Bayesian updating of reliability by cross entropy-based importance sampling

Oindrila Kanjilal*, Iason Papaioannou, Daniel Straub

Engineering Risk Analysis Group, Technische Universität München, München 80290, Germany

Abstract

Bayesian analysis enables a consistent updating of the failure probability of engineering systems when new data is available. To this end, we introduce an adaptive importance sampling (IS) method based on the principle of cross entropy (CE) minimization. The key contribution is a novel IS density associated with the posterior probability density function (PDF) of the uncertain parameters, that facilitates efficient sampling from the important region of the failure domain, especially when the failure event is rare. The IS density is designed via a two-step procedure. The first step involves construction of a sample-based approximation of the posterior, which we build using the CE method. Here the aim is to determine the parameters of a chosen parametric distribution family that minimize its Kullback-Leibler divergence from the posterior PDF. The second step of the proposed method constructs the desired IS density for sampling the failure domain through a second round of CE minimization, starting from the approximate posterior obtained in the first step. An adaptive, multi-level approach is employed to solve the two CE optimization problems. The IS densities deduced in the two steps are then applied to construct an efficient estimator for the posterior probability of failure. Through numerical studies, we investigate and demonstrate the efficacy of the method in accurately estimating the reliability of engineering systems with rare failure events.

Keywords: Reliability updating, Bayesian analysis, Importance sampling, Cross entropy

1. Introduction

The prediction of reliability lies at the heart of model-based safety assessment of engineering systems. An accurate assessment requires appropriate characterization of the uncertain model parameters, taking into account all available data. Once an engineering system comes into existence, it is possible to obtain information on the system properties and performance through measurements, monitoring and other means of observations. This information can

*Corresponding author

Email addresses: oindrila.kanjilal@tum.de (Oindrila Kanjilal), iason.papaioannou@tum.de (Iason Papaioannou), straub@tum.de (Daniel Straub)

be used to update the model parameters to provide improved estimates of the system's reliability. A consistent framework for assimilating the new information into the models is provided by Bayesian analysis.

Consider the model, $\mathcal{M}$, of an engineering system, characterized by a set of model parameters and boundary conditions that describe the geometry, material properties, loads etc. In many practical applications, some of these parameters are uncertain. Uncertain parameters are modeled by random variables gathered in a random vector $\mathbf{\Theta}$ of dimension $n_{\boldsymbol{\theta}}$. Let the probability density function (PDF) $p_{\mathbf{\Theta}}(\boldsymbol{\theta})$ denote one's prior belief about the distribution of $\mathbf{\Theta}$, i.e., before new information becomes available. The prior probability of failure is given by the multi-dimensional integral

$$P_F = \int_{\boldsymbol{\theta} \in \mathbb{R}^{n_{\boldsymbol{\theta}}}} \mathrm{I}_F(\boldsymbol{\theta}) p_{\mathbf{\Theta}}(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}, \tag{1}$$

where $F = \{\boldsymbol{\theta} \in \mathbb{R}^{n_{\boldsymbol{\theta}}} : g(\boldsymbol{\theta}) \leq 0\}$ is the failure event, defined in terms of the so called limit state function $g(\boldsymbol{\theta})$, and $\mathrm{I}_F(\boldsymbol{\theta})$ is the indicator function such that $\mathrm{I}_F(\boldsymbol{\theta}) = 1$ if $g(\boldsymbol{\theta}) \leq 0$ and $\mathrm{I}_F(\boldsymbol{\theta}) = 0$ otherwise. Computation of the limit state function for an outcome $\boldsymbol{\theta}$ of the uncertain parameters requires evaluation of the system model $\mathcal{M}$. The probability of the complement of F is the reliability of the system.

When new data from the engineering system is available, it can be used to learn the uncertain parameters, thereby updating the prior PDF. The data can be direct observations of the uncertain parameters, $\mathbf{\Theta}$, or measurements of the system response, e.g., measurements of stress condition or deformation. Let $\mathbf{d}$ denote the data that is available in the form of measurements or observations. The updated/posterior PDF of $\mathbf{\Theta}$ that incorporates the data information in the context of $\mathcal{M}$ is given by Bayes' theorem as

$$p_{\mathbf{\Theta}|\mathbf{d}}(\boldsymbol{\theta}) = c_E^{-1} L(\boldsymbol{\theta}|\mathbf{d}) p_{\mathbf{\Theta}}(\boldsymbol{\theta}). \tag{2}$$

$L(\boldsymbol{\theta}|\mathbf{d})$ is the likelihood function that expresses the plausibility of observing $\mathbf{d}$ given a certain $\boldsymbol{\theta}$ and c_E is the normalizing constant that ensures that $p_{\mathbf{\Theta}|\mathbf{d}}(\boldsymbol{\theta})$ integrates to one. It is commonly referred to as the marginal likelihood (or evidence) and is defined as

$$c_E = \int_{\boldsymbol{\theta} \in \mathbb{R}^{n_{\boldsymbol{\theta}}}} L(\boldsymbol{\theta}|\mathbf{d}) p_{\mathbf{\Theta}}(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}. \tag{3}$$

When the data contains measurements of the system response, $L(\boldsymbol{\theta}|\mathbf{d})$ includes the system model $\mathcal{M}$ and Eq. (2) is termed a Bayesian inverse problem.The probability of failure conditional on the data $\mathbf{d}$ is obtained by replacing the prior PDF in Eq. (1) by the posterior PDF

$$P_{F|\mathbf{d}} = \int_{\boldsymbol{\theta} \in \mathbb{R}^{n_{\boldsymbol{\theta}}}} \mathrm{I}_F(\boldsymbol{\theta}) p_{\mathbf{\Theta}|\mathbf{d}}(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \tag{4}$$

Evaluation of the posterior probability of failure requires repeated computation of the limit state function and the likelihood over the outcome space of $\mathbf{\Theta}$, and is challenging due to several reasons. The functions $L(\boldsymbol{\theta}|\mathbf{d})$ and $g(\boldsymbol{\theta})$ are typically evaluated numerically, i.e., they are treated as black-box models. As a consequence, it is impossible to analytically evaluate $P_{F|\mathbf{d}}$ except for some special cases. Additionally, numerical integration of the integrals in Eqs. (3) and (4) is often not feasible due to the large number of random variables involved. Approaches to approximate the posterior probability of failure using the first- and second- order reliability methods [30] or Laplace's asymptotic approximation [32] have been suggested. These methods require evaluations of the first and second derivatives of $L(\boldsymbol{\theta}|\mathbf{d})$ and $g(\boldsymbol{\theta})$, and, hence, might be computationally challenging, especially if the number of model parameters is large or evaluation of $\mathcal{M}$ is costly. Moreover, they are often inaccurate in cases where the data is not informative.

Monte Carlo simulation (MCS) methods offer a robust alternative to numerically evaluate $P_{F|\mathbf{d}}$. Here one can first perform a Bayesian analysis to learn the posterior PDF $p_{\boldsymbol{\Theta}|\mathbf{d}}(\boldsymbol{\theta})$ of the uncertain parameters, and then use samples generated from $p_{\boldsymbol{\Theta}|\mathbf{d}}(\boldsymbol{\theta})$ to evaluate $P_{F|\mathbf{d}}$. Beck and Au [5] propose an adaptive Metropolis-Hastings algorithm to generate samples from $p_{\boldsymbol{\Theta}|\mathbf{d}}(\boldsymbol{\theta})$ and then use these samples to update the reliability by evaluating the reliability conditional on each of these samples. This approach becomes inefficient when the number of uncertain model parameters $n_{\boldsymbol{\theta}}$ is high and the posterior probability of failure is small. Ching and Hsieh [9] propose a method to update the reliability by combining Bayes' theorem with maximum entropy theory. This approach uses standard MCS to fit a set of sampling distributions by the maximum entropy method. The method is suited for high dimensions $n_{\boldsymbol{\theta}}$, but it remains inefficient for small target probabilities.

The limitation of the aforementioned simulation-based approaches in estimating small posterior failure probabilities can be overcome by combining the methods for Bayesian analysis with advanced Monte Carlo methods for rare event estimation. Ching and Beck [7] propose a method for online reliability updating based on an efficient importance sampling technique of Au and Beck [2]. Sundar and Manohar [41] suggest an approach to estimate the posterior probability of failure by applying Girsanov's transformation based importance sampling [29]. The methods in [7, 41] are applicable only if the system is dynamic and the model uncertainties are due to the unknown loading. Efficient approaches to update the reliability in the presence of both structural parameter and loading uncertainties are suggested in [21, 19, 4]. Jensen et al. [21] and Hadjidoukas et al. [19] propose to first update the prior PDF of the model parameters by applying the transitional Markov chain Monte Carlo method [8]. Subsequently, subset simulation [1] is employed for evaluating the conditional probability in Eq. (4) starting from samples of the posterior. The application of subset simulation in conjunction with a Gibbs sampling-based method for Bayesian model updating [6] is explored in Bansal and Cheung [4]. Straub et al. [39, 40] present an approach that enables estimation of the updated failure probability without resorting to posterior samples. In this procedure, termed BUS (Bayesian updating with structural reliability methods), the integrals appearing in the definition of $P_{F|\mathbf{d}}$ in Eq. (4) are converted into equivalent reliability integrals by means of appropriate transformations. These integrals can then be evaluated with any sampling-based reliability estimation method, such as importance sampling [39] or other advanced Monte Carlo techniques [33, 40]

In this contribution, we introduce a novel simulation-based method to update the reliability of engineering systems using data. The proposed procedure uses an importance sampling (IS) method that is developed based on the principle of cross entropy (CE) minimization. The key contribution is a novel IS density associated with the posterior PDF of the uncertain parameters, which facilitates efficient sampling of the important region of the failure domain, particularly for a small posterior probability of failure. The IS density is designed via a two-step procedure. The first step involves construction of a sample-based approximation of the posterior PDF, which we build using the CE method. Here the aim is to determine the parameters of a chosen parametric distribution family that minimize its Kullback-Leibler divergence from $p_{\boldsymbol{\Theta}|\mathbf{d}}(\boldsymbol{\theta})$. The approximation leads to an efficient IS density for estimating the marginal likelihood. In the second step of the proposed method, we use the approximation of $p_{\boldsymbol{\Theta}|\mathbf{d}}(\boldsymbol{\theta})$ as a building block to construct the desired IS density for sampling the failure domain, through a second round of CE minimization. An adaptive, multi-level approach is employed to solve the CE optimization problem in each step. The IS densities deduced in the two steps are then applied to construct an efficient estimator for the posterior failure probability.

2. Importance sampling approach for reliability updating

The posterior probability of failure is defined in terms of the marginal likelihood, c_E, and the likelihood function, $L(\boldsymbol{\theta}|\mathbf{d})$, as

$$P_{F|\mathbf{d}} = \frac{1}{c_E} \int_{\boldsymbol{\theta} \in \mathbb{R}^{n_{\boldsymbol{\theta}}}} \mathrm{I}_F(\boldsymbol{\theta}) L(\boldsymbol{\theta}|\mathbf{d}) p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}. \tag{5}$$

One can evaluate $P_{F|\mathbf{d}}$ by standard Monte Carlo simulation, via a rejection sampling scheme, wherein independent samples of the uncertain parameters $\boldsymbol{\Theta}$ generated from the prior PDF are used to estimate c_E and the posterior probability $P_{F|\mathbf{d}}$. When the posterior PDF of $\boldsymbol{\Theta}$ differs significantly from the prior, or the failure event under the posterior probability measure is a rare event, this method requires a large number of samples to yield accurate estimates. A classical approach to address this drawback is to apply importance sampling.

Design of an efficient importance sampling scheme to evaluate the posterior probability of failure requires two main ingredients: (i) an IS density to estimate the marginal likelihood, and (ii) an IS density to integrate the un-normalized posterior PDF over the failure domain. Let $q_{\boldsymbol{\Theta}}^{(1)}(\boldsymbol{\theta})$ and $q_{\boldsymbol{\Theta}}^{(2)}(\boldsymbol{\theta})$ denote these two IS densities, respectively. Accordingly, Eq. (4) is written in the modified form

$$P_{F|\mathbf{d}} = \frac{1}{c_E} \int_{\boldsymbol{\theta} \in \mathbb{R}^{n_{\boldsymbol{\theta}}}} \mathrm{I}_F(\boldsymbol{\theta}) W_2(\boldsymbol{\theta}) q_{\boldsymbol{\Theta}}^{(2)}(\boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{6}$$

with

$$c_E = \int_{\boldsymbol{\theta} \in \mathbb{R}^{n_{\boldsymbol{\theta}}}} L(\boldsymbol{\theta}|\mathbf{d}) W_1(\boldsymbol{\theta}) q_{\boldsymbol{\Theta}}^{(1)}(\boldsymbol{\theta}) d\boldsymbol{\theta}. \tag{7}$$

In the preceding equations, $W_1(\boldsymbol{\theta}) = \frac{p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}{q_{\boldsymbol{\Theta}}^{(1)}(\boldsymbol{\theta})}$ and $W_2(\boldsymbol{\theta}) = \frac{L(\boldsymbol{\theta}|\mathbf{d}) p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}{q_{\boldsymbol{\Theta}}^{(2)}(\boldsymbol{\theta})}$ are the importance weight functions.

We develop an adaptive sampling strategy to determine the IS densities $q_{\boldsymbol{\Theta}}^{(1)}(\boldsymbol{\theta})$ and $q_{\boldsymbol{\Theta}}^{(2)}(\boldsymbol{\theta})$. The method is built on the principle of cross entropy (CE) minimization [36], a classical approach for constructing near-optimal IS densities for Monte Carlo integration. In the subsequent sections, we put forward a novel procedure to adapt this principle for the reliability updating problem. The proposed method is comprised of two steps. In the first step, described in Section 3, we determine the IS density $q_{\boldsymbol{\Theta}}^{(1)}(\boldsymbol{\theta})$. We construct $q_{\boldsymbol{\Theta}}^{(1)}(\boldsymbol{\theta})$ as a near-optimal approximation of the posterior PDF, $p_{\boldsymbol{\Theta}|\mathbf{d}}(\boldsymbol{\theta})$. We adopt the approach developed in our recent work [15], where we approximate $p_{\boldsymbol{\Theta}|\mathbf{d}}(\boldsymbol{\theta})$ by a parametric density that minimizes the cross entropy (CE) loss between $p_{\boldsymbol{\Theta}|\mathbf{d}}(\boldsymbol{\theta})$ and a chosen family of parametric distributions. In the next step, we construct $q_{\boldsymbol{\Theta}}^{(2)}(\boldsymbol{\theta})$ as an approximation of the optimal IS density for integrating the un-normalized posterior PDF over the failure domain. The procedure requires a second round of CE minimization. The approach, developed in Section 4, leverages upon the approximation of $p_{\boldsymbol{\Theta}|\mathbf{d}}(\boldsymbol{\theta})$ of the first step, and a smooth approximation of the indicator of the failure event, used earlier by [34, 35], to efficiently solve the CE optimization problem. We provide the proposed IS estimator of the posterior failure probability and discuss its statistical properties. In Section 5, we discuss the choice of the parametric density in the CE method, which is followed by numerical investigations in Section 6 that demonstrate the performance of our method.

3. Approximation of the posterior PDF

As already noted, it is straightforward to evaluate the marginal likelihood, c_E, by standard Monte Carlo simulation (MCS): one generates independent samples $\left\{\boldsymbol{\theta}^{(i)}, i = 1, \ldots, N_1\right\}$ from the prior PDF $p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ and computes the sample mean of the likelihood function values $\left\{L\left(\boldsymbol{\theta}^{(i)}|\mathbf{d}\right), i = 1, \ldots, N_1\right\}$. However, if the data is highly informative, the posterior PDF tends to differ significantly from the prior PDF, necessitating a very large number of samples N_1 to obtain an accurate estimate, i.e., an estimate with a small coefficient of variation (CoV). Importance sampling provides a path to overcome the drawback of standard MCS. The IS density should be selected such that the IS estimator has a smaller coefficient of variation (CoV) compared to the estimator in standard Monte Carlo. Following (7), one can show that if the posterior PDF is selected as the IS density, i.e., if $q_{\boldsymbol{\Theta}}^{(1)}(\boldsymbol{\theta}) = p_{\boldsymbol{\Theta}|\mathbf{d}}(\boldsymbol{\theta})$, the CoV of the IS estimator of the marginal likelihood reduces to zero. In the context of importance sampling, such a density is termed the optimal IS density [36]. The optimal IS density $p_{\boldsymbol{\Theta}|\mathbf{d}}(\boldsymbol{\theta})$ requires knowledge of the target quantity c_E, and hence cannot be directly applied. However, it is possible to construct an IS density that closely resembles $p_{\boldsymbol{\Theta}|\mathbf{d}}(\boldsymbol{\theta})$, and subsequently apply it to estimate c_E. In [15], we construct an approximation of the posterior PDF by fitting parametric density models using the CE method. The approach is summarized in the following.

3.1. Multi-level cross entropy method for posterior approximation

Consider a family of parametric densities $q_{\mathbf{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu})$ defined by the parameter vector $\boldsymbol{\nu} \in \mathcal{V}$. We select $q_{\mathbf{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu})$ such that it contains the prior PDF of the uncertain parameters, i.e., $q_{\mathbf{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}_0) = p_{\mathbf{\Theta}}(\boldsymbol{\theta})$ for $\hat{\boldsymbol{\nu}}_0 \in \mathcal{V}$. The choice of the family $q_{\mathbf{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu})$ is detailed in Section 5. The CE method aims at constructing a near-optimal IS density by minimizing the Kullback-Leibler (KL) divergence between the optimal IS density and the chosen parametric family [36]. The KL divergence between $p_{\mathbf{\Theta}|\mathbf{d}}(\boldsymbol{\theta})$ and $q_{\mathbf{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu})$ is a measure of distance between the two PDFs and is defined as

$$\begin{aligned} D_{KL}\left(p_{\mathbf{\Theta}|\mathbf{d}}(\boldsymbol{\theta})||q_{\mathbf{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu})\right) &= \mathrm{E}_{p_{\mathbf{\Theta}|\mathbf{d}}}\left[\ln\left(\frac{p_{\mathbf{\Theta}|\mathbf{d}}(\boldsymbol{\theta})}{q_{\mathbf{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu})}\right)\right] \\ &= \frac{1}{c_E}\mathrm{E}_{p_{\mathbf{\Theta}}}\left[L(\boldsymbol{\theta}|\mathbf{d})\ln\left(p_{\mathbf{\Theta}|\mathbf{d}}(\boldsymbol{\theta})\right)\right] - \frac{1}{c_E}\mathrm{E}_{p_{\mathbf{\Theta}}}\left[L(\boldsymbol{\theta}|\mathbf{d})\ln\left(q_{\mathbf{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu})\right)\right] \end{aligned} \tag{8}$$

Since the first expectation on the right-hand side of Eq. (8) is not a function of $\boldsymbol{\nu}$, minimizing $D_{KL}\left(p_{\mathbf{\Theta}|\mathbf{d}}(\boldsymbol{\theta})||q_{\mathbf{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu})\right)$ is equivalent to solving the stochastic optimization problem:

$$\boldsymbol{\nu}^*_{c_E} = \underset{\boldsymbol{a}\in\mathcal{V}}{\mathrm{argmax}}\ \mathrm{E}_{p_{\mathbf{\Theta}}}\left[L(\boldsymbol{\theta}|\mathbf{d})\ln\left(q_{\mathbf{\Theta}}(\boldsymbol{\theta};\boldsymbol{a})\right)\right] \tag{9}$$

The parametric density defined by $\boldsymbol{\nu}^*_{c_E}$ is a near-optimal approximation of the posterior PDF $p_{\mathbf{\Theta}|\mathbf{d}}(\boldsymbol{\theta})$. The above optimization can be solved by approximating the expectation in Eq. (9) with a set of samples drawn from $p_{\mathbf{\Theta}}(\boldsymbol{\theta})$. However, the number of samples required to obtain a good sample approximation is large when $p_{\mathbf{\Theta}|\mathbf{d}}(\boldsymbol{\theta})$ differs significantly from $p_{\mathbf{\Theta}}(\boldsymbol{\theta})$. In such cases, directly solving Eq. (9) is computationally challenging. To address this challenge, in [15] we develop a multi-level version of the CE method that approaches the target density $p_{\mathbf{\Theta}|\mathbf{d}}(\boldsymbol{\theta})$ step-wise through a sequence of parametric densities defined by $\{\boldsymbol{\nu}_k, k = 1,\ldots,L_1\}$.

We consider a sequence of intermediate target densities $\{h_1^k(\boldsymbol{\theta}), k = 0,\ldots,L_1\}$ that starts from the prior PDF $p_{\mathbf{\Theta}}(\boldsymbol{\theta})$ and gradually approaches the posterior PDF $p_{\mathbf{\Theta}|\mathbf{d}}(\boldsymbol{\theta})$. The distribution sequence is constructed by tempering the likelihood function (i.e., by taking it to be the power of γ_k):

$$h_1^k(\boldsymbol{\theta}) = \frac{1}{C_k}L(\boldsymbol{\theta}|\mathbf{d})^{\gamma_k}p_{\mathbf{\Theta}}(\boldsymbol{\theta}). \tag{10}$$

Here C_k is the normalizing constant of $h_1^k(\boldsymbol{\theta})$ and $0 = \gamma_0 < \gamma_1 < \cdots < \gamma_{L_1} = 1$ are tempering parameters which ensure a smooth transition between the prior and posterior PDFs of $\mathbf{\Theta}$. Note that $h_1^0(\boldsymbol{\theta}) = p_{\mathbf{\Theta}}(\boldsymbol{\theta})$ and $h_1^{L_1}(\boldsymbol{\theta}) = p_{\mathbf{\Theta}|\mathbf{d}}(\boldsymbol{\theta})$. This distribution sequence has been used in [31, 10, 8] to design Markov chain-based sequential Monte Carlo samplers for Bayesian analysis. We approach the posterior PDF $p_{\mathbf{\Theta}|\mathbf{d}}(\boldsymbol{\theta})$ gradually, by solving the CE optimization problem sequentially for each intermediate target density. This leads to a sequence of parameter vectors $\{\boldsymbol{\nu}_k, k = 1,\ldots,L_1\}$ such that the final parameter $\boldsymbol{\nu}_{L_1}$ is a good approximation of the optimal parameter $\boldsymbol{\nu}^*_{c_E}$. We determine $\boldsymbol{\nu}_k$ by minimizing the KL divergence between $h_1^k(\boldsymbol{\theta})$ and $q_{\mathbf{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu})$:

$$
\begin{aligned}
\boldsymbol{\nu}_k &= \underset{\boldsymbol{a}\in\mathcal{V}}{\operatorname{argmin}} D_{KL}\left(h_1^k(\boldsymbol{\theta})||q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu})\right) \\
&= \underset{\boldsymbol{a}\in\mathcal{V}}{\operatorname{argmax}}\ \mathrm{E}_{p_{\boldsymbol{\Theta}}}\left[L(\boldsymbol{\theta}|\mathbf{d})^{\gamma_k}\ln\left(q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\boldsymbol{a})\right)\right]
\end{aligned} \tag{11}
$$

The objective function of the corresponding optimization problem, i.e., the expectation $\mathrm{E}_{p_{\boldsymbol{\Theta}}}\left[L(\boldsymbol{\theta}|\mathbf{d})^{\gamma_k}\ln\left(q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\boldsymbol{a})\right)\right]$, is approximated by importance sampling using a set of samples $\{\boldsymbol{\theta}^{(i)}, i=1,\ldots,N\}$ distributed according to $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}_{k-1})$, where $\hat{\boldsymbol{\nu}}_{k-1}$ is the estimate of $\boldsymbol{\nu}_{k-1}$ determined in the previous level. At $k=0$, the sampling density $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}_0)$ corresponds to the prior PDF $p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$. This leads to the following stochastic optimization problem to be solved in each intermediate level :

$$
\hat{\boldsymbol{\nu}}_k = \underset{\boldsymbol{a}\in\mathcal{V}}{\operatorname{argmax}}\ \frac{1}{N}\sum_{i=1}^{N}\widetilde{W}_k^1\left(\boldsymbol{\theta}^{(i)},\hat{\boldsymbol{\nu}}_{k-1}\right)\ln\left(q_{\boldsymbol{\Theta}}\left(\boldsymbol{\theta}^{(i)};\boldsymbol{a}\right)\right), \tag{12}
$$

where $\widetilde{W}_k^1\left(\boldsymbol{\theta}^{(i)},\hat{\boldsymbol{\nu}}_{k-1}\right) = L\left(\boldsymbol{\theta}^{(i)}|\mathbf{d}\right)^{\gamma_k}\frac{p_{\boldsymbol{\Theta}}\left(\boldsymbol{\theta}^{(i)}\right)}{q_{\boldsymbol{\Theta}}\left(\boldsymbol{\theta}^{(i)};\hat{\boldsymbol{\nu}}_{k-1}\right)}$ is the importance weight of a sample $\boldsymbol{\theta}^{(i)}$.

The accuracy and computational cost of this procedure depends on the choice of the tempering parameters $\{\gamma_k, k=1,\ldots,L_1\}$, which determine the change between the respective target densities. In order to obtain a good estimate of $\hat{\boldsymbol{\nu}}_k$ with a limited number of samples, the intermediate PDF $h_1^k(\boldsymbol{\theta})$ should not differ largely from the parametric density $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}_{k-1})$. To ensure this, we adopt the criterion suggested in [34] and select the tempering parameter γ_k adaptively, on the fly, such that the sample CoV $\hat{\delta}_{\widetilde{W}_k^1}$ of the weights $\left\{\widetilde{W}_k^1\left(\boldsymbol{\theta}^{(i)}\right), i=1,\ldots,N\right\}$ adheres to a target value $\delta_\gamma^* = 1.5$:

$$
\gamma_k = \underset{\gamma\in(\gamma_{k-1},1)}{\operatorname{argmin}}\ \left(\hat{\delta}_{\widetilde{W}_k^1}(\gamma)-\delta_\gamma^*\right)^2. \tag{13}
$$

The adaptive procedure terminates when the value of γ_k determined based on Eq. (13) equals 1. After termination, the final parameter vector $\hat{\boldsymbol{\nu}}_{L_1}$ is determined by solving Eq. (12) with $\gamma_{L_1}=1$. $\hat{\boldsymbol{\nu}}_{L_1}$ closely approximates the optimal parameter $\boldsymbol{\nu}_{c_E}^*$ in Eq. (9). Thus, $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}_{L_1})$ is a close approximation of the posterior PDF for the chosen parametric family, and is taken as the IS density for estimating the marginal likelihood, c_E.

4. Estimation of the posterior probability of failure

The parametric density $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}_{L_1})$ describing the posterior PDF could be applied to estimate the posterior probability of failure by importance sampling. However, if the failure event, F, is rare under the posterior probability measure, the samples from $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}_{L_1})$ do not represent well the failure domain, resulting in a high sampling CoV of the associated IS estimator. In contrast, the optimal IS density that perfectly describes F, and leads to an IS estimator with sampling CoV equal to zero, is given by

$$q^*_{P_{F|\mathbf{d}}}(\boldsymbol{\theta}) = \frac{1}{P_{F|\mathbf{d}}} I_F(\boldsymbol{\theta}) p_{\boldsymbol{\Theta}|\mathbf{d}}(\boldsymbol{\theta}). \tag{14}$$

The above IS density, however, cannot be applied in practice, as it requires knowledge of the target probability of failure. We develop an extension of the multi-level CE method described in the previous section, to construct an IS density $q^{(2)}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ that is a close approximation of the optimal IS density $q^*_{P_{F|\mathbf{d}}}(\boldsymbol{\theta})$. The proposed IS density is able to adequately describe the rare failure region, and, together with $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \hat{\boldsymbol{\nu}}_{L_1})$, it leads to an efficient IS estimator for the posterior failure probability.

Consider the parametric density family $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \boldsymbol{\nu}); \boldsymbol{\nu} \in \mathcal{V}$ introduced in the previous section. An approximation of the optimal IS density $q^*_{P_{F|\mathbf{d}}}(\boldsymbol{\theta})$ is deduced by the CE method, by minimizing the KL divergence between $q^*_{P_{F|\mathbf{d}}}(\boldsymbol{\theta})$ and $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \boldsymbol{\nu})$, i.e., by solving the CE optimization problem

$$\begin{aligned} \boldsymbol{\nu}^*_{P_{F|\mathbf{d}}} &= \underset{\boldsymbol{a} \in \mathcal{V}}{\operatorname{argmin}}\ D_{KL}\left(q^*_{P_{F|\mathbf{d}}}(\boldsymbol{\theta}) || q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \boldsymbol{a})\right) \\ &= \underset{\boldsymbol{a} \in \mathcal{V}}{\operatorname{argmax}}\ \mathrm{E}_{p_{\boldsymbol{\Theta}|\mathbf{d}}}\left[I_F(\boldsymbol{\theta}) \ln\left(q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \boldsymbol{a})\right)\right]. \end{aligned} \tag{15}$$

One can solve the above optimization in a single step, after approximating the expectation through importance sampling using samples from the parametric density $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \hat{\boldsymbol{\nu}}_{L_1})$ describing the posterior PDF, $p_{\boldsymbol{\Theta}|\mathbf{d}}(\boldsymbol{\theta})$. This approach, however, requires a large number of samples when the failure event is rare. By analogy with Section 3, we lay out a multi-level procedure that approaches the target density $q^*_{P_{F|\mathbf{d}}}(\boldsymbol{\theta})$ step-wise, by approximating a sequence of target densities $\{h^k_2(\boldsymbol{\theta}), k = 1, \ldots, L_2\}$ residing between the posterior PDF $p_{\boldsymbol{\Theta}|\mathbf{d}}(\boldsymbol{\theta})$ and $q^*_{P_{F|\mathbf{d}}}(\boldsymbol{\theta})$. The approach, described in Section 4.1, results in an extended sequence of parameter vectors $\{\hat{\boldsymbol{\nu}}_{k+L_1}, k = 1, \ldots, L_2\}$ where the final parameter $\hat{\boldsymbol{\nu}}_{L_2+L_1}$ is a good approximation of $\boldsymbol{\nu}^*_{P_{F|\mathbf{d}}}$. The parametric density $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \hat{\boldsymbol{\nu}}_{L_2+L_1})$, which is close to the optimal IS density $q^*_{P_{F|\mathbf{d}}}(\boldsymbol{\theta})$, is then applied to estimate the posterior probability of failure by importance sampling.

4.1. Multi-level CE method for estimation of the posterior failure probability

In the standard multi-level CE method for rare event estimation [36], the intermediate target densities correspond to the optimal IS densities of intermediate reliability integrals, defined by a sequence of failure events that gradually approach the rare failure event F. To enable better use of the samples generated at each level, Papaioannou et al. [35] proposed to characterize the intermediate densities using a smooth approximation of $\mathrm{I}_F(\boldsymbol{\theta})$ based on the standard normal cumulative distribution function $\Phi(\cdot)$. We follow the distribution sequence suggested in [35] and define the intermediate target densities for estimating the posterior failure probability as

$$h^k_2(\boldsymbol{\theta}) = \frac{1}{P_k} \Phi\left(-\frac{g(\boldsymbol{\theta})}{\sigma_k}\right) p_{\boldsymbol{\Theta}|\mathbf{d}}(\boldsymbol{\theta}), \tag{16}$$

where $\sigma_1 > \sigma_2 > \cdots > \sigma_{L_2} > 0$ are smoothing parameters and P_k is the normalizing constant of $h_2^k(\boldsymbol{\theta})$. Note that $\lim_{\sigma\to 0}\Phi\left(-\frac{g(\boldsymbol{\theta})}{\sigma}\right) = \mathrm{I}\{g(\boldsymbol{\theta})\leq 0\}$. Hence, with decreasing σ, the above sequence converges to the optimal IS density $q^*_{P_{F|\mathbf{d}}}(\boldsymbol{\theta})$ in Eq. (14).

Starting from the parametric density $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}_{L_1})$ approximating the posterior PDF $p_{\boldsymbol{\Theta}|\mathbf{d}}(\boldsymbol{\theta})$, we construct a sequence of densities $\{q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu}_{k+L_1}), k=1,\ldots,L_2\}$ such that $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu}_{k+L_1})$ has the minimum KL divergence from $h_2^k(\boldsymbol{\theta})$ within the parametric family. The parameter vector $\hat{\boldsymbol{\nu}}_{k+L_1}$ is determined by solving the sample counter-part of the CE optimization:

$$\boldsymbol{\nu}_{k+L_1} = \underset{\boldsymbol{a}\in\mathcal{V}}{\operatorname{argmax}}\ \mathrm{E}_{p_{\boldsymbol{\Theta}|\mathbf{d}}}\left[\Phi\left(-\frac{g(\boldsymbol{\theta})}{\sigma_k}\right)\ln\left(q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\boldsymbol{a})\right)\right]. \tag{17}$$

We approximate the expectation in Eq. (17) through importance sampling using samples $\left\{\boldsymbol{\theta}^{(i)}, i=1,\ldots,N\right\}$ generated from $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}_{k+L_1-1})$, to arrive at the following optimization problem:

$$\hat{\boldsymbol{\nu}}_{k+L_1} = \underset{\boldsymbol{a}\in\mathcal{V}}{\operatorname{argmax}}\ \sum_{i=1}^{N}\widetilde{W}_k^2\left(\boldsymbol{\theta}^{(i)},\hat{\boldsymbol{\nu}}_{k+L_1-1}\right)\ln\left(q_{\boldsymbol{\Theta}}\left(\boldsymbol{\theta}^{(i)};\boldsymbol{a}\right)\right), \tag{18}$$

with $\widetilde{W}_k^2(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}_{k+L_1-1}) = \Phi\left(-\frac{g(\boldsymbol{\theta})}{\sigma_k}\right)\frac{L(\boldsymbol{\theta}|\mathbf{d})p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}{q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}_{k+L_1-1})}$. To ensure that a good estimate of $\boldsymbol{\nu}_{k+L_1}$ is obtained with a reasonable number of samples drawn from $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}_{k+L_1-1})$, in each level the smoothing parameter is selected such that the sample CoV $\hat{\delta}_{\widetilde{W}_k^2}$ of the weights $\left\{\widetilde{W}_k^2\left(\boldsymbol{\theta}^{(i)},\hat{\boldsymbol{\nu}}_{k+L_1-1}\right), i=1,\right.$ adheres to a target value δ^*_σ:

$$\sigma_k = \underset{\sigma\in(0,\sigma_{k-1})}{\operatorname{argmin}}\ \left(\hat{\delta}_{\widetilde{W}_k^2}(\sigma)-\delta^*_\sigma\right)^2. \tag{19}$$

We select $\delta^*_\sigma = 1.5$ [34]. The adaptive procedure terminates when the CoV of the weights of the current smooth approximation with respect to the optimal IS density $\left\{\frac{\mathrm{I}\left\{g\left(\boldsymbol{\theta}^{(i)}\right)\leq 0\right\}}{\Phi\left(-\frac{g\left(\boldsymbol{\theta}^{(i)}\right)}{\sigma_k}\right)}, i=1,\ldots,N\right\}$ is smaller than δ^*_σ.

4.2. Estimator for the posterior probability of failure

The fitted IS density $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}_{L_2+L_1})$ is applied to evaluate the posterior probability of failure by importance sampling. Accordingly, we write Eq. (4) in the modified form

$$\begin{aligned}P_{F|\mathbf{d}} &= \int_{\boldsymbol{\theta}\in\mathbb{R}^{n_{\boldsymbol{\theta}}}}\mathrm{I}_F(\boldsymbol{\theta})\frac{p_{\boldsymbol{\Theta}|\mathbf{d}}(\boldsymbol{\theta})}{q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}_{L_2+L_1})}q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}_{L_2+L_1})\mathrm{d}\boldsymbol{\theta}\\ &= \frac{1}{c_E}\int_{\boldsymbol{\theta}\in\mathbb{R}^{n_{\boldsymbol{\theta}}}}\mathrm{I}_F(\boldsymbol{\theta})\frac{L(\boldsymbol{\theta}|\mathbf{d})p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}{q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}_{L_2+L_1})}q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}_{L_2+L_1})\mathrm{d}\boldsymbol{\theta}.\end{aligned} \tag{20}$$

The marginal likelihood, c_E, is evaluated by importance sampling using the IS density

$q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}_{L_1})$. This leads to

$$P_{F|\mathbf{d}} = \frac{\int_{\boldsymbol{\theta}\in\mathbb{R}^{n_{\boldsymbol{\theta}}}} \mathrm{I}_F(\boldsymbol{\theta}) W_2(\boldsymbol{\theta}) q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}_{L_2+L_1})\mathrm{d}\boldsymbol{\theta}}{\int_{\boldsymbol{\theta}\in\mathbb{R}^{n_{\boldsymbol{\theta}}}} L(\boldsymbol{\theta}|\mathbf{d}) W_1(\boldsymbol{\theta}) q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}_{L_1})\mathrm{d}\boldsymbol{\theta}}, \tag{21}$$

where $W_1(\boldsymbol{\theta}) = \frac{p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}{q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}_{L_1})}$ and $W_2(\boldsymbol{\theta}) = \frac{L(\boldsymbol{\theta}|\mathbf{d})p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}{q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}_{L_2+L_1})}$ are IS weights. The corresponding estimator is

$$\hat{P}^{\mathrm{IS}}_{F|\mathbf{d}} = \frac{\frac{1}{N_{\mathrm{IS},2}}\sum_{i=1}^{N_{\mathrm{IS},2}} \mathrm{I}_F\left(\boldsymbol{\theta}^{(2,i)}\right) W_2\left(\boldsymbol{\theta}^{(2,i)}\right)}{\frac{1}{N_{\mathrm{IS},1}}\sum_{i=1}^{N_{\mathrm{IS},1}} L\left(\boldsymbol{\theta}^{(1,i)}|\mathbf{d}\right) W_1\left(\boldsymbol{\theta}^{(1,i)}\right)}, \tag{22}$$

where $\left\{\boldsymbol{\theta}^{(1,i)}, i=1,\dots,N_{\mathrm{IS},1}\right\}$ and $\left\{\boldsymbol{\theta}^{(2,i)}, i=1,\dots,N_{\mathrm{IS},2}\right\}$ are independent samples generated from the IS densities $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}_{L_1})$ and $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}_{L_2+L_1})$, respectively.

4.3. Statistics of the proposed estimator

The bias and CoV of the estimator of the posterior failure probability are given by the following two propositions. We denote the estimators in the denominator and numerator of Eq. (22) by $\hat{P}_1$ and $\hat{P}_2$, respectively. Let P_1 and P_2, respectively, denote the true values of $\hat{P}_1$ and $\hat{P}_2$, i.e., $P_{F|\mathbf{d}} = \frac{P_2}{P_1}$. For simplicity we set $N_{\mathrm{IS},1} = N_{\mathrm{IS},2} = N$.

Proposition 1. $\hat{P}^{\mathrm{IS}}_{F|\mathbf{d}}$ is biased for finite N. The fractional bias is given by:

$$\mathrm{E}\left[\frac{\hat{P}^{\mathrm{IS}}_{F|\mathbf{d}} - P_{F|\mathbf{d}}}{P_{F|\mathbf{d}}}\right] = \delta_1^2 - \rho_{12}\delta_1\delta_2 + \mathrm{o}(1/N) = \mathrm{O}(1/N), \tag{23}$$

where δ_1 and δ_2, respectively, denote the CoV of $\hat{P}_1$ and $\hat{P}_2$, and ρ_{12} denotes the correlation coefficient between the estimators. $\hat{P}^{\mathrm{IS}}_{F|\mathbf{d}}$ is thus asypmtotically unbiased and the bias is $\mathrm{O}(1/N)$.

Proof. Since $\hat{P}^{\mathrm{IS}}_{F|\mathbf{d}} = \frac{\hat{P}_2}{\hat{P}_1}$ and $P_{F|\mathbf{d}} = \frac{P_2}{P_1}$, it follows that

$$\frac{\hat{P}^{\mathrm{IS}}_{F|\mathbf{d}} - P_{F|\mathbf{d}}}{P_{F|\mathbf{d}}} = \left(\frac{\hat{P}_2}{P_2} - \frac{\hat{P}_1}{P_1}\right)\frac{P_1}{\hat{P}_1}. \tag{24}$$

Taylor series expansion of $\frac{P_1}{\hat{P}_1}$ around P_1 leads to

$$\begin{aligned}\frac{\hat{P}^{\mathrm{IS}}_{F|\mathbf{d}} - P_{F|\mathbf{d}}}{P_{F|\mathbf{d}}} &= \left(\frac{\hat{P}_2}{P_2} - \frac{\hat{P}_1}{P_1}\right)\left(1 - \frac{\hat{P}_1 - P_1}{P_1} + \left(\frac{\hat{P}_1 - P_1}{P_1}\right)^2 + \cdots\right) \\ &= \frac{\hat{P}_2}{P_2} - \frac{\hat{P}_1}{P_1} + \frac{\hat{P}_1\left(\hat{P}_1 - P_1\right)}{P_1^2} - \frac{\hat{P}_2\left(\hat{P}_1 - P_1\right)}{P_1 P_2} + \cdots\end{aligned} \tag{25}$$

Taking expectation on both sides of Eq. (25) and noting that the estimators $\hat{P}_1$ and $\hat{P}_2$ are unbiased, i.e., $\mathrm{E}[\hat{P}_1] = P_1$ and $\mathrm{E}[\hat{P}_2] = P_2$, proves the required proposition. □

Proposition 2. The CoV $\delta_{\hat{P}^{\mathrm{IS}}_{F|\mathbf{d}}}$ of $\hat{P}^{\mathrm{IS}}_{F|\mathbf{d}}$ is given by:

$$\delta^2_{\hat{P}^{\mathrm{IS}}_{F|\mathbf{d}}} = \mathrm{E}\left[\frac{\hat{P}^{\mathrm{IS}}_{F|\mathbf{d}} - P_{F|\mathbf{d}}}{P_{F|\mathbf{d}}}\right]^2 = \delta_1^2 + \delta_2^2 - \rho_{12}\delta_1\delta_2 + \mathrm{o}(1/N) = \mathrm{O}(1/N) \tag{26}$$

where δ_1 and δ_2, respectively, denote the CoV of $\hat{P}_1$ and $\hat{P}_2$, and ρ_{12} denotes the correlation coefficient between the estimators. $\hat{P}^{\mathrm{IS}}_{F|\mathbf{d}}$ is thus a consistent estimator and its CoV $\delta_{\hat{P}^{\mathrm{IS}}_{F|\mathbf{d}}}$ is $\mathrm{O}(1/\sqrt{N})$.

Proof. From Eq. (25)

$$\begin{aligned}
\mathrm{E}\left[\frac{\hat{P}^{\mathrm{IS}}_{F|\mathbf{d}} - P_{F|\mathbf{d}}}{P_{F|\mathbf{d}}}\right]^2 &= \mathrm{E}\left[\frac{\hat{P}_2}{P_2} - \frac{\hat{P}_1}{P_1} + \frac{\hat{P}_1\left(\hat{P}_1 - P_1\right)}{P_1^2} - \frac{\hat{P}_2\left(\hat{P}_1 - P_1\right)}{P_1P_2} + \cdots\right]^2 \\
&= \mathrm{E}\left[\frac{\hat{P}_2 - P_2}{P_2} - \frac{\hat{P}_1 - P_1}{P_1} + \frac{\hat{P}_1\left(\hat{P}_1 - P_1\right)}{P_1^2} - \frac{\hat{P}_2\left(\hat{P}_1 - P_1\right)}{P_1P_2} + \cdots\right]^2 \\
&= \mathrm{E}\left[\frac{\hat{P}_1 - P_1}{P_1}\right]^2 + \mathrm{E}\left[\frac{\hat{P}_2 - P_2}{P_2}\right]^2 - \mathrm{E}\left[\left(\frac{\hat{P}_1 - P_1}{P_1}\right)\left(\frac{\hat{P}_2 - P_2}{P_2}\right)\right] + \mathrm{o}\left(\frac{1}{N}\right)
\end{aligned} \tag{27}$$

Hence the proposition. □

In practice, it is reasonable to assume that the estimators $\hat{P}_1$ and $\hat{P}_2$ are uncorrelated. Then, we can use the first two terms on the R.H.S of Eq. (26) to obtain an approximate estimate of the CoV of the probability of failure estimator:

$$\hat{\delta}^2_{\hat{P}^{\mathrm{IS}}_{F|\mathbf{d}}} \approx \hat{\delta}_1^2 + \hat{\delta}_2^2, \tag{28}$$

where $\hat{\delta}_{\hat{P}^{\mathrm{IS}}_{F|\mathbf{d}}}$, $\hat{\delta}_1$ and $\hat{\delta}_2$ denote sample estimates of $\delta_{\hat{P}^{\mathrm{IS}}_{F|\mathbf{d}}}$, δ_1 and δ_2, respectively. The estimates of δ_1 and δ_2 are obtained according to [28]

$$\hat{\delta}_1^2 = \frac{1}{\hat{P}_1^2}\frac{1}{N_{\mathrm{IS},1} - 1}\left[\frac{1}{N_{\mathrm{IS},1}}\sum_{i=1}^{N_{\mathrm{IS},1}}\left\{L\left(\boldsymbol{\theta}^{(1,i)}|\mathbf{d}\right)W_1\left(\boldsymbol{\theta}^{(1,i)}\right)\right\}^2 - \hat{P}_1^2\right] \tag{29}$$

and

$$\hat{\delta}_2^2 = \frac{1}{\hat{P}_2^2}\frac{1}{N_{\mathrm{IS},2}-1}\left[\frac{1}{N_{\mathrm{IS},2}}\sum_{i=1}^{N_{\mathrm{IS},2}} I_F\left(\boldsymbol{\theta}^{(2,i)}\right)\left\{W_2\left(\boldsymbol{\theta}^{(2,i)}\right)\right\}^2 - \hat{P}_2^2\right]. \tag{30}$$

To investigate the influence of the number of samples $N_{\mathrm{IS},1}$ and $N_{\mathrm{IS},2}$ on the CoV of $\hat{P}^{\mathrm{IS}}_{F|\mathbf{d}}$, we consider two cases. In the first case, we take $N_{\mathrm{IS},1}$ and $N_{\mathrm{IS},2}$ equal to the number of samples employed per level in the multi-level CE method, i.e., $N_{\mathrm{IS},1} = N_{\mathrm{IS},2} = N$. In the second case, we select the number of samples to ensure that $\hat{\delta}_{\hat{P}^{\mathrm{IS}}_{F|\mathbf{d}}}$ adheres to a target value δ^*. For this, we vary $N_{\mathrm{IS},1}$ and $N_{\mathrm{IS},2}$ adaptively such that the sample estimates $\hat{\delta}_1$ and $\hat{\delta}_2$ are, respectively, equal to target values δ_1^* and δ_2^* with $\delta_1^* + \delta_2^* = \delta^*$. A choice of $\delta_1^* = \delta_2^* = \delta^*/\sqrt{2}$ is employed in the present study which ensures that $\delta_{\hat{P}^{\mathrm{IS}}_{F|\mathbf{d}}} \lesssim \delta^*$.

4.4. Separation of uncertainty

In many problems, the data contains information on only a sub-group of the random variables appearing in the limit state function. For example, $\boldsymbol{\Theta}$ can contain uncertain future forcing variables, which cannot be learned. Let $\boldsymbol{\Theta}_A$ denote the group of random variables in $\boldsymbol{\Theta}$ that cannot be learned and $\boldsymbol{\Theta}_B$ denote the remaining random variables. In principle, one can consider the likelihood function to be simply constant with respect to all random variables in $\boldsymbol{\Theta}_A$. The methods described in the preceding sections then remain applicable, and the posterior probability of failure can be estimated based on Eq. (22). However, in certain applications it is convenient to evaluate the probability of failure conditional on specific instances of $\boldsymbol{\Theta}_B$ separately, using analytical or simulation-based methods [16, 11, 14]. In such cases, it is advantageous to express the posterior probability of failure in an alternative form. Let $p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = p_{\boldsymbol{\Theta}_A|\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_A|\boldsymbol{\theta}_B)p_{\boldsymbol{\Theta}_B|\mathbf{d}}(\boldsymbol{\theta}_B)$ be the prior PDF of $\boldsymbol{\Theta}$. The posterior PDF of $\boldsymbol{\Theta}$ is then given by $p_{\boldsymbol{\Theta}|\mathbf{d}}(\boldsymbol{\theta}) = p_{\boldsymbol{\Theta}_A|\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_A|\boldsymbol{\theta}_B)p_{\boldsymbol{\Theta}_B|\mathbf{d}}(\boldsymbol{\theta}_B)$, where $p_{\boldsymbol{\Theta}_B|\mathbf{d}}(\boldsymbol{\theta}_B)$ is the posterior PDF of $\boldsymbol{\Theta}_B$ defined in analogy to Eq. (2). One can write the posterior probability of failure in terms of $\boldsymbol{\Theta}_A$ and $\boldsymbol{\Theta}_B$ as

$$P_{F|\mathbf{d}} = \int_{\boldsymbol{\theta}_B \in \mathbb{R}^{n_{\boldsymbol{\theta}_B}}} P_{F|\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B)p_{\boldsymbol{\Theta}_B|\mathbf{d}}(\boldsymbol{\theta}_B)\mathrm{d}\boldsymbol{\theta}_B, \tag{31}$$

where the conditional failure probability $P_{F|\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B)$ is given by

$$P_{F|\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B) = \int_{\boldsymbol{\theta}_A \in \mathbb{R}^{n_{\boldsymbol{\theta}_A}}} \mathrm{I}_F(\boldsymbol{\theta}_A, \boldsymbol{\theta}_B)p_{\boldsymbol{\Theta}_A|\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_A|\boldsymbol{\theta}_B)\mathrm{d}\boldsymbol{\theta}_A. \tag{32}$$

In Eqs. (31) and (32), $n_{\boldsymbol{\theta}_A}$ and $n_{\boldsymbol{\theta}_B}$ denote the dimension of $\boldsymbol{\Theta}_A$ and $\boldsymbol{\Theta}_B$, respectively. The formulation in the above equations offers two advantages. Firstly, the CE optimization problem for updating model parameters $\boldsymbol{\Theta}_B$, that leads to the approximation of the posterior PDF $p_{\boldsymbol{\Theta}_B|\mathbf{d}}(\boldsymbol{\theta}_B)$, is now solved in a lower-dimensional space. This reduces the computational cost required for optimization and can help to address the degeneracy of the importance sampling weights in high dimensions. Secondly, it enables one to evaluate $P_{F|\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B)$ by tailor-made approaches specific to the application at hand, thereby reducing the uncertainty of the posterior failure probability estimator. The posterior probability of failure is then estimated by evaluating the expectation of $P_{F|\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B)$ with respect to $p_{\boldsymbol{\Theta}_B|\mathbf{d}}(\boldsymbol{\theta}_B)$. The parametric density $q_{\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B; \hat{\boldsymbol{\nu}}_{L_1})$ for approximating $p_{\boldsymbol{\Theta}_B|\mathbf{d}}(\boldsymbol{\theta}_B)$ and estimating the marginal likelihood can be constructed by the CE method through the procedure described in Section 3. However, the challenge lies in constructing an IS density associated with $p_{\boldsymbol{\Theta}_B|\mathbf{d}}(\boldsymbol{\theta}_B)$ to efficiently perform the reliability integration, i.e., the expectation of $P_{F|\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B)$, for which the procedure in Section 4.1 cannot be directly applied.

The optimal IS density for evaluating the integral in Eq. (31) is given by

$$q^*_{P_{F|\mathbf{d}}}(\boldsymbol{\theta}_B) = \frac{1}{P_{F|\mathbf{d}}} P_{F|\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B) p_{\boldsymbol{\Theta}_B|\mathbf{d}}(\boldsymbol{\theta}_B). \tag{33}$$

Note that this optimal density is different from the one in Eq. (14). Hence, to determine an approximation of the above density by the multi-level CE method, one needs to consider an alternative distribution sequence that defines a smooth transition from $p_{\boldsymbol{\Theta}_B|\mathbf{d}}(\boldsymbol{\theta}_B)$, or its approximation $q_{\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B; \hat{\boldsymbol{\nu}}_{L_1})$, to $q^*_{P_{F|\mathbf{d}}}(\boldsymbol{\theta}_B)$ in Eq. (33). Such a distribution sequence can be constructed by tempering the conditional probability function [24]:

$$h_2^k(\boldsymbol{\theta}_B) = \frac{1}{C_k} P_{F|\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B)^{\alpha_k} p_{\boldsymbol{\Theta}_B|\mathbf{d}}(\boldsymbol{\theta}_B), \tag{34}$$

where $\{\alpha_k, k = 0, \dots, L_2\}$ are the tempering parameters satisfying $0 = \alpha_0 < \alpha_1 < \dots < \alpha_{L_2} = 1$ and C_k is the normalizing constant of $h_2^k(\boldsymbol{\theta}_B)$. The parametric IS density $q_{\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B; \hat{\boldsymbol{\nu}}_{L_2+L_1})$ is determined by approximating the above distribution sequence in a step-wise manner. The associated CE optimization problems are solved sequentially, following similar steps as in Section 3. An IS estimator of the posterior probability of failure is obtained as

$$\hat{P}^{IS}_{F|\mathbf{d}} = \frac{\frac{1}{N_{\mathrm{IS},2}} \sum_{i=1}^{N_{\mathrm{IS},2}} \hat{P}_{F|\boldsymbol{\Theta}_B}\left(\boldsymbol{\theta}_B^{(2,i)}\right) W_{2,B}\left(\boldsymbol{\theta}_B^{(2,i)}\right)}{\frac{1}{N_{\mathrm{IS},1}} \sum_{i=1}^{N_{\mathrm{IS},1}} L\left(\boldsymbol{\theta}_B^{(1,i)}|\mathbf{d}\right) W_{1,B}\left(\boldsymbol{\theta}_B^{(1,i)}\right)}, \tag{35}$$

where $W_{1,B}(\boldsymbol{\theta}_B) = \frac{p_{\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B)}{q_{\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B; \hat{\boldsymbol{\nu}}_{L_1})}$ and $W_{2,B}(\boldsymbol{\theta}_B) = \frac{L(\boldsymbol{\theta}_B|\mathbf{d}) p_{\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B)}{q_{\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B; \hat{\boldsymbol{\nu}}_{L_2+L_1})}$ are IS weights, $\left\{\boldsymbol{\theta}_B^{(1,i)}, i = 1, \dots, N_{\mathrm{IS},1}\right\}$ and $\left\{\boldsymbol{\theta}_B^{(2,i)}, i = 1, \dots, N_{\mathrm{IS},2}\right\}$ are independent samples generated from the IS densities $q_{\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B; \hat{\boldsymbol{\nu}}_{L_1})$ and $q_{\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B; \hat{\boldsymbol{\nu}}_{L_2+L_1})$, respectively, and $\hat{P}_{F|\boldsymbol{\Theta}_B}\left(\boldsymbol{\theta}_B^{(2,i)}\right)$ is the estimate of the conditional probability for sample $\boldsymbol{\theta}_B^{(2,i)}$. In the following, we discuss the evaluation of $\hat{P}_{F|\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B)$ for a special case.

4.4.1. Updating the first-passage failure probability of uncertain linear systems

A prominent example where the above formulation is useful is the estimation of first-passage probability of systems subjected to random dynamic loads. In this context, F denotes the first-passage failure event, $\boldsymbol{\Theta}_A$ denotes the random variables characterizing the future random excitation and $\boldsymbol{\Theta}_B$ denotes uncertain system parameters. Typically, $\boldsymbol{\Theta}_A$ is high dimensional and is independent of $\boldsymbol{\Theta}_B$. It is well-known that applying standard importance sampling with parametric IS density in a high-dimensional random variable space can lead to poor estimates [3, 26]. This is related to the degeneracy of the IS weights in high dimensions. Hence, advanced Monte Carlo methods that are designed for high dimensions [37, 38] are commonly applied to estimate the first-passage probability. Evaluating the posterior probability of failure based on Eqs. (31) and (32) enables the integration of these methods into the framework of cross entropy-based Bayesian analysis to update first-passage probabilities of engineering systems under future excitation.

The conditional probability $P_{F|\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B)$ denotes the first-passage failure probability of the deterministic system corresponding to a specific outcome $\boldsymbol{\theta}_B$ of the system parameters. If the random excitation is a Gaussian process, $\boldsymbol{\Theta}_A$ is comprised of independent standard Gaussian random variables. The conditional first-passage probability $P_{F|\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B)$ can be estimated by importance sampling from the outcome space of $\boldsymbol{\Theta}_A$. In [2], an efficient IS density, $q_{\boldsymbol{\Theta}_A|\boldsymbol{\Theta}_B=\boldsymbol{\theta}_B}(\boldsymbol{\theta}_A)$, of $\boldsymbol{\Theta}_A$ is suggested for the particular case where the system is linear, which is defined by a weighted sum of Gaussian PDFs truncated on the failure domain of the deterministic system defined by $\boldsymbol{\theta}_B$. Accordingly, the conditional first-passage probability is expressed by the modified integral

$$P_{F|\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B) = \int_{\boldsymbol{\theta}_A \in \mathbb{R}^{n_{\boldsymbol{\theta}_A}}} \mathrm{I}_F(\boldsymbol{\theta}_A, \boldsymbol{\theta}_B) W_{2,A}(\boldsymbol{\theta}_A) q_{\boldsymbol{\Theta}_A|\boldsymbol{\Theta}_B=\boldsymbol{\theta}_B}(\boldsymbol{\theta}_A) \mathrm{d}\boldsymbol{\theta}_A, \tag{36}$$

where $W_{2,A}(\boldsymbol{\theta}_A) = \frac{p_{\boldsymbol{\Theta}_A}(\boldsymbol{\theta}_A)}{q_{\boldsymbol{\Theta}_A|\boldsymbol{\Theta}_B=\boldsymbol{\theta}_B}(\boldsymbol{\theta}_A)}$ is the IS weight. By employing a one-sample estimator of the above integral, the IS estimator for evaluating the posterior first-passage failure probability is given by:

$$\hat{P}^{IS}_{F|\mathbf{d}} = \frac{\frac{1}{N_{\mathrm{IS},2}} \sum_{i=1}^{N_{\mathrm{IS},2}} \mathrm{I}_F\left(\boldsymbol{\theta}_A^{(2,i)}, \boldsymbol{\theta}_B^{(2,i)}\right) W_{2,A}\left(\boldsymbol{\theta}_A^{(2,i)}\right) W_{2,B}\left(\boldsymbol{\theta}_B^{(2,i)}\right)}{\frac{1}{N_{\mathrm{IS},1}} \sum_{i=1}^{N_{\mathrm{IS},1}} L\left(\boldsymbol{\theta}_B^{(1,i)}|\mathbf{d}\right) W_{1,B}\left(\boldsymbol{\theta}_B^{(1,i)}\right)}, \tag{37}$$

where $\boldsymbol{\theta}_A^{(2,i)}$ denotes a sample of the random vector $\boldsymbol{\Theta}_A$ characterizing the Gaussian excitation, generated from the IS density $q_{\boldsymbol{\Theta}_A|\boldsymbol{\Theta}_B=\boldsymbol{\theta}_B^{(2,i)}}(\boldsymbol{\theta}_A)$ suggested in [2]. The IS density $q_{\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B; \hat{\boldsymbol{\nu}}_{L_2+L_1})$ is determined by applying the multi-level CE method on the distribution sequence in Eq. (34). For first-passage problems of linear systems, one can construct $q_{\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B; \hat{\boldsymbol{\nu}}_{L_2+L_1})$ efficiently, based on the framework introduced in [24, 25]. Here, an analytical approximation of the conditional first-passage probability, $P_{F|\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B)$, deduced based on Rice's formula, is employed to solve the CE optimization problem. The use of the analytical approximation facilitates smooth convergence of the CE method and reduces the optimization effort without compromising much on accuracy. The IS estimator of $P_{F|\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B)$ is applied for evaluating the posterior failure probability according to Eq. (37).

5. Choice of parametric density

In the CE method, the parametric density $q_{\mathbf{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu})$ is typically chosen such that it contains the nominal density of the uncertain model parameters. In the context of the Bayesian updating problem, the nominal density corresponds to the prior PDF $p_{\mathbf{\Theta}}(\boldsymbol{\theta})$. Without loss of generality, we assume that the prior distribution of the random vector $\mathbf{\Theta} = \{\Theta_1; \ldots; \Theta_{n_{\boldsymbol{\theta}}}\}$ representing the uncertain model parameters is the independent standard Gaussian distribution. Then the prior PDF is given by $p_{\mathbf{\Theta}}(\boldsymbol{\theta}) = \prod_{j=1}^{n_{\boldsymbol{\theta}}} p_{\Theta_j}(\theta_j)$, where for every j, $p_{\Theta_j}(\theta_j)$ is a one-dimensional standard Gaussian PDF for Θ_j. When the model parameters are a-priori non-Gaussian and dependent, they are generated from the standard Gaussian random vector $\mathbf{\Theta}$ by means of the Nataf transformation [12] or the Rosenblatt transformation [20].

The Gaussian distribution family is a standard choice of the parametric family in the CE method [36, 27]. To allow for efficient representation of multi-modal posterior distributions, we consider a multivariate Gaussian mixture (GM) model as the parametric density. The PDF of a GM model is defined as the sum of a number of Gaussian PDFs, each of them multiplied by a weighing factor:

$$q_{\mathbf{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu}) = \sum_{s=1}^{n_{GM}} \pi_s f_{\mathrm{G}}(\boldsymbol{\theta}; \boldsymbol{\mu}_s, \mathbf{\Sigma}_s), \tag{38}$$

where $f_{\mathrm{G}}(\boldsymbol{\theta}; \boldsymbol{\mu}_s, \mathbf{\Sigma}_s)$ is the s-th variate Gaussian PDF with mean $\boldsymbol{\mu}_s$ and covariance matrix $\mathbf{\Sigma}_s$ and $\{\pi_s; s = 1, \ldots, n_{GM}\}$ are normalized weights satisfying the condition $\sum_{s=1}^{n_{GM}} \pi_s = 1$. In Eq. (38), n_{GM} denotes the number of modes, which can be fixed a-priori or selected on the fly [17]. The parameter vector is given by $\boldsymbol{\nu} = \{\pi_s, \boldsymbol{\mu}_s, \mathbf{\Sigma}_s; s = 1, \ldots, n_{GM}\}$, where π_s is scalar-valued, $\boldsymbol{\mu}_s$ is a vector of dimension $n_{\boldsymbol{\theta}}$ and $\mathbf{\Sigma}_s$ is an $n_{\boldsymbol{\theta}} \times n_{\boldsymbol{\theta}}$ symmetric matrix. This results in a total of $n_{GM}\frac{n_{\boldsymbol{\theta}}(n_{\boldsymbol{\theta}}+3)}{2} + (n_{GM} - 1)$ unknown parameters in the parametric density. For the uni-modal case, i.e., $n_{GM} = 1$, closed form analytical expressions for the parameter update in Eqs. (12) and (18) are available [36]. For the general case of $n_{GM} > 1$, the parameters are updated by means of an expectation-maximization (EM) algorithm. The EM procedure and the updating rules for the parameters of the GM model are described in [17] and are not further discussed here.

It is noted that in high dimensional problems, i.e., in problems where the number $n_{\boldsymbol{\theta}}$ of uncertain model parameters is large, the CE method with Gaussian densities performs poorly. This is due to two reasons: the first is the degeneracy of the importance sampling weight in high dimensions [3, 26]. The second reason is the number of parameters in the GM model, which increases quadratically with $n_{\boldsymbol{\theta}}$. This results in a rapid increase in the number of samples per level N required to obtain an adequate estimate of the optimal parameter values. In such cases, it is beneficial to consider alternative parametric densities, such as the von-Mises-Fisher-Nakagami distribution family [43, 35], within the CE method.

6. Numerical illustrations

We investigate the performance of the proposed CE-based reliability updating (CEIS-RelUp) method by means of three numerical examples. The first example considers the reliability of a structural component subjected to fatigue updated with measurements of the crack size. Here we update the reliability of an infinite size plate with fatigue crack based on measurements of the crack size. The second example considers a geotechnical engineering problem. Here we apply CEIS-RelUp to update the reliability of an infinite clay slope based on measurements of the undrained shear strength. The third example involves dynamic reliability updating, where the first-passage probability of a randomly excited two-story moment-resisting frame is updated based on modal data.

The performance of the CEIS-RelUp method is assessed in terms of the sample mean and sample CoV of the estimates of $P_{F|\mathbf{d}}$, denoted by $\hat{P}_{F|\mathbf{d}}$ and $\hat{\delta}$ in this section, and in terms of the required computational effort. The sampling variance of the estimators $\hat{P}_1$ and $\hat{P}_2$ in the denominator and numerator of Eq. (22) contribute to the variability of $\hat{P}_{F|\mathbf{d}}$. The sample CoV of $\hat{P}_1$ and $\hat{P}_2$, denoted by $\hat{\delta}_1$ and $\hat{\delta}_2$, as well as the estimates of the marginal likelihood, denoted by $\hat{c}_E$, are also reported. The computational effort is assessed in terms of the number of samples of $\boldsymbol{\Theta}$ expended for CE optimization and reliability estimation. $N_{\mathrm{CE},1}$ and $N_{\mathrm{CE},2}$, respectively, denote the CE optimization effort required to construct the parametric IS densities $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \hat{\boldsymbol{\nu}}_{L_1})$ and $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \hat{\boldsymbol{\nu}}_{L_2+L_1})$. $N_{\mathrm{IS},1}$ and $N_{\mathrm{IS},2}$, respectively, denote the number of samples employed in the IS estimators $\hat{P}_1$ and $\hat{P}_2$ during reliability estimation. The sample size in the reliability estimation step is selected using two approaches. In the first approach, the sample size is taken equal to the number of samples per level for CE optimization, i.e., $N_{\mathrm{IS},1} = N_{\mathrm{IS},2} = N$. In the second approach, $N_{\mathrm{IS},1}$ and $N_{\mathrm{IS},2}$ are selected adaptively on the fly to ensure that an estimate of the CoV of the IS estimate of $P_{F|\mathbf{d}}$ adheres to a specified target value δ^*. The adaptive variant of the IS estimator is implemented according to the procedure described in [24]. The performance measures are averaged over 500 independent simulation runs in Examples 6.1 and 6.2 and 50 simulation runs in Example 6.3.

6.1. Fatigue crack growth

We consider an infinite size plate with fatigue crack, adapted from [13, 39]. The objective is to update the reliability of the plate based on measurements of the crack size. The rate of crack growth is described by Paris' Law as

$$\frac{\mathrm{d}a(n)}{\mathrm{d}n} = C[\Delta S\sqrt{\pi a(n)}]^m, \tag{39}$$

where a is the size of the crack, n is the number of stress cycles, ΔS is the stress range per cycle (constant stress amplitude is assumed) and C and m are empirically determined model parameters. The crack size as a function of the number of stress cycles n is given by [13]:

$$a(n) = \left[(1 - \frac{m}{2})C\Delta S^m \pi^{\frac{m}{2}} n + {a_0}^{1-\frac{m}{2}}\right]^{\frac{1}{1-\frac{m}{2}}}, \tag{40}$$

where a_0 denotes the initial crack size. The failure event is defined in terms of the number of stress cycles to failure. The performance function is given by

$$g = n_c - n_f, \tag{41}$$

where n_c denotes the number of stress cycles required to reach a critical crack size of a_c and n_f denotes the number of stress cycles at which the reliability is estimated. For an infinite plate, n_c is given by

$$\begin{aligned} n_c &= \frac{2}{(m-2)C(\sqrt{\pi}\Delta S)^m}\left[\frac{1}{a_0{}^{\frac{m-2}{2}}} - \frac{1}{a_c{}^{\frac{m-2}{2}}}\right], m \neq 2 \\ &= \frac{1}{\pi C \Delta S^2}\ln\left(\frac{a_c}{a_0}\right), m = 2. \end{aligned} \tag{42}$$

The prior probabilistic description of the uncertain model parameters is given in Table 1.

Table 1: Prior probabilistic description of the parameters of the crack growth problem in Example 6.1.

Parameter	Distribution	Mean	Standard deviation	Correlation
a_0[mm]	Exponential	1	1	-
a_c[mm]	Deterministic	50	-	-
ΔS[Nmm^{-2}]	Normal	60	10	-
$(\ln(C)[\text{N}], m[\text{mm}])$	Bi-Normal	(-33,3.5)	(0.47,0.3)	$\rho_{\ln(C),m} = -0.9$

We estimate the reliability at $n_f = 8 \times 10^5$ stress cycles. The prior value of the probability of failure, based on 2×10^5 standard Monte Carlo samples, is 9.2×10^{-3}. The failure probability is updated via the likelihood function

$$L(\boldsymbol{\theta}|\mathbf{d}) = \prod_{i=1}^{n_M} \exp\left(-\frac{1}{2}\left(\frac{a(\boldsymbol{\theta}, n_i) - a_{m,i}}{\sigma_n}\right)^2\right), \tag{43}$$

where n_M is the number of measurements, σ_n is the standard deviation of the measurement noise, n_i is the number of stress cycles up to the i-th measurement and $a_{m,i}$ are the crack size measurements. We implement CEIS-RelUp with a uni-modal Gaussian distribution as the parametric family. In the present example, where the number of uncertain model parameters is $n_{\boldsymbol{\theta}} = 4$, the Gaussian density is comprised of 14 unknown parameters, which are updated analytically during CE optimization. We investigate the influence of the number of measurements and the standard deviation of the measurement noise on the performance of the method. The results for the two case studies are summarized in the following.

6.1.1. Case study: Effect of standard deviation of measurement noise

We consider two measurements of crack size:

$$\begin{aligned} a_{m,1} &= 1.7\text{mm at } n_1 = 10^5 \text{ stress cycles} \\ a_{m,2} &= 1.8\text{mm at } n_2 = 3 \times 10^5 \text{ stress cycles} \end{aligned} \tag{44}$$

The posterior probability of failure, $P_{F|\mathbf{d}}$, is estimated with $\sigma_n = 0.5$mm, 0.25mm and 0.125mm. In our experiment, the sample size, per level, for CE optimization is varied

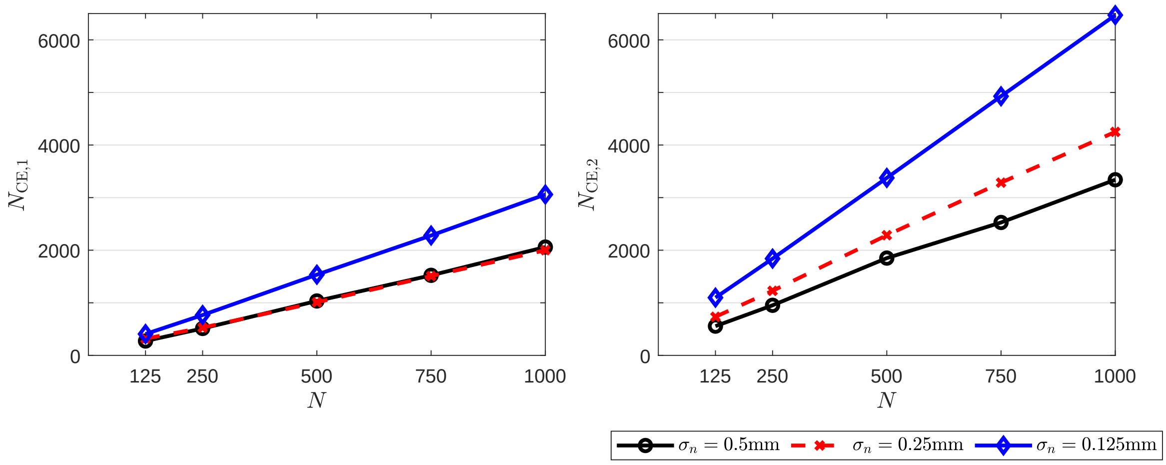

Figure 1: Cross entropy optimization effort for infinite size plate with fatigue crack

between $N = 125$ and $N = 1000$. Fig. 1 shows the total number of samples, or equivalently the number of model evaluations, required to construct the IS densities $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \hat{\boldsymbol{\nu}}_{L_1})$ and $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \hat{\boldsymbol{\nu}}_{L_2+L_1})$. It is observed that the computational effort required for CE optimization increases with decrease in σ_n. With decreasing standard deviation of the measurement noise, the likelihood function gets more concentrated; consequently the target densities $q^*_{P_{F|\mathbf{d}}}(\boldsymbol{\theta})$ and $p_{\boldsymbol{\Theta}|\mathbf{d}}(\boldsymbol{\theta})$ have lower standard deviation and their difference to the prior increases. This results in an increase in the number of levels, and hence the number of samples, required to reach the target densities by the multi-level CE method. Furthermore, we observe that $N_{\mathrm{CE},2}$ is larger than $N_{\mathrm{CE},1}$, which indicates that the number of levels required by the CE method to converge to $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \hat{\boldsymbol{\nu}}_{L_2+L_1})$ is more than that required for $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \hat{\boldsymbol{\nu}}_{L_1})$. For the values of σ_n considered, the number of levels range, on average, from $L_1 = 2$ to 3 and $L_2 =$ 3 to 9 for the target densities $p_{\boldsymbol{\Theta}|\mathbf{d}}(\boldsymbol{\theta})$ and $q^*_{P_{F|\mathbf{d}}}(\boldsymbol{\theta})$, respectively. The required number of levels indicates that the posterior is closer to the prior as compared to the optimal IS density of the failure domain and the posterior.

We estimate the posterior failure probability with the non-adaptive (N_{IS}-NonAdap) and adaptive (N_{IS}-Adap) variants of the IS estimator, $\hat{P}_{F|\mathbf{d}}$. In the latter case, the target CoV of $\hat{P}_{F|\mathbf{d}}$ is set to $\delta^* = 0.10$ and 0.05. Recall that the contribution to the CoV of $\hat{P}_{F|\mathbf{d}}$ comes from the two IS estimators, $\hat{P}_1$ and $\hat{P}_2$, in the denominator and numerator of Eq. (22). In the adaptive case, the sample sizes of these estimators, i.e., $N_{\mathrm{IS},1}$ and $N_{\mathrm{IS},2}$, are selected adaptively such that the respective sample CoVs adhere to the target values δ_1^* and δ_2^*, where $\delta_1^* = \delta_2^* = \delta^*/\sqrt{2}$. These target CoVs are equal to 0.071 for $\delta^* = 0.10$ and 0.035 for $\delta^* = 0.05$. The Monte Carlo estimate of the failure probability, using 2×10^5 samples obtained from the posterior PDF through rejection sampling, is 3.7×10^{-3}(CoV $\approx 3.7\%$), 1.7×10^{-3} (CoV $\approx 5.4\%$) and 9.5×10^{-5} (CoV $\approx 23\%$) for $\sigma_n = 0.5$mm, 0.25mm and 0.125mm, respectively. The simulation results for $N = 250$ and 500 are reported in Table 2. The sample mean of the posterior failure probability estimates are comparable with the reference solution, for all σ_n. However, the sampling variability of $\hat{P}_{F|\mathbf{d}}$ changes significantly with σ_n. As σ_n decreases, the posterior PDF becomes significantly different from the prior, and the failure event under the posterior probability measure becomes increasing rare. These factors increase the number of levels for convergence due to the reduced ability of the parametric family in describing the target densities $q^*_{P_{F|\mathbf{d}}}(\boldsymbol{\theta})$ and $p_{\boldsymbol{\Theta}|\mathbf{d}}(\boldsymbol{\theta})$, thereby leading to an increase in the sample CoV of the IS estimators for small σ_n. Hence, when the sample size of these estimators is fixed, i.e., for the non-adaptive case with $N_{\text{IS},1} = N_{\text{IS},2} = N$, we observe a monotonic increase in the respective sample CoVs, i.e., $\hat{\delta}_1$ and $\hat{\delta}_2$ in Table 2, with decrease in σ_n. The increase is significant for $\hat{\delta}_2$. Accordingly, when the sample size is selected adaptively, the IS estimators require a larger number of samples to achieve the target CoV for small values of σ_n.

Table 2: Posterior failure probability estimates for fatigue crack growth, with two measurements, for different standard deviation of measurement noise. Reference value of the probability of failure, based on 2×10^5 samples obtained through rejection sampling, is 3.7×10^{-3}, 1.7×10^{-3} and 9.5×10^{-5} for $\sigma_n = 0.5$mm, 0.25mm and 0.125mm, respectively.

			$\hat{c}_E$	$\hat{P}_{F\|\mathbf{d}}$	$\hat{\delta}$	$\hat{\delta}_1$	$\hat{\delta}_2$	$N_{\text{IS},1}$	$N_{\text{IS},2}$	N_{T}
$\sigma_n = 0.500$	$N = 250$	N_{IS}-NonAdap	0.159	3.86×10^{-3}	0.09	0.03	0.10	250	250	1965
		N_{IS}-Adap ($\delta^* = 0.10$)	0.159	3.85×10^{-3}	0.11	0.07	0.08	28	432	1926
		N_{IS}-Adap ($\delta^* = 0.05$)	0.159	3.82×10^{-3}	0.07	0.04	0.06	120	1150	2735
	$N = 500$	N_{IS}-NonAdap	0.161	3.84×10^{-3}	0.08	0.01	0.07	500	500	3885
		N_{IS}-Adap ($\delta^* = 0.10$)	0.159	3.87×10^{-3}	0.08	0.06	0.06	19	568	3472
		N_{IS}-Adap ($\delta^* = 0.05$)	0.160	3.83×10^{-3}	0.06	0.04	0.05	71	989	3945
$\sigma_n = 0.250$	$N = 250$	N_{IS}-NonAdap	0.076	1.65×10^{-3}	0.21	0.05	0.20	250	250	2255
		N_{IS}-Adap ($\delta^* = 0.10$)	0.074	1.65×10^{-3}	0.12	0.08	0.09	48	564	2367
		N_{IS}-Adap ($\delta^* = 0.05$)	0.074	1.64×10^{-3}	0.07	0.04	0.06	195	1473	3423
	$N = 500$	N_{IS}-NonAdap	0.075	1.64×10^{-3}	0.07	0.02	0.07	500	500	4295
		N_{IS}-Adap ($\delta^* = 0.10$)	0.074	1.66×10^{-3}	0.10	0.07	0.07	26	608	3929
		N_{IS}-Adap ($\delta^* = 0.05$)	0.075	1.65×10^{-3}	0.06	0.04	0.05	104	1220	4619
$\sigma_n = 0.125$	$N = 250$	N_{IS}-NonAdap	0.034	9.75×10^{-5}	0.30	0.05	0.30	250	250	3108
		N_{IS}-Adap ($\delta^* = 0.10$)	0.034	9.54×10^{-5}	0.12	0.06	0.10	61	868	3536
		N_{IS}-Adap ($\delta^* = 0.05$)	0.034	9.54×10^{-5}	0.09	0.04	0.08	274	2609	5490
	$N = 500$	N_{IS}-NonAdap	0.034	9.68×10^{-5}	0.26	0.02	0.26	500	500	5910
		N_{IS}-Adap ($\delta^* = 0.10$)	0.034	9.63×10^{-5}	0.10	0.06	0.08	45	723	5678
		N_{IS}-Adap ($\delta^* = 0.05$)	0.034	9.67×10^{-5}	0.07	0.04	0.06	157	2032	7099

The performance of CEIS-RelUp is assessed for different sample size N during CE optimization. We observe that the sample mean of the posterior failure probability estimates is broadly similar for all N. However, there is significant change in the sampling variability of the estimators and the required computational effort. The variation in the sample CoV of $\hat{P}_{F|\mathbf{d}}$ and the total computational effort, for $\sigma_n = 0.5$mm and 0.125mm, $\delta^* = 0.10$ and 0.05, are shown in Fig. 2. For the non-adaptive variant of the IS estimator, N_{IS}-NonAdap, an increase in N reduces the sample CoV, $\hat{\delta}$, of the posterior failure probability estimator. This

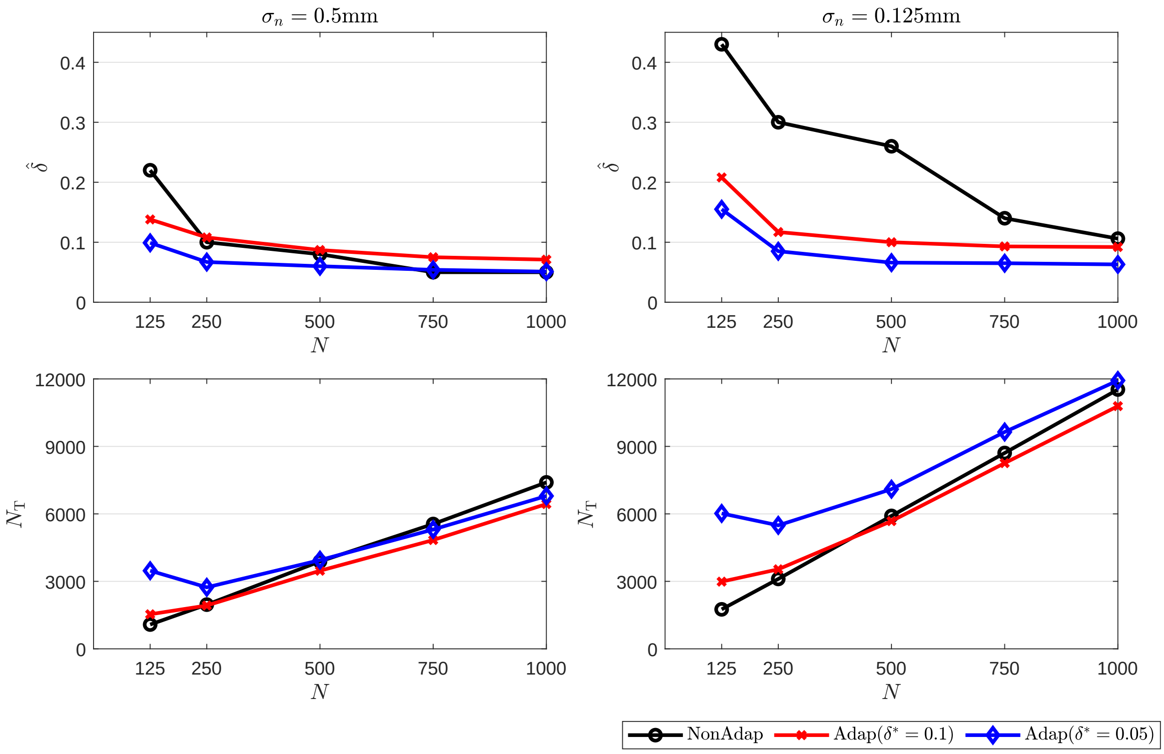

Figure 2: Coefficient of variation of posterior failure probability estimates and total computational effort for infinite size plate with fatigue crack

behavior is due to two factors. First, the number of effective samples available to fit the parametric densities increases with N. This results in improved estimates of the parameter vectors $\hat{\boldsymbol{\nu}}_{L_1}$ and $\hat{\boldsymbol{\nu}}_{L_2+L_1}$, and better approximation of the respective optimal IS densities. Second, an increase in N implies a monotonic increase in the sample size of the IS estimators $\hat{P}_1$ and $\hat{P}_2$, which leads to a reduction in the sample CoVs $\hat{\delta}_1$ and $\hat{\delta}_2$, respectively. The decrease in sampling fluctuations for N_{IS}-NonAdap is, however, at the expense of increased computational effort. In case of N_{IS}-Adap, the sample CoV δ_1 is close to the target value, i.e., $\delta_1^* = 0.071$ for $\delta^* = 0.10$ and $\delta_1^* = 0.035$ for $\delta^* = 0.05$, for all N. The estimates of δ_2 are initially large for $\delta^* = 0.10$, but they gradually reduce to 0.071 as N increases. For $\delta^* = 0.05$, however, $\hat{\delta}_2$ remains larger than the target value 0.035. This is due to inaccuracy in the estimator of δ_2, in Eq. (30), used for the adaptive selection of $N_{\text{IS},2}$. The sub-optimality in the IS density $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \hat{\boldsymbol{\nu}}_{L_2+L_1})$ for small σ_n induces possible bias in the estimator, due to which it decays faster than the true value. We note that in the adaptive variant of the method, the number of samples for estimation of the denominator can be smaller than N, whereas for the numerator it is always greater than N. This is because checking the termination criterion for the numerator requires N samples from the final density [35]. Overall, it is observed that the performance of N_{IS}-NonAdap and N_{IS}-Adap are comparable for $\sigma_n = 0.5\text{mm}$. For

$\sigma_n = 0.125$mm, the adaptive variant of the IS estimator exhibits superior performance. In the latter case, N_{IS}-NonAdap requires 11530 samples to yield a sample CoV of 10% of the IS estimator $\hat{P}_{F|\mathbf{d}}$, whereas N_{IS}-Adap yields a sample CoV of 10% and 6% with approximately 5500 and 7000 samples, respectively.

6.1.2. Case study: Effect of number of measurements

To investigate the influence of the number of measurements, n_M, on the performance of CEIS-RelUp, we consider two additional observations of the crack size:

$$\begin{aligned} a_{m,3} &= 1.9\text{mm at } n_3 = 4 \times 10^5 \text{ stress cycles} \\ a_{m,4} &= 2.1\text{mm at } n_4 = 5 \times 10^5 \text{ stress cycles} \end{aligned} \tag{45}$$

The posterior probability of failure, $P_{F|\mathbf{d}}$, for σ_n =0.5mm is 5.1×10^{-4} (CoV $\approx$ 10%). The reference solution is evaluated based on 2×10^5 samples obtained through rejection sampling. We evaluate $P_{F|\mathbf{d}}$ by CEIS-RelUp, using non-adaptive and adaptive selection of the sample size of the IS estimator. The results with N_{IS}-Adap correspond to $\delta^* = 0.10$. We select $N = 250$ and 500 samples per level during CE optimization. It is observed that the CE optimization effort, i.e., the number of levels required to construct the IS densities $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \hat{\boldsymbol{\nu}}_{L_1})$ and $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \hat{\boldsymbol{\nu}}_{L_1+L_2})$, increases with the number of measurements. The increase is marginal for $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \hat{\boldsymbol{\nu}}_{L_1})$, but approximately twice for $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \hat{\boldsymbol{\nu}}_{L_1+L_2})$. The results of reliability estimation are reported in Table 3. The sample mean of the posterior failure probability estimator, $\hat{P}_{F|\mathbf{d}}$, compares well with the reference solution. For N_{IS}-NonAdap, the sample CoV, $\hat{\delta}$, of $\hat{P}_{F|\mathbf{d}}$ decreases with increase in N, which was also observed in Table 2. For N_{IS}-Adap, $\hat{\delta}$ remains close to the specified target value. We observe that an increase in the number of measurements causes a significant increase in the sample CoV of the IS estimators $\hat{P}_1$ and $\hat{P}_2$, leading to larger sampling fluctuations in $\hat{P}_{F|\mathbf{d}}$. When the sample size of these estimators, i.e., $N_{\text{IS},1}$ and $N_{\text{IS},2}$, are fixed to N, the respective sample CoVs $\hat{\delta}_1$ and $\hat{\delta}_2$ are approximately twice of those obtained for $n_M = 2$. Similarly, when $N_{\text{IS},1}$ and $N_{\text{IS},2}$ are selected adaptively, the number of samples required to meet the target CoV increases with n_M. Finally, we observe that N_{IS}-Adap remains more efficient than N_{IS}-NonAdap, since it yields a smaller sample CoV of $\hat{P}_{F|\mathbf{d}}$ with comparable total computational effort.

Table 3: Posterior failure probability estimates for fatigue crack growth by CEIS-RelUp, with $n_M = 4$ and $\sigma_n = 0.5$mm. Reference value of the probability of failure, based on 2×10^5 samples obtained through rejection sampling, is 5.1×10^{-4}.

		$\hat{c}_E$	$\hat{P}_{F\mid\mathbf{d}}$	$\hat{\delta}$	$\hat{\delta}_1$	$\hat{\delta}_2$	$N_{\text{IS},1}$	$N_{\text{IS},2}$	N_{T}
$N = 250$	N_{IS}-NonAdap	0.087	5.38×10^{-4}	0.23	0.05	0.22	250	250	2582
	N_{IS}-Adap ($\delta^* = 0.10$)	0.085	5.42×10^{-4}	0.12	0.08	0.09	57	617	2756
$N = 500$	N_{IS}-NonAdap	0.086	5.41×10^{-4}	0.18	0.02	0.17	500	500	4918
	N_{IS}-Adap ($\delta^* = 0.10$)	0.085	5.30×10^{-4}	0.10	0.07	0.08	37	773	4728

6.2. Stability of an infinite clay slope

In this example, we apply the CEIS-RelUp approach to update the reliability of a saturated (infinite) clay slope under undrained conditions. The slope, shown in Fig. 3, has a height of H =5m, a slope angle of $\beta = 15°$ and a saturated unit weight of $\gamma = 20\text{kN/m}^3$. The short-term shear strength of the clay is characterized by the undrained shear strength, which is assumed to vary with depth from the soil surface. The factor of safety governing the slope stability is given by [18]

$$FS(z) = \frac{s_u(z)}{\gamma z \sin\beta\cos\beta}, \tag{46}$$

where $s_u(z)$ denotes the undrained shear strength at a depth z below the ground surface.

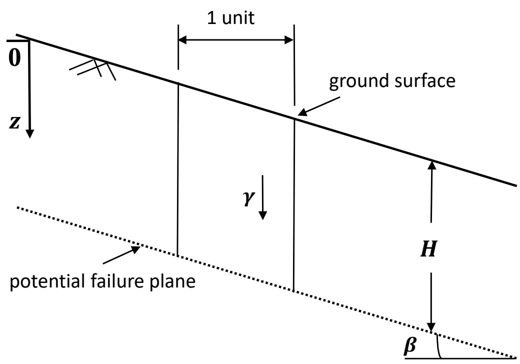

Figure 3: Infinite clay slope in Example 6.2

The failure event F of the slope is defined as $FS_{\min}$ being < 1.0, where $FS_{\min}$ is the minimum factor of safety over the height of the slope. The depth dependent nature of the undrained shear strength is characterized by the non-stationary random field model [22]

$$s_u(z) = s_{u0} + b\gamma z \exp[w(z)], \tag{47}$$

where s_{u0} is the undrained shear strength at the ground surface, b is a trend parameter that determines the rate of increase of strength with soil depth and $w(z)$ is the randomly fluctuating component of s_u, which is modeled as a one-dimensional zero mean Gaussian random field with constant standard deviation, $\sigma_w = 0.24$. To characterize the spatial correlation of s_u, we assume an exponential auto-correlation function of $w(z)$, with a correlation length of 1.9m [23]. $w(z)$ is numerically represented in terms of a finite number of random variables through the Karhunen-Loève (KL) expansion:

$$w(z) = \sum_{i=1}^{n_{KL}} \sqrt{\lambda_i}\phi_i(z)\theta_i^{KL}, \tag{48}$$

where $\{(\lambda_i, \phi_i(z)), i = 1, \dots, n_{KL}\}$ are eigenpairs of the auto-covariance function, arranged in decreasing order of magnitude of the eigenvalues, and $\left\{\theta_i^{KL}, i = 1, \dots, n_{KL}\right\}$ are independent standard Gaussian random variables. We consider $n_{KL} = 10$ eigenmodes in the KL expansion. Following [23], we model the prior distribution of the parameters s_{u0} and b by lognormal random variables, with means $\mu_{s_{u0}} = 14.67$kPa and $\mu_b = 0.272$, and standard deviations $\sigma_{s_{u0}} = 4.04$kPa and $\sigma_b = 0.189$. In this way, a total of $n_{\boldsymbol{\theta}} = 12$ random variables are required to represent the non-stationary random field $s_u(z)$.

For the purpose of reliability analysis, we discretize the soil profile into 100 equal slices of height $\Delta h = H/100$. The factor of safety is evaluated at the base of each slice, resulting in 100 different factors of safety $\{FS(z_i), i = 1, \dots, 100\}$, where $z_i = i\Delta h$. The minimum factor of safety is evaluated as $FS_{\min} = \min_{i=1,\dots,100} FS(z_i)$. Without measurements, the prior probability of slope failure is 1.49×10^{-1} as obtained from 10^5 standard Monte Carlo samples. The following measurements of the undrained shear strength are used to update the failure probability:

$$
\begin{aligned}
s_{u_{m,1}} &= 17.8\text{kPa at } z_{m,1} = 1.5\text{m} \\
s_{u_{m,2}} &= 24.5\text{kPa at } z_{m,2} = 3.0\text{m} \\
s_{u_{m,3}} &= 30.5\text{kPa at } z_{m,3} = 4.5\text{m}
\end{aligned}
\tag{49}
$$

The measurement result $s_{u_{m,i}}$ at a given location $z_{m,i}$ is related to the true value by means of independent multiplicative error $\epsilon_{m,i}$, that is assumed to follow a lognormal distribution with median equal to one and constant standard deviation. With this assumption, the likelihood function is given by [40]

$$
L(\boldsymbol{\theta}|\mathbf{d}) = \exp\left(-\sum_{i=1}^{n_M} \frac{\left[\ln s_{u_{m,i}} - \ln s_u(z_{m,i}, \boldsymbol{\theta})\right]^2}{2\sigma^2_{\ln\epsilon_{m,i}}}\right), \tag{50}
$$

where $\sigma_{\ln\epsilon_{m,i}} = \sqrt{\ln(1 + \text{CoV}^2_{\epsilon_{m,i}})}$ is the standard deviation of $\ln\epsilon_{m,i}$. The coefficient of variation of $\epsilon_{m,i}$ is set to $\text{COV}_{\epsilon_{m,i}} = 5\%$ in this example. A reference value of the posterior probability of failure based on 10^6 samples obtained through rejection sampling is 6.37×10^{-4}(CoV $\approx 4\%$).

We implement CEIS-RelUp with a uni-modal Gaussian distribution as the parametric family. In the present example, where the number of uncertain model parameters is $n_{\boldsymbol{\theta}} = 12$, the Gaussian density is comprised of 90 unknown parameters. In Fig. 4 we show the prior and posterior statistics of the undrained shear strength, s_u, and the factor of safety, FS. The estimates of the prior statistics are obtained through standard Monte Carlo simulation from the prior PDF. The estimates of the posterior statistics are obtained from CEIS-RelUp, through importance sampling from the fitted IS densities $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \hat{\boldsymbol{\nu}}_{L_1})$ and $q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \hat{\boldsymbol{\nu}}_{L_2+L_1})$. A comparison of the estimates from CEIS-RelUp with the ones obtained through rejection sampling (RS) demonstrates good agreement. The mean of the posterior of s_u conditional on the domain of the failure event, F, is close to the mean of the unconditional posterior, which indicates that the PDF of the uncertain model parameters shifts towards the failure domain after the updating. There is a reduction in the spread of the updated PDF, as indicated by smaller standard deviation of the posterior of s_u in comparison to the prior, which, in turn, results in a lower posterior probability of failure. The variation in the mean of FS indicates that failure is more likely to occur.

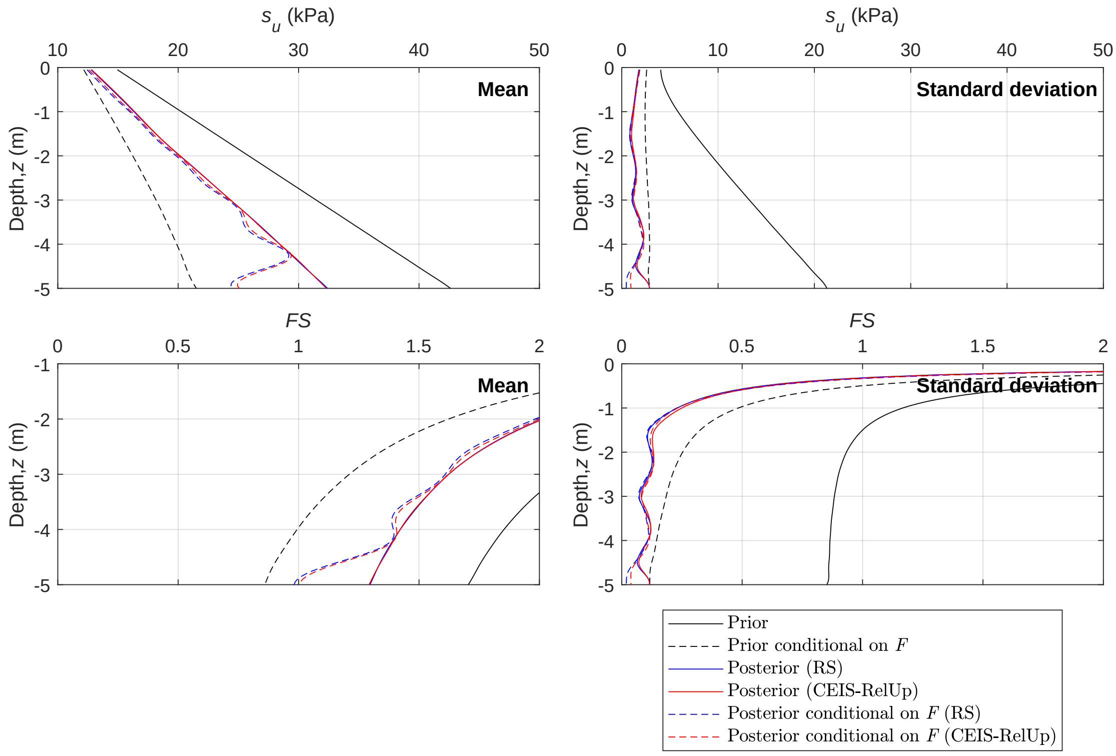

Figure 4: Prior and posterior statistics of the undrained shear strength, s_u, and the factor of safety, FS. Top-left: Variation of mean of s_u with depth. Top-right: Variation of standard deviation of s_u with depth. Bottom-left: Variation of mean of FS with depth. Bottom-right: Variation of standard deviation of FS with depth.

The results of reliability estimation, as well as the computational effort required to construct the IS densities $q_{\mathbf{\Theta}}(\boldsymbol{\theta}; \hat{\boldsymbol{\nu}}_{L_1})$ and $q_{\mathbf{\Theta}}(\boldsymbol{\theta}; \hat{\boldsymbol{\nu}}_{L_2+L_1})$, are reported in Table 4. We employ the non-adaptive (N_{IS}-NonAdap) and adaptive (N_{IS}-Adap) variants of the IS estimator, $\hat{P}_{F|\mathbf{d}}$, to estimate the posterior failure probability. In the latter case, the target CoV of $\hat{P}_{F|\mathbf{d}}$ is set to $\delta^* = 0.10$, which corresponds to target values $\delta_1^* = \delta_2^* = 0.071$ of the CoV of the estimators $\hat{P}_1$ and $\hat{P}_2$, in the denominator and numerator of Eq. (22). The estimates of $N_{\text{CE},1}$ and $N_{\text{CE},2}$ indicate that the required CE optimization effort is higher in the second stage, i.e., the number of levels required to construct $q_{\mathbf{\Theta}}(\boldsymbol{\theta}; \hat{\boldsymbol{\nu}}_{L_2+L_1})$ is approximately twice of that required for $q_{\mathbf{\Theta}}(\boldsymbol{\theta}; \hat{\boldsymbol{\nu}}_{L_1})$. As the sample size per level N increases, we observe a decrease in the required number of levels for convergence of the CE method. This is due to an increase in the number of effective samples available to fit the parametric densities, which facilitates faster convergence of the CE method.

With $N = 500$ samples per level, we observe an underestimation in the posterior failure probability estimates as obtained from CEIS-RelUp. This is due to bias in the estimates of the numerator $\hat{P}_2$. The estimate of the marginal likelihood, although not reported, is observed to be accurate for all N. In comparison, approximating the optimal IS density of the numerator, i.e., $q^*_{P_{F|\mathbf{d}}}(\boldsymbol{\theta})$, is more challenging. For a small N, the available number of effective samples is not sufficient to adequately approximate $q^*_{P_{F|\mathbf{d}}}(\boldsymbol{\theta})$, leading to bias in the failure probability estimates. With increase in N, we also observe a reduction in the sampling variability of the IS estimators. With N_{IS}-NonAdap, there is a gradual decrease in the sample CoV of the IS estimators $\hat{P}_1$ and $\hat{P}_2$, and hence of the estimator $\hat{P}_{F|\mathbf{d}}$. The results with non-adaptive and adaptive variants of the IS estimator indicate that $\hat{P}_2$ has a larger variability than $\hat{P}_1$, which indicates reduced flexibility of the parametric density in describing the posterior PDF over the failure domain. Overall, both variants of the IS estimator require similar total computational effort, N_T. Hence, selecting the sample size of the IS estimators adaptively does not offer a clear advantage in this example. This is because a large number of samples per level N is required to obtain an adequate parametric IS density for estimating the updated failure probability. Both variants of the method employ at least N samples for estimation and when N is large the CoV of the probability estimate is already small enough with N samples.

Table 4: Posterior failure probability estimates of infinite clay slope by CEIS-RelUp. Reference value of the posterior probability of failure based on 10^6 samples obtained through rejection sampling is 6.37×10^{-4}.

		$N_{\text{CE},1}$	$N_{\text{CE},2}$	$\hat{P}_{F\|\mathbf{d}}$	$\hat{\delta}$	$\hat{\delta}_1$	$\hat{\delta}_2$	$N_{\text{IS},1}$	$N_{\text{IS},2}$	N_{T}
$N = 500$	N_{IS}-NonAdap	2275	5319	5.71×10^{-4}	0.35	0.09	0.34	500	500	8594
	N_{IS}-Adap ($\delta^* = 0.10$)	2275	5319	5.94×10^{-4}	0.18	0.07	0.17	437	3412	11443
$N = 750$	N_{IS}-NonAdap	2798	6250	6.23×10^{-4}	0.15	0.09	0.14	750	750	10548
	N_{IS}-Adap ($\delta^* = 0.10$)	2798	6250	6.30×10^{-4}	0.13	0.07	0.11	237	1911	11196
$N = 1000$	N_{IS}-NonAdap	3399	7282	6.26×10^{-4}	0.11	0.05	0.10	1000	1000	12681
	N_{IS}-Adap ($\delta^* = 0.10$)	3399	7282	6.38×10^{-4}	0.11	0.07	0.08	201	1694	12576
$N = 2000$	N_{IS}-NonAdap	6253	12898	6.36×10^{-4}	0.08	0.03	0.07	2000	2000	23151
	N_{IS}-Adap ($\delta^* = 0.10$)	6253	12898	6.39×10^{-4}	0.09	0.06	0.06	140	2364	21746

6.3. First-passage failure of a two-story moment-resisting frame

We apply CEIS-RelUp to update the first-passage failure probability of a two-story moment-resisting frame, earlier studied in [5], using its identified natural frequencies. A two degree-of-freedom shear building model, shown in Fig. 5, is used to model the structure in order to identify the inter-story stiffnesses and story masses, and to predict the reliability. The inter-story stiffnesses are parameterized as $k_1 = \alpha_1 \overline{k}_1$ and $k_2 = \alpha_2 \overline{k}_2$, where α_1 and α_2 are the stiffness parameters to be identified, and $\overline{k}_1 = \overline{k}_2 = 29.7 \times 10^6$N/m are the nominal values for the inter-story stiffnesses of the first and second stories, respectively. The story masses are parameterized as $m_1 = \alpha_3 \overline{m}_1$ and $m_2 = \alpha_4 \overline{m}_2$, where α_3 and α_4 are the mass parameters to be updated, and $\overline{m}_1 = 16.5 \times 10^3$kg and $\overline{m}_2 = 16.1 \times 10^3$kg are the nominal values for the first- and second-story masses, respectively. The prior PDF for α_1 to α_4 is given by the product of four lognormal PDFs with most probable values 1.3, 0.8, 0.95 and 0.95, and standard deviations 1, 1, 0.1 and 0.1, respectively.

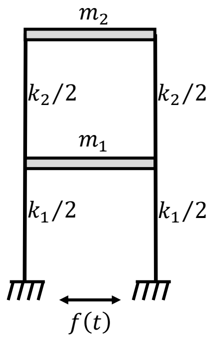

Figure 5: Two degree-of-freedom shear building model in Example 6.3

The first-passage failure probability of the structure subjected to a stochastic ground excitation is predicted using the shear building model. The response of interest is the inter-story drift between the first and the second stories. Failure is defined as the event that the inter-story drift exceeds a threshold level of h^* within a duration of $T = 10$s. The structure is assumed to be subjected to earthquake motion, $f(t)$, modeled by stationary Gaussian white noise with spectral intensity $S = 1 \times 10^{-2}\text{m}^2/\text{s}^3$. The response of the structure is computed at the discrete time instants $\{t_k = (k-1)\Delta t, k = 1, \ldots, n_T\}$, where the time step size is assumed to be $\Delta t = 0.005$s. Hence, the number of time instants is $n_T = T/\Delta t + 1 = 2001$. The stochastic excitation $f(t)$ is characterized by a sequence of independent standard normal random variables $\{\Xi_k, k = 1, \ldots, n_T\}$ that generate the white noise at the discrete time instants, i.e., $\left\{f(t_k) = \sqrt{2\pi S/\Delta t}\Xi_k, k = 1, \ldots, n_T\right\}$. The reliability is predicted for two response thresholds, $h^* = 0.030$m and $h^* = 0.035$m.

In this example, there is a total of $n_{\boldsymbol{\theta}} = n_T + 4 = 2005$ random parameters, of which four parameters (two stiffness parameters α_1 and α_2 and two mass parameters α_3 and α_4) are updated. Using noisy simulated response time histories, the identified natural frequencies are $\tilde{f}_1 = 3.13$Hz and $\tilde{f}_2 = 9.83$Hz, which are used as the data $\mathbf{d}$ in the updating. We evaluate the posterior probability of failure according to the procedure described in Section 4.4. $\boldsymbol{\Theta}_B = \{\alpha_1,\ \alpha_2,\ \alpha_3,\ \alpha_4\}$ are the random variables that are updated based on the data, and $\boldsymbol{\Theta}_A = \{\Xi_1, \ldots, \Xi_{n_T}\}$ are the remaining random variables characterizing the future excitation. Using the modal data $\mathbf{d}$, the likelihood function for updating $\boldsymbol{\Theta}_B$ is formulated as [42]

$$L(\boldsymbol{\theta}_B|\mathbf{d}) = \exp\left(-\frac{1}{2\epsilon^2}\sum_{j=1}^{2}\lambda_j^2\left[\frac{f_j^2(\boldsymbol{\theta}_B)}{\tilde{f}_j^2} - 1\right]^2\right), \tag{51}$$

where $\lambda_1 = \lambda_2 = 1$ are the means and $\epsilon = \frac{1}{16}$ is the standard deviation of the prediction error between each $\tilde{f}_j^2$ and the corresponding model squared frequency $f_j^2(\boldsymbol{\theta}_B)$.

We select a two-component Gaussian mixture (GM) model as the parametric density family. In the present example, where the number of uncertain structural parameters to be updated is 4, a two-component GM distribution is described by 29 unknown parameters that are determined by CE optimization. The parametric IS density $q_{\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B; \hat{\boldsymbol{\nu}}_{L_1})$ for evaluating the marginal likelihood and approximating the posterior PDF of $\boldsymbol{\Theta}_B$ is constructed based on the procedure in Section 3. The parametric density $q_{\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B; \hat{\boldsymbol{\nu}}_{L_2+L_1})$ approximating the optimal IS density of the posterior probability of failure is constructed by applying the multi-level CE method on the distribution sequence in Eq. (34), according to the procedure outlined in [24]. Fig. 6 shows the samples of $\boldsymbol{\Theta}_B$ obtained from the fitted parametric densities. The four components of the samples are shown in two groups: α_1 versus α_2 in the first column and α_3 versus α_4 in the second column of Fig. 6. The posterior joint distribution of the stiffness parameters is bimodal, however, only one of the modes contributes to first-passage failure. For the mass parameters, there is no significant change between the posterior density and the optimal IS density over the failure domain. In both cases, the distribution of the samples obtained from the parametric densities fitted through the CE method compare well with the reference solution obtained through rejection sampling.

We vary the number of samples per level, N, during CE optimization between 250 and 1000. Fig. 7 shows the computational effort needed to fit the parametric IS densities $q_{\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B; \hat{\boldsymbol{\nu}}_{L_1})$ and $q_{\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B; \hat{\boldsymbol{\nu}}_{L_2+L_1})$ by CE optimization. The required optimization effort, i.e., the number of samples $N_{\mathrm{CE},1}$ and $N_{\mathrm{CE},2}$, indicates a marginal decrease in the number of levels to convergence for increasing N, which is attributed to the larger number of effective samples available to fit the parametric densities. We observe that the computational effort for constructing $q_{\boldsymbol{\Theta}_B}(\boldsymbol{\theta}_B; \hat{\boldsymbol{\nu}}_{L_2+L_1})$ is larger for the higher threshold level h^*, as the failure event under the posterior probability measure gets rarer with increase in h^*. This leads to an increase in the number of levels required to estimate the optimal parameters of the IS density that best describe the failure domain.

The posterior first-passage probability of failure is evaluated based on the IS estimator in Eq. (37), wherein the IS density $q_{\boldsymbol{\Theta}_A|\boldsymbol{\Theta}_B=\boldsymbol{\theta}_B}(\boldsymbol{\theta}_A)$ of $\boldsymbol{\Theta}_A$ is selected as suggested in [2]. The results of reliability estimation are reported in Table 5. The simulation results are obtained based on IS densities constructed with $N = 500$ samples per level during CE optimization. The estimates from the adaptive variant of the IS estimator correspond to $\delta^* = 0.10$ and 0.05. The reference solution, evaluated based on 5×10^7 samples obtained through rejection sampling, is 1.85×10^{-4} (CoV $\approx 1\%$) for $h^* = 0.030$m and 4.52×10^{-6} (CoV $\approx 6.7\%$) for $h^* = 0.035$m. The estimates of the marginal likelihood and posterior failure probability obtained through CEIS-RelUp compare well with the reference value for both response thresholds. For a higher h^*, there is an increase in the sample CoV of the posterior failure probability estimates. This is due to an increase in variability of the estimator $\hat{P}_2$ with the threshold level. When the sample size of the IS estimators is fixed, i.e., for N_{IS}-NonAdap, the sample estimate $\hat{\delta}_2$ of the CoV of $\hat{P}_2$ is larger for $h^* = 0.035$m. When the sample size is selected adaptively to meet a prescribed target CoV, $\hat{P}_2$ requires more samples to converge for a higher response threshold, as indicated by the larger values of $N_{\mathrm{IS},2}$ for $h^* = 0.035$m.

We investigate the effect of the number of samples per level during CE optimization

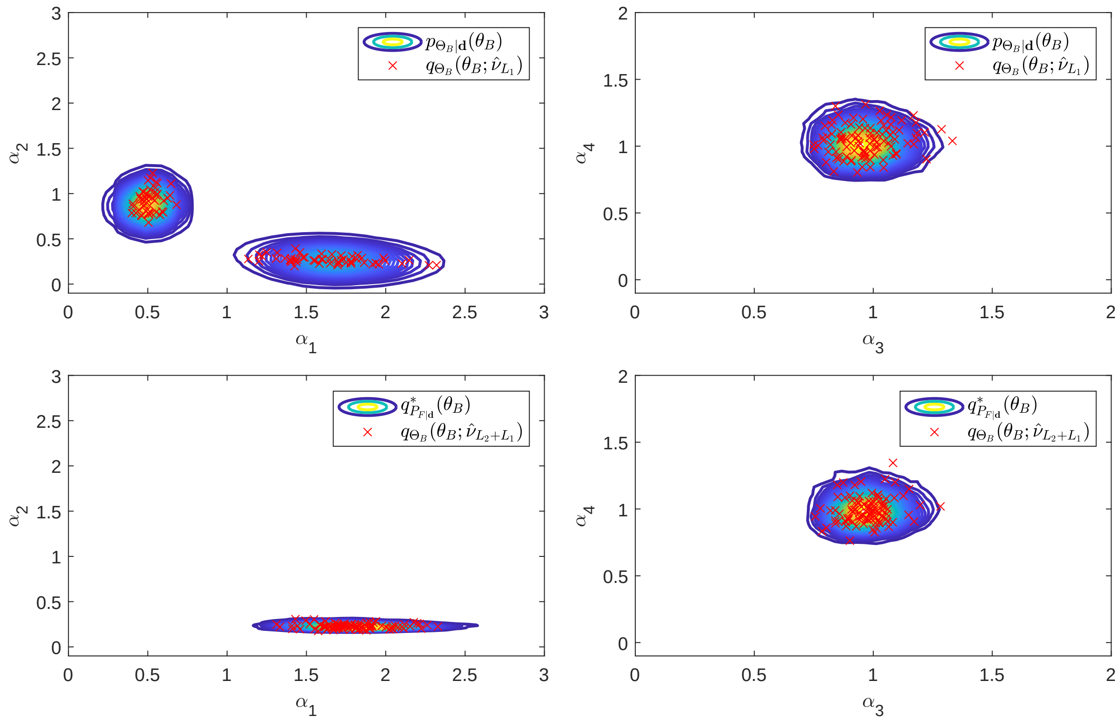

Figure 6: Samples of the stiffness parameters α_1 and α_2 and mass parameters α_3 and α_4. Top: Joint posterior PDF of the parameters. Bottom: Joint optimal IS density over the failure domain. Scattered points denote samples from the parametric densities fitted by the CE method. Solid lines denote contours of the joint PDFs constructed based on samples obtained through rejection sampling.

on the performance of CEIS-RelUp. For different values of N, the sample mean of the posterior first-passage probability estimates are similar to the values in Table 5, and hence are not reported. Fig. 8 shows the variation in the sample CoV of the failure probability estimates and the total computational effort with N. For N_{IS}-NonAdap, it is observed that the sample CoV $\hat{\delta}$ decreases with increasing N. However, once the parameters of the IS density become sufficiently optimal for larger values of N, the rate of decrease reduces. For sufficiently large N, the variation of the parameters of the fitted IS density becomes small and the rate of decrease is proportional to $1/\sqrt{N}$ (cf. Proposition 2). This is because in the non-adaptive variant, N samples are used for estimation. The estimates of the sample CoV with N_{IS}-Adap closely adhere to the prescribed target, except for $N = 250$ where we observe higher variability in the estimates. This is attributed to the sub-optimality in the parameters of the IS density due to the inadequate number of effective samples available for fitting the parametric density with $N = 250$. The total computational effort shows that selecting the sample size of the IS estimators adaptively is more efficient. For $h^* = 0.03$m and $h^* = 0.035$m, N_{IS}-NonAdap requires $N_{\mathrm{T}} = 6870$ and $N_{\mathrm{T}} = 7125$ samples, respectively, to achieve a sample CoV of 5% of the failure probability estimates. With N_{IS}-Adap, the

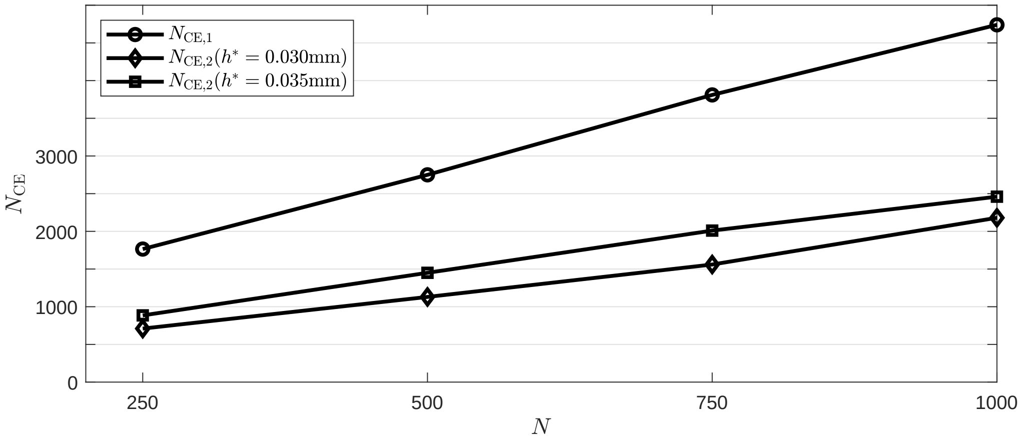

Figure 7: Cross entropy optimization effort for two-story moment-resisting frame

Table 5: Posterior first-passage probability estimates of two-story moment-resisting frame by CEIS-RelUp. CE optimization performed using $N = 500$ samples per level. Reference value of the posterior first-passage probability, based on 5×10^7 samples obtained through rejection sampling, is 1.85×10^{-4} and 4.52×10^{-6} for $h^* = 0.030$m and $h^* = 0.035$m, respectively. Reference value of the marginal likelihood is 1.42×10^{-3}.

		$\hat{c}_E$	$\hat{P}_{F\|\mathbf{d}}$	$\hat{\delta}$	$\hat{\delta}_1$	$\hat{\delta}_2$	$N_{\mathrm{IS},1}$	$N_{\mathrm{IS},2}$	N_{T}
$h^* = 0.030$m	N_{IS}-NonAdap	1.43×10^{-3}	1.81×10^{-4}	0.07	0.04	0.06	500	500	4880
	N_{IS}-Adap ($\delta^* = 0.10$)	1.39×10^{-3}	1.88×10^{-4}	0.10	0.06	0.07	98	155	4133
	N_{IS}-Adap ($\delta^* = 0.05$)	1.42×10^{-3}	1.81×10^{-4}	0.04	0.03	0.03	640	688	5208
$h^* = 0.035$m	N_{IS}-NonAdap	1.43×10^{-3}	4.67×10^{-6}	0.13	0.04	0.11	500	500	5200
	N_{IS}-Adap ($\delta^* = 0.10$)	1.39×10^{-3}	4.56×10^{-6}	0.10	0.06	0.06	98	182	4480
	N_{IS}-Adap ($\delta^* = 0.05$)	1.42×10^{-3}	4.68×10^{-6}	0.05	0.04	0.04	640	806	5646

same is achieved with $N_{\mathrm{T}} = 5208$ samples for $h^* = 0.03$m and $N_{\mathrm{T}} = 5646$ for $h^* = 0.035$m.

7. Concluding remarks

This contribution proposes a novel importance sampling (IS) method to update the failure probability of engineering systems based on data. An effective IS density of the uncertain model parameters is introduced to estimate the marginal likelihood of the data. The IS density is determined by minimizing the cross entropy (CE) between the posterior probability density function (PDF) of the uncertain parameters and a chosen parametric family of probability distributions. The IS density for marginal likelihood estimation leads to a sample-based approximation of the posterior PDF, which is subsequently used as a building block to construct an efficient IS density for estimating the posterior failure probability through a second round of CE minimization. The novel contribution lies in the development of a two-step adaptive multi-level approach to efficiently solve the two CE optimization

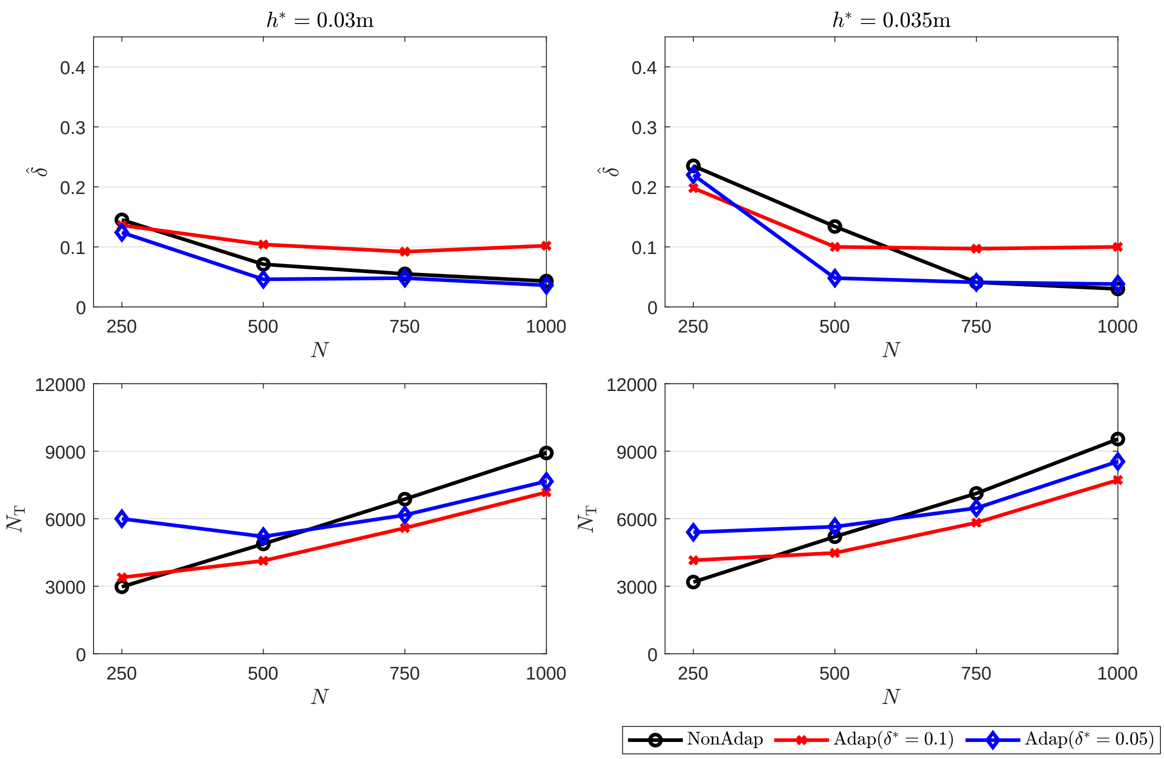

Figure 8: Coefficient of variation of posterior first-passage probability estimates and total computational effort for two-story moment-resisting frame

problems. Numerical studies on a range of engineering problems demonstrate that the proposed method gives accurate estimates of the updated reliability with reasonable total number of samples.

We discuss two approaches to select the sample size of the IS estimator for the posterior probability of failure. In the first approach, the number of samples is fixed to a certain value. The second approach considers selecting the sample size adaptively to ensure that an estimate of the sample CoV of the IS estimator adheres to a specified target. Results from numerical studies demonstrate that the adaptive variant of the estimator is more efficient.

The performance of the CE method depends on the choice of the parametric density. We consider the Gaussian density and Gaussian mixture (GM) as the parametric families, which are able to adequately represent a wide range of posterior distributions. However, the number of distribution parameters to be learnt by CE optimization increases quadratically with the number of uncertain model parameters. In an ongoing work, we explore sparse learning approaches to accelerate the learning and improve the efficiency of the method in high dimensions. In the numerical studies, the number of the terms in the GM model is chosen prior to the simulation. We intend to explore adaptive approaches that estimate the number of GM terms on the fly during CE optimization.

Acknowledgements

This work was supported by the Alexander von Humboldt Foundation.

References

[1] S. K. Au and J. L. Beck. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, 16(4):263–277, 2001.
[2] S. K. Au and J. L. Beck. First excursion probabilities for linear systems by very efficient importance sampling. *Probabilistic Engineering Mechanics*, 16(3):193–207, 2001.
[3] S. K. Au and J. L. Beck. Importance sampling in high dimensions. *Structural Safety*, 25(2):139–163, 2003.
[4] S. Bansal and S. H. Cheung. A new stochastic simulation algorithm for updating robust reliability of linear structural dynamic systems subjected to future Gaussian excitations. *Computer Methods in Applied Mechanics and Engineering*, 326:481–504, 2017.
[5] J. L. Beck and S. K. Au. Bayesian updating of structural models and reliability using Markov chain Monte Carlo simulation. *ASCE Journal of Engineering Mechanics*, 128(4):380–391, 2002.
[6] S. H. Cheung and S. Bansal. A new Gibbs sampling based algorithm for Bayesian model updating with incomplete complex modal data. *Mechanical Systems and Signal Processing*, 92:156–172, 2017.
[7] J. Ching and J. Beck. Real-time reliability estimation for serviceability limit states in structures with uncertain dynamic excitation and incomplete output data. *Probabilistic Engineering Mechanics*, 22(1): 50–62, 2007.
[8] J. Ching and Y. C. Chen. Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging. *ASCE Journal of Engineering Mechanics*, 133(7):816–832, 2007.
[9] J. Ching and Y. H. Hsieh. Updating reliability of instrumented geotechnical systems via simple Monte Carlo simulation. *Journal of GeoEngineering*, 1(2):71–78, 2006.
[10] P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436, 2006.
[11] A. Der Kiureghian and O. Ditlevsen. Aleatory or epistemic? does it matter? *Structural Safety*, 31(2): 105–112, 2009.
[12] A. Der Kiureghian and P. L. Liu. Structural reliability under incomplete probability information. *ASCE Journal of Engineering Mechanics*, 112(1):85–104, 1986.
[13] O. Ditlevsen and H. O. Madsen. *Structural Reliability Methods*, volume 178. Wiley New York, 1996.
[14] M. Ehre, I. Papaioannou, K. E. Willcox, and D. Straub. Conditional reliability analysis in high dimensions based on controlled mixture importance sampling and information reuse. *Computer Methods in Applied Mechanics and Engineering*, 381:113826, 2021.
[15] M. Engel, O. Kanjilal, I. Papaioannou, and D. Straub. Bayesian updating and maximum likelihood estimation by cross entropy-based importance sampling. *Under review, Preprint available at* *http://go.tum.de/808237*, 2021.
[16] M. H. Faber. On the treatment of uncertainties and probabilities in engineering decision analysis. *Journal of Offshore Mechanics and Arctic Engineering*, 127:243–248, 2005.
[17] S. Geyer, I. Papaioannou, and D. Straub. Cross entropy-based importance sampling using Gaussian densities revisited. *Structural Safety*, 76:15–27, 2019.
[18] D. Griffiths, J. Huang, and G. A. Fenton. Probabilistic infinite slope analysis. *Computers and Geotechnics*, 38(4):577–584, 2011.
[19] P. E. Hadjidoukas, P. Angelikopoulos, C. Papadimitriou, and P. Koumoutsakos. π4u: A high performance computing framework for Bayesian uncertainty quantification of complex models. *Journal of Computational Physics*, 284:1–21, 2015.
[20] M. Hohenbichler and R. Rackwitz. Non-normal dependent vectors in structural safety. *ASCE Journal of the Engineering Mechanics Division*, 107(6):1227–1238, 1981.

[21] H. Jensen, C. Vergara, C. Papadimitriou, and E. Millas. The use of updated robust reliability measures in stochastic dynamical systems. *Computer Methods in Applied Mechanics and Engineering*, 267:293–317, 2013.
[22] S. H. Jiang and J. Huang. Modeling of non-stationary random field of undrained shear strength of soil for slope reliability analysis. *Soils and Foundations*, 58(1):185–198, 2018.
[23] S. H. Jiang, I. Papaioannou, and D. Straub. Bayesian updating of slope reliability in spatially variable soils with in-situ measurements. *Engineering Geology*, 239:310–320, 2018.
[24] O. Kanjilal, I. Papaioannou, and D. Straub. Cross entropy-based importance sampling for first-passage probability estimation of randomly excited linear structures with parameter uncertainty. *Structural Safety*, 91:102090, 2021.
[25] O. Kanjilal, I. Papaioannou, and D. Straub. Series system reliability of uncertain linear structures under Gaussian excitation by cross entropy-based importance sampling. *ASCE Journal of Engineering Mechanics*, 148(1):04021136, 2021.
[26] L. S. Katafygiotis and K. M. Zuev. Geometric insight into the challenges of solving high-dimensional reliability problems. *Probabilistic Engineering Mechanics*, 23(2-3):208–218, 2008.
[27] N. Kurtz and J. Song. Cross entropy-based adaptive importance sampling using Gaussian mixture. *Structural Safety*, 42:35–44, 2013.
[28] M. Lemaire. *Structural Reliability*. John Wiley & Sons, 2013.
[29] M. Macke and C. Bucher. Importance sampling for randomly excited dynamical systems. *Journal of Sound and Vibration*, 268(2):269–290, 2003.
[30] H. Madsen. Model updating in reliability theory. *ICASP 5, Int. Conf. on Appl. of Statistics and Prob. in Soil and Struct*, pages 565–577, 1987.
[31] R. M. Neal. Annealed importance sampling. *Statistics and Computing*, 11(2):125–139, 2001.
[32] C. Papadimitriou, J. L. Beck, and L. S. Katafygiotis. Updating robust reliability using structural test data. *Probabilistic Engineering Mechanics*, 16(2):103–113, 2001.
[33] I. Papaioannou and D. Straub. Reliability updating in geotechnical engineering including spatial variability of soil. *Computers and Geotechnics*, 42:44–51, 2012.
[34] I. Papaioannou, C. Papadimitriou, and D. Straub. Sequential importance sampling for structural reliability analysis. *Structural Safety*, 62:66–75, 2016.
[35] I. Papaioannou, S. Geyer, and D. Straub. Improved cross entropy-based importance sampling with a flexible mixture model. *Reliability Engineering & System Safety*, 191:106564, 2019.
[36] R. Y. Rubinstein and D. P. Kroese. *Simulation and the Monte Carlo method*, volume 10. John Wiley & Sons, 2016.
[37] G. Schuëller, H. Pradlwarter, and P. S. Koutsourelakis. A critical appraisal of reliability estimation procedures for high dimensions. *Probabilistic Engineering Mechanics*, 19(4):463–474, 2004.
[38] G. I. Schuëller and H. J. Pradlwarter. Benchmark study on reliability estimation in higher dimensions of structural systems–an overview. *Structural Safety*, 29(3):167–182, 2007.
[39] D. Straub. Reliability updating with equality information. *Probabilistic Engineering Mechanics*, 26(2): 254–258, 2011.
[40] D. Straub, I. Papaioannou, and W. Betz. Bayesian analysis of rare events. *Journal of Computational Physics*, 314:538–556, 2016.
[41] V. Sundar and C. Manohar. Time variant reliability model updating in instrumented dynamical systems based on Girsanov's transformation. *International Journal of Non-Linear Mechanics*, 52:32–40, 2013.
[42] M. W. Vanik, J. L. Beck, and S. K. Au. Bayesian probabilistic approach to structural health monitoring. *Journal of Engineering Mechanics*, 126(7):738–745, 2000.
[43] Z. Wang and J. Song. Cross entropy-based adaptive importance sampling using von Mises-Fisher mixture for high dimensional reliability analysis. *Structural Safety*, 59:42–52, 2016.