mfEGRA: Multifidelity Efficient Global Reliability Analysis through Active Learning for Failure Boundary Location

Anirban Chaudhuri,* Alexandre N. Marques†
Massachusetts Institute of Technology, Cambridge, MA, 02139, USA
Karen E. Willcox‡
University of Texas at Austin, Austin, TX, 78712, USA

Abstract

This paper develops mfEGRA, a multifidelity active learning method using data-driven adaptively refined surrogates for failure boundary location in reliability analysis. This work addresses the issue of prohibitive cost of reliability analysis using Monte Carlo sampling for expensive-to-evaluate high-fidelity models by using cheaper-to-evaluate approximations of the high-fidelity model. The method builds on the Efficient Global Reliability Analysis (EGRA) method, which is a surrogate-based method that uses adaptive sampling for refining Gaussian process surrogates for failure boundary location using a single-fidelity model. Our method introduces a two-stage adaptive sampling criterion that uses a multifidelity Gaussian process surrogate to leverage multiple information sources with different fidelities. The method combines expected feasibility criterion from EGRA with one-step lookahead information gain to refine the surrogate around the failure boundary. The computational savings from mfEGRA depends on the discrepancy between the different models, and the relative cost of evaluating the different models as compared to the high-fidelity model. We show that accurate estimation of reliability using mfEGRA leads to computational savings of ∼46% for an analytic multimodal test problem and 24% for a three-dimensional acoustic horn problem, when compared to single-fidelity EGRA. We also show the effect of using *a priori* drawn Monte Carlo samples in the implementation for the acoustic horn problem, where mfEGRA leads to computational savings of 45% for the three-dimensional case and 48% for a rarer event four-dimensional case as compared to single-fidelity EGRA.

Keywords: multi-fidelity, adaptive sampling, probability of failure, contour location, classification, Gaussian process, kriging, multiple information sources, EGRA, surrogate

1 Introduction

The presence of uncertainties in the manufacturing and operation of systems make reliability analysis critical for system safety. The reliability analysis of a system requires estimating the probability of failure, which can be computationally prohibitive when the high fidelity model is expensive to evaluate. In this work, we develop a method for efficient reliability estimation by leveraging multiple sources of information with different fidelities to build a multifidelity approximation for the limit state function.

Reliability analysis for strongly non-linear systems typically require Monte Carlo sampling that can incur substantial cost because of numerous evaluations of expensive-to-evaluate high fidelity models as seen in Figure 1 (a). There are several methods that improve the convergence rate of Monte Carlo methods to decrease computational cost through Monte Carlo variance reduction, such as, importance sampling [1, 2], cross-entropy method [3], subset simulation [4, 5], etc. However, such methods are outside the scope of this paper and will not be discussed further. Another class of methods reduce the computational cost by using approximations for the failure boundary or the entire limit state function. The popular methods that fall in the first category are first- and second-order reliability methods (FORM and SORM), which approximate

---

*Postdoctoral Associate, Department of Aeronautics and Astronautics, anirbanc@mit.edu.
†Postdoctoral Associate, Department of Aeronautics and Astronautics, noll@mit.edu.
‡Director, Oden Institute for Computational Engineering and Sciences, kwillcox@oden.utexas.edu

the failure boundary with linear and quadratic approximations around the most probable failure point [6, 7]. The FORM and SORM methods can be efficient for mildly nonlinear problems and cannot handle systems with multiple failure regions. The methods that fall in the second category reduce computational cost by replacing the high-fidelity model evaluations in the Monte Carlo simulation by cheaper evaluations from adaptive surrogates for the limit state function as seen in Figure 1 (b).

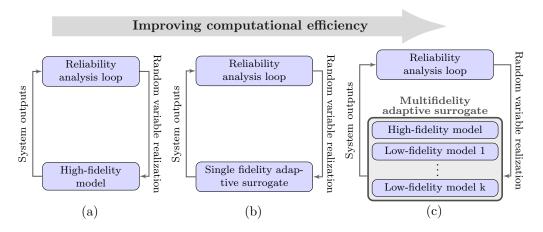**Improving computational efficiency**

Figure 1: Reliability analysis with (a) high-fidelity model, (b) single fidelity adaptive surrogate, and (c) multifidelity adaptive surrogate.

Estimating reliability requires accurately classifying samples to fail or not, which needs surrogates that accurately predict the limit state function around the failure boundary. Thus, the surrogates need to be refined only in the region of interest (in this case, around the failure boundary) and do not require global accuracy in prediction of the limit state function. The development of sequential active learning methods for refining the surrogate around the failure boundary has been addressed in the literature using only a single high-fidelity information source. Such methods fall in the same category as adaptively refining surrogates for identifying stability boundaries, contour location, classification, sequential design of experiment (DOE) for target region, etc. Typically, these methods are divided into using either Gaussian process (GP) surrogates or support vector machines (SVM). Adaptive SVM methods have been implemeted for reliability analysis and contour location [8, 9, 10]. In this work, we focus on GP-based methods (sometimes referred to as kriging-based) that use the GP prediction mean and prediction variance to develop greedy and lookahead adaptive sampling methods. Efficient Global Reliability Analysis (EGRA) adaptively refines the GP surrogate around the failure boundary by sequentially adding points that have maximum expected feasibility [11]. A weighted integrated mean square criterion for refining the kriging surrogate was developed by Picheny et al. [12]. Echard et al. [13] proposed an adaptive Kriging method that refines the surrogate in the restricted set of samples defined by a Monte Carlo simulation. Dubourg et al. [14] proposed a population-based adaptive sampling technique for refining the kriging surrogate around the failure boundary. One-step lookahead strategies for GP surrogate refinement for estimating probability of failure was proposed by Bect et al. [15] and Chevalier et al. [16]. A review of some surrogate-based methods for reliability analysis can be found in Ref. [17]. However, all the methods mentioned above use a single source of information, which is the high-fidelity model as illustrated in Figure 1 (b). This work presents a novel multifidelity active learning method that adaptively refines the surrogate around the limit state function failure boundary using multiple sources of information, thus, further reducing the active learning computational effort as seen in Figure 1 (c).

For several applications, in addition to an expensive high-fidelity model, there are potentially cheaper lower fidelity models, such as, simplified physics models, coarse-grid models, data-fit models, reduced order models, etc. that are readily available or can be built. This necessitates the development of multifidelity methods that can take advantage of these multiple information sources [18]. Various multifidelity methods have been developed in the context of GP-based Bayesian optimization [19, 20, 21]. While Bayesian optimization also uses GP models and adaptive sampling [22, 23], we note that Bayesian optimization targets a different problem to GP-based reliability analysis. In particular, the reliability analysis problem targets

the entire limit state function failure contour in the random variable space, whereas Bayesian optimization targets finding the optimal design. Thus, the sampling criteria used for failure boundary location as compared to optimization are different, and the corresponding needs and opportunities for multifidelity methods are different. In the context of reliability analysis using active learning surrogates, there are few multifidelity methods available. Dribusch et al. [24] proposed a hierarchical bi-fidelity adaptive SVM method for locating failure boundary. The recently developed CLoVER [25] method is a multifidelity active learning algorithm that uses a one-step lookahead entropy-reduction-based adaptive sampling strategy for refining GP surrogates around the failure boundary. In this work, we develop a multifidelity extension of the EGRA method [11] as EGRA has been rigorously tested on a wide range of reliability analysis problems.

We propose **mfEGRA** (multifidelity EGRA) that leverages multiple sources of information with different fidelities and cost to accelerate active learning of surrogates for failure boundary identification. For single-fidelity methods, the adaptive sampling criterion chooses where to sample next to refine the surrogate around the failure boundary. The challenge in developing a multifidelity adaptive sampling criterion is that we now have to answer two questions – (i) where to sample next, and (ii) what information source to use for evaluating the next sample. This work proposes a new adaptive sampling criterion that allows the use of multiple fidelity models. In our mfEGRA method, we combine the expected feasibility function used in EGRA with a proposed weighted lookahead information gain to define the adaptive sampling criterion for multifidelity case. We use the Kullback-Leibler (KL) divergence to quantify the information gain and derive a closed-form expression for the multifidelity GP case. The key advantage of the mfEGRA method is the reduction in computational cost compared to single-fidelity active learning methods because it can utilize additional information from multiple cheaper low-fidelity models along with the high-fidelity model information. We demonstrate the computational efficiency of the proposed mfEGRA using a multimodal analytic test problem and an acoustic horn problem with disjoint failure regions.

The rest of the paper is structured as follows. Section 2 provides the problem setup for reliability analysis using multiple information sources. Section 3 describes the details of the proposed mfEGRA method along with the complete algorithm. The effectiveness of mfEGRA is shown using an analytical multimodal test problem and an acoustic horn problem in Section 4. The conclusions are presented in Section 5.

2   Problem Setup

The inputs to the system are the N_z random variables $Z \in \Omega \subseteq \mathbb{R}^{N_z}$ with the probability density function π, where Ω denotes the random sample space. The vector of a realization of the random variables Z is denoted by $\boldsymbol{z}$.

The probability of failure of the system is $p_{\mathrm{F}} = \mathbb{P}(g(Z) > 0)$, where $g : \Omega \mapsto \mathbb{R}$ is the limit state function. In this work, without loss of generality, the failure of the system defined as $g(\boldsymbol{z}) > 0$. The failure boundary is defined as the zero contour of the limit state function, $g(\boldsymbol{z}) = 0$, and any other failure boundary, $g(\boldsymbol{z}) = c$, can be reformulated as a zero contour (i.e., $g(\boldsymbol{z}) - c = 0$).

One way to estimate the probability of failure for nonlinear systems is Monte Carlo simulation. The Monte Carlo estimate of the probability of failure $\hat{p}_{\mathrm{F}}$ is

$$\hat{p}_{\mathrm{F}} = \frac{1}{m} \sum_{i=1}^{m} \mathbb{I}_{\mathcal{G}}(\boldsymbol{z}_i), \tag{1}$$

where $\boldsymbol{z}_i, i = 1, \ldots, m$ are m samples from probability density π, $\mathcal{G} = \{\boldsymbol{z} \mid \boldsymbol{z} \in \Omega, g(\boldsymbol{z}) > 0\}$ is the failure set, and $\mathbb{I}_{\mathcal{G}} : \Omega \mapsto \{0, 1\}$ is the indicator function defined as

$$\mathbb{I}_{\mathcal{G}}(\boldsymbol{z}) = \left\{ \begin{array}{ll} 1, & \boldsymbol{z} \in \mathcal{G} \\ 0, & \text{else.} \end{array} \right. \tag{2}$$

The probability of failure estimation requires many evaluations of the expensive-to-evaluate high-fidelity model for the limit state function g, which can make reliability analysis computationally prohibitive. The computational cost can be substantially reduced by replacing the high-fidelity model evaluations with cheap-to-evaluate surrogate model evaluations. However, to make accurate estimations of $\hat{p}_{\mathrm{F}}$ using a surrogate model, the zero-contour of the surrogate model needs to approximate the failure boundary well. Adaptively

refining the surrogate around the failure boundary, while trading-off global accuracy, is an efficient way of addressing the above.

The goal of this work is to make the adaptive refinement of surrogate models around the failure boundary more efficient by using multiple models with different fidelities and costs instead of only using the high-fidelity model. We develop a multifidelity active learning method that utilizes multiple information sources to efficiently refine the surrogate to accurately locate the failure boundary. Let $g_l : \Omega \mapsto \mathbb{R}, l \in \{0, \ldots, k\}$ be a collection of $k+1$ models for g with associated cost $c_l(\boldsymbol{z})$ at location $\boldsymbol{z}$, where the subscript l denotes the information source. We define the model g_0 to be the high-fidelity model for the limit state function. The k low-fidelity models of g are denoted by $l = 1, \ldots, k$. We use a multifidelity surrogate to simultaneously approximate all information sources while encoding the correlations between them. The adaptively refined multifidelity surrogate model predictions are used for the probability of failure estimation. The Monte Carlo estimate of the probability of failure is then estimated using the refined multifidelity surrogate and is denoted here by $\hat{p}_{\mathrm{F}}^{\mathrm{MF}}$. Next we describe the multifidelity surrogate model used in this work and the multifidelity active learning method used to sequentially refine the surrogate around the failure boundary.

3 mfEGRA: Multifidelity EGRA with Information Gain

In this section, we introduce multifidelity EGRA (mfEGRA) that leverages the $k+1$ information sources to efficiently build an adaptively refined multifidelity surrogate to locate the failure boundary.

3.1 mfEGRA method overview

The proposed mfEGRA method is a multifidelity extension to the EGRA method [11]. Section 3.2 briefly describes the multifidelity GP surrogate used in this work to combine the different information sources. The multifidelity GP surrogate is built using an initial DOE and then the mfEGRA method refines the surrogate using a two-stage adaptive sampling criterion that:

1. selects the next location to be sampled using an expected feasibility function as described in Section 3.3;

2. selects the information source to be used to evaluate the next sample using a weighted lookahead information gain criterion as described in Section 3.4.

The adaptive sampling criterion developed in this work enables us to use the surrogate prediction mean and the surrogate prediction variance to make the decision of where and which information source to sample next. Note that both of these quantities are available from the multifidelity GP surrogate used in this work. Section 3.5 provides the implementation details and the algorithm for the proposed mfEGRA method. Figure 2 shows a flowchart outlining the mfEGRA method.

3.2 Multifidelity Gaussian process

We use the multifidelity GP surrogate introduced by Poloczek et al. [19], which built on earlier work by Lam et al. [20], to combine information from the $k+1$ information sources into a single GP surrogate, $\widehat{g}(l, \boldsymbol{z})$, that can simultaneously approximate all the information sources. The multifidelity GP surrogate can provide predictions for any information source l and random variable realization $\boldsymbol{z}$.

The multifidelity GP is built by making two modeling choices: (1) a GP approximation for the high-fidelity model g_0 as given by $\widehat{g}(0, \boldsymbol{z}) \sim \mathrm{GP}(\mu_0, \Sigma_0)$, and (2) independent GP approximations for the model discrepancy between the high-fidelity and the lower-fidelity models as given by $\delta_l \sim \mathrm{GP}(\mu_l, \Sigma_l)$ for $l = 1, \ldots, k$. μ_l denotes the mean function and Σ_l denotes the covariance kernel for $l = 0, \ldots, k$.

Then the surrogate for model l is constructed by using the definition $\widehat{g}(l, \boldsymbol{z}) = \widehat{g}(0, \boldsymbol{z}) + \delta_l(\boldsymbol{z})$. These modeling choices lead to the surrogate model $\widehat{g} \sim \mathrm{GP}(\mu_{\mathrm{pr}}, \Sigma_{\mathrm{pr}})$ with prior mean function μ_{pr} and prior covariance kernel Σ_{pr}. The priors for $l = 0$ are

$$
\begin{aligned}
\mu_{\mathrm{pr}}(0, \boldsymbol{z}) &= \mathbb{E}[\widehat{g}(0, \boldsymbol{z})] = \mu_0(\boldsymbol{z}), \\
\Sigma_{\mathrm{pr}}((0, \boldsymbol{z}), (l', \boldsymbol{z}')) &= \mathrm{Cov}\left(\widehat{g}(0, \boldsymbol{z}), \widehat{g}(0, \boldsymbol{z}')\right) = \Sigma_0(\boldsymbol{z}, \boldsymbol{z}'),
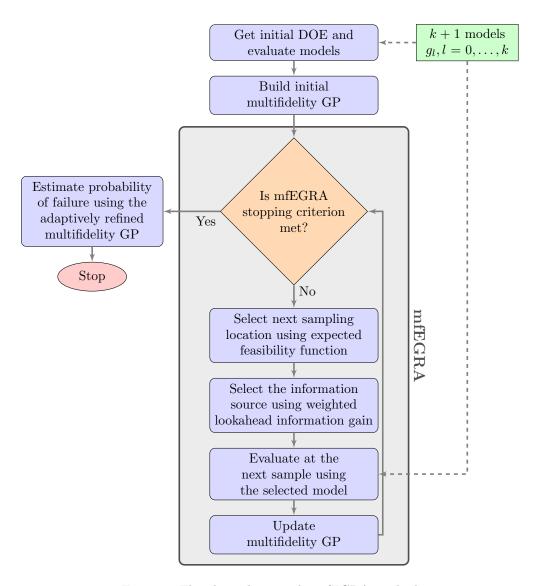\end{aligned}
\tag{3}
$$

Figure 2: Flowchart showing the mfEGRA method.

and priors for $l = 1, \ldots, k$ are

$$
\begin{aligned}
\mu_{\mathrm{pr}}(l, \boldsymbol{z}) &= \mathbb{E}[\widehat{g}(l, \boldsymbol{z})] = \mathbb{E}[\widehat{g}(0, \boldsymbol{z})] + \mathbb{E}[\delta_l(\boldsymbol{z})] = \mu_0(\boldsymbol{z}) + \mu_l(\boldsymbol{z}), \\
\Sigma_{\mathrm{pr}}((l, \boldsymbol{z}), (l', \boldsymbol{z}')) &= \mathrm{Cov}\left(\widehat{g}(0, \boldsymbol{z}) + \delta_l(\boldsymbol{z}), \widehat{g}(0, \boldsymbol{z}') + \delta_{l'}(\boldsymbol{z}')\right) \\
&= \Sigma_0(\boldsymbol{z}, \boldsymbol{z}') + \mathbb{1}_{l,l'} \Sigma_l(\boldsymbol{z}, \boldsymbol{z}'),
\end{aligned}
\tag{4}
$$

where $l' \in 0, \ldots, k$ and $\mathbb{1}_{l,l'}$ denotes the Kronecker delta. Once the prior mean function and the prior covariance kernels are defined using Equations (3) and (4), we can compute the posterior using standard rules of GP regression [26]. A more detailed description about the assumptions and the implementation of the multifidelity GP surrogate can be found in Ref. [19].

At any given $\boldsymbol{z}$, the surrogate model posterior distribution of $\widehat{g}(l, \boldsymbol{z})$ is defined by the normal distribution with posterior mean $\mu(l, \boldsymbol{z})$ and posterior variance $\sigma^2(l, \boldsymbol{z}) = \Sigma((l, \boldsymbol{z}), (l, \boldsymbol{z}))$. Consider that n samples $\{[l^i, \boldsymbol{z}^i]\}_{i=1}^n$ have been evaluated and these samples are used to fit the present multifidelity GP surrogate. Note that $[l, \boldsymbol{z}]$ is the augmented vector of inputs to the multifidelity GP. Then the surrogate is refined around the failure boundary by sequentially adding samples. The next sample $\boldsymbol{z}^{n+1}$ and the next information source l^{n+1} used to refine the surrogate are found using the two-stage adaptive sampling method mfEGRA as described below.

3.3 Location selection: Maximize expected feasibility function

The first stage of mfEGRA involves selecting the next location z^{n+1} to be sampled. The expected feasibility function (EFF), which was used as the adaptive sampling criterion in EGRA [11], is used in this work to select the location of the next sample z^{n+1}. The EFF defines the expectation of the sample lying within a band around the failure boundary (here, $\pm\epsilon(z)$ around the zero contour of the limit state function). The prediction mean $\mu(0, z)$ and the prediction variance $\sigma(0, z)$ at any z are provided by the multifidelity GP for the high-fidelity surrogate model. The multifidelity GP surrogate prediction at z is the normal distribution $\mathcal{Y}_z \sim \mathcal{N}(\mu(0, z), \sigma^2(0, z))$. Then the feasibility function at any z is defined as being positive within the ϵ-band around the failure boundary and zero otherwise as given by

$$F(z) = \epsilon(z) - \min(|y|, \epsilon(z)), \tag{5}$$

where y is a realization of $\mathcal{Y}_z$. The EFF is defined as the expectation of being within the ϵ-band around the failure boundary as given by

$$\mathbb{E}_{\mathcal{Y}_z}[F(z)] = \int_{-\epsilon(z)}^{\epsilon(z)} (\epsilon(z) - |y|)\mathcal{Y}_z(y)\mathrm{d}y. \tag{6}$$

We will use $\mathbb{E}[F(z)]$ to denote $\mathbb{E}_{\mathcal{Y}_z}[F(z)]$ in the rest of the paper. The integration in Equation (6) can be solved analytically to obtain [11]

$$
\begin{aligned}
\mathbb{E}[F(z)] = \mu(0, z) &\left[2\Phi\left(\frac{-\mu(0, z)}{\sigma(0, z)}\right) - \Phi\left(\frac{-\epsilon(z) - \mu(0, z)}{\sigma(0, z)}\right) - \Phi\left(\frac{\epsilon(z) - \mu(0, z)}{\sigma(0, z)}\right) \right] \\
- \sigma(0, z) &\left[2\phi\left(\frac{-\mu(0, z)}{\sigma(0, z)}\right) - \phi\left(\frac{-\epsilon(z) - \mu(0, z)}{\sigma(0, z)}\right) - \phi\left(\frac{\epsilon(z) - \mu(0, z)}{\sigma(0, z)}\right) \right] \\
+ \epsilon(z) &\left[\Phi\left(\frac{\epsilon(z) - \mu(0, z)}{\sigma(0, z)}\right) - \Phi\left(\frac{-\epsilon(z) - \mu(0, z)}{\sigma(0, z)}\right) \right],
\end{aligned}
\tag{7}
$$

where Φ is the cumulative distribution function and ϕ is the probability density function of the standard normal distribution. Similar to EGRA [11], we define $\epsilon(z) = 2\sigma(0, z)$ to balance exploration and exploitation. As noted before, we describe the method considering the zero contour as the failure boundary for convenience but the proposed method can be used for locating failure boundary at any contour level.

The location of the next sample is selected by maximizing the EFF as given by

$$z^{n+1} = \arg\max_{z\in\Omega} \mathbb{E}[F(z)]. \tag{8}$$

3.4 Information source selection: Maximize weighted lookahead information gain

Given the location of the next sample at z^{n+1} obtained using Equation (8), the second stage of mfEGRA selects the information source l^{n+1} to be used for simulating the next sample by maximizing the information gain. Information-gain-based approaches have been used previously for global optimization [27, 28, 21, 29], optimal experimental design [30, 31], and uncertainty propagation in coupled systems [32]. Ref. [21] used an information-gain-based approach for selecting the location and the information source for improving the global accuracy of the multifidelity GP approximation for the constraints in global optimization using a double-loop Monte Carlo sample estimate of the information gain. Our work differs from previous efforts in that we develop a weighted information-gain-based sampling strategy for failure boundary identification utilizing multiple fidelity models. In the context of information gain criterion for multifidelity GP, the two specific contributions of our work are: (1) deriving a closed-form expression for KL divergence for the multifidelity GP, which does not require double-loop Monte Carlo sampling, thus improving the robustness and decreasing the cost of estimation of the acquisition function, and (2) using weighting strategies to address the goal of failure boundary location for reliability analysis using multiple fidelity models.

The next information source is selected by using a weighted one-step lookahead information gain criterion. This adaptive sampling strategy selects the information source that maximizes the information gain in the GP surrogate prediction defined by the Gaussian distribution at any z. We measure the KL divergence between

the present surrogate predicted GP and a hypothetical future surrogate predicted GP when a particular information source is used to simulate the sample at z^{n+1} to quantify the information gain.

We represent the present GP surrogate built using the n available training samples by the subscript P for convenience as given by $\widehat{g}_P(l, z) = \widehat{g}(l, z \mid \{l^i, z^i\}_{i=1}^n)$. Then the present surrogate predicted Gaussian distribution at any z is

$$G_P(z) \sim \mathcal{N}(\mu_P(0, z), \sigma_P^2(0, z)),$$

where $\mu_P(0, z)$ is the posterior mean and $\sigma_P^2(0, z)$ is the posterior prediction variance of the present GP surrogate for the high-fidelity model built using the available training data till iteration n.

A hypothetical future GP surrogate can be understood as a surrogate built using the current GP as a generative model to create hypothetical future simulated data. The hypothetical future simulated data $y_F \sim \mathcal{N}(\mu_P(l_F, z^{n+1}), \sigma_P^2(l_F, z^{n+1}))$ is obtained from the present GP surrogate prediction at the location z^{n+1} using a possible future information source $l_F \in \{0, \ldots, k\}$. We represent a hypothetical future GP surrogate by the subscript F. Then a hypothetical future surrogate predicted Gaussian distribution at any z is

$$G_F(z|z^{n+1}, l_F, y_F) \sim \mathcal{N}(\mu_F(0, z|z^{n+1}, l_F, y_F), \sigma_F^2(0, z|z^{n+1}, l_F, y_F)).$$

The posterior mean of the hypothetical future GP is affine with respect to y_F and thus is distributed normally as given by

$$\mu_F(0, z|z^{n+1}, l_F, y_F) \sim \mathcal{N}(\mu_P(0, z), \bar{\sigma}^2(z|z^{n+1}, l_F)),$$

where $\bar{\sigma}^2(z|z^{n+1}, l_F) = (\Sigma_P((0, z), (l_F, z^{n+1})))^2/\Sigma_P((l_F, z^{n+1}), (l_F, z^{n+1}))$[19]. The posterior variance of the hypothetical future GP surrogate $\sigma_F^2(0, z|z^{n+1}, l_F, y_F)$ depends only on the location z^{n+1} and the source l_F, and can be replaced with $\sigma_F^2(0, z|z^{n+1}, l_F)$. Note that we don't need any new evaluations of the information source for constructing the future GP. The total lookahead information gain is obtained by integrating over all possible values of y_F as described below.

Since both G_P and G_F are Gaussian distributions, we can write the KL divergence between them explicitly. The KL divergence between G_P and G_F for any z is

$$
\begin{aligned}
&D_{KL}(G_P(z) \parallel G_F(z|z^{n+1}, l_F, y_F)) \\
&= \log\left(\frac{\sigma_F(0, z|z^{n+1}, l_F)}{\sigma_P(0, z)}\right) + \frac{\sigma_P^2(0, z) + (\mu_P(0, z) - \mu_F(0, z|z^{n+1}, l_F, y_F))^2}{2\sigma_F^2(0, z|z^{n+1}, l_F)} - \frac{1}{2}.
\end{aligned}
\tag{9}
$$

The total KL divergence can then be calculated by integrating $D_{KL}(G_P(z) \parallel G_F(z|z^{n+1}, l_F, y_F))$ over the entire random variable space Ω as given by

$$
\begin{aligned}
&\int_\Omega D_{KL}(G_P(z) \parallel G_F(z|z^{n+1}, l_F, y_F))\mathrm{d}z \\
&= \int_\Omega \left[\log\left(\frac{\sigma_F(0, z|z^{n+1}, l_F)}{\sigma_P(0, z)}\right) + \frac{\sigma_P^2(0, z) + (\mu_P(0, z) - \mu_F(0, z|z^{n+1}, l_F, y_F))^2}{2\sigma_F^2(0, z|z^{n+1}, l_F)} - \frac{1}{2}\right]\mathrm{d}z.
\end{aligned}
\tag{10}
$$

The total lookahead information gain for any z can then be calculated by taking the expectation of Equation (10) over all possible values of y_F as given by

$$
\begin{aligned}
D_{IG}(z^{n+1}, l_F) &= \mathbb{E}_{y_F}\left[\int_\Omega D_{KL}(G_P(z) \parallel G_F(z|z^{n+1}, l_F, y_F))\mathrm{d}z\right] \\
&= \int_\Omega \left[\log\left(\frac{\sigma_F(0, z|z^{n+1}, l_F)}{\sigma_P(0, z)}\right) + \frac{\sigma_P^2(0, z) + \mathbb{E}_{y_F}\left[(\mu_P(0, z) - \mu_F(0, z|z^{n+1}, l_F, y_F))^2\right]}{2\sigma_F^2(0, z|z^{n+1}, l_F)} - \frac{1}{2}\right]\mathrm{d}z \\
&= \int_\Omega \left[\log\left(\frac{\sigma_F(0, z|z^{n+1}, l_F)}{\sigma_P(0, z)}\right) + \frac{\sigma_P^2(0, z) + \bar{\sigma}^2(z|z^{n+1}, l_F)}{2\sigma_F^2(0, z|z^{n+1}, l_F)} - \frac{1}{2}\right]\mathrm{d}z \\
&= \int_\Omega D(z \mid z^{n+1}, l_F)\mathrm{d}z,
\end{aligned}
\tag{11}
$$

where

$$D(z \mid z^{n+1}, l_F) = \log\left(\frac{\sigma_F(0, z|z^{n+1}, l_F)}{\sigma_P(0, z)}\right) + \frac{\sigma_P^2(0, z) + \bar{\sigma}^2(z|z^{n+1}, l_F)}{2\sigma_F^2(0, z|z^{n+1}, l_F)} - \frac{1}{2}.$$

7

In practice, we choose a discrete set $\mathcal{Z} \subset \Omega$ via Latin hypercube sampling to numerically integrate Equation (11) as given by

$$D_{\mathrm{IG}}(\boldsymbol{z}^{n+1}, l_{\mathrm{F}}) = \int_{\Omega} D(\boldsymbol{z} \mid \boldsymbol{z}^{n+1}, l_{\mathrm{F}}) \mathrm{d}\boldsymbol{z} \propto \sum_{\mathcal{Z} \in \Omega} D(\boldsymbol{z} \mid \boldsymbol{z}^{n+1}, l_{\mathrm{F}}). \tag{12}$$

The total information gain for the multifidelity GP can be estimated using single-loop Monte Carlo sampling instead of double-loop Monte Carlo sampling because of the closed-form expression derived in Equation (11). This improves the robustness and decreases the cost of estimation of the acquisition function.

The total lookahead information gain evaluated using Equation (12) gives a metric of global information gain over the entire random variable space. However, we are interested in gaining more information around the failure boundary. In order to give more importance to gaining information around the failure boundary we use a weighted version of the lookahead information gain normalized by the cost of the information source. In this work, we explore three different weighting strategies: (i) no weights $w(\boldsymbol{z}) = 1$, (ii) weights defined by the EFF, $w(\boldsymbol{z}) = \mathbb{E}[F(\boldsymbol{z})]$, and (iii) weights defined by the probability of feasibility (PF), $w(\boldsymbol{z}) = \mathbb{P}[F(\boldsymbol{z})]$. The PF of the sample to lie within the $\pm\epsilon(\boldsymbol{z})$ bounds around the zero contour is

$$\mathbb{P}[F(\boldsymbol{z})] = \Phi\left(\frac{\epsilon(\boldsymbol{z}) - \mu(0, \boldsymbol{z})}{\sigma(0, \boldsymbol{z})}\right) - \Phi\left(\frac{-\epsilon(\boldsymbol{z}) - \mu(0, \boldsymbol{z})}{\sigma(0, \boldsymbol{z})}\right). \tag{13}$$

Weighting the information gain by either expected feasibility or probability of feasibility gives more importance to gaining information around the target region, in this case, the failure boundary.

The next information source l^{n+1} is selected by maximizing the weighted lookahead information gain normalized by the cost of the information source as given by

$$l^{n+1} = \underset{l \in \{0, \ldots, k\}}{\arg \max} \sum_{\boldsymbol{z} \in \Omega} \frac{1}{c_l(\boldsymbol{z})} w(\boldsymbol{z}) D_{\mathrm{IG}}(\boldsymbol{z}|\boldsymbol{z}^{n+1}, l_{\mathrm{F}} = l). \tag{14}$$

Note that the optimization problem in Equation (14) is a one-dimensional discrete variable problem. In this case, we only need $k + 1$ (number of available models) evaluations of the objective function to solve the optimization problem exactly and typically k is a small number.

3.5   Algorithm and implementation details

An algorithm describing the mfEGRA method is given in Algorithm 1. In this work, we evaluate all the models at the initial DOE. We generate the initial samples $\boldsymbol{z}$ using Latin hypercube sampling and run all the models at each of those samples to get the initial training set $\{\boldsymbol{z}^i, l^i\}_{i=1}^n$. The initial number of samples, n, can be decided based on the user's preference (in this work, we use cross-validation error). The EFF maximization problem given by Equation (8) is solved using `patternsearch` function followed by using multiple starts of a local optimizer through the `GlobalSearch` function in MATLAB. In practice, we choose a fixed set of realizations $\mathcal{Z} \in \Omega$ at which the information gain is evaluated as shown in Equation (12) for all iterations of mfEGRA. Due to the typically high cost associated with the high-fidelity model, we chose to evaluate all the $k + 1$ models when the high-fidelity model is selected as the information source and update the GP hyperparameters in our implementation. All the $k+1$ model evaluations can be done in parallel. The algorithm is stopped when the maximum value of EFF goes below 10^{-10}. However, other stopping criteria can also be explored.

Although in this work we did not encounter any case of failed model evaluations, numerical solvers can sometimes fail to provide a converged result. In the context of reliability analysis, a failed model evaluation can be treated as failure of the system (defined here as $g_l(\boldsymbol{z}) > 0$) at the particular random variable realization $\boldsymbol{z}$. One possibility to handle failed model evaluations would be to let the value of the limit state function $g_l(\boldsymbol{z})$ go to an upper limit in order to indicate failure of the system.

A potential limitation of any GP-based method is dealing with the curse of dimensionality for high-dimensional problems, where the number of samples to cover the space grows exponentially and the cost of training GPs scales as the cube of the number of samples. The multifidelity method presented here alleviates the cost of exploring the space by using cheaper low-fidelity model evaluations and restricts its queries of

the high-fidelity model to lie mostly around the failure boundary. The issue of the cost of training GPs with increasing number of samples is not addressed here but can be potentially tackled through GP sparsification techniques [33, 34]. Another strategy for reducing cost of training is through adaptive sampling strategies that exploit parallel computing. Advancements in parallel computing have led to several parallel adaptive sampling strategies for global optimization [35] and some parallel adaptive sampling methods for contour location [36, 16]. In addition, parallel methods for multifidelity adaptive sampling have increased difficulty and needs to be explored in both the fields of global optimization and contour location.

Algorithm 1 Multifidelity EGRA
___
Input: Initial DOE $\mathcal{X}_0 = \{\boldsymbol{z}^i, l^i\}_{i=1}^n$, cost of each information source c_l
Output: Refined multifidelity GP $\widehat{g}$
 1: **procedure** mfEGRA($\mathcal{X}_0$)
 2:     $\mathcal{X} = \mathcal{X}_0$ ▷ set of training samples
 3:     Build initial multifidelity GP $\widehat{g}$ using the initial set of training samples $\mathcal{X}_0$
 4:     **while** stopping criterion is not met **do**
 5:         Select next sampling location $\boldsymbol{z}^{n+1}$ using Equation (8)
 6:         Select next information source l^{n+1} using Equation (14)
 7:         Evaluate at sample $\boldsymbol{z}^{n+1}$ using information source l^{n+1}
 8:         $\mathcal{X} = \mathcal{X} \cup \{\boldsymbol{z}^{n+1}, l^{n+1}\}$
 9:         Build updated multifidelity GP $\widehat{g}$ using $\mathcal{X}$
10:         $n \leftarrow n + 1$
11:     **end while**
12:     **return** $\widehat{g}$
13: **end procedure**
___

4 Results

In this section, we demonstrate the effectiveness of the proposed mfEGRA method on an analytic multimodal test problem and two different cases for an acoustic horn application. The probability of failure is estimated through Monte Carlo simulation using the adaptively refined multifidelity GP surrogate.

4.1 Analytic multimodal test problem

The analytic test problem used in this work has two inputs and three models with different fidelities and costs. This test problem has been used before in the context of reliability analysis in Ref. [11]. The high-fidelity model of the limit state function is

$$g_0(\boldsymbol{z}) = \frac{(z_1^2 + 4)(z_2 - 1)}{20} - \sin\left(\frac{5z_1}{2}\right) - 2, \tag{15}$$

where $z_1 \sim \mathcal{U}(-4, 7)$ and $z_2 \sim \mathcal{U}(-3, 8)$ are uniformly distributed random numbers. The domain of the function is $\Omega = [-4, 7] \times [-3, 8]$. The two low-fidelity models are

$$g_1(\boldsymbol{z}) = g_0(\boldsymbol{z}) + \sin\left(\frac{5z_1}{22} + \frac{5z_2}{44} + \frac{5}{4}\right), \tag{16}$$

$$g_2(\boldsymbol{z}) = g_0(\boldsymbol{z}) + 3\sin\left(\frac{5z_1}{11} + \frac{5z_2}{11} + \frac{35}{11}\right). \tag{17}$$

The cost of each fidelity model is taken to be constant over the entire domain and is given by $c_0 = 1, c_1 = 0.01$ and $c_2 = 0.001$. In this case, there is no noise in the observations from the different fidelity models. The failure boundary is defined by the zero contour of the limit state function ($g_0(\boldsymbol{z}) = 0$) and the failure of the system is defined by $g_0(\boldsymbol{z}) > 0$. Figure 3 shows the contour plot of $g(\boldsymbol{z})$ for the three models used for the analytic test problem along with the failure boundary for each of them.
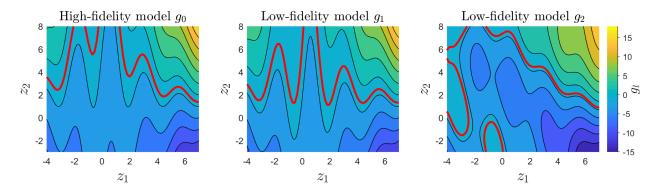
Figure 3: Contours of $g_l(\boldsymbol{z})$ using the three fidelity models for the analytic test problem. Solid red line represents the zero contour that denotes the failure boundary.

We use an initial DOE of size 10 generated using Latin hypercube sampling. All the models are evaluated at these 10 samples to build the initial multifidelity surrogate. The reference probability of failure is estimated to be $\hat{p}_F = 0.3021$ using 10^6 Monte Carlo samples of g_0 model. The relative error in probability of failure estimate using the adaptively refined multifidelity GP surrogate, defined by $|\hat{p}_F - \hat{p}_F^{MF}|/\hat{p}_F$, is used to assess the accuracy and computational efficiency of the proposed method. We repeat the calculations for 100 different initial DOEs to get the confidence bands on the results.

We first compare the accuracy of the method when different weights are used for the information gain criterion in mfEGRA as seen in Figure 4. We can see that using weighted information gain (both EFF and PF) performs better than the case when no weights are used when comparing the error confidence bands. EFF-weighted information gain leads to only marginally lower errors in this case as compared to PF-weighted information gain. Since we don't see any significant advantage of using PF as weights and we use the EFF-based criterion to select the sample location, we propose using EFF-weighted information gain to make the implementation more convenient. Note that for other problems, it is possible that PF-weighted information gain may be better. From hereon, mfEGRA is used with the EFF-weighted information gain.
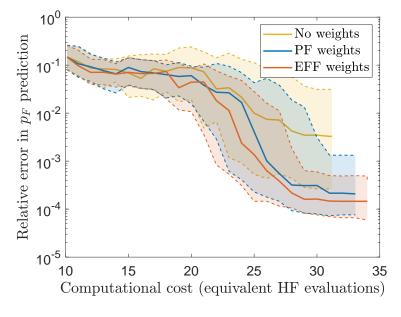

Figure 4: Effect of different weights for information gain criterion in mfEGRA for analytic test problem in terms of convergence of relative error in p_F prediction (shown in log-scale) for 100 different initial DOEs. Solid lines represent the median and dashed lines represent the 25 and 75 percentiles.

The comparison of mfEGRA with single-fidelity EGRA shows considerable improvement in accuracy at substantially lower computational cost as seen in Figure 5. In this case, to reach a median relative error of below 10^{-3} in p_F prediction, mfEGRA requires a computational cost of 26 compared to EGRA that requires a computational cost of 48 ($\sim$46% reduction). Note that we start both cases with the same 100 sets of initial samples. We also note that the original paper for the EGRA method [11] reports a computational cost of 35.1 for the mean relative error from 20 different initial DOEs to reach below 5×10^{-3}. We report the computational cost for the EGRA algorithm to reach a median relative error from 100 different initial DOEs below 10^{-3} to be 48 (in our case, the computational cost for the median relative error for EGRA to reach below 5×10^{-3} is 40). The difference in results can be attributed to the different sets of initial DOEs, the GP implementations, different statistics of reported results, and different probability distributions used for the random variables.
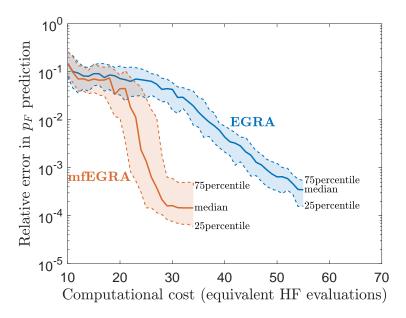

Figure 5: Comparison of mfEGRA vs single-fidelity EGRA for analytic test problem in terms of convergence of relative error in p_F prediction (shown in log-scale) for 100 different initial DOEs.

Figure 6 shows the evolution of the expected feasibility function and the weighted lookahead information gain, which are the two stages of the adaptive sampling criterion used in mfEGRA. These metrics along with the relative error in probability of failure estimate can be used to define an efficient stopping criterion, specifically when the adaptive sampling needs to be repeated for different sets of parameters (e.g., in reliability-based design optimization). Figure 7 shows the progress of mfEGRA at several iterations for a particular initial DOE. mfEGRA explores most of the domain using the cheaper g_1 and g_2 models in this case. The algorithm is stopped after 69 iterations when the expected feasibility function reached below 10^{-10}; we can see that the surrogate contour accurately traces the true failure boundary defined by the high-fidelity model. As noted before, we evaluate all the three models when the high-fidelity model is selected as the information source. In this case, mfEGRA makes a total of 21 evaluations of g_0, 77 evaluations of g_1, and 23 evaluations of g_2 including the initial DOE, to reach a value of EFF below 10^{-10}.

4.2 Acoustic horn

We demonstrate the effectiveness of mfEGRA for the reliability analysis of an acoustic horn for a three-dimensional case and a rarer event probability of failure four-dimensional case. The acoustic horn model used in this work has been used in the context of robust optimization by Ng et al. [37] An illustration of the acoustic horn is shown in Figure 8.
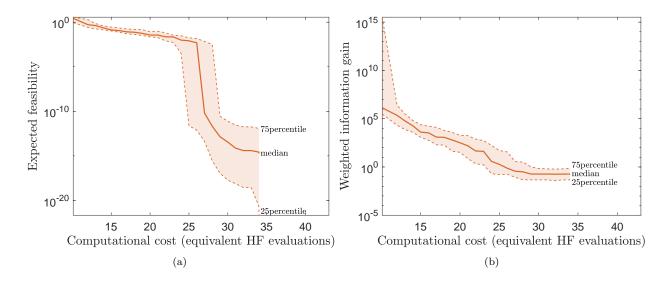
Figure 6: Evolution of adaptive sampling criteria (a) expected feasibility function, and (b) weighted information gain used in mfEGRA for 100 different initial DOEs.
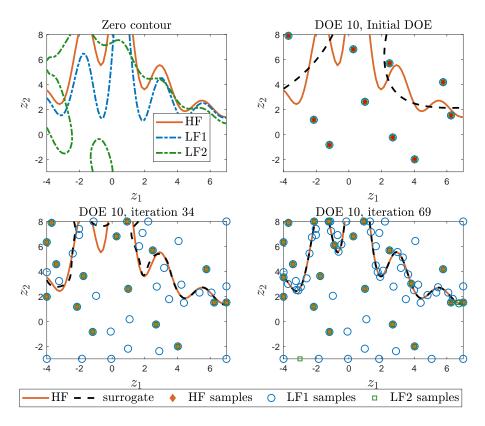

Figure 7: Progress of mfEGRA at several iterations showing the surrogate prediction and the samples from different models for a particular initial DOE. HF refers to high-fidelity model g_0, LF1 refers to low-fidelity model g_1, and LF2 refers to low-fidelity model g_2.

4.2.1 Three-dimensional case

The inputs to the system are the three random variables listed in Table 1.

Table 1: Random variables used in the three-dimensional acoustic horn problem.

Random variable	Description	Distribution	Lower bound	Upper bound	Mean	Standard deviation
k	wave number	Uniform	1.3	1.5	–	–
Z_u	upper horn wall impedance	Normal	–	–	50	3
Z_l	lower horn wall impedance	Normal	–	–	50	3

Figure 8: Two-dimensional acoustic horn geometry with $a = 0.5, b = 3, L = 5$ and shape of the horn flare described by six equally-spaced half-widths $b_1 = 0.8, b_2 = 1.2, b_3 = 1.6, b_4 = 2, b_5 = 2.3, b_6 = 2.65$. [37]

The output of the model is the reflection coefficient s, which is a measure of the horn's efficiency. We define the failure of the system to be $s(\boldsymbol{z}) > 0.1$. The limit state function is defined as $g(\boldsymbol{z}) = s(\boldsymbol{z}) - 0.1$, which defines the failure boundary as $g(\boldsymbol{z}) = 0$. We use a two-dimensional acoustic horn model governed by the non-dimensional Helmholtz equation. In this case, a finite element model of the Helmholtz equation is the high-fidelity model g_0 with 35895 nodal grid points. The low-fidelity model g_1 is a reduced basis model with $N = 100$ basis vectors [37, 38]. In this case, the cost of evaluating the low-fidelity model is 40 times faster than evaluating the high-fidelity model. The cost of evaluating the different models is taken to be constant over the entire random variable space. A more detailed description of the acoustic horn models used in this work can be found in Ref. [37].

The reference probability of failure is estimated to be $p_\mathrm{F} = 0.3812$ using 10^5 Monte Carlo samples of the high-fidelity model. We repeat the mfEGRA and the single-fidelity EGRA results using 10 different initial DOEs with 10 samples in each (generated using Latin hypercube sampling) to get the confidence bands on the results. The comparison of convergence of the relative error in the probability of failure is shown in Figure 9 for mfEGRA and single-fidelity EGRA. In this case, mfEGRA needs 19 equivalent high-fidelity solves to reach a median relative error value of below 10^{-3} as compared to 25 required by single-fidelity EGRA leading to 24% reduction in computational cost. The reduction in computational cost using mfEGRA is driven by the discrepancy between the models and the relative cost of evaluating the models. In the acoustic horn case, we see computational savings of 24% as compared to around 46% seen in the analytic test problem in Section 4.1.

This can be explained by the substantial difference in relative costs – 40 times cheaper low-fidelity model for the acoustic horn problem as compared to two low-fidelity models that are 100-1000 times cheaper than the high-fidelity model for the analytic test problem. The evolution of the mfEGRA adaptive sampling criteria can be seen in Figure 10.
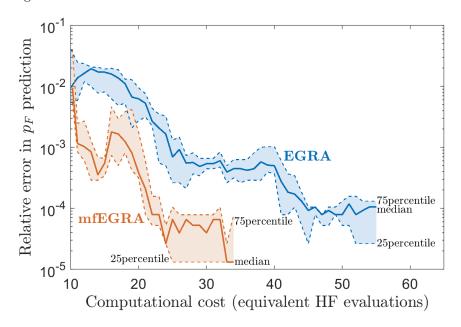

Figure 9: Comparing relative error in the estimate of probability of failure (shown in log-scale) using mfEGRA and single-fidelity EGRA for the three-dimensional acoustic horn application with 10 different initial DOEs.
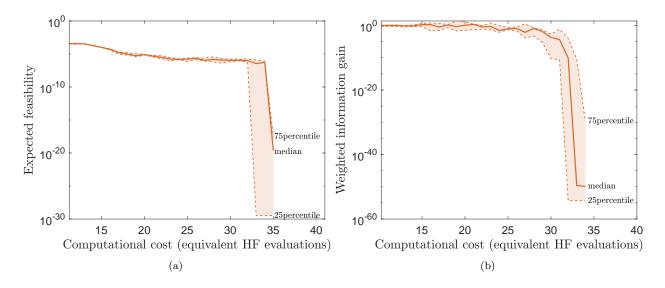

Figure 10: Evolution of adaptive sampling criteria (a) expected feasibility function, and (b) weighted information gain for the three-dimensional acoustic horn application with 10 different initial DOEs.

Figure 11 shows that classification of the Monte Carlo samples using the high-fidelity model and the adaptively refined surrogate model for a particular initial DOE lead to very similar results. It also shows that in the acoustic horn application there are two disjoint failure regions and the method is able to accurately capture both failure regions. The location of the samples from the different models when mfEGRA is used to

refine the multifidelity GP surrogate for a particular initial DOE can be seen in Figure 12. The figure shows that most of the high-fidelity samples are selected around the failure boundary. For this DOE, mfEGRA requires 31 evaluations of the high-fidelity model and 76 evaluations of the low-fidelity model to reach an EFF value below 10^{-10}.
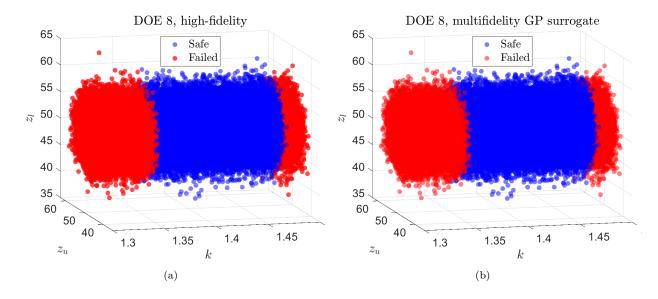

Figure 11: Classification of Monte Carlo samples using (a) high-fidelity model, and (b) the final refined multifidelity GP surrogate for a particular initial DOE using mfEGRA for the three-dimensional acoustic horn problem.
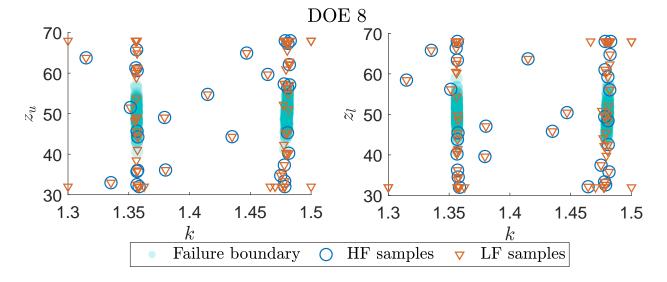

Figure 12: Location of samples from different fidelity models using mfEGRA for the three-dimensional acoustic horn problem for a particular initial DOE. The cloud of points are the high-fidelity Monte Carlo samples near the failure boundary.

Similar to the work in Refs. [13, 15], EGRA and mfEGRA can also be implemented by limiting the search space for adaptive sampling location in Equation (8) to the set of Monte Carlo samples (here, 10^5) drawn from the given random variable distribution. The convergence of relative error in probability of

failure estimate using this method improves for both mfEGRA and single-fidelity EGRA as can be seen in Figure 13. In this case, mfEGRA requires 12 equivalent high-fidelity solves as compared to 22 high-fidelity solves required by single-fidelity EGRA to reach a median relative error below 10^{-3} leading to computational savings of around 45%.
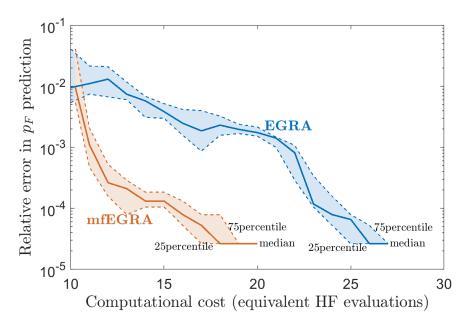

Figure 13: Comparing relative error in the estimate of probability of failure (shown in log-scale) using mfEGRA and single-fidelity EGRA by limiting the search space for adaptive sampling location to a set of Monte Carlo samples drawn from the given random variable distribution for the three-dimensional acoustic horn application with 10 different initial DOEs.

4.2.2 Four-dimensional case

For the four-dimensional acoustic horn problem, the inputs to the system are the three random variables used before along with a random variable ξ defined by a truncated normal distribution representing manufacturing uncertainty as listed in Table 2. The parameters defining the geometry of the acoustic horn (see Figure 8) are now given by $b_i + \xi, i = 1, \ldots, 6$ to account for manufacturing uncertainty. In this case, we define failure of the system to be $s(\boldsymbol{z}) > 0.16$ to make the failure a rarer event. The limit state function is defined as $g(\boldsymbol{z}) = s(\boldsymbol{z}) - 0.16$, which defines the failure boundary as $g(\boldsymbol{z}) = 0$. The reference probability of failure is estimated to be $p_F = 7.2 \times 10^{-3}$ using 10^5 Monte Carlo samples of the high-fidelity model. Note that the p_F in the four-dimensional case is two orders of magnitude lower than the three-dimensional case. The complexity of the problem increases because of the higher dimensionality as well as the rarer event probability of failure to be estimated.

In this case, we present the results for EGRA and mfEGRA implemented by limiting the search space to *a priori* Monte Carlo samples (here, 10^5) drawn from the given random variable distribution. Note that the lower probability of failure estimation required here necessitates the use of *a priori* drawn Monte Carlo samples to efficiently achieve the required accuracy. The computational efficiency can be further improved by combining EGRA and mfEGRA with Monte Carlo variance reduction techniques, especially for problems with even lower probabilities of failure. We repeat the mfEGRA and the single-fidelity EGRA results using 10 different initial DOEs with 15 samples in each (generated using Latin hypercube sampling) to get the confidence bands on the results. The comparison of convergence of the relative error in the probability of failure is shown in Figure 14 for mfEGRA and single-fidelity EGRA. In this case, mfEGRA requires 25 equivalent high-fidelity solves as compared to 48 high-fidelity solves required by single-fidelity EGRA to reach a median relative error below 10^{-3} leading to computational savings of around 48%.

Table 2: Random variables used in the four-dimensional acoustic horn problem.

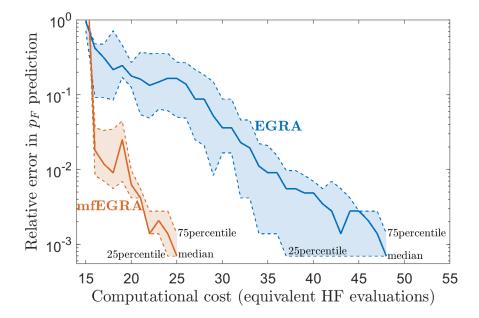| Random variable	Description	Distribution	Lower bound	Upper bound	Mean	Standard deviation
k	wave number	Uniform	1.3	1.5	–	–
Z_u	upper horn wall impedance	Normal	–	–	50	3
Z_l	lower horn wall impedance	Normal	–	–	50	3
ξ	manufacturing uncertainty	Truncated Normal	-0.1	0.1	0	0.05

Figure 14: Comparing relative error in the estimate of probability of failure (shown in log-scale) using mfEGRA and single-fidelity EGRA by limiting the search space for adaptive sampling location to a set of Monte Carlo samples drawn from the given random variable distribution for the four-dimensional acoustic horn application with 10 different initial DOEs.

5 Concluding remarks

This paper introduces the mfEGRA (multifidelity EGRA) method that refines the surrogate to accurately locate the limit state function failure boundary (or any contour) while leveraging multiple information sources with different fidelities and costs. The method selects the next location based on the expected feasibility function and the next information source based on a weighted one-step lookahead information gain criterion to refine the multifidelity GP surrogate of the limit state function around the failure boundary.

We show through three numerical examples that mfEGRA efficiently combines information from different models to reduce computational cost. The mfEGRA method leads to computational savings of ~46% for a multimodal test problem and 24% for a three-dimensional acoustic horn problem over the single-fidelity EGRA method when used for estimating the probability of failure. The mfEGRA method when

17

implemented by restricting the search-space to *a priori* drawn Monte Carlo samples showed even more computational efficiency with 45% reduction in computational cost compared to single-fidelity method for the three-dimensional acoustic horn problem. We see that using *a priori* drawn Monte Carlo samples improves the efficiency of both EGRA and mfEGRA, and the importance is further highlighted through the four-dimensional implementation of the acoustic horn problem, which requires estimating a rarer event probability of failure. For the four-dimensional acoustic horn problem, mfEGRA leads to computational savings of 48% as compared to the single-fidelity method. The driving factors for the reduction in computational cost for the method are the discrepancy between the high- and low-fidelity models, and the relative cost of the low-fidelity models compared to the high-fidelity model. These information are directly encoded in the mfEGRA adaptive sampling criterion helping it make the most efficient decision.

Acknowledgements

References

[1] Melchers, R., "Importance sampling in structural systems," *Structural Safety*, Vol. 6, No. 1, 1989, pp. 3–10.

[2] Liu, J. S., *Monte Carlo strategies in scientific computing*, Springer Science & Business Media, 2008.

[3] Kroese, D. P., Rubinstein, R. Y., and Glynn, P. W., "The cross-entropy method for estimation," *Handbook of Statistics*, Vol. 31, Elsevier, 2013, pp. 19–34.

[4] Au, S.-K. and Beck, J. L., "Estimation of small failure probabilities in high dimensions by subset simulation," *Probabilistic Engineering Mechanics*, Vol. 16, No. 4, 2001, pp. 263–277.

[5] Papaioannou, I., Betz, W., Zwirglmaier, K., and Straub, D., "MCMC algorithms for subset simulation," *Probabilistic Engineering Mechanics*, Vol. 41, 2015, pp. 89–103.

[6] Hohenbichler, M., Gollwitzer, S., Kruse, W., and Rackwitz, R., "New light on first-and second-order reliability methods," *Structural Safety*, Vol. 4, No. 4, 1987, pp. 267–284.

[7] Rackwitz, R., "Reliability analysis–a review and some perspectives," *Structural Safety*, Vol. 23, No. 4, 2001, pp. 365–395.

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

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

[10] Lecerf, M., Allaire, D., and Willcox, K., "Methodology for dynamic data-driven online flight capability estimation," *AIAA Journal*, Vol. 53, No. 10, 2015, pp. 3073–3087.

[11] Bichon, B. J., Eldred, M. S., Swiler, L. P., Mahadevan, S., and McFarland, J. M., "Efficient global reliability analysis for nonlinear implicit performance functions," *AIAA Journal*, Vol. 46, No. 10, 2008, pp. 2459–2468.

[12] Picheny, V., Ginsbourger, D., Roustant, O., Haftka, R. T., and Kim, N.-H., "Adaptive designs of experiments for accurate approximation of a target region," *Journal of Mechanical Design*, Vol. 132, No. 7, 2010, pp. 071008.

[13] Echard, B., Gayton, N., and Lemaire, M., "AK-MCS: an active learning reliability method combining Kriging and Monte Carlo simulation," *Structural Safety*, Vol. 33, No. 2, 2011, pp. 145–154.

[14] Dubourg, V., Sudret, B., and Bourinet, J.-M., "Reliability-based design optimization using kriging surrogates and subset simulation," *Structural and Multidisciplinary Optimization*, Vol. 44, No. 5, 2011, pp. 673–690.

[15] Bect, J., Ginsbourger, D., Li, L., Picheny, V., and Vazquez, E., "Sequential design of computer experiments for the estimation of a probability of failure," *Statistics and Computing*, Vol. 22, No. 3, 2012, pp. 773–793.

[16] Chevalier, C., Bect, J., Ginsbourger, D., Vazquez, E., Picheny, V., and Richet, Y., "Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set," *Technometrics*, Vol. 56, No. 4, 2014, pp. 455–465.

[17] Moustapha, M. and Sudret, B., "Surrogate-assisted reliability-based design optimization: a survey and a unified modular framework," *Structural and Multidisciplinary Optimization*, 2019, pp. 1–20.

[18] Peherstorfer, B., Willcox, K., and Gunzburger, M., "Survey of multifidelity methods in uncertainty propagation, inference, and optimization," *SIAM Review*, Vol. 60, No. 3, 2018, pp. 550–591.

[19] Poloczek, M., Wang, J., and Frazier, P., "Multi-information source optimization," *Advances in Neural Information Processing Systems*, 2017, pp. 4291–4301.

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

[21] Ghoreishi, S. F. and Allaire, D., "Multi-information source constrained Bayesian optimization," *Structural and Multidisciplinary Optimization*, Vol. 59, No. 3, 2019, pp. 977–991.

[22] Frazier, P. I., "A tutorial on bayesian optimization," *arXiv preprint arXiv:1807.02811*, 2018.

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

[24] Dribusch, C., Missoum, S., and Beran, P., "A multifidelity approach for the construction of explicit decision boundaries: application to aeroelasticity," *Structural and Multidisciplinary Optimization*, Vol. 42, No. 5, 2010, pp. 693–705.

[25] Marques, A., Lam, R., and Willcox, K., "Contour location via entropy reduction leveraging multiple information sources," *Advances in Neural Information Processing Systems*, 2018, pp. 5217–5227.

[26] Rasmussen, C. E. and Nickisch, H., "Gaussian processes for machine learning (GPML) toolbox," *Journal of Machine Learning Research*, Vol. 11, No. Nov, 2010, pp. 3011–3015.

[27] Villemonteix, J., Vazquez, E., and Walter, E., "An informational approach to the global optimization of expensive-to-evaluate functions," *Journal of Global Optimization*, Vol. 44, No. 4, 2009, pp. 509–534.

[28] Hennig, P. and Schuler, C. J., "Entropy search for information-efficient global optimization," *The Journal of Machine Learning Research*, Vol. 13, No. 1, 2012, pp. 1809–1837.

[29] Hernández-Lobato, J. M., Hoffman, M. W., and Ghahramani, Z., "Predictive entropy search for efficient global optimization of black-box functions," *Advances in Neural Information Processing Systems*, 2014, pp. 918–926.

[30] Huan, X. and Marzouk, Y. M., "Simulation-based optimal Bayesian experimental design for nonlinear systems," *Journal of Computational Physics*, Vol. 232, No. 1, 2013, pp. 288–317.

[31] Villanueva, D. and Smarslok, B. P., "Using Expected Information Gain to Design Aerothermal Model Calibration Experiments," *17th AIAA Non-Deterministic Approaches Conference, Kissimmee, FL, USA*, 2015.

[32] Chaudhuri, A., Lam, R., and Willcox, K., "Multifidelity uncertainty propagation via adaptive surrogates in coupled multidisciplinary systems," *AIAA Journal*, 2018, pp. 235–249.

[33] Williams, C. K. and Rasmussen, C. E., *Gaussian processes for machine learning*, Vol. 2, MIT press Cambridge, MA, 2006.

[34] Burt, D., Rasmussen, C. E., and Van Der Wilk, M., "Rates of Convergence for Sparse Variational Gaussian Process Regression," *International Conference on Machine Learning*, 2019, pp. 862–871.

[35] Haftka, R. T., Villanueva, D., and Chaudhuri, A., "Parallel surrogate-assisted global optimization with expensive functions–a survey," *Structural and Multidisciplinary Optimization*, Vol. 54, No. 1, 2016, pp. 3–13.

[36] Viana, F. A., Haftka, R. T., and Watson, L. T., "Sequential sampling for contour estimation with concurrent function evaluations," *Structural and Multidisciplinary Optimization*, Vol. 45, No. 4, 2012, pp. 615–618.

[37] Ng, L. W. and Willcox, K. E., "Multifidelity approaches for optimization under uncertainty," *International Journal for Numerical Methods in Engineering*, Vol. 100, No. 10, 2014, pp. 746–772.

[38] Eftang, J. L., Huynh, D., Knezevic, D. J., and Patera, A. T., "A two-step certified reduced basis method," *Journal of Scientific Computing*, Vol. 51, No. 1, 2012, pp. 28–58.