CAMERA: A Method for Cost-aware, Adaptive, Multifidelity, Efficient Reliability Analysis

S. Ashwin Renganathan

The University of Utah,1495 E 100 S, Salt Lake City, 84121, UT, USA

Vishwas Rao

Argonne National Laboratory,9700 S Cass Ave, Lemont, 60439, IL, USA

Ionel M. Navon

Florida State University,600 W College Ave, Tallahassee, 32306, FL, USA

Abstract

Estimating probability of failure in aerospace systems is a critical requirement for flight certification and qualification. Failure probability estimation involves resolving tails of probability distribution, and Monte Carlo sampling methods are intractable when expensive high-fidelity simulations have to be queried. We propose a method to use models of multiple fidelities that trade accuracy for computational efficiency. Specifically, we propose the use of multifidelity Gaussian process models to efficiently fuse models at multiple fidelity, thereby offering a cheap surrogate model that emulates the original model at all fidelities. Furthermore, we propose a novel sequential *acquisition function*-based experiment design framework that can automatically select samples from appropriate fidelity models to make predictions about quantities of interest in the highest fidelity. We use our proposed approach in an importance sampling setting and demonstrate our method on the failure level set estimation and probability estimation on synthetic test functions as well as two real-world applications, namely, the reliability analysis of a gas turbine engine blade using a finite element method and a transonic aerodynamic wing test case using Reynolds-averaged Navier–Stokes equations. We demonstrate that our method predicts the failure boundary and probability more accurately and computationally efficiently while using varying fidelity models compared with using just a single expensive high-fidelity model.

Keywords: Gaussian process regression, sequential experiment design, reliability analysis, aircraft design and certification

1. Introduction

Flight certification by analysis is a new paradigm in which airworthiness certificates for aircraft and handling qualities compliance for aircraft subsystems can be obtained via calculations with quantifiable accuracy.[1] This opens up possibilities for leveraging

---

[1]See §25.21 of Chapter 1, Title 14 in the electronic Code for Regulations of the Federal Aviation Administrations.

physics-based high-fidelity models in aircraft design, certification, and qualification. However, the high computational costs associated with high-fidelity models necessitates the development of advanced mathematical methods that judiciously spend computational resources on problems where such models may have to be queried many ($\geq \mathcal{O}(10^2) - \mathcal{O}(10^6)$) times.

One such problem we are interested in is reliability analysis [36]. Specifically, we want to estimate probabilities of failure in aerospace subsystems, whose performance are governed by models that can be queried by perturbing some control parameter. Rigorous quantification of the probability of subsystem failure is a critical requirement for flight certification. Failure probability estimation (FPE) involves accurately resolving failure boundaries (contours or level sets) in the design space and tails of input probability distributions. The popular Monte Carlo (MC) [15] sampling-based methods are intractable when expensive high-fidelity simulations have to be queried. Furthermore, the high dimensionality of the control parameters exacerbates the tractability issue. In this work we focus on a method that circumvents direct MC and instead seeks to develop a surrogate model of the mapping between the input parameters and the quantity of interest. Our hope is that a surrogate model can be constructed with far fewer queries to the expensive high-fidelity model, following which we use conventional variance-reducing sampling-based techniques for FPE.

Computational models typically include one or more customizable *fidelity* parameters that allow trading model accuracy for a gain in computational speedup. For example, the exact Navier–Stokes equations simulating fluid flows require numerical mesh resolutions on the order of the fine Kolmogorov scale, which shares an inverse nonlinear power relationship to the flow Reynolds number [58], resulting in a large number of degrees of freedom that in turn makes the computational cost of evaluating such models intractable. However, filtering out the smaller scales, leading to simplified models such as Reynolds-averaged Navier–Stokes and large eddy simulations, significantly reduces computational cost for a (typically acceptable) loss of accuracy. In general, physics-based models governing conservation laws can be computationally simplified by using coarse-grained mesh resolutions and/or fewer solver iterations for convergence (in the iterative solver) to achieve a similar effect. We exploit the availability of models with tunable fidelities and propose multifidelity methods [57, 21, 23, 56] that can judiciously utilize the cost-accuracy trade-off in models to estimate rare-event probabilities in aerospace design.

Surrogate models for conservation laws can be efficiently constructed via reduced-order models [42, 41, 43, 46] for state variables and general regression/interpolation techniques [44, 45, 37, 38] for scalar quantities of interest (QoI). In this work we are interested in a setting where a data-driven surrogate model for the QoI is adaptively constructed; that is, starting with an initial seed sample set (observations of the QoI), the surrogate model is sequentially updated with new samples that are optimally chosen to improve the prediction of the the failure boundaries in the design space. Furthermore, we account for the fact that our model of the subsystem has tunable fidelity parameters that trade predictive accuracy for computational cost. Therefore, our framework develops a surrogate model that is cost-aware while learning the relationships between the predictions between any two given fidelities. Additionally, the adaptive model construction chooses the appropriate control parameter and fidelity level to query at each step of the model-building process. For these reasons and since

2

we are interested in reliability analysis with scalar QoI in this work, we use a Gaussian process (GP) [40] regression model to learn the mapping between the QoI and the augmented input-fidelity space.

The use of surrogate models for FPE is not new. Li and Xiu [26] show that a hybrid approach of using both surrogate models and high-fidelity function evaluations can lead to computationally cheap failure probability estimates. Moreover, they show that if the chosen surrogate model is of known accuracy, their proposed approach provides a consistent failure probability estimate (i.e., approaching the true function in the limit). Extending this work, Li et al. [27] show that the same hybrid approach can be used with importance sampling with a cross-entropy method to estimate biasing distributions. However, neither [26] nor [27] addresses the choice of high-fidelity model queries necessary to construct the surrogate model (that is, they do not actively train their surrogate model); furthermore they do not consider the availability of multiple-fidelity models. Peherstorfer et al. [32] demonstrate the idea of using a surrogate model to construct biasing distributions for importance sampling (IS), but that does not include details on adaptively constructing the surrogate model.

In the context of using GP regression as surrogate models, a large body of work exists, of which we briefly review what we believe are most relevant to our work. Seminal work on using adaptive GP models for contou plots and FPE was performed by Ranjan et al. [39], who use Jones et al. [19]'s expected improvement framework to propose an improvement function that seeks to place points close (within a user-specified distance) to the target level set. They show that their acquisition function efficiently balances exploration and exploitation, and they prove that their approach asymptotically predicts the contour exactly. Bichon et al. [5] then modify the formulation of [39] to derive a closed-form expression; both [39] and [5] have essentially the same meaning. As we will show later, our framework uses the acquisition function from [5] (and hence indirectly [39])—which we consider single-fidelity acquisition functions—to develop multifidelity extensions. Indeed, our acquisition function can extend those beyond [5] and [39] to the multifidelity setting.

Other noteworthy works that propose an acquisition function for adaptive GP construction are as follows. Bect et al. [4] propose a lookahead acquisition function using GP models, called stepwise uncertainty reduction. Echard et al. [10] use a sequential sampling strategy that greedily (myopically) seeks points with high probability of being in the failure domain. Picheny et al. [34] propose a weighted integrated mean-squared prediction error criterion, where the weight function is designed to be very high in regions where the GP mean is close to the target level set with high certainty, or when the GP uncertainty is high. Gotovos et al. [14] use GP to construct an acquisition function that picks the smallest distance to the target level set from one of the confidence bounds. They leverage the theoretical results of GP-UCB [53] to derive a bound on the number of samples required to achieve a certain accuracy in resolving the level set. Dubourg et al. [9] use an adaptive Kriging surrogate model along with importance sampling to construct a quasi-optimal biasing distribution. Similar to [26], they propose a hybrid approach for FPE that combines the use of both the true limit state function and its (cheap) surrogate prediction. Wang et al. [60] use GP to approximate the forward model. Then they use Bayesian experimental design to sequentially add points toward the contour estimation. They have to use SPSA [52] to appropriately optimize their stochastic acquisition function.

3

More recently, Marques et al. [30] develop an entropy-based acquisition function that seeks to reduce the entropy of the failure contour. Furthermore, their method (CLoVER) allows for multiple information sources with varying costs of querying, and their acquisition function allows them to choose the information source. Their method, however, involve an integral in the input-dimension space, which can become intractable in higher dimensions. Recently, Cole et al. [7] improve upon [30] by developing a closed-form entropy expression for the contour that is (relatively) easier to optimize. While their method does not naturally include accounting for multiple information sources, they show that they outperform CLoVER on several problems.

Our work differs from the aforementioned work (that use GP models) in the following way. We propose the use of multifidelity GP [40] models to efficiently fuse models at multiple fidelity, where the fidelity space can be continuous or discrete. Even though [30] allows for multifidelity models, their approach depends on a Kennedy–O'Hagan [22]-like framework (see also [35]), where they learn a linear correction factor between each fidelity level and the highest-fidelity model. This has two drawbacks: (i) when the fidelity space is continuous, the approach can become intractable, and (ii) even for discrete fidelity spaces, they have to query each fidelity at a regular grid to learn the discrepancies. Our approach does not face either of these limitations. The method proposed in [24] involves the fusion of unbiased estimators from multiple fidelity models; however, the authors do not consider the choice of input and fidelities to query for an adaptive construction of the surrogate model, as we do. An adaptive construction of the surrogate model potentially results in a more judicious choice of input-fidelity pairs to query and thus a more parsimonious use of the computational resources. Furthermore, as stated previously, we propose a novel *acquisition function* that can provide a multifidelity extension to any single-fidelity acquisition function,[2], for example, [39, 5]. Our acquisition function optimization allows for sequentially selecting samples from appropriate fidelity models to make predictions about QoI in the highest fidelity. We use our proposed approach in an importance sampling setting, similar to [32], where biasing distributions are constructed with the cheap multifidelity surrogate model, to predict failure probabilities. We demonstrate our method on several synthetic test functions as well as two real-world experiments, namely, the static stress analysis of a gas turbine blade and the transonic flow past an aircraft wing. Our main contributions are summarized as follows:

1. Introduction of a generalized multifidelity acquisition framework that extends existing single-fidelity acquisition functions in GP-based approaches for reliability analysis

2. Introduction of a framework that applies to both discrete and continuous fidelity spaces

3. Theoretical demonstration that our estimates are unbiased and our surrogate model is asymptotically consistent

4. Introduction of novel multifidelity synthetic test functions that can be leveraged to benchmark multifidelity methods in general

5. Demonstration of our method on synthetic as well as real-world application problems

---

[2]although ones with a closed-form expressions suit us better

4

The remainder of the article is organized as follows. In Section 2 we provide the background and preliminaries including details about FPE and multifidelity GP models. In Section 3 we provide details of our method and discuss key theoretical properties in Section 4. In Section 5 we show results of our method demonstrated on the synthetic test functions and a real-world stress analysis of a gas turbine blade. We provide concluding remarks in Section 6.

2. Background and Preliminaries

Let $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^d$ be the uncertain input to the function $f : \mathcal{X} \to \mathbb{R}$. We assume that evaluations of $f(\mathbf{x})$ are computationally expensive. We define the *limit state function* $g : \mathbb{R} \to \mathbb{R}$ that operates on $f(\mathbf{x})$; and we define failure to occur, without loss of generality, when $g(f(\mathbf{x})) < 0$. The failure level set is then defined as

$$\mathcal{C} = \{\mathbf{x} \in \mathcal{X} : g(f(\mathbf{x})) = 0\}. \tag{1}$$

For example, for a given failure threshold $a \in \mathbb{R}$, $g(\mathbf{x}) = f(\mathbf{x}) - a$.

Remark 1. *We restrict our analysis to affine limit state functions of the form* $g(\mathbf{x}) = \rho f(\mathbf{x}) - a$, $\rho \in \mathbb{R}$, $a \in \mathbb{R}$, *in order to preserve the Gaussianity of their sample paths, when we place a GP prior on f. This is not a restrictive assumption, however, because for nonaffine limit state functions we simply place a GP prior on $g(\mathbf{x})$ instead of $f(\mathbf{x})$.*

2.1. Failure probability estimation

Consider the probability space $(\mathfrak{S}, \mathfrak{F}, \mathbb{P}) = (\mathcal{X}, \mathfrak{B}(\mathcal{X}), \mathbb{P})$, where $\mathfrak{S}$ denotes the sample space that in our setting is $\mathcal{X} \subset \mathbb{R}^d$; $\mathfrak{B}(\mathcal{X})$ denotes the Borel sets in $\mathcal{X}$; and $\mathbb{P}$ denotes a probability measure. We model the uncertain parameter $\mathbf{x}$ as a random variable in $\mathbb{R}^d$ and assume that $\mathbf{x}$ is distributed according to the distribution $\mathbb{P}_{\mathbf{x}}$, which has density $q_{\mathbf{x}}$. The failure hypersurface defined by $\mathcal{C}$ divides the parameter space into a failure set $\mathcal{F} = \{\mathbf{x} : g(f(\mathbf{x})) \le 0\}$ and safe set $\mathcal{F}' := \mathcal{X} \backslash \mathcal{F}$. We are interested in estimating the probability of failure $p_{\mathcal{F}}$ that is defined by

$$p_{\mathcal{F}} = \int_{\mathbf{x} \in \mathcal{F}} d\mathbb{P}_{\mathbf{x}} = \int_{\mathbf{x} \in \mathcal{X}} \mathbb{I}_{\{\mathbf{x} \in \mathcal{F}\}} d\mathbb{P}_{\mathbf{x}} = \int_{\mathcal{X}} \mathbb{I}_{\{\mathbf{x} \in \mathcal{F}\}} q_{\mathbf{x}}(\mathbf{x}) \, d\mathbf{x}, \tag{2}$$

where $\mathbb{I}_{\{\mathbf{x} \in \mathcal{F}\}}$ denotes the indicator function that takes the value 1 when $\mathbf{x} \in \mathcal{F}$ and 0 otherwise. MC methods are often used to estimate $p_{\mathcal{F}}$; the idea is to draw N independent and identically distributed (iid) samples of $\mathbf{x}$, $\{\mathbf{x}^i, i = 1, \dots, N\}$, $\mathbf{x}^i \sim \mathbb{P}_{\mathbf{x}}$ and approximate $p_{\mathcal{F}}$ as

$$p_{\mathcal{F}} \approx p_{\mathcal{F}}^{\mathrm{MC}} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}_{\{\mathbf{x}^i \in \mathcal{F}\}}, \tag{3}$$

where $\mathbf{x}^i$ is the ith realization of $\mathbf{x}$. The MC estimator is unbiased, as shown below.

$$\mathbb{E}[p_{\mathcal{F}}^{\mathrm{MC}}] = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}[\mathbb{I}_{\{\mathbf{x}^i \in \mathcal{F}\}}]$$

$$= \frac{1}{N} \sum_{i=1}^{N} \int_{\mathcal{X}} \mathbb{I}_{\{\mathbf{x} \in \mathcal{F}\}} q_{\mathbf{x}}(\mathbf{x}) \, d\mathbf{x} = \frac{1}{N} \sum_{i=1}^{N} p_{\mathcal{F}} = p_{\mathcal{F}}$$

The variance of the MC estimator is given by

$$\mathbb{V}(p_{\mathcal{F}}^{MC}) = \frac{p_{\mathcal{F}}^{MC}(1 - p_{\mathcal{F}}^{MC})}{N}.$$

Using the approach described by (3) requires a large number of samples in order to obtain a failure probability estimate of desired accuracy. For instance, consider $p_{\mathcal{F}} = 10^{-3}$ and a root mean-squared error in the estimate of 10^{-4} (and therefore a variance of 10^{-8}). Then the number of samples necessary to achieve this accuracy via MC is $\approx 10^5$. This could be prohibitively expensive in the reliability analysis of complex engineered systems with expensive forward models. A common means to (partially) circumvent this shortcoming is to employ an auxiliary distribution that concentrates the samples around the failure region [20, 54]; this approach is referred to as importance sampling. Equation (2) can be equivalently written as

$$p_{\mathcal{F}} = \int_{\mathbf{x} \in \mathcal{X}} \mathbb{I}_{\{\mathbf{x} \in \mathcal{F}\}} \frac{q_{\mathbf{x}}(\mathbf{x})}{q_{\mathbf{x}}'(\mathbf{x})} q_{\mathbf{x}}'(\mathbf{x}) d\mathbf{x} = \mathbb{E}_{\mathbb{P}_{\mathbf{x}}'} \left[\mathbb{I}_{\{\mathbf{x} \in \mathcal{F}\}} \frac{q_{\mathbf{x}}(\mathbf{x})}{q_{\mathbf{x}}'(\mathbf{x})} \right] \tag{4}$$

where $q_{\mathbf{x}}'$ is the density of the biasing distribution $\mathbb{P}_{\mathbf{x}}'$. The key idea is that samples drawn from the biasing distribution are more likely to be from the set $\mathcal{F}$. From (4), the IS estimate of $p_{\mathcal{F}}$ can be written as

$$p_{\mathcal{F}}^{\text{IS}} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}_{\{\mathbf{x}^i \in \mathcal{F}\}} w(\mathbf{x}^i), \tag{5}$$

where $w(\mathbf{x}^i) = \frac{q_{\mathbf{x}}(\mathbf{x}^i)}{q_{\mathbf{x}}'(\mathbf{x}^i)}$ are the importance weights and $\mathbf{x}^i \sim \mathbb{P}_{\mathbf{x}}'$. The estimator in (5) is also unbiased, provided that $q_{\mathbf{x}}'(\mathbf{x}) > 0$ whenever $\mathbb{I}_{\{\mathbf{x} \in \mathcal{F}\}} \times q_{\mathbf{x}}(\mathbf{x}) > 0$. Let $\mathcal{Q} \subset \mathcal{X} := \{\mathbf{x} \in \mathcal{X} : \mathbb{I}_{\{\mathbf{x} \in \mathcal{F}\}} \times q_{\mathbf{x}}(\mathbf{x}) > 0\}$. Then

$$\mathbb{E}[p_{\mathcal{F}}^{\text{IS}}] = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}[\mathbb{I}_{\{\mathbf{x}^i \in \mathcal{F}\}} w(\mathbf{x}^i)]$$

$$= \frac{1}{N} \sum_{i=1}^{N} \int_{\mathcal{X}} \mathbb{I}_{\{\mathbf{x} \in \mathcal{F}\}} \frac{q_{\mathbf{x}}(\mathbf{x})}{q_{\mathbf{x}}'(\mathbf{x})} q_{\mathbf{x}}(\mathbf{x}) \, d\mathbf{x} = \frac{1}{N} \sum_{i=1}^{N} \int_{\mathcal{Q}} \mathbb{I}_{\{\mathbf{x} \in \mathcal{F}\}} \frac{q_{\mathbf{x}}(\mathbf{x})}{q_{\mathbf{x}}'(\mathbf{x})} q_{\mathbf{x}}(\mathbf{x}) \, d\mathbf{x}$$

$$= \frac{1}{N} \sum_{i=1}^{N} p_{\mathcal{F}} = p_{\mathcal{F}}.$$

As mentioned previously, we will construct the biasing distribution via a surrogate model in order to keep the computational costs tractable. As we will show later, similar to the construction of our surrogate model, our biasing distribution is also adaptively improved. We now introduce the surrogate modeling technique via GP regression.

2.2. Multifidelity Gaussian process regression

Crucially, we assume that $f(\mathbf{x})$ is approximated by models $\hat{f}(\mathbf{x}, s)$, where $\mathbf{x}$ are common inputs to all the models and $s \in \mathcal{S} \subset \mathbb{R}$ is a tunable *fidelity* parameter for each model. For the sake of simplicity, in this work we set s to be scalar valued and

without loss of generality $\mathcal{S} = [0, 1]$, where $s = 1$ and $s = 0$ represent the model at the highest and lowest fidelity, respectively; therefore,

$$\hat{f}(\mathbf{x}, 1) := f(\mathbf{x}).$$

Additionally, we assume there is known a cost function $c(s) : \mathcal{S} \to \mathbb{R}$, which is monotonic increasing in s and models the computational cost of querying $\hat{f}$ at a specific fidelity. Note that we assume $c(\cdot)$ is independent of $\mathbf{x}$ for the sake of simplicity, but our framework applies to more general $c(\mathbf{x}, s)$ as well. Overall, we are interested in computing (4) by querying $\hat{f}$, while keeping the overall cost lower than computing (4) by querying $f(\mathbf{x})$ alone. Our multifidelity method depends on learning a GP model that maps the augmented input-fidelity space $(\mathbf{x}, s) \in \mathcal{X} \times \mathcal{S}$ to output quantities of interest $\hat{f}(\mathbf{x}, s)$.

Remark 2. *We mention that by assuming a monotonic cost function $c(s)$ and the fact that fidelity increases as s increases in $[0, 1]$, we assume that a hierarchy of fidelities exists among the set of models and that high-fidelity models are more expensive. However, our method tries to reduce the computational cost only for reliability analysis. Therefore, in the absence of any known hierarchy among the models, the highest fidelity can be chosen as $\arg\max_{s \in \mathcal{S}} c(s)$. If $c(s)$ is unknown a priori, it can be learned, for example, by placing another GP prior on it and making a few observations.*

We assume that we can make evaluations of the function $\hat{f}$ at a given $\mathbf{x}$ and fidelity level s. That is,

$$\hat{y}_i = \hat{f}(\mathbf{x}_i, s_i), \quad i = 1, \ldots, n. \tag{6}$$

The key idea then is to specify GP prior distributions on the f. That is, $\hat{f}(\mathbf{x}, s) \sim \mathcal{GP}(0, k((\mathbf{x}, s), (\mathbf{x}', s')))$. The covariance function (or *kernel*) k captures the correlation between the observations in the joint $(\mathbf{x}, s)$ space; here we use the product composite form given by $k((\mathbf{x}, s), (\mathbf{x}', s')) = k_{\mathbf{x}}(\mathbf{x}, \mathbf{x}'; \boldsymbol{\gamma}_{\mathbf{x}}) \times k_s(s, s'; \boldsymbol{\gamma}_s)$, where $\boldsymbol{\gamma}_{\mathbf{x}} \in \mathbb{R}_+^d$ and $\boldsymbol{\gamma}_s \in \mathbb{R}_+^d$ parameterize the covariance functions for $\mathbf{x}$ and s, respectively. We estimate the GP hyperparameters $\boldsymbol{\Omega} = \{\boldsymbol{\gamma}_{\mathbf{x}}, \boldsymbol{\gamma}_s\}$ from data by maximizing the marginal likelihood. We choose an anisotropic Matérn-type kernel for both $k_{\mathbf{x}}$ and k_s. The posterior predictive distribution of the output Y, conditioned on available observations from the oracle, is given by [40]

$$Y(\mathbf{x}, s) | \mathcal{D}_n, \boldsymbol{\Omega} \sim \mathcal{GP}(\mu_n(\mathbf{x}, s), \sigma_n^2(\mathbf{x}, s)), \quad \mathcal{D}_n := \{(\mathbf{x}_i, s_i), \hat{y}_i\}_{i=1}^n$$
$$\mu_n(\mathbf{x}, s) = \mathbf{k}_n^\top [\mathbf{K}_n]^{-1} \mathbf{y}_n \tag{7}$$
$$\sigma_n^2(\mathbf{x}, s) = k((\mathbf{x}, s), (\mathbf{x}, s)) - \mathbf{k}_n^\top [\mathbf{K}_n]^{-1} \mathbf{k}_n,$$

where $\mathbf{k}_n$ is a vector of covariance between $(\mathbf{x}, s)$ and all observed points in $\mathcal{D}_n$; $\mathbf{K}_n$ is a sample covariance matrix of observed points in $\mathcal{D}_n$; $\mathbf{I}$ is the identity matrix; and $\mathbf{y}_n$ is the vector of all observations in $\mathcal{D}_n$. Equation (7) is then used as a surrogate model for f in FPE. Note that μ_n and σ_n^2 are the posterior mean and variance of the GP, respectively, where the subscript n implies the conditioning based on n past observations. Our surrogate model for the limit state function then is given as

$$\tilde{g}_n(\mathbf{x}, s) := g(\mu_n(\mathbf{x}, s)).$$

An important component of our proposed method is that the data required to build the surrogate model is adaptively selected. This ensures that the inputs and fidelities are judiciously chosen to be tailored to our specific goals. We call this adaptive approach "sequential experimental design," which we discuss next.

Algorithm 1: Generic acquisition-based sequential design

---

1 **Given:** $\mathcal{D}_n = \{(\mathbf{x}_i, s_i), \hat{y}_i\}_{i=1}^n$, total budget q, and GP hyperparameters $\mathbf{\Omega}$
   Result: $(\mathbf{x}_q, \hat{y}_q)$
2 **for** $i = n + 1, \ldots, q$, **do**
3 $\quad$ Find $\mathbf{x}_i, s_i \in \underset{(\mathbf{x},s) \in \boldsymbol{\mathcal{X}} \times \mathcal{S}}{\arg\max} \; \alpha(\mathbf{x}, s)$ $\qquad$ (acquisition function maximization)
4 $\quad$ Observe $\hat{y}_i = f(\mathbf{x}_i, s_i) + \epsilon_i$
5 $\quad$ Append $\mathcal{D}_i = \mathcal{D}_{i-1} \cup \{(\mathbf{x}_i, s_i), \hat{y}_i\}$
6 $\quad$ Update GP hyperparameters $\mathbf{\Omega}$

2.3. Acquisition-based sequential experiment design

We use principles from Bayesian optimization (BO) [50, 12, 51] to sequentially update the GP to improve our knowledge about our goal—in this case, estimation of $\mathcal{C}$ and the subsequent FPE. Akin to BO, we proceed by defining an acquisition function $\alpha(\mathbf{x}, s)$, in terms of the GP posterior that is optimized (typically maximized) to select—given previous n observations—the next point $\mathbf{x}_{n+1}$ at s_{n+1}, and the process continues until a budget for computation is reached; see Algorithm 1. The choice of $\alpha(\mathbf{x}, s)$ is key to the performance of the quality of the sequential point selection, which we discuss in the next section.

Naturally, the acquisition function is designed according to the specific needs of the sequential process. Classical acquisition functions for optimization include the probability of improvement (`PI`) [25], the expected improvement (`EI`) [19, 31], and the GP upper confidence bound (`UCB`) [53]; these approaches work with a probabilistic estimate of the optimum via the GP. Other acquisition functions for optimization include entropy-based approaches [61, 16, 59]. Similarly, acquisition functions for active learning (purely learning the function $f(\mathbf{x})$) also exist [28, 6]. Therefore, acquisition functions to learn a level set (e.g., failure boundary) can also be defined, such as the expected level set improvement of Ranjan et al. [39] and Bichon et al. [5], entropy-based approaches of Cole et al. [7] and Marques et al. [30], stepwise uncertainty reduction by Bect et al. [4], and other methods as in Picheny et al. [34].

In all the aforementioned approaches for failure level set estimation, we point out that application to multifidelity models is not straightforward. Although [30] show their applicability with multiple information sources, that requires learning individual discrepancies between models at each fidelity and the highest fidelity model. This can be computationally demanding in high-dimensional problems or when some of the lower-fidelity models are not necessarily cheap (compared with the highest-fidelity models). Consequently, such methods are restricted to discrete fidelity spaces with small cardinality. Furthermore, outside of the reliability literature, autoregressive approaches such as that of [22] and [33] suffer from the same limitation. Therefore, in this work we propose a new acquisition approach that applies to both discrete and continuous fidelity spaces. Additionally, our approach can extend any existing acquisition function to the multifidelity setting, as we will show later. We provide more details in the following section.

3. Proposed Method

3.1. Multifidelity acquisition functions

Drawing from the work of Renganathan et al. [44], where they propose an acquisition function for time-dependent problems with a finite horizon, we define a multifidelity *value* function framework that we use to develop our multifidelity acquisition function. Given the data from n observations $\boldsymbol{\mathcal{D}}_n$, we define the value function v by

$$v_n(\mathbf{x}, s; \boldsymbol{\omega}) := \mathbb{E}_{y \sim Y|\boldsymbol{\mathcal{D}}_n} \left[h(y(\mathbf{x}, s); \boldsymbol{\omega}) \right], \qquad (8)$$

where h is a scalar-valued function, $\boldsymbol{\omega} \in \mathbb{R}^p$ parameterizes h, $Y|\boldsymbol{\mathcal{D}}_n$ is the GP posterior given n observations, and $y(\mathbf{x}, s)$ is a sample path realized from $Y|\boldsymbol{\mathcal{D}}_n$. For example, some of the existing acquisition functions for GP-based reliability analysis can be obtained from defining h as shown in Table 1, where a is the failure threshold and

$\boldsymbol{\omega}$	$h(\mathbf{x}, s)$	$v_n(\mathbf{x}, s)$	acquisition function
$\{a, \eta\}$	$\delta^2(\mathbf{x}, s) - \min\{(y(\mathbf{x}, s) - a)^2, \delta^2(\mathbf{x}, s)\}$	$EI_r(\mathbf{x}, s)$	Ranjan et al. [39]
$\{a, \eta\}$	$\delta(\mathbf{x}, s) - \min\{|(y(\mathbf{x}, s) - a)|, \delta(\mathbf{x}, s)\}$	$EI_b(\mathbf{x}, s)$	Bichon et al. [5]

Table 1: Parametric representation of existing acquisition functions via our framework (8).

$\delta^2(\mathbf{x}) = \eta\sigma^2(\mathbf{x})$, $\eta \in \mathbb{R}_+$ defines the width of the uncertainty band around the GP posterior mean prediction. We use the formulation in (8) because we can then directly extend acquisition functions in Table 1 to the multifidelity setting, as we show below.

We now define our multifidelity acquisition function as

$$\alpha(\mathbf{x}, s) = \int \max_{\mathbf{x}' \in \boldsymbol{\mathcal{X}}} \left[\underbrace{v(\mathbf{x}', 1; \boldsymbol{\omega})}_{\text{value at highest fid.}} \mid \underbrace{\boldsymbol{\mathcal{D}}_n}_{\text{prev. obs.}} \bigcup \{ \underbrace{(\mathbf{x}, s), y}_{\text{candidate obs.}} \} \right] q_y(y|\boldsymbol{\mathcal{D}}_n) \, dy, \qquad (9)$$

where $q_y(y|\boldsymbol{\mathcal{D}}_n)$ is the probability density of the GP posterior distribution conditioned on $\boldsymbol{\mathcal{D}}_n$. Notice that in the inner maximization we use $v(\cdot, 1)$; this indicates that we place our value on the highest fidelity. In essence, (9) computes the expected value of the maximum value function at the highest fidelity due to a candidate input-fidelity pair $(\mathbf{x}, s)$, where the expectation is with respect to the posterior GP. We write (9) more concisely as

$$\alpha(\mathbf{x}, s) = \mathbb{E}_n \left[\max_{\mathbf{x}' \in \boldsymbol{\mathcal{X}}} v_{n,1}(\mathbf{x}', 1; \boldsymbol{\omega}) \right]. \qquad (10)$$

In (10), note that $v_{n,1} = v|\boldsymbol{\mathcal{D}}_n \bigcup \{(\mathbf{x}, s), y\}$, which is the value function given the union of $\boldsymbol{\mathcal{D}}_n$ and the $n + 1$th observation beinga drawn from the GP posterior $Y|\boldsymbol{\mathcal{D}}_n$. In general, (10) is not guaranteed to admit a closed-form expression. Hence we make the MC approximation

$$\alpha(\mathbf{x}, s) \approx \hat{\alpha}(\mathbf{x}, s) := \frac{1}{N} \sum_{j=1}^{N} \max_{\mathbf{x}' \in \boldsymbol{\mathcal{X}}} v_{n,1}^j(\mathbf{x}', 1), \qquad (11)$$

where $v_{n,1}^j(\mathbf{x}', 1) = v(\mathbf{x}', 1)|\boldsymbol{\mathcal{D}}_n \bigcup \{(\mathbf{x}, s), y^j\}$ with the N iid samples $\{y^j, \forall j = 1, \ldots, N\}$ drawn from $Y|\mathcal{D}_n$ at $(\mathbf{x}, s)$.

The inner maximization in (11) can be approximated by evaluating the value function $v_{n,1}$ on a discrete set $\hat{\boldsymbol{\mathcal{X}}}$ and the maximization approximated as (for $\hat{\mathbf{x}}_i \in \hat{\boldsymbol{\mathcal{X}}}$)

$$\max_{\mathbf{x}' \in \boldsymbol{\mathcal{X}}} v_{n,1}^j(\mathbf{x}', 1) \approx \max_i v_{n,1}^j(\hat{\mathbf{x}}_i, 1).$$

As noted above, the value function is a function of the GP only, and hence its evaluation on $\hat{\boldsymbol{\mathcal{X}}}$ is computationally cheap. We now sequentially choose points by maximizing the cost-normalized acquisition function defined as

$$(\mathbf{x}_{n+1}, s_{n+1}) = \arg\max_{(\mathbf{x},s) \in \boldsymbol{\mathcal{X}} \times \boldsymbol{\mathcal{S}}} \frac{\hat{\alpha}_2(\mathbf{x}, s)}{c(s)}. \tag{12}$$

In (12), since we normalize the acquisition function by the cost of evaluations $c(s)$, the optimization penalizes decisions at high s, preventing expensive high-fidelity evaluations unless they offer a significant increase in the maximum value function compared with lower fidelities. For this reason, we consider our approach *cost-aware* and thus name it CAMERA: cost-aware, adaptive, multifidelity, efficient reliability analysis. We now discuss the choice of the value function.

3.2. Choice of value function

As previously mentioned, we propose a multifidelity acquisition function that provides a framework to existing acquisition functions, defined for a single-fidelity setting, for a multifidelity setting. Therefore, while the choice of the value function can be practically any existing acquisition function, for the experiments in this work we choose the *improvement*-based acquisition functions for contour estimation by [39] and [5]. Ranjan et al. [39] propose the following improvement function, which we extrapolate to the multifidelity setting as

$$I_r(\mathbf{x}, s) = \delta^2(\mathbf{x}, s) - \min\{(y(\mathbf{x}, s) - a)^2, \delta^2(\mathbf{x}, s)\}, \tag{13}$$

where a is the target level and $\delta^2(\mathbf{x}, s) = \eta \sigma^2(\mathbf{x}, s)$, $\eta \in \mathbb{R}_+$ defines the width of the uncertainty band around the GP posterior mean prediction. The intuition behind (13) is that improvement is defined as getting points closer to the target level set or in places where the uncertainty of the GP predictions is large, while still being within a certain *confidence region* of the GP prediction. Consequently, $I_r(\mathbf{x}) = 0$, when $(y(\mathbf{x}) - a)^2 > \epsilon^2(\mathbf{x})$ and therefore regions far away from the potential contour locations are avoided. Bichon et al. [5] modify (13) as

$$I_b(\mathbf{x}, s) = \delta(\mathbf{x}, s) - \min\{|(y(\mathbf{x}, s) - a)|, \delta(\mathbf{x}, s)\}, \tag{14}$$

which results in a closed-form expression for the expected improvement

$$\begin{aligned} EI_b(\mathbf{x}, s) =& \delta(\mathbf{x}, s) \left[\Phi(z^+) - \Phi(z^-)\right] - \sigma(\mathbf{x}, s) \left[2\phi(z) - \phi(z^-) - \phi(z^+)\right] \\ & + (\mu(\mathbf{x}) - a) \left[2\Phi(z) - \Phi(z^-) - \Phi(z^+)\right], \end{aligned} \tag{15}$$

where $z = \frac{a - \mu(\mathbf{x}, s)}{\sigma(\mathbf{x}, s)}$, $z^\pm = \frac{a^\pm - \mu(\mathbf{x}, s)}{\sigma(\mathbf{x}, s)}$, and $a^\pm = a \pm \delta(\mathbf{x}, s)$. Note that (13) and (14) practically have the same meaning, but (14) leads to a more convenient analytical form upon taking the expectation, as opposed to that in (13). We therefore propose to specify our value function, also extending it to the multifidelity setting, as

$$v(\mathbf{x}, 1) = EI_b(\mathbf{x}, 1), \tag{16}$$

and define our acquisition function as follows:

$$\alpha(\mathbf{x}, s) = \mathbb{E}_{y \sim Y_n} \left[\max_{\mathbf{x}' \in \mathcal{X}} EI_b(\mathbf{x}', 1) | \mathcal{D}_n \bigcup \{(\mathbf{x}, s), y\} \right], \qquad (17)$$

which essentially quantifies the *average* increase in the maximum EI_b at the highest fidelity ($s = 1$) due to a candidate decision-fidelity pair $(\mathbf{x}, s)$. Note that we extend the acquisition function in [5] to general multifidelity cases, where the fidelity space could be continuous or discrete. The acquisition function in (17) is then plugged into Algorithm 1 to obtain a final surrogate model $\tilde{g}_q(\mathbf{x}, 1)$ of $\hat{f}(\mathbf{x}, 1)$.

3.3. Importance sampling

The last step of our approach is the estimation of the failure probability. The multifidelity surrogate model that is adaptively constructed is now used to construct a biasing distribution. This distribution is then used to generate a sample set using which the failure probability is estimated.

Let $q_{\mathbf{x}}^{'(n)}$ be the density of the biasing distribution, $\mathbb{P}_{\mathbf{x}}^{'(n)}$, after n rounds of updating our multifidelity surrogate model. Then, the importance sampling estimate is given by

$$p_{\mathcal{F}}^{\text{IS}} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}_{\{\mathbf{x}^i \in \mathcal{F}\}} \frac{q(\mathbf{x}^i)}{q_{\mathbf{x}}^{'(n)}}, \qquad (18)$$

where $\mathbf{x}^i$, $i = 1, \ldots, N$ are iid samples drawn from $\mathbb{P}_{\mathbf{x}}^{'(n)}$.

3.4. Overall `CAMERA` algorithm

We are now ready to summarize our overall algorithm. We use a Gaussian mixture model (GMM) [18] to learn the biasing density $q_{\mathbf{x}}^{'(n)}$ from a sample set generated from the GP. For all the cases, we set the number of components of the GMM to 25 and restrict the covariance matrix to be diagonal. A few trials indicated that these settings are sufficient to get a good approximation to the biasing density across all the experiments. The expectation-maximization algorithm [8, 62] is used to learn the hyperparameters of the GMM. Once fit, this distribution is sampled to generate a biasing set that is evaluated on the high-fidelity model $\hat{f}(\cdot, 1)$ to compute the IS estimate of failure probability. The exact steps are summarized in Algorithm 2.

Note that M in step 6a in Algorithm 2 can be picked arbitrarily large since it does not involve any high-fidelity model queries, only the surrogate model. On the other hand, N must be tractable since it involves high-fidelity queries. Instead of directly setting M, we first uniformly sample $\mathcal{X}$ with 10^7 samples and filter out the sample set to obtain $\mathcal{X}'$ as in step 2 using the surrogate model.

4. Theoretical Properties

Recall that our surrogate-based IS estimator is unbiased. We now show that our adaptive surrogate model is consistent; that is, with high probability the predicted level set converges to the true level set asymptotically. We first demonstrate that the points selected by maximizing our multifidelity acquisition function (12) asymptotically reduce the GP posterior variance to 0 everywhere; see Lemma 2 and Lemma 3. Then, in Theorem 1, we show that our surrogate model is asymptotically consistent.

We begin by making a smoothness assumption on our unknown function f, followed by an assumption on the limit state function.

Algorithm 2: CAMERA: Cost-aware Adaptive Multifidelity Efficient Reliability Analysis.

1 Input:

- Surrogate model: Multifidelity GP posterior of f, $\mathcal{GP}(\mu_n(\mathbf{x}, s), \sigma_n^2(\mathbf{x}, s))$ and observed data $\boldsymbol{\mathcal{D}}_n$.

- Cumulative cost $C_n = \sum_{j=1}^n c(s_j)$, total budget B.

- High-fidelity sample size N.

- Biasing sample size M.

Result: $p_{\mathcal{F}}^{\mathrm{IS}}$

2 while $C_i \leq B, \ \forall \ i = n+1, \ldots,$ **do**

    1. Find $\mathbf{x}_i, s_i \in \underset{(\mathbf{x}, s) \in \boldsymbol{\mathcal{X}} \times \mathcal{S}}{\arg\max} \ \alpha(\mathbf{x}, s)$ # acquisition function maximization

    2. Observe $\hat{y}_i = f(\mathbf{x}_i, s_i)$ # new multifidelity observation

    3. Cumulative cost $C_i = C_{i-1} + c(s_i)$

    4. Append $\boldsymbol{\mathcal{D}}_i = \boldsymbol{\mathcal{D}}_{i-1} \cup \{(\mathbf{x}_i, s_i), \hat{y}_i\}$ # append data set

    5. Update GP hyperparameters $\boldsymbol{\Omega}$ # recompute GP hyperparameters

end

Let $q > n$ be the last iteration of updating the surrogate model.

  6. Fit a biasing distribution $\mathbb{P}_{\mathbf{x}}^{\prime(q)}$ with density $q_{\mathbf{x}}^{\prime(q)}$

    (a) Obtain a sample set of cardinality M, using the *high fidelity* surrogate model

$$\boldsymbol{\mathcal{X}}' = \{\mathbf{x}_i \in \boldsymbol{\mathcal{X}} : \tilde{g}_q(\mathbf{x}_i, 1) - a = \mu_q(\mathbf{x}_i, 1) - a \leq 0, \ i = 1, \ldots, M\}.$$

    (b) Fit a GMM with $\boldsymbol{\mathcal{X}}'$ to obtain the biasing distribution $\mathbb{P}_{\mathbf{x}}^{\prime(q)}$ with density $q_{\mathbf{x}}^{\prime(q)}$.

  7. Obtain a biasing sample set

$$\boldsymbol{\mathcal{X}}_{\mathrm{biasing}}' = \{\mathbf{x}_k \sim \bar{\mathbb{P}}_{\mathbf{x}}^{\prime q}, \ k = 1, \ldots, N\}.$$

  8. Compute indicator function (with *high fidelity* samples) and IS weights

$$\mathbb{I} = \{\mathbb{I}_{\{\mathbf{x}_k \in \mathcal{F}\}}\}, \quad \mathbf{w} = \left\{\frac{q_{\mathbf{x}}(\mathbf{x}_k)}{q_{\mathbf{x}}^{\prime(q)}(\mathbf{x}_k)}\right\}, \ k = 1, \ldots, N,$$

  9. compute the IS failure probability estimate

$$p_{\mathcal{F}}^{\mathrm{IS}} = \frac{\mathbb{I}^\top \mathbf{w}}{N}$$

Assumption 1. *We assume that the sample paths drawn from the GP prior on f are twice continuously differentiable. Such a GP is achieved by choosing the covariance kernel to be four times differentiable [13, Theorem 5], for example, the squared-exponential or Matérn-class kernel. Then, the following holds (for $L \in \mathbb{R}_+$):*

$$\Pr\left[|f(\mathbf{x}) - f(\mathbf{x}')| > L\|\mathbf{x} - \mathbf{x}'\|\right] \le e^{-L^2/2}.$$

Assumption 2. *The limit state function g is an affine function of f, defined as $g(f(\mathbf{x})) = \rho f(\mathbf{x}) - a$. W.l.o.g., we pick $\rho = 1$.*

First, we show that our proposed two-step lookahead acquisition function will never choose the same point twice. We present this in Lemmas 1 and 2.

Lemma 1. *Without the* max *operator in (9), the multifidelity acquisition function reduces to the value function. That is,*

$$\mathbb{E}_y[\upsilon(\mathbf{x}', 1)|\mathcal{D}_n \bigcup \{\mathbf{x}, y\}] = \upsilon(\mathbf{x}', 1)|\mathcal{D}_n$$

.

Proof. Let us write

$$\mathbb{E}_Y[\upsilon(\mathbf{x}', 1)|\mathcal{D}_n \bigcup \{(\mathbf{x}, s), Y\}] = \mathbb{E}_{Y(\mathbf{x},s)}[\upsilon_n(\mathbf{x}', 1)|Y(\mathbf{x}, s)],$$

where $\upsilon_n(\cdot, \cdot) = \upsilon(\cdot, \cdot)|\mathcal{D}_n$. Using the definition of the value function, we have

$$
\begin{aligned}
\mathbb{E}_{Y(\mathbf{x},s)}[\upsilon_n(\mathbf{x}', 1)|Y(\mathbf{x}, s)] &= \mathbb{E}_{Y(\mathbf{x},s)}[\mathbb{E}_{Y(\mathbf{x}',1)}[h(Y(\mathbf{x}', 1), \boldsymbol{\omega})]|y(\mathbf{x}, s)] \\
&= \mathbb{E}_{Y(\mathbf{x},s)}[\mathbb{E}_{Y(\mathbf{x}',1)}[h(Y(\mathbf{x}', 1), \boldsymbol{\omega})|y(\mathbf{x}, s)]] \\
&= \mathbb{E}_{Y(\mathbf{x}',s')}[h(Y(\mathbf{x}', 1), \boldsymbol{\omega})] \\
&= \upsilon_n(\mathbf{x}', 1).
\end{aligned}
\tag{19}
$$

$\square$

Lemma 2. $\forall i, j \in \{1, \ldots, n\}$, $(\mathbf{x}_i, s_i) \neq (\mathbf{x}_j, s_j)$, *where $(\mathbf{x}_i, s_i)$ and $(\mathbf{x}_j, s_j)$ are the input-fidelity pairs that maximize $\alpha(\mathbf{x}, s)|\mathcal{D}_{i-1}$ and $\alpha(\mathbf{x}, s)|\mathcal{D}_{j-1}$, respectively.*

Proof. For a given $(\mathbf{x}, s) \in \mathcal{X} \times \mathcal{S}$, let $\sigma_n^2(\mathbf{x}, s) = 0$; in other words, i a noise-free observation was previously made at $(\mathbf{x}, s)$. Then,

$$\alpha(\mathbf{x}, s) = \mathbb{E}_{Y(\mathbf{x},s)}[\max_{\mathbf{x}' \in \mathcal{X}} \upsilon(\mathbf{x}', 1)|\{(\mathbf{x}, s), Y(\mathbf{x}, s)\}] = \max_{\mathbf{x}' \in \mathcal{X}} \upsilon(\mathbf{x}', 1)|\{(\mathbf{x}, s), \mu_n(\mathbf{x})\} = \max_{\mathbf{x}' \in \mathcal{X}} \upsilon(\mathbf{x}', 1),$$

which follows from the fact that $Y(\mathbf{x}, s)|(\sigma_n^2(\mathbf{x}, s) = 0) = \mu_n(\mathbf{x}, s)$ and that the value function conditioned on a previously observed point is unchanged. Now consider a candidate point (i.e., an unobserved point) $(\mathbf{x}'', s'') \in (\mathcal{X} \times \mathcal{S}) \backslash (\mathbf{x}, s)$,

$$\alpha(\mathbf{x}'', s'') = \mathbb{E}_{Y(\mathbf{x}'',s'')}[\max_{\mathbf{x}' \in \mathcal{X}} \upsilon(\mathbf{x}', 1)|\{(\mathbf{x}'', s''), Y(\mathbf{x}'', s'')\}] \ge \max_{\mathbf{x}' \in \mathcal{X}} \upsilon(\mathbf{x}', 1) = \alpha(\mathbf{x}, s),$$

where the inequality follows from Jensen's inequality [47, Ch. 3] and we have also used Lemma 1. Furthermore, the equality holds only if $(\mathbf{x}, s) = (\mathbf{x}'', s'')$ or $\sigma_n^2(\mathbf{x}, s) = 0 = \sigma_n^2(\mathbf{x}'', s'')$, and therefore $(\mathbf{x}, s) \neq (\mathbf{x}'', s'')$, since $(\mathbf{x}'', s'')$ is unobserved and hence $\sigma_n^2(\mathbf{x}'', s'') > 0$. $\square$

Now, using Lemma 2, we will show that the multifidelity acquisition strategy reduces the GP posterior variance to 0 at $s = 1$, $\forall \mathbf{x} \in \boldsymbol{\mathcal{X}}$.

Lemma 3. *Let $\{\mathbf{x}_i\}$, $i = 1, \ldots, n$ be a sequence of points selected via our multifidelity acquisition strategy. Then, $\lim_{n \to \infty} \sup_{\mathbf{x} \in \boldsymbol{\mathcal{X}}} \sigma_n^2(\mathbf{x}, 1) = 0$.*

Proof. W.l.o.g, let $k((\mathbf{x}, s), (\mathbf{x}, s)) = 1$. Recall that the posterior variance of the GP is given by

$$\sigma_n^2(\mathbf{x}, s) = k((\mathbf{x}, s), (\mathbf{x}, s)) - \mathbf{k}_n^\top \mathbf{K}_n^{-1} \mathbf{k}_n.$$

Note that as $n \to \infty$, and if Lemma 2 holds, then $\mathbf{k}_n \to \mathbf{K}_n^{:,i}$, where by $\mathbf{K}_n^{:,i}$ we mean the ith column of $\mathbf{K}_n$, $i \in [n]$, and therefore $\mathbf{k}_n^\top \mathbf{K}_n^{-1} = \mathbf{e}_i$, where $\mathbf{e}_i$ is an n-vector with 1 at the ith location and 0 elsewhere. Then

$$\lim_{n \to \infty} \sigma_n^2(\mathbf{x}, 1) = 1 - \mathbf{K}_n^{:,i \top} \mathbf{e}_i = 1 - k((\mathbf{x}_i, 1), (\mathbf{x}_i, 1)) = 0.$$

$\square$

We are now ready to show that with high probability, the predicted failure level set via our multifidelity surrogate model converges to the true failure level set. For this purpose we assume a discrete set $\tilde{\mathcal{C}} \subset \mathcal{C}$ and show that the absolute difference between the surrogate prediction at $\tilde{\mathcal{C}}$ and the true failure level set is bounded with high probability.

Theorem 1 (Asymptotic consistency). *Let $\delta \in (0, 1)$ and Assumptions 1 and 2 hold, with $L^2 = 2 \log\left(\frac{2}{\delta}\right)$. Let $\tilde{\mathcal{C}}$ denote a discretization of $\mathcal{C}$ with $|\tilde{\mathcal{C}}| < \infty$, where we evaluate the true failure boundary. Let $\epsilon \geq 0$, and $a_n = \sqrt{2 \log(b_n |\tilde{\mathcal{C}}|)/\delta}$, where $\sum_{n=1}^{\infty} 1/b_n = 1$. Then the predicted failure level set converges to the true failure level set with high probability; that is, the following holds:*

$$\Pr\left[\forall \mathbf{x} \in \boldsymbol{\mathcal{X}}, \lim_{n \to \infty} |g(f(\mathbf{x}, 1)) - g(\mu_{n-1}(\mathbf{x}, 1))| \leq \epsilon\right] > 1 - \delta.$$

Proof. The proof is based on [53], but we differ in the choice of constants. We first present a simple identity that bounds the probability that a Gaussian random variable is greater than some value. Given $r \sim \mathcal{N}(0, 1)$ and $c > 0$, consider the following:

$$
\begin{aligned}
\Pr(r > c) &= \int_c^\infty \frac{1}{\sqrt{2\pi}} e^{-r^2/2} dr = e^{-c^2/2} \int_c^\infty \frac{1}{\sqrt{2\pi}} e^{-(r-c)^2/2} e^{-c(r-c)} dr \\
&\leq e^{-c^2/2} \int_c^\infty \frac{1}{\sqrt{2\pi}} e^{-(r-c)^2/2} dr \leq \frac{1}{2} e^{-c^2/2},
\end{aligned}
\tag{20}
$$

where the second line is due to the fact that $0 \leq e^{-c(r-c)} \leq 1$ for $r \geq c$ and $c > 0$ and the last line from the fact that the integrand is the density of $\mathcal{N}(c, 1)$. Using this result and the prior assumption that the unknown function f has a GP prior, we state, with $c = a_n$ and a fixed $\mathbf{x}_i \in \tilde{\mathcal{C}}$,

$$\Pr\left[|g(f(\mathbf{x}_i, 1)) - g(\mu_{n-1}(\mathbf{x}_i, 1))| > a_n \sigma_{n-1}(\mathbf{x}_i, 1)\right] \leq \frac{1}{2} e^{-a_n^2/2} = \frac{\delta}{2 b_n |\tilde{\mathcal{C}}|}. \tag{21}$$

14

Applying the union bound $\forall i$, we get

$$\Pr \bigcup_{i=1}^{|\tilde{\mathcal{C}}|} |g(f(\mathbf{x}_i, 1)) - g(\mu_{n-1}(\mathbf{x}_i, 1))| > a_n \sigma_{n-1}(\mathbf{x}_i, 1) \le$$

$$\sum_{i=1}^{\tilde{\mathcal{C}}} \Pr\left[|g(f(\mathbf{x}_i, 1)) - g(\mu_{n-1}(\mathbf{x}_i, 1))| > a_n \sigma_{n-1}(\mathbf{x}_i, 1)\right]$$

$$\le \sum_{i=1}^{\tilde{\mathcal{C}}} \frac{\delta}{2b_n |\tilde{\mathcal{C}}|} = \frac{\delta}{2b_n}.$$

Now we apply the union bound $\forall n$:

$$\Pr\left[\forall \tilde{\mathbf{x}} \in \tilde{\mathcal{C}}, \ \bigcup_{n=1}^{\infty} |g(f(\tilde{\mathbf{x}}, 1)) - g(\mu_{n-1}(\tilde{\mathbf{x}}, 1))| > a_n \sigma_{n-1}(\tilde{\mathbf{x}}, 1)\right] \le \sum_{n=1}^{\infty} \frac{\delta}{2b_n} = \frac{\delta}{2}. \quad (22)$$

We would like to extend the proof to a general $\mathbf{x} \in \mathcal{X}$. For some $\mathbf{x} \in \mathcal{X}$, let $\tilde{\mathbf{x}}$ denote a point $\in \tilde{\mathcal{C}}$ that is nearest (in Euclidean distance) to $\mathbf{x}$. Then, per Assumption 1 and by setting $L^2 = 2 \log\left(\frac{2}{\delta}\right)$, the following statement holds:

$$\Pr\left[|g(f(\mathbf{x}, 1)) - g(f(\tilde{\mathbf{x}}, 1))| > \sqrt{2 \log\left(\frac{2}{\delta}\right)} \|\mathbf{x} - \tilde{\mathbf{x}}\|\right] \le \frac{\delta}{2}. \quad (23)$$

Combining (22) and (23), the following statement holds:

$$\Pr\left[\forall \mathbf{x} \in \mathcal{X}, \forall n, \ |g(f(\mathbf{x}, 1)) - g(\mu_{n-1}(\tilde{\mathbf{x}}, 1))| \le a_n \sigma_{n-1}(\mathbf{x}, 1) + \sqrt{2 \log\left(\frac{2}{\delta}\right)} \|\mathbf{x} - \tilde{\mathbf{x}}\|\right] > 1 - \delta.$$

From Lemma 3, the $\lim_{n \to \infty} \sup_{\mathbf{x} \in \mathcal{X}} \sigma_{n-1}(\mathbf{x}, 1) = 0$, and thus

$$\Pr\left[\forall \mathbf{x} \in \mathcal{X}, \lim_{n \to \infty} |g(f(\mathbf{x}, 1)) - g(\mu_{n-1}(\tilde{\mathbf{x}}, 1))| \le \epsilon\right] > 1 - \delta,$$

where $\epsilon = \sqrt{2 \log\left(\frac{2}{\delta}\right)} \|\mathbf{x} - \tilde{\mathbf{x}}\|$. $\qquad \square$

5. Numerical Experiments

We first demonstrate our proposed method on synthetic test functions of varying dimensions in Section 5.1 and then on two real-world applications: a gas turbine blade stress analysis test case using finite element analysis in Section 5.2 and a transonic wing aerodynamics test case using a finite volume method in Section 5.3. Our goal is to compare the multifidelity predictions of the failure probability against the predictions with the high-fidelity model only. For the synthetic test cases and the gas turbine blade test case, we estimate the *true* failure probability using a naive MC with a set of points drawn uniformly at random from $\mathcal{X}$; note that we use this estimate as the ground truth. Then we compute the failure probability error as $\frac{|p_{\mathcal{F}} - p_{\mathcal{F}}^{\text{IS}}|}{p_{\mathcal{F}}}$. For the transonic wing test case, the true failure probability is unknown and is prohibitive to

15

compute; we therefore report only the comparative predictions between single fidelity and multifidelity.

For all the experiments, we provide a set of seed points $\mathcal{D}_n = \{(\mathbf{x}_i, s_i), y_i, \ i = 1, \ldots, n\}$ to start the algorithm. Note that we start both the single-fidelity and multifidelity runs with the same *budget* of information; therefore, for any given experiment, the n for single-fidelity and multifidelity runs are different in order to maintain the same total computational cost with which we start the experiment. Naturally, whereas for the multifidelity experiments the seed points are distributed uniformly in $\mathcal{S}$, for the single-fidelity experiments the seed points are constrained to $\mathcal{S} = \{1\}$. All the synthetic experiments are repeated a total of 20 times with randomized starting seed points. The turbine stress analysis test case is repeated only 5 times, keeping in consideration the higher computational cost of this experiment. There is no repetition in the transonic wing test case because that would be computationally prohibitive.

For all the experiments, the following cost model is assumed:

$$ c(s)/c_0 = 0.1 + \exp\left(-c_1 \times (1 - s)\right), $$

where c_0 is a normalizing constant that is required to ensure that the acquisition function optimization is not biased by the actual costs of the models; note that c_0 is also the cost of the high-fidelity ($s = 1$) model, for which we fix $c_0 = 500$. The $c_1 > 0$ determines the steepness of the cost model; we set $c_1 = 10$ for all the experiments. Finally, $c_2 > 0$ is a constant to ensure that the cost model is bounded away from zero and is roughly the cost of the lowest-fidelity model ($\hat{f}(\cdot, 0)$); we fix $c_2 = 0.1$. We note that our cost model—by virtue of setting $c_1 = 10$—is designed to emphasize the disparity between the computational costs of lower- and higher-fidelity model queries; see Figure 1. In Figure 16 we show the effect of the cost model with varying c_1 and discuss its effect on the predictions.

We evaluate our predictions qualitatively for the 2D cases, in other words, by comparing the failure level sets estimated via the final GP mean $\mu_q(\mathbf{x}, 1)$ against the true level sets in contour plots. For all the synthetic cases (including the 2D cases), we then compute the error in failure probability, as mentioned previously. The true failure probability is estimated via a naive MC with $N = 10^6$ points sampled uniformly from $\mathcal{X}$. For the turbine case, we first evaluate the high-fidelity model at a total of 150 parameter snapshots and use the observed QoIs to construct a radial basis function (RBF) interpolant. We then perform a naive MC on the RBF interpolant to obtain the true $p_{\mathcal{F}}$. We begin by discussing the synthetic experiments in the next section.

5.1. Synthetic experiments

We introduce several new multifideity test functions that can be used to benchmark multifidelity methods beyond the scope of this work. Specifically, we introduce the 2D multifidelity multimodal function [5], the 2D multifidelity four-branches function [49], and the 3D multifidelity Ishigami function [17]. Additionally, we use the 6D multifidelity Hartmann function provided in [55]. The specifics of these test functions are discussed as follows.

5.1.1. 2D multimodal function

We extend the test function in [5] to multifidelity setting as shown in (24),
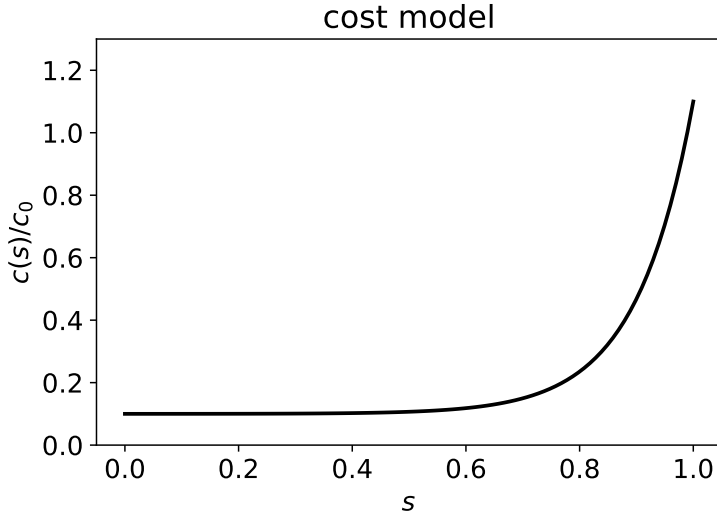
Figure 1: Multifidelity cost model $c(s)$ with $c_1 = 10$, which emphasizes the cost disparity across fidelities as one approaches $s = 1$.

$$f(\mathbf{x}, s) = \frac{(x_1^2 + 4)(x_2 - 1)}{20} - s \times \sin\left(\frac{5x_1}{2}\right) - 2, \qquad (24)$$

where $\mathbf{x} = [x_1, x_2]$, the domain $\boldsymbol{\mathcal{X}} = [-4, 7] \times [-3, 8]$, and the second term in the right-hand side introduces a fidelity dependence. This simple extension results in a correlated fidelity dependence, which is continuously differentiable in $\mathcal{X} \times \mathcal{S}$; see Figure 2 for snapshots.

5.1.2. 2D multifidelity four-branches function

A system with four distinct component limit-states [49] is a commonly used test problem in reliability analysis to demonstrate the efficacy of the algorithms in the presence of multiple failure regions. Here, in contrast to (24), we introduce fidelity by making a coordinate transformation as $x_i = x_i' - a_0 \times s$, where $a_0 \in \mathbb{R}$. The response function can be written as

$$f(\mathbf{x}, s) = \min \begin{cases} 3 + 0.1(x_1' - x_2')^2 - \frac{x_1' + x_2'}{\sqrt{2}}, \\ 3 + 0.1(x_1' - x_2')^2 + \frac{x_1' + x_2'}{\sqrt{2}}, \\ x_1' - x_2' + \frac{7}{\sqrt{2}}, \\ x_2' - x_1' + \frac{7}{\sqrt{2}}. \end{cases} \qquad (25)$$

$$\text{and } x_i' = x_i - 5.0 \times s, \ i = 1, 2, 3, \qquad (26)$$

where $\mathbf{x} = [x_1, x_2]$ and $\boldsymbol{\mathcal{X}} = [-8, 8]^2$. The multifidelity extension we provide offers a translation of the level sets across fidelities, where we set $a_0 = -5.0$, and is shown in Figure 3.

5.1.3. 3D Multifidelity Ishigami function

We modify the 3D Ishigami test function [17], similar to the four-branches function, by introducing a shift in the variables using the fidelity parameter s, as shown in (27). Note that the input $\mathbf{x}$ is normalized to be in $[0, 1]^3$ before introducing the translation:

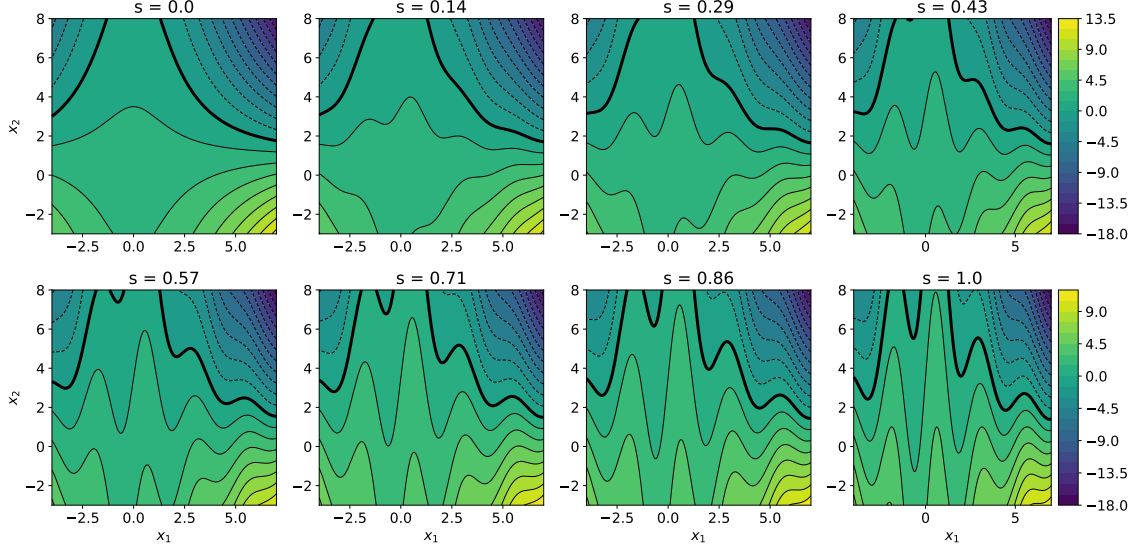$$f(\mathbf{x}, s) = \sin(x_1 - s) + 7.0 \sin^2(x_2 - s) + 0.1x_3^4 \sin(x_1 - s), \qquad (27)$$

17

Figure 2: Multifidelity multimodal function; the thick black line represents the limit state level set $g(f(\mathbf{x}, \cdot)) = 0$.
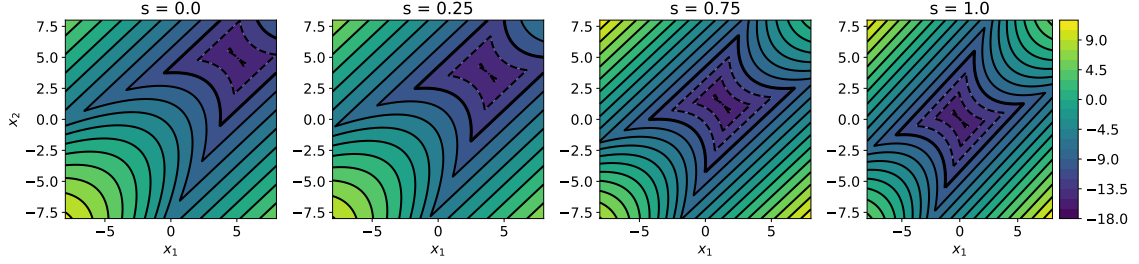

Figure 3: Multifidelity four-branches test function. The thick black line represents the failure boundary $g(f(\mathbf{x})) = 0$.

where $\mathbf{x} = [x_1, x_2, x_3]$ and $\boldsymbol{\mathcal{X}} = [-\pi, \pi]^3$. The resulting function is shown for two slices $x_3 = -\pi$ and $x_3 = 0$ in Figure 4.

5.1.4. 6D Hartmann function

For a higher-dimensional test function, we use the "augmented" Hartmann 6D function introduced in [3] and restated below:

$$f(\mathbf{x}, s) = -(\beta_1 - 0.1 * (1 - s)) * \exp\left(-\sum_{j=1}^{6} A_{1j}(x_j - P_{1j})^2\right) - \sum_{i=2}^{4} \beta_i \exp\left(-\sum_{j=1}^{6} A_{ij}(x_j - P_{ij})^2\right),$$

(28)

where $\mathbf{x} = [x_1, \ldots, x_6]$ and $\boldsymbol{\mathcal{X}} = [0, 1]^6$ and

$$A = \begin{bmatrix} 10 & 3 & 17 & 3.5 & 1.7 & 8 \\ 0.05 & 10 & 17 & 0.1 & 8 & 14 \\ 3 & 3.5 & 1.7 & 10 & 17 & 8 \\ 17 & 8 & 0.05 & 10 & 0.1 & 14 \end{bmatrix}$$

18
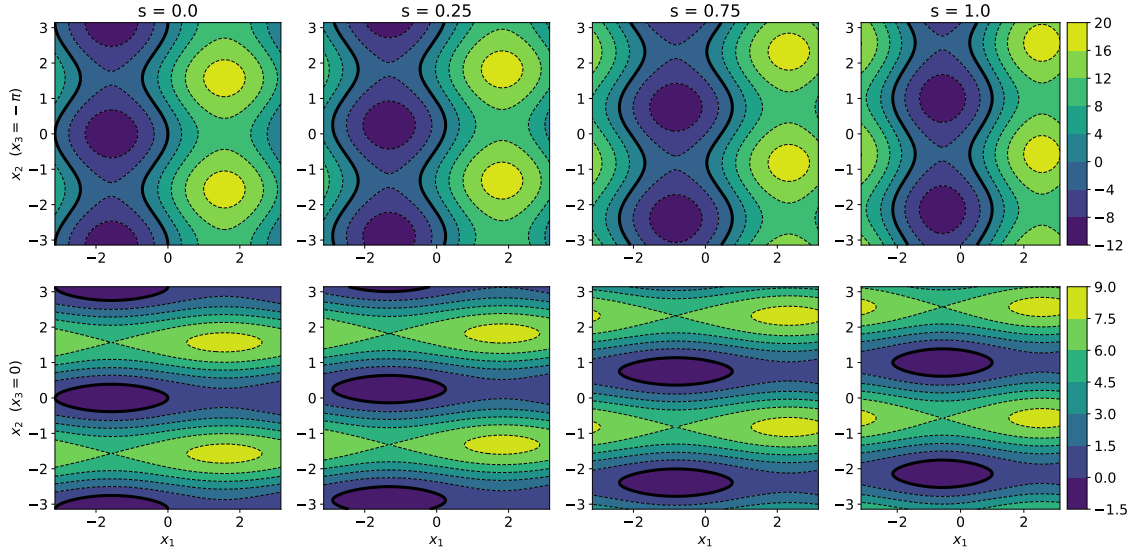
Figure 4: Multifidelity extension of the Ishigami [17] test function.

(a) Ishigami
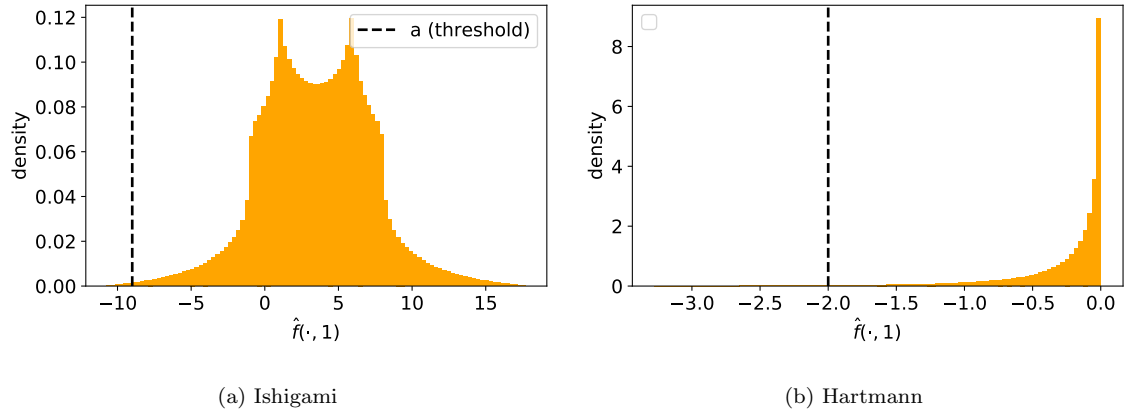
(b) Hartmann

Figure 5: Histograms of the high-fidelity output $\hat{f}(\cdot, 1)$ of the multifidelity 3D Ishigami (left) and 6D Hartmann test functions. Vertical lines indicate failure thresholds.

and

$$
P = \begin{bmatrix}
1312 & 1696 & 5569 & 124 & 8283 & 5886 \\
2329 & 4135 & 8307 & 3736 & 1004 & 9991 \\
2348 & 1451 & 3522 & 2883 & 3047 & 6650 \\
4047 & 8828 & 8732 & 5743 & 1091 & 381
\end{bmatrix}.
$$

For the 2D test functions previously discussed, the failure threshold was set as $a = 0.$. For the Ishigami and Hartmann functions, we show the histogram of the output at $s = 1$ along with the failure thresholds (vertical lines), which are set to $a = -9.0$ and $a = -2.0$, respectively, in Figure 5; these thresholds were chosen particularly to make the failure probabilities on the order of 1%.

The results of the 2D test functions are shown in terms of the predicted contours in Figure 6 and Figure 7; in both figures, whereas the true level set is plotted in the red dashed-dotted line, the background (filled contours) shows the final GP posterior mean $\mu_q(\mathbf{x}, 1)$ with the predicted level set as the thick black line. We hope that readers are able to visually compare the predictions against the true failure boundary. Note that

(a) Single-fidelity prediction   (b) Multifidelity prediction
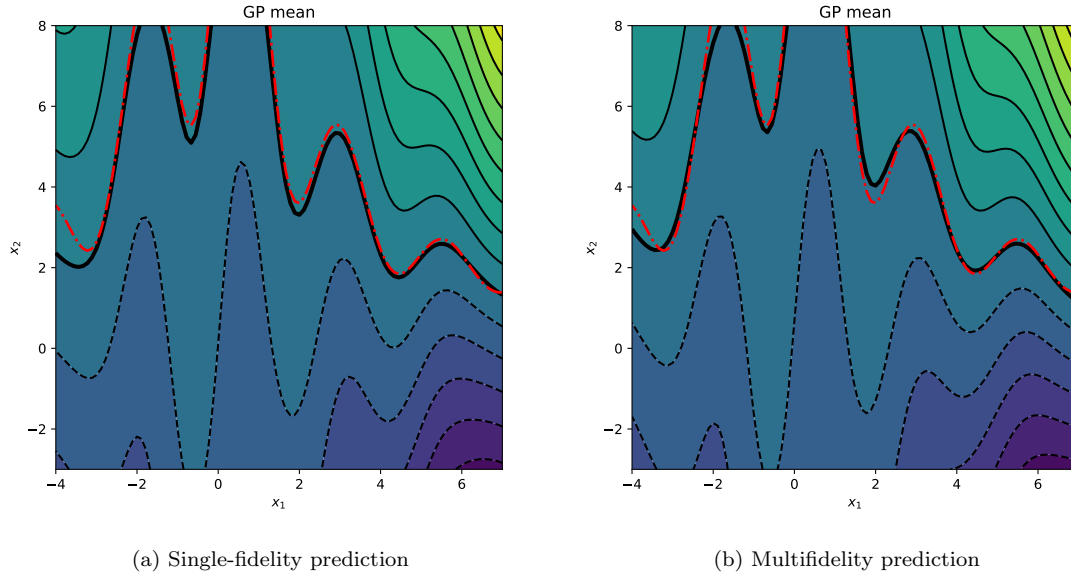
Figure 6: Results of the 2D multimodal test function; red dash-dotted line represents the true level set. Cost for multifidelity prediction is $30,010$ and for single-fidelity prediction $55,000$. Both figures share the same colormap.

in both figures we show only the predictions of one of the twenty random repetitions. Whereas for the multimodal function the predictions from the single- and multifidelity approaches are similar, the multifidelity achieves this prediction with a computational cost that is more than 45% cheaper. For the four-branches test function, the prediction of the multifidelity approach is visually much better than that of the single-fidelity approach, albeit using only half the computational cost.

We use the contours to provide a visual interpretation of the results, since they are feasible for the 2D cases. For a quantitative evaluation, however, we also show the error in computing the failure probability. The absolute errors in computing the failure volume are shown in Figure 8 for all the synthetic test functions. We run both the single- and multifidelity experiments for 100 iterations after providing them with seed samples, as previously mentioned. In all four cases, we observe that the multifidelity approach is able to predict the failure volume with greater accuracy compared with the single-fidelity approach, for a certain computational budget.

5.2. Turbine reliability analysis with discrete fidelity space

We now demonstrate our method on the structural analysis of a gas turbine blade at steady-state operating conditions. Gas turbines have a radial arrangement of blades downstream of the combustor. Hence, these blades are subject to very high pressures and temperatures that can cause concentrated zones of high stress that can in turn lead to structural deformation and potential catastrophic damage of the engine parts. Therefore, we are interested in the probability that the maximum stress acting on the thermal blade exceeds a certain threshold.
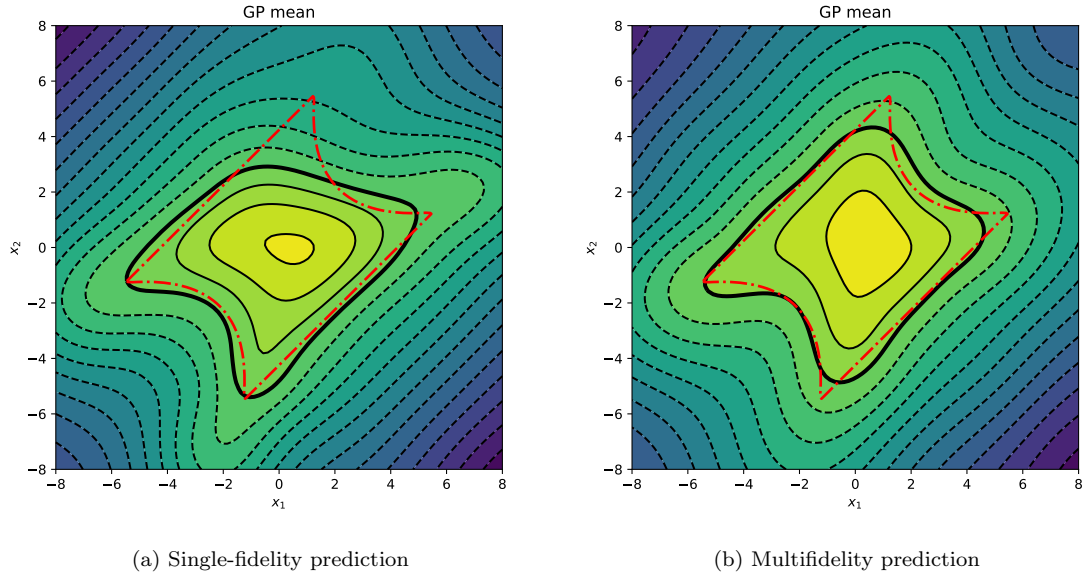
20

(a) Single-fidelity prediction   (b) Multifidelity prediction

Figure 7: Results of the 2D four-branches test function; red dash-dotted line represents the true level set. Cost for multifidelity prediction is $30,005$ and for single-fidelity prediction $55,000$. Both figures share the same colormap.

$$x_1 : \text{Pressure-side loading}$$
$$x_2 : \text{Suction-side loading}$$
$$s : \text{mesh resolution,}$$

In this regard we vary the pressure-side and suction-side pressures as boundary conditions, as shown above, while ignoring thermal effects, to observe the maximum von Mises stress acting on the blade; that is, $f(\mathbf{x}) = \max_{\mathbf{z}} u(\mathbf{z})$, where $u(\mathbf{z})$, $\mathbf{z} \in \partial \mathbf{z}$ is the Von Mises stress distribution on the blade surface $\partial \mathbf{z}$. Note that we negate the output, $f(\mathbf{x}) \times -1$, to comply with our definition of failure. The material properties are fixed to the following constants: Young's modulus $227E + 9$ Pa, Poisson's ratio 0.27, and coefficient of thermal expansion $12.7E - 6$. The governing equations are solved using a finite element method using Matlab's PDE toolbox. The computational mesh density is determined via a minimum element size which also serves as the fidelity parameter. Specifically, we use the minimum mesh element sizes $\{0.05, 0.025, 0.01\}$, which correspond to fidelities $\mathcal{S} = \{0.0, 0.5, 1.0\}$, respectively; note that this case corresponds to a discrete fidelity space. The mesh resolutions for the specified three fidelity levels are shown in Figure 9. As previously mentioned, the maximum von Mises stress is the quantity of interest in this experiment; see Figure 10 for a couple of snapshots of the von Mises stress distribution on the blade corresponding to two different pressure loadings.

The suction-side x_1 and pressure-side x_2 pressures are varied in the following ranges: $x_1 \in [201160, 698390]$, $x_2 \in [301460, 598020]$. A total of $n = 40$ solution snapshots are generated and supplied as seed points for the multifidelity run; the equivalent n for the single-fidelity run is determined to match the multifidelity seed budget. Both experiments are then run for 100 iterations and are repeated 5 times with randomized seed points.
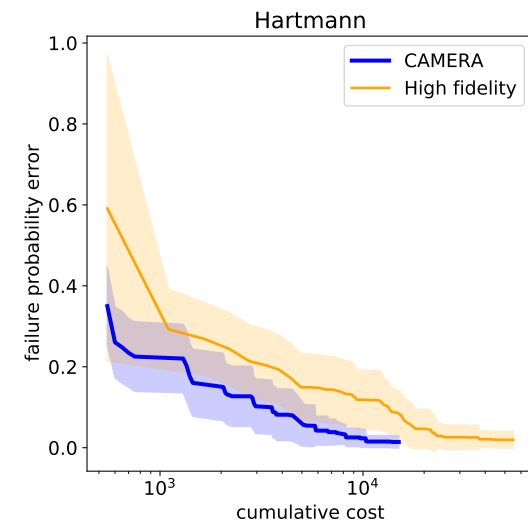
21
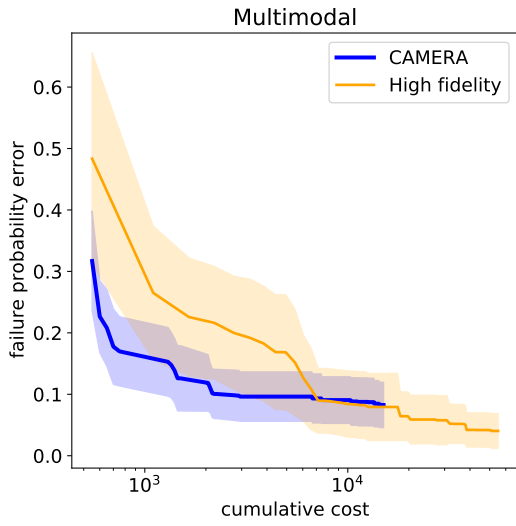
(a) $a = 0.0$, $p_{\mathcal{F}} = 0.30215$
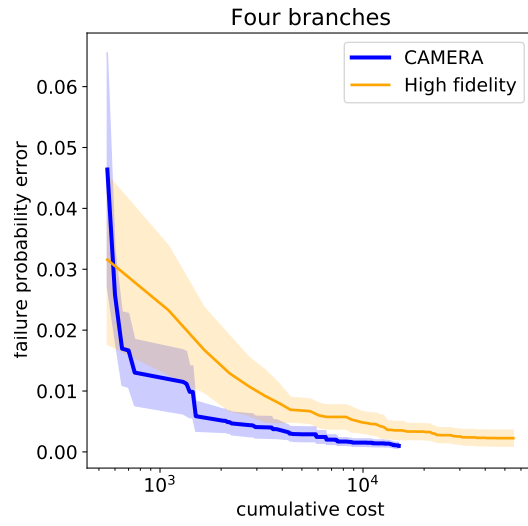
(b) $a = 0.0$, $p_{\mathcal{F}} = 0.1689$
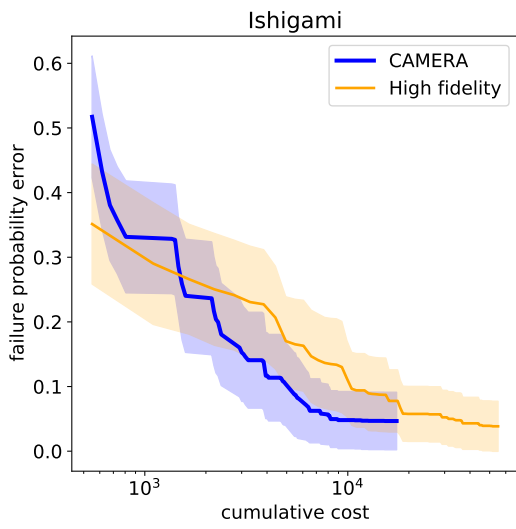
(c) $a = -9.0$, $p_{\mathcal{F}} = 0.0011$

(d) $a = -2.0$, $p_{\mathcal{F}} = 0.00737$

Figure 8: Error in computing the failure probability for the synthetic test functions.

The contour plots of maximum von Mises stress are shown in Figure 11 for each of the three fidelities considered. The failure boundary is set as the level set corresponding to 95% of the maximum von Mises stress across a total of 150 high-fidelity simulations that were computed offline. This is shown by the thick black line in Figure 11 (far right) and corresponds to a probability of failure of 0.004236. Notice that for the low-fidelity case ($s = 0.0$) this level set is outside the domain of $\hat{f}(\mathbf{x}, 0.0)$ and for the medium-fidelity case ($s = 0.5$) this level set is at the top-left corner of the domain. Therefore, this test case poses a unique challenge of the fidelity levels leading to very different function domains; for example, according to the low-fidelity model, the probability of failure is 0.

To evaluate our algorithm, we compute the predicted failure probability via IS at the end of each iteration and take the absolute difference from the true failure probability. This is plotted against the cumulative computational costs for the single- and multifidelity cases in Figure 12. From the figure, notice that for any given cumulative cost, the multifidelity approach results in a failure probability error that is better than or equal to the single-fidelity approach. Because of the availability of only three discrete fidelity levels and the assumed cost model, the discrepancies in the cost between low and high fidelities are very pronounced; the single fidelity curve starts at a cumulative cost of ≈ 550 (which is the cost of a high-fidelity query) and the multifidelity curve starts at a cumulative cost of ≈ 50 (which is the cost of the lowest-fidelity query).

(a) Low fidelity ($s = 0.0$)   (b) Medium fidelity ($s = 0.5$)   (c) High fidelity ($s = 1.0$)
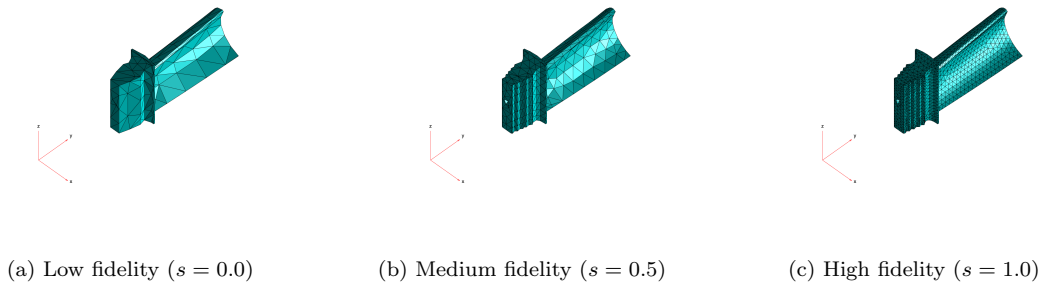
Figure 9: Mesh resolutions for the three discrete fidelities considered.

5.3. ONERA transonic wing

The ONERA-M6 wing [48] is an analytical (shape prescribed using mathematical equations) wing created in the 1970s by the French Aerospace Laboratory for the purposes of understanding three-dimensional airflow in transonic speeds and high Reynolds numbers.[3] It is a semi-span, swept wing with no twist, and at transonic Mach numbers is characterized by local supersonic flow, shocks, and turbulent boundary layer separation. Therefore, this wing has been frequently used for benchmarking computational software for compressible turbulent flow [29]. Because of its relevance in the commercial and military aircraft flight regimes, which make it an important candidate for reliability analysis, we use it as one of the test cases in this work.

---

[3]https://www.onera.fr/en/news/onera-m6-wing-star-of-cfd
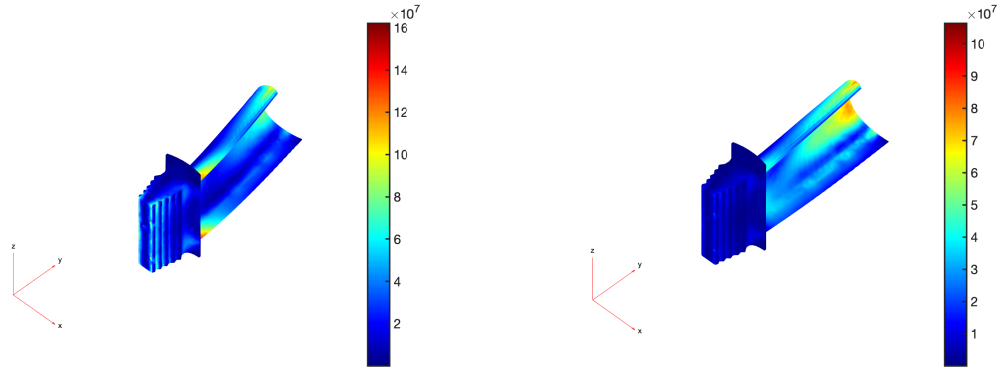
(a) Solution at $\mathbf{x} = [7E5, 4.5E5]$ (b) Solution at $\mathbf{x} = [5E5, 4.5E5]$

Figure 10: Von Mises stress distribution and deformation on the gas turbine blade, under two different pressure loading boundary conditions.
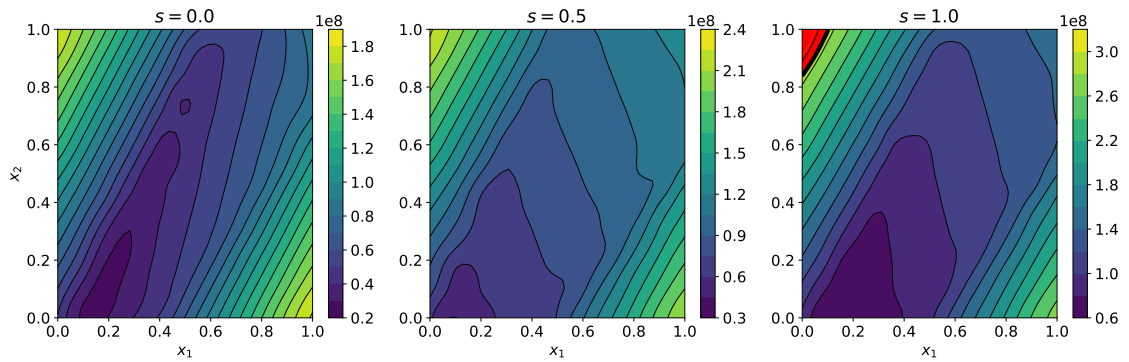

Figure 11: Contours of *maximum* von Mises stress. The failure boundary (thick black line) is outside the range of the low-fidelity approximations $\hat{f}(\mathbf{x}, 0.0)$ and $\hat{f}(\mathbf{x}, 0.5)$.
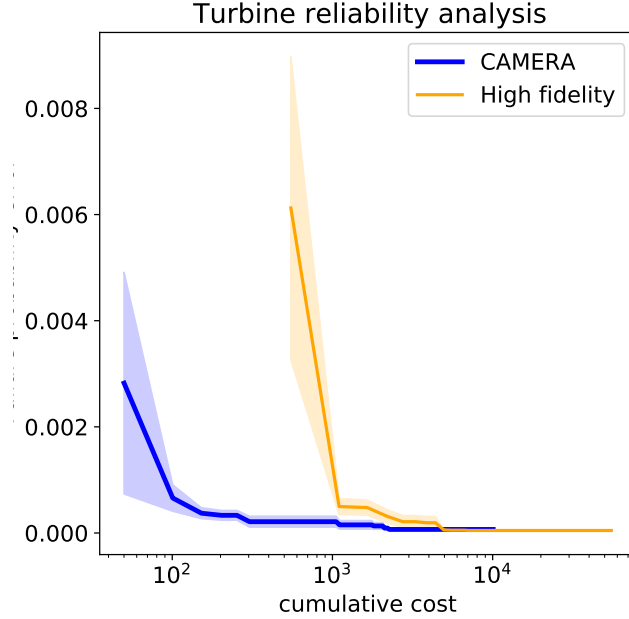
Figure 12: Results of the reliability analysis of the turbine blade.

The wing, shown in Figure 13a, is surrounded by a fluid domain that is a quadrant of an ellipsoid, as shown in Figure 13b. The two external sides of the fluid domain are designated as "freestream" boundary conditions, with a prescribed ambient temperature, flow speeds, and angle of attack (that is, the incidence angle between the flight direction and the wing mean aerodynamic chord). The fluid domain is discretized with a mesh of approximately $56,000$ hexahedral elements with local refinement to resolve boundary layers, shocks, and shear layers. The three-dimensional compressible RANS equations are solved with a finite-volume method with second-order spatial discretization and Euler implicit discretization. The turbulence closure is provided by the Spalart–Allmaras [1] turbulence model. All the simulations are performed with the open-source, finite-volume-based multiphysics code SU2. [4] [11]

For the reliability analysis, we consider the following input parameters:

$$x_1 :\text{Reynolds number}$$
$$x_2 :\text{Mach number}$$
$$x_3 :\text{Angle of attack}$$
$$s :\text{Iterations to converge.}$$

Therefore, $\mathbf{x} = [x_1, x_2, x_3]$, and our $f(\mathbf{x})$ is the lift-to-drag ratio L/D, where L is the net aerodynamic force acting perpendicular to the flight path and D is in the direction parallel to the flight path. Note that the L/D is also known as the "aerodynamic efficiency," and hence the higher its value, the better. However, the maximum lift obtainable during flight is limited by the structural limits of the aircraft wing [2, Ch. 4], and hence we want to ensure the L/D does not exceed a certain safety threshold,

---

[4]https://su2code.github.io/

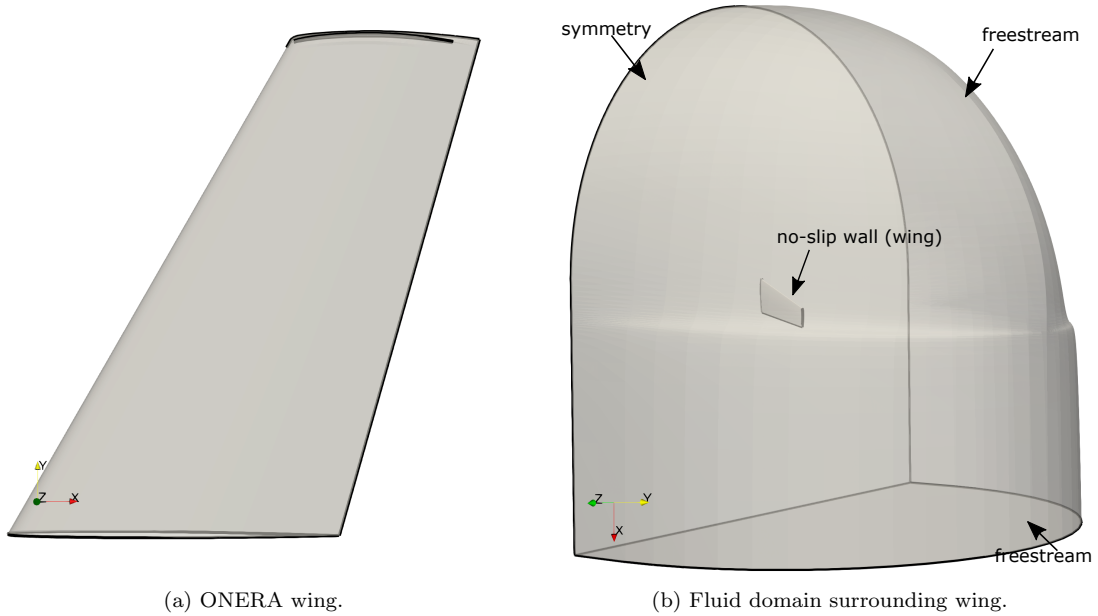(a) ONERA wing. (b) Fluid domain surrounding wing.

Figure 13: Fluid domain and boundary conditions for the ONERA transonic wing.

over a range of operating conditions.

The fidelity parameter s is the number of iterations allowed for the solver to converge; fewer iterations result in a faster (and hence cheaper) solution that is considered low fidelity and vice versa. We set the following bounds on the control and fidelity parameters: $x_1 \in [5 \times 10^6, 11 \times 10^6]$, $x_2 \in [0.7, 0.85]$, $x_3 \in [1, 3]$, and $\mathcal{S} \in [0, 1]$ (which corresponds to the number of iterations $\in [500, 5000]$). Note that we round any fractional values for the number of iterations, recommended via our algorithm, to the nearest integer. We show the pressure coefficient distribution on the wing for the different values of s in Figure 14. Notice that while the visual differences in the flow field are rather subtle, the L/D value can be drastically different between an unconverged and fully converged solution. As in the turbine test case, we negate the output of the function f and set the failure threshold $a = -18$ and therefore are interested in estimating the probability that $g(\mathbf{x}) = f(\mathbf{x}) - a \leq 0$.

We begin our algorithm with $n = 100$ seed points obtained via running the simulations at a Latin hypercube design of experiments across the joint $\mathcal{X} \times \mathcal{S}$ space, and we run the algorithm for 100 iterations. For the single-fidelity run, n is set to keep the total cost of seed points identical to that of the multifidelity run. We keep the same cost model as in Figure 1. Note that for this experiment we compute the failure probability only once—at the conclusion of running Algorithm 2 for 100 iterations. Furthermore, we do not repeat the experiment with randomized seed points. We fix $N = 400$ high-fidelity samples to evaluate $p_{\mathcal{F}}^{\mathrm{IS}}$.

We show the results in Figure 15, where the predictions of $p_{\mathcal{F}}^{\mathrm{IS}}$ are shown in fig. 15a. While we do not know the $p_{\mathcal{F}}$ (truth) for this case, we see that the predictions from CAMERA are with a smaller confidence interval compared with using the high-fidelity model only. The overall cumulative cost of the high-fidelity run is 44,667 compared with 38,614 for CAMERA. Furthermore, in fig. 15b we show the histograms of the high-fidelity samples sampled per the biasing distributions fit from the high-fidelity-only run and CAMERA. Notice that whereas CAMERA biases points toward the failure domain (red

26

(a) Unconverged (500 iterations). $L/D = 14.70$ (b) Unconverged (2500 iterations). $L/D = 14.67$

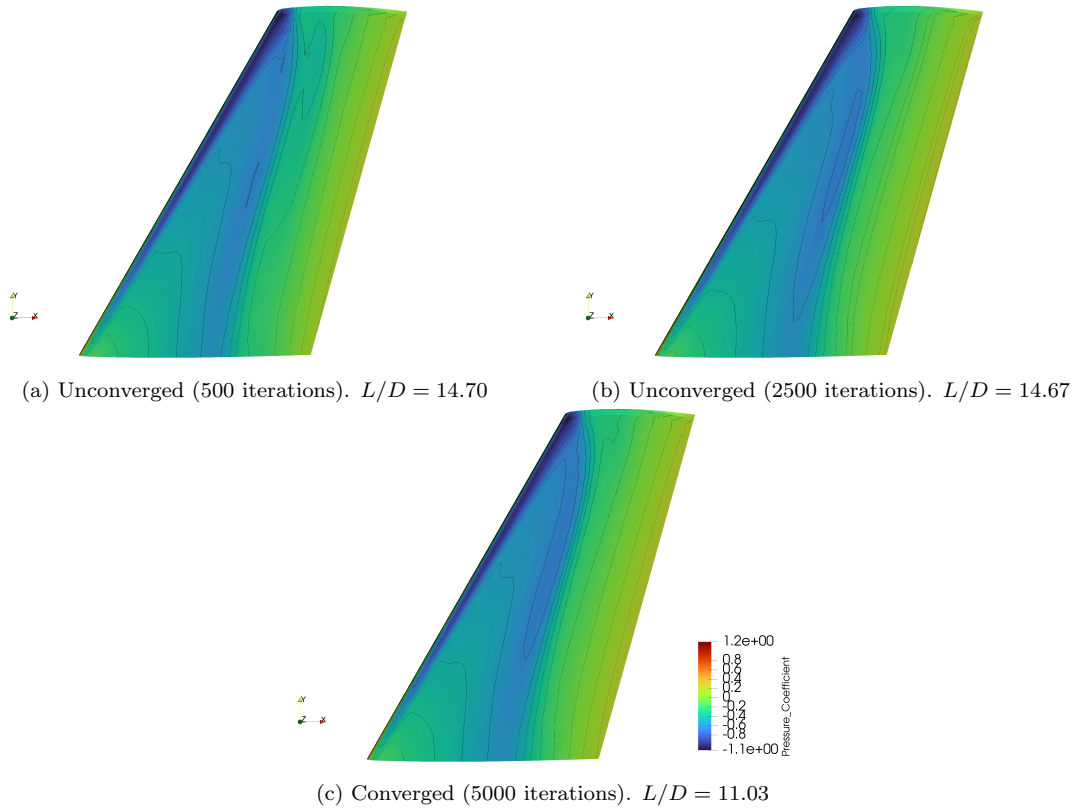(c) Converged (5000 iterations). $L/D = 11.03$

Figure 14: Surface pressure coefficient distribution for the ONERA wing at 3 different levels of convergence and fixed boundary conditions.

shaded region in the plot), the high-fidelity-only run results in a bi-modal distribution that is not expected. We use this as anecdotal evidence that our CAMERA algorithm performs better for this experiment.
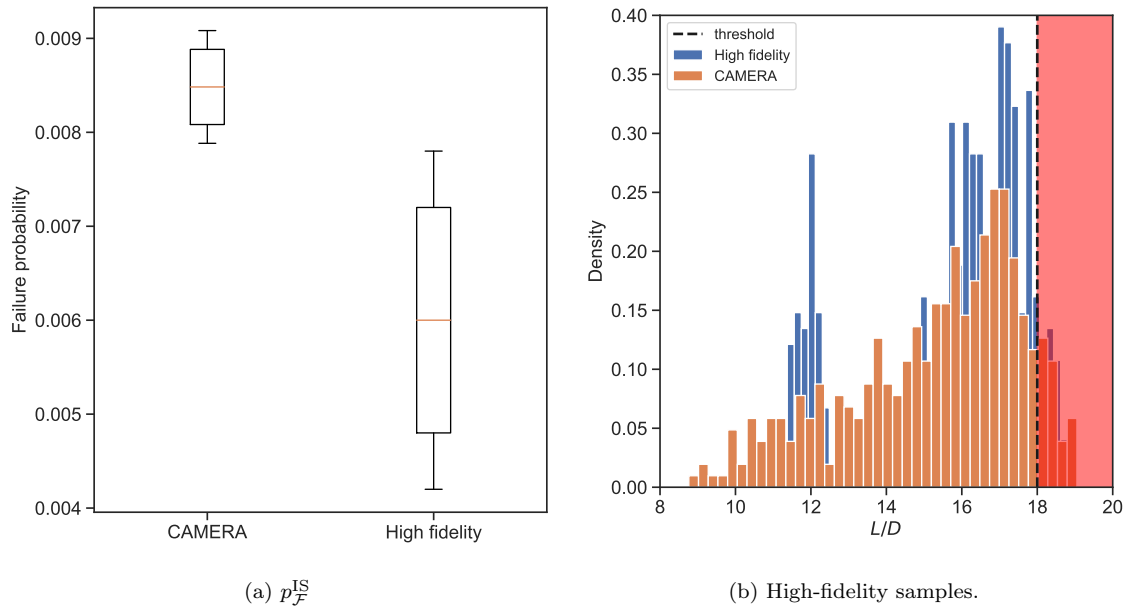
(a) $p_{\mathcal{F}}^{\mathrm{IS}}$

(b) High-fidelity samples.

Figure 15: Predicted $p_{\mathcal{F}}^{\mathrm{IS}}$ via CAMERA and using the high-fidelity model only (left) and the histograms of high-fidelity samples sampled according to the biasing distribution (right).

6. Conclusion

In this work we address the problem of failure probability estimation for complex engineered systems with a computationally expensive model. Additionally, we account for the fact that the expensive model has tunable fidelity parameters that can trade computational cost for predictive accuracy. We propose an approach based on Gaussian process models that can learn the correlations between the different fidelities, given a continuous or discrete fidelity space, and therefore predict the high-fidelity function with potentially fewer queries to the expensive high-fidelity model that are supplemented by cheaper low-fidelity model queries. Further, we propose an acquisition-based sequential experiment design approach to adaptively build the multifidelity surrogate model, so that only the most *useful* fidelities and inputs are queried. The main highlight of our proposed acquisition function is that it provides a generalized multifidelity framework that can extend existing (single-fidelity) acquisition functions to the multifidelity setting and furthermore applies to both discrete and continuous fidelity spaces. Additionally, we demonstrate that our surrogate modeling approach is asymptotically consistent, meaning that the predicted failure level set approaches the true level set with high probability as the number of iterations approaches infinity. We also show that estimates of the failure probability, predicted by using only the developed multifidelity surrogate model, produces very high accuracy at a lower computational cost, compared with using the high-fidelity model only.

With the sequential surrogate model building approach, we observe that the predicted failure level sets are very close to the true level set. Furthermore, we observe that in the presence of multiple fidelity models, our approach performs even better by requiring only $\approx 50\%$ of the computational budget. In terms of predicting the failure probability, our approach shows a decrease in the prediction error with increasing evaluations of the models, while the multifidelity approach always outperforms the single-fidelity case. In addition to the synthetic cases, we demonstrate our method

on a real-world turbine reliability case, with discrete fidelities, where the failure probability is $\approx 0.4\%$, and a transonic wing test case. We observe results consistent with the synthetic cases where we are able to obtain very high accuracy in failure probability error, with the multifidelity approach significantly outperforming using the high-fidelity model only. Overall, our experiments reveal that our approach can lead to cost-effective yet accurate predictions for reliability analysis, for example, using only $N = 400$ high-fidelity samples for estimating $p_{\mathcal{F}}^{\mathrm{IS}}$ in the transonic wing example.

We anticipate conducting follow-up studies of this work in several directions. First, the `CAMERA` algorithm can be modified to adaptively provide an estimate for $p_{\mathcal{F}}^{\mathrm{IS}}$; currently, it is estimated at the conclusion of the adaptive surrogate model construction. This is useful for real-world applications, where running expensive high-fidelity simulations in multiple batches is less risky and therefore more practical. Along the same lines, we plan on evaluating the performance of our sequential experiment design of adaptive surrogate model construction, in making batch evaluations at every step. Furthermore, whereas we have demonstrated our approach in estimating $p_{\mathcal{F}}$ in $\mathcal{O}(10^{-4})$, it would be interesting to evaluate the approach on problems with smaller orders of magnitude in $p_{\mathcal{F}}$.

Appendix

We vary the choice of the constant c_1 in the cost model to observe its impact on the algorithm performance. Recall that c_1 determines the rate of change of the computational cost of the models with respect to fidelity. In this regard, smaller values of c_1 result in "flatter" cost models, as shown in fig. 16a. We use the Ishigami test function to evaluate the impact of the choice of cost model on the performance of `CAMERA`, whose results, with 20 repetitions with randomized seed points, are shown in fig. 16b. We observe no appreciable difference in the algorithm performance between a choice of $c_1 = 10$ and $c_1 = 5$. At $c_1 = 1.0$, `CAMERA` still outperforms the high-fidelity-only run, although suffering a higher failure probability error at lower cumulative costs. As c_1 approaches 0, the performance approaches that of the high-fidelity-only run, as expected. Therefore, we conclude that the choice of the cost model, while expected to impact the performance of the algorithm, does not necessarily offer an undue advantage to the algorithm. This experiment justifies the use of $c_1 = 10.0$ in all of our experiments.

Acknowledgments

References

[1] Steven R Allmaras and Forrester T Johnson. Modifications and clarifications for the implementation of the spalart-allmaras turbulence model. In *Seventh international conference on computational fluid dynamics (ICCFD7)*, volume 1902. Big Island, HI, 2012.
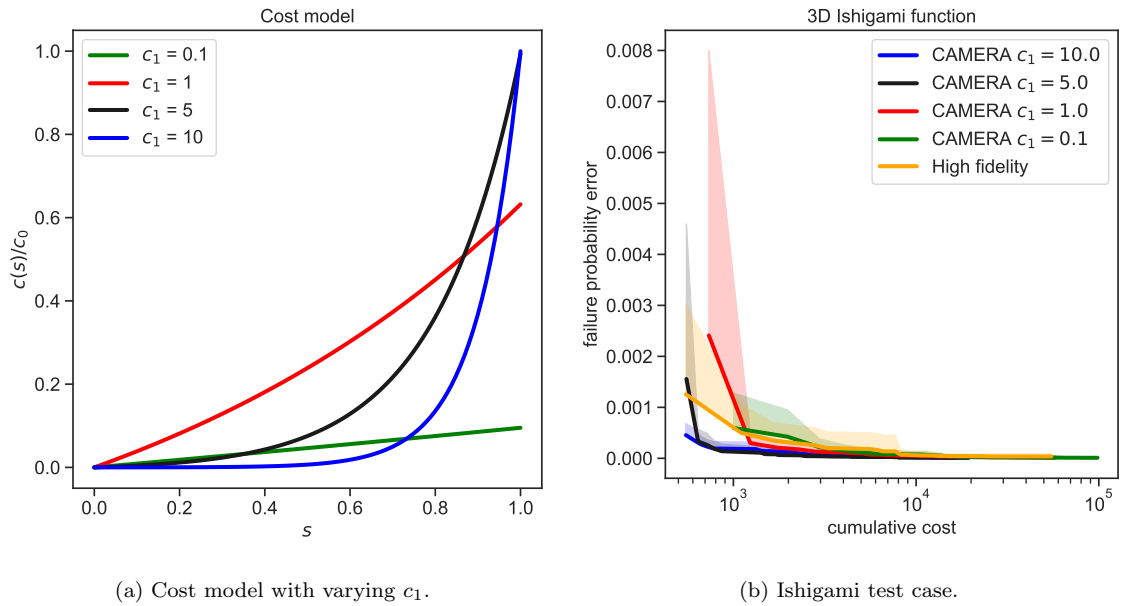
(a) Cost model with varying c_1.

(b) Ishigami test case.

Figure 16: Impact of the choice of cost model on the algorithm performance.

[2] John David Anderson. *Aircraft performance and design*, volume 1. WCB/McGraw-Hill Boston, 1999.

[3] Maximilian Balandat, Brian Karrer, Daniel R Jiang, Samuel Daulton, Benjamin Letham, Andrew Gordon Wilson, and Eytan Bakshy. BoTorch: Programmable Bayesian optimization in PyTorch. *arXiv:1910.06403*, 2019.

[4] Julien Bect, David Ginsbourger, Ling Li, Victor Picheny, and Emmanuel Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22(3):773–793, 2012.

[5] Barron J Bichon, Michael S Eldred, Laura Painton Swiler, Sandaran Mahadevan, and John M McFarland. Efficient global reliability analysis for nonlinear implicit performance functions. *AIAA Journal*, 46(10):2459–2468, 2008.

[6] Mickaël Binois, Jiangeng Huang, Robert B Gramacy, and Mike Ludkovski. Replication or exploration? sequential design for stochastic simulation experiments. *Technometrics*, 61(1):7–23, 2019.

[7] D Austin Cole, Robert B Gramacy, James E Warner, Geoffrey F Bomarito, Patrick E Leser, and William P Leser. Entropy-based adaptive design for contour finding and estimating reliability. *arXiv preprint arXiv:2105.11357*, 2021.

[8] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.

[9] Vincent Dubourg, Bruno Sudret, and Franois Deheeger. Metamodel-based importance sampling for structural reliability analysis. *Probabilistic Engineering Mechanics*, 33:47–57, 2013.

[10] B Echard, N Gayton, and M Lemaire. Kriging-based Monte Carlo simulation to compute the probability of failure efficiently: AK-MCS method. *6emes Journées Nationales de Fiabilité, 24–26 mars, Toulouse, France*, 2010.

[11] Thomas D Economon, Francisco Palacios, Sean R Copeland, Trent W Lukaczyk, and Juan J Alonso. SU2: An open-source suite for multiphysics simulation and design. *Aiaa Journal*, 54(3):828–846, 2016.

[12] Peter I Frazier. A tutorial on Bayesian optimization. *arXiv preprint arXiv:1807.02811*, 2018.

[13] Subhashis Ghosal and Anindya Roy. Posterior consistency of Gaussian process prior for nonparametric binary regression. *The Annals of Statistics*, 34(5):2413–2429, 2006.

[14] Alkis Gotovos, Nathalie Casati, Gregory Hitz, and Andreas Krause. Active learning for level set estimation. In *Twenty-Third International Joint Conference on Artificial Intelligence*, 2013.

[15] John Hammersley. *Monte carlo methods*. Springer Science & Business Media, 2013.

[16] Philipp Hennig and Christian J. Schuler. Entropy search for information-efficient global optimization. *Journal of Machine Learning Research*, 13(Jun):1809–1837, 2012.

[17] Tsutomu Ishigami and Toshimitsu Homma. An importance quantification technique in uncertainty analysis for computer models. In *[1990] Proceedings. First International Symposium on Uncertainty Modeling and Analysis*, pages 398–403. IEEE, 1990.

[18] Friedman Jerome, Trevor Hastie, and Robert Tibshirani. The elements of statistical learning. In *Springer series in statistics*, volume 1. Springer, 2001.

[19] Donald R. Jones, Matthias Schonlau, and William J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4):455–492, 1998. doi: 10.1023/A:1008306431147.

[20] Herman Kahn and Andy W Marshall. Methods of reducing sample size in Monte Carlo computations. *Journal of the Operations Research Society of America*, 1 (5):263–278, 1953.

[21] Kirthevasan Kandasamy, Gautam Dasarathy, Jeff Schneider, and Barnabás Póczos. Multi-fidelity Bayesian optimisation with continuous approximations. *arXiv:1703.06240*, 2017.

[22] Marc C Kennedy and Anthony O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3): 425–464, 2001.

[23] Aaron Klein, Stefan Falkner, Simon Bartels, Philipp Hennig, and Frank Hutter. Fast Bayesian optimization of machine learning hyperparameters on large datasets. In *Artificial Intelligence and Statistics*, pages 528–536. PMLR, 2017.

[24] Boris Kramer, Alexandre Noll Marques, Benjamin Peherstorfer, Umberto Villa, and Karen Willcox. Multifidelity probability estimation via fusion of estimators. *Journal of Computational Physics*, 392:385–402, 2019.

[25] Harold J. Kushner. A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. *Journal of Basic Engineering*, 86(1): 97–106, 1964. doi: 10.1115/1.3653121.

[26] Jing Li and Dongbin Xiu. Evaluation of failure probability via surrogate models. *Journal of Computational Physics*, 229(23):8966–8980, 2010.

[27] Jing Li, Jinglai Li, and Dongbin Xiu. An efficient surrogate-based method for computing rare failure probability. *Journal of Computational Physics*, 230(24): 8683–8697, 2011.

[28] David JC MacKay. Information-based objective functions for active data selection. *Neural Computation*, 4(4):590–604, 1992. doi: 10.1162/neco.1992.4.4.590.

[29] M Mani, J Ladd, A Cain, R Bush, M Mani, J Ladd, A Cain, and R Bush. An assessment of one-and two-equation turbulence models for internal and external flows. In *28th Fluid Dynamics Conference*, page 2010, 1997.

[30] Alexandre N Marques, Remi R Lam, and Karen E Willcox. Contour location via entropy reduction leveraging multiple information sources. *arXiv preprint arXiv:1805.07489*, 2018.

[31] Jonas Mockus, Vytautas Tiešis, and Antanas Žilinskas. The application of Bayesian methods for seeking the extremum. *Towards Global Optimization*, 2: 117–129, 1978.

[32] Benjamin Peherstorfer, Tiangang Cui, Youssef Marzouk, and Karen Willcox. Multifidelity importance sampling. *Computer Methods in Applied Mechanics and Engineering*, 300:490–509, 2016.

[33] Paris Perdikaris, Maziar Raissi, Andreas Damianou, Neil D Lawrence, and George Em Karniadakis. Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2198):20160751, 2017.

[34] Victor Picheny, David Ginsbourger, Olivier Roustant, Raphael T Haftka, and Nam-Ho Kim. Adaptive designs of experiments for accurate approximation of a target region. 2010.

[35] Matthias Poloczek, Jialei Wang, and Peter I Frazier. Multi-information source optimization. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 4291–4301, 2017.

[36] Rüdiger Rackwitz. Reliability analysis – a review and some perspectives. *Structural Safety*, 23(4):365–395, 2001.

[37] Dushhyanth Rajaram, Tejas G Puranik, Ashwin Renganathan, Woong Je Sung, Olivia J Pinon-Fischer, Dimitri N Mavris, and Arun Ramamurthy. Deep gaussian process enabled surrogate models for aerodynamic flows. In *AIAA Scitech 2020 Forum*, page 1640, 2020.

[38] Dushhyanth Rajaram, Tejas G Puranik, S Ashwin Renganathan, WoongJe Sung, Olivia Pinon Fischer, Dimitri N Mavris, and Arun Ramamurthy. Empirical assessment of deep Gaussian process surrogate models for engineering problems. *Journal of Aircraft*, 58(1):182–196, 2021.

[39] Pritam Ranjan, Derek Bingham, and George Michailidis. Sequential experiment design for contour estimation from complex computer codes. *Technometrics*, 50 (4):527–541, 2008.

[40] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006. doi: 10.7551/mitpress/3206.001.0001.

[41] S Ashwin Renganathan. Koopman-based approach to nonintrusive reduced order modeling: Application to aerodynamic shape optimization and uncertainty propagation. *AIAA Journal*, 58(5):2221–2235, 2020.

[42] S Ashwin Renganathan, Yingjie Liu, and Dimitri N Mavris. Koopman-based approach to nonintrusive projection-based reduced-order modeling with black-box high-fidelity models. *AIAA Journal*, 56(10):4087–4111, 2018.

[43] S Ashwin Renganathan, Romit Maulik, and Vishwas Rao. Machine learning for nonintrusive model order reduction of the parametric inviscid transonic flow past an airfoil. *Physics of Fluids*, 32(4):047110, 2020.

[44] S Ashwin Renganathan, Jeffrey Larson, and Stefan M Wild. Lookahead acquisition functions for finite-horizon time-dependent Bayesian optimization and application to quantum optimal control. *arXiv preprint arXiv:2105.09824*, 2021.

[45] S Ashwin Renganathan, Romit Maulik, and Jai Ahuja. Enhanced data efficiency using deep neural networks and gaussian processes for aerodynamic design optimization. *Aerospace Science and Technology*, 111:106522, 2021.

[46] Sudharshan Ashwin Renganathan. *A Methodology for Non-Intrusive projection-based model reduction of expensive black-box PDE-based systems and application in the many-query context*. PhD thesis, Georgia Institute of Technology, 2018.

[47] Walter Rudin. *Real and Complex Analysis P. 2*. McGraw-Hill, 1970.

[48] V Schmitt. Pressure distributions on the ONERA M6-wing at transonic mach numbers, experimental data base for computer program assessment. *AGARD AR-138*, 1979.

[49] Roland Schöbi, Bruno Sudret, and Stefano Marelli. Rare event estimation using polynomial-chaos kriging. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 3(2):D4016002, 2017.

[50] Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando De Freitas. Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2015. doi: 10.1109/jproc.2015.2494218.

[51] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical Bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems*, 25, 2012.

[52] James C Spall. *Introduction to stochastic search and optimization: estimation, simulation, and control*, volume 65. John Wiley & Sons, 2005.

[53] Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. In *Proceedings of the International Conference on Machine Learning*, 2009.

[54] Rajan Srinivasan. *Importance sampling: Applications in communications and detection*. Springer Science & Business Media, 2002.

[55] S. Surjanovic and D. Bingham. Virtual library of simulation experiments: Test functions and datasets. Retrieved May 23, 2020, from `http://www.sfu.ca/~ssurjano`.

[56] Kevin Swersky, Jasper Snoek, and Ryan Prescott Adams. Multi-task Bayesian optimization. 2013.

[57] Shion Takeno, Hitoshi Fukuoka, Yuhki Tsukada, Toshiyuki Koyama, Motoki Shiga, Ichiro Takeuchi, and Masayuki Karasuyama. Multi-fidelity Bayesian optimization with max-value entropy search. *arXiv:1901.08275*, 2019.

[58] Hendrik Tennekes and John L Lumley. *A first course in turbulence*. MIT Press, 2018.

[59] Julien Villemonteix, Emmanuel Vazquez, and Eric Walter. An informational approach to the global optimization of expensive-to-evaluate functions. *Journal of Global Optimization*, 44(4):509–534, 2009.

[60] Hongqiao Wang, Guang Lin, and Jinglai Li. Gaussian process surrogates for failure detection: A Bayesian experimental design approach. *Journal of Computational Physics*, 313:247–259, 2016.

[61] Zi Wang and Stefanie Jegelka. Max-value entropy search for efficient Bayesian optimization. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pages 3627–3635, 2017.

[62] CF Jeff Wu. On the convergence properties of the em algorithm. *The Annals of statistics*, pages 95–103, 1983.