Conditional reliability analysis in high dimensions based on controlled mixture importance sampling and information reuse

Max Ehre[a,*], Iason Papaioannou[a], Karen E. Willcox[b], Daniel Straub[a]

[a] *Technische Universität München, Arcisstr. 21, 80290 München, Germany*
[b] *University of Texas at Austin, Austin, TX, 78712, USA*

Abstract

In many contexts, it is of interest to assess the impact of selected parameters on the failure probability of a physical system. To this end, one can perform conditional reliability analysis, in which the probability of failure becomes a function of these parameters. Computing conditional reliability requires recomputing failure probabilities for a sample sequence of the parameters, which strongly increases the already high computational cost of conventional reliability analysis. We alleviate these costs by reusing information from previous reliability computations in each subsequent reliability analysis of the sequence. The method is designed using two variants of importance sampling and performs information transfer by reusing importance densities from previous reliability analyses in the current one. We put forward a criterion for selecting the most informative importance densities, which is robust with respect to the input space dimension, and use a recently proposed density mixture model for constructing effective importance densities in high dimensions. The method controls the estimator coefficient of variation to achieve a prescribed accuracy. We demonstrate its performance by means of two engineering examples featuring a number of pitfall features such as strong non-linearity, high dimensionality and very small failure probabilities (10^{-5} to 10^{-9}).

1. Introduction

In order to accurately predict model behaviour with confidence, it is vital to account for uncertainties influencing the model and its output. Reliability analysis is concerned with quantifying the extremal behaviour of a model under uncertainty by computing its probability of failure, i.e., the probability of an unacceptable model response. Often it is of interest to repeatedly perform the analysis on a series of parametrised reliability problems. Such situations arise in reliability-based design optimization (RBDO), where the parametrisation is given by the design parameters, or whenever it is desirable to separate the model inputs into two categories. Hereafter, we refer to these categories as type A and type B. One example of such a separation is in reducible (epistemic) and irreducible (aleatory) uncertainty [1, 2, 3, 4, 5]. The general goal of separating inputs in this way is to establish a distinct relationship between type B variables and the probability of failure conditional on type B variables. By conditioning the probability of failure on type B variables, one obtains a measure for the influence of these variables on the probability of failure. In general, the concept applies to any target that can be cast in terms of an expected value. Such a formulation is useful in many contexts — it may, e.g., be used to provide estimates of credibility bounds, dispersion measures or the distribution of the probability of failure conditional on type B variables thus quantifying lack of knowledge/confidence caused by these variables. It gives rise to global sensitivity measures of the conditional failure probability with respect to the type B variables (e.g., Sobol' indices [6]) and facilitates the computation of the partial value of (im-)perfect information for eliciting optimal decisions based on conditional failure probabilities [7]. We have demonstrated how to construct surrogate models mapping the type B-variables to the conditional probability of failure to obtain global sensitivity measures at significantly reduced computational cost [6].

*Corresponding author
Email address: max.ehre@tum.de (Max Ehre)

Conventional reliability analysis is a challenging task as failure probabilities are typically associated with rare events [8, 9] and thus assume very small values. Simultaneously, it has received considerable attention due to its relevance to engineering and financial applications. Structural reliability methods (SRM) can be categorized into approximation-based methods, such as the first- (FORM) and second-order reliability method (SORM) [10] and sampling-based methods (importance sampling [11, 12, 13], sequential importance sampling [14, 15], subset simulation [16], cross-entropy importance sampling [17, 18, 19], line-sampling [20, 21], multi-level Monte Carlo (MC) [22] and multi-fidelity MC [23]). Conditional reliability analysis is considerably more expensive compared to its conventional counterpart as it requires the solution of a sequence of reliability problems rather than a single one. A number of sampling approaches have been developed for computing the probability of multiple correlated failure events efficiently. Ref. [24] introduces parallel subset simulation to estimate failure probabilities of several failure events simultaneously by defining a principle variable that is correlated with each failure event. In Refs. [25, 26] subset simulation is applied to a parallel system with each system component representing one reliability problem in the sequence. The corresponding failure probabilities can be estimated based on the failure probability of the system and the conditional samples from each subset. Refs. [24, 25, 26] are efficient if all considered failure events are strongly correlated, but will encounter difficulties if there are failure events occurring in the sequence that are not correlated with any of the other events. This implies that these methods are not suited for conditional reliability analysis if the number of type B-samples is large and/or if the type B-samples contribute a large fraction of variability to the conditional probability of failure, since in either case, the probability of disjunct failure events in the sequence is considerable. Ref. [27] proposes a method for robust optimization problems, i.e., design optimization under probabilistic constraints that include the mean and variance of the model response. Control variates are used to recycle information stemming from the previous optimizer iteration to accelerate the MC constraint computation in each optimization step (but the first). The design parameters, which parametrize the model and which change in each design iteration, can be viewed as a deterministic counterpart to type B-variables (see Section 3.2). The same is true for RBDO, where the design optimization is carried out under constraints on system reliability rather than response moments. Ref. [14] proposes bridge importance sampling to solve RBDO: Importance densities of reliability computations at previous steps in the design optimization are used to initialize a bridging step towards the current optimal importance density. Suitable density candidates are identified based on a heuristic that has inspired an earlier approach to conditional reliability analysis [28] as well as this contribution (Section 3.3). Ref. [29] proposes to solve RBDO with importance sampling where information from previous design iterations is incorporated in the choice of the importance density of the current reliability problem.

Here, we propose a method for solving the conditional reliability problem in high dimensions efficiently through information reuse. In Section 2, we briefly recap conventional reliability analysis and popular solution approaches before formally introducing the conditional reliability problem in Section 3. We then discuss our approach to information reuse that consists of a selection strategy for informative importance densities and two importance samplers (mixed and controlled) that serve to exploit the selected densities in the current reliability estimate. In Section 4, we present comprehensive investigations of the method's performance in two engineering examples and provide a detailed discussion of the results. Conclusions are given in Section 5.

2. Conventional reliability analysis

In this section, we set up the reliability problem formulation and discuss well-etablished approaches to its solution. In the second part, we discuss the cross-entropy method (CE) [17, 30] and a recently introduced improved version thereof (iCE) [31].

2.1. Problem Statement

Consider a system that is modelled by $\mathcal{Y} : \mathbb{D}_{\boldsymbol{\Theta}} \to \mathbb{R}$ with d-dimensional random input vector $\boldsymbol{\Theta} : \Omega \to \mathbb{D}_{\boldsymbol{\Theta}} \subseteq \mathbb{R}^d$, where Ω is the sample space of $\boldsymbol{\Theta}$. $F_{\boldsymbol{\Theta}}$ is the joint cumulative distribution function (CDF) of $\boldsymbol{\Theta}$. $\mathcal{Y}$ maps to the system response $Y = \mathcal{Y}(\boldsymbol{\theta})$ with the model input $\boldsymbol{\theta} \in \mathbb{D}_{\boldsymbol{\Theta}}$. Based on the response Y,

unacceptable system states are defined by means of the limit-state function LSF $\tilde{g}(Y)$. Defining $g(\boldsymbol{\theta}) = \tilde{g} \circ \mathcal{Y}(\boldsymbol{\theta})$ and introducing the convention

$$g(\boldsymbol{\theta}) = \begin{cases} \leq 0, \text{Failure} \\ > 0, \text{Safety}, \end{cases}$$

the failure event of the system is defined as $\mathrm{F} = \{\boldsymbol{\theta} \in \mathbb{D}_{\boldsymbol{\Theta}} : g(\boldsymbol{\theta}) \leq 0\}$. The probability of failure is given by [8]

$$P = \mathbb{P}(\mathrm{F}) = \int_{\mathbb{D}_{\boldsymbol{\Theta}}} \mathrm{I}[g(\boldsymbol{\theta}) \leq 0] f(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} = \mathbb{E}\left[\mathrm{I}(g(\boldsymbol{\Theta}) \leq 0)\right], \tag{1}$$

where f is the joint probability density function (PDF) of $\boldsymbol{\Theta}$ and the indicator function $\mathrm{I}[\cdot]$ equals 1 if true and 0 otherwise. Without loss of generality, one may formulate an equivalent reliability problem with respect to the standard-normal probability space using the random vector $\boldsymbol{U} : \Omega \to \mathbb{R}^d$. Given an isoprobabilistic transformation $T : \mathbb{D}_{\boldsymbol{\Theta}} \to \mathbb{R}^d$, such that $\boldsymbol{U} = T(\boldsymbol{\Theta})$ [32] and defining $G = g(T^{-1}(\boldsymbol{U}))$, one can write Eq. (1) as

$$P = \int_{\mathbb{R}^d} \mathrm{I}[G(\boldsymbol{u}) \leq 0] \varphi_d(\boldsymbol{u}) \mathrm{d}\boldsymbol{u} = \mathbb{E}\left[\mathrm{I}(G(\boldsymbol{U}) \leq 0)\right], \tag{2}$$

where φ_d denotes the d-dimensional independent standard-normal PDF.

2.2. Standard MC

The standard MC estimate of integral (2) reads

$$\widehat{p}_{\mathrm{MC}} = \frac{1}{n} \sum_{k=1}^{n} \mathrm{I}[G(\boldsymbol{u}^k) \leq 0], \quad \boldsymbol{u}^k \overset{i.i.d.}{\sim} \varphi_d.$$

This estimate is unbiased and has coefficient of variation (CoV)

$$\delta_{\mathrm{MC}} = \sqrt{\frac{1-P}{nP}}.$$

Its costs in terms of g-evaluations ($= n$) are independent of the model dimension d. If $P \ll 1$, δ_{MC} scales approximately inversely with the square root of the failure probability and n grows very large for small values of P. Namely, for a target δ_0, at least n_0 evaluations of G are required, where

$$n_0 = \frac{1-P}{\delta_0^2 P}.$$

Thus, while independent of the model input dimension, the standard MC estimate is not suited for estimating rare events if evaluating $\mathcal{Y}$ is not extremely cheap.

2.3. Importance sampling

One of the most commonly used techniques to alleviate the above restriction on the sample size while achieving a prescribed CoV is the importance sampling (IS) method. Let h be a density, such that $h(\boldsymbol{u}) > 0$ whenever $\mathrm{I}(G(\boldsymbol{u}) \leq 0)\varphi_d(\boldsymbol{u}) \neq 0$. Then, we can rewrite integral (1)

$$P = \int_{\mathbb{R}^d} \mathrm{I}(G(\boldsymbol{u}) \leq 0) \overbrace{\frac{\varphi_d(\boldsymbol{u})}{h(\boldsymbol{u})}}^{w(\boldsymbol{u})} h(\boldsymbol{u}) \mathrm{d}\boldsymbol{u} = \mathbb{E}_h\left[\mathrm{I}(G(\boldsymbol{U}) \leq 0) w(\boldsymbol{U})\right], \tag{3}$$

where h is termed the *importance*, *auxiliary*, *instrumental* or *biasing* density and w is the *likelihood ratio* or *IS weight*. In the context of importance sampling, φ_d is often referred to as the *nominal* density. The corresponding estimate of the probability of failure is given by

$$\widehat{p}_{\mathrm{IS}} = \frac{1}{n} \sum_{k=1}^{n} \mathrm{I}[G(\boldsymbol{u}^k) \leq 0] w(\boldsymbol{u}^k), \quad \boldsymbol{u}^k \overset{i.i.d.}{\sim} h. \tag{4}$$

$\widehat{p}_{\mathrm{IS}}$ is an unbiased estimate of P and its variance is

$$\mathbb{V}[\widehat{p}_{\mathrm{IS}}] = \frac{1}{n}\mathbb{V}_h[\mathrm{I}[G(\boldsymbol{U}) \leq 0]w(\boldsymbol{U})].$$

Estimating the above based on a set of samples drawn from h, we obtain an estimate for the CoV of $\widehat{p}_{\mathrm{IS}}$ as

$$\widehat{\delta}_{\mathrm{IS}} = \frac{1}{\widehat{p}_{\mathrm{IS}}}\sqrt{\frac{1}{n(n-1)}\sum_{k=1}^{n}(\mathrm{I}[G(\boldsymbol{u}^k) \leq 0]w(\boldsymbol{u}^k) - \widehat{p}_{\mathrm{IS}})^2}, \quad \boldsymbol{u}^k \overset{i.i.d.}{\sim} h. \tag{5}$$

There exists an optimal importance density h^* such that $\mathbb{V}[\widehat{p}_{\mathrm{IS}}] = 0$:

$$h^*(\boldsymbol{u}) = \frac{1}{P}\mathrm{I}[G(\boldsymbol{u}) \leq 0]\varphi_d(\boldsymbol{u}). \tag{6}$$

Note, that h^* requires knowledge of the target quantity P. Thus it cannot be used immediately to compute integral (3). However, it gives rise to a variety of approaches that aim at approximating h^* by propagating a sequence of distributions from φ_d towards h^*, e.g., via conditional sampling using Markov Chain Monte Carlo [12, 15] or through fitting parametric density models [11, 17, 30].

2.4. The iCE method (iCE)

Here, we discuss a recently proposed version of sequential importance sampling that is based on the classical CE method proposed in Ref. [30] and has been demonstrated to work well in high dimensions [31]. Consider a parametric version of the importance density $h(\boldsymbol{u}, \boldsymbol{v})$, which is defined by the parameter vector $\boldsymbol{v} \in \mathcal{V}$. The parameter space $\mathcal{V}$ contains $\boldsymbol{v}_0$, where $h(\boldsymbol{u}, \boldsymbol{v}_0) = \varphi_d(\boldsymbol{u})$. The standard CE method aims at minimizing the Kullback-Leibler (KL) divergence $D_{KL}(h^*(\boldsymbol{u})||h(\boldsymbol{u}, \boldsymbol{v}))$ between $h^*(\boldsymbol{u})$ and $h(\boldsymbol{u}, \boldsymbol{v})$ over the parameter space $\mathcal{V}$, which is defined as [17]

$$\begin{aligned} D_{KL}(h^*(\boldsymbol{u})||h(\boldsymbol{u}, \boldsymbol{v})) &= \mathbb{E}_{h^*}\left[\ln\left(\frac{h^*(\boldsymbol{u})}{h(\boldsymbol{u}, \boldsymbol{v})}\right)\right] \\ &\overset{(6)}{=} \frac{1}{P}\mathbb{E}_{\varphi_d}[\mathrm{I}[G(\boldsymbol{u}) \leq 0]\ln(h^*(\boldsymbol{u}))] - \frac{1}{P}\mathbb{E}_{\varphi_d}[\mathrm{I}[G(\boldsymbol{u}) \leq 0]h(\boldsymbol{u}, \boldsymbol{v})]. \end{aligned} \tag{7}$$

The first summand on the right-hand side of Eq. (7) is not a function of $\boldsymbol{v}$, so that minimizing $D_{KL}(h^*(\boldsymbol{u})||h(\boldsymbol{u}, \boldsymbol{v}))$ can be expressed as

$$\boldsymbol{v}^* = \arg\max_{\boldsymbol{v}\in\mathcal{V}} \mathbb{E}_{\varphi_d}[\mathrm{I}[G(\boldsymbol{U}) \leq 0]\ln(h(\boldsymbol{U}, \boldsymbol{v}))] \tag{8}$$

and its sample-based approximation reads

$$\widehat{\boldsymbol{v}}^* = \arg\max_{\boldsymbol{v}\in\mathcal{V}} \frac{1}{n}\sum_{k=1}^{n}[\mathrm{I}[G(\boldsymbol{u}^k) \leq 0]\ln(h(\boldsymbol{u}^k, \boldsymbol{v}))], \quad \boldsymbol{u}^k \overset{i.i.d.}{\sim} \varphi_d. \tag{9}$$

For fixed $\boldsymbol{v}$, the objective function in program (9) is equivalent to a weighted version of $\widehat{p}_{MC}$. That is, to approximate $\boldsymbol{v}^*$ well with $\widehat{\boldsymbol{v}}^*$, n has to be large if F is a rare event. The CE method circumvents this problem by approaching h^* stepwise with a sequence of parametric distributions defined by $\{\boldsymbol{v}_i, i = 1, \ldots, m\}$. The failure event F is represented by a series of more probable intermediate events $\{F_i, i = 1, \ldots, m\}$ that are defined by manipulating their associated threshold ξ_i s.t. $F_i = \{\boldsymbol{u} \in \mathbb{R}^d : G(\boldsymbol{u}) \leq \xi_i\}$, where $\xi_1 > \xi_2 > \cdots > \xi_{m-1} > \xi_m$. Starting from $h(\boldsymbol{u}, \boldsymbol{v}_0) = \varphi_d(\boldsymbol{u})$, the threshold ξ_i is determined as the lower ρ-quantile of the LSF based on samples from the parametric density associated with F_{i-1}, $h(\boldsymbol{u}, \boldsymbol{v}_{i-1})$, with typical choices for the quantile value $\rho = [10^{-2}, 10^{-1}]$. The i-th parametric density is then found through minimizing the KL divergence between $h_i(\boldsymbol{u})$ and $h(\boldsymbol{u}, \boldsymbol{v}_i)$, where $h_i(\boldsymbol{u})$ is the optimal importance sampling density associated with the threshold ξ_i. Once all $n\rho$ new samples lie within the failure domain, i.e., G evaluated at these samples is always negative, the algorithm is stopped. Solving program (9) based on samples from $h(\boldsymbol{u}, \widehat{\boldsymbol{v}}_{i-1})$ rather than φ_d introduces the weight $W(\boldsymbol{u}, \widehat{\boldsymbol{v}}_{i-1}) = \frac{\varphi_d(\boldsymbol{u})}{h(\boldsymbol{u}, \widehat{\boldsymbol{v}}_{i-1})}$:

$$\widehat{\boldsymbol{v}}_i = \arg\max_{\boldsymbol{v}\in\mathcal{V}} \frac{1}{n}\sum_{k=1}^{n}\mathrm{I}[G(\boldsymbol{u}^k) \leq \xi_i]\ln(h(\boldsymbol{u}^k, \boldsymbol{v}))W(\boldsymbol{u}^k, \widehat{\boldsymbol{v}}_{i-1}), \quad \boldsymbol{u}^k \overset{i.i.d.}{\sim} h(\boldsymbol{u}, \widehat{\boldsymbol{v}}_{i-1}). \tag{10}$$

Computing the new parameter set $\widehat{\boldsymbol{v}}_i$ based on $n\rho$ samples in each iteration effectively leaves a fraction of $1-\rho$ $(90-99\%)$ of the samples unused and motivates the first of two major points of departure of iCE from CE: Within iCE, h_i is re-parametrized using a smooth approximation of $\mathrm{I}[g \leq 0]$ based on the standard-normal CDF $\Phi(\cdot)$:

$$h_i(\boldsymbol{u}) = \frac{1}{P_i}\Phi\left(-\frac{G(\boldsymbol{u})}{\sigma_i}\right)\varphi_d(\boldsymbol{u}) = \frac{1}{P_i}\eta_i(\boldsymbol{u}), \tag{11}$$

where $P_i = \mathbb{E}_{\varphi_d}[\Phi(-G(\boldsymbol{U})/\sigma_i)]$ is a normalizing constant and σ_i is a smoothing parameter. This distribution sequence has been used to construct adaptive importance sampling-based approaches to reliability analysis [15, 33] , reliability sensitivity analysis [34] and RBDO [14]. The CE-program now reads

$$\widehat{\boldsymbol{v}}_i = \underset{\boldsymbol{v}\in\mathcal{V}}{\arg\max}\, \frac{1}{n}\sum_{k=1}^{n} \ln(h(\boldsymbol{u}^k, \boldsymbol{v}))W(\boldsymbol{u}^k, \widehat{\boldsymbol{v}}_{i-1}), \quad \boldsymbol{u}^k \overset{i.i.d.}{\sim} h(\boldsymbol{u}, \widehat{\boldsymbol{v}}_{i-1}), \tag{12}$$

where $W(\boldsymbol{u}, \widehat{\boldsymbol{v}}_{i-1}) = \frac{\eta_i(\boldsymbol{u})}{h(\boldsymbol{u}, \widehat{\boldsymbol{v}}_{i-1})}$. In program (12), all samples available from $h(\boldsymbol{u}, \widehat{\boldsymbol{v}}_{i-1})$ will be used with their respective modified weight. Then, in each step, the current σ_t is identified such that the sample CoV of the weights $\{W(\boldsymbol{u}^k, \widehat{\boldsymbol{v}}_{i-1}), k = 1, \ldots, n\}$ adheres to a target value δ_{target}:

$$\sigma_i = \underset{\sigma\in[0,\sigma_{i-1}]}{\arg\min} \left(\widehat{\delta}_W(\sigma) - \delta_{\text{target}}\right)^2. \tag{13}$$

The algorithm terminates, when the approximated and optimal importance density, h_i and h^*, are sufficiently close in the sense that $\mathrm{CoV}[\mathrm{I}[G(\boldsymbol{u})]/\Phi(-g(\boldsymbol{u})/\sigma_i)] \leq \delta_{\text{target}}$, with $\delta_{\text{target}} = 1.5$ a typical choice [31]. Note, that δ_{target} is computed with $h_i(\boldsymbol{u})$ rather than the parametric $h(\boldsymbol{u}, \widehat{\boldsymbol{v}}_i)$, since it is more robust with respect to the flexibility of the parametric model. After termination, additional samples can be drawn from the final parametric importance density to achieve a prescribed estimator CoV according to Eq. (5). Algorithm 1 describes the iCE procedure in detail.

The second point of departure of iCE from CE is given by the parametric density model choice. When working in standard-normal space, typical choices for $h(\boldsymbol{u}, \boldsymbol{v})$ are the d-dimensional single Gaussian density (SG) [35] or a d-dimensional Gaussian mixture (GM) [19, 35]. For the single Gaussian, programs (10) & (12) can be solved analytically and for the Gaussian mixture, the solution is identified through the expectation maximization (EM) algorithm. However, within importance sampling algorithms, both models perform poorly in higher-dimensional problems ($d \geq 20$). A detailed discussion of the issue can be found in Refs. [35, 36, 37]. In Ref. [38], a von Mises-Fisher model for the direction in $\boldsymbol{U}$-space is proposed to remedy these issues. For the iCE method, this model is extended by a Nakagami distribution for the radius of any point in standard-normal space, which yields the von Mises-Fisher-Nakagami-mixture model (vMFNM) for $h(\boldsymbol{u}, \boldsymbol{v})$, where $\boldsymbol{v} = [\boldsymbol{m}, \boldsymbol{\Omega}, \boldsymbol{\mu}, \boldsymbol{\kappa}, \boldsymbol{\alpha}]$ [31]. $\boldsymbol{\alpha} \in \mathbb{R}^K$ are the mixture weights of the K components, $\boldsymbol{m} \in \mathbb{R}^K$ and $\boldsymbol{\Omega} \in \mathbb{R}^K$ are the shape and spread parameters of the K Nakagami distributions and $\boldsymbol{\kappa} \in \mathbb{R}^K$ and $\boldsymbol{\mu} \in \mathbb{R}^{K\times d}$ are the concentration and mean direction parameters of the K von Mises-Fisher distributions. Program (12) can be solved through a weighted expectation-maximization algorithm. The number of components in the mixture K can be either prescribed through prior knowledge of the reliability problem (e.g., knowledge of the number of disjunct failure regions) or - in moderate dimensions - identified through a clustering algorithm such as DBSCAN [39]. For details, see [31, 35].

Algorithm 1 The iCE method

Input LSF $G(\boldsymbol{u})$, input space dimension d, target CoV δ_{target}, samples per level N

Output estimate $\widehat{p}_{\text{iCE}}$, compute estimate CoV $\widehat{\delta}_{\text{iCE}}$, no. of levels m, importance densities $\{h(\boldsymbol{u}, \widehat{\boldsymbol{v}}_i), i = 1, \ldots, m\}$

procedure iCE(g, δ_{target}, N, d)
  Set converged $= false$
  Set $i = 1$
  Select $\widehat{\boldsymbol{v}}_0$ ▷ e.g., s.t. $h(\boldsymbol{u}, \widehat{\boldsymbol{v}}_0) = \varphi_d(\boldsymbol{u})$
  while ¬converged **do**
    Sample $\mathbf{U} = \{\boldsymbol{u}^k, k = 1, \cdots, N\} \in \mathbb{R}^{N \times d}$ ▷ $\boldsymbol{u}^k \overset{i.i.d.}{\sim} h(\boldsymbol{u}, \widehat{\boldsymbol{v}}_{i-1})$
    $\mathbf{G} = G(\mathbf{U}) \in \mathbb{R}^{N \times 1}$
    if $\text{CoV}[\text{I}[\mathbf{G}]/\Phi(-\mathbf{G}/\sigma_i)] \leq \delta_{\text{target}}$ **then** ▷ CoV of likelihood ratio of h_* and $\eta_i(\boldsymbol{u})$
      Set $m = i - 1$
      $\boldsymbol{W} = \text{I}[\mathbf{G}^k \leq 0]\varphi_d(\boldsymbol{u}^k)/h(\boldsymbol{u}^k, \widehat{\boldsymbol{v}}_m)$ ▷ Likelihood ratio of h_* and $h(\boldsymbol{u}, \widehat{\boldsymbol{v}}_m)$
      Estimate the failure probability

$$\widehat{p}_{\text{iCE}} = \widehat{\mathbb{E}}(\boldsymbol{W})$$

      Compute the failure probability estimate's CoV

$$\widehat{\delta}_{\text{iCE}} = \sqrt{\frac{\widehat{\mathbb{V}}(\boldsymbol{W})}{N\widehat{\mathbb{E}}(\boldsymbol{W})^2}}$$

      Set converged $= true$
    else
      Compute σ_i from Eq. (13)
      Compute $\widehat{\boldsymbol{v}}_i$ from Eq. (12)
      $i = i + 1$
  return $\widehat{p}_{\text{iCE}}$, $\widehat{\delta}_{\text{iCE}}$, m, $\{h(\boldsymbol{u}, \widehat{\boldsymbol{v}}_i), i = 1, \ldots, m\}$

3. Conditional reliability analysis

3.1. Problem Statement

The interest is in computing the failure probability conditional on a d_B-dimensional subset of the input random vector $\boldsymbol{\Theta}$. Let this subset be $\boldsymbol{\Theta}_B : \Omega \to \mathbb{D}_B \subseteq \mathbb{R}^{d_B}$ with joint CDF F_B. Further let $\boldsymbol{\Theta}_A : \Omega \to \mathbb{D}_A \subseteq \mathbb{R}^{d_A}$ with joint CDF F_A be the complement of $\boldsymbol{\Theta}_B$ over $\boldsymbol{\Theta}$ such that we may reorder the inputs and write $\boldsymbol{\Theta} = [\boldsymbol{\Theta}_A, \boldsymbol{\Theta}_B]^T$. The failure probability conditional on $\boldsymbol{\Theta}_B$ is defined by the integral

$$P_{\mathrm{F}}(\boldsymbol{\Theta}_B) = \mathbb{P}(\mathrm{F}|\boldsymbol{\Theta}_B) = \int_{\mathbb{D}_A} \mathrm{I}[g(\boldsymbol{\theta}_A; \boldsymbol{\Theta}_B) \leq 0] f(\boldsymbol{\theta}_A|\boldsymbol{\Theta}_B) \mathrm{d}\boldsymbol{\theta}_A = \mathbb{E}_{f(\boldsymbol{\theta}_A|\boldsymbol{\Theta}_B)} \left[\mathrm{I}(g(\boldsymbol{\Theta}) \leq 0)|\boldsymbol{\Theta}_B\right]. \tag{14}$$

Using the isoprobabilistic transformation T from Section 2.1 we recast Eq. (14) in standard-normal space:

$$P_{\mathrm{F}}(\boldsymbol{U}_B) = \mathbb{P}(\mathrm{F}|\boldsymbol{U}_B) = \int_{\mathbb{R}^{d_A}} \mathrm{I}[G(\boldsymbol{u}_A; \boldsymbol{U}_B) \leq 0] \varphi_{d_A}(\boldsymbol{u}_A) \mathrm{d}\boldsymbol{u}_A = \mathbb{E}\left[\mathrm{I}(G(\boldsymbol{U}) \leq 0)|\boldsymbol{U}_B\right]. \tag{15}$$

Note that $\boldsymbol{U}_A$ and $\boldsymbol{U}_B$ are independent and thus we have $f(\boldsymbol{u}_A|\boldsymbol{u}_B) = f(\boldsymbol{u}_A) = \varphi_{d_A}(\boldsymbol{u}_A)$.

As mentioned before, possible applications include quantiles $\underline{P}_{\mathrm{F}}/\bar{P}_{\mathrm{F}}$, surrogate models $\widehat{P}_{\mathrm{F}}(\boldsymbol{\Theta}_B)$ or a density estimate $\widehat{f}(p_{\mathrm{F}})$ of $P_{\mathrm{F}}(\boldsymbol{\Theta}_B)$. In practice, these quantities are computed based on n_B samples of $P_{\mathrm{F}}(\boldsymbol{\Theta}_B)$, $\{p^j, j = 1, \ldots, n_B\}$, where

$$p^j = P_{\mathrm{F}}(\boldsymbol{u}_B^j) = \int_{\mathbb{R}^{d_A}} \mathrm{I}(G(\boldsymbol{u}_A, \boldsymbol{u}_B^j) \leq 0) \varphi_{d_A}(\boldsymbol{u}_A) \mathrm{d}\boldsymbol{u}_A, \quad \boldsymbol{u}_B^j \overset{i.i.d.}{\sim} \varphi_{d_B}, \quad j = 1, \ldots, n_B. \tag{16}$$

The computational cost associated with this setting can be considerably higher than that of a conventional reliability analysis as n_B d_A-dimensional reliability problems of form (14) have to be solved instead of a single one. Figure 1 illustrates the estimation of conditional failure probabilities with iCE and the associ-

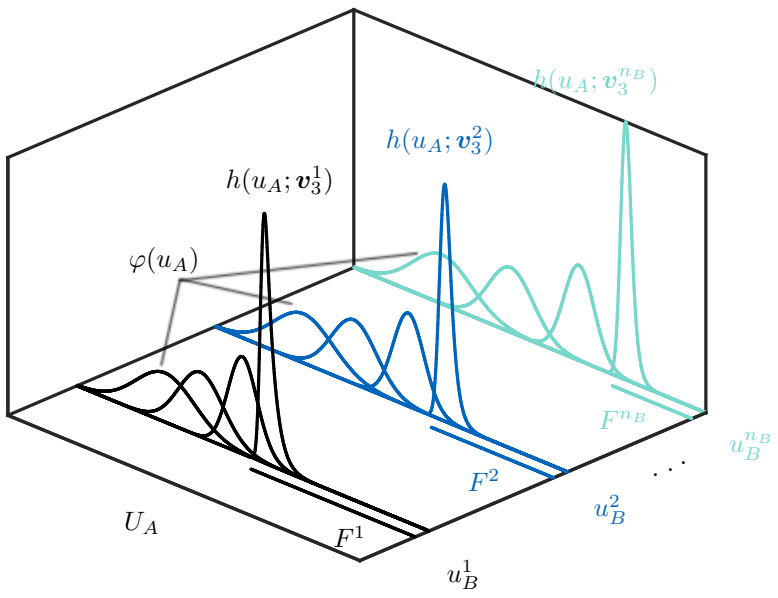

Figure 1: Illustration of iCE densities: $h(u_A; \boldsymbol{v}_i^j)$ is the importance density used at the i-th iCE step in the j-th conditional reliability problem in standard-normal space.

ated parametric importance densities constructed in the process. Therein, $\boldsymbol{v}_i^j$ is the parameter vector of the parametric importance density constructed in the i-th step of the j-th iCE run (i.e., the run solving the j-th

reliability problem). The parameter vector $\boldsymbol{v}$ now receives a superscript to identify the reliability problem to which it belongs while its subscript indicates the associated iCE step.

The main contribution of this paper is an algorithm that efficiently solves the sequence of conditional reliability problems given in Eq. (16). The connection of this problem sequence to a conceptually similar one arising in RBDO is discussed in the following subsection. Our framework is based on the iCE method for conventional reliability analysis. The basic idea is to reuse information from - or more precisely: densities constructed in - past problems to alleviate the computational cost in the current estimation. The two key tasks of such an algorithm are the identification of suitable biasing densities amongst solved reliability computations on the one hand, and the efficient integration of these densities in the estimation of the current conditional failure probability on the other hand. Identification and integration are addressed in subsections 3.3 and 3.4, respectively.

3.2. Connection to RBDO

Reliability-based design optimization (RBDO) may be defined as the minimization of a deterministic cost function under constraints on the failure probability given the system design. To this end, the design is parametrized by means of a set of design variables. Then, a problem similar to Eq. (16) arises as the failure probability of the system has to be evaluated repeatedly and conditional on several points in the design space. In such case, $\boldsymbol{\theta}_B^j$ would represent the design variables' values in the j-th iteration of the RBDO program. Due to this similarity, information reuse is also interesting for solving RBDO problems and has been put forward in this context in Ref. [29]. There, an influence hypersphere around each $\boldsymbol{\Theta}_B$-sample in $\mathbb{D}_B$ is defined to identify suitable previously constructed importance densities. An important difference to conditional reliability estimation is the fact that values of $\boldsymbol{\Theta}_B$ are not based on randomly sampling from $f(\boldsymbol{\theta}_B)$ but are inherently ordered as they are generated by an optimisation procedure. This can simplify the source identification task discussed in Section 3.3. ℓ_2-distance-based information reuse in $\mathbb{D}_B$ is a promising approach as long as the design space dimension remains moderate. [28] use a nearest neighbour search to identify such an ordering based on the ℓ_2-distance amongst $\boldsymbol{\Theta}_B$-samples in $\mathbb{D}_B$. However, in our experience, such a heuristic for the proximity of reliability problems is not robust if either g is not sufficiently well-behaved (e.g., not sufficiently linear in $\boldsymbol{\theta}_B$) or the dimension of $\mathbb{D}_B$ is large. In the latter case, the heuristic will suffer from the concentration of distance in high dimensions [40].

3.3. Source selection

We reuse information by identifying parametric importance densities constructed for previous conditional reliability problems in the sequence (16) that are, in some sense, well-suited to estimate the current conditional failure probability. Thus, each parameter set $\{\boldsymbol{v}_i^j,\ i = 1, \dots, m_j, j = 1, \dots, n_B\}$ is stored in a candidate pool during the computation of the j-th problem, with m_j denoting the number of steps in the iCE method solving the j-th reliability problem. Within importance sampling, the fitness of an importance density for a given reliability problem can be characterized in terms of its proximity to the optimal importance density, e.g., in terms of an f-divergence measure.

3.3.1. Mode search according to Beaurepaire et al.

Estimating an f-divergence measure of each density in the candidate pool from the target density would require a considerable amount of LSF evaluations per available density candidate for source identification only (in addition to the estimation cost). Instead, Ref. [14] proposes a heuristic to reduce the identification cost to a single LSF evaluation per candidate density in the context of RBDO. There, the failure probability is re-evaluated with sequential importance sampling conditional on various design parameter values provided by an optimization sequence. The heuristic is based on evaluating the current (ℓ-th) LSF $G(\boldsymbol{u}_A, \boldsymbol{u}_B^\ell)$ at the mode $\bar{\boldsymbol{u}}_i^j$ of each stored parametric density. The fittest importance density amongst all available candidates is identified as the one whose mode evaluation is closest to 0:

$$[I_\ell, J_\ell] = \underset{\substack{i=1,\dots,m_j,\\ 1j=1,\dots,n_B}}{\arg\min}\ |G(\bar{\boldsymbol{u}}_i^j, \boldsymbol{u}_B^\ell)|, \tag{17}$$

i.e., the importance density of the ℓ-th reliability problem is selected as the importance density constructed at the I_ℓ-th step of the J_ℓ-th conditional reliability problem.

3.3.2. Extension for mixtures and the CE-framework

In this work, we extend the idea presented in Ref. [14] by identifying multiple potentially suitable candidate densities and combining them into a mixture. Instead of identifying a single density amongst the members of the candidate pool, we evaluate the current LSF at the mode of each candidate density to identify a mixture density. The weight α_i^j of the candidate density constructed in the i-th iCE step of the j-th reliability problem is computed as the inverse absolute value of the LSF at the density mode. Mixture components whose weights fall below a threshold value (we choose the threshold as 0.01) are eliminated from the mixture to prevent dilution of the mixtures. The source identification procedure for mixtures is detailed

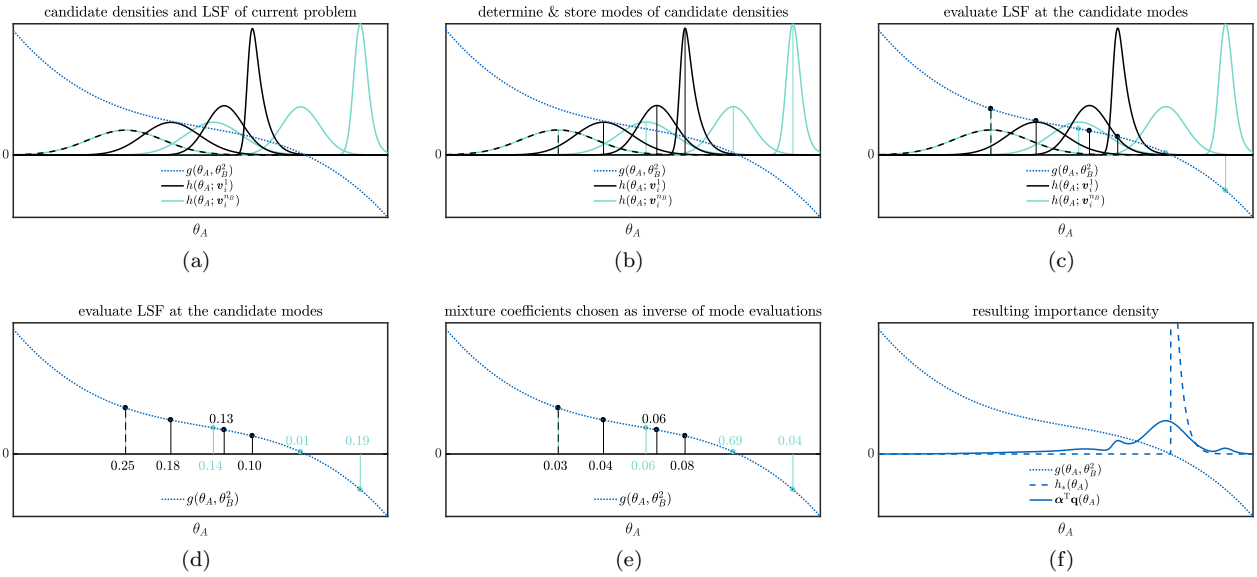

Figure 2: Step-wise illustration of the mode search algorithm (top left to bottom right). The perspective is obtained by a projection of Figure 1 along the U_B-axis. (a) Candidate densities (importance densities from reliability problems 1 and 3) along with LSF of current reliability problem (problem 2) are depicted. (b) The modes of the candidate densities are identified. (c) + (d) The LSF of the current problem is evaluated at these modes. (e) The normalized reciprokes of the mode evaluations form the mixture weights. (f) The importance density mixture for the current problem is computed.

Algorithm 2 Mode-based source identification

Input current conditional LSF $G(\boldsymbol{u}_A, \boldsymbol{u}_B^\ell)$, candidate modes $\bar{\boldsymbol{u}}_i^j$, candidate densities $h(\boldsymbol{u}_A; \boldsymbol{v}_i^j)$
Output number of mixture components M, mixture coefficients $\boldsymbol{\alpha}$, mixture densities $\boldsymbol{q}$

procedure SOURCE-ID(G,$\bar{\boldsymbol{u}}_i^j$,$h(\boldsymbol{u}_A; \boldsymbol{v}_i^j)$)
  Compute mode evaluations $\bar{g}_i^j = G(\bar{\boldsymbol{u}}_i^j, \boldsymbol{u}_B^\ell)$ ▷ $i \in 1, 2, \ldots, m_j, j \in 1, 2, \ldots, \ell - 1$
  Compute mixture coefficients $a_i^j = \frac{1/\bar{g}_i^j}{\sum_{j=1}^{\ell-1} \sum_{i=1}^{m_j} 1/\bar{g}_i^j}$
  Initialize $\boldsymbol{\alpha}$, $\boldsymbol{q}$
  for all a_i^j **do** ▷ Gather tuples identifying important candidates
    if $a_i^j > 0.01$ **then**
      Add: $\boldsymbol{\alpha} \leftarrow a_i^j$, $\boldsymbol{q} \leftarrow h(\boldsymbol{u}_A; \boldsymbol{v}_i^j)$

  Renormalize mixture coefficients $\boldsymbol{\alpha} = \boldsymbol{\alpha}/\|\boldsymbol{\alpha}\|_1 \in [0.01, 1]^{M \times 1}$ ▷ $M = \sum_{j=1}^{\ell-1} \sum_{i=1}^{m_j} \mathrm{I}(a_i^j > 0.01)$
  return M,$\boldsymbol{\alpha}$, $\boldsymbol{q}$

in Algorithm 2, which returns the importance mixture $q_{\boldsymbol{\alpha}} = \boldsymbol{\alpha}^T \boldsymbol{q}$ for each reliability problem but the first. $\boldsymbol{q}$ is the vector of all retained candidates in the current mixture and $\boldsymbol{\alpha}$ is the vector of associated mixture weights. Figure 2 illustrates the algorithm in one dimension ($d_A = 1$). The candidate density modes can be

computed exactly without additional LSF evaluations due to their parametric nature.

3.4. Information reuse for iCE: Mixture-based and controlled importance sampling

Once the mixture $q_{\boldsymbol{\alpha}}$ is identified, it can be used for importance sampling. The mixture-based importance sampling (M-IS) estimate of p^ℓ reads

$$\widehat{p}^\ell_{\mathrm{MIS}} = \sum_{k=1}^{n} \frac{\varphi_{d_A}(\boldsymbol{u}_A^k)}{q_{\boldsymbol{\alpha}}(\boldsymbol{u}_A^k)} \mathrm{I}[G(\boldsymbol{u}_A^k, \boldsymbol{u}_B^\ell) \leq 0], \quad \boldsymbol{u}_A^k \overset{i.i.d.}{\sim} q_{\boldsymbol{\alpha}}. \tag{18}$$

The M-IS estimate's CoV is found in the same way as that of the previously discussed standard IS estimate, i.e.,

$$\widehat{\delta}^\ell_{\mathrm{MIS}} = \frac{1}{\widehat{p}^\ell_{\mathrm{MIS}}} \sqrt{\frac{1}{n^2} \sum_{k=1}^{n} \left(\frac{\varphi_{d_A}(\boldsymbol{u}_A^k)}{q_{\boldsymbol{\alpha}}(\boldsymbol{u}_A^k)} \mathrm{I}[G(\boldsymbol{u}_A^k, \boldsymbol{u}_B^\ell) \leq 0] - \widehat{p}^\ell_{\mathrm{MIS}} \right)^2}. \tag{19}$$

The M-IS is one of two IS estimates employed in this work. The second IS estimate results from the application of a variance reduction technique to the M-IS estimate and is known as controlled importance sampling (C-IS) [41]. The C-IS estimate can be viewed as a control variates analogue for probability densities. The classical control variates method is a variance reduction technique that may be applied to unbiased estimators $\widehat{\mu}$ of $\mathbb{E}[f(\boldsymbol{X})]$, where $\boldsymbol{X}$ is a random vector with CDF $F_{\boldsymbol{X}}$ [17]. The idea is to use a random vector of d_C control variates $\boldsymbol{C}$ that is correlated with $\widehat{\mu}$ and has known mean $\boldsymbol{c} = \mathbb{E}[\boldsymbol{C}]$ to construct a new unbiased estimate with lower variance compared to $\widehat{\mu}$ as:

$$\widehat{\mu}_{\boldsymbol{\beta}} = \widehat{\mu} + \boldsymbol{\beta}^{\mathrm{T}}(\boldsymbol{C} - \boldsymbol{c}).$$

$\boldsymbol{\beta}$ is found by minimizing $\mathbb{V}[\widehat{\mu}_{\boldsymbol{\beta}}]$, which yields [17]

$$\boldsymbol{\beta}_{\mathrm{opt}} = \boldsymbol{\Sigma}_C^{-1} \boldsymbol{\sigma}_{\widehat{\mu}C},$$

where $\boldsymbol{\Sigma}_C$ is the covariance matrix of $\boldsymbol{C}$ and $\boldsymbol{\sigma}_{\widehat{\mu}C}$ is the d_C-dimensional vector of covariances between the components of $\boldsymbol{C}$ and $\widehat{\mu}$. The minimal variance reads

$$\mathbb{V}[\widehat{\mu}_{\boldsymbol{\beta}_{\mathrm{opt}}}] = \mathbb{V}[\widehat{\mu}] - \boldsymbol{\sigma}_{\widehat{\mu}C}^{\mathrm{T}} \boldsymbol{\Sigma}_C^{-1} \boldsymbol{\sigma}_{\widehat{\mu}C},$$

where $\boldsymbol{\sigma}_{\widehat{\mu}C}^{\mathrm{T}} \boldsymbol{\Sigma}_C^{-1} \boldsymbol{\sigma}_{\widehat{\mu}C} \geq 0$, with the expression becoming 0 only if $\boldsymbol{\sigma}_{\widehat{\mu}C} = \boldsymbol{0}$, i.e., if μ and $\boldsymbol{C}$ are uncorrelated. That is, control variates are based on exploiting knowledge about a quantity that is correlated with the estimation target, where the larger the correlation, the larger the variance reduction.

Ref. [41] introduces a second mixture $q_{\boldsymbol{\beta}}$ of control densities with coefficients $\boldsymbol{\beta}$ into the M-IS estimate and add a correction term to preserve its unbiasedness (corresponding to $\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{c}$ above). Ref. [41] states that, ideally, for the M-IS sampler these densities are the ones that constitute the importance mixture $q_{\boldsymbol{\beta}}$ (namely, $\boldsymbol{q}$). The C-IS estimate reads:

$$\widehat{p}^\ell_{\mathrm{CIS}} = \sum_{k=1}^{n} \frac{\varphi_{d_A}(\boldsymbol{u}_A^k) \mathrm{I}[G(\boldsymbol{u}_A^k, \boldsymbol{u}_B^\ell) \leq 0] - q_{\boldsymbol{\beta}}(\boldsymbol{u}_A^k)}{q_{\boldsymbol{\alpha}}(\boldsymbol{u}_A^k)} + \|\boldsymbol{\beta}\|_1, \tag{20}$$

where $q_{\boldsymbol{\beta}} = \boldsymbol{\beta}^T \boldsymbol{q}$. The second summand is the correction term that preserves unbiasedness. Since $q_{\boldsymbol{\beta}}$ is a density and thus ingrates to 1, we have $\mathbb{E}_{q_{\boldsymbol{\alpha}}}[\beta_i q_i / q_{\boldsymbol{\alpha}}] = \beta_i \ \forall \ i \in 1, 2, \ldots, M$. Optimal variance reduction can be achieved by minimizing the variance of the C-IS estimate jointly over the additional free coefficients $\boldsymbol{\beta}$ and its associated estimate $\widehat{p}^\ell_{\mathrm{CIS}}(\boldsymbol{\beta})$. The estimate's variance can be computed based on the sample from $q_{\boldsymbol{\alpha}}$ as

$$(\widehat{\sigma}^\ell_{\mathrm{CIS}})^2 = \frac{1}{n} \sum_{k=1}^{n} \left(\frac{\varphi_{d_A}(\boldsymbol{u}_A^k) \mathrm{I}[G(\boldsymbol{u}_A^k, \boldsymbol{u}_B^\ell) \leq 0] - q_{\boldsymbol{\beta}}(\boldsymbol{u}_A^k)}{q_{\boldsymbol{\alpha}}(\boldsymbol{u}_A^k)} + \|\boldsymbol{\beta}\|_1 - \widehat{p}^\ell_{\mathrm{CIS}} \right)^2. \tag{21}$$

Following Ref. [41], minimizing (21) can be cast as a multiple linear regression of the model $Y(\boldsymbol{u}) = \boldsymbol{c}^T \boldsymbol{Z}(\boldsymbol{u})$ with the extended coefficient vector $\boldsymbol{c} = [\widehat{p}^\ell_{\mathrm{CIS}}, \boldsymbol{\beta}^T]^T$ and

$$\begin{aligned} Y(\boldsymbol{u}) &= \varphi_{d_A}(\boldsymbol{u}) \mathrm{I}[G(\boldsymbol{u}, \boldsymbol{u}_B^\ell) \leq 0] / q_{\boldsymbol{\alpha}}(\boldsymbol{u}) \\ \boldsymbol{Z}(\boldsymbol{u}) &= [1, q_1(\boldsymbol{u})/q_{\boldsymbol{\alpha}}(\boldsymbol{u}) - 1, q_2(\boldsymbol{u})/q_{\boldsymbol{\alpha}}(\boldsymbol{u}) - 1, \ldots, q_M(\boldsymbol{u})/q_{\boldsymbol{\alpha}}(\boldsymbol{u}) - 1]^T. \end{aligned}$$

The multiple linear regression program reads

$$\widehat{\boldsymbol{c}} = \underset{\boldsymbol{c}\in\mathbb{R}^{1\times M+1}}{\arg\min}\, \frac{1}{n}\sum_{k=1}^{n}\left[Y(\boldsymbol{u}_A^k) - \boldsymbol{c}^T\boldsymbol{Z}(\boldsymbol{u}_A^k)\right]^2, \tag{22}$$

where $\widehat{\boldsymbol{\beta}}_{\text{opt}} = \widehat{\boldsymbol{c}}_{2:M+1}$ and $\widehat{p}^{\ell}_{\text{CIS}}(\widehat{\boldsymbol{\beta}}_{\text{opt}}) = \widehat{c}_1$. For simplicity, $\widehat{p}^{\ell}_{\text{CIS}}$ shall always denote the minimum variance estimate $\widehat{p}^{\ell}_{\text{CIS}}(\widehat{\boldsymbol{\beta}}_{\text{opt}})$ and its CoV is denoted by $\widehat{\delta}^{\ell}_{\text{CIS}}$. The latter can be computed directly from the standard error of multiple linear regression as

$$\widehat{\delta}^{\ell}_{\text{CIS}} = \frac{1}{\widehat{p}^{\ell}_{\text{CIS}}}\sqrt{\frac{1}{n(n-M-1)}\sum_{k=1}^{n}\left[Y(\boldsymbol{u}_A^k) - \widehat{\boldsymbol{c}}^T\boldsymbol{Z}(\boldsymbol{u}_A^k)\right]^2}. \tag{23}$$

Eqs. (19) & (23) provide the means to determine the accuracy of the two p^{ℓ}-estimates. In Section 4, we test the efficiency and accuracy of both the M-IS and C-IS estimates against a standard iCE run starting from the nominal distribution p. Based on Algorithm 1, this baseline estimate will have CoV $\widehat{\delta}_{\text{iCE}}$. Thus, the goal is to achieve $\widehat{\delta}_{\text{MIS/CIS}} \leq \delta_{\text{iCE}}$ at lower computational cost compared to the total cost of the iCE baseline. In iCE, δ_{target} is prescribed for the CoV of the weights of the optimal IS density with respect to its current smooth approximation h_i. This is equivalent to requiring that $\sqrt{N}\widehat{\delta}_{\text{iCE}} \lesssim \delta_{\text{target}}$. The inequality is exact if $h_i(\boldsymbol{u})$ and $h(\boldsymbol{u};\boldsymbol{v}_i)$ are equal. Hence, it is reasonable to enforce

$$\widehat{\delta}^{\ell}_{\text{MIS/CIS}} \leq \frac{\delta_{\text{target}}}{\sqrt{N}}, \quad 2 \leq \ell \leq n_B. \tag{24}$$

A straightforward way to ensure criterion (24) with as few samples as possible is to incrementally add samples drawn from the importance mixture to the estimate until convergence. In practice, we start with Δn samples and iteratively increase the number of samples by batches of Δn, where we set $\Delta n = N/100$. The maximum number of samples is set to the number of samples per level in the iCE procedure N. Convergence is likely to be achieved within N samples, if the identified importance mixture $q_{\boldsymbol{\alpha}}$ is a good approximation of the optimal importance density h^*, i.e., if $\text{CoV}[h^*/q_{\boldsymbol{\alpha}}] \leq \sqrt{N}\delta_{\text{target}}$.

3.5. Preconditioning iCE

If convergence is not achieved within N samples, it is still likely that $\text{CoV}[h^*/q_{\boldsymbol{\alpha}}] \leq \text{CoV}[h^*/\varphi_{d_A})]$, such that replacing the nominal density in Algorithm 1 with $q_{\boldsymbol{\alpha}}$ leads to a reduced number of steps in the iCE sequence, m. To precondition iCE in this way, one may use the N samples drawn from $q_{\boldsymbol{\alpha}}$ and evaluated for the M-IS/C-IS-estimate, effectively entering Algorithm (1) in line 8 with $\boldsymbol{v}_0$ corresponding to the parameter set of $q_{\boldsymbol{\alpha}}$. In Algorithm 3, this preconditioned version of iCE is called as preconditioned-iCE$(G(\boldsymbol{u}_A, \boldsymbol{u}_B^j), \delta_{\text{target}}, N, d_A, q_{\boldsymbol{\alpha}}, \mathbf{U}, \mathbf{G})$, where the three additional arguments represent a set of samples $\mathbf{U}$ and the corresponding LSF evaluations $\mathbf{G}$, drawn from $q_{\boldsymbol{\alpha}}$. If the preconditioned iCE step is performed, it is likely because the history of already computed conditional reliability problems does not contain problems that are sufficiently informative with respect to the current one. Therefore, once the current problem is solved with preconditioned iCE, the resulting importance density $h(\boldsymbol{u}_A;\boldsymbol{v}^{\ell}_{m_\ell})$ is added to the pool of candidate densities. In this way, the pool grows adaptively and with each solved conditional reliability problem, it becomes more informative for the problems left in the sequence. The final procedure is outlined in Section 3.6 by Algorithm 3.

3.6. The computational procedure

In Algorithm 3, we summarise the overall procedure to estimate a sequence of conditional reliability problems using Algorithms 1 & 2 as well as samplers (18) or (20).

Algorithm 3 Importance sampling with information reuse

Input LSF $G(\boldsymbol{u})$, input space dimension d, target CoV δ_{target}, sample increment Δn, samples per level N, a set of B-samples $\mathbf{U}_B = \{\boldsymbol{u}_B^j, j = 1, \cdots, n_B\} \in \mathbb{R}^{d_B \times n_B}$

Output conditional failure estimates $\widehat{\boldsymbol{p}} \in [0,1]^{n_B \times 1}$, estimate CoVs $\widehat{\boldsymbol{\delta}} \in \mathbb{R}^{n_B \times 1}$, no. of levels $\boldsymbol{m} \in \mathbb{N}^{n_B \times 1}$

procedure IS-IR(g, δ_{target}, N, d, $\mathbf{U}_B$)
  $j \leftarrow 1$
  $g_j(\boldsymbol{u}_A) = g(\boldsymbol{u}_A, \boldsymbol{u}_B^j)$
  $[\widehat{p}^j, \widehat{\delta}^j, m^j, \{h(\boldsymbol{u}_A, \widehat{\boldsymbol{v}}_i), i = 1, \ldots, m^j\}]$= iCE($g_j$, δ_{target}, N, d_A) ▷ Algorithm 1
  Add $\boldsymbol{q} \leftarrow \{h(\boldsymbol{u}_A, \widehat{\boldsymbol{v}}_i), i = 1, \ldots, m^j\}$
  for $j = 2, \ldots, n_B$ **do**
    $[M, \boldsymbol{\alpha}, \boldsymbol{q}]$ = Source-ID(g,$\bar{\boldsymbol{u}}_i^j$,$h(\boldsymbol{u}_A; \boldsymbol{v}_i^j)$) ▷ Algorithm 2
    $q_{\boldsymbol{\alpha}} = \boldsymbol{\alpha}^T \boldsymbol{q}$
    $n = \Delta n$
    Sample $\mathbf{U}_A = \{\boldsymbol{u}_A^k, k = 1, \ldots, n\} \in \mathbb{R}^{n \times d_A}$ ▷ $\boldsymbol{u}_A^k \overset{i.i.d.}{\sim} q_{\boldsymbol{\alpha}}$
    Compute $\mathbf{G} = G(\mathbf{U}_A, \boldsymbol{u}_B^j) \in \mathbb{R}^{n \times 1}$
    while $n \leq N$ **do**
      With $(\mathbf{U}_A, \mathbf{G})$, compute $\widehat{p}_{\text{MIS}}^j / \widehat{p}_{\text{CIS}}^j$ based on Eq. (18)/(20)
      With $(\mathbf{U}_A, \mathbf{G})$, compute $\widehat{\delta}_{\text{MIS}}^j / \widehat{\delta}_{\text{CIS}}^j$ based on Eq. (19)/(23)
      if $\widehat{\delta}_{\text{MIS/CIS}}^j \leq \delta_{\text{target}} / \sqrt{N}$ **then**
        break
      else if $n \leq N$ **then**
        Sample $\mathbf{U}_{A,\text{new}} = \{\boldsymbol{u}_{A,\text{new}}^k, k = 1, \ldots, \Delta n\} \in \mathbb{R}^{\Delta n \times d_A}$ ▷ $\boldsymbol{u}_{A,\text{new}}^k \overset{i.i.d.}{\sim} q_{\boldsymbol{\alpha}}$
        Compute $\mathbf{G}_{\text{new}} = g(\mathbf{U}_{A,\text{new}}, \boldsymbol{u}_B^j) \in \mathbb{R}^{\Delta n \times 1}$
        Add $\mathbf{U}_A \leftarrow \mathbf{U}_{A,\text{new}}$
        Add $\mathbf{G} \leftarrow \mathbf{G}_{\text{new}}$
      else if $n = N$ **and** $\widehat{\delta}_{\text{MIS/CIS}}^j > \delta_{\text{target}} / \sqrt{N}$ **then**
        $g_j(\boldsymbol{u}_A) = g(\boldsymbol{u}_A, \boldsymbol{u}_B^j)$
        $[\widehat{p}^j, \widehat{\delta}^j, m^j, h(\boldsymbol{u}_A, \widehat{\boldsymbol{v}}_{m^j}^j)]$ = preconditioned-iCE($g_j, \delta_{\text{target}}, N, d_A, q_{\boldsymbol{\alpha}}, \mathbf{U}_A, \mathbf{G}$)
        Add $\boldsymbol{q} \leftarrow h(\boldsymbol{u}_A, \widehat{\boldsymbol{v}}_{m^j}^j)$
  return $\widehat{\boldsymbol{p}}$,$\widehat{\boldsymbol{\delta}}$,$\boldsymbol{m}$

4. Numerical experiments

4.1. Parameter Study: Sequential processing chain

In the first example, we consider a sequence of processing steps. Each step is modelled as a Possion process and arrival of the first jump indicates finalisation of a step and triggers the subsequent step. The arrival time of the first jump in the i-th step is distributed exponentially with rate parameter λ_i, which is an uncertain parameter. The goal is to estimate the probability of the total processing time exceeding a threshold. The LSF thus reads

$$g(\boldsymbol{t}, \boldsymbol{\lambda}) = T - \underbrace{\sum_{i=1}^{d_A} t_i}_{\sim \text{Hypoexp}(\boldsymbol{\lambda})}, \qquad t_i \sim \text{Exp}(\lambda_i), \tag{25}$$

where Exp denotes the exponential distribution and Hypoexp denotes the hypoexponential distribution. The sum of d_A independent exponential random variables with rate parameters $\boldsymbol{\lambda} \in \mathbb{R}_+^{d_A \times 1}$ is hypoexponentially distributed [42]. The exponential and hypoexponential CDF read

$$\begin{aligned} F_{\text{Exp}}(x;\lambda) &= 1 - \mathrm{e}^{-\lambda x} \quad \forall x \geq 0, \\ F_{\text{Hyp}}(x;\boldsymbol{\lambda}) &= 1 - \sum_{i=1}^{d_A} \mathrm{e}^{-\lambda_i x} \prod_{j=1, j\neq i}^{d_A} \frac{\lambda_j}{\lambda_j - \lambda_i} \quad \forall x \geq 0. \end{aligned}$$

With known rate parameter vector $\boldsymbol{\lambda}$, the failure probability is given as

$$\mathbb{P}(\text{F}|\boldsymbol{\lambda}) = \mathbb{P}(g(\boldsymbol{t}) \leq 0) = 1 - F_{\text{Hyp}}(T;\boldsymbol{\lambda}). \tag{26}$$

We let $\boldsymbol{\theta}_A = \boldsymbol{t} \in \mathbb{R}^{d_A \times 1}$, $\boldsymbol{\Theta}_B \sim \mathcal{U}_{d_B}(0,1) \in \mathbb{R}^{d_B \times 1}$, where $d_B = n \cdot d_A,\ n \in \mathbb{N}$, and

$$\frac{1}{\lambda_i} = \frac{1}{n} \sum_{\substack{j=n\cdot i \\ -n+1}}^{n\cdot i} \theta_{B,j}, \quad i = 1, 2, \ldots, d_A. \tag{27}$$

That is, each inverse rate parameter (scale parameter) is computed as the average of n different reducible variables. With these definitions and transforming the LSF to standard-normal space, Eq. (25) reads

$$G(\boldsymbol{u}_A, \boldsymbol{u}_B) = T - \sum_{i=1}^{d_A} F_{\text{Exp}}^{-1}(\Phi(u_{A,i});\lambda_i), \quad \frac{1}{\lambda_i} = \frac{1}{n} \sum_{\substack{j=n\cdot i \\ -n+1}}^{n\cdot i} \Phi(u_{B,j}). \tag{28}$$

We define $\boldsymbol{A} \in \{0,1\}^{d_A \times d_B}$ such that $\boldsymbol{\lambda}^{-1} = \boldsymbol{A}\boldsymbol{\theta}_B/n$. $\boldsymbol{A}$ has a banded structure with elements of the main diagonal and the first $n-1$ diagonals above the main diagonal being 1 and the rest 0. An analytical expression for $P_{\text{F}}(\boldsymbol{U}_B)$ follows from Eq. (26):

$$P_{\text{F}}(\boldsymbol{U}_B) = 1 - F_{\text{Hyp}}\left(T; (\boldsymbol{A}\Phi(\boldsymbol{U}_B)/n)^{-1}\right). \tag{29}$$

By inverting Eq. (29) one may compute T_p corresponding to the first-order approximation of a fixed unconditional failure level p as

$$T_p = F_{\text{Hyp}}^{-1}\left(1-p; (\boldsymbol{A}\Phi(\mathbb{E}\left[\boldsymbol{U}_B\right])/n)^{-1}\right) = F_{\text{Hyp}}^{-1}(1-p;\mathbf{2}), \tag{30}$$

where $\mathbf{2}$ denotes a d_A-dimensional vector of twos. In the following application, if not stated otherwise, $d_B = 100$ and $n = 5$ such that $d_A = 20$. The unconditional failure probability is chosen as $\mathbb{P}(\text{F}) = 10^{-5}$ and the number of $\boldsymbol{\Theta}_B$-samples as $n_B = 100$. We use the relative root-mean-squared error (RMSE) to measure the accuracy of a failure probability estimate $\widehat{p}$ conditional on $\boldsymbol{\theta}_B$, which reads

$$e(\widehat{p}|\boldsymbol{\theta}_B) = \frac{1}{P_{\text{F}}(\boldsymbol{\theta}_B)} \sqrt{\mathbb{E}\left[\left(\widehat{p}(\boldsymbol{\theta}_B) - P_{\text{F}}(\boldsymbol{\theta}_B)\right)^2\right]}, \tag{31}$$

and plot the mean of $e(\widehat{p}|\boldsymbol{\theta}_B)$ taken over the n_B $\boldsymbol{\Theta}_B$-samples in Figure 3. The expectation in Eq. (31) is computed with 100 repetitions of the analysis using the same $\boldsymbol{\Theta}_B$ sample set. The proposed algorithms (M-IS & C-IS) are benchmarked against an iCE reference solution with equal number of samples per level $N = 1000$ and target CoV $\delta_{\text{target}} = 1.5$ for iCE, M-IS and C-IS. This corresponds to a target CoV of the conditional failure probability estimate of approximately 0.05. A series of parameter studies is devised to illustrate the behaviour of our algorithm depending on the failure magnitude of the problem ($\mathbb{P}$(F) - through choosing different T_p thresholds according to Eq. (30)), its dimensionality in A-space (d_A) and B-space (d_B) and the number of conditional reliability problems present in the sequence (n_B).

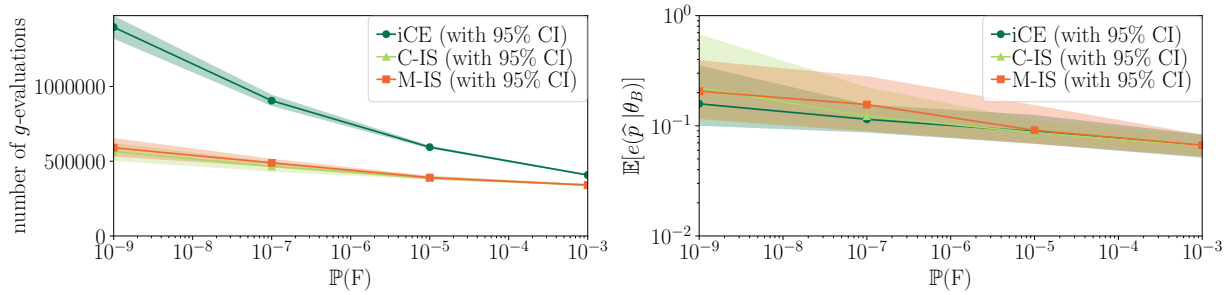

Figure 3: Failure probability estimation: computational cost (left) and RMSE mean including confidence intervals (right) with varying unconditional failure probability levels.

The estimator accuracy (mean and variance) of both M-IS and C-IS-based information reuse samplers matches that of the reference solution when varying $\mathbb{P}$(F) over a wide range (10^{-3} and 10^{-9}, see Figure 3). The computational savings of our information reuse samplers increase as $\mathbb{P}$(F) decreases and reach $> 60\%$ saved effort at $\mathbb{P}(\text{F}) = 10^{-9}$. This is somewhat intuitive as the number of required steps in the iCE sequence grows with decreasing failure probability magnitude such that the savings potential rises. C-IS does not notably outperform M-IS, i.e., the control variate effect is negligible at all investigated $\mathbb{P}$(F)-values.

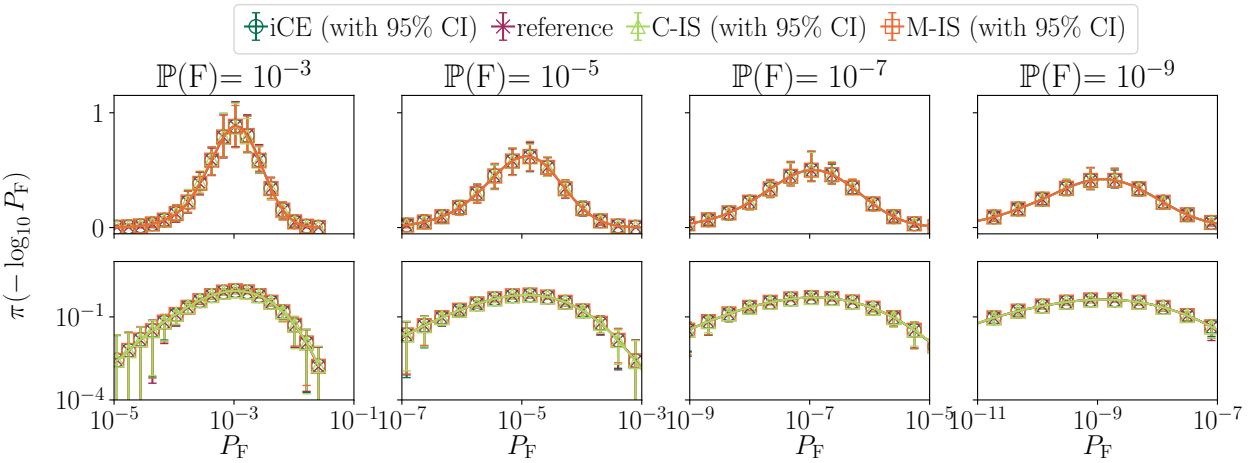

Figure 4: Conditional failure probability densities in linear (upper) and logarithmic (lower) scale with varying unconditional failure probability levels.

Figure 4 shows the density estimates corresponding to the $\mathbb{P}$(F)-study. Both mean and confidence intervals (CI) of the density estimates produced with M-IS/C-IS match exactly with the reference solution.,The confidence intervals here are again based on 100 repeated runs of iCE, M-IS and C-IS while drawing a *different* sample set $\{\boldsymbol{\theta}_B^j, j = 1, \ldots, n_B\}$ in each repetition (as opposed to the plots in Figure 3, which are based on 100 repetitions given an identical sample set). That is, the CIs in Figure 4 represent the aggregation of statistical uncertainty stemming from the failure estimation (iCE/M-IS/C-IS) *and* the density estimation based on n_B

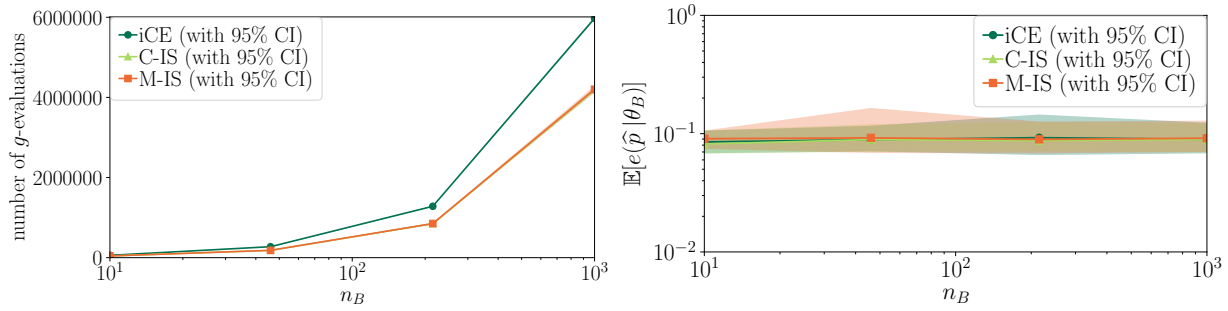

Figure 5: Computational cost (left) and mean relative RMSE of conditional failure probability estimates including confidence intervals (right) with varying number of type B-variable samples.

samples. This procedure (100 repetitions at identical type B-sample for cost/error analysis and 100 random draws of the type B -sample for computing P_F density CIs) is used for any following parameter study as well. The second parameter study considers the number of type B-samples n_B. There, the relative savings potential at $\mathbb{P}(\mathrm{F}) = 10^{-5}$ amounts to 30% (this is also evident from the first parameter study in Figure 3) at all investigated sample sizes n_B, while the absolute savings scale proportionally with the total computational effort or n_B (Figure 5, left). Although the relative RMSE of all failure probability estimators is independent of n_B (Figure 5), P_F density CIs are shrinking around the mean density estimates as n_B increases (Figure 6) due to the aggregation of failure probability and density estimation uncertainty in these computations (the uncertainty in the density estimation). As n_B rises further, we expect an increase in relative savings due to an increasingly dense population of the type B-variable space with samples. This, in turn, will produce more correlation amongst the LSFs, which facilitates more efficient information reuse. n_B does not affect the relative RMSE of the failure probability estimates (Figure 5, right). Finally, we study the influence of

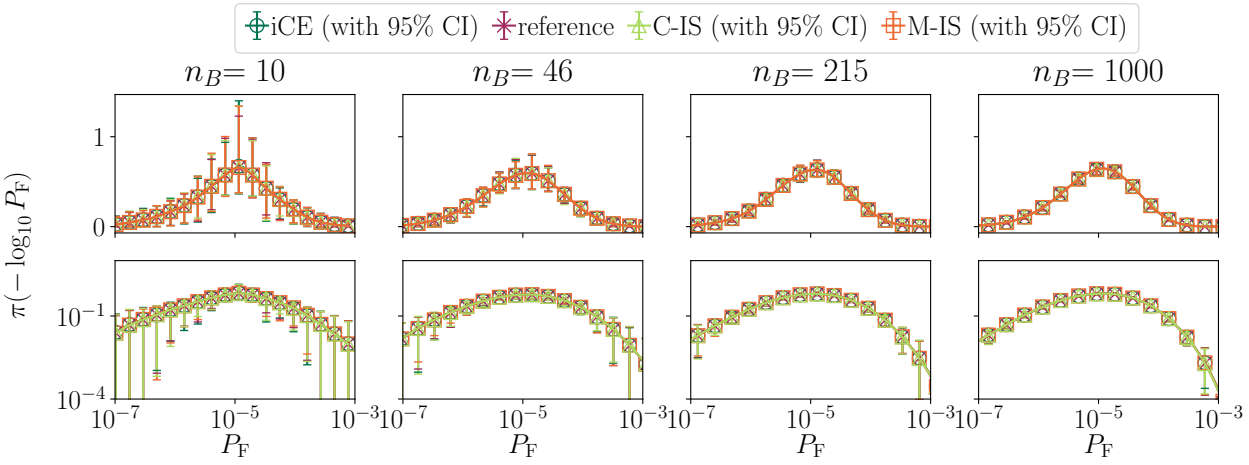

Figure 6: Conditional failure probability densities in linear (upper) and logarithmic (lower) scale with varying conditional sample size n_B.

the subspace dimensions d_A and d_B on the performance of our method. Figure 7 shows the progression of computational cost and RMSE as d_B increases and $d_A = 10$, $\mathbb{P}(\mathrm{F}) = 10^{-5}$ and $n_B = 100$ are fixed. The corresponding conditional failure probability densities - plotted in Figure 8 - reveal a decreasing variability in P_F as d_B rises. This is to be expected as every rate parameter (type A-variable) in $\boldsymbol{\lambda}$ is averaged over an increasing number of type B variables. That is, the rate parameter variance - and thus also the variance of P_F - scale inversely with d_B. As d_B increases from 10 to 1000, the computational savings increase from 45% to $> 70\%$ while the error (RMSE) decreases slightly due to the decreasing variability of P_F. The same effect is observed when increasing d_A while keeping $d_B = 100$ (and again $d_B = 100$, $\mathbb{P}(\mathrm{F}) = 10^{-5}$ and $n_B = 100$)

constant: The variance of P_F increases with rising d_A (Figure 10), which in turn causes a slight increase of the relative RMSE (Figure 9). Figure 9 also reveals an intermediate d_A regime, in which the computational savings are relatively low ($\approx 20\%$) due to a plateau in the computational cost of the reference solution between $d_A = 10$ and $d_A = 25$. In conclusion, this study demonstrates the robustness of the information

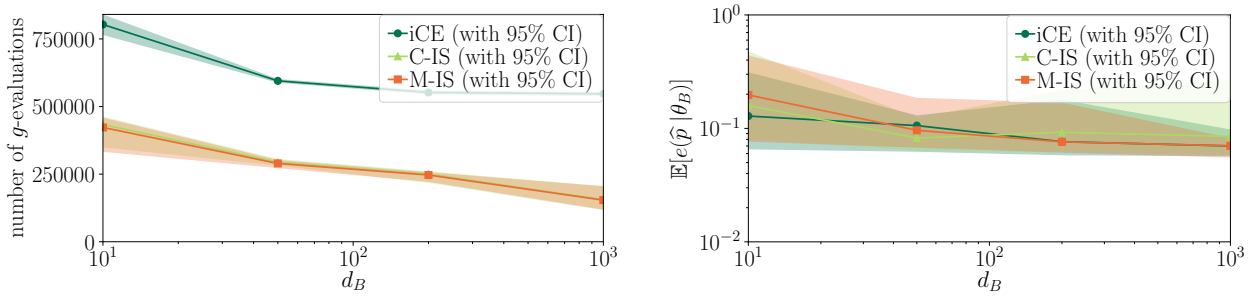

Figure 7: Failure probability estimation: computational cost (left) and RMSE mean including confidence intervals (right) with varying problem dimension d_B.

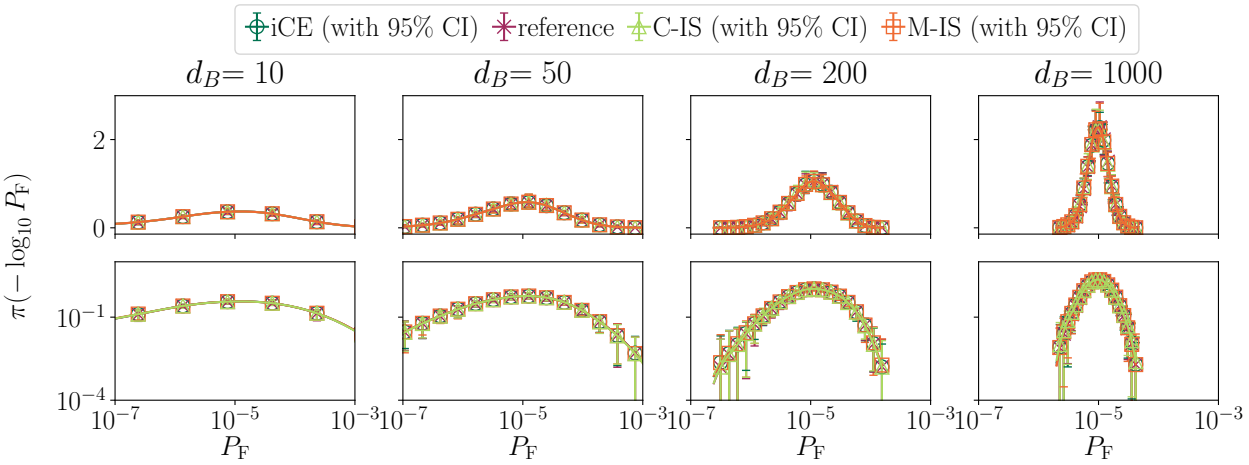

Figure 8: Conditional failure probability densities in linear (upper) and logarithmic (lower) scale with varying dimension d_B.

reuse-based M-IS/C-IS estimators with respect to the problem dimensions in both type A & B - variables while facilitating computational savings of $\approx 25\% - 75\%$.

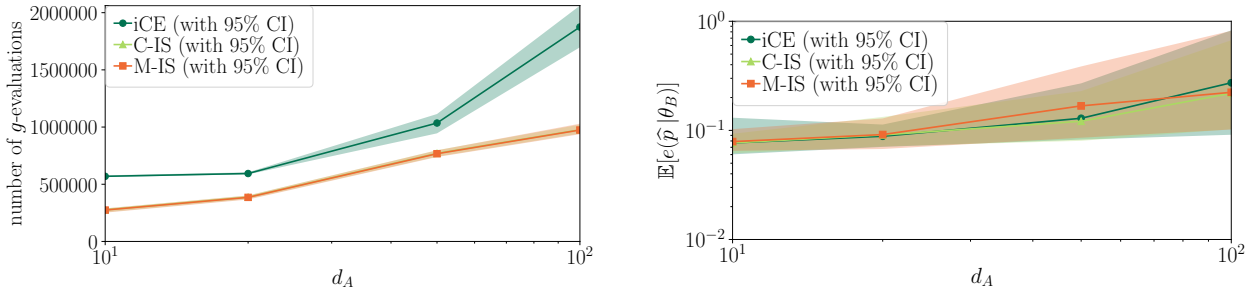

Figure 9: Failure probability estimation: computational cost (left) and RMSE mean including confidence intervals (right) with varying problem dimension d_A.

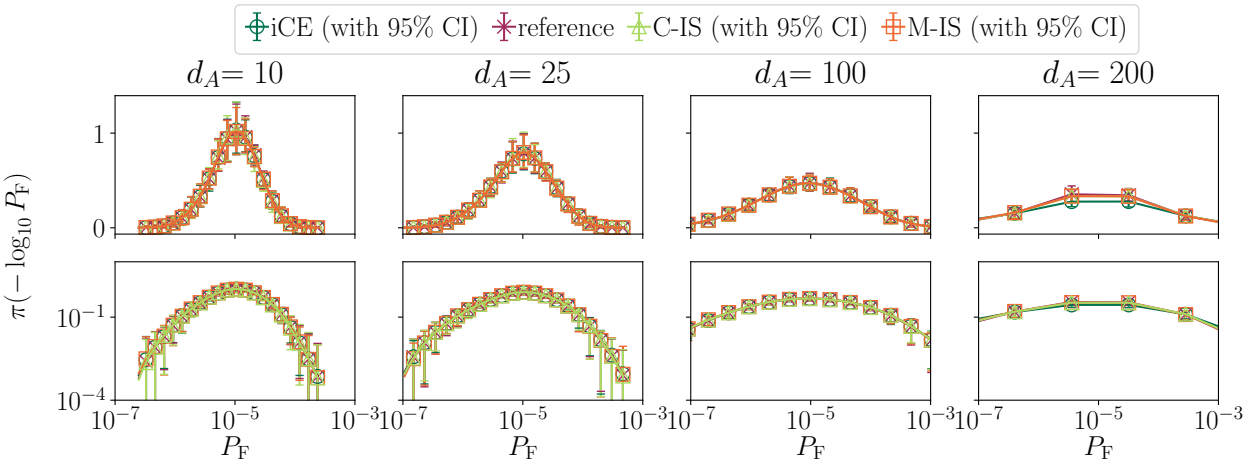

Figure 10: Conditional failure probability densities in linear (upper) and logarithmic (lower) scale with varying dimension d_A.

4.2. Case Study: Monopile foundation in plastically behaving soil

4.2.1. Problem Setup

This case study is based on a finite element model for the interaction of a monopile foundation of an offshore wind turbine (Figure 11) with stiff, plastic soil. Deterministic parameters of the monopile are its depth $L = 30$ m, diameter $D = 6$ m, wall thickness $t = 0.07$ m, Poisson ratio $\nu = 0.3$ and Young's modulus $E = 2.1 \cdot 10^5$ MPa. The uncertain inputs comprise the lateral load H as well as the undrained shear strength s of the soil and hyperparameters of both quantities. The engineering model consists of a nonlinear finite element code whose setup is described in detail in Ref. [43] and the probabilistic model considered there has been modified following Ref. [44]: s is considered both uncertain and increasing in mean with soil depth z.

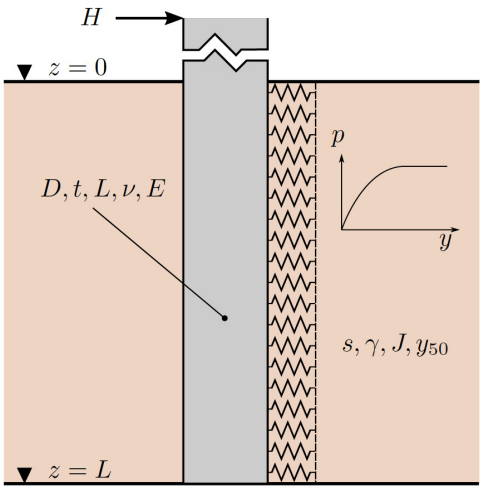

Figure 11: Wind turbine monopile foundation [43].

It is thus modelled by a random field with linear mean drift along the soil depth coordinate z. Given an underlying stationary Gaussian random field $\tilde{s}(z, \boldsymbol{\Theta})$

$$\{\tilde{s}(z) : 0 \leq z \leq L\} \sim \mathcal{N}(0, \sigma_{\tilde{s}}),$$

the non-stationary random field representing the shear strength of the soil can be expressed as

$$\begin{aligned} s(z, \boldsymbol{\Theta}) &= s_0 + s_1 \sigma'(z) \exp\{\tilde{s}(z, \boldsymbol{\Theta})\} \\ &= s_0 + s_1 \gamma z \exp\{\tilde{s}(z, \boldsymbol{\Theta})\}, \end{aligned}$$

where γ is the soil unit weight, $\sigma'(z) = \gamma z$ is the effective vertical stress, s_0 is the undrained shear strength at ground level and s_1 is the drift parameter governing the mean increase of s with increasing soil depth. $\tilde{s}(z, \boldsymbol{\Theta})$ models the intra-site variability. That is, at a given site with known deterministic s_0 and s_1, it describes the inherent variability of the undrained shear strength. In order to describe the inter-site variability, the parameters s_0 and s_1 are modelled probabilistically as well. The stationary RF $\tilde{s}$ is equipped with an exponential correlation function:

$$\rho_{\tilde{s}\tilde{s}}(z', z'') = \exp\left\{-\frac{2|z' - z''|}{\theta_{\tilde{s}}}\right\},$$

with vertical soil scale of fluctuation $\theta_{\tilde{s}} = 1.9m$ [45] and standard deviation $\sigma_{\tilde{s}} = 0.3$ [45, 46]. We assume the soil to be stiff and plastic according to the classification provided in Ref. [47]. There, the specific soil weight range is given with $17 - 19kN/m^3$, thus we set $\gamma = 18kN/m^3$. The mean cohesion range is given with $20 - 50kN/m^2$ by Ref. [47] and Ref. [48] lists the mean range of the undrained shear strength ratio s_u/σ' as $0.23 - 1.4$. We fit log-normal distributions for s_0 and s_1 by setting the 10 % and 90% quantiles of the distributions equal to the lower and upper bounds of these ranges. The resulting parameters are detailed in Table 1 along with uncertain parameters for the load H, namely μ_H and σ_H. The mean and CoV of the load Gumbel distribution in Table 1 are conditional on the parameters a_H (location parameter) and b_H (scale parameter):

$$\begin{aligned}\mu_{H|a_H,b_H} &= \mu_{a_H} + \gamma_E \mu_{b_H} \\ \delta_{H|a_H,b_H} &= \frac{\pi}{\sqrt{6}} \frac{\mu_{b_H}}{\mu_{H|a_H,b_H}},\end{aligned}$$

where γ_E is the Euler-Mascheroni constant. $\tilde{s}$ is simulated by means of the midpoint method. That is,

Table 1: Input variable definitions of the monopile foundation.

Input	Distribution	Mean μ	CoV δ
$\boldsymbol{\xi}$ $[-]$	Standard-Normal	$\mathbf{0}$	$n.d.$ $(\boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\xi}} = \boldsymbol{I}_{n\times n})$
s_0 $[kPa]$	Log-Normal	33.7094	0.3692
s_1 $[kPa]$	Log-Normal	0.7274	0.8019
H $[kN]$	Gumbel	$\mu_{P|a_H,b_H}$	$\delta_{P|a_H,b_H}$
a_H $[kN]$	Log-Normal	2274.97	0.2
b_H $[kN]$	Log-Normal	225.02	0.2

the spatial domain $[0, L]$ is discretized with n spatial elements and $\tilde{s}$ is represented by means of n random variables with joint distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\tilde{s}\tilde{s}})$. The random variables represent the random field values at the element midpoints. Thus, the covariance matrix $\boldsymbol{\Sigma}_{\tilde{s}\tilde{s}}$ is computed by evaluating $\sigma^2_{\tilde{s}}\rho_{\tilde{s}\tilde{s}}(z', z'')$ at the element midpoints. The number of elements is chosen such that 95% of the inherent RF variability is captured by the RF discretization, leading to $n = 82$ in this example. Therefore, the total input dimension is $d = 87$. As the sampling approaches are implemented in the standard normal space, the midpoint random variables are transformed to independent standard normal random variables, denoted as $\boldsymbol{\xi}$, by means of the Cholesky decomposition of $\boldsymbol{\Sigma}_{\tilde{s}\tilde{s}}$. The model output $Y = \mathcal{Y}(\boldsymbol{\theta})$ is the maximum occurring stress in the foundation. The LSF is given by

$$G(\boldsymbol{u}) = \sigma_{crit} - \mathcal{Y}(T^{-1}(\boldsymbol{u})),$$

with T the transformation from the original input probability space $\mathbb{D}_A$ to standard-normal space and $\sigma_{crit} =$ 100 MPa the stress threshold, which corresponds to a system failure probability of $\mathbb{P}(F) = 3.6{\cdot}10^{-4}$ (estimated with MC and $\widehat{\delta}_{MC} = 0.1187$). Depending on the availability of measurements and inter-site data, the assignment of inputs to either $\boldsymbol{\Theta}_A$ and $\boldsymbol{\Theta}_B$ may vary. We illustrate two cases, the first of which features a high-dimensional reducible space $\mathbb{D}_B$ as we set $\boldsymbol{\Theta}_A = [a_H, b_H, H]$and $\boldsymbol{\Theta}_B = [s_0, s_1, \boldsymbol{\xi}]$ while the second has high-dimensional $\mathbb{D}_A$, where $\boldsymbol{\Theta}_A = [\boldsymbol{\xi}, H]$ and $\boldsymbol{\Theta}_B = [s_0, s_1, a_H, b_H]$.

4.2.2. Efficiency & Accuracy

Due to the large computational cost of a single evaluation of the monopile foundation model, we compare single runs of M-IS/C-IS over a given type B-sample to the iCE reference, which is repeated 26 times over that same sample to estimate a confidence interval. For both type B-variable configurations, we set the number of samples per iCE level to $N = 500$ and choose $\delta_{\text{target}} = 1.5$ and $n_B = 100$. The monopile

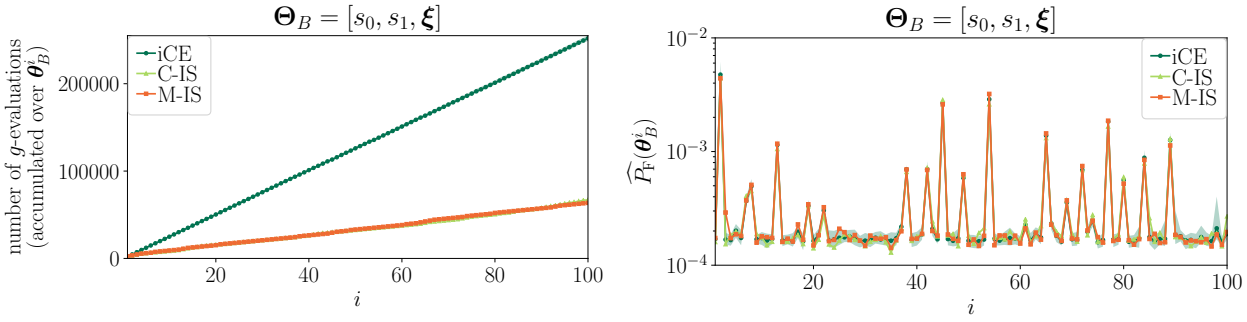

Figure 12: Monopile configuration 1: computational cost (left) and conditional failure probability estimates at $\boldsymbol{\Theta}_B$ samples (right) with iCE reference solution mean and 95% CI.

example is well-suited to demonstrate the dependence of information reuse-based savings potential on the partitioning of inputs in type A and type B. Figures (12) & (13) (both: right) show that the conditional failure probabilities are computed accurately with the information reuse estimators for both type B-variable configurations. However, while for the configuration $\boldsymbol{\Theta}_B = [s_0, s_1, \boldsymbol{\xi}]$, these results are obtained with $\approx 25\%$ of the reference computational cost, the second configuration $\boldsymbol{\Theta}_B = [s_0, s_1, a_H, b_H]$ admits almost no savings (Figures (12) & (13), left). This is mostly due to the random field hyperparameters a_H and b_H that are present in the second configuration and cause a large fraction of P_{F}-variability: The conditional failure probabilities span roughly 6 orders of magnitude in the second configuration (10^{-8}-10^{-2} , Figure (13), right) versus only 2 orders of magnitude in the first (10^{-4}-10^{-2}, Figure (12), right). This high variability delays the construction of a pool of candidate densities that are relevant for information reuse. As the number of $\boldsymbol{\Theta}_B$-samples increases we expect the computational savings to increase.

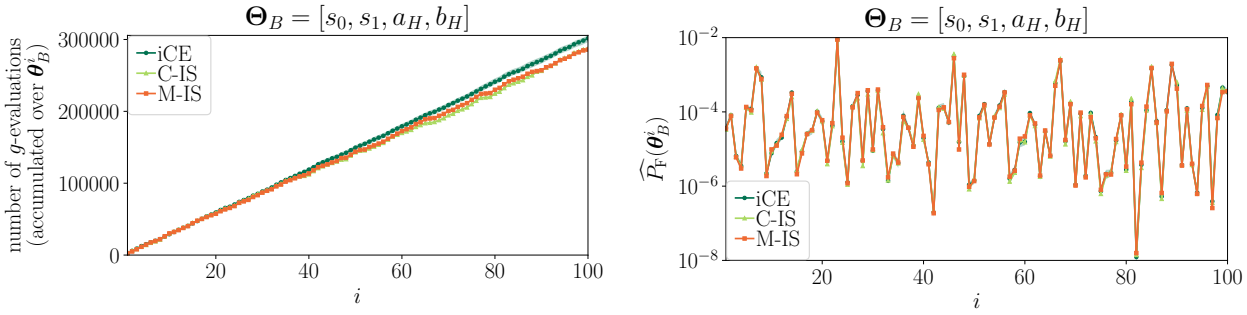

Figure 13: Monopile configuration 2: computational cost (left) and conditional failure probability estimates at $\boldsymbol{\Theta}_B$ samples (right) with iCE reference solution mean and 95% CI.

4.2.3. Conditional densities and global sensitivity analysis

Based on $n_B = 500$ (first configuration) and $n_B = 100$ (second configuration) type B-variable samples, we compute confidence intervals on P_{F} (Table 2) and a kernel density estimate of its distribution (Figures 14 & 15, left). The number of samples per iCE level is set to $N = 1000$. The intervals again demonstrate the vast increase of P_{F}-variability in between the first to the second type B-variable configuration. Moreover, we compute reliability-oriented variance-based sensitivity indices according to Ref. [6]: The n_B conditional failure probability samples are used to construct a surrogate (a partial least squares-based polynomial chaos

Table 2: P_F-confidence intervals for two type B-variable configurations.

	$\boldsymbol{\Theta}_B = [s_0, s_1, \boldsymbol{\xi}]$		$\boldsymbol{\Theta}_B = [s_0, s_1, a_H, b_H]$	
Mean	$2.537 \cdot 10^{-4}$		$2.828 \cdot 10^{-4}$	
interval width	lower CI	upper CI	lower CI	upper CI
90%	$1.530 \cdot 10^{-4}$	$4.812 \cdot 10^{-4}$	$7.053 \cdot 10^{-7}$	$1.457 \cdot 10^{-3}$
95%	$1.476 \cdot 10^{-4}$	$7.060 \cdot 10^{-4}$	$3.910 \cdot 10^{-7}$	$2.325 \cdot 10^{-3}$
99%	$1.444 \cdot 10^{-4}$	$2.789 \cdot 10^{-3}$	$1.308 \cdot 10^{-8}$	$8.658 \cdot 10^{-3}$

expansion [49]) of $\log P_\mathrm{F}(\boldsymbol{\Theta}_B)$, $\log \widehat{P_\mathrm{F}}(\boldsymbol{\Theta}_B)$. Sobol' and total-effect indices of $\log \widehat{P_\mathrm{F}}$ with respect to $\boldsymbol{\Theta}_B$ can be computed by post-processing the surrogate model coefficients [50]. The sensitivity indices indicate that the random field drift gradient s_1 contributes by far the largest variability to P_F in the first configuration while in the second, the load hyperparameters a_H and b_H dominate the random field parameters (Figures 14 & 15, right). We observe, that Sobol' and total-effect indices are much closer to one another under the second configuration indicating an absence of interaction effects. Under the first configuration, such higher-order effects play a more prominent role, which is due to the fact that in this configuration all type B-variables are random field-related. In other words, any pair of parameters belonging to the same probabilistic model is likely to exhibit stronger dependencies than a pair belonging to two different probabilistic models (here: to the random field model and the random load model). In the latter case, the pair of parameters can only interact through the FE-model, where however first-order effects seem to be dominant. Asymptotically, the mean estimates of the conditional failure probability provided in Table 2 will coincide with the unconditional failure probability of the system.

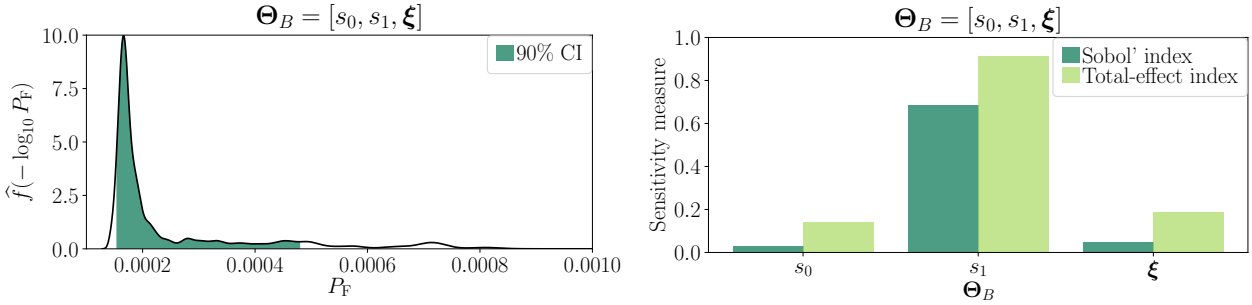

Figure 14: Monopile configuration 1: Conditional failure PDF with confidence intervals (left) and variance-based failure sensitivity indices (right) with $n_B = 500$.

The computed quantities help answer questions such as: 'Is there something to gain from gathering additional information on any of the type B inputs?' And 'If so, which parameters should we learn and update by collecting additional information on them?' Based on the confidence intervals, the answer to the first question may be based on predefined maximally admissible bounds on the interval widths or upper semi-widths (for the failure probability, naturally, the upper tail of the distribution is the decisive one). Under the second configuration, the large target variability motivates collecting additional information for uncertainty reduction, whereas, in the first configuration the target variability is already quite low. Then, in order to decide which type B-variables require an update one may resort to the variable ranking provided by the Sobol' and total-effect indices of $\log \widehat{P_\mathrm{F}}$: For the second configuration, the load dispersion parameter b_H contributes most of the variance in $\log \widehat{P_\mathrm{F}}$ and thus should receive priority in the acquisition of additional data.

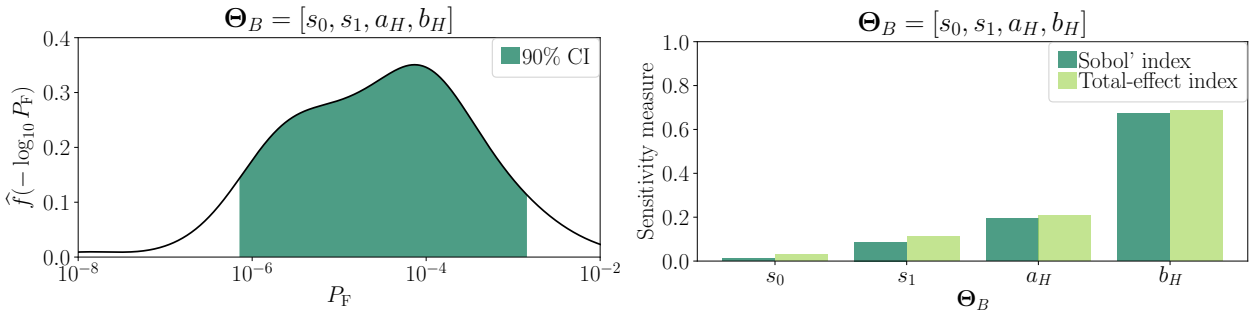

Figure 15: Monopile configuration 2: Conditional failure PDF with confidence intervals (left) and variance-based failure sensitivity indices (right) with $n_B = 100$.

5. Conclusion

This paper reviews conditional reliability analysis, i.e., estimating the probability of failure conditional on a subset of the uncertain inputs. Such Conditional probabililties of failures are random functions of the input they're conditioned upon and are useful in a multitude of contexts such as sensitivity analysis, quantification and communication of lack of knowledge and decision analysis. The estimation of conditional failure probabilities is a computationally intensive task. We present a method to efficiently perform conditional reliability estimation by reusing information throughout the computation. Information reuse is realized through importance densities from previous computation steps for importance sampling estimates of conditional failure probabilities. We propose a strategy for the selection of these densities and test two importance sampling estimators that efficiently incorporate them to reduce the estimator variance, namely: mixture importance sampling and controlled importance sampling.

In two numerical examples, we find that both mixture and controlled importance sampling perform similarly and provide up to 76% computational savings compared to a baseline method without information reuse. A parameter study reveals the robustness of the proposed method both with respect to the magnitude of the probability of failure (down to $\mathbb{P}(\mathrm{F}) = 10^{-9}$) and the dimensionality of both type A- and type B- variable spaces (scenarios with up to 200 type A-variables and 1000 type B-variables are investigated). The second example showcases the application of our method to compute the distribution and global sensitivty indices of two differently conditioned probabilities of failure for a wind turbine foundation model with 87 inputs. We find that the potential for computational savings offered by information reuse depends on the variability contributed to P_F by the type B-variables. Namely, increasing variability reduces correlation amongst the limit-state functions in the reliability problem sequence and therefore the re-usability of importance densities.

In order to further increase the potential for computational savings of information reuse, surrogate models may be used to replace the expensive engineering model at various points in the algorithm. For example, the sequence of importance densities occurring in iCE could be based on locally reconstructed surrogate models. Similar approaches have been used in the context of conventional reliability analysis [51] and Bayesian updating [52].

6. Acknowledgment

This project was supported by the German Research Foundation (DFG) through Grant STR 1140/6-1 under SPP 1886. The third author acknowledges support of the AEOLUS center under the U.S. Department of Energy Applied Mathematics MMICC award DE-SC0019303. The first author would like to thank Anirban Chaudhuri and Boris Kramer who have been very helpful in discussing information reuse concepts as well as importance sampling and its variants.

References

[1] M. H. Faber, On the treatment of uncertainties and probabilities in engineering decision analysis, Journal of Offshore Mechanics and Arctic Engineering 127 (2005) 243–248.

[2] A. O'Hagan, J. E. Oakley, Probability is perfect, but we can't elicit it perfectly, Reliability Engineering & System Safety 85 (2004) 239–248.

[3] S. Ferson, C. A. Joslyn, J. C. Helton, W. L. Oberkampf, K. Sentz, Summary from the epistemic uncertainty workshop: consensus amid diversity, Reliability Engineering & System Safety 85 (2004) 355–369.

[4] J. Helton, W. Oberkampf, Alternative representations of epistemic uncertainty, Reliability Engineering & System Safety 85 (2004) 1 – 10. Alternative Representations of Epistemic Uncertainty.

[5] A. Der Kiureghian, O. Ditlevsen, Aleatory or epistemic? Does it matter?, Structural Safety 31 (2009) 105–112.

[6] M. Ehre, I. Papaioannou, D. Straub, A framework for global reliability sensitivity analysis in the presence of multi-uncertainty, Reliability Engineering & System Safety 195 (2020) 106726.

[7] J. E. Oakley, Decision-theoretic sensitivity analysis for complex computer models, Technometrics 51 (2009) 121–129.

[8] O. Ditlevsen, H. O. Madsen, Structural reliability methods, John Wiley & Sons Ltd, 1996.

[9] M. Lemaire, A. Chateauneuf, J.-C. Mitteau, Structural reliability, Wiley-ISTE, 2009.

[10] A. Der Kiureghian, First-and second-order reliability methods, in: E. Nikolaidis, D. M. Ghiocel, S. Singhal (Eds.), Engineering Design Reliability Handbook, CRC Press, Boca Raton, FL, 2005.

[11] C. G. Bucher, Adaptive sampling an iterative fast Monte Carlo procedure, Structural Safety 5 (1988) 119 – 126.

[12] S. Engelund, R. Rackwitz, A benchmark study on importance sampling techniques in structural reliability, Structural Safety 12 (1993) 255 – 276.

[13] S. Au, J. Beck, A new adaptive importance sampling scheme for reliability calculations, Structural Safety 21 (1999) 135 – 158.

[14] P. Beaurepaire, H. Jensen, G. Schuëller, M. Valdebenito, Reliability-based optimization using bridge importance sampling, Probabilistic Engineering Mechanics 34 (2013) 48–57.

[15] I. Papaioannou, C. Papadimitriou, D. Straub, Sequential importance sampling for structural reliability analysis, Structural Safety 62 (2016) 66 – 75.

[16] S.-K. Au, J. L. Beck, Estimation of small failure probabilities in high dimensions by subset simulation, Probabilistic Engineering Mechanics 16 (2001) 263 – 277.

[17] R. Y. Rubinstein, D. P. Kroese, Simulation and the Monte Carlo method, Springer, 3 edition, 2017.

[18] D. P. Kroese, R. Y. Rubinstein, P. W. Glynn, Chapter 2 - The cross-entropy method for estimation, in: Handbook of Statistics: Machine Learning: Theory and Applications, volume 31 of *Handbook of Statistics*, Elsevier, 2013, pp. 19 – 34.

[19] N. Kurtz, J. Song, Cross-entropy-based adaptive importance sampling using gaussian mixture, Structural Safety 42 (2013) 35 – 44.

[20] M. Hohenbichler, R. Rackwitz, Improvement of second-order reliability estimates by importance sampling, Journal of Engineering Mechanics 114 (1988) 2195–2199.

[21] P. Koutsourelakis, H. Pradlwarter, G. Schuller, Reliability of structures in high dimensions, part i: algorithms and applications, Probabilistic Engineering Mechanics 19 (2004) 409 – 417.

[22] E. Ullmann, I. Papaioannou, Multilevel estimation of rare events, SIAM/ASA Journal on Uncertainty Quantification 3 (2015) 922–953.

[23] B. Peherstorfer, B. Kramer, K. Willcox, Multifidelity preconditioning of the cross-entropy method for rare event simulation and failure probability estimation, SIAM/ASA Journal on Uncertainty Quantification 6 (2018) 737–761.

[24] W.-C. Hsu, J. Ching, Evaluating small failure probabilities of multiple limit states by parallel subset simulation, Probabilistic Engineering Mechanics 25 (2010) 291 – 304.

[25] H.-S. Li, Y.-Z. Ma, Z. Cao, A generalized subset simulation approach for estimating small failure probabilities of multiple stochastic responses, Computers & Structures 153 (2015) 239 – 251.

[26] S. Bansal, S. H. Cheung, On the evaluation of multiple failure probability curves in reliability analysis with multiple performance functions, Reliability Engineering & System Safety 167 (2017) 583 – 594. Special Section: Applications of Probabilistic Graphical Models in Dependability, Diagnosis and Prognosis.

[27] L. W. Ng, K. E. Willcox, Monte Carlo information-reuse approach to aircraft conceptual design optimization under uncertainty, Journal of Aircraft 53 (2015) 427–438.

[28] M. Ehre, I. Papaioannou, D. Straub, Efficient conditional reliability updating with sequential importance sampling, PAMM 18 (2018) e201800282.

[29] A. Chaudhuri, B. Kramer, K. E. Willcox, Information reuse for importance sampling in reliability-based design optimization, Reliability Engineering & System Safety (2020) 106853.

[30] R. Y. Rubinstein, Optimization of computer simulation models with rare events, European Journal of Operational Research 99 (1997) 89–112.

[31] I. Papaioannou, S. Geyer, D. Straub, Improved cross entropy-based importance sampling with a flexible mixture model, Reliability Engineering & System Safety 191 (2019) 106564.

[32] M. Rosenblatt, Remarks on a multivariate transformation, The annals of mathematical statistics 23 (1952) 470–472.

[33] V. Dubourg, B. Sudret, F. Deheeger, Metamodel-based importance sampling for structural reliability analysis, Probabilistic Engineering Mechanics 33 (2013) 47 – 57.

[34] I. Papaioannou, K. Breitung, D. Straub, Reliability sensitivity estimation with sequential importance sampling, Structural Safety 75 (2018) 24 – 34.

[35] S. Geyer, I. Papaioannou, D. Straub, Cross entropy-based importance sampling using gaussian densities revisited, Structural Safety 76 (2019) 15–27.

[36] S. Au, J. Beck, Important sampling in high dimensions, Structural Safety 25 (2003) 139 – 163.

[37] L. Katafygiotis, K. Zuev, Geometric insight into the challenges of solving high-dimensional reliability problems, Probabilistic Engineering Mechanics 23 (2008) 208 – 218. 5th International Conference on Computational Stochastic Mechanics.

[38] Z. Wang, J. Song, Cross-entropy-based adaptive importance sampling using von mises-fisher mixture for high dimensional reliability analysis, Structural Safety 59 (2016) 42–52.

[39] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, AAAI Press, 1996, pp. 226–231.

[40] K. Beyer, J. Goldstein, R. Ramakrishnan, U. Shaft, When is "nearest neighbor" meaningful?, in: C. Beeri, P. Buneman (Eds.), Database Theory — ICDT'99, Springer Berlin Heidelberg, Berlin, Heidelberg, 1999, pp. 217–235.

[41] A. Owen, Y. Zhou, Safe and effective importance sampling, Journal of the American Statistical Association 95 (2000) 135–143.

[42] G. Bolch, S. Greiner, H. Meer, K. Trivedi, Queuing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications, 2006.

[43] I. Depina, I. Papaioannou, D. Straub, G. Eiksund, Coupling the cross-entropy with the line sampling method for risk-based design optimization, Structural and Multidisciplinary Optimization 55 (2017) 1589–1612.

[44] S.-H. Jiang, I. Papaioannou, D. Straub, Bayesian updating of slope reliability in spatially variable soils with in-situ measurements, Engineering Geology 239 (2018) 310 – 320.

[45] K.-K. Phoon, F. H. Kulhawy, Characterization of geotechnical variability, Canadian Geotechnical Journal 36 (1999) 612–624.

[46] K.-K. Phoon, F. H. Kulhawy, Evaluation of geotechnical property variability, Canadian Geotechnical Journal 36 (1999) 625–639.

[47] R. Rackwitz, Reviewing probabilistic soils modelling, Computers and Geotechnics 26 (2000) 199 – 223.

[48] Z. Cao, Y. Wang, D. Li, Quantification of prior knowledge in geotechnical site characterization, Engineering Geology 203 (2016) 107 – 116.

[49] I. Papaioannou, M. Ehre, D. Straub, PLS-based adaptation for efficient PCE representation in high dimensions, Journal of Computational Physics 387 (2019) 186 – 204.

[50] M. Ehre, I. Papaioannou, D. Straub, Global sensitivity analysis in high dimensions with PLS-PCE, Reliability Engineering & System Safety 198 (2020) 106861.

[51] I. Papaioannou, M. Ehre, D. Straub, Efficient PCE representations for reliability analysis in high dimensions, in: J. Song (Ed.), Proceedings of the 19th working conference of the IFIP Working Group 7.5 on Reliability and Optimization of Structural Systems, ETH Zürich, 2018.

[52] J. Zhang, A. Taflanidis, Bayesian posterior sampling using a metamodel–based sequential approach, in: J. Song (Ed.), Proceedings of the 19th working conference of the IFIP Working Group 7.5 on Reliability and Optimization of Structural Systems, ETH Zürich, 2018.