INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN ENGINEERING
Int. J. Numer. Meth. Engng (2014)
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/nme.4761

Multifidelity approaches for optimization under uncertainty

Leo W. T. Ng and Karen E. Willcox*,†

Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

SUMMARY

It is important to design robust and reliable systems by accounting for uncertainty and variability in the design process. However, performing optimization in this setting can be computationally expensive, requiring many evaluations of the numerical model to compute statistics of the system performance at every optimization iteration. This paper proposes a multifidelity approach to optimization under uncertainty that makes use of inexpensive, low-fidelity models to provide approximate information about the expensive, high-fidelity model. The multifidelity estimator is developed based on the control variate method to reduce the computational cost of achieving a specified mean square error in the statistic estimate. The method optimally allocates the computational load between the two models based on their relative evaluation cost and the strength of the correlation between them. This paper also develops an information reuse estimator that exploits the autocorrelation structure of the high-fidelity model in the design space to reduce the cost of repeatedly estimating statistics during the course of optimization. Finally, a combined estimator incorporates the features of both the multifidelity estimator and the information reuse estimator. The methods demonstrate 90% computational savings in an acoustic horn robust optimization example and practical design turnaround time in a robust wing optimization problem. Copyright © 2014 John Wiley & Sons, Ltd.

Received 20 September 2013; Revised 28 May 2014; Accepted 16 July 2014

KEY WORDS: multifidelity; design under uncertainty; model reduction; optimization; probabilistic methods; stochastic problems

1. INTRODUCTION

Uncertainties are present in many engineering applications, and it is important to account for their effects during the design process to achieve robust and reliable systems. However, this is often too computationally expensive in a formal optimization setting. A key challenge is to efficiently propagate uncertainties from the random inputs of a numerical model to its outputs. However, the trend toward higher fidelity models makes it increasingly expensive to perform uncertainty propagation. This is especially challenging when uncertainty propagation is nested within another layer of analysis such as optimization.

We propose to lower the computational cost of uncertainty propagation on high-fidelity models by leveraging inexpensive low-fidelity models to provide useful information about the high-fidelity model outputs. These low-fidelity models can be simpler versions or approximations of the underlying physics, less accurate numerical solutions of the governing equations, reduced-order models, and so on. Such multifidelity methods have been employed by the (deterministic) optimization community to reduce the cost of minimizing expensive objective functions. The idea is that the low-fidelity model can inexpensively supply the trend toward the general vicinity of the high-fidelity model optimum with only occasional recourse to the high-fidelity model as a correction or adjustment to ensure convergence [1–3]. For uncertainty propagation, we similarly take advantage of the trend of

*Correspondence to: Karen E. Willcox, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
†E-mail: kwillcox@mit.edu

the low-fidelity model by making use of the linear correlation between the low-fidelity model outputs and the high-fidelity model outputs to make adjustments to the estimators of the statistics of the high-fidelity model outputs.

Even with multifidelity models, optimization under uncertainty can still be prohibitively expensive because of the nested structure, where the outer loop adjusts the design variables and the inner loop samples the uncertain model parameters. We further reduce computational cost by leveraging the model autocorrelation in the design space. The idea is that the samples of model outputs at one design point provide useful information (based on linear correlation) about the samples of model outputs at a different design point.

There have been recent efforts to bring multifidelity concepts to various topics of uncertainty quantification. In rare event simulation (such as reliability analysis), an approach that is conceptually similar to multifidelity optimization is to evaluate the low-fidelity model extensively to narrow down the location of the limit-state boundary and correct the estimate using the high-fidelity model [4, 5]. Non-intrusive polynomial approximations (e.g., polynomial chaos expansion, stochastic collocation) of the high-fidelity model outputs can be constructed by combining an approximation of the low-fidelity model on a fine sparse grid with an approximation of the correction on a coarse sparse grid [6]. There are also multifidelity sampling approaches for uncertainty propagation based on Bayesian regression between the high-fidelity model outputs and low-fidelity model outputs [7, 8].

Our approach is based on the control variate technique for Monte Carlo (MC) simulation [9, 10] to estimate statistics of the high-fidelity model outputs (e.g., mean and variance) given the distributions of the uncertain model parameters as inputs. It resembles multilevel or multigrid MC methods for solving stochastic differential equations in which the estimators from coarse discretizations are used as the 'control' to reduce the sampling variance of the estimator from the fine discretization [11–13]. We consider more general engineering models, possibly as a black-box, so the complexity analysis for the multilevel methods [11] does not necessarily apply. In addition, we may not have the flexibility to easily adjust the accuracy and computational cost of the low-fidelity model as we can for stochastic differential equations using a discretization parameter. Instead, we focus on minimizing practical computation times with the assumption that the evaluation of the (black-box) models takes up most of the computational effort. The reduced-basis control-variate MC method pre-computes and stores a reduced-basis of the control variates offline to allow for efficient solution of parameterized stochastic differential equations online [14, 15]. In our approach, we do not separate the calculation of the samples into an offline and online stage. Instead, both the low-fidelity model and high-fidelity model are evaluated online with the computational resources split between the two to minimize the sampling variance of the estimator. Finally, in contrast to the StackMC method that constructs a low-fidelity model from the samples of high-fidelity model outputs using data-fit techniques [16], we develop a mathematical framework that admits a broad range of low-fidelity models, including reduced-order models and models that may arise from different modeling assumptions.

In this paper, we present our multifidelity estimator, information reuse estimator, and combined estimator for optimization under uncertainty. In Section 2, we describe the optimization under uncertainty problem and set up the general framework. In Section 3, we develop the multifidelity estimator to take advantage of multifidelity models for uncertainty propagation. In Section 4, we develop the information reuse estimator to take advantage of model autocorrelation during optimization. We discuss how to combine the two estimators in Section 5. The choices of optimization algorithms are discussed in Section 6. Numerical results are presented in Section 7 for two example problems to demonstrate the effectiveness of the estimators. Finally, we conclude in Section 8.

2. PROBLEM FORMULATION

The engineering system is described by a high-fidelity model, $M_{\text{high}}(\mathbf{x}, \mathbf{u})$, and a low-fidelity model, $M_{\text{low}}(\mathbf{x}, \mathbf{u})$, with two input vectors: design variables $\mathbf{x}$ and model parameters $\mathbf{u}$. We consider the case where there are uncertainties in the model parameters represented by some probability distribution. Therefore, $\mathbf{u}$ is a realization of the random vector $\mathbf{U}(\omega), \omega \in \Omega$ where Ω is the sample space, and

DOI: 10.1002/nme

the high-fidelity model output is a random variable defined as $A(\mathbf{x}, \omega) = M_{\text{high}}(\mathbf{x}, \mathbf{U}(\omega))$. In this setting, a statistic of the high-fidelity model output, denoted as $s_A(\mathbf{x})$, such as the mean, the variance, and so on, can be used to describe the system performance.

We are interested in the design under uncertainty of the engineering system as described by the high-fidelity model. Therefore, we consider the following general optimization problem:[‡]

$$\begin{aligned} \mathbf{x}^* = \arg\min_{\mathbf{x}} \; & f(\mathbf{x}, s_A(\mathbf{x})) \\ \text{s.t. } & g(\mathbf{x}, s_A(\mathbf{x})) \leqslant 0 \\ & h(\mathbf{x}, s_A(\mathbf{x})) = 0, \end{aligned} \tag{1}$$

where the objective and constraint functions f, g, and h may depend on the statistic $s_A(\mathbf{x})$. For example, in robust design, the objective function may be a linear combination of the mean and the standard deviation of the high-fidelity model output. Because we typically cannot compute $s_A(\mathbf{x})$ exactly, we approximate it with its estimator $\hat{s}_A(\mathbf{x})$, and so the objective and constraint functions are themselves also estimators: $\hat{f}(\mathbf{x}) = f\left(\mathbf{x}, \hat{\mathbf{s}}_A(\mathbf{x})\right), \hat{g}(\mathbf{x}) = g\left(\mathbf{x}, \hat{\mathbf{s}}_A(\mathbf{x})\right)$, and $\hat{h}(\mathbf{x}) = h\left(\mathbf{x}, \hat{\mathbf{s}}_A(\mathbf{x})\right)$.

To solve (1), we compute the estimator $\hat{s}_A(\mathbf{x}_k)$ at every step in the design space taken toward the optimum, $\mathbf{x}_k, k = 0, 1, 2, \ldots$. We focus on MC simulation because it is non-intrusive, parallelizable, broadly applicable (does not rely on smoothness of the underlying problem), and independent of the dimension of the uncertain model parameters. Thus, we have a nested setup where the design variables are adjusted by the optimization algorithm in the outer loop, and the uncertain model parameters are sampled by the MC simulation in the inner loop.

For a particular vector of design variables $\mathbf{x}_k$, consider estimating $s_A = \mathbb{E}[A(\omega)]$. We generate n independent and identically distributed (i.i.d.) samples $a_i = M_{\text{high}}(\mathbf{x}_k, \mathbf{u}_i), i = 1, 2, \ldots, n$ from the distribution of $A(\omega)$ by evaluating the high-fidelity model at n i.i.d. samples of the uncertain model parameters $\mathbf{u}_i, i, = 1, 2, \ldots, n$ drawn from the distribution of $\mathbf{U}(\omega)$. The regular MC estimator of s_A, denoted as $\bar{a}_n$, is

$$\bar{a}_n = \frac{1}{n} \sum_{i=1}^{n} a_i \tag{2}$$

and its mean square error (MSE) is given by the estimator variance

$$\text{MSE}\left[\bar{a}_n\right] = \text{Var}\left[\bar{a}_n\right] = \frac{1}{n^2} \text{Var}\left[\sum_{i=1}^{n} a_i\right] = \frac{\sigma_A^2}{n}, \tag{3}$$

where $\sigma_A^2 = \text{Var}[A(\omega)]$ is the variance of $A(\omega)$. Once we obtain the estimator with an acceptably low estimator variance, we evaluate $\hat{f}, \hat{g}$, and $\hat{h}$ and return the estimated objective and constraint functions to the optimization routine to determine the next vector of design variables, $\mathbf{x}_{k+1}$.

This approach can be computationally expensive because the high-fidelity model is evaluated at every sample of the uncertain model parameters and at every vector of design variables specified by the optimization routine. This paper focuses on lowering the computational cost of the estimator by making use of approximate information. The control variate method [9, 10] is a technique to reduce the estimator variance by making use of the correlation between the random variables $A(\omega)$ and an auxiliary random variable. We modify the control variate method to leverage the output of an inexpensive, low-fidelity model and to leverage the autocorrelation of the high-fidelity model in the design space, resulting in our multifidelity estimator and our information reuse estimator, respectively.

[‡]If the objective and constraints are functions of more than one statistic, we let the statistic s_A become a vector of statistics $\mathbf{s}_A$.

DOI: 10.1002/nme

3. MULTIFIDELITY ESTIMATOR

In this section, we develop a method to compute the estimator $\hat{s}_A$ of the statistic s_A, where the random variable $A(\omega)$ is the high-fidelity model output $M_{\text{high}(\mathbf{x}_k,\mathbf{U}(\omega))}$ at some fixed values of the design variables $\mathbf{x}_k$. We introduce the multifidelity estimator based on the control variate method to make use of the low-fidelity model output $M_{\text{low}(\mathbf{x}_k,\mathbf{U}(\omega))}$ as the auxiliary random variable $B(\omega)$.

3.1. Approach

For simplicity of explanation, consider the estimation of the mean of the high-fidelity model, that is, $s_A = \mathbb{E}[A(\omega)]$. Let the random variable of interest be $A(\omega) = M_{\text{high}}(\mathbf{x}_k, \mathbf{U}(\omega))$ and the auxiliary random variable be $B(\omega) = M_{\text{low}}(\mathbf{x}_k, \mathbf{U}(\omega))$. Given the i.i.d. samples $\mathbf{u}_i, i = 1, 2, 3, \ldots$ drawn from the distribution of the uncertain model parameters $\mathbf{U}(\omega)$, we evaluate both the high-fidelity model and the low-fidelity model to generate samples of the random variables $a_i = M_{\text{high}}(\mathbf{x}_k, \mathbf{u}_i)$ and $b_i = M_{\text{low}}(\mathbf{x}_k, \mathbf{u}_i), i = 1, 2, 3, \ldots$. The classical control variate estimator is defined as

$$\hat{s}_A = \bar{a}_n + \alpha\left(s_B - \bar{b}_n\right) = \frac{1}{n}\sum_{i=1}^{n} a_i + \alpha\left(s_B - \frac{1}{n}\sum_{i=1}^{n} b_i\right),$$

where $\alpha \in \mathbb{R}$ is the control parameter and the statistic $s_B = \mathbb{E}[B(\omega)]$.

Because, in general, we would not know exactly the statistic of the low-fidelity model output, we approximate s_B by $\bar{b}_m = \frac{1}{m}\sum_{i=1}^{m} b_i$ with $m > n$. That is, we make use of the relatively low computational cost of the low-fidelity model to calculate a more accurate estimate of s_B than the estimate with only n samples. Some other extensions to the control variate method [17, 18] generate an independent simulation with m samples to compute $\bar{b}_m$. In our case, we simply require $m - n$ additional samples of b_i beyond the n samples already available. Therefore, the multifidelity estimator of s_A is defined as

$$\hat{s}_{A,p} = \bar{a}_n + \alpha\left(\bar{b}_m - \bar{b}_n\right), \qquad m > n, \tag{4}$$

for some control parameter $\alpha \in \mathbb{R}$ and computational effort p to be defined in the succeeding text. The rationale for this formulation is to make an adjustment to the regular MC estimator $\bar{a}_n$ by the difference $\bar{b}_m - \bar{b}_n$, which may be interpreted as the error of the estimator $\bar{b}_n$ with respect to the more accurate estimator $\bar{b}_m$. Because $\mathbb{E}\left[\bar{b}_m - \bar{b}_n\right] = 0$, the second term has no effect in expectation, but we can take advantage of its effect in variance. The variance of the multifidelity estimator can be derived as

$$\begin{aligned}
\text{Var}\left[\hat{s}_{A,p}\right] &= \text{Var}\left[\bar{a}_n\right] + \alpha^2\,\text{Var}\left[\bar{b}_m\right] + \alpha^2\,\text{Var}\left[\bar{b}_n\right] \\
&\quad + 2\alpha\,\text{Cov}\left[\bar{a}_n, \bar{b}_m\right] - 2\alpha\,\text{Cov}\left[\bar{a}_n, \bar{b}_n\right] - 2\alpha^2\,\text{Cov}\left[\bar{b}_m, \bar{b}_n\right] \\
&= \frac{\sigma_A^2}{n} + \alpha^2\frac{\sigma_B^2}{m} + \alpha^2\frac{\sigma_B^2}{n} \\
&\quad + 2\alpha\frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\text{Cov}\left[a_i, b_j\right] - 2\alpha\frac{\rho_{AB}\sigma_A\sigma_B}{n} - 2\alpha^2\frac{1}{nm}\sum_{i=1}^{m}\sum_{j=1}^{n}\text{Cov}\left[b_i, b_j\right] \\
&= \frac{\sigma_A^2}{n} + \alpha^2\frac{\sigma_B^2}{m} + \alpha^2\frac{\sigma_B^2}{n} \\
&\quad + 2\alpha\frac{1}{nm}\sum_{i=1}^{n}\text{Cov}\left[a_i, b_i\right] - 2\alpha\frac{\rho_{AB}\sigma_A\sigma_B}{n} - 2\alpha^2\frac{1}{nm}\sum_{j=1}^{n}\text{Cov}\left[b_j, b_j\right] \\
&= \frac{1}{n}\left(\sigma_A^2 + \alpha^2\sigma_B^2 - 2\alpha\rho_{AB}\sigma_A\sigma_B\right) - \frac{1}{m}\left(\alpha^2\sigma_B^2 - 2\alpha\rho_{AB}\sigma_A\sigma_B\right),
\end{aligned}$$

DOI: 10.1002/nme

where $\sigma_B^2 = \mathrm{Var}[B(\omega)]$ and $\rho_{AB} = \mathrm{Corr}[A(\omega), B(\omega)]$ are the variance of $B(\omega)$ and the correlation coefficient between $A(\omega)$ and $B(\omega)$, respectively.

Computing the regular MC estimator from (2) requires n evaluations of the high-fidelity model. However, computing the multifidelity estimator $\hat{s}_{A,p}$ from (4) requires n evaluations of the high-fidelity model and m evaluations of the low-fidelity model. Therefore, to benchmark against the regular MC estimator, we define a unit of computational effort p as the 'equivalent number of high-fidelity model evaluations':

$$p = n + \frac{m}{w} = n\left(1 + \frac{r}{w}\right), \tag{5}$$

where w is the ratio of the average computation time per high-fidelity model evaluation to the average computation time per low-fidelity model evaluation,[§] and $r = m/n > 1$ is the ratio of the number of low-fidelity model evaluations to the number of high-fidelity model evaluations. Thus, for a fixed computational budget p, r is a parameter to allocate the computational resources between high-fidelity model evaluations and low-fidelity model evaluations. We rewrite the multifidelity estimator variance in terms of p, r, and w:

$$\mathrm{Var}\left[\hat{s}_{A,p}\right] = \frac{1}{p}\left(1 + \frac{r}{w}\right)\left[\sigma_A^2 + \left(1 - \frac{1}{r}\right)\left(\alpha^2\sigma_B^2 - 2\alpha\rho_{AB}\sigma_A\sigma_B\right)\right]$$

Given the computational budget p, we minimize $\mathrm{Var}\left[\hat{s}_{A,p}\right]$ in terms of both α and r. Because $1/p$ is a multiplicative factor in $\mathrm{Var}\left[\hat{s}_{A,p}\right]$, the optimal value of α and r, denoted as α^* and r^*, do not depend on p. The result is

$$\alpha^* = \rho_{AB}\frac{\sigma_A}{\sigma_B}, \qquad r^* = \sqrt{\frac{w\rho_{AB}^2}{1 - \rho_{AB}^2}}$$

and the optimal $\mathrm{Var}\left[\hat{s}_{A,p}\right]$, denoted as $\mathrm{Var}\left[\hat{s}^*_{A,p}\right]$, is

$$\mathrm{MSE}\left[\hat{s}^*_{A,p}\right] = \mathrm{Var}\left[\hat{s}^*_{A,p}\right] = \left(1 + \frac{r^*}{w}\right)\left[1 - \left(1 - \frac{1}{r^*}\right)\rho_{AB}^2\right]\frac{\sigma_A^2}{p}. \tag{6}$$

It can be seen that r^* allocates a greater proportion of the computational budget p to low-fidelity model evaluations when the low-fidelity model is inexpensive (i.e., w is large) and 'accurate', where accuracy is in terms of the correlation (i.e., ρ_{AB} is close to 1). In the extreme cases, as $w \to \infty$, that is, the low-fidelity model is almost free, then $r^* \to \infty$ and $\mathrm{Var}\left[\hat{s}^*_{A,p}\right] \to \left(1 - \rho_{AB}^2\right)\frac{\sigma_A^2}{p}$, which is the classical control variate solution where we know exactly the statistic of the low-fidelity model output s_B. The effect of replacing s_B with the estimate $\bar{b}_m$ is to reduce ρ_{AB}^2 by the factor $\left(1 - \frac{1}{r^*}\right)$. On the other hand, as $\rho_{AB} \to 1$, that is, the low-fidelity model is almost perfect, then $\mathrm{Var}\left[\hat{s}^*_{A,p}\right] \to \frac{1}{w}\frac{\sigma_A^2}{p}$. Therefore, a perfectly correlated low-fidelity model is not sufficient for variance reduction over the regular MC estimator using the same computational budget p; it must also be cheaper to evaluate than the high-fidelity model, that is, $w > 1$.

If the values of w and ρ_{AB} are such that $r^* \leqslant 1$, the correlation is not high enough and the low-fidelity model is not cheap enough for the multifidelity estimator to be worthwhile. In other words, if

$$\rho_{AB}^2 > \frac{1}{1 + w}$$

is not satisfied, it is better to switch back to the regular MC estimator.

[§]If w is not known, we simply time a few evaluations of the high-fidelity model and the low-fidelity model and take the ratio.

DOI: 10.1002/nme

3.2. Implementation

In practice, we cannot compute the multifidelity estimator $\hat{s}_{A,p}$ and its estimator variance Var $\left[\hat{s}_{A,p}\right]$ directly from (4) and (6), respectively, because σ_A, σ_B, and ρ_{AB} are typically unknown. Therefore, as in the classical control variate method, we replace these quantities with their sample estimates. At a particular computational budget p, we have n samples of $A(\omega)$ and $m > n$ samples of $B(\omega)$. Therefore, we make use of the n common samples $\{a_i, b_i\}_{i=1}^n$ to calculate

$$\hat{\alpha} = \frac{\sum_{i=1}^n (a_i - \bar{a}_n)(b_i - \bar{b}_n)}{\sum_{i=1}^n (b_i - \bar{b}_n)^2}, \tag{7a}$$

$$\hat{r} = \sqrt{\frac{w\hat{\rho}_{AB}^2}{1 - \hat{\rho}_{AB}^2}}, \tag{7b}$$

$$\hat{\rho}_{AB}^2 = \frac{\left[\sum_{i=1}^n (a_i - \bar{a}_n)(b_i - \bar{b}_n)\right]^2}{\left[\sum_{i=1}^n (a_i - \bar{a}_n)^2\right]\left[\sum_{i=1}^n (b_i - \bar{b}_n)^2\right]}, \tag{7c}$$

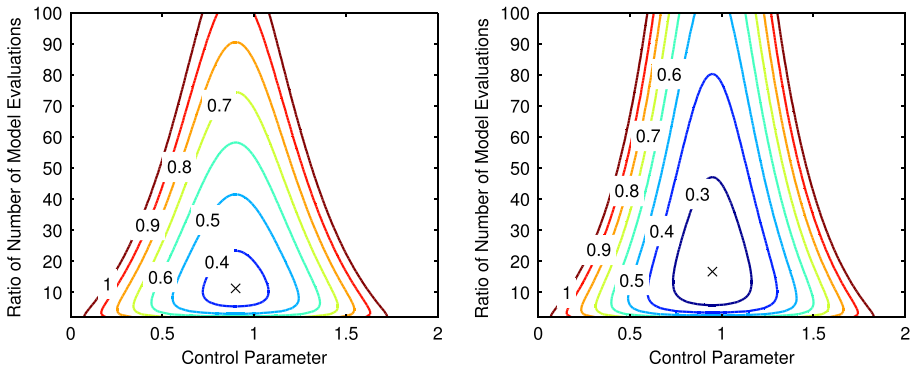
$$\hat{\sigma}_A^2 = \frac{\sum_{i=1}^n (a_i - \bar{a}_n)^2}{n-1}. \tag{7d}$$

Because we have replaced the exact parameters α^* and r^* with their sample estimates $\hat{\alpha}$ and $\hat{r}$, it is important to check if the multifidelity estimator still has a lower estimator variance than the regular MC estimator. We assess the robustness of the multifidelity estimator to the errors in $\hat{\alpha}$ and $\hat{r}$ relative to α^* and r^* by plotting the ratio

$$\frac{\text{Var}\left[\hat{s}_{A,p}\right]}{\text{Var}\left[\bar{a}_p\right]} = \left(1 + \frac{r}{w}\right)\left[1 + \left(1 - \frac{1}{r}\right)\left(\alpha^2 \frac{\sigma_B^2}{\sigma_A^2} - 2\alpha\rho_{AB}\frac{\sigma_B}{\sigma_A}\right)\right]$$

as a function of α and r in Figure 1 for some typical values of w and ρ_{AB}. A value less than one indicates the multifidelity estimator variance is lower than the regular MC estimator variance for the same computational effort. It can be seen that small errors in α^* and r^* (location indicated by the cross) is acceptable, although optimal variance reduction will not be achieved.

To compute the multifidelity estimator, we first draw n samples $\{\mathbf{u}_i\}_{i=1}^n$ from the distribution of $\mathbf{U}(\omega)$ and evaluate $M_{\text{high}}(\mathbf{x}_k, \mathbf{u}_i)$ and $M_{\text{low}}(\mathbf{x}_k, \mathbf{u}_i)$ to generate the samples $\{a_i, b_i\}_{i=1}^n$. We calculate the quantities in (7) and obtain $m = \text{round}(n\hat{r})$, allowing us to draw an additional $m - n$



(a) Correlation coefficient $\rho_{AB} = 0.9$ (b) Correlation coefficient $\rho_{AB} = 0.95$

Figure 1. Contour plots of Var $[\hat{s}_{A,p}]$ /Var $[\bar{a}_p]$ as a function of control parameter α and ratio of the number of model evaluations r with $w = 30$ and $\sigma_A/\sigma_B = 1$. The cross indicates the location of (α^*, r^*). (a) correlation coefficient $\rho_{AB} = 0.9$ and (b) correlation coefficient $\rho_{AB} = 0.95$.

DOI: 10.1002/nme

samples of $\mathbf{U}(\omega)$ and to evaluate the low-fidelity model an additional $m - n$ times to obtain the samples $\{b_i\}_{i=1}^{m}$ (including the original n samples). We can then compute the multifidelity estimator from (4) and its MSE from (6). If the MSE is too high, we increase n and repeat. We emphasize that all of the samples used in the method are generated from the same stream of uncertain model parameters, $\mathbf{u}_i, i = 1, 2, 3, \ldots$, to induce the correlation between the high-fidelity model and the low-fidelity model.

4. INFORMATION REUSE ESTIMATOR

In this section, let us assume that we do not have an inexpensive, low-fidelity model. Instead, we consider $M_{\text{high}}(\mathbf{x}_\ell, \mathbf{U}(\omega)), \mathbf{x}_\ell \neq \mathbf{x}_k$, as an approximation to $M_{\text{high}}(\mathbf{x}_k, \mathbf{U}(\omega))$ when $\|\mathbf{x}_k - \mathbf{x}_\ell\|$ is small. During optimization under uncertainty, as described in Section 2, the statistic of the high-fidelity model output is estimated at a sequence of steps in the design space toward the optimum. Therefore, we introduce the information reuse estimator that takes advantage of this setting to reuse the estimator computed at a previously visited design point.

4.1. Model autocorrelation

To motivate our approach, we interpret the high-fidelity model output $M_{\text{high}}(\mathbf{x}, \mathbf{U}(\omega))$ as a random process indexed by the vector of design variables $\mathbf{x}$. Therefore, if the random variable of interest $A(\omega)$ is the high-fidelity model output at a particular vector of design variables and the auxiliary random variable $C(\omega)$ is the high-fidelity model output at another vector of design variables, then the autocorrelation structure of $M_{\text{high}}(\mathbf{x}, \mathbf{U}(\omega))$ provides the correlation between $A(\omega)$ and $C(\omega)$ for the control variate method. Intuitively, if $M_{\text{high}}(\mathbf{x}, \mathbf{u})$ at a realization $\mathbf{u}$ of the uncertain model parameters $\mathbf{U}(\omega)$ is smooth in the $\mathbf{x}$ direction, then a small perturbation $\mathbf{x}+\Delta\mathbf{x}$ produces only a small change in the output. Considering all realizations of $\mathbf{U}(\omega)$, it is likely that $M_{\text{high}}(\mathbf{x} + \Delta\mathbf{x}, \mathbf{U}(\omega))$ is correlated with $M_{\text{high}}(\mathbf{x}, \mathbf{U}(\omega))$.

To make the argument concrete, consider the simpler case of a scalar design variable x. Let $M_{\text{high}}(x, \mathbf{u})$ be twice differentiable in x for all realizations $\mathbf{u}$ of $\mathbf{U}(\omega)$. Applying a second order Taylor expansion in x, the correlation coefficient between $M_{\text{high}}(x + \Delta x, \mathbf{U}(\omega))$ and $M_{\text{high}}(x, \mathbf{U}(\omega))$ is quadratic in Δx for $|\Delta x| \ll 1$, as derived in Appendix A:

$$\begin{aligned} &\text{Corr}\left[M_{\text{high}}(x + \Delta x, \mathbf{U}(\omega)), M_{\text{high}}(x, \mathbf{U}(\omega))\right] \\ &\approx 1 - \frac{1 - \text{Corr}\left[M'_{\text{high}}(x, \mathbf{U}(\omega)), M_{\text{high}}(x, \mathbf{U}(\omega))\right]^2}{2\text{Var}\left[M_{\text{high}}(x, \mathbf{U}(\omega))\right] / \text{Var}\left[M'_{\text{high}}(x, \mathbf{U}(\omega))\right]} \Delta x^2, \end{aligned} \tag{8}$$

where $M'_{\text{high}}(x, \mathbf{u}) = \partial M_{\text{high}}(x, \mathbf{u})/\partial x$. We see that the correlation coefficient approaches one as $\Delta x \to 0$. This motivates the use of the information reuse estimator within an optimization algorithm, where a sequence of potentially small steps in the design space readily provides good candidates for the choice of the auxiliary random variable $C(\omega)$.

4.2. Approach

Let k be the current optimization iteration and let $\{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_k\}$ be the sequence of design points visited by the optimization algorithm. Let the random variable of interest be $A(\omega) = M_{\text{high}}(\mathbf{x}_k, \mathbf{U}(\omega))$, and we again compute an estimator $\hat{s}_{A,p}$ of $s_A = \mathbb{E}[A(\omega)]$ and the estimator variance $\text{Var}\left[\hat{s}_{A,p}\right]$, where p denotes the computational effort. Furthermore, let the auxiliary random variable be $C(\omega) = M_{\text{high}}(\mathbf{x}_\ell, \mathbf{U}(\omega))$ for $\ell < k$. We assume that during optimization iteration ℓ, we have stored the estimator at that iteration and its estimator variance in a database. Therefore, at the

DOI: 10.1002/nme

current optimization iteration k, we have available the value of the estimator $\hat{s}_C$ of $s_C = \mathbb{E}[C(\omega)]$ as well as the value of the estimator variance Var $[\hat{s}_C]$.[¶] Analogous to the multifidelity estimator, we do not know exactly the statistic s_C required by the classical control variate method. Therefore, we approximate s_C by $\hat{s}_C$ to obtain the information reuse estimator of s_A

$$\hat{s}_{A,p} = \bar{a}_n + \gamma\,(\hat{s}_C - \bar{c}_n) \tag{9}$$

for some control parameter $\gamma \in \mathbb{R}$. The variance of the information reuse estimator can be derived as

$$\begin{aligned}\text{Var}\left[\hat{s}_{A,p}\right] &= \text{Var}\,[\bar{a}_n] + \gamma^2\,(\text{Var}\,[\hat{s}_C] + \text{Var}\,[\bar{c}_n]) - 2\gamma\text{Cov}\,[\bar{a}_n, \bar{c}_n] \\ &= \frac{\sigma_A^2}{n} + \gamma^2\left(\text{Var}\,[\hat{s}_C] + \frac{\sigma_C^2}{n}\right) - 2\gamma\frac{\rho_{AC}\sigma_A\sigma_C}{n} \\ &= \frac{1}{n}\left[\sigma_A^2 + \gamma^2\sigma_C^2(1+\eta) - 2\gamma\rho_{AC}\sigma_A\sigma_C\right],\end{aligned}$$

where

$$\eta = \frac{\text{Var}\,[\hat{s}_C]}{\text{Var}\,[\bar{c}_n]} = \frac{\text{Var}\,[\hat{s}_C]}{\sigma_C^2/n}, \qquad \sigma_C^2 = \text{Var}\,[C(\omega)], \qquad \rho_{AC} = \text{Corr}\,[A(\omega), C(\omega)]\,.$$

The derivation assumes that $\hat{s}_C$, the estimator from optimization iteration ℓ, is uncorrelated with $\bar{a}_n$ or $\bar{c}_n$, which are computed at the current optimization iteration k. This can be achieved in practice by ensuring that the set of realizations of $\mathbf{U}(\omega)$ used in optimization iteration k is independent of the set of realizations of $\mathbf{U}(\omega)$ used in optimization iteration ℓ. Otherwise, additional covariance terms, $\text{Cov}\,[\bar{a}_n, \hat{s}_C]$ and $\text{Cov}\,[\hat{s}_C, \bar{c}_n]$, appear in the expression. This is problematic because $\hat{s}_C$ is itself the information reuse estimator at optimization iteration ℓ with its own auxiliary random variable at yet another earlier optimization iteration and so on, resulting in a chain of covariance terms between random variables stretching back to the first optimization iteration. By requiring $\hat{s}_C$ to be independent of $\bar{a}_n$ or $\bar{c}_n$, we break this chain of dependence and simplify the expression.

Computing the information reuse estimator at optimization iteration k using the samples $\mathbf{u}_i, i = 1, 2, 3, \ldots, n$ drawn from the distribution of $\mathbf{U}(\omega)$ requires n high-fidelity model evaluations at $\mathbf{x}_k$ to calculate $\bar{a}_n$ and n high-fidelity model evaluations at $\mathbf{x}_\ell$ to calculate $\bar{c}_n$. Therefore, the computational effort is $p = 2n$. Given the computational budget p, we minimize Var $\left[\hat{s}_{A,p}\right]$ in terms of γ and obtain

$$\gamma^* = \left(\frac{\rho_{AC}}{1+\eta}\right)\frac{\sigma_A}{\sigma_C}$$

and

$$\text{MSE}\left[\hat{s}^*_{A,p}\right] = \text{Var}\left[\hat{s}^*_{A,p}\right] = 2\left(1 - \frac{\rho_{AC}^2}{1+\eta}\right)\frac{\sigma_A^2}{p}. \tag{10}$$

The information reuse estimator variance is low if the correlation between the high-fidelity model output at $\mathbf{x}_k$ and the high-fidelity model output at $\mathbf{x}_\ell$ is high. Based on the discussion in Section 4.1, we select $\ell = \arg\min_{\ell' < k} \|\mathbf{x}_k - \mathbf{x}_{\ell'}\|$. However, similar to the multifidelity estimator, the effect of replacing s_C with the estimate $\hat{s}_C$ is to reduce ρ_{AC}^2 by the factor $1/(1+\eta)$. Thus, the amount of variance reduction is degraded when the parameter η is large, which occurs when Var $[\hat{s}_C]$ is large relative to Var $[\bar{c}_n]$. This also suggests that the benefit of the information reuse estimator is reduced if the desired estimator variance at the current optimization iteration, Var $\left[\hat{s}^*_{A,p}\right]$, is much lower than the estimator variance at the previous optimization iteration, Var $[\hat{s}_C]$.

[¶]We do not indicate the computation effort of the estimator $\hat{s}_C$ because it is not relevant to the current optimization iteration.

DOI: 10.1002/nme

The information reuse estimator variance is not guaranteed to be lower than the regular MC estimator variance at the same computational effort because of the need to evaluate the high-fidelity model twice (at $\mathbf{x}_k$ and at $\mathbf{x}_\ell$) per sample. Therefore, we compare the results of (10) and (3) to decide if it is necessary to switch to the regular MC estimator as a safeguard.

4.3. Implementation

Analogous to the multifidelity estimator, we replace the unknown quantities σ_A, σ_C, and ρ_{AC} with their sample estimates using the available samples $\{a_i, c_i\}_{i=1}^n$ and obtain

$$\hat{\gamma} = \frac{1}{1+\hat{\eta}} \frac{\sum_{i=1}^n (a_i - \bar{a}_n)(c_i - \bar{c}_n)}{\sum_{i=1}^n (c_i - \bar{c}_n)^2}, \tag{11a}$$

$$\hat{\eta} = \frac{\text{Var}\,[\hat{s}_C]\, n(n-1)}{\sum_{i=1}^n (c_i - \bar{c}_n)^2}, \tag{11b}$$

$$\hat{\rho}_{AC}^2 = \frac{\left[\sum_{i=1}^n (a_i - \bar{a}_n)(c_i - \bar{c}_n)\right]^2}{\left[\sum_{i=1}^n (a_i - \bar{a}_n)^2\right]\left[\sum_{i=1}^n (c_i - \bar{c}_n)^2\right]}, \tag{11c}$$

$$\hat{\sigma}_A^2 = \frac{\sum_{i=1}^n (a_i - \bar{a}_n)^2}{n-1}. \tag{11d}$$

The information reuse estimator is intended to be used within an outer optimization loop that steps through a sequence of design points $\{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_k\}$.[‖] At the initial design point $\mathbf{x}_0$, we cannot compute the information reuse estimator and thus we start with the regular MC estimator and save the estimator and the estimator variance in a database. At each optimization iteration, we select $\mathbf{x}_\ell$ and the corresponding $\hat{s}_C$ and Var $[\hat{s}_C]$ from the database to compute the information reuse estimator. We then store the current information reuse estimator and the estimator variance in the database to provide candidates for $\hat{s}_C$ and Var $[\hat{s}_C]$ at subsequent optimization iterations.

The procedure to compute the information reuse estimator at a particular vector of design variables $\mathbf{x}_k$, once $\hat{s}_C$ and Var $[\hat{s}_C]$ have been retrieved from the database, is analogous to that of the multifidelity estimator. However, as a safeguard mechanism, we start with a small number of samples $\{a_i, c_i\}_{i=1}^{n_{\text{init}}}$, which is used to determine whether to (i) continue with the information reuse estimator or to (ii) switch to the regular MC estimator by comparing (10) and (3). The value of n_{init} is chosen by the user. If it is set too large, then it defeats the purpose of the safeguard mechanism. However, n_{init} should be large enough to make a reasonable prediction about whether to continue with the information reuse estimator or switch to the regular MC estimator. If continuing with the information reuse estimator, we then increase the sample size n, generate additional samples to obtain $\{a_i, c_i\}_{i=1}^n$, calculate the quantities in (11), and compute the estimator from (9) and the MSE from (10). If the MSE is too high, we increase n and repeat.

4.4. Correlated estimator errors

Because the results of an optimization iteration are reused at a later optimization iteration, the information reuse estimators computed during the course of optimization are correlated with each other. To illustrate, we derive an expression for the correlation coefficient between $\hat{s}_{A,p}$, the information reuse estimator at optimization iteration k, and $\hat{s}_C$, the information reuse estimator at optimization

[‖]Technically, the information reuse estimator can be used within any outer loop that generates a sequence of design points. However, it is most effective when the design points visited by the outer loop cluster together, as is often the case during optimization when it is near the optimum.

DOI: 10.1002/nme

iteration $\ell < k$. Because $\mathrm{Cov}\left[\bar{a}_n, \hat{s}_C\right] = 0$ and $\mathrm{Cov}\left[\bar{c}_n, \hat{s}_C\right] = 0$, we have

$$\begin{aligned}\mathrm{Cov}\left[\hat{s}_{A,p}, \hat{s}_C\right] &= \mathrm{Cov}\left[\bar{a}_n + \gamma^*\left(\hat{s}_C - \bar{c}_n\right), \hat{s}_C\right] \\ &= \gamma^* \mathrm{Cov}\left[\hat{s}_C, \hat{s}_C\right] \\ &= \left(\frac{\rho_{AC}}{1+\eta}\right)\frac{\sigma_A}{\sigma_C}\mathrm{Var}\left[\hat{s}_C\right].\end{aligned}$$

Furthermore,

$$\begin{aligned}\mathrm{Var}\left[\hat{s}_{A,p}\right]\mathrm{Var}\left[\hat{s}_C\right] &= \left(1 - \frac{\rho_{AC}^2}{1+\eta}\right)\frac{\sigma_A^2}{n}\mathrm{Var}\left[\hat{s}_C\right] \\ &= \left(\frac{1+\eta-\rho_{AC}^2}{1+\eta}\right)\frac{\sigma_A^2}{n}\mathrm{Var}\left[\hat{s}_C\right].\end{aligned}$$

Therefore,

$$\begin{aligned}\mathrm{Corr}\left[\hat{s}_{A,p}, \hat{s}_C\right] &= \frac{\mathrm{Cov}\left[\hat{s}_{A,p}, \hat{s}_C\right]}{\sqrt{\mathrm{Var}\left[\hat{s}_{A,p}\right]\mathrm{Var}\left[\hat{s}_C\right]}} \\ &= \rho_{AC}\sqrt{\frac{n\mathrm{Var}\left[\hat{s}_C\right]}{(1+\eta)\left(1+\eta-\rho_{AC}^2\right)\sigma_C^2}} \\ &= \rho_{AC}\sqrt{\frac{\eta}{(1+\eta)\left(1+\eta-\rho_{AC}^2\right)}} \\ &= \frac{\rho_{AC}}{\sqrt{(1+1/\eta)\left(1+\eta-\rho_{AC}^2\right)}}.\end{aligned} \tag{12}$$

A contour plot of the correlation of the estimators $\mathrm{Corr}\left[\hat{s}_{A,p}, \hat{s}_C\right]$ as a function of the correlation coefficient of the model outputs ρ_{AC} and the ratio of estimator variances $\eta = \mathrm{Var}\left[\hat{s}_C\right]/\mathrm{Var}\left[\bar{c}_n\right]$ is shown in Figure 2. As the model outputs become more correlated, that is, ρ_{AC} is high, the estimators become more correlated as well.

If we assume normality, then $\hat{s}_{A,p}$ and $\hat{s}_C$ are jointly normally distributed as

$$\begin{bmatrix}\hat{s}_{A,p} \\ \hat{s}_C\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}s_A \\ s_C\end{bmatrix}, \begin{bmatrix}\mathrm{Var}\left[\hat{s}_{A,p}\right] & \mathrm{Cov}\left[\hat{s}_{A,p}, \hat{s}_C\right] \\ \mathrm{Cov}\left[\hat{s}_{A,p}, \hat{s}_C\right] & \mathrm{Var}\left[\hat{s}_C\right]\end{bmatrix}\right),$$

where $\mathrm{Var}\left[\hat{s}_{A,p}\right]$ is given by (10). Thus, $\hat{s}_{A,p}$ conditioned on $\hat{s}_C$ is a biased estimator of s_A

$$\begin{aligned}\hat{s}_{A,p} \mid \hat{s}_C \sim \mathcal{N}\Bigg(&s_A + \mathrm{Corr}\left[\hat{s}_{A,p}, \hat{s}_C\right]\sqrt{\frac{\mathrm{Var}\left[\hat{s}_{A,p}\right]}{\mathrm{Var}\left[\hat{s}_C\right]}}\left(\hat{s}_C - s_C\right), \\ &\left(1 - \mathrm{Corr}\left[\hat{s}_{A,p}, \hat{s}_C\right]^2\right)\mathrm{Var}\left[\hat{s}_{A,p}\right]\Bigg).\end{aligned}$$

In other words, given a particular realization of the estimator at the first optimization iteration, the information reuse estimators at all subsequent optimization iterations will have a bias. For fixed values of $\mathrm{Var}\left[\hat{s}_{A,p}\right]$ and $\mathrm{Var}\left[\hat{s}_C\right]$ (i.e., keeping the MSE fixed), as $\mathrm{Corr}\left[\hat{s}_{A,p}, \hat{s}_C\right]$ increases, the estimator $\hat{s}_{A,p}$ conditioned on a given value of $\hat{s}_C$ trades estimator variance for bias. Because this conditional estimator variance manifests as noise to the optimization algorithm, trading the conditional estimator variance for conditional bias (for a particular MSE) reduces the amount of

DOI: 10.1002/nme

noise in the objective and/or constraint functions experienced by the optimization algorithm. At the same time, as correlation increases, the overall MSE, given by (10), decreases because of the variance reduction. This combination of effects improves the convergence during optimization under uncertainty. This is discussed further in the numerical example in Section 7.2.

5. COMBINED ESTIMATOR

In this section, we return to the situation where an inexpensive low-fidelity model is available and discuss how to combine the multifidelity estimator and the information reuse estimator to further reduce the computational cost of uncertainty propagation in the context of optimization under uncertainty.

5.1. Approach

Let k be the current optimization iteration and let $\ell < k$ be a past optimization iteration as described in Section 4.2. Given vectors of design variables $\mathbf{x}_k$ and $\mathbf{x}_\ell$, we define the following random variables:

$$\begin{aligned} A(\omega) &= M_{\text{high}}\left(\mathbf{x}_k, \mathbf{U}(\omega)\right), \\ B(\omega) &= M_{\text{low}}\left(\mathbf{x}_k, \mathbf{U}(\omega)\right), \\ C(\omega) &= M_{\text{high}}\left(\mathbf{x}_\ell, \mathbf{U}(\omega)\right), \\ D(\omega) &= M_{\text{low}}\left(\mathbf{x}_\ell, \mathbf{U}(\omega)\right). \end{aligned}$$

The multifidelity estimators for $s_A = \mathbb{E}[A(\omega)]$ and $s_C = \mathbb{E}[C(\omega)]$, as discussed in Section 3.1, are

$$\begin{aligned} \hat{s}_A &= \bar{a}_n + \alpha\left(\bar{b}_m - \bar{b}_n\right), \\ \hat{s}_C &= \bar{c}_n + \beta\left(\bar{d}_m - \bar{d}_n\right), \end{aligned}$$

respectively, for control parameters $\alpha, \beta \in \mathbb{R}$ and $m > n$. We apply the information reuse estimator formulation (9) but replace the regular MC estimators $\bar{a}_n$ and $\bar{c}_n$ with the aforementioned multifidelity estimators instead. Therefore, the combined estimator of s_A, denoted as $\tilde{s}_{A,p}$, is

$$\begin{aligned} \tilde{s}_{A,p} &= \hat{s}_A + \gamma\left(\tilde{s}_C - \hat{s}_C\right) \\ &= \left[\bar{a}_n + \alpha\left(\bar{b}_m - \bar{b}_n\right)\right] + \gamma\left[\tilde{s}_C - \bar{c}_n - \beta\left(\bar{d}_m - \bar{d}_n\right)\right] \end{aligned} \tag{13}$$

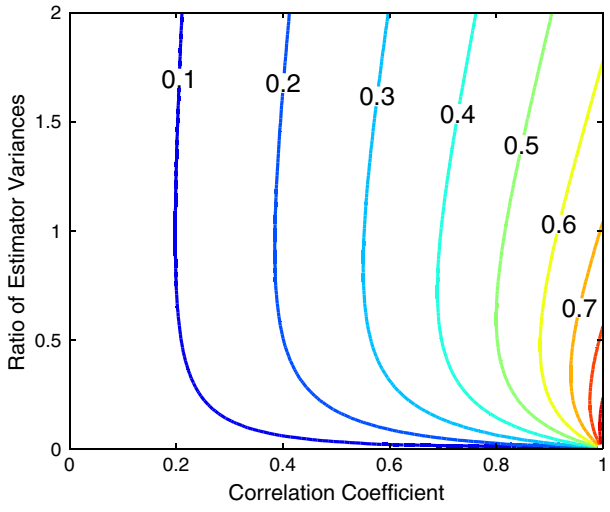


Figure 2. A contour plot of the correlation of the estimators Corr $[\hat{s}_{A,p}, \hat{s}_C]$ as a function of the correlation coefficient of the model outputs ρ_{AC} and the ratio of estimator variances $\eta = \text{Var}\left[\hat{s}_C\right] / \text{Var}\left[\bar{c}_n\right]$.

DOI: 10.1002/nme

for control parameter $\gamma \in \mathbb{R}$, where $\tilde{s}_C$ is the combined estimator at optimization iteration ℓ.

The computational effort for the combined estimator is $p = 2(n + m/w)$, where w is the ratio of the average computation time per high-fidelity model evaluation to the average computation time per low-fidelity model evaluation. If we were to follow the derivations in Sections 3.1 or 4.2, we would derive the expression for the combined estimator variance $\text{Var}\left[\tilde{s}_{A,p}\right]$ and minimize it with respect to all of the parameters α, β, γ, and $r = m/n > 1$ for fixed computational budget p. Unfortunately, there is no tractable analytical result similar to the multifidelity estimator and the information reuse estimator. Therefore, we propose a suboptimal approach to reduce the combined estimator variance whereby we determine the parameters sequentially.

We first calculate the optimal (but suboptimal overall) α and β for the multifidelity estimators $\hat{s}_A$ and $\hat{s}_C$, respectively. From Section 3.1, they are

$$\alpha^* = \rho_{AB}\frac{\sigma_A}{\sigma_B}, \qquad r^*_{AB} = \sqrt{\frac{w\rho^2_{AB}}{1-\rho^2_{AB}}}$$

and

$$\beta^* = \rho_{CD}\frac{\sigma_C}{\sigma_D}, \qquad r^*_{CD} = \sqrt{\frac{w\rho^2_{CD}}{1-\rho^2_{CD}}}.$$

To be conservative in expending computational effort to evaluate the low-fidelity model, we choose the ratio of the number of low-fidelity model evaluations to the number of high-fidelity model evaluations as

$$r^* = \frac{m}{n} = \min\left\{r^*_{AB}, r^*_{CD}\right\}.$$

Using these choices for α^*, β^*, and r^*, the variances of the multifidelity estimators, based on (5) and (6), are

$$\text{Var}\left[\hat{s}^*_A\right] = \left[1 - \left(1 - \frac{1}{r^*}\right)\rho^2_{AB}\right]\frac{\sigma^2_A}{n}$$

and

$$\text{Var}\left[\hat{s}^*_C\right] = \left[1 - \left(1 - \frac{1}{r^*}\right)\rho^2_{CD}\right]\frac{\sigma^2_C}{n}.$$

Next, we need to determine γ. Because we already have $\text{Var}\left[\hat{s}^*_A\right]$ and $\text{Var}\left[\hat{s}^*_C\right]$, the variance of the combined estimator is

$$\text{Var}\left[\tilde{s}_{A,p}\right] = \text{Var}\left[\hat{s}^*_A\right] + \gamma^2\left(\text{Var}\left[\tilde{s}_C\right] + \text{Var}\left[\hat{s}^*_C\right]\right) - 2\gamma\text{Cov}\left[\hat{s}^*_A, \hat{s}^*_C\right].$$

Minimizing the variance with respect to γ gives

$$\gamma^* = \frac{\text{Cov}\left[\hat{s}^*_A, \hat{s}^*_C\right]}{\text{Var}\left[\tilde{s}_C\right] + \text{Var}\left[\hat{s}^*_C\right]}$$

and

$$\text{MSE}\left[\tilde{s}^*_{A,p}\right] = \text{Var}\left[\tilde{s}^*_{A,p}\right] = \text{Var}\left[\hat{s}^*_A\right] - \frac{\text{Cov}\left[\hat{s}^*_A, \hat{s}^*_C\right]^2}{\text{Var}\left[\tilde{s}_C\right] + \text{Var}\left[\hat{s}^*_C\right]}. \tag{14}$$

DOI: 10.1002/nme

Lastly, to use (14), we derive the expression for the covariance of the multifidelity estimators:

$$\begin{aligned}
\mathrm{Cov}\left[\hat{s}_A^*, \hat{s}_C^*\right] &= \mathrm{Cov}\left[\bar{a}_n + \alpha^*\left(\bar{b}_m - \bar{b}_n\right), \bar{c}_n + \beta^*\left(\bar{d}_m - \bar{d}_n\right)\right] \\
&= \mathrm{Cov}\left[\bar{a}_n, \bar{c}_n\right] + \beta^*\mathrm{Cov}\left[\bar{a}_n, \bar{d}_m\right] - \beta^*\mathrm{Cov}\left[\bar{a}_n, \bar{d}_n\right] + \alpha^*\mathrm{Cov}\left[\bar{b}_m, \bar{c}_n\right] \\
&\quad - \alpha^*\mathrm{Cov}\left[\bar{b}_n, \bar{c}_n\right] + \alpha^*\beta^*\mathrm{Cov}\left[\bar{b}_m, \bar{d}_m\right] - \alpha^*\beta^*\mathrm{Cov}\left[\bar{b}_m, \bar{d}_n\right] \\
&\quad - \alpha^*\beta^*\mathrm{Cov}\left[\bar{b}_n, \bar{d}_m\right] + \alpha^*\beta^*\mathrm{Cov}\left[\bar{b}_n, \bar{d}_n\right] \\
&= \frac{1}{n}\left(\rho_{AC}\sigma_A\sigma_C - \beta^*\rho_{AD}\sigma_A\sigma_D - \alpha^*\rho_{BC}\sigma_B\sigma_C + \alpha^*\beta^*\rho_{BD}\sigma_B\sigma_D\right) \\
&\quad - \frac{1}{m}\left(-\beta^*\rho_{AD}\sigma_A\sigma_D - \alpha^*\rho_{BC}\sigma_B\sigma_C + \alpha^*\beta^*\rho_{BD}\sigma_B\sigma_D\right) \\
&= \frac{1}{n}\left[\rho_{AC}\sigma_A\sigma_C + \left(1 - \frac{1}{r^*}\right)\left(\alpha^*\beta^*\rho_{BD}\sigma_B\sigma_D - \alpha^*\rho_{BC}\sigma_B\sigma_C - \beta^*\rho_{AD}\sigma_A\sigma_D\right)\right] \\
&= \left[\rho_{AC} + \left(1 - \frac{1}{r^*}\right)\left(\rho_{AB}\rho_{CD}\rho_{BD} - \rho_{AB}\rho_{BC} - \rho_{CD}\rho_{AD}\right)\right]\frac{\sigma_A\sigma_C}{n}
\end{aligned}$$

The implementation of the combined estimator is similar to that of the multifidelity estimator and the information reuse estimator. At a particular n, we evaluate the high-fidelity model at both $\mathbf{x}_k$ and $\mathbf{x}_\ell$ n times and evaluate low-fidelity model at both $\mathbf{x}_k$ and $\mathbf{x}_\ell$ $m = nr^*$ times for a computational effort of $p = 2(n + m/w)$. Quantities such as σ_A^2, σ_C^2, ρ_{AB}, ρ_{BC}, ρ_{CD}, and so on are replaced by their sample estimates using the first n samples, similar to (7) and (11). If the MSE is too high, we increase n and repeat. Analogous to the information reuse estimator, we compare the results of (14) and (6) to decide if it is necessary to switch to the multifidelity estimator as a safeguard.

6. OPTIMIZATION METHODS

We return to optimization problem (1) introduced in Section 2 and briefly discuss the choice of optimization algorithms to use in the outer loop in conjunction with the multifidelity estimator, information reuse estimator, or the combined estimator in the inner loop. Because of the pseudo-randomness of MC sampling, the objective and constraint values returned to the optimizer at optimization iteration k, $\hat{f}(\mathbf{x}_k) = f\left(\mathbf{x}_k, \hat{\mathbf{s}}_{A,p}(\mathbf{x}_k)\right), \hat{g}(\mathbf{x}_k) = g\left(\mathbf{x}_k, \hat{\mathbf{s}}_{A,p}(\mathbf{x}_k)\right)$, and $\hat{h}(\mathbf{x}_k) = h\left(\mathbf{x}_k, \hat{\mathbf{s}}_{A,p}(\mathbf{x}_k)\right)$ are noisy with respect to the exact objective and constraint values $f\left(\mathbf{x}_k, \mathbf{s}_A(\mathbf{x}_k)\right), g\left(\mathbf{x}_k, \mathbf{s}_A(\mathbf{x}_k)\right)$, and $h\left(\mathbf{x}_k, \mathbf{s}_A(\mathbf{x}_k)\right)$ and the optimization problem becomes a stochastic optimization problem. While the level of noise can be controlled by specifying the desired estimator variance, it nevertheless poses a challenge for any optimization algorithm that is not noise tolerant. Therefore, we consider three classes of optimization algorithms:

Stochastic approximation (also known as the Robbins-Monro method) [19] generates steps in the estimated descent direction of the design space from the noisy evaluations of the objective (or gradient) functions, analogous to the steepest descent method for deterministic optimization. It converges to the true optimal design as $k \to \infty$. However, it can be difficult to incorporate nonlinear (potentially also noisy) constraints.

Sample average approximation (also known as the sample path method) [19] uses the same realizations of the uncertain model parameters $\mathbf{U}(\omega)$ to calculate the estimators for all optimization iterations. This effectively turns the objective and constraint functions into deterministic functions of $\mathbf{x}$ at the cost of introducing a bias, permitting a wide range of deterministic constrained nonlinear programming techniques to solve the optimization problem. As discussed in Section 4.2 for the information reuse estimator, the realizations of the uncertain model parameters at each optimization iteration must be independent of those at other optimization iterations. As a result, the information reuse estimator and the combined estimator are incompatible with the sample average approximation method.

DOI: 10.1002/nme

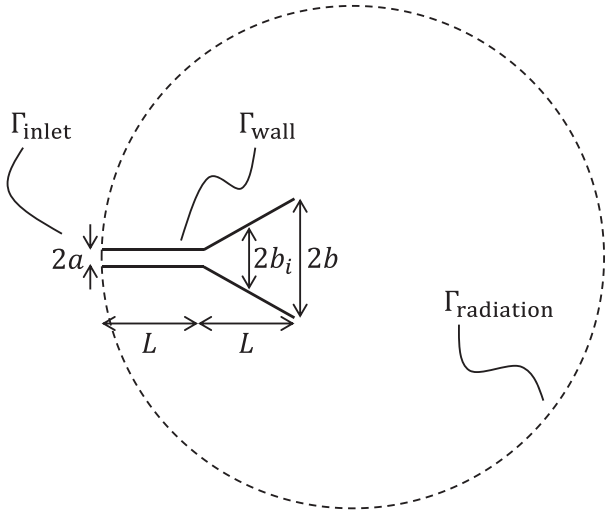


Figure 3. 2-D horn geometry where $a = 0.5$, $b = 3$, and $L = 5$. The shape of the horn flare is described by the half-widths $b_i, i = 1, \ldots, 6$ uniformly distributed along the flare.

Derivative-free optimization methods, such as mesh adaptive direct search [20], implicit filtering [21], derivative-free optimization (DFO) [22], bound optimization by quadratic approximation (BOBYQA)[**], and constrained optimization by linear approximation (COBYLA) [24], are not developed specifically for stochastic optimization problems but are typically tolerant to small levels of noise in practice [25]. Most of these methods sample the objective and constraint functions relatively widely in the design space to determine the search direction, which has the effect of smoothing out the high-frequency noise in the function evaluations provided that the magnitude of the noise is small relative to the true function values. However, there is no guarantee that the DFO methods will actually converge to the true optimal design.

In the succeeding example problems, we have selected the derivative-free optimization methods BOBYQA and COBYLA. BOBYQA solves bound-constrained optimization problems without analytical derivatives by constructing an underdetermined quadratic interpolation model of the objective function based on the least-Frobenius-norm update of the Hessian [26, 27]. The quadratic model is used in a trust-region subproblem to generate the vectors of design variables. COBYLA can also handle nonlinear constraints by constructing linear interpolation models of the objective and constraint functions using evaluations on a simplex. The vectors of design variables are obtained by solving a linear programming subproblem.

7. NUMERICAL RESULTS

We demonstrate the multifidelity estimator, the information reuse estimator, and the combined estimator with two example problems.

7.1. Acoustic horn uncertainty propagation

In this example, we model a 2-D acoustic horn governed by the non-dimensional complex Helmholtz equation $\nabla^2 u + k^2 u = 0$. An incoming wave enters the horn through the inlet and exits the outlet into the exterior domain with a truncated absorbing boundary $\Gamma_{\text{radiation}}$ [28]. The geometry of the horn is illustrated in Figure 3. The output of the model is the reflection coefficient $s = \left| \int_{\Gamma_{\text{inlet}}} u \, \mathrm{d}\Gamma - 1 \right|$, a measure of the horn's efficiency, and we estimate its mean and its variance. The three uncertain model parameters considered for this example are the wave number k, upper

[**]BOBYQA was developed by M.J.D. Powell based on his earlier work on the NEWUOA software [23].

DOI: 10.1002/nme

Table I. Distributions of the three uncertain parameters for the horn example.

Random variable	Distribution	Lower bound	Upper bound	Mean	Standard Deviation
$k(\omega)$	Uniform	1.3	1.5	–	–
$z_u(\omega)$	Normal	–	–	50	3
$z_l(\omega)$	Normal	–	–	50	3

horn wall impedance z_u, and lower horn wall impedance z_l with distributions listed in Table I. In the next section, we also consider six geometric design variables b_1 to b_6 describing the profile of the horn flare for optimization. Here, for uncertainty propagation, the design variables are fixed at a straight flare profile as shown in Figure 3.

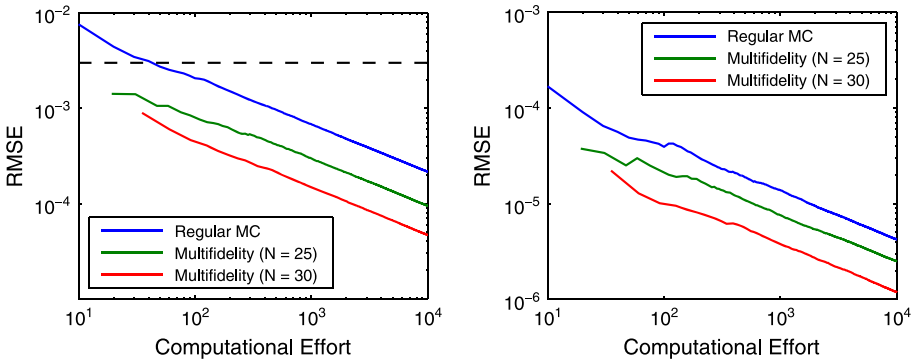
The high-fidelity model is a finite element model of the Helmholtz equation [28]. The spatial domain as depicted in Figure 3 is discretized using linear nodal basis functions, resulting in a system of 35,895 equations with 35,895 unknowns corresponding to the pressures at the nodal grid points. At a particular vector of design variables and a realization of uncertain model parameters, the system is solved for the pressures that are then used to compute the reflection coefficient of the horn.

The low-fidelity model is a reduced basis model obtained by projecting the system of equations from the high-fidelity finite element discretization onto a reduced subspace [29]. This is done by determining $N \ll 35,895$ basis vectors, which define an N-dimensional reduced subspace. Approximating the unknowns as a linear combination of these basis vectors and projecting the governing equations onto the reduced subspace lead to a system of N equations with N unknowns. The reduced system is significantly faster to solve but is less accurate because of the discarded modes. We consider two cases for the low-fidelity model: (i) a less accurate reduced basis model with $N = 25$ reduced basis functions and (ii) a more accurate reduced basis model with $N = 30$ reduced basis functions. For (i), the correlation coefficient between the high-fidelity model and the low-fidelity model (computed from (7c)) is $\hat{\rho}_{AB} = 0.959$ for the samples used to estimate the mean and is $\hat{\rho}_{AB} = 0.897$ for the samples used to estimate the variance. For (ii), it is $\hat{\rho}_{AB} = 0.998$ for the samples used to estimate the mean and is $\hat{\rho}_{AB} = 0.994$ for the samples used to estimate the variance. In both cases, the ratio of the average computation time of the high-fidelity model to that of the low-fidelity model is $w = 40$ (the increase in computation time caused by the addition of five more reduced basis functions in (ii) relative to (i) is negligible compared with the 35,895 unknowns in the high-fidelity model). This allows us to examine the effect of the correlation coefficient on the efficiency of the multifidelity estimator.

The root mean square error (RMSE) of the mean estimator and of the variance estimator are shown in Figure 4(a) and (b), respectively. The computational effort is the number of high-fidelity model evaluations for the regular MC estimator and the equivalent number of high-fidelity model evaluations for the multifidelity estimator. This example demonstrates the benefit of a good correlation between the high-fidelity model and the low-fidelity model. To achieve 10^{-5} RMSE for the variance estimate, the regular MC estimator requires about 1800 high-fidelity model evaluations, the multifidelity estimator with the less correlated low-fidelity model ($N = 25$ basis functions) requires about 600 equivalent high-fidelity model evaluations, and the multifidelity estimator with the more correlated low-fidelity model ($N = 30$ basis functions) requires about 100 equivalent high-fidelity model evaluations.

Given the high correlation between the reduced basis model with $N = 30$ basis functions and the finite element model, it appears to be simpler to throw all of the computational budget into computing a regular MC estimator using only the reduced order model. If we do this for the mean estimate, there will be a bias error of about -3×10^{-3} with respect to the true mean of the finite element model that cannot be reduced regardless of the number of cheap reduced order model evaluations used. For this problem, the bias is large relative to the RMSE that can be achieved by the multifidelity estimator (Figure 4(a)) and highlights the fact that the choice of the low-fidelity model should be on the basis of its correlation with the high-fidelity model rather than point-wise differences in the outputs.

DOI: 10.1002/nme



(a) Mean estimator (b) Variance estimator

Figure 4. Root mean square error of the estimators as a function of the computational effort for the horn example. The dashed line indicates the bias of the low-fidelity model ($N = 30$ basis functions). (a) mean estimator and (b) variance estimator.

Table II. Initial values, lower bounds, upper bounds, and optimal values of the six horn flare half-widths.

Design variables	Initial values	Lower bounds	Upper bounds	Optimal values
b_1	0.857	0.679	1.04	0.679
b_2	1.21	1.04	1.39	1.07
b_3	1.57	1.39	1.75	1.75
b_4	1.93	1.75	2.11	1.99
b_5	2.29	2.11	2.46	2.25
b_6	2.64	2.46	2.82	2.46

7.2. Acoustic horn robust optimization

Next, we demonstrate the effectiveness of the information reuse estimator in the robust optimization of the shape of the acoustic horn flare. The model we consider is the reduced basis model for the reflection coefficient s with $N = 30$ basis functions. The design variables are $\mathbf{b} = [b_1 \cdots b_6]^\top$, representing the half-widths of the horn flare as shown in Figure 3. The initial values of the design variables (corresponding to the straight flare), their lower bounds, and their upper bounds are listed in Table II. The minimization of the horn reflection coefficient is formulated as

$$\min_{\mathbf{b}_L \leqslant \mathbf{b} \leqslant \mathbf{b}_U} f(\mathbf{b}) = \mathbb{E}[s(\mathbf{b}, \omega)] + 3\sqrt{\mathrm{Var}\,[s(\mathbf{b}, \omega)]}.$$

The mean and the variance of the horn reflection coefficient, $\mathbb{E}[s(\mathbf{b}, \omega)]$ and $\mathrm{Var}\,[s(\mathbf{b}, \omega)]$, respectively, are estimated using either the regular MC estimator or the information reuse estimator. The MSE of the objective function can be approximately evaluated from the estimator variances of the regular MC estimator or the information reuse estimator based on a first-order Taylor expansion about $\mathbb{E}[s(\mathbf{b}, \omega)]$ and $\mathrm{Var}\,[s(\mathbf{b}, \omega)]$.

We employ the BOBYQA optimization algorithm described in Section 6. The optimization is conducted with the objective function evaluated using the regular MC estimator and with the objective function evaluated using the information reuse estimator. In both cases, we employ the number of model evaluations needed to satisfy the RMSE tolerance on objective function of 1×10^{-3} at every vector of design variables. Three trials were run for each case and the convergence of the objective as a function of the cumulative computational effort is shown in Figure 5, where the computational effort is the number of model evaluations. It can be seen that we locate the optimum using about 50% less computational effort with the information reuse estimator.

We examine the computational cost in more detail by plotting the computational effort used at each optimization iteration in Figure 6(a). We see that, while both estimators require about the same computational effort during the first few optimization iterations, the information reuse estimator

DOI: 10.1002/nme

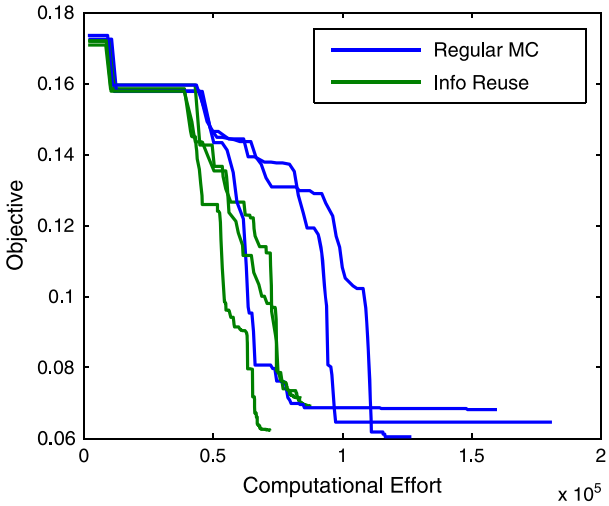
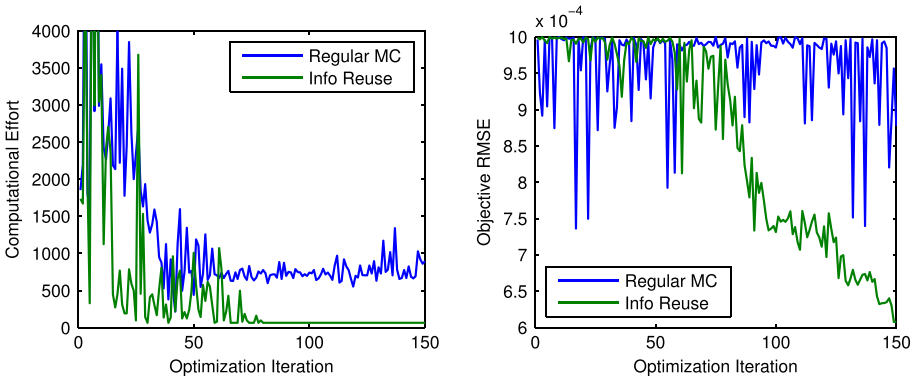


Figure 5. Comparison of convergence histories for the robust horn optimization using the regular Monte Carlo estimator and the information reuse estimator.



(a) Computational effort p (b) Objective RMSE

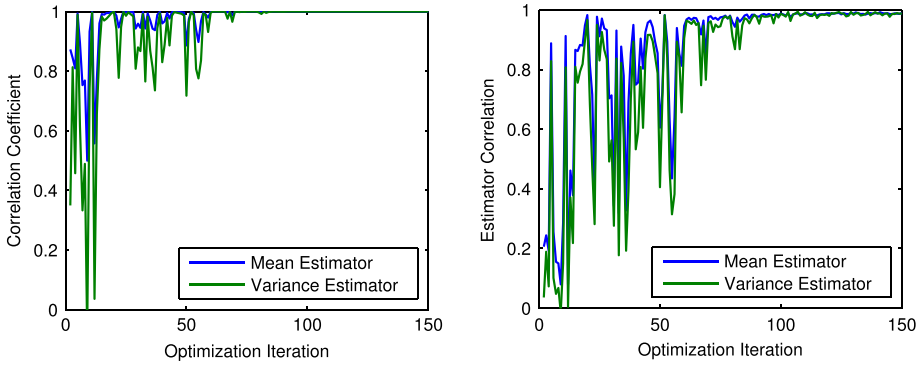
Figure 6. Computational effort per optimization iteration and the root mean square error of the objective versus optimization iteration for the robust horn optimization example. (a) computational effort and (b) objective RMSE.

requires significantly less computational effort at subsequent optimization iterations. However, we maintained a minimum of 32 samples at each optimization iteration to calculate the parameters in (11). The result is that the RMSE of the objective actually decreases and becomes smaller than the specified tolerance of 1×10^{-3} as shown in Figure 6(b).

The reduction in computational effort corresponds to the high correlation coefficient $\hat{\rho}_{AC}$ shown in Figure 7(a) due to the optimizer taking smaller and smaller steps in the design space at the later optimization iterations as it refines the vector of design variables near the optimum. As discussed in Section 4.1, the correlation between the random model output at a vector of design variables and the random model output at another vector of design variables tends to increase as the distance between the two vectors of design variables decreases.

In Figure 7(b), we plot the correlation coefficient between the estimators $\hat{s}_{A,p}$ and $\hat{s}_C$ as derived in (12). It shows that the errors in the estimators at the later optimization iterations become more and more correlated with each other. As discussed in Section 4.4, this correlation has the effect of smoothing the noise in the objective function at the cost of introducing a bias. This may help prevent the derivative-free optimization algorithm from getting stuck in the noise of the objective function by causing it to behave more like the sample average approximation method. Furthermore, as shown in Figure 6(b), the estimator variances (not conditioned on the previous estimators) decreases at the

DOI: 10.1002/nme



(a) Model output correlation $\hat{\rho}_{AC}$ (b) Estimator correlation $\mathrm{Corr}\left[\hat{s}_{A,p}, \hat{s}_C\right]$

Figure 7. The model output correlation coefficient and the estimator correlation coefficient for the information reuse estimator versus optimization iteration for the robust horn optimization example. (a) model output correlation $\hat{\rho}_{AC}$ and (b) estimator correlation $\mathrm{Corr}\left[\hat{s}_{A,p}, \hat{s}_C\right]$.

later optimization iterations. Therefore, unlike the sample average approximation method, the bias can decrease as the optimization algorithm progresses.

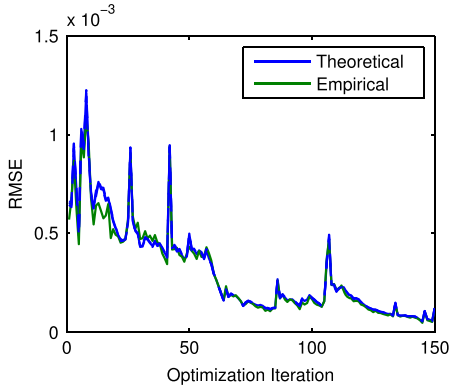
We verify our theoretical RMSE for the information reuse estimator from (10) by comparing it with the empirical RMSE. We fix the sequence of vectors of design variables generated by the optimization algorithm and calculate the information reuse estimators on this sequence of vectors of design variables, expending $p = 1000$ computational effort on every vector of design variables. We repeat this with new realizations of the random inputs 100 times to calculate the empirical RMSE of the information reuse estimators. The results are shown in Figure 8 and show good agreement.

7.3. Acoustic horn – combined estimator

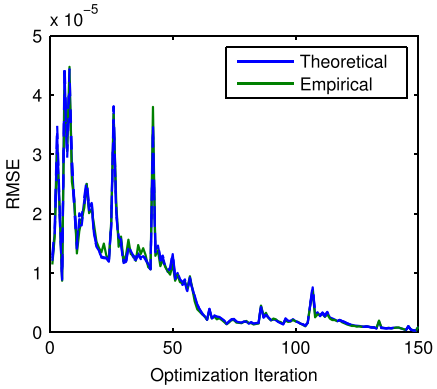
Finally, we compare the regular MC estimator, the information reuse estimator, the multifidelity estimator, and the combined estimator. The problem setup is the same as that in Section 7.2; however, we now consider the high-fidelity model to be the finite element model of the Helmholtz equation and the low-fidelity model to be the reduced basis model with $N = 30$ basis functions. Figure 9 shows the convergence of the objective with respect to the cumulative computational effort, where the computational effort is the number of high-fidelity model evaluations for the regular MC estimator and the information reuse estimator and is the equivalent number of high-fidelity model evaluations for the multifidelity estimator and the combined estimator. It can be seen that the combined estimator requires significantly less computational effort than the regular MC estimator. For this problem, the multifidelity estimator is already quite efficient and so the combined estimator does not provide much additional benefit. Nevertheless, Figure 10 shows that the combined estimator indeed combines the benefits of both the multifidelity estimator and the information reuse estimator—at the first few optimization iterations, when the optimizer takes relative large steps in the design space, the combined estimator takes about the same computational effort as the multifidelity estimator; at the last few optimization iterations, when the optimizer takes relative small steps in the design space, the combined estimator uses only the minimum number of samples as in the case of the information reuse estimator. Overall the information reuse estimator provided about 60% computational savings, the multifidelity estimator provided about 75% computational savings, and the combined estimator provided about 90% computational savings, as shown in Table III.

Table IV presents the four slightly different optimal solutions obtained using each of the four estimators and shows the challenge in solving noisy optimization problems using derivative-free optimization algorithms. While the RMSE of 1×10^{-3} is two orders of magnitude less than the initial objective value, it is only one order of magnitude less than the final objective value. As discussed in Section 6, derivative-free optimization methods are not guaranteed to converge to the true optimal solution in the presence of noise and may terminate prematurely. If computational resources allow, one may perform multiple runs of the optimization to obtain a spread of the final objective values,

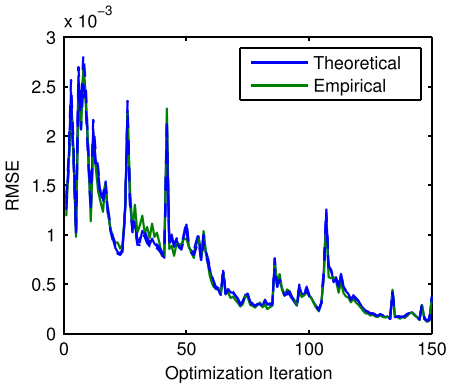
DOI: 10.1002/nme



(a) Mean estimator



(b) Variance estimator



(c) $f = \mathbb{E}[s] + 3\sqrt{\mathrm{Var}[s]}$ estimator

Figure 8. Comparison of the theoretical root mean square errors to the empirical root mean square errors of the information reuse estimators for the robust horn optimization example. (a) mean estimator, (b) variance estimator, and (c) $f = \mathbb{E}[s] + 3\sqrt{\mathrm{Var}[s]}$ estimator.

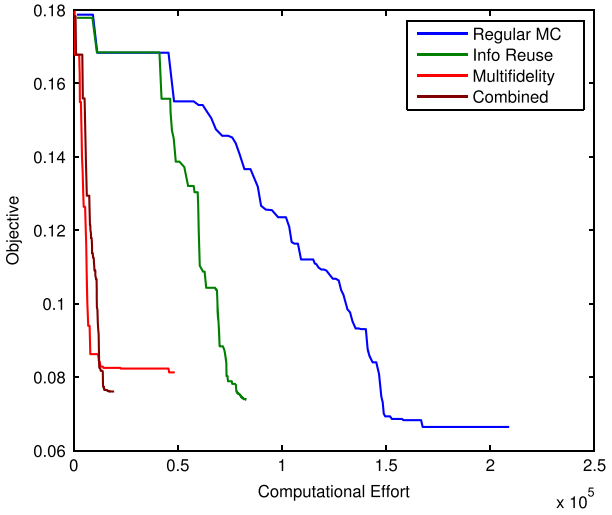


Figure 9. Comparison of convergence histories for the robust horn optimization using the regular Monte Carlo estimator, the information reuse estimator, the multifidelity estimator, and the combined estimator.

DOI: 10.1002/nme

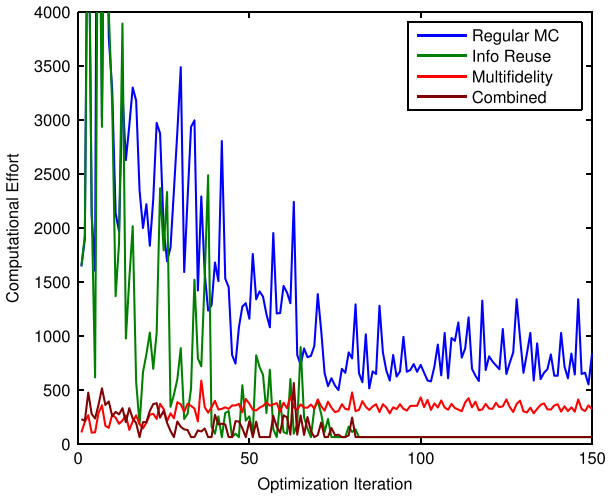


Figure 10. Computational effort per optimization iteration versus optimization iteration for the robust horn optimization example.

Table III. Comparison of the total computational efforts for the robust horn optimization using the regular Monte Carlo estimator, the information reuse estimator, the multifidelity estimator, and the combined estimator.

	Total computational effort
Regular Monte Carlo estimator	209,096
Information reuse estimator	82,928
Multifidelity estimator	48,595
Combined estimator	19,449

Table IV. Comparison of the final design variables for the robust horn optimization using the regular Monte Carlo (MC) estimator, the information reuse estimator, the multifidelity estimator, and the combined estimator.

Design variables	Regular MC	Info reuse	Multifidelity	Combined
b_1	0.679	0.679	0.679	0.679
b_2	1.06	1.08	1.10	1.06
b_3	1.75	1.67	1.66	1.69
b_4	1.96	1.86	1.89	1.84
b_5	2.26	2.11	2.11	2.14
b_6	2.46	2.46	2.52	2.48

similar to what is sometimes done for the sample average approximation method. For the acoustic horn problem, the spread of final objective values for the different estimators would overlap, as can be seen in Figure 5.

7.4. Multifidelity robust wing optimization

For the second example problem, we consider the robust optimization of a wing whose geometry is loosely based on the Bombardier Q400 aircraft. The numerical model is a coupled aerostructural solver using Tripan for the aerodynamic analysis and the Toolkit for the Analysis of Composite Structures (TACS) for the structural analysis [30]. Tripan is an unstructured panel method using source and doublet singularities for inviscid, incompressible, external lifting flow. TACS is a second-order finite element method using linear or nonlinear strain relationships for thin-walled structures. The aerostructural solver is parallelized at both the coupling level and at the individual analysis

DOI: 10.1002/nme

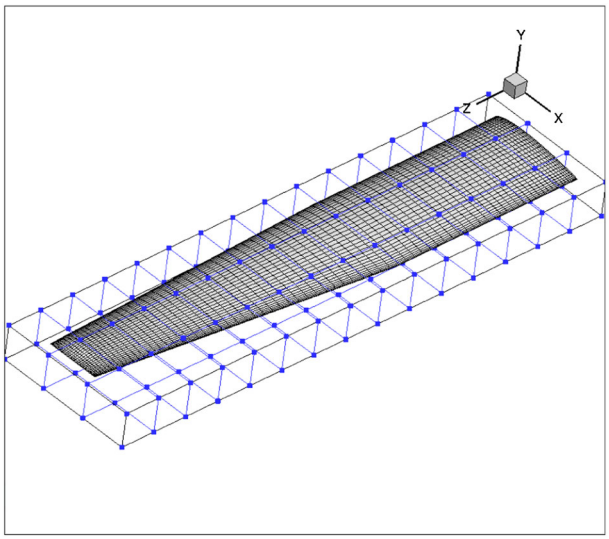


Figure 11. Geometry and the free-form deformation control points of the wing [32, Figure 6-8].

Table V. Geometry data of the wing.

Planform area [m^2]	68.44
Wing span [m]	28.40
Aspect ratio	11.78
Taper ratio	0.54
Taper break [% of span]	40
Airfoil stack	NACA 2412
Twist	None
Number of ribs	20
Forward spar location [% of chord]	15
Aft spar location [% of chord]	50

level to take advantage of multiple processors. A computer-aided-design-free (CAD-free) approach based on free-form deformations [31] is used to parameterize the wing geometry. The wing and the free-form deformation control points are shown in Figure 11, and some data for the initial wing geometry are listed in Table V.

In the deterministic optimization problem [30, 32], the drag of the wing is minimized subject to 1-g cruise lift equals weight constraint, four Kreisselmeier-Steinhauser (KS) stress constraints for the ribs, the spars, the upper surface skin, and the lower surface skin, and 72 constraints on the magnitude of the change in thicknesses of spars and skins along the span for a total of 77 constraints. The design variables include the angle of attack, eight wing twist angles along the span, and 19 thicknesses of each of the forward spar, the aft spar, the ribs, the upper surface skin, and the lower surface skin for a total of 104 variables.[††] Because only the wing is optimized, any savings in wing weight is subtracted from the total aircraft weight for the lift constraint.

We introduce uncertainties into the cruise condition and structural properties using triangular distributions listed in Table VI. However, the deterministic optimization problem cannot be translated into a robust optimization problem in a straightforward manner because it is not clear how to apply the lift equals weight equality constraint when the lift and the weight are random outputs of the numerical model. To resolve this issue, we wrap the secant root-finding method around the aerostructural solver to find the angle of attack that satisfies the equality constraint. This effectively moves the equality constraint out of the optimization problem and into the augmented numerical

[††]The aspect ratio is a natural choice for the design variables as it provides a trade-off between aerodynamics and structural loads, but it is not available in the current free-form deformation parameterization of the wing.

DOI: 10.1002/nme

Table VI. Triangular distributions of the seven uncertain model parameters for the robust wing optimization problem.

Random variable	Lower limit	Mode	Upper limit
Maximum take-off weight [kg]	29257×0.9	29257	29257×1.1
Fraction of weight not including wing	0.83	0.86	0.89
Cruise Mach number	0.55	0.6	0.65
Material density [kg/m^3]	2810×0.95	2810	2810×1.05
Material elastic modulus [GPa]	70×0.9	70	70×1.1
Material poisson ratio	0.31	0.33	0.35
Material yield stress [MPa]	370×0.9	370	370×1.1

model and eliminates the angle of attack from the design variables. The interpretation is that each set of design variables (wing twists and thicknesses) along with a realization of the uncertain model parameters define a wing for the aircraft that is trimmed for level flight and the augmented numerical model outputs the drag and stresses for this wing at level flight. This illustrates that a numerical model for deterministic optimization may not always make sense in the presence of uncertainties.

We further modify the problem by removing the rib and skin thicknesses from the design variables and their corresponding change-in-thickness constraints to reduce the size of the problem. Therefore, the robust optimization problem has 46 design variables (8 twist angles, 19 forward spar thicknesses, and 19 aft spar thicknesses) and 40 constraints (4 stress constraints and 36 change in thickness constraints). Let the vector of design variables be $\mathbf{x}$ with lower and upper bounds $\mathbf{x}_L$ and $\mathbf{x}_U$, respectively, and let $\mathbf{u}$ be a realization of the uncertain model parameters $\mathbf{U}(\omega)$ listed in Table VI. The four stress constraints are denoted as

$$\begin{aligned}
\mathrm{ks}_1(\mathbf{x},\mathbf{u}) &= \text{KS function for rib stress} - \text{yield stress} \leqslant 0\\
\mathrm{ks}_2(\mathbf{x},\mathbf{u}) &= \text{KS function for spar stress} - \text{yield stress} \leqslant 0\\
\mathrm{ks}_3(\mathbf{x},\mathbf{u}) &= \text{KS function for upper surface skin stress} - \text{yield stress} \leqslant 0\\
\mathrm{ks}_4(\mathbf{x},\mathbf{u}) &= \text{KS function for lower surface skin stress} - \text{yield stress} \leqslant 0.
\end{aligned}$$

The change-in-thickness constraints are deterministic and linear and are represented as the matrix $\mathbf{K}$. In the presence of uncertain model parameters, the robust wing optimization problem is formulated as

$$\begin{aligned}
\min_{\mathbf{x}_L \leqslant \mathbf{x} \leqslant \mathbf{x}_U} \quad & \mathbb{E}\left[\mathrm{drag}\left(\mathbf{x},\mathbf{U}(\omega)\right)\right] + \lambda\sqrt{\mathrm{Var}\left[\mathrm{drag}\left(\mathbf{x},\mathbf{U}(\omega)\right)\right]}\\
\text{s.t.} \quad & \mathbb{E}\left[\mathrm{ks}_1\left(\mathbf{x},\mathbf{U}(\omega)\right)\right] + \lambda\sqrt{\mathrm{Var}\left[\mathrm{ks}_1\left(\mathbf{x},\mathbf{U}(\omega)\right)\right]} \leqslant 0\\
& \mathbb{E}\left[\mathrm{ks}_2\left(\mathbf{x},\mathbf{U}(\omega)\right)\right] + \lambda\sqrt{\mathrm{Var}\left[\mathrm{ks}_2\left(\mathbf{x},\mathbf{U}(\omega)\right)\right]} \leqslant 0\\
& \mathbb{E}\left[\mathrm{ks}_3\left(\mathbf{x},\mathbf{U}(\omega)\right)\right] + \lambda\sqrt{\mathrm{Var}\left[\mathrm{ks}_3\left(\mathbf{x},\mathbf{U}(\omega)\right)\right]} \leqslant 0\\
& \mathbb{E}\left[\mathrm{ks}_4\left(\mathbf{x},\mathbf{U}(\omega)\right)\right] + \lambda\sqrt{\mathrm{Var}\left[\mathrm{ks}_4\left(\mathbf{x},\mathbf{U}(\omega)\right)\right]} \leqslant 0\\
& \mathbf{K}\mathbf{x} \leqslant 0,
\end{aligned}$$

where we chose $\lambda = 2$.

We employ the COBYLA optimization algorithm [24] described in Section 6. The discretizations of the wing for the aerostructural solver are shown in Table VII with the high-fidelity model utilizing the fine discretization and the low-fidelity model utilizing the coarse discretization. The ratio of average computational cost between the high-fidelity model and the low-fidelity model, w, is approximately 4. The optimization is conducted with the objective and constraint functions evaluated using the combined estimator and with the objective and constraint functions evaluated using the information reuse estimator. We do not have the regular MC estimator as a comparison because of unaffordable computational cost. In both cases, the tolerance on the RMSE of the objective and constraint function estimators is fixed at 2×10^{-4}.

The convergence of the objective as a function of the cumulative computational effort is shown in Figure 12. The computational effort is the number of high-fidelity model evaluations for the

DOI: 10.1002/nme

Table VII. Number of degrees of freedom for different discretizations of the aerodynamic and structural elements. The approximate evaluation time for the aerostructural solver wrapped with the secant method is based on a PC with 16 processors at 2.93 GHz each.

		Coarse	Fine
Aerodynamic panels	Chordwise	30	34
	Spanwise	20	40
	Wake	20	40
	Total panels	1000	2960
Structural elements	Chordwise	5	8
	Spanwise	30	80
	Thickness	4	6
	Total d.o.f.	5624	14288
Approximate evaluation time [s]		6	24

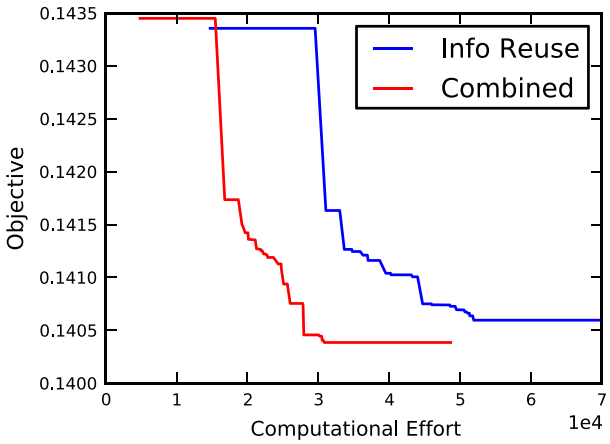
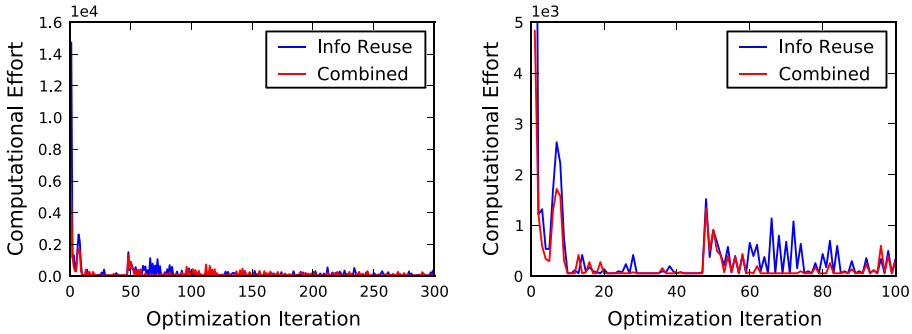


Figure 12. Comparison of convergence histories for the robust wing optimization problem using the information reuse estimator and the combined estimator. The objective is mean $+\lambda$ standard deviations of the drag coefficient.



(a) Original axes scales (b) Expanded axes scales

Figure 13. Computational effort per optimization iteration versus optimization iteration for the robust wing optimization problem. The scales of the axes have been expanded in (b) to show the comparison more clearly. (a) original axes scales and (b) expanded axes scales.

information reuse estimator and the equivalent number of high-fidelity model evaluations for the combined estimator. The initial wing drag coefficient has a mean of 0.1254 and a standard deviation of 0.008974 and the final wing drag coefficient has a mean of 0.1239 and a standard deviation of 0.008515—a reduction of about 1.2% in expected wing drag and a reduction of 5.1% in the

DOI: 10.1002/nme

standard deviation. In Figure 13, we plot the computational effort required to compute the combined estimator and the information reuse estimator at each optimization iteration. As discussed in previous examples, the information reuse estimator tends to reduce computational effort only in the later optimization iterations. It can be seen that the main benefit of the combined estimator here is to reduce the computational effort during the first few optimization iterations by leveraging the cheaper low-fidelity model. The solution was obtained in 13.4 days using the information reuse estimator and in 9.7 days using the combined estimator on a PC with 16 processors at 2.93 GHz each. If the 90% computational savings of the combined estimator over the regular MC estimator seen in the acoustic horn robust optimization example in Section 7.3 is extended to this problem, the solution would have taken about 3.2 months to obtain using the regular MC estimator.

8. CONCLUSIONS

In this paper, we developed the multifidelity estimator, the information reuse estimator, and the combined estimator to address the need for a general optimization under uncertainty approach that makes use of inexpensive, approximate information to reduce computational cost. The quality of the approximate information is measured in terms of its linear correlation with the high-fidelity model output—a strong correlation indicates that a statistical correction calculated from the approximate information is likely to be beneficial to the estimator of the statistic of the high-fidelity model output.

The estimators presented are not much more difficult to implement than the regular MC estimator. They can make use of models that are provided as closed 'black-boxes'. They are embarrassingly parallelizable. Furthermore, they do not require a priori knowledge about the models. Although all of the estimators share the same slow convergence rate as the regular MC estimator, the reduction in the MSE can be significant enough, as demonstrated in the numerical examples, to mitigate this disadvantage unless very high accuracy is needed. On the other hand, the methods are suitable for problems with a very large number of uncertain parameters.

An apparent extension to the estimators is to consider two or more sources of approximate information, for example, using multiple low-fidelity models and reusing multiple past optimization iterations. However, the benefit is not immediately clear—the multiple sources of approximate information are likely to be correlated with each other, thus providing less unique new information for the high-fidelity model output. It may also be advantageous to generalize the approach to methods beyond MC simulation, particularly to those that exhibit faster convergence rate for smooth problems such as quasi-MC simulation. The challenge is in determining which aspect of the high-fidelity model output should be correlated with the approximate information [33].

The robust wing optimization problem represents a first step in large-scale, high-fidelity engineering design under uncertainty. Further development of this problem includes extending the free-form deformation parameterization of the wing to cover other geometric changes and incorporating distributed uncertainties in the structural properties along the wing. Distributed uncertainties is a more realistic scenario than uncertainties in the bulk material properties and may also be used to represent degradation in the wing structure. The estimators developed in this paper are well-suited to handle this high-dimensional (in the stochastic space) problem.

APPENDIX A: AUTOCORRELATION DERIVATION

In this appendix, we derive the result presented in Section 4.1:

$$\begin{aligned}&\operatorname{Corr}\left[M_{\text{high}}(x+\Delta x,\mathbf{U}(\omega)),M_{\text{high}}(x,\mathbf{U}(\omega))\right]\\&\approx 1-\frac{1-\operatorname{Corr}\left[M'_{\text{high}}\left(x,\mathbf{U}(\omega)\right),M_{\text{high}}\left(x,\mathbf{U}(\omega)\right)\right]^2}{2\operatorname{Var}\left[M_{\text{high}}\left(x,\mathbf{U}(\omega)\right)\right]/\operatorname{Var}\left[M'_{\text{high}}\left(x,\mathbf{U}(\omega)\right)\right]}\Delta x^2,\end{aligned}$$

for $|\Delta x| \ll 1$. To simplify notation, for the remainder of this appendix we drop the subscript 'high' and the argument $\mathbf{U}(\omega)$ from M.

DOI: 10.1002/nme

Let the model be twice differentiable in x for all realizations of $\mathbf{U}(\omega)$. For some small Δx, we apply a second-order Taylor expansion in x to obtain

$$M(x+\Delta x) \approx M(x) + M'(x)\Delta x + M''(x)\frac{\Delta x^2}{2},$$

where $M'(x) = \frac{\partial M(x)}{\partial x}$ and $M''(x) = \frac{\partial^2 M(x)}{\partial x^2}$. Taking the expectation of both sides, we obtain

$$\mu_M(x+\Delta x) \approx \mu_M(x) + \mu_{M'}(x)\Delta x + \mu_{M''}(x)\frac{\Delta x^2}{2},$$

where $\mu_M(x) = \mathbb{E}[M(x)]$, $\mu_{M'} = \mathbb{E}[M'(x)]$, and $\mu_{M''} = \mathbb{E}[M''(x)]$.

We first derive the covariance between $M(x+\Delta x)$ and $M(x)$:

$$\begin{aligned}
&\mathrm{Cov}\left[M(x+\Delta x), M(x)\right] \\
&= \mathbb{E}\left[\{M(x+\Delta x) - \mu_M(x+\Delta x)\}\{M(x) - \mu_M(x)\}\right] \\
&\approx \mathbb{E}\left[\{M(x) - \mu_M(x)\}^2\right] + \mathbb{E}\left[\left\{M'(x) - \mu_{M'}(x)\right\}\{M(x) - \mu_M(x)\}\right]\Delta x \\
&\quad + \mathbb{E}\left[\left\{M''(x) - \mu_{M''}(x)\right\}\{M(x) - \mu_M(x)\}\right]\frac{\Delta x^2}{2} \\
&= \mathrm{Var}\left[M(x)\right] + \mathrm{Cov}\left[M'(x), M(x)\right]\Delta x + \mathrm{Cov}\left[M''(x), M(x)\right]\frac{\Delta x^2}{2} \\
&= \mathrm{Var}\left[M(x)\right]\left\{1 + \frac{\mathrm{Cov}\left[M'(x), M(x)\right]}{\mathrm{Var}\left[M(x)\right]}\Delta x + \frac{\mathrm{Cov}\left[M''(x), M(x)\right]}{\mathrm{Var}\left[M(x)\right]}\frac{\Delta x^2}{2}\right\}.
\end{aligned}$$

Next, we derive the variance of $M(x+\Delta x)$ in a similar manner:

$$\begin{aligned}
&\mathrm{Var}\left[M(x+\Delta x)\right] \\
&= \mathbb{E}\left[\{M(x+\Delta x) - \mu_M(x+\Delta x)\}^2\right] \\
&\approx \mathrm{Var}\left[M(x)\right] + \mathrm{Var}\left[M'(x)\right]\Delta x^2 + \mathrm{Var}\left[M''(x)\right]\frac{\Delta x^4}{4} + 2\,\mathrm{Cov}\left[M'(x), M(x)\right]\Delta x \\
&\quad + 2\,\mathrm{Cov}\left[M''(x), M(x)\right]\frac{\Delta x^2}{2} + 2\,\mathrm{Cov}[M''(x), M'(x)]\frac{\Delta x^3}{2}.
\end{aligned}$$

Using the formula for the Taylor expansion of the inverse square root

$$\frac{1}{\sqrt{a + bz + cz^2 + dz^3 + ez^4}} \approx \frac{1}{a^{1/2}} - \frac{b}{2a^{3/2}}z + \frac{3b^2 - 4ac}{8a^{5/2}}z^2,$$

we obtain

$$\begin{aligned}
&\frac{1}{\sqrt{\mathrm{Var}\left[M(x+\Delta x)\mathrm{Var}\left[M(x)\right]\right]}} \\
&\approx \frac{1}{\mathrm{Var}\left[M(x)\right]} - \frac{\mathrm{Cov}\left[M'(x), M(x)\right]}{\mathrm{Var}\left[M(x)\right]^2}\Delta x \\
&\quad + \frac{3\,\mathrm{Cov}\left[M'(x), M(x)\right]^2 - \mathrm{Var}\left[M(x)\right]\{\mathrm{Var}\left[M'(x)\right] + \mathrm{Cov}\left[M''(x), M(x)\right]\}}{2\mathrm{Var}\left[M(x)\right]^3}\Delta x^2
\end{aligned}$$

DOI: 10.1002/nme

Therefore, the correlation coefficient between $M(x+\Delta x)$ and $M(x)$, omitting terms of order Δx^3 or higher, is

$$
\begin{aligned}
&\mathrm{Corr}\left[M(x+\Delta x), M(x)\right] \\
&= \frac{\mathrm{Cov}\left[M(x+\Delta x), M(x)\right]}{\sqrt{\mathrm{Var}\left[M(x+\Delta x)\mathrm{Var}\left[M(x)\right]\right]}} \\
&\approx 1 - \frac{\mathrm{Cov}\left[M'(x), M(x)\right]}{\mathrm{Var}\left[M(x)\right]}\Delta x \\
&\quad + \frac{3\ \mathrm{Cov}\left[M'(x), M(x)\right] - \mathrm{Var}\left[M(x)\right]\left\{\mathrm{Var}\ \left[M'(x)\right] + \mathrm{Cov}\left[M''(x), M(x)\right]\right\}}{2\mathrm{Var}\left[M(x)\right]^2}\Delta x^2 \\
&\quad + \frac{\mathrm{Cov}\left[M'(x), M(x)\right]}{\mathrm{Var}\left[M(x)\right]}\Delta x - \frac{\mathrm{Cov}\left[M'(x), M(x)\right]^2}{\mathrm{Var}\left[M(x)\right]^2}\Delta x^2 + \frac{\mathrm{Cov}\left[M''(x), M(x)\right]}{2\mathrm{Var}\left[M(x)\right]}\Delta x^2 \\
&= 1 - \frac{\mathrm{Var}\ \left[M'(x)\right]\mathrm{Var}\left[M(x)\right] - \mathrm{Cov}\left[M'(x), M(x)\right]^2}{2\mathrm{Var}\left[M(x)\right]^2}\Delta x^2 \\
&= 1 - \frac{1 - \mathrm{Corr}\left[M'(x), M(x)\right]^2}{2\mathrm{Var}\left[M(x)\right]/\mathrm{Var}\ \left[M'(x)\right]}\Delta x^2.
\end{aligned}
$$

ACKNOWLEDGEMENTS

This work was supported by the AFOSR MURI on Uncertainty Quantification, grant FA9550-09-0613, program manager F. Fahroo. The authors thank D.B.P. Huynh for his development of the acoustic horn finite element and reduced basis models. The authors thank G. Kennedy and J. Martins for supplying the aerostructural wing models.

REFERENCES

1. Alexandrov NM, Lewis RM, Gumbert CR, Green LL, Newman PA. Approximation and Model Management in Aerodynamic Optimization with Variable Fidelity Models. *AIAA Journal of Aircraft* 2001; **38**(6):1093–1101.
2. Booker AJ, Dennis JE, Frank PD, Serafini DB, Torczon V, Trosset MW. A Rigorous Framework for Optimization of Expensive Functions by Surrogates. *Structural and Multidisciplinary Optimization* 1999; **17**(1):1–13.
3. Eldred MS, Giunta AA, Collis SS. Second-Order Corrections for Surrogate-Based Optimization with Model Hierarchies. *10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Albany, NY, 2004; 1–15.
4. Dribusch C, Missoum S, Beran P. A Multifidelity Approach for the Construction of Explicit Decision Boundaries: Application to Aeroelasticity. *Structural and Multidisciplinary Optimization* 2010; **42**(5):693–705.
5. Li J, Xiu D. Evaluation of Failure Probability via Surrogate Models. *Journal of Computational Physics* 2010; **229**(23):8966–8980.
6. Ng LWT, Eldred MS. Multifidelity Uncertainty Quantification Using Nonintrusive Polynomial Chaos and Stochastic Collocation. *14th AIAA Non-Deterministic Approaches Conference*, Honolulu, HI, 2012; 1–15.
7. Kennedy MC, O'Hagan A. Predicting the Output from a Complex Computer Code When Fast Approximations Are Available. *Biometrika* 2000; **87**(1):1–13.
8. Koutsourelakis PS. Accurate Uncertainty Quantification Using Inaccurate Computational Models. *SIAM Journal on Scientific Computing* 2009; **31**(5):3274–3300.
9. Hammersley JM, Handscomb DC. *Monte Carlo Methods*. Methuen: London, UK, 1964.
10. Nelson BL. On Control Variate Estimators. *Computers & Operations Research* 1987; **14**(3):219–225.
11. Giles MB. Multilevel Monte Carlo Path Simulation. *Operations Research* 2008; **56**(3):607–617.
12. Speight A. A Multilevel Approach to Control Variates. *Journal of Computational Finance* 2009; **12**(4):3–27.
13. Teckentrup AL, Scheichl R, Giles MB, Ullmann E. Further Analysis of Multilevel Monte Carlo Methods for Elliptic PDEs with Random Coefficients. *Numerische Mathematik* 2013; **125**(3):569–600.
14. Boyaval S, Lelièvre T. A Variance Reduction Method for Parametrized Stochastic Differential Equations Using the Reduced Basis Paradigm. *Communications in Mathematical Sciences* 2010; **8**(3):735–762.
15. Boyaval S. A Fast Monte-Carlo Method with a Reduced Basis of Control Variates Applied to Uncertainty Propagation and Bayesian Estimation. *Computer Methods in Applied Mechanics and Engineering* 2012; **241**:190–205.
16. Tracey B, Wolpert D, Alonso JJ. Using Supervised Learning to Improve Monte Carlo Integral Estimation. *AIAA Journal* 2013; **51**(8):2015–2023.

DOI: 10.1002/nme

17. Emsermann M, Simon B. Improving Simulation Efficiency with Quasi Control Variates. *Stochastic Models* 2002; **18**(3):425–448.
18. Pasupathy R, Schmeiser BW, Taaffe MR, Wang J. Control-Variate Estimation Using Estimated Control Means. *IIE Transactions* 2012; **44**(5):381–385.
19. Spall JC. *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. John Wiley & Sons: Hoboken, NJ, 2003.
20. Audet C, Dennis Jr. JE. Mesh Adaptive Direct Search Algorithms for Constrained Optimization. *SIAM Journal on Optimization* 2006; **17**(1):188–217.
21. Kelley CT. *Implicit Filtering*. SIAM: Philadelphia, PA, 2011.
22. Conn AR, Scheinberg K, Toint PL. A Derivative Free Optimization Algorithm in Practice. *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, St. Louis, MO, 1998; 129–139.
23. Powell MJD. The NEWUOA Software for Unconstrained Optimization Without Derivatives. In *Large-Scale Nonlinear Optimization*. Springer, 2006; 255–297.
24. Powell MJD. A Direct Search Optimization Method That Models the Objective and Constraint Functions by Linear Interpolation. *Advances in Optimization and Numerical Analysis* 1994; **7**:51–67.
25. Conn AR, Scheinberg K, Vicente LN. *Introduction to Derivative-Free Optimization*. SIAM: Philadelphia, PA, 2009.
26. Powell MJD. Least Frobenius Norm Updating of Quadratic Models that Satisfy Interpolation Conditions. *Mathematical Programming* 2004; **100**(1):183–215.
27. Powell MJD. On the Use of Quadratic Models in Unconstrained Minimization Without Derivatives. *Optimization Methods and Software* 2004; **19**(3–4):399–411.
28. Eftang JL, Huynh DBP, Knezevic DJ, Patera AT. A Two-Step Certified Reduced Basis Method. *SIAM Journal of Scientific Computing* 2012; **51**(1):28–58.
29. Rozza G, Huynh DBP, Patera AT. Reduced Basis Approximation and A Posteriori Error Estimation for Affinely Parametrized Elliptic Coercive Partial Differential Equations. *Archives of Computational Methods in Engineering* 2008; **15**(3):229–275.
30. Kennedy GJ, Martins JRRA. Parallel Solution Methods for Aerostructural Analysis and Design Optimization. *13th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Fort Worth, TX, 2010; 1–19.
31. Kenway GKW, Kennedy GJ, Martins JRRA. A CAD-Free Approach to High-Fidelity Aerostructural Optimization. *13th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Fort Worth, TX, 2010; 1–18.
32. March AI. Multifidelity Methods for Multidisciplinary System Design. *Ph.D. Thesis*, Cambridge, MA, 2012.
33. Hickernell FJ, Lemieux C, Owen AB. Control Variates for Quasi-Monte Carlo. *Statistical Science* 2005; **20**(1):1–31.

DOI: 10.1002/nme