*, 210, 307 – 357, 1911.

[8] Maute, K., Nikbay, M., and Farhat, C., “Coupled analytical sensitivity analysis and optimization of three-dimensional nonlinear aeroelastic systems,” *AIAA Journal*, 39(11), 2051–2061, 2001.