Cross entropy-based importance sampling for first-passage probability estimation of randomly excited linear structures with parameter uncertainty

Oindrila Kanjilal*, Iason Papaioannou, Daniel Straub

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

Abstract

We propose an efficient importance sampling (IS) method for estimating the first-passage probability of linear structures with uncertain parameters and subjected to Gaussian process excitations. The method evaluates the reliability through integrating the conditional first-passage probability given the uncertain structural parameters. We develop an adaptive IS strategy to efficiently perform this integration based on an IS density that is constructed using the cross entropy (CE) method. The CE method determines the IS density by adaptively minimizing the Kullback-Leibler divergence between the theoretically optimal sampling density and a chosen parametric family of probability distributions. The CE optimization problem is solved for a series of target densities that gradually approach the optimal IS density of the structural parameters. To define the intermediate densities, a smoothening of the conditional first-passage probabilities is employed. Once the IS density of the uncertain structural parameters is obtained, an effective IS density of the random excitations conditional on the structural parameters is introduced to estimate the failure probability of the structure. Unlike other tailored methods for solving this problem, the proposed IS approach does not require any prior analysis of the dynamic system and can be applied as a black-box method. Numerical examples demonstrate that the proposed method can calculate the first-passage probability with remarkable efficiency.

Keywords: First-passage probability, Uncertain linear structure, Gaussian excitation, Importance sampling, Cross entropy method

1. Introduction

Reliability assessment of dynamically excited structures, involving structural parameter uncertainties as well as randomness in the external excitation, has gained increasing attention. In structural dynamic systems, failure is usually defined by the first passage of a

*Corresponding author

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

response quantity over a prescribed threshold. This paper focuses on the problem of estimating the first-passage probability for uncertain linear structures subjected to Gaussian process excitations.

In early studies of first-passage problems, the reliability was evaluated using either out-crossing theory based on Rice's formula or diffusion theory through a numerical solution of the Kolmogorov equation [30, 31]. The potential of these methods in dealing with systems involving uncertain structural parameters and random excitations has been investigated in [48, 34, 4, 17]. The solution based on out-crossing theory is approximate, as it is based on heuristic assumptions on the properties of the out-crossing event. The difficulty in solving the multi-dimensional Kolmogorov equation makes the diffusion theory method applicable to only small size structural systems. Some recent research efforts focus on the development of the probability density evolution method for estimating the first-passage probability of dynamical systems [9, 29, 49]. When the number of random variables is large, which is often the case for the problem at hand, the application of this method is not straightforward.

In contrast, Monte Carlo methods are more robust and represent a powerful alternative to the aforementioned approaches. The direct Monte Carlo (DMC) estimator, which is universally applicable, has a sampling coefficient of variation (c.o.v.) that is inversely proportional to the sample size. When the probability of failure is small, the method requires a large computational effort to estimate the reliability with acceptable confidence. In view of this, several advanced sampling techniques have been developed for reducing the sampling c.o.v. of the probability of failure estimator. Approaches applicable to general dynamic systems with both parameter uncertainties and randomness in the excitation include the standard subset simulation method [3, 11, 5], spherical subset simulation [21, 22] and the line sampling method [24, 39]. A comprehensive review of such variance reduction methods can be found in [43, 16].

When the dynamical system is linear, more efficient sampling techniques can be designed by taking advantage of the linear relationship between the structural response and the applied loading. For the particular case where the structural parameters are characterized as deterministic and the excitation is modeled by a Gaussian random process, alternative approaches to efficiently estimate the failure probability are proposed in [2, 38]. These methods have been extended to deal with structures involving parameter uncertainties. Thereby, the first-passage probability conditional on a specific value of the uncertain parameters is evaluated according to the procedures in [2, 20, 38]. The unconditional first-passage probability is then computed through integration of the conditional probability over the space of uncertain structural parameters using importance sampling (IS) [19, 45] or line sampling [38]. These methods are non-adaptive in nature and require system-specific information to facilitate the probability estimation. In the methods suggested in [19, 38], a psuedo-design point with respect to the uncertain structural parameters needs to be identified. In [19] the design point is used to generate the sampling density and to approximate the system response in terms of the uncertain structural parameters to alleviate computational time. In the approach based on line sampling [38], the design point provides the reference direction for generating random lines in the uncertain parameter space to efficiently explore the failure domain. The aforementioned methods are thus effective if the important region

contributing to the failure probability lies in the vicinity of a unique design point, and if the design point is easy to identify. When the failure domain has multiple regions of importance, an appropriate reference point is often difficult to identify. In the IS method presented in [45], the construction of the IS density is based on a surrogate function used to represent the conditional instantaneous failure probabilities, i.e. probabilities of out-crossing at specific time instants. The performance of the method thus depends on a proper choice of the surrogate model, whose determination is not straightforward when the number of uncertain parameters is large, or the dependence of the conditional instantaneous failure probabilities on these parameters is strongly non-linear.

The aim of this contribution is to develop an adaptive IS approach for estimating the first-passage probability of uncertain linear structures subjected to Gaussian random processes. In the proposed method, the randomness in the external excitation is tackled using the IS strategy presented in [2]. The main contribution lies in the introduction of a novel approach to obtain a near-optimal IS density related to the uncertain structural parameters, which is based on the cross-entropy (CE) method. In the CE method, an efficient sampling density is constructed by minimizing the Kullback-Leibler (KL) divergence between the optimal IS density of the structural parameters and a chosen parametric family of probability distributions through a few rounds of small-size pre-sampling [41]. In the initial sampling steps, we solve the CE optimization for a sequence of target densities. To choose the intermediate densities for the first-passage problem, a smoothing parameter for the first-passage probability conditional on the uncertain structural parameters is introduced. The parameter value is updated adaptively in each sampling iteration to ensure a smooth transition to the optimal IS density. In terms of robustness, the proposed method can be used as a black-box algorithm as it is completely adaptive and does not require any prior investigations of the dynamical system. Therefore, in comparison to the methods suggested in [19, 38, 45], the approach developed in this study is more generally applicable to any linear dynamic system. The performance of the proposed method is demonstrated through numerical examples in section 5.

2. Problem formulation

2.1. Linear Dynamic System

Consider a linear structural system comprising n degrees of freedom. The state of the structure under dynamic load is described by a mass matrix $\mathbf{M}$, damping matrix $\mathbf{C}$ and stiffness matrix $\mathbf{K}$, each of dimension $n \times n$. Let $\boldsymbol{\Theta} = \{\Theta_1; \ldots; \Theta_{n_{\boldsymbol{\theta}}}\}$ be a vector of basic random variables that model the uncertain structural parameters describing $\mathbf{M}$, $\mathbf{C}$ and $\mathbf{K}$. Consider an l-dimensional load vector $\boldsymbol{f}(t) = \{f_1(t); \ldots; f_l(t)\}$ acting on the structure over a time duration $[0, T]$. It is assumed that the components of $\boldsymbol{f}(t)$ are independent Gaussian random processes. Then one can characterize $\boldsymbol{f}(t)$ as a linear function of independent and identically distributed standard normal variables, which we collectively express as $\boldsymbol{\Xi} = \{\Xi_1; \ldots; \Xi_{n_{\boldsymbol{\xi}}}\}$. The mean of the input random processes is taken to be zero, without loss of

generality. The equation of motion of the structure subjected to the stochastic excitation $\boldsymbol{f}(t, \boldsymbol{\Xi})$ is given by

$$\mathbf{M}(\boldsymbol{\Theta})\ddot{\boldsymbol{X}}(t) + \mathbf{C}(\boldsymbol{\Theta})\dot{\boldsymbol{X}}(t) + \mathbf{K}(\boldsymbol{\Theta})\boldsymbol{X}(t) = \mathbf{D}\boldsymbol{f}(t, \boldsymbol{\Xi}). \tag{1}$$

This semi-discretized equation of motion can be obtained, for instance, from spatial discretization of the continuum model of the structure by the finite element method. In Eq. (1), $\ddot{\boldsymbol{X}}$, $\dot{\boldsymbol{X}}$ and $\boldsymbol{X}$ denote the acceleration, velocity and displacement vectors of dimension $n \times 1$, and $\mathbf{D}$ is an $n \times l$ matrix that couples the random excitation $\boldsymbol{f}(t, \boldsymbol{\Xi})$ with the degrees of freedom of the structure. Without loss of generality, zero initial conditions are assumed at $t = 0$.

Let $h(t, \boldsymbol{\Theta}, \boldsymbol{\Xi})$ denote a critical response that is of interest, e.g. displacements, strains, accelerations, inter-storey drifts, etc. The structure is considered safe if the response $h(t, \boldsymbol{\Theta}, \boldsymbol{\Xi})$ fulfills certain performance criteria. For a particular value of the structural parameters $\boldsymbol{\Theta} = \boldsymbol{\theta}$, it is assumed that the relationship between the input excitation $\boldsymbol{f}(t, \boldsymbol{\Xi})$ and the response $h(t, \boldsymbol{\theta}, \boldsymbol{\Xi})$ is linear. Therefore, the input-output relation can be written in terms of a convolution integral as

$$h(t, \boldsymbol{\theta}, \boldsymbol{\Xi}) = \sum_{j=1}^{l} \int_0^t K_j(t - \tau; \boldsymbol{\theta}) f_j(\tau, \boldsymbol{\Xi}) d\tau = \int_0^t \boldsymbol{K}^{\mathrm{T}}(t - \tau; \boldsymbol{\theta}) \boldsymbol{f}(\tau, \boldsymbol{\Xi}) d\tau, \tag{2}$$

where $K_j(t; \boldsymbol{\theta})$ is the response at time t due to a unit impulse applied at the j-th input at time $t = 0$. It requires l dynamic analyses to obtain the whole set of impulse response functions $\{K_j(t; \boldsymbol{\theta}), j = 1, \ldots, l\}$, which completely define the input-output relationship. If the input excitation $\boldsymbol{f}(t)$ consists of filtered non-white processes, an augmented structural model comprising of the original structural system and the filters is considered, and K_j is taken as the convolution of the impulse response function of the original structural system and the filter producing the j-th component of $\boldsymbol{f}(t)$. The dynamic response is then computed by convoluting the impulse response functions of the augmented linear system with the white noise excitations applied to the filters.

In practical applications, often a discrete-time formulation is adopted, wherein the dynamic response of the structure is calculated by numerical integration using the values of the input at the discrete time instants. Let $\{t_1, \ldots, t_{n_T}\}$ be a set of discrete time points at a uniform time spacing $\Delta t = T/(n_T - 1)$ over the duration $[0, T]$. Let $\{\boldsymbol{f}(\boldsymbol{\Xi}, t_1), \ldots, \boldsymbol{f}(\boldsymbol{\Xi}, t_{n_T})\}$ denote the stochastic excitation in discrete time. One can represent $\boldsymbol{f}(\boldsymbol{\Xi}, t_k)$ by a linear combination of $\boldsymbol{\Xi}$ as $\boldsymbol{f}(\boldsymbol{\Xi}, t_k) = \mathbf{G}_k \boldsymbol{\Xi}$, where $\{\mathbf{G}_k, k = 1, \ldots, n_T\}$ are appropriate deterministic matrices of dimension $l \times n_{\boldsymbol{\xi}}$. Then, by analogy with Eq. (2), the input-output relationship is given by

$$h(t_k, \boldsymbol{\theta}, \boldsymbol{\Xi}) = \sum_{s=1}^{k} c_s \boldsymbol{K}^{\mathrm{T}}(t_k - t_s; \boldsymbol{\theta}) \boldsymbol{f}(t_s, \boldsymbol{\Xi}) \Delta t = \boldsymbol{r}_k^{\mathrm{T}}(\boldsymbol{\theta}) \boldsymbol{\Xi}, \tag{3}$$

where $\boldsymbol{r}_k^{\mathrm{T}}(\boldsymbol{\theta}) = \sum_{s=1}^{k} c_s \boldsymbol{K}^{\mathrm{T}}(t_k - t_s; \boldsymbol{\theta}) \mathbf{G}_s \Delta t$ and c_s is a coefficient that depends on the particular numerical integration scheme used to integrate Eq. (2).

2.2. First-passage failure probability

Consider a safe domain D_S for the critical response $h(t, \boldsymbol{\Theta}, \boldsymbol{\Xi})$. In first passage problems, the structure is regarded as safe if the response remains within the domain D_S over the interval $[0, T]$. Hence, the safe event is expressed as

$$S = \{\boldsymbol{\theta} \in \mathbb{R}^{n_{\boldsymbol{\theta}}}, \boldsymbol{\xi} \in \mathbb{R}^{n_{\boldsymbol{\xi}}} : h(t_k, \boldsymbol{\theta}, \boldsymbol{\xi}) \in D_S \; \forall k \in \{1, \ldots, n_T\}\}. \tag{4}$$

Whenever a response trajectory exits the safe domain, failure takes place. The first-passage failure event F is therefore the complement of S.

For a single-sided barrier specified by a response threshold h^*, F is defined as

$$F = \left\{\boldsymbol{\theta} \in \mathbb{R}^{n_{\boldsymbol{\theta}}}, \boldsymbol{\xi} \in \mathbb{R}^{n_{\boldsymbol{\xi}}} : \max_{k=1,\ldots,n_T} h(t_k, \boldsymbol{\theta}, \boldsymbol{\xi}) \geq h^*\right\}. \tag{5}$$

In the case of a double-sided barrier $\pm h^*$, F is the union of two disjoint events. The first one corresponds to an up-crossing of the response over the threshold h^*, the second one corresponds to a down-crossing of the response below the threshold $-h^*$. This leads to the following definition of F:

$$\begin{aligned} F = &\left\{\boldsymbol{\theta} \in \mathbb{R}^{n_{\boldsymbol{\theta}}}, \boldsymbol{\xi} \in \mathbb{R}^{n_{\boldsymbol{\xi}}} : \max_{k=1,\ldots,n_T} h(t_k, \boldsymbol{\theta}, \boldsymbol{\xi}) \geq h^*\right\} \\ &\bigcup \left\{\boldsymbol{\theta} \in \mathbb{R}^{n_{\boldsymbol{\theta}}}, \boldsymbol{\xi} \in \mathbb{R}^{n_{\boldsymbol{\xi}}} : \min_{k=1,\ldots,n_T} h(t_k, \boldsymbol{\theta}, \boldsymbol{\xi}) \leq -h^*\right\} \\ = &\left\{\boldsymbol{\theta} \in \mathbb{R}^{n_{\boldsymbol{\theta}}}, \boldsymbol{\xi} \in \mathbb{R}^{n_{\boldsymbol{\xi}}} : \max_{k=1,\ldots,n_T} |h(t_k, \boldsymbol{\theta}, \boldsymbol{\xi})| \geq h^*\right\}, \end{aligned} \tag{6}$$

where $|\cdot|$ denotes absolute value. Following Eqs. (5) and (6), the probability of first-passage failure is defined via a multi-dimensional integral of the form

$$P_F = \int_{\boldsymbol{\theta} \in \mathbb{R}^{n_{\boldsymbol{\theta}}}} \int_{\boldsymbol{\xi} \in \mathbb{R}^{n_{\boldsymbol{\xi}}}} \mathrm{I}\{(\boldsymbol{\theta}, \boldsymbol{\xi}) \in F\} p_{\boldsymbol{\Xi}}(\boldsymbol{\xi}) p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) d\boldsymbol{\xi} d\boldsymbol{\theta}. \tag{7}$$

Here $p_{\boldsymbol{\Xi}}(\boldsymbol{\xi})$ and $p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$, respectively, denote the joint probability density function (PDF) of $\boldsymbol{\Xi}$ and $\boldsymbol{\Theta}$, and $\mathrm{I}\{(\boldsymbol{\theta}, \boldsymbol{\xi}) \in F\}$ is the indicator function for the failure event which takes the value 1 if $(\boldsymbol{\theta}, \boldsymbol{\xi}) \in F$ and is 0 otherwise. The above probability can be estimated by direct Monte Carlo (DMC) simulation. However, it is well known that the number of samples needed by the DMC estimator to achieve a desired coefficient of variation (c.o.v.) is inversely proportional to the magnitude of P_F. Therefore, when the probability of failure is small, which is typically the case in engineering applications, this approach requires an excessively large number of samples to provide accurate estimates. A standard approach to reduce the sampling variance of the DMC estimator is importance sampling (IS).

The basic idea in IS is to draw samples of $\boldsymbol{\Theta}$ and $\boldsymbol{\Xi}$ from an importance sampling density $h_{\boldsymbol{\Theta},\boldsymbol{\Xi}}(\boldsymbol{\theta},\boldsymbol{\xi})$. Accordingly, the reliability integral in Eq. (7) is modified to

$$P_F = \int_{\boldsymbol{\theta}\in\mathbb{R}^{n_{\boldsymbol{\theta}}}} \int_{\boldsymbol{\xi}\in\mathbb{R}^{n_{\boldsymbol{\xi}}}} \mathrm{I}\{(\boldsymbol{\theta},\boldsymbol{\xi}) \in F\} \frac{p_{\boldsymbol{\Xi}}(\boldsymbol{\xi}) p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}{h_{\boldsymbol{\Theta},\boldsymbol{\Xi}}(\boldsymbol{\theta},\boldsymbol{\xi})} h_{\boldsymbol{\Theta},\boldsymbol{\Xi}}(\boldsymbol{\theta},\boldsymbol{\xi}) d\boldsymbol{\xi} d\boldsymbol{\theta}. \tag{8}$$

An IS estimator of P_F based on the above integral is given by

$$\hat{P}_F = \frac{1}{N} \sum_{i=1}^{N} \mathrm{I}\{(\boldsymbol{\theta}^i,\boldsymbol{\xi}^i) \in F\} \frac{p_{\boldsymbol{\Xi}}(\boldsymbol{\xi}^i) p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}^i)}{h_{\boldsymbol{\Theta},\boldsymbol{\Xi}}(\boldsymbol{\theta}^i,\boldsymbol{\xi}^i)}, \tag{9}$$

where $\{(\boldsymbol{\theta}^i,\boldsymbol{\xi}^i), i = 1,\ldots,N\}$ are independent samples distributed according to $h_{\boldsymbol{\Theta},\boldsymbol{\Xi}}(\boldsymbol{\theta},\boldsymbol{\xi})$. The above estimator is unbiased provided the support of $h_{\boldsymbol{\Theta},\boldsymbol{\Xi}}(\boldsymbol{\theta},\boldsymbol{\xi})$ contains the failure domain. The c.o.v. of the IS estimator depends not only on the number of samples, but also on the IS density. Hence, an appropriate selection of $h_{\boldsymbol{\Theta},\boldsymbol{\Xi}}(\boldsymbol{\theta},\boldsymbol{\xi})$ can lead to significantly smaller c.o.v. compared to that of the DMC estimator. It is straightforward to show that the sampling variance becomes zero when

$$h_{\boldsymbol{\Theta},\boldsymbol{\Xi}}(\boldsymbol{\theta},\boldsymbol{\xi}) = h^*_{\boldsymbol{\Theta},\boldsymbol{\Xi}}(\boldsymbol{\theta},\boldsymbol{\xi}) = \frac{1}{P_F} \mathrm{I}\{(\boldsymbol{\theta},\boldsymbol{\xi}) \in F\} p_{\boldsymbol{\Xi}}(\boldsymbol{\xi}) p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}), \tag{10}$$

where $h^*_{\boldsymbol{\Theta},\boldsymbol{\Xi}}(\boldsymbol{\theta},\boldsymbol{\xi})$ represents the optimal choice of the IS density. However, this optimal IS density cannot be used in practice because it requires prior knowledge of P_F. The main challenge in the application of IS schemes therefore lies in the selection of a sub-optimal IS density $h_{\boldsymbol{\Theta},\boldsymbol{\Xi}}(\boldsymbol{\theta},\boldsymbol{\xi})$ that is a good approximation to $h^*_{\boldsymbol{\Theta},\boldsymbol{\Xi}}(\boldsymbol{\theta},\boldsymbol{\xi})$.

In order to construct an effective IS density for the first-passage problem, we express the probability of failure in Eq. (7) alternatively as

$$P_F = \int_{\boldsymbol{\theta}\in\mathbb{R}^{n_{\boldsymbol{\theta}}}} P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta}) p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{11}$$

where

$$P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta}) = \int_{\boldsymbol{\xi}\in\mathbb{R}^{n_{\boldsymbol{\xi}}}} \mathrm{I}\{(\boldsymbol{\theta},\boldsymbol{\xi}) \in F\} p_{\boldsymbol{\Xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} \tag{12}$$

is the first-passage probability conditional on $\boldsymbol{\Theta} = \boldsymbol{\theta}$. One can evaluate this conditional probability of failure using existing methods for first-passage probability estimation of deterministic structures. In particular, an efficient IS density that allows estimation of $P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta})$ by IS has been developed in [2]. Therefore, in formulating an IS strategy to estimate P_F, the key task is the design of an IS density related to the uncertain structural parameters that would enable efficient estimation of the integral in Eq. (11). In Sec. 3 we briefly review two existing approaches to estimate the conditional first-passage probability $P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta})$. Subsequently, in Sec. 4 we develop a novel IS procedure for evaluating Eq. (11) and discuss the resulting IS estimator for the first-passage probability P_F.

3. Conditional first-passage failure probability

In this section, the first-passage problem is discussed for the case where the vector $\boldsymbol{\Theta}$ modeling structural parameter uncertainties assumes a known value $\boldsymbol{\Theta} = \boldsymbol{\theta}$. The setting corresponds to the case of a deterministic structure subjected to Gaussian process excitation. As discussed before, the input excitation is represented by a linear function of independent standard normal random variables $\boldsymbol{\Xi} = \{\Xi_1; \ldots; \Xi_{n_{\boldsymbol{\xi}}}\}$. The conditional first-passage probability is defined as the probability that the critical response $h(t, \boldsymbol{\theta}, \boldsymbol{\Xi})$ out-crosses at least once the threshold h^* during the time span $[0, T]$. Consider the discrete-time system defined in Eq. (3). Let $F_k(\boldsymbol{\theta}), k \in \{1, \ldots, n_T\}$ denote the event of out-crossing the threshold at the k-th time instant. The instantaneous out-crossing events are termed as 'elementary failure events' [2]. Occurrence of any one of the elementary failure events represents failure of the structure. Therefore, the first-passage failure event conditional on $\boldsymbol{\Theta} = \boldsymbol{\theta}$ is defined as

$$F(\boldsymbol{\theta}) = \bigcup_{k=1}^{n_T} F_k(\boldsymbol{\theta}). \tag{13}$$

Let $F_k^+(\boldsymbol{\theta})$ and $F_k^-(\boldsymbol{\theta})$, respectively, denote the up-crossing and down-crossing events at the k-th time step, i.e. $F_k^+(\boldsymbol{\theta}) = \{\boldsymbol{\xi} \in \mathbb{R}^{n_{\boldsymbol{\xi}}} : h(t_k, \boldsymbol{\theta}, \boldsymbol{\xi}) \geq h^*\}$ and $F_k^-(\boldsymbol{\theta}) = \{\boldsymbol{\xi} \in \mathbb{R}^{n_{\boldsymbol{\xi}}} : h(t_k, \boldsymbol{\theta}, \boldsymbol{\xi}) \leq -h^*\}$. Then, for the single- and double-barrier problems defined in Eqs. (5) and (6), the elementary failure events are given by $F_k(\boldsymbol{\theta}) = F_k^+(\boldsymbol{\theta})$ and $F_k(\boldsymbol{\theta}) = F_k^+(\boldsymbol{\theta}) \cup F_k^-(\boldsymbol{\theta})$, respectively. It is noted that the events $F_k^+(\boldsymbol{\theta})$ and $F_k^-(\boldsymbol{\theta})$ are mutually exclusive. Due to the linear relationship between $h(t_k, \boldsymbol{\theta}, \boldsymbol{\Xi})$ and $\boldsymbol{\Xi}$ given in Eq. (3), the boundaries of the failure domains $F_k^+(\boldsymbol{\theta})$ and $F_k^-(\boldsymbol{\theta})$ are hyperplanes in the $n_{\boldsymbol{\xi}}$-dimensional space of $\boldsymbol{\Xi}$. The probability content of $F_k^+(\boldsymbol{\theta})$ and $F_k^-(\boldsymbol{\theta})$ is given by

$$\mathrm{P}\left[F_k^+(\boldsymbol{\theta})\right] = \mathrm{P}\left[F_k^-(\boldsymbol{\theta})\right] = \Phi\left(-\frac{h^*}{\|\boldsymbol{r}_k(\boldsymbol{\theta})\|}\right), \tag{14}$$

where $\boldsymbol{r}_k(\boldsymbol{\theta})$ is as defined in Eq. (3) and $\Phi(\cdot)$ is the standard normal cumulative distribution function (CDF).

3.1. Analytical approximation based on out-crossing theory

A classical approach to evaluate the probability of occurrence of $F(\boldsymbol{\theta})$ is based on the Poisson hypothesis for the number of out-crossings. Let $\eta(h^*; \boldsymbol{\theta}, 0, T)$ denote the number of of times the critical response $h(t, \boldsymbol{\theta}, \boldsymbol{\Xi})$ out-crosses a safe domain defined by the threshold h^*. If the individual out-crossing events $\{F_k(\boldsymbol{\theta}), k = 1, \ldots, n_T\}$ are assumed to be independent, the number of out-crossings $\eta(h^*; \boldsymbol{\theta}, 0, T)$ can be approximated by a Poisson random variable. Based on this approximation, an analytical solution of the conditional first-passage probability, for at-rest initial condition, is obtained as [32]

$$P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta}) = 1 - \mathrm{P}\left[\{\eta(h^*; \boldsymbol{\theta}, 0, T) = 0\}\right] = 1 - \exp\left(-\int_0^T \alpha(t; h^*, \boldsymbol{\theta}) dt\right), \tag{15}$$

where $\alpha(t; h^*, \boldsymbol{\theta})$ denotes the rate of out-crossing the safe domain at time t.

Let $\alpha^-(t; -h^*, \boldsymbol{\theta})$ and $\alpha^+(t; h^*, \boldsymbol{\theta})$ denote the rates of down- and up-crossings of the process $h(t, \boldsymbol{\theta}, \boldsymbol{\Xi})$ across the levels $-h^*$ and h^*, respectively. These level-crossing rates are given by the Rice formula [40]. The out-crossing rate across the safe domain is given as $\alpha(t; h^*, \boldsymbol{\theta}) = \alpha^+(t; h^*, \boldsymbol{\theta})$ for a single-sided barrier, and $\alpha(t; h^*, \boldsymbol{\theta}) = \alpha^-(t; -h^*, \boldsymbol{\theta}) + \alpha^+(t; h^*, \boldsymbol{\theta})$ for a double-sided barrier. For a linear system subjected to a zero mean Gaussian process excitation, the response process $h(t, \boldsymbol{\theta}, \boldsymbol{\Xi})$ is a Gaussian random process (this follows directly from Eq. (2)). Furthermore, due to zero initial condition, the response process has a zero mean. In this situation, $\alpha^-(t; -h^*, \boldsymbol{\theta}) = \alpha^+(t; h^*, \boldsymbol{\theta})$ holds.

For the discrete-time system in Eq. (3), the level-crossing rates at the discrete time points $\{t_k, k = 1, \ldots, n_T\}$ are given by the expression [44, 28]

$$\begin{aligned}\alpha^-(t_k; -h^*, \boldsymbol{\theta}) = \alpha^+(t_k; h^*, \boldsymbol{\theta}) = &\frac{\sigma_{Z_2}\sqrt{1-\rho_{Z_2Z_1}{}^2}}{\sigma_{Z_1}}\phi\left(\frac{h^*}{\sigma_{Z_1}}\right)\left\{\phi\left(\frac{\rho_{Z_2Z_1}h^*}{\sigma_{Z_1}\sqrt{1-\rho_{Z_2Z_1}{}^2}}\right)+\right.\\ &\left.\frac{\rho_{Z_2Z_1}h^*}{\sigma_{Z_1}\sqrt{1-\rho_{Z_2Z_1}{}^2}}\Phi\left(\frac{\rho_{Z_2Z_1}h^*}{\sigma_{Z_1}\sqrt{1-\rho_{Z_2Z_1}{}^2}}\right)\right\}.\end{aligned} \tag{16}$$

where Z_1 and Z_2 are random variables that, respectively, denote $h(t, \boldsymbol{\theta}, \boldsymbol{\Xi})$ and its time derivative at $t = t_k$, σ_{Z_1} and σ_{Z_2} are, respectively, the standard deviations of Z_1 and Z_2, $\rho_{Z_2Z_1}$ is the correlation coefficient between Z_1 and Z_2, and $\phi(\cdot)$ is the standard normal PDF. The statistics σ_{Z_1}, σ_{Z_2} and $\rho_{Z_2Z_1}$ are computed by direct analysis of Eq. (3). The conditional first-passage probability in Eq. (15) can then be evaluated through the discrete-time formulation as

$$P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta}) = 1 - \exp\left(-\sum_{k=2}^{n_T}\alpha(t_k; h^*, \boldsymbol{\theta})\Delta t\right). \tag{17}$$

Note that, although the Poisson approximation is a convenient way to calculate the conditional first-passage probability, it is based on the assumption of independent out-crossings, which is often not met by real-life response processes. An improved formula that approximately accounts for the dependence between the out-crossing events is proposed in [46]. The improvement is derived primarily for stationary narrow-band Gaussian processes.

3.2. Monte Carlo simulation based on Importance Sampling

In view of the limitations of the analytical approximation presented in the preceding section, we additionally consider an IS evaluation of the conditional first-passage probability. Therein an IS density $h_{\boldsymbol{\Xi}}(\boldsymbol{\xi}|\boldsymbol{\theta})$, conditional on $\boldsymbol{\Theta} = \boldsymbol{\theta}$, is introduced, and the probability in Eq. (12) is written as

$$P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta}) = \int_{\boldsymbol{\xi}\in\mathbb{R}^{n_{\boldsymbol{\xi}}}} \mathrm{I}\left\{(\boldsymbol{\theta}, \boldsymbol{\xi}) \in F\right\} \frac{p_{\boldsymbol{\Xi}}(\boldsymbol{\xi})}{h_{\boldsymbol{\Xi}}(\boldsymbol{\xi}|\boldsymbol{\theta})} h_{\boldsymbol{\Xi}}(\boldsymbol{\xi}|\boldsymbol{\theta}) d\boldsymbol{\xi}. \tag{18}$$

As mentioned before, the key challenge in the application of IS schemes lies in the selection of an appropriate IS density that guarantees low variability of the IS estimator. For the particular case of deterministic linear systems subjected to Gaussian process excitations, an efficient IS density is proposed in [2]. This IS density is expressed as a weighted sum of the conditional PDFs $p_{\boldsymbol{\Xi}}(\boldsymbol{\xi}|F_k(\boldsymbol{\theta})) = p_{\boldsymbol{\Xi}}(\boldsymbol{\xi})\mathrm{I}\{(\boldsymbol{\theta},\boldsymbol{\xi}) \in F_k(\boldsymbol{\theta})\}/P(F_k(\boldsymbol{\theta}))$, which are the joint PDFs of $\boldsymbol{\Xi}$ truncated on the elementary failure domains $F_k(\boldsymbol{\theta})$:

$$h_{\boldsymbol{\Xi}}(\boldsymbol{\xi}|\boldsymbol{\theta}) = \sum_{k=1}^{n_T} w_k(\boldsymbol{\theta}) p_{\boldsymbol{\Xi}}(\boldsymbol{\xi}|F_k(\boldsymbol{\theta})) = \sum_{k=1}^{n_T} w_k(\boldsymbol{\theta}) \frac{p_{\boldsymbol{\Xi}}(\boldsymbol{\xi})\mathrm{I}\{(\boldsymbol{\theta},\boldsymbol{\xi}) \in F_k(\boldsymbol{\theta})\}}{\mathrm{P}\left[F_k(\boldsymbol{\theta})\right]}. \tag{19}$$

In Eq. (19), $\{w_k(\boldsymbol{\theta}), k = 1, \ldots, n_T\}$ are the normalized weights associated with the elementary failure events. The weights are chosen to be proportional to $\mathrm{P}\left[F_k(\boldsymbol{\theta})\right]$, i.e. $w_k(\boldsymbol{\theta}) = \mathrm{P}\left[F_k(\boldsymbol{\theta})\right] / \sum_{j=1}^{n_T} \mathrm{P}\left[F_j(\boldsymbol{\theta})\right]$. For single- and double-sided barriers, the probability of occurrence of $F_k(\boldsymbol{\theta})$ is given by $\mathrm{P}\left[F_k(\boldsymbol{\theta})\right] = \mathrm{P}\left[F_k^+(\boldsymbol{\theta})\right]$ and $\mathrm{P}\left[F_k(\boldsymbol{\theta})\right] = \mathrm{P}\left[F_k^+(\boldsymbol{\theta})\right] + \mathrm{P}\left[F_k^-(\boldsymbol{\theta})\right]$, respectively, where $\mathrm{P}\left[F_k^+(\boldsymbol{\theta})\right]$ and $\mathrm{P}\left[F_k^-(\boldsymbol{\theta})\right]$ are as defined in Eq. (14). Upon substituting the expression of Eq. (19) for $h_{\boldsymbol{\Xi}}(\boldsymbol{\xi}|\boldsymbol{\theta})$ in Eq. (18), the following integral for the conditional first-passage probability is obtained

$$\begin{aligned} P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta}) =& \tilde{P}(\boldsymbol{\theta}) \int_{\boldsymbol{\xi}\in\mathbb{R}^{n_{\boldsymbol{\xi}}}} \frac{\mathrm{I}\{(\boldsymbol{\theta},\boldsymbol{\xi}) \in F(\boldsymbol{\theta})\}}{\sum_{k=1}^{n_T} \mathrm{I}\{(\boldsymbol{\theta},\boldsymbol{\xi}) \in F_k(\boldsymbol{\theta})\}} h_{\boldsymbol{\Xi}}(\boldsymbol{\xi}|\boldsymbol{\theta}) d\boldsymbol{\xi} \\ =& \tilde{P}(\boldsymbol{\theta}) \int_{\boldsymbol{\xi}\in\mathbb{R}^{n_{\boldsymbol{\xi}}}} \frac{1}{\sum_{k=1}^{n_T} \mathrm{I}\{(\boldsymbol{\theta},\boldsymbol{\xi}) \in F_k(\boldsymbol{\theta})\}} h_{\boldsymbol{\Xi}}(\boldsymbol{\xi}|\boldsymbol{\theta}) d\boldsymbol{\xi}. \end{aligned} \tag{20}$$

Here $\tilde{P}(\boldsymbol{\theta}) = \sum_{j=1}^{n_T} \mathrm{P}\left[F_j(\boldsymbol{\theta})\right]$ denotes the sum of the probabilities of the elementary failure events. The second equality in the above equation is due to the fact that $\mathrm{I}\{(\boldsymbol{\theta},\boldsymbol{\xi}) \in F(\boldsymbol{\theta})\} = 1$ for every sample $\boldsymbol{\xi}$ simulated according to $h_{\boldsymbol{\Xi}}(\boldsymbol{\xi}|\boldsymbol{\theta})$. Estimation of the conditional first-passage probability based on Eq. (20) requires samples of the vector $\boldsymbol{\Xi}$ distributed according to the IS density $h_{\boldsymbol{\Xi}}(\boldsymbol{\xi}|\boldsymbol{\theta})$. An algorithm for generating these samples is discussed in [2], and is provided in Appendix A for completeness.

4. First-passage probability considering random structural parameters

Estimation of the unconditional first-passage probability, considering both structural parameter uncertainties and random excitation, requires integration of the conditional failure probability $P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta})$ over the whole domain of the PDF $p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$. This leads to the integral $P_F = \int_{\boldsymbol{\theta}\in\mathbb{R}^{n_{\boldsymbol{\theta}}}} P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta}) p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) d\boldsymbol{\theta}$, introduced earlier in Eq. (11). When the dimension of $\boldsymbol{\Theta}$ is large, numerical integration is not feasible. In principle, one can estimate the integral by direct Monte Carlo sampling, wherein independent samples $\{(\boldsymbol{\theta}^i, i = 1, \ldots, N_S\}$ are generated from the PDF $p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$. The associated conditional first-passage probabilities $P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta}^i)$ are determined based on Eq. (20), and the unconditional first-passage probability P_F is

estimated as the sample average of $\left\{P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta}^i), i = 1, \ldots, N_S\right\}$. Such an approach is efficient only if the parameter uncertainty is low. For high parameter uncertainty, the variance of $P_{F|\boldsymbol{\Theta}}(\boldsymbol{\Theta})$ is large, and hence a large number of Monte Carlo samples are necessary to achieve acceptable accuracy.

For the purpose of estimating the first-passage probability more efficiently, we introduce an IS density $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ for $\boldsymbol{\Theta}$. The integral in Eq. (11) is modified to

$$P_F = \int_{\boldsymbol{\theta}\in\mathbb{R}^{n_{\boldsymbol{\theta}}}} P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta}) W(\boldsymbol{\theta}) h_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{21}$$

where $W(\boldsymbol{\theta}) = p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})/h_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ is the importance weight function. The density $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ should be tailored to generate more samples of $\boldsymbol{\Theta}$ from the 'important region', i.e. the region over which the value of $P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta})p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ is large, while ensuring a low variability of $W(\boldsymbol{\theta})$. The optimal IS density that guarantees this is given by the expression

$$h^*_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = \frac{1}{P_F} P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta}) p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}). \tag{22}$$

We use the CE method to construct an IS density $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ that is a close approximation of $h^*_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$.

4.1. Construction of importance sampling density for $\boldsymbol{\Theta}$ by the cross entropy method

The CE method is an adaptive approach that determines a near-optimal IS density by minimising the KL divergence between the optimal IS density $h^*_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ and a chosen parametric family of distributions. Define a family of parametric densities $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu})$, where $\boldsymbol{\nu} \in \mathcal{V}$ is the parameter vector. The KL divergence between $h^*_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ and $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu})$ is a measure of the difference between the two PDFs, and is defined as [41]

$$\begin{aligned} D(h^*_{\boldsymbol{\Theta}}(\boldsymbol{\theta}), h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu})) &= \mathrm{E}_{h^*_{\boldsymbol{\Theta}}}\left[\ln\left(\frac{h^*_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}{h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu})}\right)\right] \\ &= \mathrm{E}_{h^*_{\boldsymbol{\Theta}}}\left[\ln\left(h^*_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\right)\right] - \mathrm{E}_{h^*_{\boldsymbol{\Theta}}}\left[\ln\left(h_{\boldsymbol{\Theta}}\left(\boldsymbol{\theta};\boldsymbol{\nu}\right)\right)\right]. \end{aligned} \tag{23}$$

It holds that $D(h^*_{\boldsymbol{\Theta}}(\boldsymbol{\theta}), h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu})) \geq 0$, with $D(h^*_{\boldsymbol{\Theta}}(\boldsymbol{\theta}), h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu})) = 0$ if and only if $h^*_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu})$. The CE method aims at finding the parameters $\boldsymbol{\nu}^*$ that minimize $D(h^*_{\boldsymbol{\Theta}}(\boldsymbol{\theta}), h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu}))$, i.e. it solves the following optimization problem:

$$\boldsymbol{\nu}^* = \underset{\boldsymbol{q}\in\mathcal{V}}{\operatorname{argmin}}\ D(h^*_{\boldsymbol{\Theta}}(\boldsymbol{\theta}), h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\boldsymbol{q})) \tag{24}$$

Since the parametric density appears only in the second term in Eq. (23), the above equation is equivalent to

$$\boldsymbol{\nu}^* = \underset{\boldsymbol{q}\in\mathcal{V}}{\operatorname{argmax}}\ \mathrm{E}_{h^*_{\boldsymbol{\Theta}}}\left[\ln\left(h_{\boldsymbol{\Theta}}\left(\boldsymbol{\theta};\boldsymbol{q}\right)\right)\right]. \tag{25}$$

Substituting $h^*_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ in Eq. (25) with the expression of Eq. (22), one gets

$$\boldsymbol{\nu}^* = \underset{\boldsymbol{q}\in\mathcal{V}}{\operatorname{argmax}}\ \mathrm{E}_{p_{\boldsymbol{\Theta}}}\left[P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta})\ln\left(h_{\boldsymbol{\Theta}}\left(\boldsymbol{\theta};\boldsymbol{q}\right)\right)\right]. \tag{26}$$

The expectation in the above equation can be estimated using a set of samples $\{\boldsymbol{\theta}^i, i = 1,\ldots,N_{CE}\}$ from $p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$, which leads to the sample counterpart of the CE optimization problem:

$$\hat{\boldsymbol{\nu}}^* = \underset{\boldsymbol{q}\in\mathcal{V}}{\operatorname{argmax}}\ \frac{1}{N_{CE}}\sum_{i=1}^{N_{CE}} P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta}^i)\ln\left(h_{\boldsymbol{\Theta}}\left(\boldsymbol{\theta}^i;\boldsymbol{q}\right)\right). \tag{27}$$

Obtaining a good sample approximation of Eq.(26) with Eq.(27) requires a considerable number of samples in the high probability region of $h^*_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$, i.e. a large fraction of the samples $\{\boldsymbol{\theta}^i, i = 1,\ldots,N_{CE}\}$ drawn from $p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ should belong to the region where the value of $P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta})p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ is large. When P_F is small, this region has a small probability volume under $p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$. This is illustrated in Fig. 1. Moreover, if the variance of $P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta})$ is high, $h^*_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) \propto P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta})p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ differs significantly from $p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$. In these cases, solving Eq. (27) is impractical, because a large number of samples would be required to obtain a good sample approximation. To circumvent this problem, we solve the CE optimization using a multi-level approach. We introduce a sequence of intermediate target densities $h^k_{\boldsymbol{\Theta}}(\boldsymbol{\theta}), k = 0,\ldots,L$, that start from $p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ and gradually converge to the optimal IS density $h^*_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$. The CE optimization problem is solved sequentially for each intermediate target density, leading to a sequence of parameter vectors $\{\boldsymbol{\nu}^k, k = 1,\ldots,L\}$. The ultimate goal is to have $\boldsymbol{\nu}^L$ close to $\boldsymbol{\nu}^*$, so that the IS density $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu_L})$ can be used to estimate P_F. This multi-level approach has been extensively used within the CE method for rare-event simulation (see [41] for the general methodology, and [25, 47, 15, 37] for applications in structural reliability). In contrast to existing works, which utilize the CE method to estimate the expectation of the indicator function $\mathrm{I}\{\boldsymbol{\theta}\in F\}$ (the failure probability), we utilize the CE mehod to estimate the expectation of the (smooth) conditional probability $P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta})$.

To bridge the gap between the nominal density $p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ and the optimal IS density $h^*_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$, we define the intermediate target densities $\{h^k_{\boldsymbol{\Theta}}(\boldsymbol{\theta}), k = 0,\ldots,L\}$ as

$$h^k_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = \frac{1}{C_k}P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta})^{\gamma_k}p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}), \tag{28}$$

where $0 = \gamma_0 < \gamma_1 < \ldots < \gamma_L = 1$ and $C_k = \int_{\boldsymbol{\theta}\in\mathbb{R}^{n_{\boldsymbol{\theta}}}} P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta})^{\gamma_k}p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})d\boldsymbol{\theta}$ is the normalization constant of the k-th distribution in the sequence. Note that $h^0_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ and $h^L_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = h^*_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$. This distribution sequence, which is illustrated in Fig. 1, has been previously used to derive sequential Monte Carlo approaches for sampling posterior distributions and for estimating normalizing constants in the context of Bayesian analysis [33, 12, 7, 13, 10]. In Eq. (28), $\{\gamma_k, k = 1,\ldots,L\}$ are smoothing parameters that define the smooth transition between $p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ and $h^*_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$. The rationale is that although $h^*_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ can differ largely from $p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$, the change between two consecutive intermediate densities can be made small through

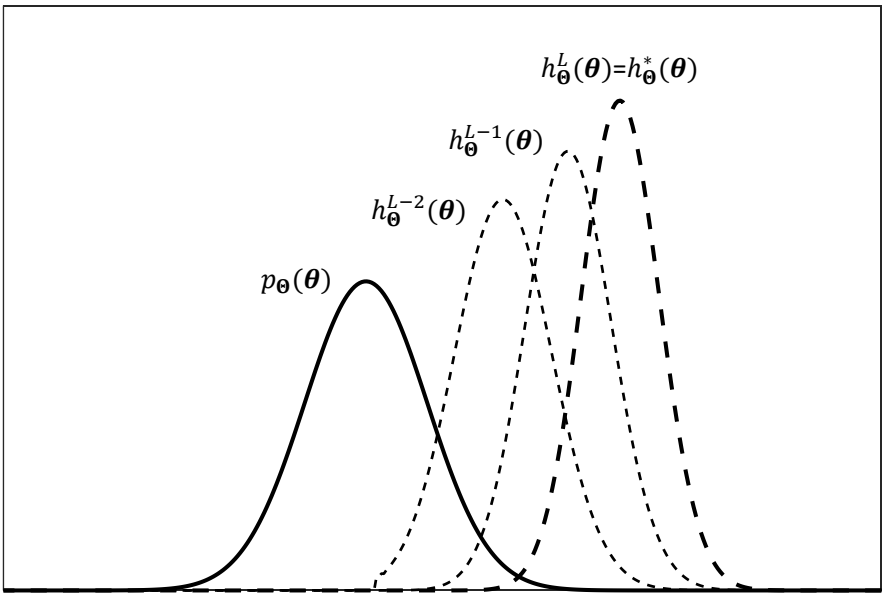

Figure 1: Representation of $p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ and $h^*_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$, where $h^*_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ peaks in the tail region of $p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$. The figure also shows the sequence of intermediate target densities that define a smooth transition from $p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ to $h^*_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ in the CE method.

an appropriate selection of the parameters $\{\gamma_k, k = 1, \ldots, L\}$. This small change makes it possible for the samples generated from $h^{k-1}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$, or a close approximation of $h^{k-1}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$, to lie in the region of high probability density of $h^k_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$.

In the k-th step of the multi-level approach, we determine the parameter vector $\boldsymbol{\nu}^k$ that minimizes the KL divergence between $h^k_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ and the parametric density $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \boldsymbol{\nu})$. The stochastic optimization problem to be solved at each intermediate level is thus given by

$$\begin{aligned} \boldsymbol{\nu}^k &= \underset{\boldsymbol{q}\in\mathcal{V}}{\text{argmin}}\ D(h^k_{\boldsymbol{\Theta}}(\boldsymbol{\theta}), h_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \boldsymbol{q})) \\ &= \underset{\boldsymbol{q}\in\mathcal{V}}{\text{argmax}}\ \mathrm{E}_{h^k_{\boldsymbol{\Theta}}}\left[\ln\left(h_{\boldsymbol{\Theta}}\left(\boldsymbol{\theta}; \boldsymbol{q}\right)\right)\right] \\ &= \underset{\boldsymbol{q}\in\mathcal{V}}{\text{argmax}}\ \mathrm{E}_{p_{\boldsymbol{\Theta}}}\left[P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta})^{\gamma_k} \ln\left(h_{\boldsymbol{\Theta}}\left(\boldsymbol{\theta}; \boldsymbol{q}\right)\right)\right]. \end{aligned} \tag{29}$$

The expectation in the objective function of Eq. (29) is estimated by IS, using a set of samples $\{\boldsymbol{\theta}^i, i = 1, \ldots, N_{CE}\}$ distributed according to $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \hat{\boldsymbol{\nu}}^{k-1})$, $\hat{\boldsymbol{\nu}}^{k-1}$ being the solution of the optimization in the previous step. Hence, the importance sampling counterpart of the optimization problem in Eq. (29) is given by

$$\hat{\boldsymbol{\nu}}^k = \underset{\boldsymbol{q}\in\mathcal{V}}{\text{argmax}}\ \frac{1}{N_{CE}} \sum_{i=1}^{N_{CE}} \widetilde{W}_k\left(\boldsymbol{\theta}^i, \hat{\boldsymbol{\nu}}^{k-1}\right) \ln\left(h_{\boldsymbol{\Theta}}\left(\boldsymbol{\theta}^i; \boldsymbol{q}\right)\right), \tag{30}$$

where $\widetilde{W}_k\left(\boldsymbol{\theta}, \hat{\boldsymbol{\nu}}^{k-1}\right) = P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta})^{\gamma_k} \frac{p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}{h_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \hat{\boldsymbol{\nu}}^{k-1})}$. A default choice for $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \hat{\boldsymbol{\nu}}^0)$ is the nominal

density of $\boldsymbol{\Theta}$, i.e. $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}^0) = p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$. Note that for $k \geq 2$, $\hat{\boldsymbol{\nu}}^{k-1}$ minimizes the KL divergence between $h_{\boldsymbol{\Theta}}^{k-1}(\boldsymbol{\theta})$ and $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu})$, and, therefore, $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}^{k-1})$ is a close approximation of $h_{\boldsymbol{\Theta}}^{k-1}(\boldsymbol{\theta})$. Typically, the objective functions in Eqs. (29) and (30) are convex and differentiable with respect to $\boldsymbol{q}$. Therefore, the optimization problem can be solved by setting the gradient of the objective function to zero.

To ensure that each density $h_{\boldsymbol{\Theta}}^{k}(\boldsymbol{\theta})$, and hence the objective function in the corresponding optimization problem, can be reasonably approximated using a limited number of samples drawn from $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}^{k-1})$, we select the parameter γ_k adaptively such that the sample c.o.v. $\hat{\delta}_{\widetilde{W}_k}$ of the weights $\left\{\widetilde{W}_k\left(\boldsymbol{\theta}^i,\hat{\boldsymbol{\nu}}^{k-1}\right), i = 1,\ldots,N_{CE}\right\}$ reaches a target value δ_{target}. Hence, at each intermediate sampling step one solves the optimization problem

$$\gamma_k = \underset{\gamma\in(\gamma_{k-1},1)}{\operatorname{argmin}} \left(\hat{\delta}_{\widetilde{W}_k}(\gamma) - \delta_{target}\right)^2. \tag{31}$$

We note that bounding the c.o.v. of the weights according to Eq. (31) is equivalent to requiring that the number of effective samples (ESS) used to fit the parametric model takes a target value [26]. The ESS is expressed in terms of the c.o.v. of the weights as ESS $= N_{CE}/\left(1+\hat{\delta}^2_{\widetilde{W}_k}(\gamma)\right)$.

The sequential procedure is stopped when the smoothing parameter for the target density at the current sampling step is equal to one. After convergence at the L-th step, the final parameter vector $\hat{\boldsymbol{\nu}}^L$ is determined by solving the optimization problem in Eq. (30) with $\gamma_L = 1$. The resulting density $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}^L)$ is the closest approximation of $h^*_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ for the chosen parametric model based on the samples drawn from $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}^{L-1})$, and it is taken as the IS density for estimating the first-passage probability. Bounding the sample c.o.v. of the weights according to Eq. (31) is equivalent to bounding the c.o.v. of the IS estimate of the normalizing constant of the respective distribution in the sequence of Eq. (28) [35, 36]. For example, choosing $\delta_{target} = 1.5$, as suggested in [35], corresponds to a target c.o.v. of the IS estimate of approximately 0.05 for $N_{CE} = 1000$. Therefore, the applied stopping criteria ensures that the c.o.v. of the IS estimate of the unconditional first-passage probability is smaller than a certain target value.

Determination of the IS density for $\boldsymbol{\Theta}$ based on the above described procedure requires repeated evaluations of the conditional first-passage probability $P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta})$. One could evaluate $P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta})$ by IS according to Eq. (20). However, for the parameter vectors obtained by solving Eq. (30) at each intermediate step to converge smoothly to the optimal parameter value, the sampling variance of the estimates of $P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta})$ should be small. To ensure this, the IS estimator of $P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta})$ will require a large number of samples, which would significantly increase the overall computational effort needed for CE optimization. Hence, we instead evaluate $P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta})$ analytically by the Poisson approximation stated in Eq. (17). The IS estimator of $P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta})$ is only applied once the final IS density for $\boldsymbol{\Theta}$ is obtained. The analytical approximation reduces the computational effort at the expense of accuracy, as the fitted IS density may be sub-optimal. However, numerical studies show that the IS density obtained based on this approach gives fairly accurate estimates of the unconditional

Algorithm 1: Determination of IS density for $\boldsymbol{\Theta}$ by the CE method

input:
- Sample size N_{CE}.
- Choice of parametric density $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu})$.
- Target c.o.v. of the weights at each intermediate step, δ_{target}.

initialization:
- Set $k=0$.
- Select $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}^k)$ as the nominal density $p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$.

repeat:
- Set $k=k+1$.
- Generate independent samples $\left\{\boldsymbol{\theta}^i, i=1,\ldots,N_{CE}\right\}$ from $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}^{k-1})$.
- Evaluate the conditional first-passage probabilities $\left\{P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta}^i), i=1,\ldots,N_{CE}\right\}$ based on the analytical approximation stated in Eq. (17).
- Compute the likelihood ratio $\left\{\frac{p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}^i)}{h_{\boldsymbol{\Theta}}(\boldsymbol{\theta}^i;\hat{\boldsymbol{\nu}}^{k-1})}, i=1,\ldots,N_{CE}\right\}$ for the random samples.
- Solve the optimization problem in Eq. (31) to determine γ_k.
  Note that the conditional first-passage probabilities and the likelihood ratios computed in the previous steps are used to evaluate the sample c.o.v. of the weights $\left\{\widetilde{W}_k\left(\boldsymbol{\theta}^i,\hat{\boldsymbol{\nu}}^{k-1}\right), i=1,\ldots,N_{CE}\right\}$. Further simulations are not needed in this step.
- Determine $\hat{\boldsymbol{\nu}}^k$ by solving the optimization problem in Eq. (30).

while $\gamma_k<1$

output:
- $L=k$, and $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}^L)=$ IS density for Θ.

first-passage failure probability. The resulting procedure for determining the IS density for $\boldsymbol{\Theta}$ based on the CE method is described in Algorithm 1.

4.2. Estimator for the unconditional first-passage failure probability

Using the IS density for $\boldsymbol{\Theta}$ introduced in the previous section, the unconditional first-passage probability P_F is obtained by computing the expectation of the conditional failiure probability:

$$\begin{aligned} P_F &= \int_{\boldsymbol{\theta}\in\mathbb{R}^{n_{\boldsymbol{\theta}}}} P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta}) p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= \int_{\boldsymbol{\theta}\in\mathbb{R}^{n_{\boldsymbol{\theta}}}} P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta}) \frac{p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}{h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}^L)} h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}^L) d\boldsymbol{\theta} \end{aligned} \tag{32}$$

Substitution of the integral expression for $P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta})$ in Eq. (20) into the above equation yields

$$P_F = \int_{\boldsymbol{\theta}\in\mathbb{R}^{n_{\boldsymbol{\theta}}}} \int_{\boldsymbol{\xi}\in\mathbb{R}^{n_{\boldsymbol{\xi}}}} \left\{ \frac{\tilde{P}(\boldsymbol{\theta})}{\sum_{k=1}^{n_T} \mathrm{I}\{(\boldsymbol{\theta},\boldsymbol{\xi}) \in F_k(\boldsymbol{\theta})\}} W(\boldsymbol{\theta}) \right\} h_{\boldsymbol{\Theta},\boldsymbol{\Xi}}(\boldsymbol{\theta},\boldsymbol{\xi}) d\boldsymbol{\xi} d\boldsymbol{\theta}, \tag{33}$$

where $W(\boldsymbol{\theta}) = \frac{p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}{h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}^L)}$ and $h_{\boldsymbol{\Theta},\boldsymbol{\Xi}}(\boldsymbol{\theta},\boldsymbol{\xi}) = h_{\boldsymbol{\Xi}}(\boldsymbol{\xi}|\boldsymbol{\theta}) h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}^L)$ is the joint IS density of $\boldsymbol{\Theta}$ and $\boldsymbol{\Xi}$. Here $h_{\boldsymbol{\Xi}}(\boldsymbol{\xi}|\boldsymbol{\theta})$ is the IS density for $\boldsymbol{\Xi}$ conditional on $\boldsymbol{\Theta} = \boldsymbol{\theta}$ defined in Eq. (19), and $\tilde{P}(\boldsymbol{\theta}) = \sum_{k=1}^{n_T} \mathrm{P}\left[F_k(\boldsymbol{\theta})\right]$ is the sum of the probabilities of the elementary failure events $\{F_k(\boldsymbol{\theta}), k = 1,\ldots,n_T\}$. As discussed in Sec. 3.2, $\{\mathrm{P}\left[F_k(\boldsymbol{\theta})\right], k = 1,\ldots,n_T\}$ can be calculated analytically using Eq. (14). The IS estimator for P_F is expressed as

$$\hat{P}_F = \frac{1}{N} \sum_{i=1}^{N} \frac{\tilde{P}(\boldsymbol{\theta}^i)}{\sum_{k=1}^{n_T} \mathrm{I}\{(\boldsymbol{\theta}^i,\boldsymbol{\xi}^i) \in F_k(\boldsymbol{\theta}^i)\}} W(\boldsymbol{\theta}^i), \tag{34}$$

where $\left\{\left(\boldsymbol{\theta}^i,\boldsymbol{\xi}^i\right), i = 1,\ldots,N\right\}$ are samples of the structural parameters and excitation distributed according to $h_{\boldsymbol{\Theta},\boldsymbol{\Xi}}(\boldsymbol{\theta},\boldsymbol{\xi}) = h_{\boldsymbol{\Xi}}(\boldsymbol{\xi}|\boldsymbol{\theta}) h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}^L)$. The variance estimate of $\hat{P}_F$ is [27]

$$\widehat{\mathrm{Var}}(\hat{P}_F) = \frac{1}{N-1} \left[\frac{1}{N} \sum_{i=1}^{N} \left(\frac{\tilde{P}(\boldsymbol{\theta}^i)}{\sum_{k=1}^{n_T} \mathrm{I}\{(\boldsymbol{\theta}^i,\boldsymbol{\xi}^i) \in F_k(\boldsymbol{\theta}^i)\}} W(\boldsymbol{\theta}^i) \right)^2 - \hat{P}_F^2 \right]. \tag{35}$$

The sample size N of the IS estimator determines the sampling c.o.v. of the probability of failure estimates. In the present study, two choices for N are investigated. In the first case, N is taken equal to N_{CE}, i.e. the number of sample employed in the intermediate steps of the CE optimization problem. In the second case, N is adapted on the fly to ensure that, in each independent simulation run, the estimator of the c.o.v. of $\hat{P}_F$, denoted by $\hat{\delta}_{\hat{P}_F}$, takes a value smaller than a prescribed threshold $\delta^*_{\hat{P}_F}$. $\hat{\delta}_{\hat{P}_F}$ is given by the expression $\hat{\delta}_{\hat{P}_F} = \hat{\mathrm{Var}}(\hat{P}_F)/\hat{P}_F$, and is computed after every sample increment. We observed that the variance of the estimator $\hat{\delta}_{\hat{P}_F}$ is large, especially when N is small. To obtain a robust convergence criterion, we check for convergence after every M_0 samples, wherein the average of the previous m_0 values of $\hat{\delta}_{\hat{P}_F}$, denoted by δ_{avg}, is compared with $\delta^*_{\hat{P}_F}$. This ensures that δ_{avg} has a gradually decreasing behaviour. Ideally, m_0 should be chosen smaller than or equal to M_0. In the present study we take $M_0 = 100$ and $m_0 = 25$. The procedure to estimate P_F using the adaptive approach is described in Algorithm 2.

Algorithm 2: Implementation of IS estimator for P_F with adaptive choice of N

input:

Nominal density $p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ and the IS density $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}^L)$ for $\boldsymbol{\Theta}$.

IS density $h_{\boldsymbol{\Xi}}(\boldsymbol{\xi}|\boldsymbol{\theta})$ for $\boldsymbol{\Xi}$ conditional on $\boldsymbol{\Theta}=\boldsymbol{\theta}$.

Number of sample increments M_0 after which convergence is checked.

The window length m_0 for averaging the values of $\hat{\delta}_{\hat{P}_F}$ to check convergence.

Target value of the c.o.v. of $\hat{P}_F$, $\delta^*_{\hat{P}_F}$.

initialization:

Set $i=0$.

repeat:

Set $i=i+1$.

Generate a sample $\boldsymbol{\theta}^i$ distributed according to $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\hat{\boldsymbol{\nu}}^L)$.

Generate $\boldsymbol{\xi}^i$ distributed according to $h_{\boldsymbol{\Xi}}(\boldsymbol{\xi}|\boldsymbol{\theta}^i)$ using the algorithm in Appendix A.

Compute the dynamic response $h\left(t,\boldsymbol{\theta}^i,\boldsymbol{\xi}^i\right)$ at the discrete time instants $\{t_k, k=1,\ldots,n_T\}$ using Eq. (3).

Evaluate the indicator functions $\left\{\mathrm{I}\left\{\left(\boldsymbol{\theta}^i,\boldsymbol{\xi}^i\right)\in F_k\left(\boldsymbol{\theta}^i\right)\right\}, k=1,\ldots,n_T\right\}$.

Compute the likelihood ratio $W\left(\boldsymbol{\theta}^i\right)=\frac{p_{\boldsymbol{\Theta}}\left(\boldsymbol{\theta}^i\right)}{h_{\boldsymbol{\Theta}}\left(\boldsymbol{\theta}^i;\hat{\boldsymbol{\nu}}^L\right)}$.

Compute $\tilde{P}\left(\boldsymbol{\theta}^i\right)=\sum_{k=1}^{n_T}\mathrm{P}\left[F_k\left(\boldsymbol{\theta}^i\right)\right]$, where $\mathrm{P}\left[F_k\left(\boldsymbol{\theta}^i\right)\right]$ is evaluated using Eq. (14).

Compute $\hat{P}_F=\frac{1}{i}\sum_{j=1}^{i}\frac{\tilde{P}\left(\boldsymbol{\theta}^j\right)}{\sum_{k=1}^{n_T}\mathrm{I}\left\{\left(\boldsymbol{\theta}^j,\boldsymbol{\xi}^j\right)\in F_k\left(\boldsymbol{\theta}^j\right)\right\}}W\left(\boldsymbol{\theta}^j\right)$.

Compute

$$\hat{\delta}_{\hat{P}_F}(i)=\frac{1}{\hat{P}_F\sqrt{i-1}}\sqrt{\frac{1}{i}\sum_{j=1}^{i}\left(\frac{\tilde{P}\left(\boldsymbol{\theta}^j\right)}{\sum_{k=1}^{n_T}\mathrm{I}\left\{\left(\boldsymbol{\theta}^j,\boldsymbol{\xi}^i\right)\in F_k\left(\boldsymbol{\theta}^j\right)\right\}}W\left(\boldsymbol{\theta}^j\right)\right)^2-\hat{P}_F^2}.$$

if *i mod* M_0 *= 0* **then**

Compute the sample average $\delta_{avg}=\frac{1}{m_0}\sum_{j=i-m_0+1}^{i}\hat{\delta}_{\hat{P}_F}(j)$.

end

while $i\le M_0$ *or* $\delta_{avg}>\delta^*_{\hat{P}_F}$

output:

$N=i$, and $\hat{P}_F=$ estimate of the first-passage probability.

4.3. Choice of the parametric density in the cross entropy method

In this section, we discuss the choice of the parametric distribution family $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu})$. Typically, $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu})$ is chosen such that it contains the nominal density $p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$. It is recalled that $\boldsymbol{\Theta}=\{\Theta_1;\ldots;\Theta_{n_{\boldsymbol{\theta}}}\}$ is the vector of basic random variables that model the uncertain

structural parameters. In reliability analysis, it is common practice to consider the components of $\boldsymbol{\Theta}$ are independent and standard normally distributed. If the structural parameters are mutually dependent and (or) follow a non-Gaussian distribution, they can be generated by an iso-probabilistic transformation of independent standard normal random variables [18, 14]. Therefore, without loss of generality, we assume that $\boldsymbol{\Theta}$ is an $n_{\boldsymbol{\theta}}$-dimensional standard normal random vector, i.e. $p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = \prod_{j=1}^{n_{\boldsymbol{\theta}}} p_{\Theta_j}(\theta_j)$, where for every j, $p_{\Theta_j}(\theta_j)$ is a one-dimensional standard normal PDF for Θ_j.

4.3.1. Multi-variate normal distribution

A standard choice of the distribution family is the multi-variate normal distribution [25, 41, 15]. This parametric family is completely described by its first two (joint) moments, the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$. The unknown parameter vector to be determined by CE optimization is thus given by $\boldsymbol{\nu} = [\boldsymbol{\mu}\ \boldsymbol{\Sigma}]$. This leads to a total of $n_{\boldsymbol{\theta}}(n_{\boldsymbol{\theta}} + 3)/2$ scalar parameters to be estimated at every sampling iteration of the CE method. The multi-variate normal distribution belongs to the exponential family, for which an analytical solution to the optimization problem in Eq. (30) can be derived [41]. Substituting $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \boldsymbol{\nu}) = \mathcal{N}(\boldsymbol{\theta}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ in Eq. (30), and setting the derivative of the objective function with respect to $\boldsymbol{\nu}$ equal to $\mathbf{0}$, yields the following parameter updates:

$$\hat{\boldsymbol{\mu}}^k = \frac{\sum_{i=1}^{N_{CE}} \boldsymbol{\theta}^i \widetilde{W}_k\left(\boldsymbol{\theta}^i, \hat{\boldsymbol{\nu}}^{k-1}\right)}{\sum_{i=1}^{N_{CE}} \widetilde{W}_k\left(\boldsymbol{\theta}^i, \hat{\boldsymbol{\nu}}^{k-1}\right)} \tag{36}$$

$$\hat{\boldsymbol{\Sigma}}^k = \frac{\sum_{i=1}^{N_{CE}} \widetilde{W}_k\left(\boldsymbol{\theta}^i, \hat{\boldsymbol{\nu}}^{k-1}\right)\left(\boldsymbol{\theta}^i - \hat{\boldsymbol{\mu}}^k\right)\left(\boldsymbol{\theta}^i - \hat{\boldsymbol{\mu}}^k\right)^{\mathrm{T}}}{\sum_{i=1}^{N_{CE}} \widetilde{W}_k\left(\boldsymbol{\theta}^i, \hat{\boldsymbol{\nu}}^{k-1}\right)} \tag{37}$$

The number of parameters in the multi-variate normal distribution increases quadratically with $n_{\boldsymbol{\theta}}$. This results in a rapid increase in the number of samples per level required to obtain a good estimate of the optimal parameter values. Therefore, the CE method with multi-variate normal densities performs poorly in high dimensional problems, i.e. in problems where the number of basic random variables representing structural parameter uncertainties is large.

4.3.2. von Mises-Fisher-Nakagami distribution

An alternative choice of the parametric density that is more suitable for high dimensions is the von Mises-Fisher-Nakagami (vMFN) distribution, as proposed in [37]. This distribution model is defined in terms of the polar coordinates of the standard normal random vector $\boldsymbol{\Theta}$. The polar representation of $\boldsymbol{\Theta}$ is given by

$$\boldsymbol{\Theta} = R\boldsymbol{A}, \tag{38}$$

where $\boldsymbol{A}$ is a random unit vector uniformly distributed on the $n_{\boldsymbol{\theta}}$-dimensional unit hypersphere and R is a scalar random variable, independent of $\boldsymbol{A}$, that follows the chi-distribution

with $n_{\boldsymbol{\theta}}$ degrees of freedom. The sample pair $\{r\ \boldsymbol{a}\}$ represents the radius and direction of a corresponding standard normal sample $\boldsymbol{\theta}$.

The vMFN distribution in polar coordinates is given by [37]

$$f_{\text{vMFN}}([r\ \boldsymbol{a}];\boldsymbol{\mu},\kappa,m,\Omega) = f_{\text{N}}(r;m,\Omega) f_{\text{vMF}}(\boldsymbol{a};\boldsymbol{\mu},\kappa). \tag{39}$$

Here $f_{\text{vMF}}(\boldsymbol{a};\boldsymbol{\mu},\kappa)$ is the PDF of the von Mises-Fisher distribution and $f_{\text{N}}(r;m,\Omega)$ is the PDF of the Nakagami distribution. $f_{\text{vMF}}(\boldsymbol{a};\boldsymbol{\mu},\kappa)$ is defined in terms of the mean direction $\boldsymbol{\mu}$ (with $\|\boldsymbol{\mu}\| = 1$) and the concentration parameter $\kappa \geq 0$, which characterizes the concentration around $\boldsymbol{\mu}$, and is given by

$$f_{\text{vMF}}(\boldsymbol{a};\boldsymbol{\mu},\kappa) = C_{n_{\boldsymbol{\theta}}}(\kappa) \exp\left(\kappa \boldsymbol{\mu}^{\text{T}} \boldsymbol{a}\right). \tag{40}$$

Here $C_{n_{\boldsymbol{\theta}}}(\kappa) = \frac{\kappa^{\frac{n_{\boldsymbol{\theta}}}{2}-1}}{(2\pi)^{\frac{n_{\boldsymbol{\theta}}}{2}} I_{\frac{n_{\boldsymbol{\theta}}}{2}-1}(\kappa)}$ is the normalizing constant, and I_k denotes the modified Bessel function of the first kind at order k. $f_{\text{N}}(r;m,\Omega)$ is defined in terms of a shape parameter $m \geq 0.5$ and a spread parameter $\Omega > 0$. $f_{\text{N}}(r;m,\Omega)$ is given by

$$f_{\text{N}}(r;m,\Omega) = \frac{2m^m}{\Gamma(m)\Omega^m} r^{2m-1} \exp\left(-\frac{m}{\Omega} r^2\right), \tag{41}$$

where $\Gamma(k)$ is the gamma function. The nominal density of the polar coordinates in the $n_{\boldsymbol{\theta}}$-dimensional $\boldsymbol{\Theta}$-space is retrieved by setting $[\boldsymbol{\mu},\kappa,m,\Omega] = [\boldsymbol{a},0,\frac{n_{\boldsymbol{\theta}}}{2},n_{\boldsymbol{\theta}}]$, where $\boldsymbol{a}$ is an arbitrary direction.

When the vMFN distribution is used within the CE method, the unknown parameter vector to be estimated by CE optimization is given by $\boldsymbol{\nu} = [\boldsymbol{\mu},\kappa,m,\Omega]$. Here all parameters are scalar-valued, with the exception of $\boldsymbol{\mu}$, which is a vector of dimension $n_{\boldsymbol{\theta}}$. Thus, the total number of parameters to be estimated at each sampling iteration is $n_{\boldsymbol{\theta}} + 3$, which increases only linearly with $n_{\boldsymbol{\theta}}$. The parameter updates at each step are obtained by substituting $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta};\boldsymbol{\nu}) = f_{\text{vMFN}}([r\ \boldsymbol{a}];\boldsymbol{\mu},\kappa,m,\Omega)$ in Eq. (30), and equating the derivate of the objective function with respect to $\boldsymbol{\nu}$ to $\mathbf{0}$. This leads to closed-form analytical updating rules for $\boldsymbol{\mu}$ and Ω, which are given by

$$\hat{\Omega}^k = \frac{\sum_{i=1}^{N_{CE}} \widetilde{W}_k\left(r^i \boldsymbol{a}^i, \hat{\boldsymbol{\nu}}^{k-1}\right)(r^i)^2}{\sum_{i=1}^{N_{CE}} \widetilde{W}_k\left(r^i \boldsymbol{a}^i, \hat{\boldsymbol{\nu}}^{k-1}\right)} \tag{42}$$

$$\hat{\boldsymbol{\mu}}^k = \frac{\sum_{i=1}^{N_{CE}} \boldsymbol{a}^i \widetilde{W}_k\left(r^i \boldsymbol{a}^i, \hat{\boldsymbol{\nu}}^{k-1}\right)}{\|\sum_{i=1}^{N_{CE}} \boldsymbol{a}^i \widetilde{W}_k\left(r^i \boldsymbol{a}^i, \hat{\boldsymbol{\nu}}^{k-1}\right)\|}. \tag{43}$$

The updating rules for κ and m cannot be derived analytically in closed-form, they require the solution of a non-linear equation [47, 8]. One could use a numerical scheme, such as the Newton-Raphson method to determine these parameters. However, in order to have an

efficient updating process, we employ an approximate solution of the resulting non-linear equation to update κ and m. The approximate updating rule for the concentration parameter κ of the von Mises-Fisher distribution reads [6]

$$\hat{\kappa}^k = \frac{\chi \cdot n_{\boldsymbol{\theta}} - \chi^3}{1 - \chi^2}, \tag{44}$$

where χ is defined as

$$\chi = \min\left(\frac{\| \sum_{i=1}^{N_{CE}} \boldsymbol{a}^i \widetilde{W}_k \left(r^i \boldsymbol{a}^i, \hat{\boldsymbol{\nu}}^{k-1} \right) \|}{\sum_{i=1}^{N_{CE}} \widetilde{W}_k \left(r^i \boldsymbol{a}^i, \hat{\boldsymbol{\nu}}^{k-1} \right)}, \chi_{max} \right). \tag{45}$$

with $\chi_{max} < 1$ typically selected as $\chi_{max} = 0.95$. The shape parameter m of the Nakagami distribution is approximated with the inverse normalized variance estimator [1, 23]

$$\hat{m}^k = \frac{(\hat{\Omega}^k)^2}{\mu_4 - (\hat{\Omega}^k)^2}, \tag{46}$$

where μ_4 is defined as

$$\mu_4 = \frac{\sum_{i=1}^{N_{CE}} \widetilde{W}_k \left(r^i \boldsymbol{a}^i, \hat{\boldsymbol{\nu}}^{k-1} \right) (r^i)^4}{\sum_{i=1}^{N_{CE}} \widetilde{W}_k \left(r^i \boldsymbol{a}^i, \hat{\boldsymbol{\nu}}^{k-1} \right)}. \tag{47}$$

4.3.3. Remark on the use of mixture models as the parametric density

When the failure domain is complex or has multiple design points in the structural parameter space, the uni-modal parametric densities described above may not be flexible enough to fit the complex shape and orientation of the optimal IS density. In such cases, a multi-modal model for the parametric density $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \boldsymbol{\nu})$ is required. Implementation of the multi-modal counterpart of the two aforementioned distribution models, including the associated parameter updates, is described in the literature (see [25, 15] for studies on multivariate normal mixtures and [47, 37] for studies on the vMFN mixture model). However, for the numerical examples considered in this paper, a uni-modal choice of $h_{\boldsymbol{\Theta}}(\boldsymbol{\theta}; \boldsymbol{\nu})$ is sufficient. Hence, multi-modal parametric densities are not investigated further here.

5. Numerical illustrations

The performance of the proposed CE-based importance sampling (CEIS) method is demonstrated by means of two numerical examples: the first considers a single degree-of-freedom (DOF) linear oscillator subjected to a stationary Gaussian white noise (GWN); and the second considers a 10-DOF linear structure excited by a filtered non-stationary Gaussian excitation. The second problem is one of the benchmark reliability examples studied in the

existing literature [42]. In the first example, the structure has two uncertain parameters. Since the problem is low dimensional with respect to the number of uncertain structural parameters, the multi-variate normal distribution is chosen as the parametric density for implementing the CE method. The second example demonstrates the performance of the method in a setting where the dimension of the structural parameter uncertainty is high. Both choices of the parametric density, namely, the multi-variate normal distribution and the vMFN model, are implemented and their relative performance is investigated.

In the CE optimization step, the target c.o.v. of the weights for the intermediate distributions is set to $\delta_{target} = 1.5$. Recall from Sec. 4.2 that the sample size N in the final reliability estimation step is either fixed or chosen adaptively on the fly. In the subsequent sections, the following nomenclature is used to denote the choice of N as well as the choice of the parametric density within the CE method:

- CEIS-mvn-fixN denotes the estimator with a fixed N, equal to the number samples per level of the CE optimization, and the multi-variate normal distribution as the parametric density.
- CEIS-vMFN-fixN is same as above, but uses the von Mises-Fisher-Nakagami distribution as the parametric density.
- CEIS-mvn-adap and CEIS-vMFN-adap are the variants of the above estimators where N is chosen adaptively to ensure that the sample estimate of the c.o.v. of $\hat{P}_F$ in each independent simulation run is smaller than a prescribed threshold $\delta^*_{\hat{P}_F}$.

We measure the performance of the proposed method in terms of the sample mean and c.o.v. of the estimates of the probability of failure and the total computational effort needed to obtain the estimates. In the subsequent sections, these performance measures are denoted by $\hat{P}_F$, $\delta^*_{\hat{P}_F}$ and N_T, respectively. The main computational effort lies in computing the dynamic system response, which requires evaluation and post-processing of the impulse response function. The number of evaluations of the impulse response function is chosen as the measure for the computational cost. In the considered examples, the dimension of the input excitation is $l = 1$. Hence, each evaluation of the impulse response function requires one dynamic analysis. The impulse response function of the critical response needs to be evaluated for every sample realization $\boldsymbol{\theta}$ of the uncertain parameter vector $\boldsymbol{\Theta}$. During CE optimization, the impulse response function is post-processed to determine the Poisson approximation of the conditional first-passage probability $P_{F|\boldsymbol{\Theta}}(\boldsymbol{\theta})$. In the reliability estimation step, the impulse response function for every sample $\boldsymbol{\theta}$ is convoluted with a sample realization of the input excitation to obtain a realization of the response time-history, which is subsequenly used for evaluating the IS estimator for P_F in Eq. (34). In Tables 1-6, the computational effort, i.e. the number of evaluations of the impulse response function, for finding the optimal IS density of the structural parameters by CE optimization and that for reliability estimation by IS are noted separately within parenthesis. The performance measures are estimated from 100 independent simulation runs. The reference values for the failure probabilities of Examples 5.1 and 5.2 are evaluated with large-scale DMC simulation.

The reference solution is denoted by $P_{F,DMCS}$ in Tables 1-6, and the associated c.o.v. given by $\delta_{P_{F,DMCS}} = \sqrt{\frac{1-P_{F,DMCS}}{N_T P_{F,DMCS}}}$ is also noted.

5.1. SDOF oscillator under stationary Gaussian white noise

Consider a single degree-of-freedom (DOF) oscillator with natural frequency ω and damping ratio η subjected to white noise excitation:

$$\ddot{X}(t) + 2\eta\omega\dot{X}(t) + \omega^2 X(t) = f(t) \tag{48}$$

The system is assumed to start from rest, i.e. $X(0) = 0$ and $\dot{X}(0) = 0$. The structural parameters ω and η are modeled as a pair of independent lognormal random variables. ω has mean 2π rad/s and standard deviation 0.2π rad/s, and η has mean 0.05 and standard deviation 0.005. The structure is subjected to a stochastic force $f(t)$ of duration $T = 20$s, which is characterized as a zero mean Gaussian white noise with auto-correlation function $\langle f(t)f(t+\tau)\rangle = I\delta(\tau)$. Here I denotes the intensity of the white noise, which is taken equal to 1 $\mathrm{m}^2/\mathrm{s}^3$. Failure is defined as the maximum value of the displacement $X(t)$ exceeding a threshold h^* within the time span $[0, T]$. 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 $\Delta t = 0.01$s. Hence, the number of time instants is $n_T = 2001$. The random vector $\boldsymbol{\Xi}$ characterizing $f(t)$ consists of the sequence of i.i.d. standard normal random variables $\{\Xi_k, k = 1, \ldots, n_T\}$ that generate the white noise at the discrete time instants, i.e. $\left\{f(t_k, \boldsymbol{\Xi}) = \sqrt{I/\Delta t}\Xi_k, k = 1, \ldots, n_T\right\}$.

The sample mean and c.o.v. of the estimates of the first-passage probability for different h^*, as well as the average computational effort needed to obtain the estimates with the proposed CEIS method, are reported in Table 1. The simulation results are obtained using the multi-variate normal distribution as the parametric density. A sample size of $N_{CE} = 500$ samples per level is employed to solve the CE optimization problem. The estimates from the adaptive variant of the IS estimator (CEIS-mvn-adap) correspond to $\delta^*_{\hat{P}_F} = 0.05$. The reference value for the probability of failure, also shown in Table 1, is computed from 10^8 DMC samples. The sample mean value of the probability estimates obtained using the proposed estimators compare well with the reference solution. There is a small underestimation for the case $h^* = 1.1$m, which might well be due to the sampling variance of the reference solution. The required computational effort for CE optimization shows that the average number of sampling iterations increases as the failure probability becomes smaller.

A parametric study is conducted to investigate the effect of the number of samples per level on the performance of the CEIS method. To this end, values of N_{CE} in the range 100-1000, and $\delta^*_{\hat{P}_F} = 0.05$ and 0.025 are considered. For different values of N_{CE}, the sample means of the probability of failure estimates across all threshold levels h^* are similar to the values in Table 1, and hence are not reported. Fig. 2 shows that the sample c.o.v. $\delta_{\hat{P}_F}$ and the total computational effort N_T change significantly for varying N_{CE}. The dotted lines in Fig. 2 indicate the average computational effort needed to solve the CE optimization

Table 1: Failure probability estimates for Example 5.1 using CEIS-mvn-fixN and CEIS-mvn-adap with $N_{CE} = 500$. Results from CEIS-mvn-adap obtained with $\delta^*_{\hat{P}_F} = 0.05$. Reference solution obtained from large-scale DMC simulation.

h^* (m)	CEIS-mvn-fixN			CEIS-mvn-adap			DMC ($N_T = 10^8$)	
	$\hat{P}_F$	$\delta_{\hat{P}_F}$	N_T	$\hat{P}_F$	$\delta_{\hat{P}_F}$	N_T	$P_{F,DMCS}$	$\delta_{P_{F,DMCS}}$
0.7	1.81×10^{-3}	0.040	1595 (1095 + 500)	1.81×10^{-3}	0.046	1476 (1110 + 366)	1.82×10^{-3}	0.002
0.8	2.74×10^{-4}	0.037	1910 (1410 + 500)	2.73×10^{-4}	0.046	1763 (1440 + 323)	2.80×10^{-4}	0.006
0.9	4.00×10^{-5}	0.058	2045 (1545 + 500)	3.91×10^{-5}	0.048	1878 (1560 + 318)	4.08×10^{-5}	0.016
1.1	7.80×10^{-7}	0.040	2485 (1985 + 500)	7.78×10^{-7}	0.043	2332 (2035 + 297)	9.06×10^{-7}	0.105

problem for different values of N_{CE}. The difference between the vertical coordinates of the solid lines and the dotted line is the average computational effort required in the final reliability estimation step. The optimization effort increases monotonically with N_{CE}. For the estimator CEIS-mvn-fixN, an increase in N_{CE} also implies an increase in the sample size of the IS estimator, and hence we observe an increase in the total computational effort and a decrease in the sampling c.o.v. of the probability estimates. In case of the estimator CEIS-mvn-adap, we observe that the computational effort N_T initially decreases before increasing. This behaviour is more pronounced for $\delta^*_{\hat{P}_F} = 0.025$. The reason for this is the number of effective samples (ESS) that are available for fitting the parametric density in every sample iteration of the CE method. For a fixed value of δ_{target}, the ESS decreases with decrease in the number of samples per level. Hence the parameter vector determined by CE optimization using a small N_{CE} is sub-optimal in comparison to that obtained using higher values of N_{CE}. Consequently, a larger number of samples are required during reliability estimation to meet the prescribed $\delta^*_{\hat{P}_F}$. The increase in the sample size in the final step increases the overall computational effort. As N_{CE} increases, we obtain improved estimates of the parameter vector, and the number of samples needed for reliability estimation starts decreasing. Beyond a certain value of N_{CE}, which for this example is $N_{CE} = 250$, the sample size required in the IS estimator becomes nearly constant. This indicates that the parametric density fitted using 250 samples per level is sufficiently optimal, and a further increase in the number of samples per level does not give any additional advantage. As expected, the sample c.o.v. of the estimator CEIS-mvn-adap remains close to the prescribed threshold $\delta^*_{\hat{P}_F}$ for different values of N_{CE}. Finally, Fig. 2 demonstrates that the total computational effort required by the estimator CEIS-mvn-adap to achieve a sample c.o.v. of 0.025 is approximately half of that required by the estimator CEIS-mvn-fixN. Hence, if the goal is to ensure that the IS estimates of the failure probability achieve a given level of c.o.v., using the adaptive variant of the estimator is computationally more efficient.

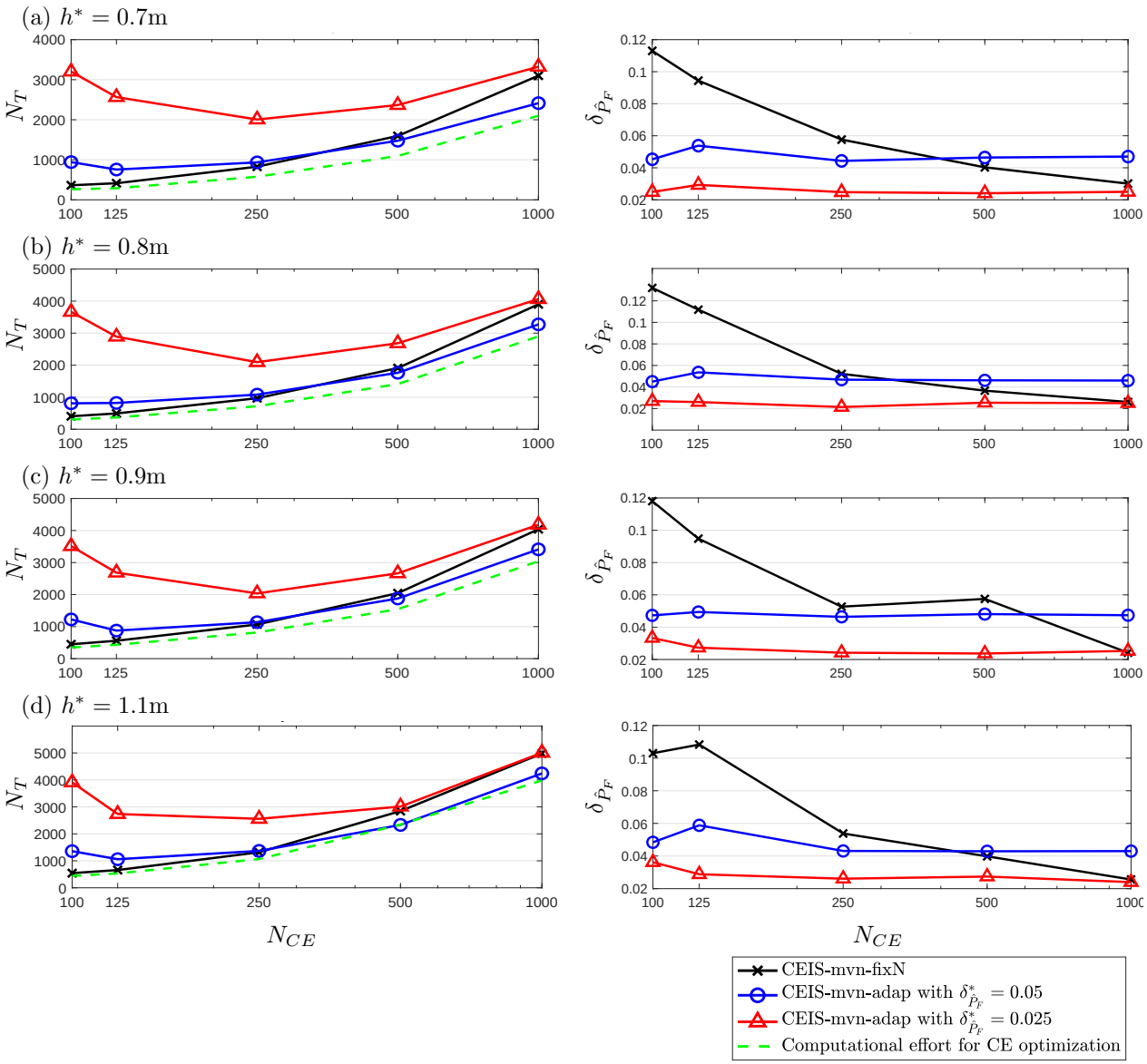

Figure 2: Variation in the total computational effort N_T and the sample c.o.v. $\delta_{\hat{P}_F}$ for different values of N_{CE} in Example 5.1. The rows corresponds to the threshold levels (a) $h^* = 0.7$m; (b) $h^* = 0.8$m; (c) $h^* = 0.9$m; (d) $h^* = 1.1$m. Note that the dashed line does not reflect N_T, but the computational effort needed only for CE optimization. The values corresponding to CEIS-mvn-fixN are plotted.

5.2. 10-story linear frame under filtered non-stationary Gaussian excitation

The second example considers a ten-story linear structure with uncertain material properties and subject to a stochastic ground acceleration. This system has been studied in [42]. The structure is idealized as a 10-DOF mass-spring-dashpot system with lumped masses $\{m_i, i = 1, \ldots, 10\}$, inter-storey stiffness coefficients $\{k_i, i = 1, \ldots, 10\}$ and damping ratios $\{\eta_i, i = 1, \ldots, 10\}$. The governing equation is given by

$$\mathbf{M}\ddot{\boldsymbol{X}}(t) + \mathbf{C}\dot{\boldsymbol{X}}(t) + \mathbf{K}\boldsymbol{X}(t) = \{m_1, \ldots, m_{10}\}^{\mathrm{T}} f(t), \tag{49}$$

with initial conditions $\boldsymbol{X}(0) = \mathbf{0}$ and $\dot{\boldsymbol{X}}(0) = \mathbf{0}$. In Eq. (49), $\boldsymbol{X}(t) = \{\boldsymbol{X}_1(t), \ldots, \boldsymbol{X}_{10}(t)\}^{\mathrm{T}}$ is the displacement vector, where $\boldsymbol{X}_i(t)$ denotes the relative displacement between the i-th floor and the support, and $\mathbf{M}$, $\mathbf{C}$ and $\mathbf{K}$ are the mass, damping and stiffness matrices which are, respectively, given by

$$\mathbf{M} = \begin{bmatrix} m_1 & 0 & 0 & \ldots & 0 \\ 0 & m_2 & 0 & \ldots & 0 \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ 0 & \ldots & 0 & 0 & m_{10} \end{bmatrix} \tag{50}$$

$$\mathbf{C} = \begin{bmatrix} c_1 + c_2 & -c_2 & 0 & \ldots & 0 \\ -c_2 & c_2 + c_3 & -c_3 & \ldots & 0 \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ 0 & \ldots & 0 & -c_{10} & c_{10} \end{bmatrix} \tag{51}$$

$$\mathbf{K} = \begin{bmatrix} k_1 + k_2 & -k_2 & 0 & \ldots & 0 \\ -k_2 & k_2 + k_3 & -k_3 & \ldots & 0 \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ 0 & \ldots & 0 & -k_{10} & k_{10} \end{bmatrix} \tag{52}$$

The damping coefficients in Eq. (51) are defined as $\{c_i = 2\eta_i\sqrt{m_i k_i}, i = 1, \ldots, 10\}$.

The random support excitation $f(t)$ in Eq. (49) is modeled by a modulated filtered GWN:

$$f(t) = \omega_d^2 x_d(t) + 2\eta_d\omega_d\dot{x}_d(t) - \omega_g^2 x_g(t) - 2\eta_g\omega_g\dot{x}_g(t), \tag{53}$$

where $\{x_d(t) \;\; \dot{x}_d(t) \;\; x_g(t) \;\; \dot{x}_g(t)\}^{\mathrm{T}}$ are the states of the filter defined by the linear system

$$\begin{aligned} &\ddot{x}_d(t) + 2\eta_d\omega_d\dot{x}_d(t) + \omega_d^2 x_d(t) = N(t) \\ &\ddot{x}_g(t) + 2\eta_g\omega_g\dot{x}_g(t) + \omega_g^2 x_g(t) = 2\eta_d\omega_d\dot{x}_d(t) + \omega_d^2 x_d(t) \\ &x_d(0) = 0, \;\; \dot{x}_d(0) = 0, \;\; x_g(t), \;\; \dot{x}_g(0) = 0. \end{aligned} \tag{54}$$

In the above equation, $N(t)$ is a GWN with zero mean and auto-correlation function $\langle N(t)N(t+\tau)\rangle = Ie^2(t)\delta(\tau)$. Here I denotes the intensity of the white noise, and $e(t)$ is the modulating function given by

$$e(t) = \begin{cases} 0 & \text{for } t \le 0\text{s} \\ t/2 & \text{for } 0\text{s} \le t \le 2\text{s} \\ 1 & \text{for } 2\text{s} \le t \le 10\text{s} \\ \exp(-0.1(t-10)) & \text{for } t \ge 10\text{s} \end{cases} \tag{55}$$

The numerical values of the filter parameters are assumed to be $\omega_d = 15$ rad/s, $\eta_d = 0.8$, $\omega_g = 0.3$ rad/s and $\eta_g = 0.995$, and the white noise intensity is taken as $I = 0.08$ m^2/s^3. Independent normally distributed impulses with zero mean and standard deviation $e(t_k)\sqrt{I\Delta t}$ are applied to the filter at each discrete time steps $\{t_k, k = 1, \ldots, n_T\}$. The duration of the excitation and the sampling interval are taken as $T = 20$s and $\Delta t = 0.005$s, respectively. Therefore the total number of random variables required to characterize the uncertain excitation is 4001.

In addition to the uncertainty in the input excitation, the structural parameters m_i, η_i and k_i, $i = 1, \ldots, 10$ are assumed to be uncertain and are modeled as random variables. The following two cases are considered.

- Case 1: the stiffness coefficients $\{k_i, i = 1, \ldots, 10\}$ are modeled by independent Gaussian random variables. The lumped masses $\{m_i, i = 1, \ldots, 10\}$ and damping ratios $\{\eta_i, i = 1, \ldots, 10\}$ are assumed to be deterministic with their values set to $m_1 = \cdots = m_{10} = 10$ Mg and $\eta_1 = \cdots = \eta_{10} = 0.04$, respectively.

- Case 2: in addition to $\{k_i, i = 1, \ldots, 10\}$, the parameters $\{m_i, i = 1, \ldots, 10\}$ and $\{\eta_i, i = 1, \ldots, 10\}$ are modeled by independent Gaussian random variables. This setting corresponds to Case 2 of Problem 2 considered in [42].

The statistical properties of the uncertain structural parameters in Cases 1 and 2 are listed in Table 2. It is noted that Gaussian random variables for structural properties are censored when the deviation from the mean exceeds five times the standard deviation.

Table 2: Statistical properties of the uncertain structural parameters in Example 5.2.

	Parameters	Distribution	Mean	Standard deviation
Case 1	k_1, k_2, k_3	censored Normal	40 MN/m	4.0 MN/m
	k_4, k_5, k_6	censored Normal	36 MN/m	3.6 MN/m
	k_7, k_8, k_9, k_{10}	censored Normal	32 MN/m	3.2 MN/m
	$m_1, \cdots, m_{10}$	Deterministic	10 Mg	–
	$\eta_1, \cdots, \eta_{10}$	Deterministic	0.04	–
Case 2	$k_1, \ldots, k_{10}$	censored Normal	same as in Case 1	
	$m_1, \ldots, m_{10}$	censored Normal	10 Mg	1.0 Mg
	$\eta_1, \ldots, \eta_{10}$	censored Normal	0.04	0.004

In both cases, two response quantities are considered: (i) the first floor displacement, given by $h_1(t, \boldsymbol{\Theta}, \boldsymbol{\Xi}) = X_1(t)$, and (ii) the inter-storey drift between the ninth and the tenth floors, given by $h_2(t, \boldsymbol{\Theta}, \boldsymbol{\Xi}) = X_{10}(t) - X_9(t)$. The objective is to determine the probability that the maximum value of a response $h_i(t, \boldsymbol{\Theta}, \boldsymbol{\Xi})$ exceeds a prescribed threshold h_i^* over the time interval [0s 20s].

5.2.1. Case 1: 10 uncertain structural parameters

The structural system in this case comprises 10 uncertain parameters $\{k_i, i = 1, \ldots, 10\}$, which are modeled as independent normal variables with statistical properties shown in Table 2. The CEIS method is implemented using the multi-variate normal distribution that has a total of 65 scalar parameters for $n_{\boldsymbol{\theta}} = 10$, and the vMFN model that has 13 scalar parameters.

The simulation results for $h_1^* = 0.052$m and $h_2^* = 0.01$m, obtained using the estimators CEIS-mvn-fixN and CEIS-vMFN-fixN, are reported in Tables 3 and 4, respectively. These results illustrate the effect of varying the number of samples per level on the performance of the method. While using the vMFN model within the CEIS method results in sufficiently accurate estimates for any value of N_{CE}, the sample means of the estimates obtained using the multi-variate normal distribution deviate significantly from the reference solution for $N_{CE} = 125$. In terms of the c.o.v. of the estimates $\delta_{\hat{P}_F}$, the estimator CEIS-vMFN-fixN is superior for sample sizes lower than 1000 samples per level. In case of the multi-variate normal distribution the CE optimization problem requires additional steps to converge. This results in a larger computational effort N_T for the estimator CEIS-mvn-fixN. The performance gap between the two estimators reduces with increase in the number of samples per levels, and with $N_{CE} = 1000$ both approaches give comparable results. The poor performance of the multi-variate normal distribution is due to the larger number of parameters that are required to be updated in every sampling iteration of the CE method. For a small N_{CE}, the available number of effective samples per level is insufficient to obtain an adequate estimate of the optimal parameter values during CE optimization.

Table 3: Failure probability estimates for Example 5.2-Case 1 for $h_1^* = 0.052$m. Comparison between the estimators CEIS-mvn-fixN and CEIS-vMFN-fixN. Reference probability of failure obtained from 10^6 DMC samples is $P_{F,DMCS} = 3.88 \times 10^{-4} (\delta_{P_{F,DMCS}} = 0.051)$.

N_{CE}	CEIS-mvn-fixN			CEIS-vMFN-fixN		
	$\hat{P}_F$	$\delta_{\hat{P}_F}$	N_T	$\hat{P}_F$	$\delta_{\hat{P}_F}$	N_T
125	1.70×10^{-4}	0.821	653 (528 + 125)	3.84×10^{-4}	0.165	406 (281 + 125)
250	3.66×10^{-4}	0.200	1178 (928 + 250)	3.82×10^{-4}	0.115	805 (555 + 250)
500	3.82×10^{-4}	0.134	1800 (1300 + 500)	3.79×10^{-4}	0.079	1540 (1040 + 500)
1000	3.79×10^{-4}	0.053	3180 (2180 + 1000)	3.78×10^{-4}	0.060	3020 (2020 + 1000)

We investigate the performance of the adaptive variant of the IS estimator for different sample sizes per level. The results from CEIS-mvn-adap and CEIS-vMFN-adap obtained with $N_{CE} = 250$, 500, 1000, and $\delta^*_{\hat{P}_F} = 0.05$ are reported in Table 5. With 250 samples per level, the CEIS method using the multi-variate normal distribution fails to con-

Table 4: Failure probability estimates for Example 5.2-Case 1 for $h_2^* = 0.01$m. Comparison between the estimators CEIS-mvn-fixN and CEIS-vMFN-fixN. Reference probability of failure obtained from 10^6 DMC samples is $P_{F,DMCS} = 1.27 \times 10^{-3} (\delta_{P_{F,DMCS}} = 0.028)$.

N_{CE}	CEIS-mvn-fixN			CEIS-vMFN-fixN		
	$\hat{P}_F$	$\delta_{\hat{P}_F}$	N_T	$\hat{P}_F$	$\delta_{\hat{P}_F}$	N_T
125	2.21×10^{-3}	7.135	644 (519 + 125)	1.21×10^{-3}	0.205	404 (279 + 125)
250	1.18×10^{-3}	0.237	1135 (885 + 250)	1.22×10^{-3}	0.111	783 (533 + 250)
500	1.23×10^{-3}	0.104	1755 (1255 + 500)	1.21×10^{-3}	0.063	1540 (1040 + 500)
1000	1.25×10^{-3}	0.053	3090 (2090 + 1000)	1.22×10^{-3}	0.054	3000 (2000 + 1000)

verge to the prescribed $\delta^*_{\hat{P}_F}$, even after expending considerable computational effort. For $N_{CE} = 500$ and 1000, the estimator CEIS-mvn-adap requires greater computational effort to converge in the final reliability estimation step. The sample mean and c.o.v. of the probability estimates $\hat{P}_F$ obtained using the two estimators are comparable.

Table 5: Failure probability estimates for Example 5.2-Case 1 using CEIS-mvn-adap and CEIS-vMFN-adap. Results obtained with $\delta^*_{\hat{P}_F} = 0.05$.

		CEIS-mvn-adap			CEIS-vMFN-adap		
		$\hat{P}_F$	$\delta_{\hat{P}_F}$	N_T	$\hat{P}_F$	$\delta_{\hat{P}_F}$	N_T
$h_1^* = 0.052$m	$N_{CE} = 250$	–	–	–	3.73×10^{-4}	0.050	1690 (550 + 1140)
	$N_{CE} = 500$	3.74×10^{-4}	0.061	2622 (1230 + 1392)	3.77×10^{-4}	0.053	2144 (1045 + 1099)
	$N_{CE} = 1000$	3.78×10^{-4}	0.050	3430 (2240 + 1190)	3.78×10^{-4}	0.051	3062 (2020 + 1042)
$h_2^* = 0.010$m	$N_{CE} = 250$	–	–	–	1.22×10^{-3}	0.044	1682 (543 + 1139)
	$N_{CE} = 500$	1.23×10^{-3}	0.049	2722 (1250 + 1472)	1.22×10^{-3}	0.050	2048 (1020 + 1028)
	$N_{CE} = 1000$	1.22×10^{-3}	0.051	3190 (2080 + 1110)	1.22×10^{-3}	0.046	3012 (2000 + 1012)

Fig. 3 shows the variation in the total computational effort N_T and the sample c.o.v. $\delta_{\hat{P}_F}$ of the estimators CEIS-vMFN-fixN and CEIS-vMFN-adap for different number of samples per level. The results from CEIS-vMFN-adap are obtained using $\delta^*_{\hat{P}_F} = 0.05$ and 0.025. For the estimator CEIS-vMFN-fixN we observe an increase in N_T and a decrease in $\delta_{\hat{P}_F}$ with increasing values of N_{CE}. For the estimator CEIS-vMFN-adap with $\delta^*_{\hat{P}_F} = 0.025$,

the computational effort required to obtain the probability estimates using $N_{CE} = 125$ is significantly higher compared to that using $N_{CE} = 250$. The required computational effort with $\delta^*_{\hat{P}_F} = 0.05$ remains nearly constant for $N_{CE} = 125,\ 250$. For both values of $\delta^*_{\hat{P}_F}$, a gradual increase in N_T is observed for $N_{CE} \geq 250$. The sample c.o.v. $\delta_{\hat{P}_F}$ of the estimator CEIS-vMFN-adap remains close to the prescribed values of $\delta^*_{\hat{P}_F}$ for all values of N_{CE}. These results are consistent with the ones of Fig. 2 of Example 5.1. Finally it is noted that the estimator CEIS-vMFN-fixN requires a total computational effort of approximately $N_T = 3000$ to achieve a sample c.o.v. of approximately 0.05 for both considered thresholds. The estimator CEIS-vMFN-adap achieves this sample c.o.v. with an approximate effort of $N_T = 1690$ for both response thresholds, and hence is more efficient.

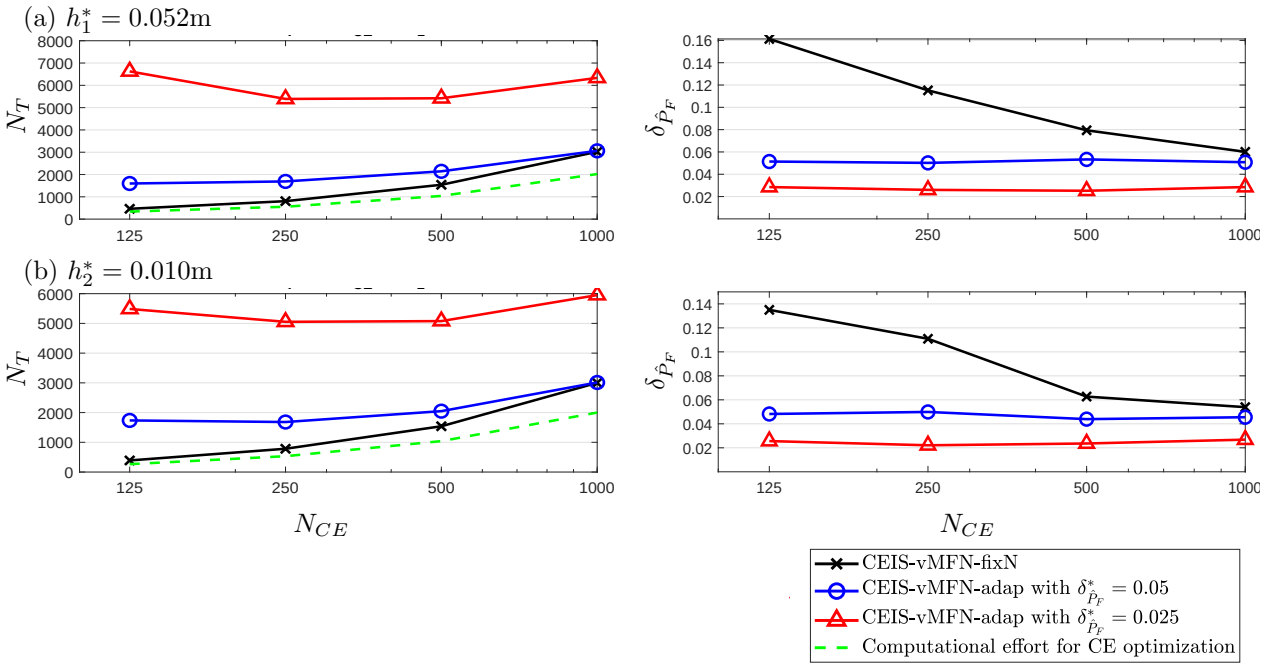

Figure 3: Variation in the total computational effort N_T and the sample c.o.v. $\delta_{\hat{P}_F}$ for different values of N_{CE} in Example 5.2-Case 1. The rows corresponds to the threshold levels (a) $h_1^* = 0.052$m; (b) $h_2^* = 0.010$m. Note that the dashed line does not reflect N_T, but the computational effort needed only for CE optimization. The values corresponding to CEIS-vMFN-fixN are plotted.

5.2.2. Case 2: 30 uncertain structural parameters

In this case, the parameters $\{m_i, k_i, \eta_i, i = 1, \ldots, 10\}$ are modeled as independent normally distributed random variables with statistical properties shown in Table 2. The dynamical system thus comprises of $n_{\boldsymbol{\theta}} = 30$ uncertain structural parameters. Choosing the multivariate normal distribution as the parametric density leads to a total of $n_{\boldsymbol{\theta}}(n_{\boldsymbol{\theta}} + 3)/2 = 495$ scalar parameters that need to be estimated at every sampling iteration of the CE method. In order to obtain sufficiently accurate estimates of the failure probability, the number of samples per level N_{CE} should be chosen such that an adequate number of effective samples

(ESS) are available for fitting the parametric model. The required ESS depends on the number of parameters to be estimated. In [37] it is discussed that an ESS of approximately 1.5 times the number of parameters is a good choice. For a multi-variate normal distribution in 30 dimensions, this leads to a required sample size of $N_{CE} \gtrapprox 2400$. Using such large sample size per level is inefficient. Therefore, for this case, the vMFN distribution is chosen as the parametric density in the CE method. The vMFN model has $n_{\boldsymbol{\theta}} + 3 = 33$ parameters and gives fairly accurate estimates with $N_{CE} \gtrapprox 165$.

The simulation results for $h_1^* = 0.057$m, 0.073m and $h_2^* = 0.013$m, 0.017m obtained using the estimators CEIS-vMFN-fixN and CEIS-vMFN-adap with $N_{CE} = 250$ are reported in Table 6. The results from CEIS-vMFN-adap are obtained using $\delta^*_{\hat{P}_F} = 0.05$. The reference solution obtained from 3.5×10^7 DMC samples is also shown in the table. The dynamic system in Case 2 is one of the benchmark examples considered in [42]. Hence results from other existing variance reduction methods are available for this case. These results are also reported in Table 6. The following nomenclature is used to denote the alternative methods: SubSim/MCMC denotes subset simulation with Markov chain Monte Carlo, SubSim/Hybrid denotes hybrid subset simulation, CMA denotes importance sampling using approximate representations of performance functions, S^3 denotes spherical subset simulation and LS denotes line sampling. The results show that the sample mean of the probability estimates obtained using the proposed methods are broadly comparable with that obtained using other variance reduction schemes. The CEIS-vMFN-fixN and CEIS-vMFN-adap methods significantly outperform SubSim/MCMC, SubSim/Hybrid and S^3 in terms of the sample c.o.v. of the estimates and the computational effort. The estimates obtained using CMA and LS have smaller variability and require smaller computational effort. It is noted that the superior performance of CMA and LS comes at the expense of reduced robustness. These methods make use of certain algorithmic parameters for reliability estimation, whose selection requires prior knowledge of the system behaviour and possibly additional investigations. LS requires specification of an important direction to sample from, which corresponds to the important region of the failure domain. CMA uses IS to estimate the failure probability, wherein a pseudo-design point needs to de identified to generate the sample density. The design point is further used to obtain an approximate representation of the dynamic response in terms of the uncertain structural parameters. Such an approximation can also lead to a bias in the reliability estimates. In this regard, the proposed CEIS method is advantageous as it is completely adaptive, requires no system specific information, and can be utilized as a black-box method.

Fig. 4 demonstrates the variation in the total computational effort and the sample c.o.v. of the estimator CEIS-vMFN-adap for different number of samples per level. The observations are broadly similar to that in Fig. 3. The results in Fig. 4 show that if the target c.o.v. $\delta^*_{\hat{P}_F}$ is about 0.1, one could expect to get satisfactory results with a total computational effort of approximately 1000 samples, i.e. 125-250 samples per level. Moreover, it is observed that even with 125 samples per level, CEIS-vMFN-fixN and CEIS-vMFN-adap outperform other sampling approaches like SubSim/MCMC, SubSim/Hybrid and S^3.

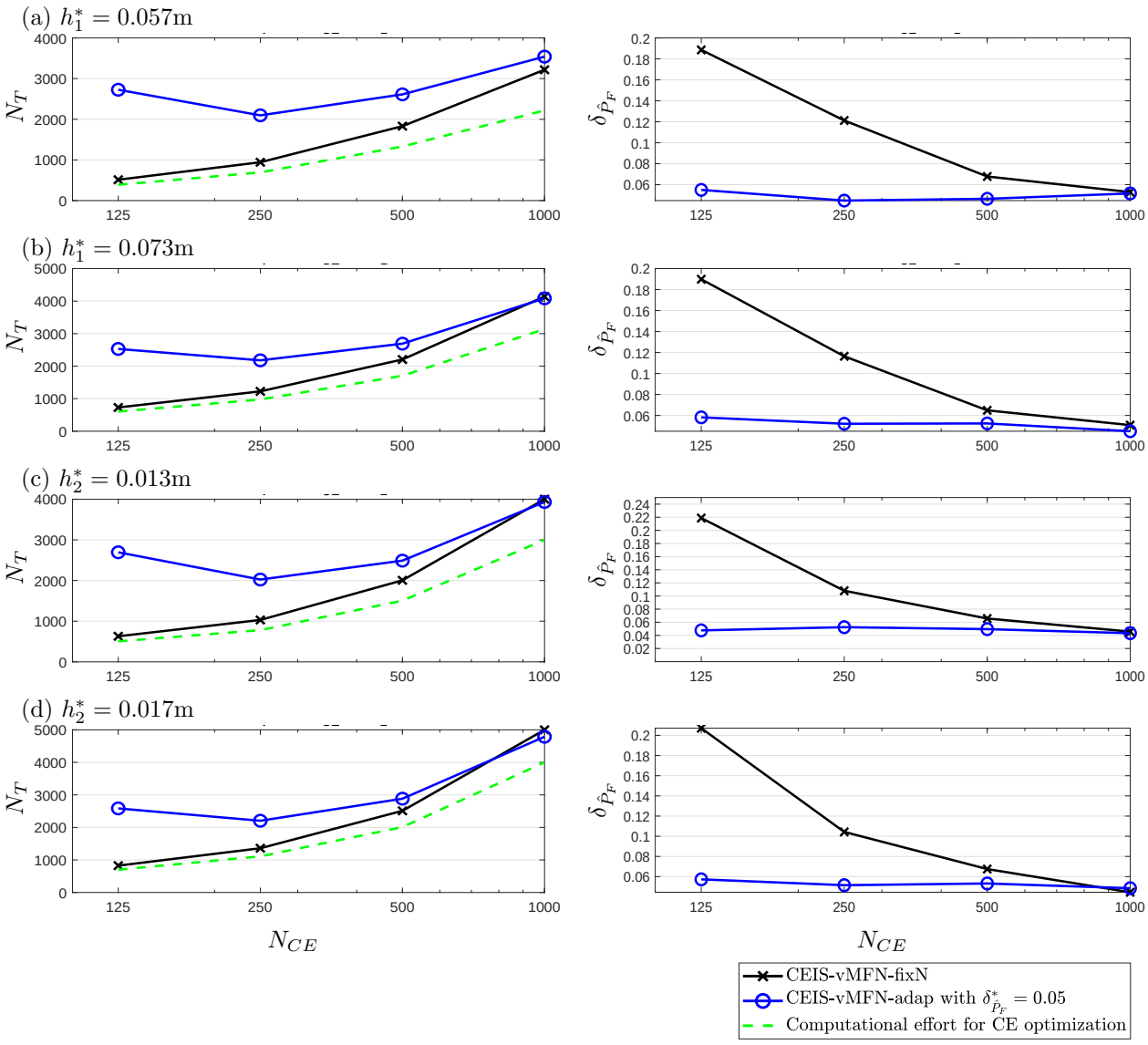

Figure 4: Variation in the total computational effort N_T and the sample c.o.v. $\delta_{\hat{P}_F}$ for different values of N_{CE} in Example 5.2-Case 2. The rows corresponds to the threshold levels (a) $h_1^* = 0.057$m; (b) $h_1^* = 0.073$m; (c) $h_2^* = 0.013$m; (d) $h_2^* = 0.017$m. Note that the dashed line does not reflect N_T, but the computational effort needed only for CE optimization. The values corresponding to CEIS-vMFN-fixN are plotted.

Table 6: Failure probability estimates for Example 5.2-Case 2 using CEIS-vMFN-fixN and CEIS-vMFN-adap and other existing methods. Results from CEIS-vMFN-fixN and CEIS-vMFN-adap obtained with $N_{CE} = 250$. In CEIS-vMFN-adap, $\delta^*_{\hat{P}_F} = 0.05$. Results from SubSim/MCMC, SubSim/Hybrid, CMA, S^3 and LS taken from [42].

		DMC	SubSim/MCMC	SubSim/Hybrid	CMA	S^3	LS	CEIS-vMFN-fixN	CEIS-vMFN-adap
$h_1^* = 0.057$m	$\hat{P}_F$	9.55×10^{-5}	1.2×10^{-4}	1.1×10^{-4}	1.0×10^{-4}	9.2×10^{-5}	9.8×10^{-5}	9.51×10^{-5}	9.51×10^{-5}
	$\delta_{\hat{P}_F}$	0.017	0.770	0.410	0.100	0.350	0.050	0.121	0.045
	N_T	3.5×10^7	1850	2128	300	3070	360	943	2096 (733 + 1363)
$h_1^* = 0.073$m	$\hat{P}_F$	1.11×10^{-6}	1.0×10^{-6}	1.1×10^{-6}	9.8×10^{-7}	8.8×10^{-7}	9.7×10^{-7}	8.68×10^{-7}	8.58×10^{-7}
	$\delta_{\hat{P}_F}$	0.160	0.990	0.770	0.100	0.680	0.080	0.116	0.052
	N_T	3.5×10^7	2750	3163	300	4200	360	1225	2179 (932 + 1247)
$h_2^* = 0.013$m	$\hat{P}_F$	5.20×10^{-5}	6.6×10^{-5}	5.9×10^{-5}	6.0×10^{-5}	4.6×10^{-5}	6.0×10^{-5}	5.21×10^{-5}	5.22×10^{-5}
	$\delta_{\hat{P}_F}$	0.023	0.580	0.460	0.140	0.580	0.120	0.108	0.053
	N_T	3.5×10^7	2300	2645	300	3250	360	1030	2024 (792 + 1232)
$h_2^* = 0.017$m	$\hat{P}_F$	2.28×10^{-7}	4.7×10^{-7}	3.2×10^{-7}	4.6×10^{-7}	5.3×10^{-7}	4.6×10^{-7}	3.72×10^{-7}	3.76×10^{-7}
	$\delta_{\hat{P}_F}$	0.354	0.780	0.740	0.140	0.820	0.160	0.104	0.052
	N_T	3.5×10^7	2750	3680	300	4900	360	1360	2205 (1087 + 1118)

6. Concluding remarks

We develop an adaptive importance sampling method to estimate the first-passage probability of uncertain linear structures subjected to dynamic loads modeled by Gaussian random processes. The main novelty lies in the construction of an effective IS density for the uncertain structural parameters, which is accomplished using the multi-level CE method. In the multi-level approach, the CE optimization is solved sequentially for a series of intermediate target densities that gradually approach the optimal IS density of the uncertain parameters. The distribution sequence is constructed by introducing a smoothening of the first-passage probability conditional on the uncertain parameter vector. During CE optimization, the conditional first-passage probability is evaluated using an analytical approximation. This approach significantly reduces the computational effort needed for optimization, without compromising much on accuracy. The IS density of the uncertain structural parameters is finally combined with an efficient IS density for the random excitation proposed in [2] to estimate the first-passage probability.

The sample size in the IS estimator of the failure probability is chosen using two approaches. In the first case, the number of samples is fixed to a certain value, and in the second case the sample size is chosen adaptively on the fly to ensure that an estimate of the c.o.v. of the IS estimator is less than a specified threshold. Results from numerical examples demonstrate that for achieving a desired level of the sample c.o.v. of the probability estimates, the adaptive variant of the estimator is computationally more efficient. Implementation of the CE method requires specification of a parametric density. In the present study, the multi-variate normal distribution is used when the number of uncertain structural parameters is small, while for high dimensional problems the von Mises-Fisher-Nakagami distribution is employed. The numerical studies show that the proposed IS method performs significantly better than other sampling-based approaches in terms of the sample c.o.v. of the estimates and the computational effort. Furthermore, the proposed approach is a black-box method that requires no prior investigations of the dynamical system.

Two different directions of future research are envisioned. The first involves investigating how the proposed line of work can be extended to problems of system reliability analysis, where failure of the structure is expressed in terms of multiple component level first-passage events. The second involves determining if the proposed approach can be further developed to estimate the reliability of non-linear structures.

Acknowledgements

This work was supported by the Alexander von Humboldt Foundation.

References

[1] A. Abdi and M. Kaveh. Performance comparison of three different estimators for the Nakagami m parameter using Monte Carlo simulation. *IEEE Communications Letters*, 4(4):119–121, 2000.

[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. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, 16(4):263–277, 2001.
[4] S. K. Au, C. Papadimitriou, and J. L. Beck. Reliability of uncertain dynamical systems with multiple design points. *Structural Safety*, 21(2):113–133, 1999.
[5] S. K. Au, J. Ching, and J. L. Beck. Application of subset simulation methods to reliability benchmark problems. *Structural Safety*, 29(3):183–193, 2007.
[6] A. Banerjee, I. S. Dhillon, J. Ghosh, and S. Sra. Clustering on the unit hypersphere using von Mises-Fisher distributions. *Journal of Machine Learning Research*, 6:1345–1382, 2005.
[7] 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.
[8] N. Bouhlel and A. Dziri. Maximum likelihood parameter estimation of Nakagami-Gamma shadowed fading channels. *IEEE Communications Letters*, 19(4):685–688, 2015.
[9] J. B. Chen and J. Li. The extreme value distribution and dynamic reliability analysis of nonlinear structures with uncertain parameters. *Structural Safety*, 29(2):77–93, 2007.
[10] 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.
[11] J. Ching, J. L. Beck, and S. K. Au. Hybrid subset simulation method for reliability estimation of dynamical systems subject to stochastic excitation. *Probabilistic Engineering Mechanics*, 20(3):199–214, 2005.
[12] N. Chopin. A sequential particle filter method for static models. *Biometrika*, 89(3):539–552, 2002.
[13] 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.
[14] A. Der Kiureghian and P. L. Liu. Structural reliability under incomplete probability information. *ASCE Journal of Engineering Mechanics*, 112(1):85–104, 1986.
[15] S. Geyer, I. Papaioannou, and D. Straub. Cross entropy-based importance sampling using Gaussian densities revisited. *Structural Safety*, 76:15–27, 2019.
[16] B. Goller, H. J. Pradlwarter, and G. I. Schuëller. Reliability assessment in structural dynamics. *Journal of Sound and Vibration*, 332(10):2488–2499, 2013.
[17] S. Gupta and C. S. Manohar. Reliability analysis of randomly vibrating structures with parameter uncertainties. *Journal of Sound and Vibration*, 297(3-5):1000–1024, 2006.
[18] M. Hohenbichler and R. Rackwitz. Non-normal dependent vectors in structural safety. *ASCE Journal of the Engineering Mechanics Division*, 107(6):1227–1238, 1981.
[19] H. A. Jensen and M. A. Valdebenito. Reliability analysis of linear dynamical systems using approximate representations of performance functions. *Structural Safety*, 29(3):222–237, 2007.
[20] L. Katafygiotis and S. H. Cheung. Domain decomposition method for calculating the failure probability of linear dynamic systems subjected to Gaussian stochastic loads. *ASCE Journal of Engineering Mechanics*, 132(5):475–486, 2006.
[21] L. S. Katafygiotis and S. H. Cheung. Application of spherical subset simulation method and auxiliary domain method on a benchmark reliability study. *Structural Safety*, 29(3):194–207, 2007.
[22] L. S. Katafygiotis, S. H. Cheung, and K. V. Yuen. Spherical subset simulation S^3 for solving nonlinear dynamical reliability problems. *International Journal of Reliability and Safety*, 4(2/3):122–138, 2010.
[23] R. Kolar, R. Jirik, and J. Jan. Estimator comparison of the Nakagami-m parameter and its application in echocardiography. *Radioengineering*, 13(1):8–12, 2004.
[24] P. S. Koutsourelakis, H. J. Pradlwarter, and G. I. Schuëller. Reliability of structures in high dimensions, part I: algorithms and applications. *Probabilistic Engineering Mechanics*, 19(4):409–417, 2004.
[25] N. Kurtz and J. Song. Cross-entropy-based adaptive importance sampling using Gaussian mixture. *Structural Safety*, 42:35–44, 2013.
[26] J. Latz, I. Papaioannou, and E. Ullmann. Multilevel sequential Monte Carlo for Bayesian inverse

problems. *Journal of Computational Physics*, 368:154–178, 2018.
[27] M. Lemaire. *Structural Reliability*. John Wiley & Sons, 2013.
[28] C. Q. Li and R. E. Melchers. Outcrossings from convex polyhedrons for nonstationary Gaussian processes. *ASCE Journal of Engineering Mechanics*, 119(11):2354–2361, 1993.
[29] J. Li. Probability density evolution method: background, significance and recent developments. *Probabilistic Engineering Mechanics*, 44:111–117, 2016.
[30] Y. K. Lin and G. Q. Cai. *Probabilistic structural dynamics: advanced theory and applications*. McGraw-Hill, 1995.
[31] L. D. Lutes and S. Sarkani. *Random vibrations: analysis of structural and mechanical systems*. Butterworth-Heinemann, 2004.
[32] R. E. Melchers and A. T. Beck. *Structural reliability analysis and prediction*. John Wiley & Sons, 2018.
[33] R. M. Neal. Annealed importance sampling. *Statistics and Computing*, 11(2):125–139, 2001.
[34] C. Papadimitriou, J. L. Beck, and L. S. Katafygiotis. Asymptotic expansions for reliability and moments of uncertain systems. *ASCE Journal of Engineering Mechanics*, 123(12):1219–1229, 1997.
[35] I. Papaioannou, C. Papadimitriou, and D. Straub. Sequential importance sampling for structural reliability analysis. *Structural Safety*, 62:66–75, 2016.
[36] I. Papaioannou, K. Breitung, and D. Straub. Reliability sensitivity estimation with sequential importance sampling. *Structural Safety*, 75:24–34, 2018.
[37] 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.
[38] H. J. Pradlwarter and G. I. Schuëller. Uncertain linear structural systems in dynamics: Efficient stochastic reliability assessment. *Computers & Structures*, 88(1-2):74–86, 2010.
[39] H. J. Pradlwarter, G. I. Schueller, P. S. Koutsourelakis, and D. C. Charmpis. Application of line sampling simulation method to reliability benchmark problems. *Structural Safety*, 29(3):208–221, 2007.
[40] S. O. Rice. Mathematical analysis of random noise. *Bell System Technical Journal*, 23(3):282–332, 1944.
[41] R. Y. Rubinstein and D. P. Kroese. *Simulation and the Monte Carlo method*, volume 10. John Wiley & Sons, 2016.
[42] 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.
[43] G. I. Schuëller, H. J. Pradlwarter, and P. S. Koutsourelakis. A critical appraisal of reliability estimation procedures for high dimensions. *Probabilistic Engineering Mechanics*, 19(4):463–474, 2004.
[44] M. Shinozuka and W. F. Wu. On the first passage problem and its application to earthquake engineering. In *Proc. 9th World Conf. on Earthquake Engineering*, volume 8, pages 767–72, 1988.
[45] M. A. Valdebenito, H. A. Jensen, and A. A. Labarca. Estimation of first excursion probabilities for uncertain stochastic linear systems subject to Gaussian load. *Computers & Structures*, 138:36–48, 2014.
[46] E. H. Vanmarcke. On the distribution of the first-passage time for normal stationary random processes. *ASME Journal of Applied Mechanics*, 42(1):215–220, 1975.
[47] 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.
[48] Y. K. Wen and H. C. Chen. On fast integration for time variant structural reliability. *Probabilistic Engineering Mechanics*, 2(3):156–162, 1987.
[49] Q. Zhou, W. Fan, Z. Li, and M. Ohsaki. Time-variant system reliability assessment by probability density evolution method. *ASCE Journal of Engineering Mechanics*, 143(11):04017131, 2017.

Appendix A. Sample generation from the IS density $h_{\boldsymbol{\Xi}}(\boldsymbol{\xi}|\boldsymbol{\theta})$ in Eq. (19)

The steps for generating a sample $\boldsymbol{\xi}$ distributed according to the IS density $h_{\boldsymbol{\Xi}}(\boldsymbol{\xi}|\boldsymbol{\theta})$ defined in Eq. (19) are summarized in Algorithm 3. The algorithm is as proposed in [2].

Algorithm 3: Sample generation from the IS density $h_{\boldsymbol{\Xi}}(\boldsymbol{\xi}|\boldsymbol{\theta})$

input:

- Weights of IS density $h_{\boldsymbol{\Xi}}(\boldsymbol{\xi}|\boldsymbol{\theta})$: $\left\{w_k(\boldsymbol{\theta}) = \mathrm{P}\left[F_k(\boldsymbol{\theta})\right] / \sum_{j=1}^{n_T} \mathrm{P}\left[F_j(\boldsymbol{\theta})\right], k = 1, \ldots, n_T\right\}$. The probabilities of the elementary failure events $\{F_k(\boldsymbol{\theta}), k = 1, \ldots, n_T\}$ are calculated analytically based on Eq. (14).
- Design points associated with the boundaries of the elementary failure events: $\left\{\boldsymbol{\xi}_k^*(\boldsymbol{\theta}) = h^* \frac{\boldsymbol{r}_k(\boldsymbol{\theta})}{\|\boldsymbol{r}_k(\boldsymbol{\theta})\|^2}, k = 1, \ldots, n_T\right\}$. h^* is the threshold value of the critical response and $\boldsymbol{r}_k(\boldsymbol{\theta})$ is as defined in Eq. (3).

sampling:

- Draw an index k from the set $\{1, \ldots, n_T\}$. Each index k is selected with a probability proportional to $w_k(\boldsymbol{\theta})$.
- Simulate $\boldsymbol{\xi}'$ as a $n_{\boldsymbol{\xi}}$-dimensional standard Gaussian vector with independent components.
- Simulate U_1 and U_2 as uniform variables on $[0, 1]$.
- Compute $\alpha = \Phi^{-1}\left[U_1 + (1 - U_1)\Phi\left(\beta_k(\boldsymbol{\theta})\right)\right]$. $\beta_k(\boldsymbol{\theta}) = \|\boldsymbol{\xi}_k^*(\boldsymbol{\theta})\|$ is the Euclidean norm of the the design point $\boldsymbol{\xi}_k^*(\boldsymbol{\theta})$.
- Compute the unit vector $\boldsymbol{u}_k^* = \boldsymbol{\xi}_k^*(\boldsymbol{\theta}) / \|\boldsymbol{\xi}_k^*(\boldsymbol{\theta})\|$.
- For a single-sided barrier, set $\boldsymbol{\xi} = \boldsymbol{\xi}' + (\alpha - \boldsymbol{\xi}'^{\mathrm{T}} \boldsymbol{u}_k^*)\boldsymbol{u}_k^*$

  For a double-sided barrier, set $\boldsymbol{\xi} = \begin{cases} \boldsymbol{\xi}' + \left(\alpha - \boldsymbol{\xi}'^{\mathrm{T}} \boldsymbol{u}_k^*\right) \boldsymbol{u}_k^* & \text{if } U_2 \leq 1/2 \\ -\boldsymbol{\xi}' - \left(\alpha - \boldsymbol{\xi}'^{\mathrm{T}} \boldsymbol{u}_k^*\right) \boldsymbol{u}_k^* & \text{otherwise} \end{cases}$

output:

- $\boldsymbol{\xi}$, a random sample distributed according to $h_{\boldsymbol{\Xi}}(\boldsymbol{\xi}|\boldsymbol{\theta})$.