A theoretically-consistent parallel enrichment strategy for Bayesian active learning reliability analysis

Tong Zhou, Tong Guo, Xujia Zhu, Masaru Kitahara, Jize Zhang

A theoretically-consistent parallel enrichment strategy for Bayesian active learning reliability analysis

Tong Zhou[a], Tong Guo[b], Xujia Zhu[c], Masaru Kitahara[d], Jize Zhang[a,*]

[a]*Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong*
[b]*School of Civil Engineering, Southeast University, Nanjing 210098, China*
[c]*Laboratoire des signaux et systèmes, Université Paris-Saclay, CNRS, CentraleSupélec, Gif-sur-Yvette 91190, France*
[d]*Department of Civil Engineering, The University of Tokyo, Tokyo 113-8656, Japan*

Abstract

Although parallel active learning reliability analysis is promising and has been widely studied, there remains an open question regarding how to achieve better theoretical consistency and avoid reliance on empirical practices heavily. A new parallel Bayesian active learning reliability method is developed in this study. First, in Bayesian failure probability estimation, a metric called integrated probability of misclassification (IPM) is defined from the upper bound of mean absolute deviation of failure probability. Then, a multi-point learning function called k-point integrated probability of misclassification reduction (k-IPMR) is proposed to guide the selection of a batch of $k(\geq 1)$ new samples that maximize the expected reduction of IPM. To further reduce the computational overhead, fast k-IPMR-guided parallel Bayesian active learning reliability analysis is conducted through four key workarounds. (i) The k-IPMR is substituted by its theoretically analogous but computationally cheaper variant. (ii) A stepwise maximization of k-IPMR is deployed to replace the cumbersome direct maximization approach. (iii) The number of new samples added per iteration is identified in an adaptive manner. (iv) A hybrid convergence criterion is specified according to the actual reduction of IPM at each iteration. Owing to the core role of IPM, we fuse the three major ingredients, i.e., Bayesian inference of failure probability, multi-point enrichment process, and convergence criterion, in a theoretically consistent way. The performance of the proposed method is testified on four examples of varying complexity. The results indicate that the proposed approach needs a fewer number of iterations than those existing ones and thus is more computationally efficient, particularly when dealing with time-intensive complex reliability problems.

Keywords: Multi-point learning function, Parallel enrichment, Stepwise maximization, Theoretical consistency, Bayesian active learning, Reliability analysis

Nomenclature

AK–MCS	Adaptive Kriging – Monte Carlo simulation	IPMR	integrated probability of misclassification reduction
ALR	Active learning reliability	MCS	Monte Carlo simulation
BALR	Bayesian active learning reliability	PABQ	parallel adaptive Bayesian quadrature
CDF	cumulative distribution function	PDEM	probability density evolution method
CV	coefficient of variation	PDF	probability density function
ED	experimental design	PM	probability of misclassification
EM	ensemble of metamodels	SUR	stepwise uncertainty reduction
IPM	integrated probability of misclassification	UBVC	upper-bound variance contribution
		VAIS	variance-amplified importance sampling

1. Introduction

In probabilistic reliability analysis, a major task is to compute the probability of failure of an engineering system, taking into account various sources of uncertainty inherent in its physical properties, external loads, operating conditions, etc [1]. Basically, those uncertainties can be represented through a vector of d random variables $\boldsymbol{X} = \{X_1, \ldots, X_d\} \in \mathbb{X} \subset \mathbb{R}^d$, with a known joint probability density function (PDF) $f_{\boldsymbol{X}}(\boldsymbol{x})$. Here, $\boldsymbol{X}$

*Corresponding author.
Email addresses: tong.ce.zhou@outlook.com (Tong Zhou), guotong@seu.edu.cn (Tong Guo), xujia.zhu@l2s.centralesupelec.fr (Xujia Zhu), kitahara@bridge.t.u-tokyo.ac.jp (Masaru Kitahara), cejize@ust.hk (Jize Zhang)

is assumed to be a standard Gaussian vector, and a transformation (e.g., isoprobabilistic, Nataf, or Rosenblatt transformation) can be employed to project the original random vector to a standard Gaussian one, if necessary. Then, the state of an engineering system is encoded by the so-called performance function $\mathcal{G}(\boldsymbol{x}) : \mathbb{R}^d \to \mathbb{R}^1$. Conventionally, the system is assumed to be in a failure state (resp. safe state) when $\mathcal{G}(\boldsymbol{x}) \leq 0$ (resp. $\mathcal{G}(\boldsymbol{x}) > 0$). Then, the failure probability P_f can be defined as [1]

$$P_f = \mathbb{P}\left(\mathcal{G}(\boldsymbol{X}) \leq 0\right) = \int_{\mathbb{X}} \mathbf{1}_{\mathcal{G}}\left(\boldsymbol{x}\right) f_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}, \tag{1}$$

where $\mathbb{P}\left(\cdot\right)$ denotes the probability measure; $\mathbf{1}_{\mathcal{G}}\left(\boldsymbol{x}\right)$ is the failure indicator function, expressed as

$$\mathbf{1}_{\mathcal{G}}\left(\boldsymbol{x}\right) = \begin{cases} 1, & \mathcal{G}(\boldsymbol{x}) \leq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Eq. (1) generally has no analytical solution in realistic scenarios, and a variety of numerical reliability methods have been developed in the past decades. They can be basically categorized into the following five groups. (i) *Approximation methods*, such as the first-/second-order reliability methods [2, 3]. (ii) *Sampling methods*, comprising the well-known Monte Carlo simulation (MCS) and those variance reduction techniques, e.g., importance sampling [4], subset simulation [5], and line sampling [6]. (iii) *Moment methods*, consisting of integer moment-based and fractional moment-based methods [7, 8]. (iv) *Probability preservation-based methods*, including probability density evolution method (PDEM) [9, 10] and its variants, e.g., direct probability integral method [11]. (v) *Active learning reliability (ALR) methods* [12, 13], which have recently gained increasing popularity due to their unique advantage of offering substantial savings in computational budget. Two pioneering contributions in this category are the efficient global reliability analysis [14] and the Adaptive Kriging – Monte Carlo simulation (AK–MCS) [15]. This is also the main focus of this study.

The ALR paradigm mainly consists of four key ingredients, namely surrogate model, reliability estimation algorithm, learning function, and convergence criterion [13]. (i) Commonly-used surrogate models include Kriging [16, 17], polynomial chaos expansion [18], support vector regression [19], radial basis function [20, 21], and ensemble of metamodels (EM) [22], to name just a few. Kriging is arguably the most popular one, owing to its exact interpolation property and the inherent ability to quantify epistemic uncertainty over predictions. (ii) Simulation methods [15, 23] and PDEM [24, 25] are often advocated to be used with a surrogate model to provide the failure probability estimate, owing to their fair guarantee of accuracy. (iii) Learning functions act as an "engine" of the ALR paradigm, distinguishing it from other categories of reliability methods. Typical Kriging-based learning functions include the U function [15], the expected feasibility function [14], the margin probability function [26], and so on. Essentially, those learning functions are designed in a spirit of achieving a pointwise trade-off between the exploitation (closeness of Kriging mean to the limit state) and exploration (high Kriging variance). Hence, they are collectively called pointwise learning functions in [27]. (iv) Convergence criterion is often defined according to the maximum/minimum value of a pointwise learning function. However, this practice is often too conservative [28] due to the fact that a pointwise learning function only reflects local uncertainty of failure probability. Then, some convergence criteria are defined with a direct link to the failure probability estimate, such as those defined based on the confidence bound or stabilization of failure probability estimate [28, 13, 25]. The readers are referred to [12, 13] for a comprehensive review of recent advances in the ALR framework.

More recently, Bayesian active learning reliability (BALR) method [29, 30], acting as a sub-category of the general ALR framework, has emerged as a promising approach. By treating Kriging-based failure probability estimation as a Bayesian inference problem, it seeks to infer the posterior statistics of failure probability. The posterior mean is a natural estimate of failure probability, while the posterior variance acts as a global measure of residual uncertainty about failure probability. Such a residual uncertainty can guide the design of both learning function and stopping condition. Obviously, a fair fusion of Bayesian inference and active learning distinguishes the BALR methods from other traditional ALR methods [31]. Initially, a learning function called stepwise uncertainty reduction (SUR) [29] was proposed with the direct aim of reducing the posterior variance of failure probability. Unfortunately, both the posterior variance of failure probability and the SUR are expressed as analytically-intractable double integral; hence, they remain too computationally demanding to use in practice [32]. To alleviate such heavy computational burden, some alternative measures of residual uncertainty of failure probability were proposed, expressed as computationally cheaper single integrals. Typical ones include the integrated Bernoulli variance [32], the integrated probability of misclassification [33], the integrated margin probability [34], and the limit-state margin volume [35], etc.

According to those residual uncertainty measures above, the basic principles of designing a learning function in the literature can be mainly classified into the following three categories. (i) *The integrand approach.* The learning function is simply defined as the integrand of a residual uncertainty measure, with an intuitive belief that adding the point with the maximum integrand will contribute the most to reducing this residual

uncertainty measure. Typical ones include the upper-bound variance contribution (UBVC) [30], the expected misclssification probability contribution [33], and the left-/right-shifted contribution [34]. In essence, they belong to the category of pointwise learning functions. Hence, this intuitive approach loses the power of explicitly quantifying the reduction of residual uncertainty of failure probability due to adding a new sample. (ii) *Plain double integral approach*. The learning function is defined as the expected reduction of a residual uncertainty measure incurred by adding a new sample and is crudely expressed as a double integral. Typical ones include the expected uncertainty reduction [36], the expected integrated error reduction [37], and the sampling-based expected uncertainty reduction [38]. Due to analytical intractability, those learning functions have to be numerically computed as a nested summation, which is very computationally cumbersome. (iii) *Look-ahead approach*. Similar to the second category, the look-ahead learning function is designed to select a new sample that reduces, at most, a specific residual uncertainty measure of failure probability in expectation. With the help of Kriging update formulas, the look-ahead learning function could be reduced to a single integral. Typical ones comprise the second version of SUR [32], the expected margin volume reduction [35], the integrated probability of misclassification reduction [39], the stepwise margin reduction [40], the SUR-Bichon function [41], etc. Obviously, those look-head learning functions come with much lower computational cost than those in the second category. However, those aforementioned learning functions are generally applicable to the single-point scenario.

With the widespread availability of distributed computing facilities, parallel (B)ALR methods have gained increasing attention in the past decade. They intend to select a batch of new training samples at which to evaluate the costly performance function in parallel, so as to maximize the resource utilization and save the total computational time. Those pointwise learning functions have difficulty in extending their analytical expressions to multi-point scenario, due to their point-to-point nature. Hence, existing parallel (B)ALR methods are often implemented by combining a single-point learning function with some additional multi-point selection strategies. Those additional strategies can be grossly grouped into the following four categories. (i) *Clustering approach*, e.g., the K-means clustering method [28, 42, 30] or the K-medoids clustering method [43, 23]. (ii) *Kriging retraining approach*, such as Kriging believer strategy [44, 45], or constant liar strategy [46]. (iii) *Pseudo learning function approach*, e.g., the kernel function-based methods [47, 48, 49]. (iv) *EM-based approach*, such as [50, 22]. Notably, the number, K, of new samples added per iteration has to be prescribed in those strategies, and it is very challenging to specify a rational value of K *a priori*. Due to the inherent flaw of pointwise learning functions, such practice is also unable to quantify the potential reduction of residual uncertainty measure of failure probability brought by a batch of new samples. Hence, the resulting batch of new samples is not optimal for reducing the residual uncertainty measure to some extent [31]. By comparison, the look-ahead learning functions have the potential of directly measuring the impact of adding a batch of new samples on reducing the residual uncertainty measure of failure probability. Hence, it is feasible to conduct the multi-point enrichment process through a multi-point learning function directly, eliminating the need for additional batch selection strategies. If so, the parallel BALR framework will be more theoretically consistent and numerically elegant.

To fill in the research gap, a new parallel Bayesian active learning reliability method is proposed. It is devoted to developing a multi-point look-ahead learning function based on a computationally cheap residual uncertainty measure of failure probability, whereby the parallel enrichment process can be efficiently implemented. The main contributions of this study are summarized as follows.

- The metric integrated probability of misclassification (IPM) is rigorously proved as the upper bound of mean absolute deviation of failure probability within Bayesian inference. Hence, it can be viewed as a global measure of residual uncertainty of failure probability and is much computationally cheaper than the posterior variance of failure probability.
- A multi-point learning function called k-point integrated probability of misclassification reduction (k-IPMR) is proposed. It enables quantifying the gain of adding a batch of $k(\geq 1)$ new samples on reducing the IPM in expectation.
- The multi-point enrichment process can be conducted based on the k-IPMR solely. Moreover, an adaptive scheme is proposed to specify the rational number of new samples added per iteration, instead of the traditional prescribed manner.
- A hybrid stopping condition is developed from the actual reduction of IPM. Thanks to the central role of IPM, the three key ingredients, i.e., Bayesian inference of failure probability, learning function (and parallel enrichment), and stopping condition, are fused in a theoretically consistent way.

The remainder of this paper is structured as follows. Section 2 revisits several related studies. Section 3 devotes to both theoretical derivation and numerical computation of a multi-point learning function k-IPMR. Section 4 details how to efficiently conduct the k-IPMR-guided multi-point enrichment process. The proposed approach is illustrated in Section 5 through four distinct examples. This paper closes with some concluding remarks in Section 6.

2. Preliminaries

Section 2.1 reviews Bayesian inference of failure probability. Then, Section 2.2 outlines the multi-point Kriging update formulas.

2.1. Bayesian inference of failure probability

Treating the performance function $\mathcal{G}(\boldsymbol{x})$ as one sample path of Kriging $\widehat{\mathcal{G}}_n(\boldsymbol{x}) \sim \mathcal{GP}\left(\mu_n(\boldsymbol{x}), c_n(\boldsymbol{x}, \boldsymbol{x}')\right)$, we can infer the epistemic uncertainty of $\widehat{\mathbf{1}}_n(\boldsymbol{x})$ and further $\widehat{P}_{f,n}$ arising from the lack of exact knowledge of $\mathcal{G}(\boldsymbol{x})$, as enclosed by the yellow box in Fig. 1. Further, to reduce the epistemic uncertainty of $\widehat{P}_{f,n}$, two key ingredients in the ALR, i.e., learning function and convergence criterion, can be designed in a goal-oriented manner, as visualized by the red arrows in Fig. 1. Obviously, Bayesian inference and active learning are well fused in the BALR framework.

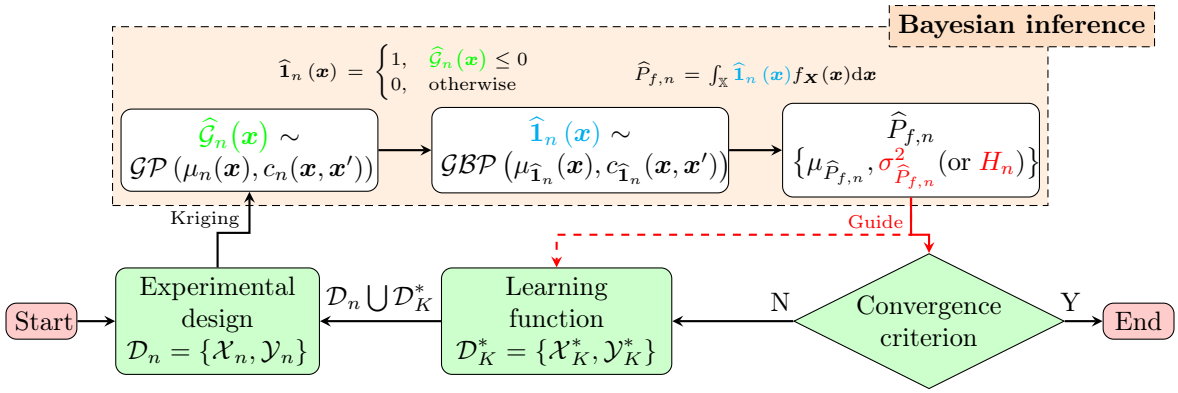


Figure 1: Basic workflow of Bayesian active learning reliability analysis

For brevity, the basics of Kriging are given in Appendix A. When providing an experimental design (ED) of size n, denoted as $\mathcal{D}_n = \{\mathcal{X}_n, \mathcal{Y}_n\} = \left\{\left(\boldsymbol{x}^{(i)}, y^{(i)}\right), i = 1, \ldots, n\right\}$, a Kriging $\widehat{\mathcal{G}}_n(\boldsymbol{x})$ can be trained, with its mean $\mu_n(\boldsymbol{x})$, variance $\sigma_n^2(\boldsymbol{x})$, and covariance $c_n(\boldsymbol{x}, \boldsymbol{x}')$ given by Eqs. (A.7), (A.8), and (A.9), respectively. Then, substituting the Kriging $\widehat{\mathcal{G}}_n(\boldsymbol{x})$ to Eq. (2), the failure indicator function $\mathbf{1}_{\mathcal{G}}(\boldsymbol{x})$ is recast as

$$\widehat{\mathbf{1}}_n(\boldsymbol{x}) = \begin{cases} 1, & \widehat{\mathcal{G}}_n(\boldsymbol{x}) \le 0, \\ 0, & \text{otherwise}, \end{cases} \tag{3}$$

which follows a generalized Bernoulli process, that is, $\widehat{\mathbf{1}}_n(\boldsymbol{x}) \sim \mathcal{GBP}\left(\mu_{\widehat{\mathbf{1}}_n}(\boldsymbol{x}), c_{\widehat{\mathbf{1}}_n}(\boldsymbol{x}, \boldsymbol{x}')\right)$, with its mean $\mu_{\widehat{\mathbf{1}}_n}(\boldsymbol{x})$, variance $\sigma^2_{\widehat{\mathbf{1}}_n}(\boldsymbol{x})$, and covariance $c_{\widehat{\mathbf{1}}_n}(\boldsymbol{x}, \boldsymbol{x}')$ expressed as [33]

$$\mu_{\widehat{\mathbf{1}}_n}(\boldsymbol{x}) = \Phi\left(-\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right), \tag{4}$$

$$\sigma^2_{\widehat{\mathbf{1}}_n}(\boldsymbol{x}) = \Phi\left(-\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right)\Phi\left(\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right), \tag{5}$$

$$c_{\widehat{\mathbf{1}}_n}(\boldsymbol{x}, \boldsymbol{x}') = \Phi_2\left(\begin{bmatrix}0\\0\end{bmatrix}; \begin{bmatrix}\mu_n(\boldsymbol{x})\\ \mu_n(\boldsymbol{x}')\end{bmatrix}, \begin{bmatrix}\sigma_n^2(\boldsymbol{x}) & c_n(\boldsymbol{x}, \boldsymbol{x}')\\ c_n(\boldsymbol{x}', \boldsymbol{x}) & \sigma_n^2(\boldsymbol{x}')\end{bmatrix}\right) - \Phi\left(-\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right)\Phi\left(-\frac{\mu_n(\boldsymbol{x}')}{\sigma_n(\boldsymbol{x}')}\right), \tag{6}$$

respectively, where $\Phi(\cdot)$ denotes the cumulative distribution function (CDF) of a standard Gaussian variable; $\Phi_2(\cdot; \boldsymbol{\mu}, \mathcal{C})$ represents the joint CDF of a bivariate Gaussian vector with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathcal{C}$. Note that $\Phi_2(\cdot; \cdot, \cdot)$ has no analytical solution and needs numerical calculation.

Further, substituting Eq. (3) into Eq. (1), the failure probability is estimated as

$$\widehat{P}_{f,n} = \int_{\mathbb{X}} \widehat{\mathbf{1}}_n(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}, \tag{7}$$

which is still a random variable through $\widehat{\mathbf{1}}_n(\cdot)$. Although the exact distribution of $\widehat{P}_{f,n}$ is unavailable, its mean $\mu_{\widehat{P}_{f,n}}$ and variance $\sigma^2_{\widehat{P}_{f,n}}$ can be obtained as [33]

$$\mu_{\widehat{P}_{f,n}} = \int_{\mathbb{X}} \Phi\left(-\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right) f_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}, \tag{8}$$

$$\sigma^2_{\widehat{P}_{f,n}} = \iint_{\mathbb{X}\times\mathbb{X}} \Phi_2\left(\begin{bmatrix}0\\0\end{bmatrix}; \begin{bmatrix}\mu_n(\boldsymbol{x})\\ \mu_n(\boldsymbol{x}')\end{bmatrix}, \begin{bmatrix}\sigma_n^2(\boldsymbol{x}) & c_n(\boldsymbol{x}, \boldsymbol{x}')\\ c_n(\boldsymbol{x}, \boldsymbol{x}') & \sigma_n^2(\boldsymbol{x}')\end{bmatrix}\right) f_{\boldsymbol{X}}(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}') \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{x}' - \mu^2_{\widehat{P}_{f,n}}, \tag{9}$$

respectively. Then, $\mu_{\widehat{P}_{f,n}}$ is a natural estimate of $\widehat{P}_{f,n}$, while $\sigma^2_{\widehat{P}_{f,n}}$ can be viewed as a measure of residual uncertainty of $\widehat{P}_{f,n}$.

Generally, both $\mu_{\widehat{P}_{f,n}}$ and $\sigma^2_{\widehat{P}_{f,n}}$ have no analytical solution and require numerical integration. Take the variance-amplified importance sampling (VAIS) method for example, $\mu_{\widehat{P}_{f,n}}$ can be numerically computed as

$$\mu_{\widehat{P}_{f,n}} = \int_{\mathbb{X}} \Phi\left(-\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right) \frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{h_{\boldsymbol{X}}(\boldsymbol{x})} h_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \approx \frac{1}{Q_1} \sum_{q=1}^{Q_1} \Phi\left(-\frac{\mu_n\left(\boldsymbol{x}^{(q)}\right)}{\sigma_n\left(\boldsymbol{x}^{(q)}\right)}\right) \frac{f_{\boldsymbol{X}}\left(\boldsymbol{x}^{(q)}\right)}{h_{\boldsymbol{X}}\left(\boldsymbol{x}^{(q)}\right)} = \tilde{\mu}_{\widehat{P}_{f,n}}, \tag{10}$$

where $h_{\boldsymbol{X}}(\boldsymbol{x})$ denotes the importance sampling density function and can be simply specified as $\boldsymbol{\phi}(\boldsymbol{x}; \boldsymbol{0}, \alpha^2 \boldsymbol{I}_d)$, with $\boldsymbol{\phi}(\cdot)$ the joint PDF of a d-variate Gaussian vector, $\boldsymbol{I}_d$ an identity matrix of size $d \times d$, α the amplification coefficient and taken as 1.5 here [30]; $\mathcal{X}_{Q_1} = \left\{\boldsymbol{x}^{(q)}, q = 1, \ldots, Q_1\right\}$ denotes a set of Q_1 quadrature points drawn from $h_{\boldsymbol{X}}(\boldsymbol{x})$. Moreover, the sample variance of $\tilde{\mu}_{\widehat{P}_{f,n}}$ is expressed as

$$\mathrm{Var}\left[\tilde{\mu}_{\widehat{P}_{f,n}}\right] = \frac{1}{Q_1 \times (Q_1 - 1)} \sum_{q=1}^{Q_1} \left[\Phi\left(-\frac{\mu_n\left(\boldsymbol{x}^{(q)}\right)}{\sigma_n\left(\boldsymbol{x}^{(q)}\right)}\right) \frac{f_{\boldsymbol{X}}\left(\boldsymbol{x}^{(q)}\right)}{h_{\boldsymbol{X}}\left(\boldsymbol{x}^{(q)}\right)} - \tilde{\mu}_{\widehat{P}_{f,n}}\right]^2, \tag{11}$$

then, the sample coefficient of variation (CV) is given by $\mathrm{CV}\left[\tilde{\mu}_{\widehat{P}_{f,n}}\right] = \sqrt{\mathrm{Var}\left[\tilde{\mu}_{\widehat{P}_{f,n}}\right]} / \tilde{\mu}_{\widehat{P}_{f,n}}$. Basically, the sample size Q_1 is considered sufficient when $\mathrm{CV}\left[\tilde{\mu}_{\widehat{P}_{f,n}}\right] \leq \varepsilon_{\mathrm{Q}}$, with the tolerance ε_{Q} set as 5% here.

Unlike $\mu_{\widehat{P}_{f,n}}$, $\sigma^2_{\widehat{P}_{f,n}}$ is expressed as a double integral of the bivariate Gaussian CDF $\Phi_2(\cdot;\cdot,\cdot)$. Its calculation is very computationally demanding, due to the complexity and non-analytical nature of $\Phi_2(\cdot;\cdot,\cdot)$. To this end, a computationally cheaper alternative of $\sigma^2_{\widehat{P}_{f,n}}$ will be proposed in Section 3.

2.2. Multi-point Kriging update formulas

To define a learning function that quantifies the epistemic uncertainty reduction of failure probability, we need to explicitly measure the impact of adding a batch of new samples on the posterior of Kriging. This can be realized through the so-called Kriging update formulas [51].

Specifically, suppose that a batch of k new samples $\mathcal{X}_k^+ = \left\{\boldsymbol{x}_+^{(1)}, \ldots, \boldsymbol{x}_+^{(k)}\right\}$ and their model responses $\mathcal{Y}_k^+ = \left\{y_+^{(1)}, \ldots, y_+^{(k)}\right\}$ are added into the ED $\mathcal{D}_n$, the Kriging update formulas are expressed as [51]

$$\mu_{n+k}(\boldsymbol{x}) = \mu_n(\boldsymbol{x}) + c_n\left(\boldsymbol{x}, \mathcal{X}_k^+\right)^\top \left(\mathcal{C}_k^+\right)^{-1} \left(\mathcal{Y}_k^+ - \mu_n\left(\mathcal{X}_k^+\right)\right), \tag{12}$$

$$\sigma^2_{n+k}(\boldsymbol{x}) = \sigma^2_n(\boldsymbol{x}) - c_n\left(\boldsymbol{x}, \mathcal{X}_k^+\right)^\top \left(\mathcal{C}_k^+\right)^{-1} c_n\left(\boldsymbol{x}, \mathcal{X}_k^+\right), \tag{13}$$

$$c_{n+k}(\boldsymbol{x}, \boldsymbol{x}') = c_n(\boldsymbol{x}, \boldsymbol{x}') - c_n\left(\boldsymbol{x}, \mathcal{X}_k^+\right)^\top \left(\mathcal{C}_k^+\right)^{-1} c_n\left(\boldsymbol{x}', \mathcal{X}_k^+\right), \tag{14}$$

which are referred to as the *look-ahead* mean, variance, and covariance of Kriging, respectively. $c_n\left(\boldsymbol{x}, \mathcal{X}_k^+\right) \coloneqq \left[c_n\left(\boldsymbol{x}, \boldsymbol{x}_+^{(1)}\right), \ldots, c_n\left(\boldsymbol{x}, \boldsymbol{x}_+^{(k)}\right)\right]^\top$ is a $k \times 1$ vector of covariances between $\boldsymbol{x}$ and all new points in $\mathcal{X}_k^+$, $\mathcal{C}_k^+ \coloneqq \left[c_n\left(\boldsymbol{x}_+^{(i)}; \boldsymbol{x}_+^{(j)}\right)\right]_{1 \leq i,j \leq k}$ is a $k \times k$ matrix of covariances among all new points in $\mathcal{X}_k^+$.

Clearly, $\mu_{n+k}(\boldsymbol{x})$, $\sigma^2_{n+k}(\boldsymbol{x})$, and $c_{n+k}(\boldsymbol{x}, \boldsymbol{x}')$ are directly obtained from some matrix manipulations on the current posterior of Kriging, without re-optimizing its kernel parameters $\boldsymbol{\theta}$ [35]. Hence, the look-ahead posterior of Kriging is much computationally cheaper than the retraining of Kriging. Thanks to both analytical tractability and low computational cost of Kriging update formulas, a theoretically sound and computationally fast multi-point learning function will be proposed in Section 3.

3. A multi-point learning function *k*-IPMR

Section 3.1 outlines a measure of residual uncertainty of failure probability estimated by Kriging. On this basis, a multi-point learning function called k-IPMR is defined in Section 3.2 and its upper bound is then given in Section 3.3. Subsequently, Section 3.4 presents the setting of single numerical evaluation of k-IPMR. Finally, Section 3.5 elucidates the significant computational burden faced by the direct k-IPMR-guided multi-point enrichment process.

3.1. A residual uncertainty measure of failure probability

Proposition 1. *Denote the integrated probability of misclassification (IPM) as*

$$H_n = \mathbb{E}_{\boldsymbol{X}}\left[\pi_n(\boldsymbol{x})\right] = \int_{\mathbb{X}} \pi_n(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}, \tag{15}$$

where $\mathbb{E}_{\boldsymbol{X}}[\cdot]$ stands for the expectation with respect to $\boldsymbol{X}$; $\pi_n(\boldsymbol{x})$ is called probability of misclassification (PM), expressed as

$$\pi_n(\boldsymbol{x}) = \Phi\left(-\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}\right), \tag{16}$$

which denotes the probability of mis-classifying the failure/safe state at $\boldsymbol{x}$ according to the Kriging mean $\mu_n(\boldsymbol{x})$, as illustrated in Fig. 2.

Then, the mean absolute deviation of $\widehat{P}_{f,n}$ satisfies

$$\mathbb{E}_n\left[\left|\widehat{P}_{f,n} - \mu_{\widehat{P}_{f,n}}\right|\right] \leq 2H_n, \tag{17}$$

where $\mathbb{E}_n[\cdot]$ denotes the expectation with respect to $\widehat{\mathcal{G}}_n(\boldsymbol{x})$.

The proof of Proposition 1 is given in Appendix B. Notably, $\pi_n(\boldsymbol{x})$ is exactly the theoretical underpinning of the well-known U function [15, 28]. Although similar definitions to the IPM (but different names) were empirically given in [33, 49], a rigorous proof that it acts as a global residual uncertainty measure of failure probability has not been provided yet.

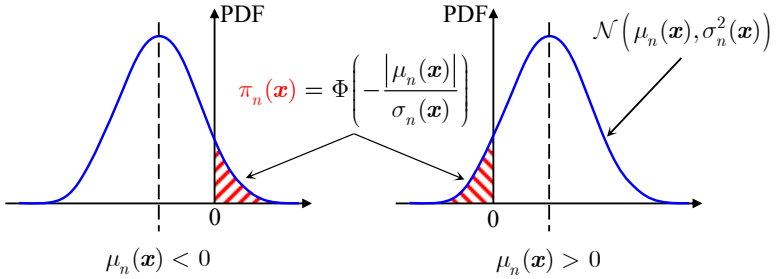


Figure 2: Schematic representation of the PM $\pi_n(\boldsymbol{x})$

Proposition 1 indicates that when the well-defined IPM $H_n \to 0$, $\mu_{\widehat{P}_{f,n}}$ converges to the actual failure probability P_f in expectation. Hence, similar to the variance $\sigma^2_{\widehat{P}_{f,n}}$, the IPM H_n can be viewed as an epistemic uncertainty measure of $\widehat{P}_{f,n}$. More importantly, the H_n is only expressed as a single integral of $\Phi(\cdot)$ and is much computationally cheaper than $\sigma^2_{\widehat{P}_{f,n}}$. Similarly, the VAIS-based estimation of H_n can be given as

$$H_n = \int_{\mathbb{X}} \pi_n(\boldsymbol{x}) \frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{h_{\boldsymbol{X}}(\boldsymbol{x})} h_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \approx \frac{1}{Q_2} \sum_{q=1}^{Q_2} \pi_n\left(\boldsymbol{x}^{(q)}\right) \frac{f_{\boldsymbol{X}}\left(\boldsymbol{x}^{(q)}\right)}{h_{\boldsymbol{X}}\left(\boldsymbol{x}^{(q)}\right)} = \tilde{H}_n, \tag{18}$$

where $\mathcal{X}_{\mathrm{Q}_2} = \left\{\boldsymbol{x}^{(q)}, q = 1, \ldots, Q_2\right\}$ denotes a set of Q_2 samples drawn from $h_{\boldsymbol{X}}(\boldsymbol{x})$. Further, the $\mathrm{CV}\left[\tilde{H}_n\right]$ can be computed, and the sample size Q_2 is then considered sufficient when $\mathrm{CV}\left[\tilde{H}_n\right] \leq \varepsilon_{\mathrm{Q}}$.

According to the definition of IPM H_n, a learning function can be designed to select a sequence of new samples that reduce, at most, the IPM in expectation, which will be detailed in Section 3.2.

3.2. Basic definition of the learning function k-IPMR

Suppose that a batch of k new samples and their $\mathcal{G}$-function responses, i.e., $\left\{\mathcal{X}_k^+, \mathcal{Y}_k^+\right\} = \left\{\left(\boldsymbol{x}_+^{(i)}, y_+^{(i)}\right), i = 1, \ldots, k\right\}$, are added into the current ED $\mathcal{D}_n$, the look-ahead mean $\mu_{n+k}(\boldsymbol{x})$, variance $\sigma^2_{n+k}(\boldsymbol{x})$ and covariance $c_{n+k}(\boldsymbol{x}, \boldsymbol{x}')$ of Kriging are given by Eqs. (12), (13) and (14), respectively.

Then, owing to the addition of $\left\{\mathcal{X}_k^+, \mathcal{Y}_k^+\right\}$, the look-ahead IPM $H_{n+k}\left(\mathcal{X}_k^+, \mathcal{Y}_k^+\right)$ can be expressed as

$$H_{n+k}\left(\mathcal{X}_k^+, \mathcal{Y}_k^+\right) = \mathbb{E}_{\boldsymbol{X}}\left[\pi_{n+k}\left(\boldsymbol{x}; \mathcal{X}_k^+, \mathcal{Y}_k^+\right)\right], \tag{19}$$

where the look-ahead PM $\pi_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+,\mathcal{Y}_k^+\right)$ is given as

$$\begin{aligned}\pi_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+,\mathcal{Y}_k^+\right) &= \Phi\left(-\frac{|\mu_{n+k}(\boldsymbol{x})|}{\sigma_{n+k}(\boldsymbol{x})}\right),\\ &= \Phi\left(-\frac{\left|\mu_n(\boldsymbol{x})+c_n\left(\boldsymbol{x},\mathcal{X}_k^+\right)^\top\left(\mathcal{C}_k^+\right)^{-1}\left(\mathcal{Y}_k^+-\mu_n\left(\mathcal{X}_k^+\right)\right)\right|}{\sigma_{n+k}(\boldsymbol{x})}\right),\\ &= \Phi\left(-\left|\frac{\mu_n(\boldsymbol{x})}{\sigma_{n+k}(\boldsymbol{x})}+\frac{c_n\left(\boldsymbol{x},\mathcal{X}_k^+\right)^\top\left(\mathcal{C}_k^+\right)^{-1}}{\sigma_{n+k}(\boldsymbol{x})}\left(\mathcal{Y}_k^+-\mu_n\left(\mathcal{X}_k^+\right)\right)\right|\right),\\ &= \Phi\left(-\left|a(\boldsymbol{x})+\boldsymbol{b}(\boldsymbol{x})^\top\mathcal{U}_k^+\right|\right),\end{aligned} \tag{20}$$

where $a(\boldsymbol{x})=\frac{\mu_n(\boldsymbol{x})}{\sigma_{n+k}(\boldsymbol{x})}$ and $\boldsymbol{b}(\boldsymbol{x})=\frac{\left(\mathcal{C}_k^+\right)^{-1}c_n\left(\boldsymbol{x},\mathcal{X}_k^+\right)}{\sigma_{n+k}(\boldsymbol{x})}$ are a scalar quantity and a $k\times 1$ vector independent of $\mathcal{Y}_k^+$, respectively; then, $\mathcal{U}_k^+=\mathcal{Y}_k^+-\mu_n\left(\mathcal{X}_k^+\right)$ is a $k\times 1$ vector hinging on $\mathcal{Y}_k^+$.

Since the responses $\mathcal{Y}_k^+$ cannot be known exactly without evaluating the actual $\mathcal{G}$-function on $\mathcal{X}_k^+$, the $\pi_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+,\mathcal{Y}_k^+\right)$ in Eq. (20) and the $H_{n+k}\left(\mathcal{X}_k^+,\mathcal{Y}_k^+\right)$ in Eq. (19) are unknown as well. To this end, the responses $\mathcal{Y}_k^+$ are replaced by the Kriging predictions evaluated on $\mathcal{X}_k^+$, i.e., $\boldsymbol{Y}_k^+=\left(\widehat{\mathcal{G}}_n\left(\boldsymbol{x}_+^{(1)}\right),\ldots,\widehat{\mathcal{G}}_n\left(\boldsymbol{x}_+^{(k)}\right)\right)^\top\sim\mathcal{N}_k\left(\mu_n\left(\mathcal{X}_k^+\right),\mathcal{C}_k^+\right)$. Then, the $H_{n+k}\left(\mathcal{X}_k^+,\mathcal{Y}_k^+\right)$ in Eq. (19) becomes a function of $\mathcal{X}_k^+$ solely, that is,

$$\mathcal{H}_{n+k}\left(\mathcal{X}_k^+\right)=\mathbb{E}_{\boldsymbol{X}}\left[\Pi_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)\right], \tag{21}$$

where $\Pi_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)$ is expressed as

$$\Pi_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)=\Phi\left(-\left|a(\boldsymbol{x})+\boldsymbol{b}(\boldsymbol{x})^\top\boldsymbol{U}_k^+\right|\right), \tag{22}$$

with $\boldsymbol{U}_k^+=\boldsymbol{Y}_k^+-\mu_n\left(\mathcal{X}_k^+\right)\sim\mathcal{N}\left(\boldsymbol{0},\mathcal{C}_k^+\right)$ following a d-variate centered Gaussian distribution with the covariance matrix $\mathcal{C}_k^+$.

In this way, the possible reduction of IPM brought by $\mathcal{X}_k^+$ is given as

$$\Delta\mathcal{H}_{n+k}\left(\mathcal{X}_k^+\right)=H_n-\mathcal{H}_{n+k}\left(\mathcal{X}_k^+\right), \tag{23}$$

which is actually a random quantity through $\boldsymbol{U}_k^+$ and still cannot be a deterministic criterion.

To this end, the learning function is defined by taking the expectation of $\Delta\mathcal{H}_{n+k}\left(\mathcal{X}_k^+\right)$ with respect to $\boldsymbol{U}_k^+$, that is,

$$\begin{aligned}\mathrm{IPMR}_{n+k}\left(\mathcal{X}_k^+\right) &= \mathbb{E}_{\boldsymbol{U}_k^+}\left[\Delta\mathcal{H}_{n+k}\left(\mathcal{X}_k^+\right)\right],\\ &= H_n-\mathbb{E}_{\boldsymbol{U}_k^+}\left[\mathbb{E}_{\boldsymbol{X}}\left[\Pi_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)\right]\right],\\ &= \mathbb{E}_{\boldsymbol{X}}\left[\pi_n(\boldsymbol{x})-\mathbb{E}_{\boldsymbol{U}_k^+}\left[\Pi_{n+k}(\boldsymbol{x};\mathcal{X}_k^+)\right]\right],\\ &= \mathbb{E}_{\boldsymbol{X}}\left[\Gamma_{n+k}(\boldsymbol{x};\mathcal{X}_k^+)\right],\end{aligned} \tag{24}$$

where $\Gamma_{n+k}(\boldsymbol{x};\mathcal{X}_k^+)=\pi_n(\boldsymbol{x})-\mathbb{E}_{\boldsymbol{U}_k^+}\left[\Pi_{n+k}(\boldsymbol{x};\mathcal{X}_k^+)\right]$ corresponds to the reduction of PM at any point $\boldsymbol{x}$ brought by $\mathcal{X}_k^+$ in expectation.

In essence, Eq. (24) can be interpreted as the average reduction of IPM brought by $\mathcal{X}_k^+$, conditional on the Kriging assumption of their $\mathcal{G}$-function responses. Hence, a batch of k best next points $\mathcal{X}_k^*=\left\{\boldsymbol{x}^{(n+1)},\ldots,\boldsymbol{x}^{(n+k)}\right\}$ can be selected as

$$\mathcal{X}_k^*=\underset{\mathcal{X}_k^+\in\mathcal{X}_C}{\arg\max}\,\mathrm{IPMR}_{n+k}(\mathcal{X}_k^+), \tag{25}$$

where $\mathcal{X}_C=\left\{\boldsymbol{x}^{(i)},i=1,\ldots,C\right\}$ denotes a candidate pool of size C.

As more and more batches of k best next samples are sequentially added by Eq. (25), the IPM will be reduced step by step. Hence, the learning function in Eq. (24) is called k-point integrated probability of misclassification reduction (k-IPMR) here.

Eq. (24) indicates that $\mathrm{IPMR}_{n+k}(\cdot)$ consists of two integrals, i.e., the inner integral of $\boldsymbol{U}_k^+$ and the outer integral of $\boldsymbol{X}$. Then, the computation of $\mathrm{IPMR}_{n+k}(\cdot)$ will be further detailed in Section 3.3.

3.3. Analysis and bound establishment for k-IPMR

Proposition 2. *In* $\mathrm{IPMR}_{n+k}(\mathcal{X}_k^+)$*, its inner integral* $\Gamma_{n+k}(\boldsymbol{x};\mathcal{X}_k^+)$ *is analytically expressed as*

$$\Gamma_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)=\Phi\left(\frac{-\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}}{\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)}\right)-2\Phi_2\left(\begin{bmatrix}\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}\\ -\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}/\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)\end{bmatrix};\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1 & -\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)\\ -\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right) & 1\end{bmatrix}\right), \tag{26}$$

which is a function of both $\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}$ *and* $\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)$*. Then, the* $\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)$ *is given by*

$$\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)=\frac{\sqrt{c_n\left(\boldsymbol{x},\mathcal{X}_k^+\right)^{\top}\left(\mathcal{C}_k^+\right)^{-1}c_n\left(\boldsymbol{x},\mathcal{X}_k^+\right)}}{\sigma_n(\boldsymbol{x})}\in[0,1], \tag{27}$$

which can be interpreted as the correlation coefficient between $\widehat{\mathcal{G}}_n(\boldsymbol{x})$ *and the subspace spanned by the random vector* $\widehat{\mathcal{G}}_n\left(\mathcal{X}_k^+\right)=\left\{\widehat{\mathcal{G}}_n\left(\boldsymbol{x}_+^{(1)}\right),\ldots,\widehat{\mathcal{G}}_n\left(\boldsymbol{x}_+^{(k)}\right)\right\}$*. The closer between* $\boldsymbol{x}$ *and* $\mathcal{X}_k^+$ *is, the greater* $\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)$ *will be, and vice versa.*

The proof of Proposition 2 is provided in Appendix C, and the geometrical interpretation of $\rho_{n+k}(\boldsymbol{x},\mathcal{X}_k^+)$ has been proved in [52].

Proposition 3. *The lower and upper bounds of* $\Gamma_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)$*, denoted as* $\underline{\Gamma}_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)$ *and* $\overline{\Gamma}_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)$*, are expressed as*

$$\begin{cases}\underline{\Gamma}_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)=0,\\ \overline{\Gamma}_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)=\Phi\left(\dfrac{-\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}}{\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)}\right).\end{cases} \tag{28}$$

The proof of Proposition 3 is given in Appendix D. Then, according to Proposition 3, there exists

$$0=\underline{\Gamma}_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)\le\Gamma_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)\le\overline{\Gamma}_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)\le\pi_n(\boldsymbol{x}), \tag{29}$$

which indicates that $\pi_n(\boldsymbol{x})$ is a looser upper bound of $\Gamma_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)$.

Fig. 3 illustrates the comparison between $\Gamma_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)$ and $\overline{\Gamma}_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)$ under different combinations of $\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}$ and $\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)$. Clearly, $\overline{\Gamma}_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)$ inflates the expected reduction of PM at point $\boldsymbol{x}$ when compared with $\Gamma_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)$, particularly in the vicinity of $\left\{\boldsymbol{x}\in\mathbb{X}:\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}=0\right\}$. Despite that, the global trend of $\overline{\Gamma}_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)$ agrees well with that of $\Gamma_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)$.

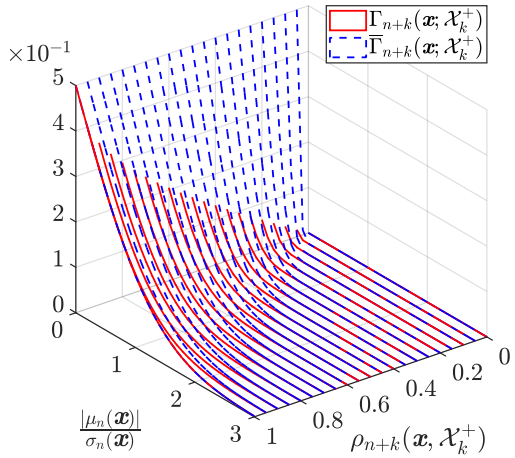


Figure 3: Comparison between $\Gamma_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)$ and $\overline{\Gamma}_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)$ in $\mathrm{IPMR}_{n+k}\left(\mathcal{X}_k^+\right)$

Substituting Eq. (26) into Eq. (24), $\mathrm{IPMR}_{n+k}\left(\mathcal{X}_k^+\right)$ finally reduces to a single integral, that is,

$$\begin{aligned}\mathrm{IPMR}_{n+k}\left(\mathcal{X}_k^+\right)&=\mathbb{E}_{\boldsymbol{X}}\left[\Gamma_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)\right],\\&=\mathbb{E}_{\boldsymbol{X}}\left[\Phi\left(\frac{-\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}}{\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)}\right)-2\Phi_2\left(\begin{bmatrix}\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}\\ -\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}/\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)\end{bmatrix};\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1 & -\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)\\ -\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right) & 1\end{bmatrix}\right)\right].\end{aligned} \tag{30}$$

Besides, as per Eq. (28), the upper bound of $\mathrm{IPMR}_{n+k}\left(\mathcal{X}_k^+\right)$ is given as

$$\overline{\mathrm{IPMR}}_{n+k}\left(\mathcal{X}_k^+\right)=\mathbb{E}_{\boldsymbol{X}}\left[\overline{\Gamma}_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)\right]=\mathbb{E}_{\boldsymbol{X}}\left[\Phi\left(\frac{-\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}}{\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)}\right)\right]. \tag{31}$$

Then, according to Eq. (29), there holds

$$0\le\mathrm{IPMR}_{n+k}\left(\mathcal{X}_k^+\right)\le\overline{\mathrm{IPMR}}_{n+k}\left(\mathcal{X}_k^+\right)\le H_n. \tag{32}$$

3.4. Numerical evaluation of k-IPMR

Since $\mathrm{IPMR}_{n+k}\left(\mathcal{X}_k^+\right)$ (Eq. (30)) generally has no analytical solution, its VAIS-based estimation can be expressed as

$$\mathrm{IPMR}_{n+k}\left(\mathcal{X}_k^+\right) \approx \frac{1}{Q_3}\sum_{q=1}^{Q_3}\Gamma_{n+k}\left(\boldsymbol{x}^{(q)};\mathcal{X}_k^+\right)\frac{f_{\boldsymbol{X}}\left(\boldsymbol{x}^{(q)}\right)}{h_{\boldsymbol{X}}\left(\boldsymbol{x}^{(q)}\right)}, \tag{33}$$

where $\mathcal{X}_{\mathrm{Q}_3}=\left\{\boldsymbol{x}^{(q)}, q=1,\ldots,Q_3\right\}$ denotes a set of Q_3 quadrature points drawn from $h_{\boldsymbol{X}}(\boldsymbol{x})$. Intuitively, the sample size Q_3 can be determined according to the sample CV value for each $\mathcal{X}_k^+$, i.e., $\mathrm{CV}\left[\mathrm{IPMR}_{n+1}\left(\mathcal{X}_k^+\right)\right]$. However, when selecting $\mathcal{X}_k^*$ from the candidate pool $\mathcal{X}_{\mathrm{C}}$ according to Eq. (25), the $\mathrm{IPMR}_{n+k}(\cdot)$ has to be evaluated on all possible choices of $\mathcal{X}_k^+$ in $\mathcal{X}_{\mathrm{C}}$, i.e., $\mathrm{IPMR}_{n+k}\left(\mathcal{X}_k^+\right), \forall\, \mathcal{X}_k^+ \in \mathcal{X}_{\mathrm{C}}$. Such a great amount of candidate choices make it infeasible to specify different values of Q_3 for distinct choices of $\mathcal{X}_k^+ \in \mathcal{X}_{\mathrm{C}}$. To this end, a tractable solution is developed as follows.

Recall that in each iteration of BALR (Fig. 1), a total of 3 single integrals need to be numerically computed based on the VAIS, consisting of $\mu_{\widehat{P}_{f,n}}$ in Eq. (10), H_n in Eq. (18), and $\mathrm{IPMR}_{n+k}(\cdot)$ in Eq. (33). Unlike $\mathrm{IPMR}_{n+k}(\cdot)$, the former two only need to be computed once per iteration. Hence, for ease of numerical implementation, the sizes, Q_1, Q_2, Q_3, of their quadrature point sets can be jointly specified as

$$Q_1 = Q_2 = Q_3 = Q, \tag{34}$$

where the Q is determined according to

$$\left\{\mathrm{CV}\left[\tilde{\mu}_{\widehat{P}_{f,n}}\right] \le \varepsilon_{\mathrm{Q}}\right\} \bigcap \left\{\mathrm{CV}\left[\tilde{H}_n\right] \le \varepsilon_{\mathrm{Q}}\right\}. \tag{35}$$

Obviously, Eq. (35) ensures the integral precision of $\tilde{\mu}_{\widehat{P}_{f,n}}$ and $\tilde{H}_n$ readily; then, Eq. (32) shows that the H_n is the upper bound of $\mathrm{IPMR}_{n+k}(\mathcal{X}_k^+)$, and the promising batches are naturally those with greater values of $\mathrm{IPMR}_{n+k}(\mathcal{X}_k^+)$, that is, being closer to H_n. Hence, the quadrature point set $\mathcal{X}_{\mathrm{Q}}$ sufficient to H_n shall also suffice for the computation of $\mathrm{IPMR}_{n+k}(\cdot)$ on those promising candidate batches, justifying the rationality of Eq. (34).

Remark 1. *In the BALR's workflow, the Q value at each iteration can be customized based on Eq. (35) such that $Q = Q_{\mathrm{init}} + Q_{\mathrm{add}} \times n_{\mathrm{add}}$, where Q_{init} denotes the initial size and is taken as the size of the final quadrature point set in the previous iteration; Q_{add} and n_{add} are the number of quadrature points added per step and the number of steps to be needed, respectively. Here, Q_{add} is set as 2×10^5, and n_{add} is determined by sequentially adding Q_{add} quadrature points to $\mathcal{X}_{\mathrm{Q}}^{\mathrm{init}}$ until satisfying Eq. (35). Generally, after several initial iterations, the Q value remains the same in the subsequent iterations, which avoids frequently tuning the quadrature set of BALR to great extent.*

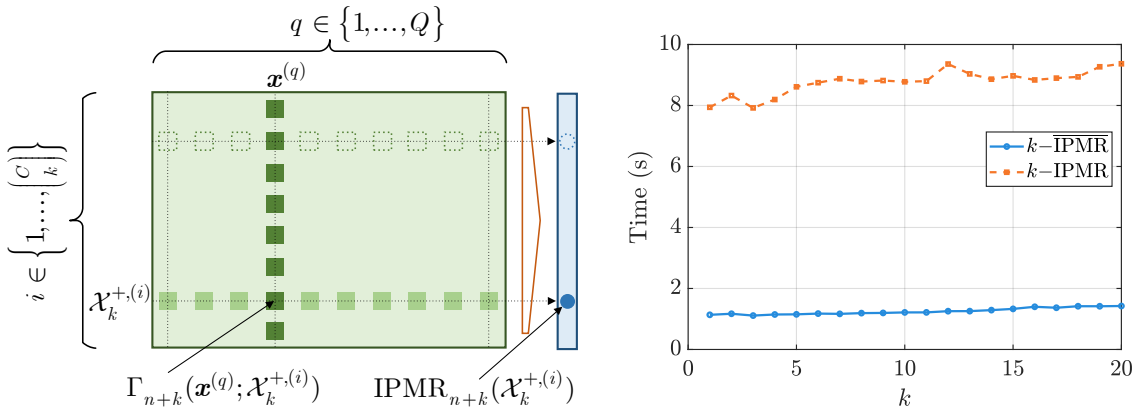
Remark 2. *With three distinct integrals to be numerically computed in the BALR, the numerical integration methods shall be as efficient and robust as possible. Clearly, the optimal importance distributions for the three integrals are different from each other and are totally unknown a priori. Then, the VAIS can be viewed as an advanced varient of MCS. When the amplification coefficient $\alpha = 1$, the VAIS is exactly the crude MCS. When α is too significant, e.g., $\alpha \ge 2$, certain improper samples are prone to be generated. Hence, the rational value of α shall be in [1, 2]. The setting of $\alpha = 1.5$ has been justified in the existing literature [33, 34] and is adopted in this study. More advanced numerical integration methods will be explored in the future study.*

3.5. Computational challenge of the direct k-IPMR-guided multi-point enrichment process

The quadrature point set $\mathcal{X}_{\mathrm{Q}}$ (Eq. (34)) will serve as the candidate pool $\mathcal{X}_{\mathrm{C}}$, i.e., $\mathcal{X}_{\mathrm{C}} = \mathcal{X}_{\mathrm{Q}}$, in each iteration of BALR. Intuitively, the multi-point enrichment can be easily conducted through selecting a batch of k best next samples $\mathcal{X}_k^*$ from $\mathcal{X}_{\mathrm{C}}$ via maximizing $\mathrm{IPMR}_{n+k}(\cdot)$, as per Eq. (25). However, intensive computational burden will be encountered in this crude multi-point enrichment process, owing to the following two issues.

- A single evaluation of $\mathrm{IPMR}_{n+k}(\cdot)$ (Eq. (33)) can be costly due to it entailing a total of Q calls to $\Gamma_{n+k}(\cdot;\mathcal{X}_k^+)$ point by point, as visualized as the boxes in the horizontal direction in Fig. 4(a). Note that the computational burden primarily lies on the evaluation of $\Phi_2(\cdot;\cdot,\cdot)$, whose evaluation is complex and can not be easily vectorized.
- When carrying out the maximization in Eq. (25), a total of $\binom{Q}{k}$ evaluations of $\mathrm{IPMR}_{n+k}(\cdot)$ are involved, as visualized as the boxes in the vertical direction in Fig. 4(a). The many-queries nature (up to $\binom{Q}{k}\times Q$ calls to $\Gamma_{n+k}(\cdot;\mathcal{X}_k^+)$) renders the direct maximization of k-IPMR very time-consuming, even resorting to parallel computing.

To this end, great endeavors will be made in Section 4 to alleviate such heavy computational burden.



(a) Direct maximization of $\mathrm{IPMR}_{n+k}(\cdot)$ (b) $\mathrm{IPMR}_{n+k}(\cdot)$ vs. $\overline{\mathrm{IPMR}}_{n+k}(\cdot)$

Figure 4: Illustration of the intensive computational burden of k-IPMR

4. Efficient k-IPMR-guided multi-point enrichment process

To reduce the overhead of k-IPMR-based multi-point enrichment process, three effective workarounds are sequentially developed here. Section 4.1 elucidates the fast single evaluation of k-IPMR. Section 4.2 details the efficient stepwise maximization of k-IPMR. Section 4.3 discusses the identification of rational size of batch of new samples per iteration. Finally, Section 4.4 summarizes the workflow of the proposed parallel Bayesian active learning reliability method.

4.1. Fast single evaluation of k-IPMR

To accelerate a single evaluation of k-IPMR, the following two steps are conducted. (i) Substitute $\mathrm{IPMR}_{n+k}(\mathcal{X}_k^+)$ by its upper bound $\overline{\mathrm{IPMR}}_{n+k}(\mathcal{X}_k^+)$, avoiding the evaluation of $\Phi_2(\cdot;\cdot,\cdot)$; (ii) Prune $\mathcal{X}_\mathrm{Q}$ in the computation of $\overline{\mathrm{IPMR}}_{n+k}(\mathcal{X}_k^+)$, alleviating the computer memory issue.

First, similar to Eq. (33), the VAIS-based estimation of $\overline{\mathrm{IPMR}}_{n+k}(\mathcal{X}_k^+)$ can be expressed as

$$\overline{\mathrm{IPMR}}_{n+k}\left(\mathcal{X}_k^+\right) \approx \frac{1}{Q}\sum_{q=1}^{Q}\overline{\Gamma}_{n+k}\left(\boldsymbol{x}^{(q)};\mathcal{X}_k^+\right)\frac{f_{\boldsymbol{X}}\left(\boldsymbol{x}^{(q)}\right)}{h_{\boldsymbol{X}}\left(\boldsymbol{x}^{(q)}\right)}. \tag{36}$$

where $\left\{\overline{\Gamma}_{n+k}\left(\boldsymbol{x}^{(q)};\mathcal{X}_k^+\right)\right\}_{q=1}^{Q}$ (Eq. (28)) can be cheaply computed in a vectorized manner, getting rid of the point-wise evaluation of $\Phi_2(\cdot;\cdot,\cdot)$. Obviously, in comparison with $\mathrm{IPMR}_{n+k}(\cdot)$ in Eq. (33), $\overline{\mathrm{IPMR}}_{n+k}(\cdot)$ in Eq. (36) also reflects the impact of adding $\mathcal{X}_k^+$ on the reduction of IPM, but comes with much lower computational cost.

Second, Eq. (36) implies that the relative contributions of $\boldsymbol{x}^{(q)}, q=1,\ldots,Q$, to $\overline{\mathrm{IPMR}}_{n+k}\left(\mathcal{X}_k^+\right)$ are different from each other. Hence, the $\mathcal{X}_\mathrm{Q}$ can be pruned to a subset of $Q_\mathrm{P}(\ll Q)$ quadrature points with greatest values of $\overline{\Gamma}_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)\frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{h_{\boldsymbol{X}}(\boldsymbol{x})}$. However, since $\overline{\Gamma}_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)$ relies on $\mathcal{X}_k^+$, the resulting pruned quadrature point set will vary with $\mathcal{X}_k^+ \in \mathcal{X}_\mathrm{C}$. To specify an identical pruned quadrature point set for any batch $\mathcal{X}_k^+ \in \mathcal{X}_\mathrm{C}$, the pruning criterion shall avoid the presence of $\mathcal{X}_k^+$.

Therefore, the pruning criterion is specified according to $\pi_n(\boldsymbol{x})\frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{h_{\boldsymbol{X}}(\boldsymbol{x})}$, rather than $\overline{\Gamma}_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)\frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{h_{\boldsymbol{X}}(\boldsymbol{x})}$ itself. First, the $\mathcal{X}_\mathrm{Q}$ is sorted in decreasing order of $\pi_n(\boldsymbol{x})\frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{h_{\boldsymbol{X}}(\boldsymbol{x})}$, resulting in $\left\{\boldsymbol{x}^{(q_j)}, j=1,\ldots,Q\right\}$, where q_j denotes the index of the point with the j-th greatest value. Then, only the first Q_P $(\ll Q)$ quadrature points in $\mathcal{X}_\mathrm{Q}$ are collected, that is,

$$\mathcal{X}_{\mathrm{Q_P}} = \left\{\boldsymbol{x}^{(q_j)}, j=1,\ldots,Q_\mathrm{P}\right\}, \tag{37}$$

with the size Q_P specified as

$$Q_\mathrm{P} = \min\left\{\mathcal{Q}\in\mathbb{N}: \frac{\sum_{j=1}^{\mathcal{Q}}\pi_n\left(\boldsymbol{x}^{(q_j)}\right)\frac{f_{\boldsymbol{X}}\left(\boldsymbol{x}^{(q_j)}\right)}{h_{\boldsymbol{X}}\left(\boldsymbol{x}^{(q_j)}\right)}}{\sum_{j=1}^{Q}\pi_n\left(\boldsymbol{x}^{(q_j)}\right)\frac{f_{\boldsymbol{X}}\left(\boldsymbol{x}^{(q_j)}\right)}{h_{\boldsymbol{X}}\left(\boldsymbol{x}^{(q_j)}\right)}} \geq \varepsilon_\mathrm{P}\right\}, \tag{38}$$

where the threshold ε_P is set as 99% here.

In this way, the $\overline{\mathrm{IPMR}}_{n+k}(\mathcal{X}_k^+)$ in Eq. (36) further reduces to

$$\overline{\mathrm{IPMR}}_{n+k}\left(\mathcal{X}_k^+\right) \approx \frac{1}{Q}\sum_{j=1}^{Q_\mathrm{P}}\overline{\Gamma}_{n+k}\left(\boldsymbol{x}^{(q_j)};\mathcal{X}_k^+\right)\frac{f_{\boldsymbol{X}}\left(\boldsymbol{x}^{(q_j)}\right)}{h_{\boldsymbol{X}}\left(\boldsymbol{x}^{(q_j)}\right)}. \tag{39}$$

Then, the $\mathcal{X}_k^*$ can be selected as

$$\mathcal{X}_k^* = \underset{\mathcal{X}_k^+ \in \mathcal{X}_{\mathrm{C}}}{\arg\max}\, \overline{\mathrm{IPMR}}_{n+k}\left(\mathcal{X}_k^+\right). \tag{40}$$

For illustration, Fig. 5 shows the pruning of $\mathcal{X}_{\mathrm{Q}}$ in a single iteration (the turbine blade example in Section 5.3). First, the size, Q, of $\mathcal{X}_{\mathrm{Q}}$ is determined as 4×10^5 in this iteration, as per Remark 1. Then, only a small fraction of quadrature points have significant values of $\pi_n(\cdot)\frac{f_{\boldsymbol{X}}(\cdot)}{h_{\boldsymbol{X}}(\cdot)}$; see the left panel of Fig. 5. According to Eq. (38), the size, Q_{P}, of $\mathcal{X}_{\mathrm{Q_P}}$ is determined as 3473, and there holds $\frac{Q_{\mathrm{P}}}{Q} = 0.87\%$; see the right panel of Fig. 5. Obviously, such small value of Q_{P} makes the single evaluation of $\overline{\mathrm{IPMR}}_{n+k}(\cdot)$ very fast.

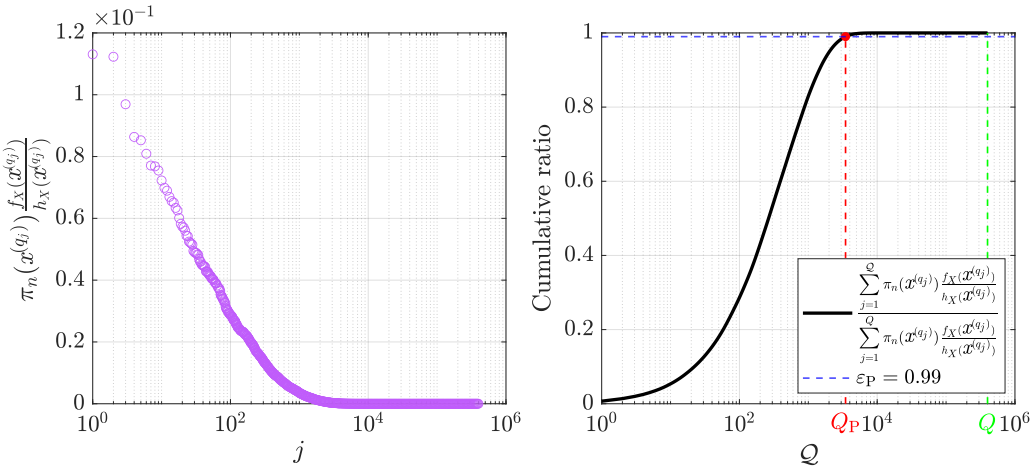


Figure 5: Illustration of the pruning of $\mathcal{X}_{\mathrm{Q}}$

Remark 3. *The rationale of Eq. (39) can be justified from Fig. E.16. Comparison between Figs. E.16(c), (f), and (i) shows that the local region with greater values of* $\overline{\Gamma}_{n+k}(\cdot;\mathcal{X}_k^+)\frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{h_{\boldsymbol{X}}(\boldsymbol{x})}$ *is always enclosed by that with greater values of* $\pi_n(\boldsymbol{x})\frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{h_{\boldsymbol{X}}(\boldsymbol{x})}$*, regardless of promising or unpromising batches. Hence, the* $\mathcal{X}_{\mathrm{Q_P}}$ *shall suffice for the computation of* $\overline{\mathrm{IPMR}}_{n+k}(\cdot)$*, especially for those promising candidate batches.*

4.2. Efficient stepwise maximization of k-IPMR

The maximization problem of dimension $k \times d$ in Eq. (40) is still expensive-to-run, due to vast number of candidate choices of $\mathcal{X}_k^+$. To this end, the following two steps are implemented. (i) Prune the candidate pool $\mathcal{X}_{\mathrm{C}}$, reducing the number of possible candidate batches. (ii) Replace the direct maximization approach with a stepwise maximization treatment, alleviating the remaining computational burden.

First, the characteristic of a promising candidate batch $\mathcal{X}_k^+$ within $\mathcal{X}_{\mathrm{C}}$ is explored in Appendix E. Briefly, the promising batch is prone to be in the local area endowed with great value of $\pi_n(\boldsymbol{x})\frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{h_{\boldsymbol{X}}(\boldsymbol{x})}$; see Fig. E.16. Hence, the entire candidate pool $\mathcal{X}_{\mathrm{C}}$ can be pruned to a subset of C_{P} candidate points with greatest values of $\pi_n(\boldsymbol{x})\frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{h_{\boldsymbol{X}}(\boldsymbol{x})}$, that is,

$$\mathcal{X}_{\mathrm{C_P}} = \left\{\boldsymbol{x}^{(q_j)}, j = 1, \ldots, C_{\mathrm{P}}\right\}, \tag{41}$$

where the size C_{P} is specified as 10^3 here.

Second, given that the number of candidate batches in the direct maximization approach, say $\binom{C_{\mathrm{P}}}{k}$, is still intensive, a stepwise maximization approach is developed here. It starts by maximizing the 1-point $\overline{\mathrm{IPMR}}$ to select the 1-st best next point $\boldsymbol{x}^{(n+1)}$; then, maximize the 2-point $\overline{\mathrm{IPMR}}$ (keeping $\boldsymbol{x}^{(n+1)}$ as the fixed argument) to select the 2-nd best next point $\boldsymbol{x}^{(n+2)}$; this process iterates until a pertinent stopping condition is satisfied.

Specifically, when $k = 1$, the 1-point $\overline{\mathrm{IPMR}}$ is expressed as

$$\overline{\mathrm{IPMR}}_{n+1}\left(\boldsymbol{x}_+\right) \approx \frac{1}{Q}\sum_{j=1}^{Q_{\mathrm{P}}} \overline{\Gamma}_{n+1}\left(\boldsymbol{x}^{(q_j)};\boldsymbol{x}_+\right)\frac{f_{\boldsymbol{X}}\left(\boldsymbol{x}^{(q_j)}\right)}{h_{\boldsymbol{X}}\left(\boldsymbol{x}^{(q_j)}\right)}, \tag{42}$$

where $\overline{\Gamma}_{n+1}\left(\cdot;\boldsymbol{x}_+\right)$ acts as a reminder that it is a function of $\boldsymbol{x}_+$ solely. In this way, the 1-st best next point $\boldsymbol{x}^{(n+1)}$ can be selected as

$$\boldsymbol{x}^{(n+1)} = \underset{\boldsymbol{x}_+ \in \mathcal{X}_{\mathrm{C_P}}}{\arg\max}\, \overline{\mathrm{IPMR}}_{n+1}\left(\boldsymbol{x}_+\right), \tag{43}$$

then, $\mathcal{X}_1^* = \left\{\boldsymbol{x}^{(n+1)}\right\}$, and $\mathcal{X}_{\mathrm{C_P}} = \mathcal{X}_{\mathrm{C_P}} \setminus \left\{\boldsymbol{x}^{(n+1)}\right\}$.

When $k \geq 2$, assume that the former $(k-1)$ best next points $\mathcal{X}_{k-1}^* = \left\{\boldsymbol{x}^{(n+1)}, \ldots, \boldsymbol{x}^{(n+k-1)}\right\}$ have been selected and are taken as the fixed arguments. Then, the k-point $\overline{\text{IPMR}}$ is recast as

$$\overline{\text{IPMR}}_{n+k}\left(\mathcal{X}_{k-1}^*, \boldsymbol{x}_+\right) \approx \frac{1}{Q}\sum_{j=1}^{Q_\text{P}} \overline{\Gamma}_{n+k}\left(\boldsymbol{x}^{(q_j)}; \mathcal{X}_{k-1}^*, \boldsymbol{x}_+\right) \frac{f_{\boldsymbol{X}}\left(\boldsymbol{x}^{(q_j)}\right)}{h_{\boldsymbol{X}}\left(\boldsymbol{x}^{(q_j)}\right)}, \tag{44}$$

where $\overline{\Gamma}_{n+k}(\cdot; \mathcal{X}_{k-1}^*, \boldsymbol{x}_+)$ acts as a reminder that it is also a function of $\boldsymbol{x}_+$ solely. In this way, the k-th best next point $\boldsymbol{x}^{(n+k)}$ is selected as

$$\boldsymbol{x}^{(n+k)} = \underset{\boldsymbol{x}_+ \in \mathcal{X}_{\text{C}_\text{P}}}{\arg\max}\, \overline{\text{IPMR}}_{n+k}\left(\mathcal{X}_{k-1}^*, \boldsymbol{x}_+\right), \tag{45}$$

then, $\mathcal{X}_k^* = \mathcal{X}_{k-1}^* \bigcup \left\{\boldsymbol{x}^{(n+k)}\right\}$, and $\mathcal{X}_{\text{C}_\text{P}} = \mathcal{X}_{\text{C}_\text{P}} \setminus \left\{\boldsymbol{x}^{(n+k)}\right\}$.

Obviously, the original maximization problem of dimension $k \times d$ in Eq. (40) is transformed to k consecutive maximization problems of dimension d in Eqs. (43) and (45). By doing so, the number of candidate choices reduces from $\binom{C_\text{P}}{k}$ to $C_\text{P} \times k$, which is much smaller to handle.

4.3. Adaptive new sample batch size identification per iteration

Another key issue in the multi-point enrichment process is how to specify the rational size, K, of batch of new points added per iteration. Traditionally, the K is set as a fixed value *a priori*; then, sequentially increasing k until K will give rise to $\mathcal{X}_K^* = \left\{\boldsymbol{x}^{(n+k)}, k = 1, \ldots, K\right\}$ readily. This is the so-called *prescribed scheme*.

By contrast, an *adaptive scheme* that specifies automatically the K value in each iteration is developed here. Specifically, according to Eqs. (43) and (45), the individual gain G_{n+k} of adding $\boldsymbol{x}^{(n+k)}$ in terms of reducing the IPM can be quantified as

$$G_{n+k} = \begin{cases} \overline{\text{IPMR}}_{n+k}\left(\mathcal{X}_k^*\right) - \overline{\text{IPMR}}_{n+k-1}\left(\mathcal{X}_{k-1}^*\right), & k \geq 2, \\ \overline{\text{IPMR}}_{n+1}\left(\mathcal{X}_1^*\right), & k = 1, \end{cases} \tag{46}$$

which generally decreases with the increase of k. When the G_{n+k} itself or the ratio $\frac{G_{n+k}}{G_{n+1}}$ becomes very small, it implies that adding $\boldsymbol{x}^{(n+k)}$ is not useful to further reduce the IPM, and the sequential increase of k can be stopped. Hence, the rational value of K in each iteration can be taken as

$$K = \min\left\{K_1, K_2, n_\text{core}\right\}, \tag{47}$$

where $K_1 = \min\left\{k \in \mathbb{N} : G_{n+k} \leq 0\right\}$; $K_2 = \min\left\{k \in \mathbb{N} : \frac{G_{n+k}}{G_{n+1}} \leq \varepsilon_\text{G}\right\}$; n_core is the number of available CPU cores; the threshold ε_G is set as 0.1.

Remark 4. *The setup of the adaptive scheme totally relies on the peculiar ability of k-$\overline{\text{IPMR}}$ to quantify the expected reduction of IPM brought by each new sample in the batch. Obviously, this adaptivity is unlikely to be achieved by the traditional practice of combining a pointwise learning function with some additional multi-point selection strategies, due to their inherent flaws.*

Algorithm 1 summarizes the fast k-$\overline{\text{IPMR}}$-based multi-point enrichment process in a single iteration. Notably, computing $\overline{\text{IPMR}}_{n+k}\left(\mathcal{X}_{k-1}^*, \boldsymbol{x}_+\right), \forall \boldsymbol{x}_+ \in \mathcal{X}_{\text{C}_\text{P}}$ (the 'for-loop' in Lines 7-11) can be conducted in parallel to accelerate this process. Clearly, the k-$\overline{\text{IPMR}}$ can be readily substituted by the cumbersome k-IPMR in Algorithm 1. Note that the running time of Algorithm 1 is mainly consumed by the $K \times C_\text{P} \times Q_\text{P}$ times of evaluation of $\bar{\Gamma}_{n+k}(\cdot;\cdot)$ or $\Gamma_{n+k}(\cdot;\cdot)$. Recall that C_P is fixed as 10^3 and K is pre-assigned in the *prescribed* scheme, the running time is actually dominated by Q_P, thereby it varying with iterations in the workflow of BALR.

Fig. 4(b) compares the running time between k-IPMR and k-$\overline{\text{IPMR}}$ in a single iteration, with the prescribed batch size $K = 20$ and the $Q_\text{P} = 8970$ (the truss example in Section 5.2). It is observed that k-$\overline{\text{IPMR}}$ consumes approximately $30 = 1.5 \times 20$ seconds, which is much smaller than that of k-IPMR, say $180 = 9 \times 20$ seconds. As a result, k-$\overline{\text{IPMR}}$ comes with much less running time than k-IPMR.

Remark 5. *Algorithm 1 is conducted based on k-$\overline{\text{IPMR}}$ itself, eliminating the need for additional parallel enrichment strategies. Moreover, the resulting batch $\mathcal{X}_K^*$ is optimal for reducing the IPM. By comparison, most of the existing parallel (B)ALR methods do not explicitly measure the impact of adding a batch of new samples, let alone ensure the optimality of the resulting batch of new samples. Hence, the proposed k-$\overline{\text{IPMR}}$ is more theoretically sound than those existing methods.*

Remark 6. *An existing learning function called SUR is outlined in Appendix F, consisting of two different forms. They are expressed either as a double integral of $\Phi_2(\cdot;\cdot,\cdot)$ or as a single integral of $\Phi_2(\cdot;\cdot,\cdot)$, both of which are expensive-to-evaluate. By comparison, the proposed k-$\overline{\text{IPMR}}$ is conceptually similar but is only expressed as a single integral of $\Phi(\cdot)$, coming with a much lower computational burden. Comparison between them will be made in Section 5.*

Algorithm 1 The fast k-$\overline{\text{IPMR}}$-guided multi-point enrichment process in a single iteration

Input: The Kriging $\widehat{\mathcal{G}}_n(\boldsymbol{x})$, the K value in the *prescribed scheme*, or the ε_{G} value in the *adaptive scheme*.

Compute $\tilde{\mu}_{\widehat{P}_{f,n}}$ and $\tilde{H}_n$, together with the values of $\text{CV}\big[\tilde{\mu}_{\widehat{P}_{f,n}}\big]$ and $\text{CV}\big[\tilde{H}_n\big]$. ▷ Eqs. (10), (18)
Specify the size, Q, of $\mathcal{X}_{\text{Q}}$ according to $\text{CV}\big[\tilde{\mu}_{\widehat{P}_{f,n}}\big]$ and $\text{CV}\big[\tilde{H}_n\big]$; then, let $\mathcal{X}_{\text{C}} = \mathcal{X}_{\text{Q}}$. ▷ Eq. (35)
Obtain a pruned quadrature point set $\mathcal{X}_{\text{Q}_\text{P}} = \big\{\boldsymbol{x}^{(q_j)}, j = 1, \ldots, Q_\text{P}\big\}$ from $\mathcal{X}_{\text{Q}}$. ▷ Eqs. (37), (38)
Obtain a pruned candidate pool $\mathcal{X}_{\text{C}_\text{P}} = \Big\{\boldsymbol{x}_+^{(i)}, i = 1, \ldots, C_\text{P}\Big\}$ from $\mathcal{X}_{\text{C}}$. ▷ Eq. (41)
Let $k \leftarrow 1$, and $\mathcal{X}_0^* = \{\}$.
while true **do**
  for $i = 1, \ldots, C_\text{P}$ **do**
    Compute $\Big\{\rho_{n+k}\big(\boldsymbol{x}^{(q_j)}; \mathcal{X}_{k-1}^*, \boldsymbol{x}_+^{(i)}\big)\Big\}_{j=1}^{Q_\text{P}}$ at $\mathcal{X}_{\text{Q}_\text{P}}$. ▷ Eq. (27)
    Compute $\Big\{\overline{\Gamma}_{n+k}\big(\boldsymbol{x}^{(q_j)}; \mathcal{X}_{k-1}^*, \boldsymbol{x}_+^{(i)}\big)\Big\}_{j=1}^{Q_\text{P}}$ at $\mathcal{X}_{\text{Q}_\text{P}}$. ▷ Eq. (28)
    Compute $\overline{\text{IPMR}}_{n+k}\big(\mathcal{X}_{k-1}^*, \boldsymbol{x}_+^{(i)}\big)$. ▷ Eqs. (42), (44)
  end for
  Select the k-th best next point $\boldsymbol{x}^{(n+k)}$ from $\mathcal{X}_{\text{C}_\text{P}}$ according to $\overline{\text{IPMR}}_{n+k}(\cdot)$. ▷ Eqs. (43), (45)
  Identify the rational value of K in the *adaptive scheme*. ▷ Eqs. (46), (47)
  if $k \geq K$ **then**
    Break.
  else
    $\mathcal{X}_k^* = \mathcal{X}_{k-1}^* \bigcup \big\{\boldsymbol{x}^{(n+k)}\big\}$, $\mathcal{X}_{\text{C}_\text{P}} = \mathcal{X}_{\text{C}_\text{P}} \setminus \big\{\boldsymbol{x}^{(n+k)}\big\}$, and $k \leftarrow k + 1$.
  end if
end while

Output: A batch of K best next samples $\mathcal{X}_K^* = \big\{\boldsymbol{x}^{(n+k)}, k = 1, \ldots, K\big\}$ selected in this iteration.

4.4. Workflow of efficient k-IPMR-guided parallel Bayesian active learning reliability analysis

Fig. 6 presents the flowchart of the proposed parallel Bayesian active learning reliability method. The main steps are summarized as follows.

Step 1: **Initial experimental design**.
To obtain an initial Kriging model with fair accuracy, the initial ED prefers to be as space-filling as possible. According to the 'four-sigma' rule, the sampling domain $\mathbb{X}_\text{s}$ is defined by

$$\mathbb{X}_\text{s} = \prod_{l=1}^{d} [-4, 4], \tag{48}$$

and generate a set of n_0 uniform samples in the $\mathbb{X}_\text{s}$, denoted as $\mathcal{X}_{n_0} = \big\{\boldsymbol{x}^{(i)}, i = 1, \ldots, n_0\big\}$ with $n_0 = \max(d + 1, 10)$, using the Latinized centroidal Voronoi tessellation sampling method [53]. Then, evaluate the $\mathcal{G}$-function on $\mathcal{X}_{n_0}$ to obtain their responses $\mathcal{Y}_{n_0} = \big\{y^{(i)}, i = 1, \ldots, n_0\big\}$. In this way, the initial ED is collected as $\mathcal{D}_{n_0} = \{\mathcal{X}_{n_0}, \mathcal{Y}_{n_0}\}$. Finally, let $n = n_0$.

Step 2: **Kriging calibration**.
Train a Kriging $\widehat{\mathcal{G}}_n(\boldsymbol{x})$ based on the current ED $\mathcal{D}_n$, offering the posterior mean $\mu_n(\boldsymbol{x})$, variance $\sigma_n(\boldsymbol{x})$, and covariance $c_n(\boldsymbol{x}, \boldsymbol{x}')$; see Eqs. (A.7), (A.8) and (A.9).

Step 3: **Failure probability estimation**.
According to the VAIS, compute both the mean failure probability $\tilde{\mu}_{\widehat{P}_{f,n}}$ (Eq. (10)) and the IPM $\tilde{H}_n$ (Eq. (18)) based on a quadrature point set $\mathcal{X}_{\text{Q}} = \big\{\boldsymbol{x}^{(q)}, q = 1, \ldots, Q\big\}$, with the size Q determined as per Remark 1. Besides, the $\mathcal{X}_{\text{Q}}$ will be taken as the candidate pool $\mathcal{X}_{\text{C}}$ in Step 5.

Step 4: **Hybrid convergence criterion**.
A hybrid convergence criterion that combines two different ones is developed here. First, recall that the IPM H_n acts as a global measure of residual uncertainty of $\tilde{\mu}_{\widehat{P}_{f,n}}$, the $\frac{\tilde{H}_n}{\tilde{\mu}_{\widehat{P}_{f,n}}}$ is a natural metric to check the convergence of BALR, expressed as

$$\frac{\tilde{H}_n}{\tilde{\mu}_{\widehat{P}_{f,n}}} \leq \varepsilon_\text{H}, \tag{49}$$

where the setting of ε_H shall consider the distinction between static and dynamic reliability problems. Hence, $\varepsilon_\text{H} = 0.2 \times \max\limits_{i \leq n} \frac{\tilde{H}_i}{\tilde{\mu}_{\widehat{P}_{f,i}}}$ (resp. $0.6 \times \max\limits_{i \leq n} \frac{\tilde{H}_i}{\tilde{\mu}_{\widehat{P}_{f,i}}}$) in static (resp. dynamic) problem.

Second, the stabilization of $\tilde{\mu}_{\widehat{P}_f}$ between consecutive iterations is also a common metric to check the convergence of BALR, expressed as

$$\Delta\tilde{\mu}_{\widehat{P}_{f,n}} = \frac{\left|\tilde{\mu}_{\widehat{P}_{f,n}} - \tilde{\mu}_{\widehat{P}_{f,n-1}}\right|}{\tilde{\mu}_{\widehat{P}_{f,n}}} \leq \varepsilon_{\mathrm{S}}, \tag{50}$$

where the tolerance ε_{S} is set as 0.1 here.

Finally, if Eq. (49) and (50) are satisfied simultaneously, skip to Step 7; otherwise, continue to Step 5.

Step 5: **Multi-point learning function**.
According to the learning function k-$\overline{\mathrm{IPMR}}$, select a batch of K best next points $\mathcal{X}_K^* = \left\{\boldsymbol{x}^{(n+k)}, k = 1, \ldots, K\right\}$ from the candidate pool $\mathcal{X}_{\mathrm{C}}$, with the batch size K identified via either the prescribed scheme or the adaptive scheme, as elucidated in Algorithm 1.

Step 6: **Enrichment and update**.
Evaluate the $\mathcal{G}$-function on $\mathcal{X}_K^*$ in parallel, resulting in $\mathcal{Y}_K^* = \left\{y^{(n+k)}, k = 1, \ldots, K\right\}$; then, conduct the following updates: $\mathcal{D}_{n+K} = \mathcal{D}_n \bigcup \{\mathcal{X}_K^*, \mathcal{Y}_K^*\}$, $n \leftarrow n + K$, $n_{\mathrm{iter}} \leftarrow n_{\mathrm{iter}} + 1$, and return to Step 2.

Step 7: **Termination**.
The final results of three performance metrics are recorded at the end of this algorithm, i.e., the failure probability estimate $\tilde{\mu}_{\widehat{P}_{f,n}}$, the total number of iterations n_{iter}, and the total number of performance function evaluations n_{eval}.

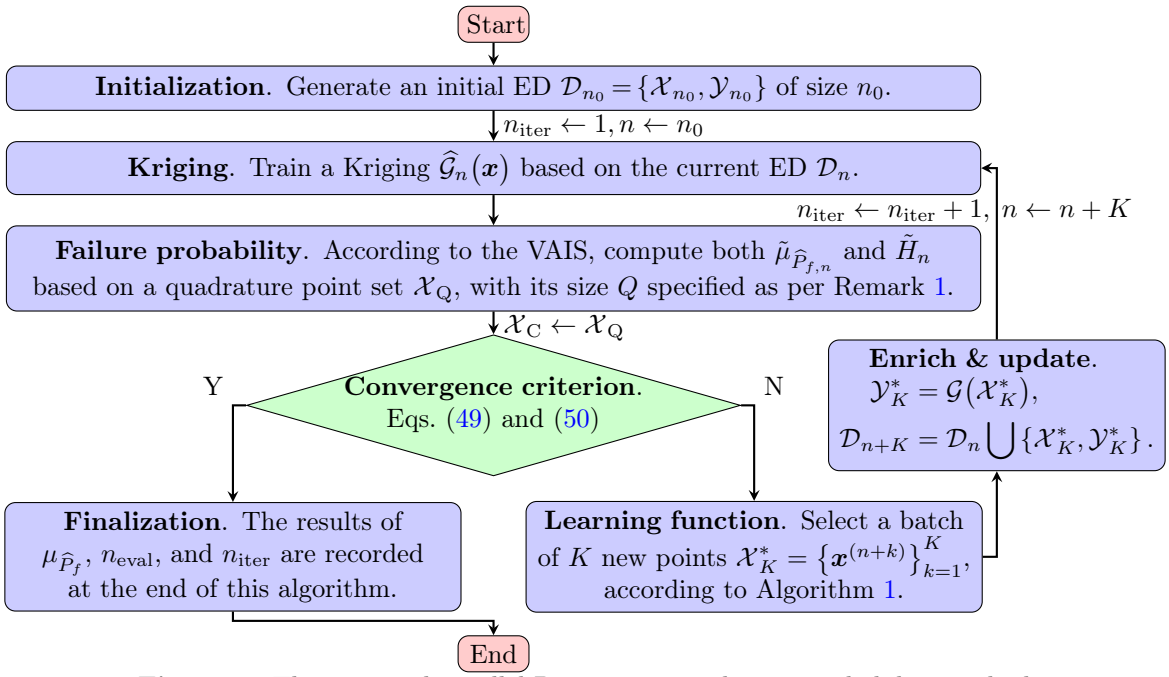


Figure 6: The proposed parallel Bayesian active learning reliability method

Remark 7. *When the prescribed scheme is adopted in Fig. 6, the relationship between* n_{call} *and* n_{iter} *is easily expressed as* $n_{\mathrm{call}} = n_0 + K \times (n_{\mathrm{iter}} - 1)$. *By contrast, since the size* K *varies with iterations in the adaptive scheme, there is no analytical formulation between* n_{call} *and* n_{iter}, *but they can be directly recorded in Fig. 6.*

5. Illustrative examples

The proposed method is testified on four examples of varying complexity. Both the prescribed scheme (K = 5, 10, 15, 20) and the adaptive scheme ($n_{\mathrm{core}} = 20$, $\varepsilon_{\mathrm{G}} = 0.1$) are considered. Besides, the two learning functions k-$\overline{\mathrm{IPMR}}$ and k-IPMR are conducted for comparison.

The MCS is conducted to provide an estimate of failure probability $\widehat{P}_f^{\mathrm{MCS}}$ for reference. Some other existing parallel (B)ALR methods are also conducted for comparison, including the ALR module in UQLab [13], the parallel adaptive Bayesian quadrature (PABQ) [30], as well as the second form of SUR. The ingredients of the ALR module in UQLab are set in default [13]: the Kriging, the subset simulation, the U function, the stopping condition based on the confidence bound of reliability index, and the K-means clustering strategy. The PABQ comprises the following ingredients [30]: the Kriging, the Bayesian inference of failure probability, the pointwise learning function UBVC, and the K-means clustering strategy. Since the numerical computation, candidate

pool, and convergence criterion of SUR were not detailed in [32], the settings of quadrature point set, the pruning rule, and convergence criterion in Section 4 are customized to the SUR here. Contrary to the proposed IPMR, the three aforementioned reliability methods only support the prescribed scheme. Besides, the results from some other reliability methods will be listed in the tables, if available in the literature.

All those reliability methods are run on an Intel i9-14900KF CPU processor equipped with 3.2 GHz, 64 GB RAM, and 20 physical cores. Then, three performance metrics, i.e., $\mu_{\widehat{P}_f}$, n_{eval}, and n_{iter}, are recorded for each method. Further, the relative error of $\mu_{\widehat{P}_f}$ with respect to $\widehat{P}_f^{\text{MCS}}$ is computed as

$$\delta_{\widehat{P}_f} = \frac{\left|\mu_{\widehat{P}_f} - \widehat{P}_f^{\text{MCS}}\right|}{\widehat{P}_f^{\text{MCS}}} \times 100\%. \tag{51}$$

Those (B)ALR methods are repeated 10 times to account for the randomness from both the initial ED and/or sampling. Due to computational cost consideration, the maximum value of n_{eval} is set as 300. Then, both the average values of the four performance metrics, i.e., $\mathbb{E}[\mu_{\widehat{P}_f}]$, $\mathbb{E}\big[n_{\text{eval}}\big]$, $\mathbb{E}\big[n_{\text{iter}}\big]$, and $\mathbb{E}[\delta_{\widehat{P}_f}]$, and the CV of $\mu_{\widehat{P}_f}$, i.e., $\text{CV}\big[\mu_{\widehat{P}_f}\big]$, are calculated. Moreover, the average value of the total computational time t_c is provided in the last three engineering examples for comparison.

5.1. A four-branch function

The first example addresses the well-known four-branch function, which is a typical benchmark in the reliability literature [15, 28, 12]. The performance function is expressed as [15]

$$\mathcal{G}(\boldsymbol{X}) = \min \left\{ \begin{array}{c} a + 0.1(X_1 - X_2)^2 - \dfrac{X_1 + X_2}{\sqrt{2}} \\ a + 0.1(X_1 - X_2)^2 + \dfrac{X_1 + X_2}{\sqrt{2}} \\ (X_1 - X_2) + \dfrac{b}{\sqrt{2}} \\ (X_2 - X_1) + \dfrac{b}{\sqrt{2}} \end{array} \right\}, \tag{52}$$

where X_1 and X_2 are two independent standard Gaussian variables; the two constants a and b are set as 3 and 6, respectively.

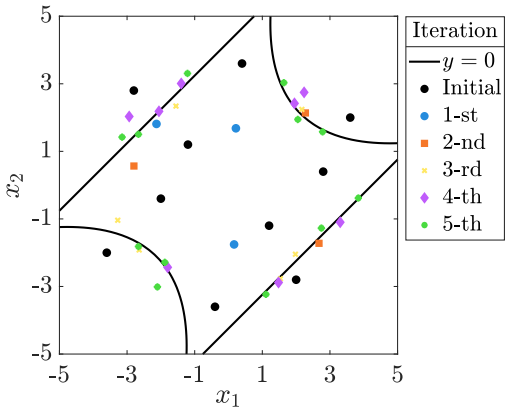
Fig. 7 illustrates a single run of the proposed k-$\overline{\text{IPMR}}$ (adaptive scheme) in the four-branch function. Except for the first iteration, the remaining batches of new samples added by the k-$\overline{\text{IPMR}}$ are mostly located in the vicinity of the actual limit state, as plotted as the colored markers in Fig. 7(a). Then, informed by the k-$\overline{\text{IPMR}}$, the expected gain G_{n+k} brought by the k-th new sample in each iteration is shown in Fig. 7(b). Basically, the G_{n+k} reduces with the increasing of k at each iteration, justifying the rationality of identifying the batch size K via the adaptive scheme. Fig. 7(c) presents the resulting size, K, of batch of new samples identified in each iteration, which gradually increases during the active learning process. Finally, Figs. 7(d) and (e) show the convergence performance of this algorithm. The $\frac{H_n}{\mu_{\widehat{P}_{f,n}}}$ reduces dramatically with the sequential addition of batches of new samples; then, the $\mu_{\widehat{P}_{f,n}}$ converges gradually to the reference value $\widehat{P}_f^{\text{MCS}}$.

Table 1 lists the results of different reliability methods for the four-branch function. In the existing parallel (B)ALR methods, the n_{iter} generally reduces with the increasing of K. The proposed k-$\overline{\text{IPMR}}$ requires fewer iterations, while providing smaller values of the relative error $\delta_{\widehat{P}_f}$, i.e., better accuracy of failure probability estimate. Further, in the prescribed scheme of k-$\overline{\text{IPMR}}$, the n_{iter} decreases slowly, or even rises, with the increasing of K. This implies that it is unwise to blindly increase the batch size K in the prescribed scheme. By comparison, the adaptive scheme identifies automatically the rational value of K in each iteration, avoiding the excessive increase of n_{eval} effectively.

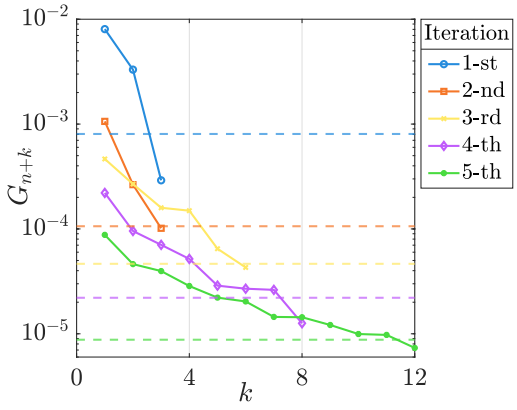
5.2. A 23-bar planar truss under vertical loads

The second example tackles with a 23-bar planar truss subjected to vertical concentrated loads, as geometrically illustrated in Fig. 8. It is also a typical benchmark in the existing reliability literature [28, 12].

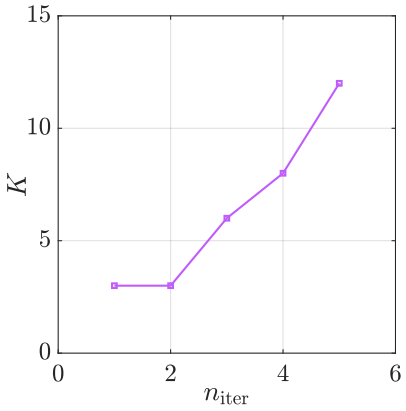
The 23 bars are categorized into two groups: the first group consists of 11 horizontal bars, with Young's modulus E_1 and cross-sectional area A_1; the second group comprises 12 diagonal bars, with Young's modulus E_2 and cross-sectional area A_2. Then, the vertical loads applied on the upper nodes from left to right are denoted as $P_1, \ldots, P_6$. Hence, a total of 10 independent random variables are considered in this truss example, i.e., $\{E_1, E_2, A_1, A_2, P_1, \ldots, P_6\}$. Table 2 lists the statistical information of those random variables.



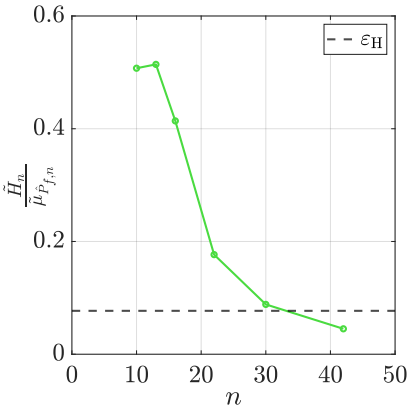
(a) batch of new samples added per iteration



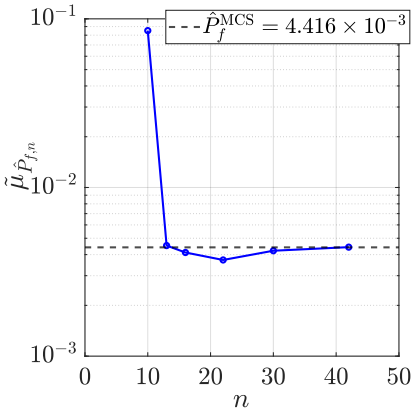
(b) expected gain G_{n+k} of $\boldsymbol{x}^{(n+k)}$ per iteration



(c) batch size K per iteration



(d) convergence of $\frac{H_n}{\mu_{\widehat{P}_{f,n}}}$



(e) convergence of $\mu_{\widehat{P}_{f,n}}$

Figure 7: The k-$\overline{\text{IPMR}}$ (adaptive scheme) in the four-branch function

Table 1: Reliability results in the four-branch function

Method		$\mathbb{E}[n_{\text{iter}}]$	$\mathbb{E}[n_{\text{eval}}]$	$\mathbb{E}[\mu_{\widehat{P}_f}]$	$\text{CV}[\mu_{\widehat{P}_f}]$	$\mathbb{E}[\delta_{\widehat{P}_f}]$	Reference
MCS		-	10^6	4.416×10^{-3}	-	-	[15]
AK-MCS	$K = 1$	62.2	73.2	4.457×10^{-3}	1.50 %	-	[28]
	$K = 6$	15.4	98.4	4.458×10^{-3}	1.50 %	-	[28]
PABQ	$K = 6$	6.6	43.6	4.440×10^{-3}	2.53 %	-	[30]
	$K = 10$	5.2	52	4.400×10^{-3}	2.22 %	-	[30]
	$K = 15$	4.6	64.7	4.440×10^{-3}	1.35 %	-	[30]
RBIK	$K = 6$	17.7	110	4.429×10^{-3}	0.07 %	-	[43]
	$K = 8$	13.7	111.9	4.428×10^{-3}	0.07 %	-	[43]
	$K = 10$	12.1	121.4	4.429×10^{-3}	0.07 %	-	[43]
P-AK-MCS	$K = 4$	15.6	70.4	4.490×10^{-3}	-	-	[48]
	$K = 8$	8.8	74.4	4.560×10^{-3}	-	-	[48]
AK-KB	$K = 3$	22.5	74.6	4.419×10^{-3}	-	-	[54]
	$K = 6$	11.5	73.1	4.411×10^{-3}	-	-	[54]
ALR in UQLab	$K = 5$	20.3	106.5	4.571×10^{-3}	1.42 %	3.51 %	-
	$K = 10$	10.2	102	4.581×10^{-3}	2.22 %	3.74 %	-
	$K = 15$	8	115	4.541×10^{-3}	2.49 %	2.99 %	-
	$K = 20$	6.9	128	4.539×10^{-3}	1.78 %	2.79 %	-
k-$\overline{\text{IPMR}}$	$K = 5$	6.9	39.5	4.401×10^{-3}	1.20 %	1.36 %	-
	$K = 10$	5.5	55	4.406×10^{-3}	1.33 %	1.12 %	-
	$K = 15$	4.7	64	4.415×10^{-3}	0.92 %	0.67 %	-
	$K = 20$	5	90	4.410×10^{-3}	0.83 %	0.63 %	-
	Adaptive	6.3	44	4.405×10^{-3}	0.99 %	0.87 %	-

The mid-span deflection of this planar truss is of concern, expressed as [55]

$$u(\boldsymbol{X}) = \frac{2\sqrt{2}P_1 + 6\sqrt{2}P_2 + 10\sqrt{2}P_3 + 10\sqrt{2}P_4 + 6\sqrt{2}P_5 + 2\sqrt{2}P_6}{16\,E_2A_2} + \frac{36P_1 + 100P_2 + 140P_3 + 140P_4 + 100P_5 + 36P_6}{E_1A_1}, \tag{53}$$

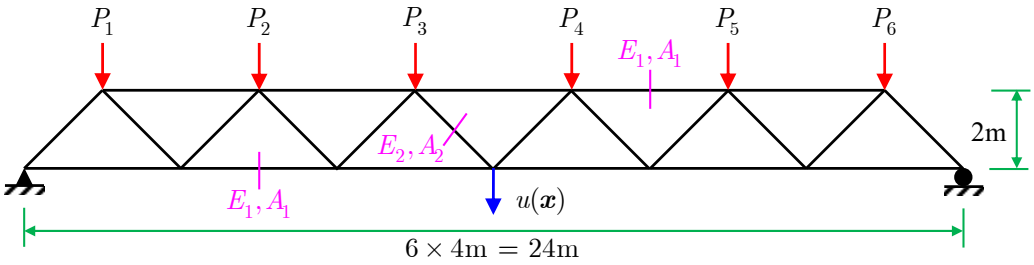


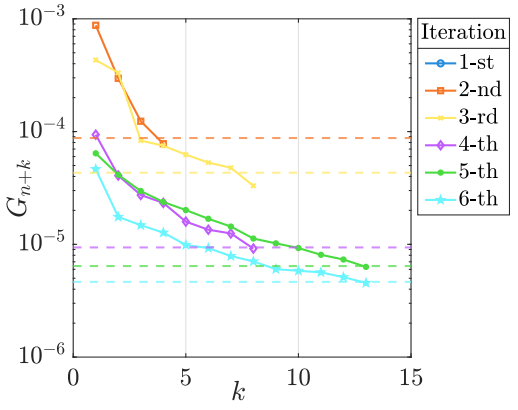
Figure 8: A 23-bar planar truss under vertical loads

Table 2: Random variables in the 23-bar planar truss [28]

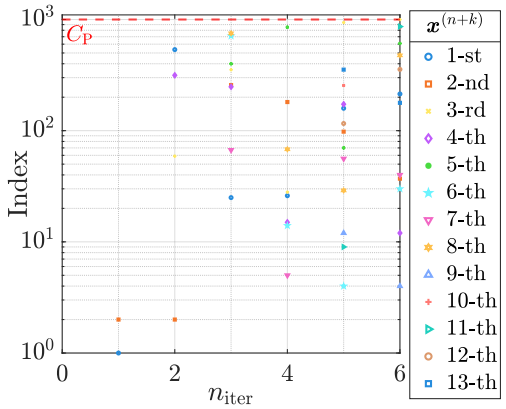
Variable	Unit	Distribution	Mean	Standard deviation
$E_1 - E_2$	Pa	Lognormal	2.1×10^{11}	2.1×10^{10}
A_1	m^2	Lognormal	2.0×10^{-3}	2.0×10^{-4}
A_2	m^2	Lognormal	1.0×10^{-3}	1.0×10^{-4}
$P_1 - P_6$	N	Gumbel	5×10^4	7.5×10^3

and the maximum allowable deflection is set as 12 cm. In this way, the performance function $\mathcal{G}(\boldsymbol{X})$ is given by

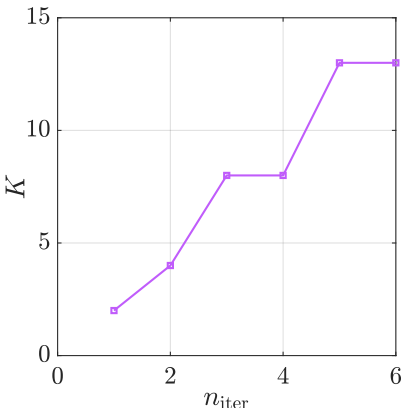
$$\mathcal{G}(\boldsymbol{X}) = 12 - u(\boldsymbol{X}) \tag{54}$$



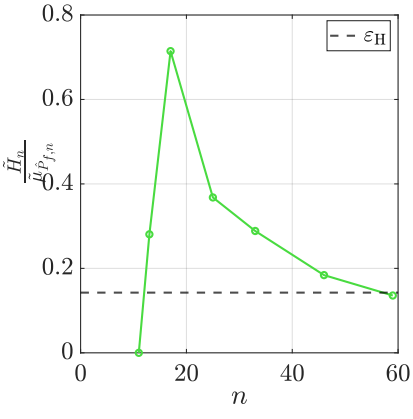
(a) expected gain G_{n+k} of $\boldsymbol{x}^{(n+k)}$ per iteration



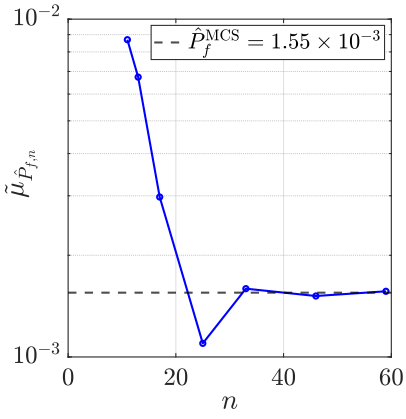
(b) selection index of $\boldsymbol{x}^{(n+k)}$ per iteration



(c) batch size K per iteration



(d) convergence of $\frac{H_n}{\mu_{\hat{P}_{f,n}}}$



(e) convergence of $\mu_{\hat{P}_{f,n}}$

Figure 9: The k-$\overline{\text{IPMR}}$ (adaptive scheme) in the 23-bar planar truss example

Fig. 9 presents the performance of one run of the proposed k-$\overline{\text{IPMR}}$ (adaptive scheme) for the 23-bar planar truss example. A total of 6 iterations are needed in the whole multi-point enrichment process, as shown in Fig. 9(a). Meanwhile, the resulting size, K, of batch of new samples added per iteration increases gradually during the multi-point enrichment process, see Fig. 9(c). Notably, Fig. 9(b) shows that most of the selection indexes of $\boldsymbol{x}^{(n+k)}$ in each iteration are far less than C_{P}, justifying the rationale of the pruning of candidate pool

in Eq. (41). Finally, the IPM H_n is reduced remarkably, and the mean failure probability $\mu_{\widehat{P}_f}$ achieves a good agreement with $\widehat{P}_f^{\mathrm{MCS}}$; see Figs. 9(d) and (e).

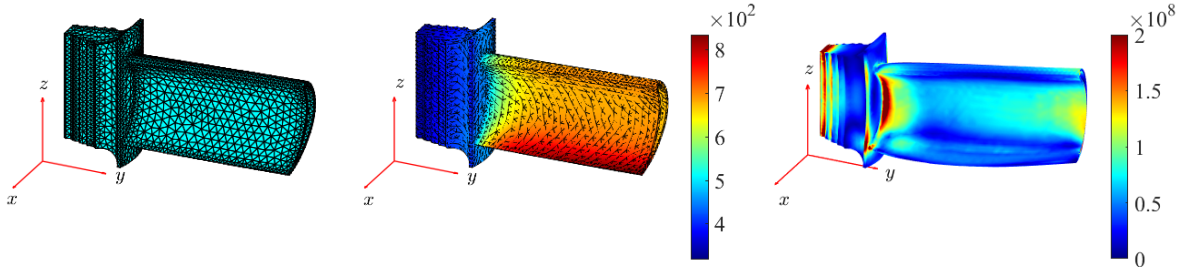
Table 3 lists the results of different reliability methods for the 23-bar planar truss example, where the total computational time t_c is also provided for comparison. All the parallel (B)ALR methods provide relatively accurate estimates of failure probability. Then, both the SUR and the proposed IPMR (k-$\overline{\mathrm{IPMR}}$ and k-IPMR) need fewer iterations than other existing methods. However, due to repeated evaluations of the bivariate Gaussian CDF $\Phi_2(\cdot)$ in both SUR and k-IPMR (Eqs. (33) and (F.4)), they entail much more computational time. By comparison, the k-$\overline{\mathrm{IPMR}}$ avoids the time-intensive evaluation of $\Phi_2(\cdot)$ and only consumes approximately 30% - 50% of computational time of k-IPMR. Admittedly, since the running time of k-$\overline{\mathrm{IPMR}}$ is generally greater than that of traditional practice (e.g., the learning function U + the K-means clustering approach), the advantage of k-$\overline{\mathrm{IPMR}}$ in terms of n_{iter} is not sufficiently transformed to that of t_c.

Table 3: Reliability results in the 23-bar planar truss example

Method		$\mathbb{E}[n_{\mathrm{iter}}]$	$\mathbb{E}[n_{\mathrm{call}}]$	$\mathbb{E}[\mu_{\widehat{P}_f}]$	$\mathrm{CV}[\mu_{\widehat{P}_f}]$	$\mathbb{E}[\delta_{\widehat{P}_f}]$	$\mathbb{E}[t_c]$(s)	References
MCS	-	-	10^6	1.550×10^{-3}	-	-	-	[28]
AK-MCS	$K = 1$	272	283	1.520×10^{-3}	-	1.94 %	-	[28]
	$K = 6$	26	162	1.530×10^{-3}	-	1.29 %	-	[28]
ALR in UQLab	$K = 5$	14.6	79	1.558×10^{-3}	1.87 %	1.51 %	148.3	-
	$K = 10$	7.4	75	1.574×10^{-3}	2.46 %	1.99 %	29.2	-
	$K = 15$	6.3	90.5	1.568×10^{-3}	2.12 %	1.82 %	31.8	-
	$K = 20$	4.8	87	1.589×10^{-3}	3.04 %	2.77 %	29.2	-
PABQ	$K = 5$	14.9	79.5	1.360×10^{-3}	5.55 %	12.24 %	230.1	-
	$K = 10$	8.2	82	1.377×10^{-3}	3.30 %	11.17 %	119.6	-
	$K = 15$	6	85	1.427×10^{-3}	4.49 %	8.31 %	89.1	-
	$K = 20$	5.1	92	1.428×10^{-3}	6.36 %	9.34 %	79.1	-
SUR	$K = 5$	9.5	53.5	1.548×10^{-3}	1.17 %	0.81 %	455.2	-
	$K = 10$	6.2	63	1.551×10^{-3}	0.93 %	0.62 %	388.1	-
	$K = 15$	5	71	1.546×10^{-3}	0.99 %	0.72 %	418.6	-
	$K = 20$	4.1	73	1.545×10^{-3}	0.90 %	0.64 %	447.9	-
k-IPMR	$K = 5$	9.3	52.5	1.555×10^{-3}	1.55 %	1.26 %	445.3	-
	$K = 10$	6.2	63	1.552×10^{-3}	1.59 %	1.35 %	366.4	-
	$K = 15$	5	71	1.550×10^{-3}	0.71 %	0.58 %	374.3	-
	$K = 20$	4.2	75	1.554×10^{-3}	1.26 %	1.15 %	413.1	-
	Adaptive	5.4	63.5	1.553×10^{-3}	1.65 %	0.76 %	404.3	-
k-$\overline{\mathrm{IPMR}}$	$K = 5$	9.5	53.5	1.549×10^{-3}	1.40 %	0.86 %	178.6	-
	$K = 10$	6.1	62	1.549×10^{-3}	1.33 %	0.89 %	125.9	-
	$K = 15$	5	71	1.554×10^{-3}	1.13 %	0.87 %	126.4	-
	$K = 20$	4.2	75	1.555×10^{-3}	0.68 %	0.71 %	155.7	-
	Adaptive	6.7	61.2	1.552×10^{-3}	1.09 %	0.91 %	162.9	-

5.3. A jet engine turbine blade under both pressure loading and thermal stress

The third example considers the turbine blade of a jet engine under both the pressure load of surrounding gases and the thermal expansion. The finite-element model of this blade is built using the PDE toolbox in MATLAB, with a maximum mesh size of 0.01 m; see Fig. 10(a).



(a) finite-element mesh (b) temperature effect (c) typical deformation and stress

Figure 10: A jet engine turbine blade under both pressure and thermal effects

The turbine blade is made of nickel-based alloy material, with Young's modulus E, Poisson's ratio ν, coefficient of thermal expansion c_{te}, and thermal conductivity κ. Then, the face of the root that is in contact with other metal is fixed. The pressure loads due to surrounding gas are applied on the pressure and suction sides of the blade, denoted as p_1 and p_2, respectively. In this way, a total of 6 independent random parameters are considered in this system, i.e., $\{E, \nu, c_{\text{te}}, \kappa, p_1, p_2\}$. The statistical information of those random variables are given in Table 4.

Table 4: Random variables in the turbine blade example

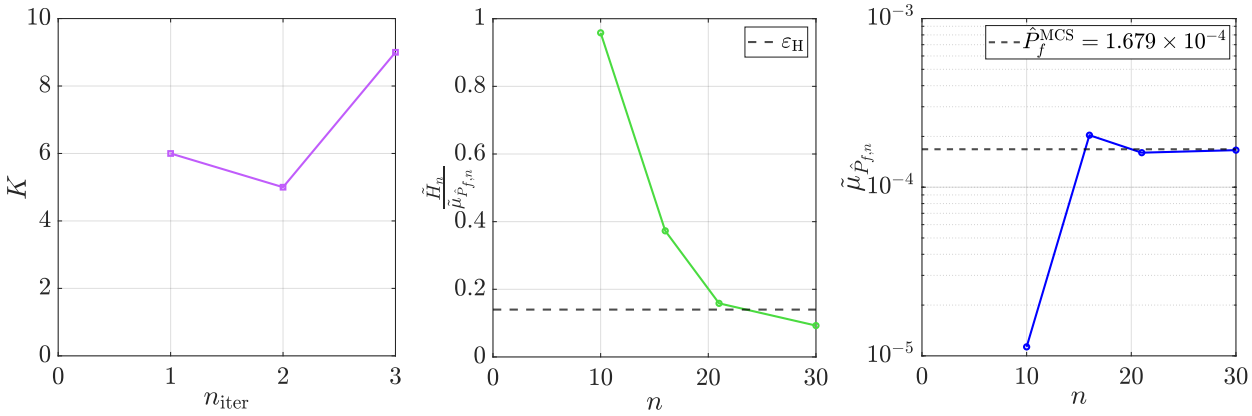
Variable	Unit	Description	Distribution	Mean	COV
E	Pa	Young's modulus	Lognormal	2.27×10^9	0.1
ν	-	Poisson's ratio	Lognormal	0.27	0.1
c_{te}	1/K	Coefficient of thermal expansion	Lognormal	1.27×10^{-5}	0.1
κ	W/m/K	Thermal conductivity	Lognormal	11.5	0.1
p_1	Pa	Pressure load	Gumbel	5×10^5	0.15
p_2	Pa	Pressure load	Gumbel	4.5×10^5	0.15

The convective heat transfer between the surrounding fluid and the faces of the blade defines the boundary conditions of thermal expansion, as shown in Fig. 10(b). Finally, the typical stress and deformation of the blade due to the combination of thermal and pressure effects is displayed in Fig. 10(c).

The maximum von Mises stress $\sigma(\boldsymbol{X})$ of the turbine blade is of concern, and the associated threshold is set as 1.15 GPa. Hence, the performance function $\mathcal{G}(\boldsymbol{X})$ is expressed as

$$\mathcal{G}(\boldsymbol{X}) = 1.15 - \sigma(\boldsymbol{X}). \tag{55}$$

Fig. 11 shows a single run of the proposed k-$\overline{\text{IPMR}}$ (adaptive scheme) in the turbine blade example. The proposed method converges with 3 iterations of multi-point enrichment process, and the corresponding batch size K in each iteration is displayed in Fig. 11(a). Then, the IPM $\frac{H_n}{\mu_{\widehat{P}_f}}$ is sequentially reduced with the addition of batch of new samples (Fig. 11(b)) and the $\mu_{\widehat{P}_f}$ finally achieves a good agreement with $\widehat{P}_f^{\text{MCS}}$, as illustrated in Fig. 11 (c).



(a) batch size K in each iteration (b) convergence of $\frac{H_n}{\mu_{\widehat{P}_{f,n}}}$ (c) convergence of $\mu_{\widehat{P}_{f,n}}$

Figure 11: The k-$\overline{\text{IPMR}}$ (adaptive scheme) in the turbine blade example

Table 5 lists the results of different reliability methods in the turbine blade example. Both the ALR module in UQLab and the PABQ provide relatively accurate estimates of failure probability $\mu_{\widehat{P}_f}$. Then, both the SUR and the k-IPMR need smaller values of n_{iter}, but their total computational time t_{c} are inversely greater. In contrast, the k-$\overline{\text{IPMR}}$ needs comparable values of n_{iter} but much smaller values of t_{c}, highlighting the advantage of k-$\overline{\text{IPMR}}$ in terms of computational time savings. Further, in the prescribed scheme of k-$\overline{\text{IPMR}}$, the n_{iter} reduces slowly and stagnates as the prescribed batch size K increases. Then, the resulting value of t_{c} increases with K, indicating that the blind increase of K does not always bring gain of reducing t_{c}. By contrast, the adaptive scheme effectively achieves a good balance between the resource consumption (n_{eval}) and the computational time t_{c}.

5.4. A reinforced concrete frame subjected to fully-nonstationary ground motion excitation

The final example considers a 3-bay, 6-storey planar reinforced concrete frame subjected to fully nonstationary stochastic ground motion excitation. Fig. 12 illustrates both geometric information and reinforcement

Table 5: Reliability results in the turbine blade example

Method		$\mathbb{E}[n_{\text{iter}}]$	$\mathbb{E}[n_{\text{eval}}]$	$\mathbb{E}[\mu_{\widehat{P}_f}]$	$\text{CV}[\mu_{\widehat{P}_f}]$	$\mathbb{E}[\delta_{\widehat{P}_f}]$	$\mathbb{E}[t_{\text{c}}]$(s)
MCS		-	1×10^6	1.679×10^{-4}	-	-	6.21×10^6
ALR in UQLab	$K = 5$	13.4	72	1.753×10^{-4}	3.99 %	5.35 %	320.8
	$K = 10$	7	70	1.796×10^{-4}	2.89 %	6.95 %	207.6
	$K = 15$	5.1	71.5	1.812×10^{-4}	2.38 %	7.90 %	180.1
	$K = 20$	4.7	84	1.810×10^{-4}	2.15 %	7.79 %	219.6
PABQ	$K = 5$	5.1	30.5	1.712×10^{-4}	3.22 %	2.95 %	136.7
	$K = 10$	4.4	44	1.706×10^{-4}	2.19 %	2.10 %	137.3
	$K = 15$	3.8	52	1.740×10^{-4}	2.56 %	3.83 %	146.6
	$K = 20$	3.8	66	1.729×10^{-4}	2.00 %	3.21 %	155.8
SUR	$K = 5$	4.1	25.5	1.661×10^{-4}	1.29 %	1.44 %	250.7
	$K = 10$	3.8	38	1.665×10^{-4}	1.14 %	1.26 %	228.5
	$K = 15$	3.4	46	1.660×10^{-4}	0.39 %	1.15 %	187.6
	$K = 20$	3.3	56	1.660×10^{-4}	0.31 %	1.15 %	293.2
k-IPMR	$K = 5$	4.4	27	1.654×10^{-4}	1.83 %	2.03 %	191.4
	$K = 10$	3.5	35	1.666×10^{-4}	1.35 %	1.27 %	182.9
	$K = 15$	3.3	44.5	1.654×10^{-4}	0.82 %	1.49 %	212.3
	$K = 20$	3.3	56	1.664×10^{-4}	0.55 %	0.95 %	275.8
	Adaptive	4.1	42.1	1.655×10^{-4}	0.30 %	1.47 %	203.4
k-$\overline{\text{IPMR}}$	$K = 5$	4.2	26	1.667×10^{-4}	0.90 %	0.86 %	76.1
	$K = 10$	3.3	33	1.656×10^{-4}	0.94 %	1.45 %	66.3
	$K = 15$	3.2	43	1.655×10^{-4}	0.54 %	1.45 %	79.3
	$K = 20$	3.2	54	1.657×10^{-4}	0.56 %	1.35 %	102.5
	Adaptive	4	29.5	1.665×10^{-4}	1.03 %	1.55 %	78.8

details of this structure. Finite-element analysis of this frame is conducted with the OpenSees software [56, 57]. Both beams and columns are modeled by the force-based elements with fiber-discretized section. The uniaxial relationships of concrete and rebar are described by the Concrete-01 and Steel-01 material models, respectively. The concrete slab on each floor is 100 mm thick, and the weight of the slab is then applied to those beams under it. Rayleigh damping is considered, with the damping ratio 5%. The related material parameters are regarded as random variables, with their statistical information given in Table 6.

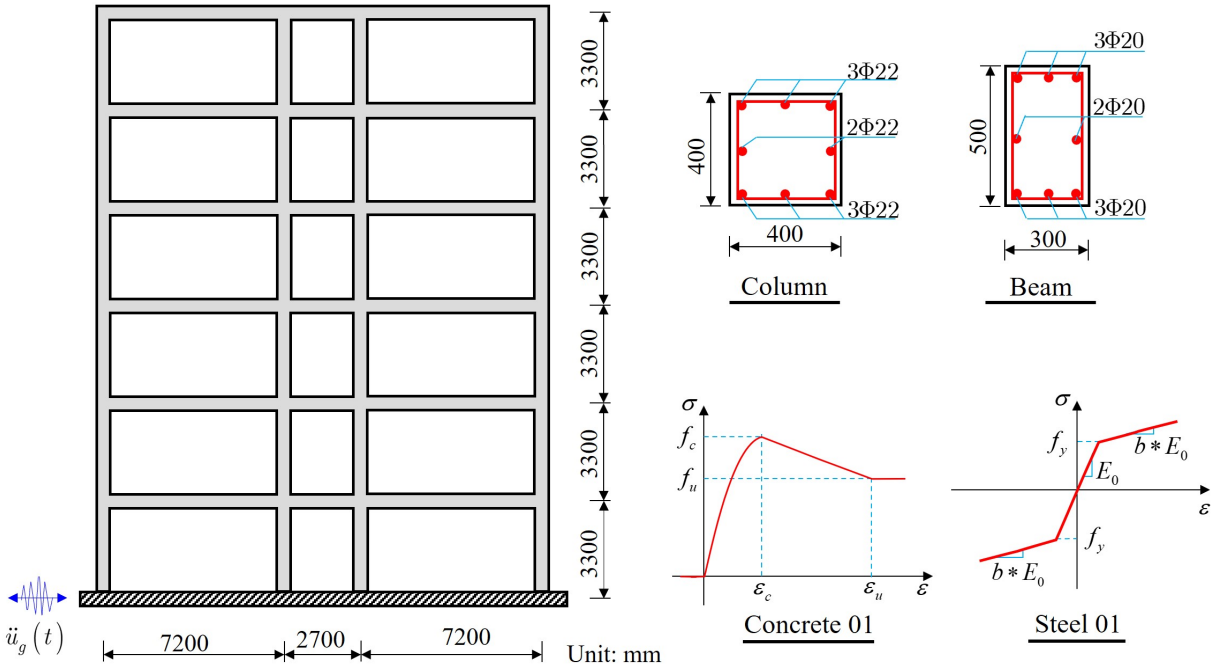


Figure 12: A planar reinforced concrete frame

A fully-nonstationary stochastic ground motion model, referred to as the modulated filtered white-noise

Table 6: Material parameters in the reinforced concrete frame

Variable	Unit	Description	Distribution	Mean	COV
f_{cc}	MPa	Maximum strength of confined concrete	Lognormal	35	0.1
$\varepsilon_{\mathrm{cc}}$	-	Strain at maximum strength of confined concrete	Lognormal	0.005	0.05
f_{cu}	MPa	Crushing strength of confined concrete	Lognormal	25	0.1
$\varepsilon_{\mathrm{cu}}$	-	Strain at crushing strength of confined concrete	Lognormal	0.02	0.05
f_{c}	MPa	Maximum strength of unconfined concrete	Lognormal	27	0.1
ε_{c}	-	Strain at maximum strength of unconfined concrete	Lognormal	0.002	0.05
f_{u}	MPa	Crushing strength of unconfined concrete	Lognormal	10	0.1
ε_{u}	-	Strain at crushing strength of unconfined concrete	Lognormal	0.006	0.05
f_{y}	MPa	Yielding strength of rebar	Lognormal	400	0.1
E_0	GPa	Young's modulus of rebar	Lognormal	200	0.1
b	-	Strain-hardening ratio of rebar	Lognormal	0.007	0.05

process model [58, 59], is considered to generate seismic accelerogram $\ddot{u}_g(t)$, formulated as

$$\ddot{u}_g(t;\Omega) = q\left(t;\Omega\right)\left\{\frac{1}{\sigma_h(t)}\int_0^t h\left[t-\tau,\boldsymbol{\lambda}(\tau;\Omega)\right]w(\tau)\mathrm{d}\tau\right\} \tag{56}$$

with a set of parameters Ω. Here, $q(t;\Omega)$ is a time-modulating function; $w(\tau)$ is a white noise process; $h\left[t-\tau,\boldsymbol{\lambda}(\tau;\Omega)\right]$ stands for the impulse response function of a filter with time-varying parameters $\boldsymbol{\lambda}(\tau;\Omega)$; $\sigma_h^2(t)=\int_0^t h^2[t-\tau,\boldsymbol{\lambda}(\tau;\Omega)]w(\tau)\mathrm{d}\tau$ is the variance of the filtered white noise process. Hence, the quantity in the curved bracket has a unit variance, and $q(t;\Omega)$ equals the standard deviation of process. In this way, $q(t;\Omega)$ defines the temporal feature of $\ddot{u}_g(t)$, while $h[t-\tau,\boldsymbol{\lambda}(\tau;\Omega)]$ defines the spectral characteristic of $\ddot{u}_g(t)$.

The 'Gamma' modulating function is adopted as [59]

$$q\left(t;\Omega\right)=\alpha_1 t^{\alpha_2-1}\exp\left(-\frac{t}{\alpha_3}\right), \tag{57}$$

where $\alpha_1,\alpha_2,\alpha_3$ control the intensity, shape, and duration of the process, respectively. The three parameters can be identified from the following three physical quantities related to the ground motion [59]: the expected Arias intensity I_a, the D_{5-95} that stands for the time interval between the instants when the 5% and 95% of I_a are reached, and the t_{mid} that represents time instant when the 45% of I_a is reached.

The $h[t-\tau,\boldsymbol{\lambda}(\tau;\Omega)]$ adopts the pseudo-acceleration response of a single-degree-of-freedom linear oscillator such that [59]

$$h\left[t-\tau,\boldsymbol{\lambda}(\tau;\Omega)\right]=\begin{cases}\frac{\omega_f(\tau)}{\sqrt{1-\zeta_f^2(\tau)}}\cdot\exp\left[-\zeta_f(\tau)\omega_f(\tau)(t-\tau)\right]\cdot\sin\left[\omega_f(\tau)\sqrt{1-\zeta_f^2(\tau)}(t-\tau)\right], & t\geq\tau,\\ 0, & \text{otherwise,}\end{cases} \tag{58}$$

where $\boldsymbol{\lambda}(\tau;\Omega)=[\omega_f(\tau),\zeta_f(\tau)]$. Then, the time-varying circular frequency $\omega_f(\tau)$ and damping ratio $\zeta_f(\tau)$ of filter are given as

$$\omega_f(\tau)=\omega_{\mathrm{mid}}+\omega'\left(\tau-t_{\mathrm{mid}}\right), \tag{59}$$

$$\zeta_f(\tau)=\zeta_f, \tag{60}$$

where ω_{mid} and ω' denote the circular frequency and the rate of change of frequency at the instant t_{mid}; ζ_f is a time-invariant damping ratio. Besides, to ensure zero residual velocity and displacement at the end of the accelerogram, the acceleration process in Eq. (56) is passed through the Butterworth high-pass filter, with the cutoff frequency $f_c=0.5\pi$. Given that ω' has a negligible effect on structural response, it is taken as a constant. Then, the statistical information of the remaining 5 parameters is listed in Table 7.

Fig. 13 shows the acceleration, velocity, and displacement time histories of three generated ground motion accelerograms. Clearly, the nonstationary characteristics in both temporal and spectral domains are well observed. Besides, zero residual velocity and displacement are fairly secured.

To summarize, a total of 16 random variables are involved, i.e., $\{f_{\mathrm{cc}},\varepsilon_{\mathrm{cc}},f_{\mathrm{cu}},\varepsilon_{\mathrm{cu}},f_{\mathrm{c}},\varepsilon_{\mathrm{c}},f_{\mathrm{u}},\varepsilon_{\mathrm{u}},f_{\mathrm{y}},E_0,b,I_a,D_{5-95},t_{\mathrm{mid}},\omega_{\mathrm{mid}}$ Figs. 14(a) and (b) present the typical stress-strain curves of concrete and rebar at the end section of the leftmost bottom column, when those random variables take their means. Then, Fig. 14(c) illustrates the relationship of restoring force vs. deformation at the bottom floor. Clearly, both material- and structure-level nonlinearity are well observed.

The inter-story drift $U_i(\boldsymbol{X},t), i=1,\ldots,6$, between the i-th and $(i-1)$-th floor is of interest. Then, the maximum allowable drift is set as $66=3300\times{}^1/_{50}$ mm. Hence, the failure probability of this dynamic system

Table 7: Random variables in the stochastic earthquake model [60]

Variable	Unit	Distribution	Parameters*	
I_a	m/s	Lognormal	1.9	0.3
D_{5-95}	s	Lognormal	2.21	0.23
t_{mid}	s	Lognormal	1.698	0.21
$\omega_{\text{mid}}/2\pi$	Hz	Uniform	2.8	4.8
ζ_f	-	Uniform	0.25	0.45

* Parameters of the random variables, rather than statistical moments.

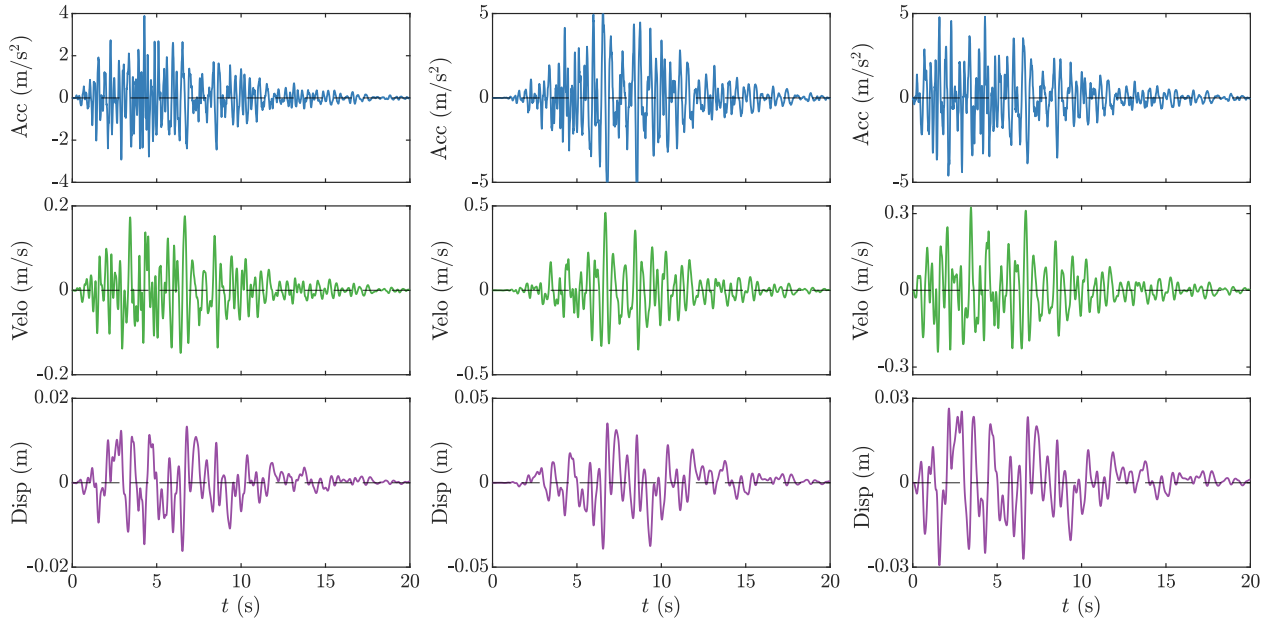
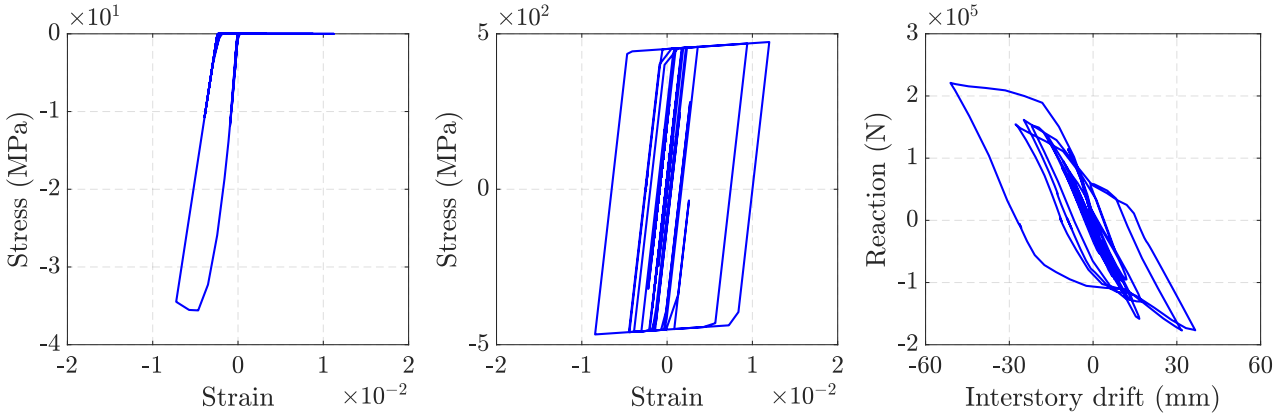


Figure 13: Generation of three typical fully-nonstationary seismic accelerograms



(a) hysteretic curve of concrete (b) hysteretic curve of rebar (c) hysteretic curve of column

Figure 14: Nonlinear behaviors of the reinforced concrete frame

is defined as [61]

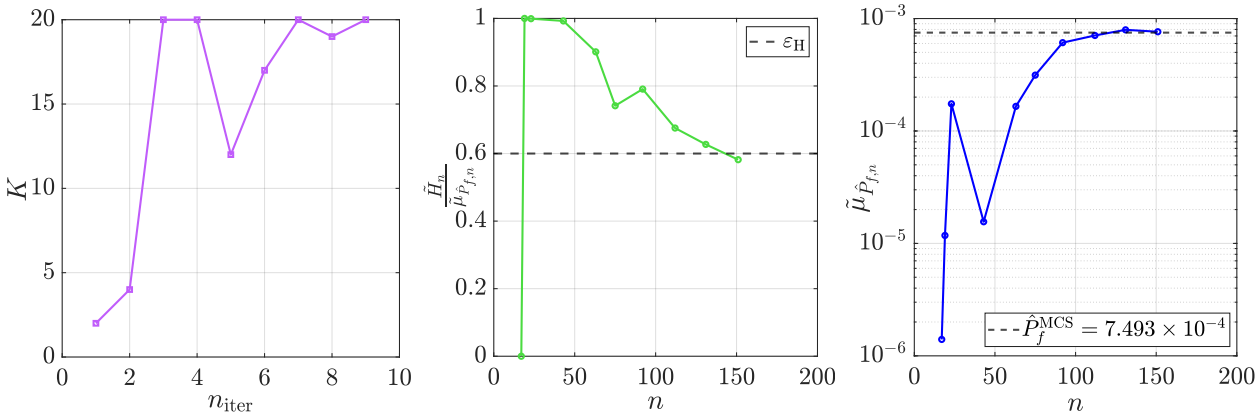
$$P_f = \mathbb{P}\left(\bigcup_{i=1}^{6}\left(\exists t \in [0, 20\text{s}], 66 - |U_i(\boldsymbol{X}, t)| \leq 0\right)\right), \tag{61}$$

which can be recast as the standard form in Eq. (1), with the performance function $\mathcal{G}(\boldsymbol{X})$ expressed as [62]

$$\mathcal{G}(\boldsymbol{X}) = 66 - \max_{1\leq i\leq 6}\left(\max_{t\in[0,20\text{s}]}|U_i(\boldsymbol{X}, t)|\right). \tag{62}$$

Fig. 15 illustrates a single run of the proposed k-$\overline{\text{IPMR}}$ (adaptive scheme) in the reinforced concrete frame example. First, the convergence of this algorithm is achieved by a total of 9 iterations of multi-point enrichment

process, with the corresponding batch size K per iteration shown in Fig. 15(a). Except for the first two iterations, at least 12 new samples are added in each iteration. Fig. 15(b) shows that the IPM $\frac{H_n}{\mu_{\widehat{P}_f}}$ is significantly reduced. Consequently, the $\mu_{\widehat{P}_f}$ converges gradually to $\widehat{P}_f^{\mathrm{MCS}}$; see Fig. 15(c).



(a) batch size K in each iteration (b) convergence of $\frac{H_n}{\mu_{\widehat{P}_{f,n}}}$ (c) convergence of $\mu_{\widehat{P}_{f,n}}$

Figure 15: The k-$\overline{\mathrm{IPMR}}$ (adaptive scheme) in the reinforced concrete frame example

Table 8 gives the comparative results of different reliability methods in the reinforced concrete frame example. First, both the ALR module in UQLab and the PABQ produce great values of $\delta_{\widehat{P}_f}$ at the cost of more than 300 runs of finite-element analysis of this frame. By contrast, both the SUR and the IPMR (k-IPMR and k-$\overline{\mathrm{IPMR}}$) achieve better accuracy of failure probability estimate with much fewer evaluations, say approximately 50% - 65%. However, both the SUR and the k-IPMR entail more computational time t_{c}, due to the time-intensive computation of themselves. Conversely, the k-$\overline{\mathrm{IPMR}}$ produces comparable accuracy of $\mu_{\widehat{P}_f}$ with much less computational time t_{c}, say around 10 %.

Table 8: Reliability results in the reinforced concrete frame example

Method		$\mathbb{E}[n_{\mathrm{iter}}]$	$\mathbb{E}[n_{\mathrm{eval}}]$	$\mathbb{E}[\mu_{\widehat{P}_f}]$	$\mathrm{CV}[\mu_{\widehat{P}_f}]$	$\mathbb{E}[\delta_{\widehat{P}_f}]$	$\mathbb{E}[t_{\mathrm{c}}]$(s)
MCS	-	-	1.2×10^6	7.493×10^{-4}	-	-	1.08×10^7
ALR in UQLab	$K = 5$	> 58	> 300	6.423×10^{-4}	4.74 %	12.02 %	> 1173.7
	$K = 10$	> 30	> 300	6.728×10^{-4}	17.01 %	13.91 %	> 882.5
	$K = 15$	> 20	> 300	6.471×10^{-4}	9.33 %	11.36 %	> 892.7
	$K = 20$	> 16	> 300	6.246×10^{-4}	9.47 %	15.03 %	> 947.3
PABQ	$K = 5$	> 58	> 300	2.803×10^{-4}	60.37 %	61.61 %	$> 3.23 \times 10^3$
	$K = 10$	> 30	> 300	3.300×10^{-4}	50.96 %	56.10 %	$> 1.83 \times 10^3$
	$K = 15$	> 20	> 300	4.161×10^{-4}	37.15 %	43.07 %	$> 1.54 \times 10^3$
	$K = 20$	> 16	> 300	3.743×10^{-4}	46.08 %	51.81 %	$> 1.34 \times 10^3$
SUR	$K = 5$	29.9	161.5	7.545×10^{-4}	7.72 %	6.54 %	6.51×10^3
	$K = 10$	17.5	182	7.420×10^{-4}	9.14 %	7.10 %	6.89×10^3
	$K = 15$	13.7	207.5	7.737×10^{-4}	5.28 %	5.12 %	8.01×10^3
	$K = 20$	13.1	259	7.654×10^{-4}	3.41 %	3.40 %	9.42×10^3
k-IPMR	$K = 5$	27	147	7.411×10^{-4}	9.04 %	5.73 %	6.23×10^3
	$K = 10$	14.9	156	7.552×10^{-4}	7.54 %	4.63 %	7.02×10^3
	$K = 15$	11.6	176	7.876×10^{-4}	11.40 %	9.94 %	7.93×10^3
	$K = 20$	9.4	185	7.741×10^{-4}	6.74 %	7.53 %	8.52×10^3
	Adaptive	10	175.2	7.375×10^{-4}	7.17 %	5.93 %	7.34×10^3
k-$\overline{\mathrm{IPMR}}$	$K = 5$	26.2	143	7.575×10^{-4}	9.15 %	7.18 %	684.3
	$K = 10$	15.8	166	7.724×10^{-4}	8.59 %	7.74 %	782.7
	$K = 15$	10.9	165.5	7.532×10^{-4}	9.01 %	4.42 %	776.8
	$K = 20$	10.3	203	7.851×10^{-4}	3.45 %	5.31 %	941.6
	Adaptive	10.5	181	7.573×10^{-4}	7.29 %	5.98 %	905.2

5.5. Discussions

Two critical discussions are made here to clarify the critical contributions of this study. First, the k-$\overline{\mathrm{IPMR}}$ is always advantageous over those existing (B)ALR methods in terms of n_{iter}. Admittedly, the inherent overhead

of k-$\overline{\text{IPMR}}$ itself per iteration is much less than that of SUR and k-IPMR, but is still greater than that of pointwise learning functions, e.g., the U function in the ALR module in UQLab and the UBVC in PABQ. Hence, in the k-$\overline{\text{IPMR}}$, whether its advantage of n_{iter} can be converted to that of t_{c} or not depends on the evaluation time of performance function $\mathcal{G}(\cdot)$ being considered. If a single evaluation of $\mathcal{G}(\cdot)$ is very expensive, its evaluation time will dominate the running time of each iteration of BALR, and the benefit of reducing n_{iter} will be readily converted to that of reducing t_{c}. In the last three examples, the evaluation time of $\mathcal{G}$-function gradually increases. Hence, the advantage of k-$\overline{\text{IPMR}}$ in terms of t_c becomes more and more prominent.

Second, a parallel active learning strategy is developed in the framework of PDEM in previous study [63]. Then, the similarities and differences between this study and previous study are outlined as follows.

- *Similarities*. Both of them adhere to a similar logical framework, i.e., multi-point look-ahead paradigm, to build the critical ingredients of parallel active learning workflow.
- *Differences*. (i) Their global measures of epistemic uncertainty of failure probability differ from each other, due to different inherent philosophies between PDEM and Bayesian inference of failure probability. (ii) The number of representative points in PDEM is only $\mathcal{O}(10^3)$. Hence, there is no need for the pruning of sample set in previous study. (iii) The PDEM is known to work well in dynamic reliability problems but degrade in small failure probabilities, say $P_f \leq \mathcal{O}(10^{-4})$. By contrast, good performance of Bayesian inference of failure probability has been verified in relatively small failure probabilities [33, 30] and is also testified in this study. Hence, it is necessary to develop efficient parallel enrichment strategy for Bayesian active learning reliability analysis, which is exactly the main motivation of this study.

In summary, the proposed parallel enrichment strategy is totally goal-oriented, so as to be as efficient as possible.

6. Concluding remarks

A new parallel Bayesian active learning reliability method is developed in this study. Starting by a well-defined global measure of residual uncertainty of failure probability, a multi-point learning function called k-$\overline{\text{IPMR}}$ is developed to select a sequence of new samples that reduce the IPM dramatically. Besides, an adaptive scheme is devised to identify the rational size of batch of new samples added in each iteration. Clearly, the multi-point enrichment process is implemented based on the k-$\overline{\text{IPMR}}$ itself, without resorting to additional parallel selection procedures. Thanks to the low computational costs of both IPM and k-$\overline{\text{IPMR}}$, the proposed approach comes with better theoretical consistency and higher computational efficiency. Some concluding remarks are summarized as follows.

- In the Bayesian inference of failure probability, the IPM is rigorously proved to be the upper bound of mean absolute deviation of failure probability, and it can be thus viewed as a computationally cheaper alternative of the posterior variance of failure probability.
- Guided by reducing the IPM, a multi-point learning function k-$\overline{\text{IPMR}}$ is analytically derived as a single integral. Besides, a hybrid convergence criterion is developed according to the actual reduction of IPM in each iteration.
- The multi-point enrichment process is efficiently performed based on three key workarounds, i.e., fast single evaluation of k-$\overline{\text{IPMR}}$, stepwise maximization of k-$\overline{\text{IPMR}}$, and adaptive scheme of identifying the batch size of new samples per iteration.
- Thanks to the core role of IPM, the three main ingredients, i.e., Bayesian inference of failure probability, multi-point learning function, and convergence criterion, are fused in a compatible and consistent way.

Admittedly, in comparison with the exact posterior variance of failure probability, the IPM is still a relatively looser metric. A much tighter residual uncertainty measure of failure probability and the resulting learning function will be further explored in the near future.

CRediT authorship contribution statement

Tong Zhou: Conceptualization, Formal analysis, Methodology, Software, Validation, Visualization, Writing - original draft, Funding acquisition. **Tong Guo**: Writing - review & editing. **Xujia Zhu**: Conceptualization, Methodology, Writing -review & editing. **Masaru Kitahara**: Writing -review & editing. **Jize Zhang**: Writing -review & editing, Funding acquisition.

Conflict of Interest

The authors declare that they have no conflicts of interest in this work.

Data Availability

Data will be made available on request.

Acknowledgements

The support of the National Natural Science Foundation of China (Grant No. 52408222), the Research Grants Council of Hong Kong (Grant No. 26200124), and Université Paris-Saclay (Welcome Research Package No. WRP394) are highly appreciated.

Appendix A. Basics of Kriging

The Kriging assumes the response of $\mathcal{G}(\boldsymbol{x})$ as one possible realization of an underlying Gaussian process, expressed as [16]

$$\mathcal{G}(\boldsymbol{x}) \approx \widehat{\mathcal{G}}_n(\boldsymbol{x}) = \boldsymbol{\beta}^\top \boldsymbol{f}(\boldsymbol{x}) + \sigma^2 \mathcal{W}(\boldsymbol{x}), \tag{A.1}$$

where the trend function $\boldsymbol{\beta}^\top \boldsymbol{f}(\boldsymbol{x}) = \beta_0 + \sum_{l=1}^d \beta_l x_l$, with $\boldsymbol{f}(\boldsymbol{x}) = \{1, x_1, \ldots, x_d\}$ a set of basis functions, and $\boldsymbol{\beta} = \{\beta_0, \beta_1, \ldots, \beta_d\}$ a set of unknown coefficients. σ^2 is the variance of Gaussian process; $\mathcal{W}(\boldsymbol{x}) \sim \mathcal{GP}(\boldsymbol{0}, R(\boldsymbol{x}, \boldsymbol{x}'))$ is a stationary Gaussian process with zero mean, unit variance and a known correlation function $R(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{\theta})$. The Matern-5/2 correlation function is expressed as [16]

$$R(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{\theta}) = \prod_{l=1}^d \left(1 + \sqrt{5}\frac{|x_l - x_l'|}{\theta_l} + \frac{5}{3}\left(\frac{|x_l - x_l'|}{\theta_l}\right)^2\right) \exp\left(-\sqrt{5}\frac{|x_l - x_l'|}{\theta_l}\right), \tag{A.2}$$

where the kernel parameters $\boldsymbol{\theta} = \{\theta_l > 0\}_{l=1}^d$.

When providing an ED of size n, denoted as $\mathcal{D}_n = \{\mathcal{X}_n, \mathcal{Y}_n\} = \left\{\left(\boldsymbol{x}^{(i)}, y^{(i)}\right), i = 1, \ldots, n\right\}$, both $\boldsymbol{\beta}$ and σ^2 can be estimated as

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{F}^\top \boldsymbol{R}^{-1} \boldsymbol{F}\right)^{-1} \boldsymbol{F}^\top \boldsymbol{R}^{-1} \mathcal{Y}_n, \tag{A.3}$$

$$\widehat{\sigma}^2 = \frac{1}{n} \left(\mathcal{Y}_n - \boldsymbol{F}\boldsymbol{\beta}\right)^\top \boldsymbol{R}^{-1} \left(\mathcal{Y}_n - \boldsymbol{F}\boldsymbol{\beta}\right), \tag{A.4}$$

where $\boldsymbol{F} \coloneqq \left[f_j\left(\boldsymbol{x}^{(i)}\right)\right]_{1 \le i \le n, 1 \le j \le d+1}$ is the information matrix; $\boldsymbol{R} \coloneqq \left[R\left(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}; \boldsymbol{\theta}\right)\right]_{1 \le i,j \le n}$ is the matrix of correlations between all samples in the $\mathcal{X}_n$. Both $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}^2$ depend on $\boldsymbol{\theta}$, which can be then estimated as [16]

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \Theta} \widehat{\sigma}^2 |\boldsymbol{R}|^{\frac{1}{n}}, \tag{A.5}$$

where Θ is the feasible domain of $\boldsymbol{\theta}$.

Finally, the Kriging $\widehat{\mathcal{G}}_n(\boldsymbol{x})$ conditional on $\mathcal{D}_n$ is still a Gaussian process, that is,

$$\widehat{\mathcal{G}}_n(\boldsymbol{x}) \sim \mathcal{GP}\left(\mu_n(\boldsymbol{x}), c_n(\boldsymbol{x}, \boldsymbol{x}')\right), \tag{A.6}$$

with its mean $\mu_n(\boldsymbol{x})$, variance $\sigma_n^2(\boldsymbol{x})$ and covariance $c_n(\boldsymbol{x}, \boldsymbol{x}')$ are expressed as [16]

$$\mu_n(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})^\top \widehat{\boldsymbol{\beta}} + \boldsymbol{r}(\boldsymbol{x})^\top \boldsymbol{R}^{-1} \left(\mathcal{Y}_n - \boldsymbol{F}\widehat{\boldsymbol{\beta}}\right), \tag{A.7}$$

$$\sigma_n^2(\boldsymbol{x}) = \widehat{\sigma}^2 \left(1 - \boldsymbol{r}(\boldsymbol{x})^\top \boldsymbol{R}^{-1} \boldsymbol{r}(\boldsymbol{x}) + \boldsymbol{u}(\boldsymbol{x})^\top \left(\boldsymbol{F}^\top \boldsymbol{R}^{-1} \boldsymbol{F}\right)^{-1} \boldsymbol{u}(\boldsymbol{x})\right), \tag{A.8}$$

$$c_n(\boldsymbol{x}, \boldsymbol{x}') = \widehat{\sigma}^2 \left(R(\boldsymbol{x}, \boldsymbol{x}') - \boldsymbol{r}(\boldsymbol{x})^\top \boldsymbol{R}^{-1} \boldsymbol{r}(\boldsymbol{x}') + \boldsymbol{u}(\boldsymbol{x})^\top \left(\boldsymbol{F}^\top \boldsymbol{R}^{-1} \boldsymbol{F}\right)^{-1} \boldsymbol{u}(\boldsymbol{x}')\right), \tag{A.9}$$

where $\boldsymbol{r}(\boldsymbol{x}) = \left[R\left(\boldsymbol{x}, \boldsymbol{x}^{(1)}\right), \ldots, R\left(\boldsymbol{x}, \boldsymbol{x}^{(n)}\right)\right]^\top$, and $\boldsymbol{u}(\boldsymbol{x}) = \boldsymbol{F}^\top \boldsymbol{R}^{-1} \boldsymbol{r}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x})$.

Appendix B. Proof of Proposition 1

Proof. First, the mean absolute deviation of $\widehat{P}_{f,n}$ satisfies the following inequality

$$\begin{aligned}
\mathbb{E}_n\left[\left|\widehat{P}_{f,n}-\mu_{\widehat{P}_{f,n}}\right|\right] &= \mathbb{E}_n\left[\left|\int_{\mathbb{X}}\widehat{\mathbf{1}}_n(\boldsymbol{x})\,f_{\boldsymbol{X}}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}-\int_{\mathbb{X}}\mu_{\widehat{\mathbf{1}}_n}(\boldsymbol{x})f_{\boldsymbol{X}}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\right|\right],\\
&= \mathbb{E}_n\left[\left|\int_{\mathbb{X}}\left(\widehat{\mathbf{1}}_n(\boldsymbol{x})-\mu_{\widehat{\mathbf{1}}_n}(\boldsymbol{x})\right)f_{\boldsymbol{X}}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\right|\right],\\
&\le \mathbb{E}_n\left[\int_{\mathbb{X}}\left|\widehat{\mathbf{1}}_n(\boldsymbol{x})-\mu_{\widehat{\mathbf{1}}_n}(\boldsymbol{x})\right|f_{\boldsymbol{X}}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\right],\\
&= \int_{\mathbb{X}}\mathbb{E}_n\left[\left|\widehat{\mathbf{1}}_n(\boldsymbol{x})-\mu_{\widehat{\mathbf{1}}_n}(\boldsymbol{x})\right|\right]f_{\boldsymbol{X}}(\boldsymbol{x})\mathrm{d}\boldsymbol{x},\\
&= \mathbb{E}_{\boldsymbol{X}}\left[\mathbb{E}_n\left[\left|\widehat{\mathbf{1}}_n(\boldsymbol{x})-\mu_{\widehat{\mathbf{1}}_n}(\boldsymbol{x})\right|\right]\right],
\end{aligned}\tag{B.1}$$

where $\mathbb{E}_n\left[\left|\widehat{\mathbf{1}}_n(\boldsymbol{x})-\mu_{\widehat{\mathbf{1}}_n}(\boldsymbol{x})\right|\right]$ denotes the mean absolute deviation of $\widehat{\mathbf{1}}_n(\boldsymbol{x})$.

Then, there exists

$$\left|\widehat{\mathbf{1}}_n(\boldsymbol{x})-\mu_{\widehat{\mathbf{1}}_n}(\boldsymbol{x})\right| = \begin{cases}\left|1-\Phi\left(-\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right)\right|, & \widehat{\mathcal{G}}_n(\boldsymbol{x})\le 0\\ \left|0-\Phi\left(-\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right)\right|, & \text{otherwise}\end{cases} = \begin{cases}\Phi\left(\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right), & \widehat{\mathcal{G}}_n(\boldsymbol{x})\le 0,\\ \Phi\left(-\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right), & \text{otherwise},\end{cases}\tag{B.2}$$

hence, $\mathbb{E}_n\left[\left|\widehat{\mathbf{1}}_n(\boldsymbol{x})-\mu_{\widehat{\mathbf{1}}_n}(\boldsymbol{x})\right|\right]$ can be computed as

$$\begin{aligned}
\mathbb{E}_n\left[\left|\widehat{\mathbf{1}}_n(\boldsymbol{x})-\mu_{\widehat{\mathbf{1}}_n}(\boldsymbol{x})\right|\right] &= \int_{-\infty}^{0}\Phi\left(\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right)f_{\widehat{Y}}(\widehat{y})\mathrm{d}\widehat{y}+\int_{0}^{+\infty}\Phi\left(-\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right)f_{\widehat{Y}}(\widehat{y})\mathrm{d}\widehat{y},\\
&= \Phi\left(\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right)\Phi\left(-\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right)+\Phi\left(-\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right)\Phi\left(\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right),\\
&= 2\Phi\left(\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right)\Phi\left(-\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right).
\end{aligned}\tag{B.3}$$

Further, given that

$$\Phi\left(\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right)\Phi\left(-\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right)\le\min\left\{\Phi\left(\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right),\Phi\left(-\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right)\right\}=\Phi\left(-\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}\right)=\pi_n(\boldsymbol{x}),\tag{B.4}$$

there exits

$$\mathbb{E}_n\left[\left|\widehat{\mathbf{1}}_n(\boldsymbol{x})-\mu_{\widehat{\mathbf{1}}_n}(\boldsymbol{x})\right|\right]\le 2\pi_n(\boldsymbol{x}).\tag{B.5}$$

Finally, substituting Eq. (B.5) into Eq. (B.1) proves the inequality in Eq. (17), that is,

$$\mathbb{E}_n\left[\left|\widehat{P}_{f,n}-\mu_{\widehat{P}_{f,n}}\right|\right]\le 2\mathbb{E}_{\boldsymbol{X}}\left[\pi_n(\boldsymbol{x})\right]=2H_n.\tag{B.6}$$

□

Appendix C. Proof of Proposition 2

Proof. First, by definition of $\Phi(\cdot)$, the term $\mathbb{E}_{\boldsymbol{U}_k^+}\left[\Pi_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)\right]$ in Eq. (24) can be expressed as

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{U}_k^+}\left[\Pi_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)\right] &= \int_{\mathbb{R}^k}\Phi\left(-\left|a(\boldsymbol{x})+\boldsymbol{b}(\boldsymbol{x})^\top\mathcal{U}_k^+\right|\right)f_{\boldsymbol{U}_k^+}(\mathcal{U}_k^+)\mathrm{d}\mathcal{U}_k^+,\\
&= \int_{\mathbb{R}^k}\mathbb{P}\left(\Lambda\le-\left|a(\boldsymbol{x})+\boldsymbol{b}(\boldsymbol{x})^\top\mathcal{U}_k^+\right|\right)f_{\boldsymbol{U}_k^+}(\mathcal{U}_k^+)\mathrm{d}\mathcal{U}_k^+,\\
&= \mathbb{P}\left(\Lambda\le-\left|a(\boldsymbol{x})+\boldsymbol{b}(\boldsymbol{x})^\top\boldsymbol{U}_k^+\right|\right),\\
&= \mathbb{P}\left(\Lambda+|W|\le 0\right),\\
&= \mathbb{P}\left(\Lambda+W\le 0,W\ge 0\right)+\mathbb{P}\left(\Lambda-W\le 0,W<0\right),\\
&= \underbrace{\mathbb{P}\left(\Lambda+W\le 0,-W\le 0\right)}_{①}+\underbrace{\mathbb{P}\left(\Lambda-W\le 0,W<0\right)}_{②},
\end{aligned}\tag{C.1}$$

where $\Lambda \sim \mathcal{N}(0,1)$ is a standard Gaussian variable independent of $\boldsymbol{U}_k^+$; then, $W = a(\boldsymbol{x}) + \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{U}_k^+$ is a Gaussian variable, with its mean μ_W and variance σ_W^2 given by

$$\begin{cases} \mu_W = \mathbb{E}\big[a(\boldsymbol{x}) + \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{U}_k^+\big] = a(\boldsymbol{x}) + \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{0} = a(\boldsymbol{x}) = \dfrac{\mu_n(\boldsymbol{x})}{\sigma_{n+k}(\boldsymbol{x})}, \\ \sigma_W^2 = \mathbb{E}\big[a(\boldsymbol{x}) + \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{U}_k^+\big] = \boldsymbol{b}(\boldsymbol{x})^\top \mathcal{C}_k^+ \boldsymbol{b}(\boldsymbol{x}) = \dfrac{c_n\left(\boldsymbol{x}, \mathcal{X}_k^+\right)^\top \left(\mathcal{C}_k^+\right)^{-1}}{\sigma_{n+k}(\boldsymbol{x})} \mathcal{C}_k^+ \dfrac{\left(\mathcal{C}_k^+\right)^{-1} c_n\left(\boldsymbol{x}, \mathcal{X}_k^+\right)}{\sigma_{n+k}(\boldsymbol{x})} = \dfrac{\sigma_n^2(\boldsymbol{x})}{\sigma_{n+k}^2(\boldsymbol{x})} - 1. \end{cases} \tag{C.2}$$

Then, the vector $\begin{bmatrix} \Lambda + W \\ -W \end{bmatrix}$ in ① follows bivariate Gaussian distribution, with the components of their mean vector and covariance matrix given as

$$\begin{cases} \mathbb{E}\big[\Lambda + W\big] = 0 + \mu_W = \mu_W, \\ \mathrm{Var}\big[\Lambda + W\big] = 1 + \mathrm{Var}\big[W\big] = 1 + \sigma_W^2, \\ \mathrm{Cov}\left[\Lambda + W, -W\right] = -\mathrm{Cov}\left[\Lambda, W\right] - \mathrm{Cov}\left[W, W\right] = -\sigma_W^2. \end{cases} \tag{C.3}$$

In this way, $\begin{bmatrix} \Lambda + W \\ -W \end{bmatrix} \sim \mathcal{N}_2\left(\begin{bmatrix} \mu_W \\ -\mu_W \end{bmatrix}, \begin{bmatrix} 1+\sigma_W^2 & -\sigma_W^2 \\ -\sigma_W^2 & \sigma_W^2 \end{bmatrix}\right)$, and the ① is equivalent to

$$\begin{aligned} ① &= \Phi_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}; \begin{bmatrix} \mu_W \\ -\mu_W \end{bmatrix}, \begin{bmatrix} 1+\sigma_W^2 & -\sigma_W^2 \\ -\sigma_W^2 & \sigma_W^2 \end{bmatrix}\right) = \Phi_2\left(\begin{bmatrix} \frac{-\mu_W}{\sqrt{1+\sigma_W^2}} \\ \frac{\mu_W}{\sigma_W} \end{bmatrix}; \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & -\frac{\sigma_W}{\sqrt{1+\sigma_W^2}} \\ -\frac{\sigma_W}{\sqrt{1+\sigma_W^2}} & 1 \end{bmatrix}\right), \\ &= \Phi_2\left(\begin{bmatrix} -\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})} \\ \frac{\mu_n(\boldsymbol{x})}{\sqrt{c_n\left(\boldsymbol{x},\mathcal{X}_k^+\right)^\top \left(\mathcal{C}_k^+\right)^{-1} c_n\left(\boldsymbol{x},\mathcal{X}_k^+\right)}} \end{bmatrix}; \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & -\frac{\sqrt{c_n\left(\boldsymbol{x},\mathcal{X}_k^+\right)^\top \left(\mathcal{C}_k^+\right)^{-1} c_n\left(\boldsymbol{x},\mathcal{X}_k^+\right)}}{\sigma_n(\boldsymbol{x})} \\ -\frac{\sqrt{c_n\left(\boldsymbol{x},\mathcal{X}_k^+\right)^\top \left(\mathcal{C}_k^+\right)^{-1} c_n\left(\boldsymbol{x},\mathcal{X}_k^+\right)}}{\sigma_n(\boldsymbol{x})} & 1 \end{bmatrix}\right), \\ &= \Phi_2\left(\begin{bmatrix} -\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})} \\ \frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})} / \rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right) \end{bmatrix}; \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & -\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right) \\ -\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right) & 1 \end{bmatrix}\right), \end{aligned} \tag{C.4}$$

where $\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)$ is defined as

$$\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right) = \frac{\sqrt{c_n\left(\boldsymbol{x},\mathcal{X}_k^+\right)^\top \left(\mathcal{C}_k^+\right)^{-1} c_n\left(\boldsymbol{x},\mathcal{X}_k^+\right)}}{\sigma_n(\boldsymbol{x})} \in [0,1]. \tag{C.5}$$

Similarly, the vector $\begin{bmatrix} \Lambda - W \\ W \end{bmatrix}$ in ② follows bivariate Gaussian distribution, with the components of their mean vector and covariance matrix expressed as

$$\begin{cases} \mathbb{E}\big[\Lambda - W\big] = 0 - \mu_W = -\mu_W, \\ \mathrm{Var}\big[\Lambda - W\big] = 1 + \mathrm{Var}\big[W\big] = 1 + \sigma_W^2, \\ \mathrm{Cov}\left[\Lambda - W, W\right] = \mathrm{Cov}\left[\Lambda, W\right] - \mathrm{Cov}\left[W, W\right] = -\sigma_W^2. \end{cases} \tag{C.6}$$

In this regard, $\begin{bmatrix} \Lambda - W \\ W \end{bmatrix} \sim \mathcal{N}_2\left(\begin{bmatrix} -\mu_W \\ \mu_W \end{bmatrix}, \begin{bmatrix} 1+\sigma_W^2 & -\sigma_W^2 \\ -\sigma_W^2 & \sigma_W^2 \end{bmatrix}\right)$, and the ② is recast as

$$\begin{aligned} ② &= \Phi_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}; \begin{bmatrix} -\mu_W \\ \mu_W \end{bmatrix}, \begin{bmatrix} 1+\sigma_W^2 & -\sigma_W^2 \\ -\sigma_W^2 & \sigma_W^2 \end{bmatrix}\right), \\ &= \Phi_2\left(\begin{bmatrix} \frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})} \\ -\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})} / \rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right) \end{bmatrix}; \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & -\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right) \\ -\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right) & 1 \end{bmatrix}\right). \end{aligned} \tag{C.7}$$

Substitute Eqs. (C.4) and (C.7) into Eq. (C.1), giving rise to

$$\begin{aligned} \mathbb{E}_{\boldsymbol{U}_k^+}\left[\Pi_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)\right] &= \Phi_2\left(\begin{bmatrix} -\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})} \\ \frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})} / \rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right) \end{bmatrix}; \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & -\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right) \\ -\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right) & 1 \end{bmatrix}\right), \\ &+ \Phi_2\left(\begin{bmatrix} \frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})} \\ -\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})} / \rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right) \end{bmatrix}; \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & -\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right) \\ -\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right) & 1 \end{bmatrix}\right), \end{aligned} \tag{C.8}$$

which is a function of $\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}$ and $\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)$. Further, Eq. (C.8) implies that $\mathbb{E}_{\boldsymbol{U}_k^+}\left[\Pi_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)\right]$ is an even function with respect to $\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}$. Hence, Eq. (C.8) can be rewritten as

$$\begin{aligned}\mathbb{E}_{\boldsymbol{U}_k^+}\left[\Pi_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)\right] &= \Phi_2\left(\begin{bmatrix}-\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}\\ \frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}/\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)\end{bmatrix};\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1 & -\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)\\ -\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right) & 1\end{bmatrix}\right),\\ &+\Phi_2\left(\begin{bmatrix}\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}\\ -\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}/\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)\end{bmatrix};\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1 & -\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)\\ -\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right) & 1\end{bmatrix}\right).\end{aligned} \tag{C.9}$$

Recall that [64]

$$\Phi_2\left(\begin{bmatrix}-h_1\\-h_2\end{bmatrix};\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1 & r\\ r & 1\end{bmatrix}\right) = \Phi_2\left(\begin{bmatrix}h_1\\h_2\end{bmatrix};\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1 & r\\ r & 1\end{bmatrix}\right) - \Phi\left(h_1\right) - \Phi\left(h_2\right) + 1, \tag{C.10}$$

then, Eq. (C.9) can be recast as

$$\begin{aligned}&\mathbb{E}_{\boldsymbol{U}_k^+}\left[\Pi_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)\right]\\ &= 2\Phi_2\left(\begin{bmatrix}\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}\\ -\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}/\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)\end{bmatrix};\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1 & -\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)\\ -\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right) & 1\end{bmatrix}\right) - \Phi\left(-\frac{\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}}{\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)}\right) - \Phi\left(\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}\right) + 1,\\ &= \Phi\left(-\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}\right) - \Phi\left(-\frac{\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}}{\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)}\right) + 2\Phi_2\left(\begin{bmatrix}\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}\\ -\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}/\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)\end{bmatrix};\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1 & -\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)\\ -\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right) & 1\end{bmatrix}\right).\end{aligned} \tag{C.11}$$

Finally, substituting Eq. (C.11) into Eq. (24) proves the analytical expression of $\Gamma_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)$ in Eq. (26), that is,

$$\begin{aligned}\Gamma_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right) &= \pi_n(\boldsymbol{x}) - \mathbb{E}_{\boldsymbol{U}_k^+}\left[\Pi_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)\right],\\ &= \Phi\left(-\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}\right) - \Phi\left(-\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}\right) + \Phi\left(\frac{-\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}}{\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)}\right)\\ &\quad - 2\Phi_2\left(\begin{bmatrix}\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}\\ -\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}/\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)\end{bmatrix};\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1 & -\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)\\ -\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right) & 1\end{bmatrix}\right),\\ &= \Phi\left(\frac{-\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}}{\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)}\right) - 2\Phi_2\left(\begin{bmatrix}\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}\\ -\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}/\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)\end{bmatrix};\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1 & -\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)\\ -\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right) & 1\end{bmatrix}\right),\end{aligned} \tag{C.12}$$

which is a function of both $\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})} \in [0,+\infty)$ and $\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right) \in [0,1]$. □

Appendix D. Proof of Proposition 3

Proof. The upper bound $\overline{\Gamma}_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)$ is a direct consequence of removing the second term in Eq. (26). Then, we will focus on the lower bound in the sequel.

Given the definition of $\pi_n(\boldsymbol{x})$ in Eq. (16), it can be expressed as

$$\pi_n(\boldsymbol{x}) = \min\left\{\mathbb{P}\left(\widehat{\mathcal{G}}_n(\boldsymbol{x}) \le 0\right), \mathbb{P}\left(\widehat{\mathcal{G}}_n(\boldsymbol{x}) > 0\right)\right\} = \min\left\{\mathbb{P}\left(\widehat{\mathcal{G}}_n(\boldsymbol{x}) \le 0\right), 1 - \mathbb{P}\left(\widehat{\mathcal{G}}_n(\boldsymbol{x}) \le 0\right)\right\}, \tag{D.1}$$

where $\widehat{\mathcal{G}}_n(\boldsymbol{x}) \sim \mathcal{N}\left(\mu_n(\boldsymbol{x}), \sigma_n^2(\boldsymbol{x})\right)$ denotes the Kriging prediction at $\boldsymbol{x}$ conditional on the current ED $\mathcal{D}_n$.

Then, let us define an auxiliary function

$$J(s) = \min\{s, 1-s\},\ s \in [0,1], \tag{D.2}$$

which is concave, due to the fact that for any $s,t \in [0,1]$ and for any $\lambda \in [0,1]$, there holds

$$\begin{aligned}\lambda J(s) + (1-\lambda)J(t) &= \min\{\lambda s, \lambda(1-s)\} + \min\{(1-\lambda)t, (1-\lambda)(1-t)\},\\ &\le \min\{\lambda s + (1-\lambda)t, \lambda(1-s) + (1-\lambda)(1-t)\},\\ &= J(\lambda s + (1-\lambda)t).\end{aligned} \tag{D.3}$$

With the help of $J(\cdot)$, we could rewrite $\pi_n(\boldsymbol{x})$ and $\pi_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+,\mathcal{Y}_k^+\right)$ (Eq. (20)) as

$$\begin{cases} \pi_n(\boldsymbol{x}) = J\left(\mathbb{P}\left(\widehat{\mathcal{G}}_n(\boldsymbol{x}) \le 0\right)\right), \\ \pi_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+,\mathcal{Y}_k^+\right) = J\left(\mathbb{P}\left(\widehat{\mathcal{G}}_{n+k}(\boldsymbol{x} \mid \mathcal{X}_k^+,\mathcal{Y}_k^+) \le 0\right)\right), \end{cases} \tag{D.4}$$

respectively.

Following Jensen's inequality, there exists

$$\begin{aligned} \mathbb{E}_{\boldsymbol{Y}_k^+}\left[\Pi_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)\right] &= \mathbb{E}_{\boldsymbol{Y}_k^+}\left[J\left(\mathbb{P}\left(\widehat{\mathcal{G}}_{n+k}(\boldsymbol{x} \mid \mathcal{X}_k^+,\boldsymbol{Y}_k^+ = \mathcal{Y}_k^+) \le 0\right)\right)\right], \\ &\le J\left(\mathbb{E}_{\boldsymbol{Y}_k^+}\left[\mathbb{P}\left(\widehat{\mathcal{G}}_{n+k}(\boldsymbol{x} \mid \mathcal{X}_k^+,\boldsymbol{Y}_k^+ = \mathcal{Y}_k^+) \le 0\right)\right]\right), \\ &\overset{*}{=} J\left(\mathbb{P}\left(\widehat{\mathcal{G}}_n(\boldsymbol{x}) \le 0\right)\right), \\ &= \pi_n(\boldsymbol{x}), \end{aligned} \tag{D.5}$$

where the equality $\overset{*}{=}$ follows the law of total probability.

Consequently, $\Gamma_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)$ in Eq. (24) is non-negative, i.e.,

$$\Gamma_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right) = \pi_n(\boldsymbol{x}) - \mathbb{E}_{\boldsymbol{Y}_k^+}\left[\Pi_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right)\right] \ge 0, \tag{D.6}$$

hence, $\underline{\Gamma}_{n+k}\left(\boldsymbol{x};\mathcal{X}_k^+\right) = 0$. □

Appendix E. Characteristic of a promising candidate batch informed by k-IPMR

Eqs. (30) and (31) indicate that the k-$\overline{\text{IPMR}}$ is a function of $\mathbb{E}_{\boldsymbol{X}}[\cdot]$, $\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}$, and $\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)$. Here, $\mathbb{E}_{\boldsymbol{X}}[\cdot]$ specifies the quadrature points and their weights; $\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}$ reflects the relative position of a quadrature point $\boldsymbol{x}$ to others; $\rho_{n+k}\left(\boldsymbol{x},\mathcal{X}_k^+\right)$ encodes the impact of adding a new batch $\mathcal{X}_k^+$. Hence, the former two can be viewed as fixed arguments, while the latter one is taken as a variable argument of k-$\overline{\text{IPMR}}$. In this way, it is feasible to unveil how the k-$\overline{\text{IPMR}}$ varies when $\mathcal{X}_k^+$ roams through the input space.

For illustration, let's consider a toy $\mathcal{G}$-function such that

$$Y = 3 - \frac{1}{\sqrt{2}}\left(X_1 + X_2\right) + \frac{1}{10}\left(X_1 - X_2\right)^2, \tag{E.1}$$

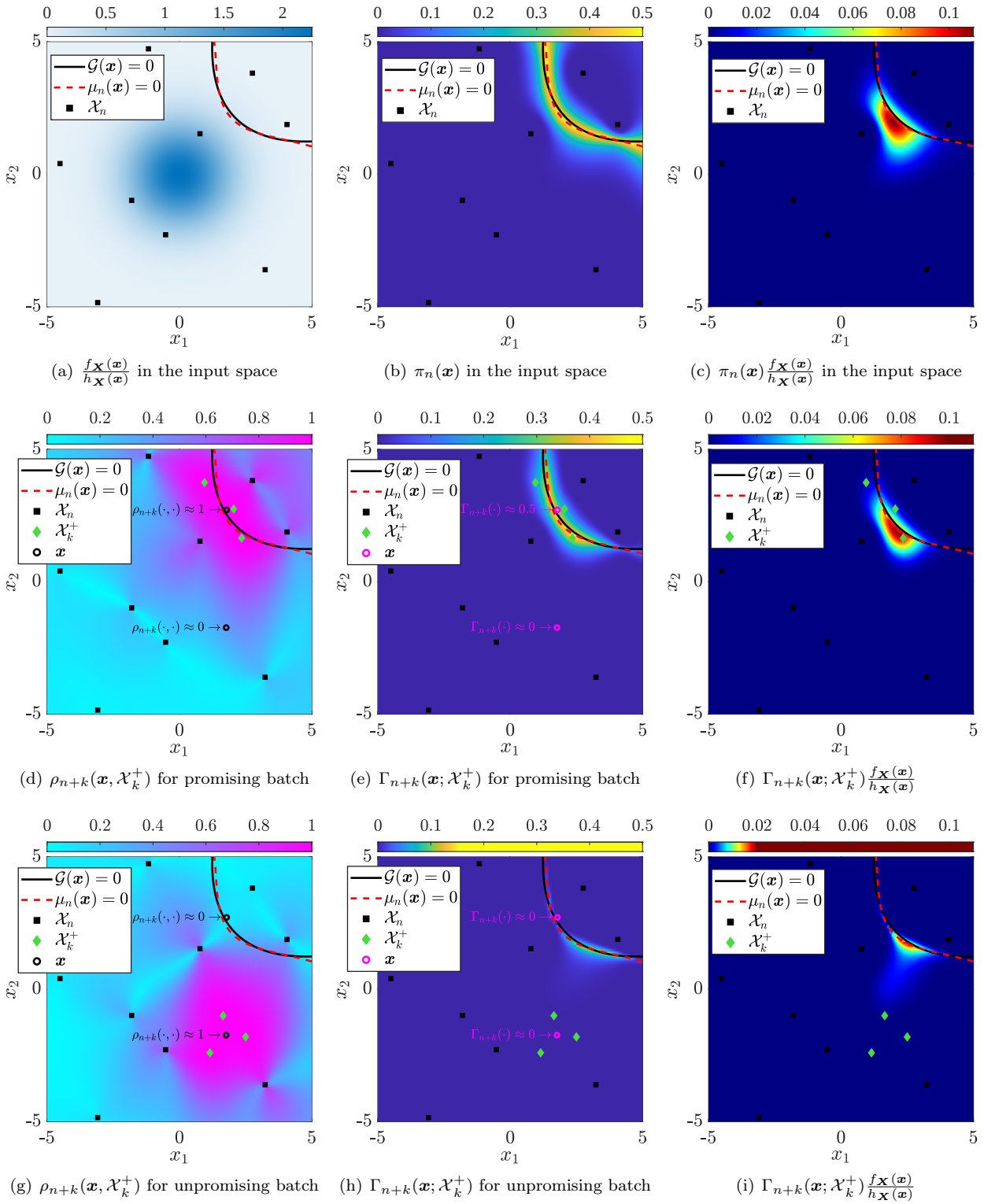
where $X_1, X_2 \sim \mathcal{N}(0,1)$. The Latin hypercube sampling technique is used to generate an ED $\mathcal{D}_n = \{\mathcal{X}_n, \mathcal{Y}_n\}$ of size $n = 9$, as marked as black squares in Fig. E.16. Then, the VAIS's weight $\frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{h_{\boldsymbol{X}}(\boldsymbol{x})}$ and the PM $\pi_n(\boldsymbol{x})$ are shown in Figs. E.16(a) and E.16(b), respectively. Further, the following two different cases of $\mathcal{X}_k^+$ are considered for comparison.

- The $\mathcal{X}_k^+$ is a batch of 3 new samples around $\left\{\boldsymbol{x} \in \mathbb{X} : \frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})} \to 0\right\}$, which are marked as the green diamonds in Figs. E.16(d) - (f).
- The $\mathcal{X}_k^+$ is a batch of 3 new samples far away from $\left\{\boldsymbol{x} \in \mathbb{X} : \frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})} \to 0\right\}$, which are plotted as the green diamonds in Figs. E.16(g) - (i).

In the former case, great values of $\rho_{n+k}(\cdot,\mathcal{X}_k^+)$ are observed around $\mathcal{X}_k^+$, i.e., the local region with small value of $\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}$ in Fig. E.16(d). Then, the resulting values of $\Gamma_{n+k}(\cdot;\mathcal{X}_k^+)$ are only significant in this area, but are nearly 0 in other regions, as exemplified by two magenta circles in Fig. E.16(e). Further, when considering the VAIS's weight, great values of $\Gamma_{n+k}(\boldsymbol{x};\mathcal{X}_k^+)\frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{h_{\boldsymbol{X}}(\boldsymbol{x})}$ are only shown in the local region with small $\frac{|\mu_n(\cdot)|}{\sigma_n(\cdot)}$ but great $\frac{f_{\boldsymbol{X}}(\cdot)}{h_{\boldsymbol{X}}(\cdot)}$, as shown in Fig. E.16(f).

In the latter case, significant values of $\rho_{n+k}(\cdot,\mathcal{X}_k^+)$ are only in the local region with great values of $\frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}$ in Fig. E.16(g), but the values of $\Gamma_{n+k}(\cdot;\mathcal{X}_k^+)$ are close to 0 in most of input space, with its maximum value only 0.1 in Fig. E.16(h). Further, when considering the VAIS's weight, the resulting values of $\Gamma_{n+k}(\boldsymbol{x};\mathcal{X}_k^+)\frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{h_{\boldsymbol{X}}(\boldsymbol{x})}$ are still very trivial, with its maximum value only 0.01 in Fig. E.16(i).

Finally, according to k-$\overline{\text{IPMR}}$, the $\mathcal{X}_k^+$ in the former case is more promising than that in the latter case. Obviously, the promising batch $\mathcal{X}_k^+$ is prone to be located in the local region featuring smaller $\frac{|\mu_n(\cdot)|}{\sigma_n(\cdot)}$ but greater $\frac{f_{\boldsymbol{X}}(\cdot)}{h_{\boldsymbol{X}}(\cdot)}$. In this way, the large-sized candidate pool $\mathcal{X}_{\mathrm{C}}$ can be pruned to some extent.



(a) $\frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{h_{\boldsymbol{X}}(\boldsymbol{x})}$ in the input space (b) $\pi_n(\boldsymbol{x})$ in the input space (c) $\pi_n(\boldsymbol{x})\frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{h_{\boldsymbol{X}}(\boldsymbol{x})}$ in the input space

(d) $\rho_{n+k}(\boldsymbol{x}, \mathcal{X}_k^+)$ for promising batch (e) $\Gamma_{n+k}(\boldsymbol{x}; \mathcal{X}_k^+)$ for promising batch (f) $\Gamma_{n+k}(\boldsymbol{x}; \mathcal{X}_k^+)\frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{h_{\boldsymbol{X}}(\boldsymbol{x})}$

(g) $\rho_{n+k}(\boldsymbol{x}, \mathcal{X}_k^+)$ for unpromising batch (h) $\Gamma_{n+k}(\boldsymbol{x}; \mathcal{X}_k^+)$ for unpromising batch (i) $\Gamma_{n+k}(\boldsymbol{x}; \mathcal{X}_k^+)\frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{h_{\boldsymbol{X}}(\boldsymbol{x})}$

Figure E.16: Illustration of a promising candidate batch informed by k-IPMR

Appendix F. Stepwise uncertainty reduction (SUR)

In the learning function SUR [29, 32], two different forms were developed and are called $\mathrm{SUR}^{(1)}(\cdot)$ and $\mathrm{SUR}^{(2)}(\cdot)$ here, respectively.

First, by quantifying the impact of adding $\mathcal{X}_k^+$ on the reduction of the posterior variance $\sigma^2_{\widehat{P}_{f,n}}$ (Eq. (9)), the $\mathrm{SUR}^{(1)}$ is eventually expressed as [32]

$$\mathrm{SUR}^{(1)}\left(\mathcal{X}_k^+\right) = \gamma_n - \iint_{\mathbb{X}\times\mathbb{X}} \Phi_2\left(\begin{bmatrix} -\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})} \\ -\frac{\mu_n(\boldsymbol{x}')}{\sigma_n(\boldsymbol{x}')} \end{bmatrix}; \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \omega_{n+k} \\ \omega_{n+k} & 1 \end{bmatrix}\right), \tag{F.1}$$

where $\gamma_n = \iint_{\mathbb{X}\times\mathbb{X}} \Phi_2\left(\begin{bmatrix} -\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})} \\ -\frac{\mu_n(\boldsymbol{x}')}{\sigma_n(\boldsymbol{x}')} \end{bmatrix}; \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_n(\boldsymbol{x},\boldsymbol{x}') \\ \rho_n(\boldsymbol{x},\boldsymbol{x}') & 1 \end{bmatrix}\right) f_{\boldsymbol{X}}(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}') \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{x}'$ is a quantity independent of $\mathcal{X}_k^+$; then, only $\omega_{n+k} = \dfrac{c_n\left(\boldsymbol{x}, \mathcal{X}_k^+\right)^\top \left(\mathcal{C}_k^+\right)^{-1} c_n\left(\boldsymbol{x}' \mathcal{X}_k^+\right)}{\sigma_n(\boldsymbol{x})\sigma_n(\boldsymbol{x}')}$ encodes the impact of adding $\mathcal{X}_k^+$ in $\mathrm{SUR}^{(1)}\left(\mathcal{X}_k^+\right)$.

In this way, the batch of k best next points $\mathcal{X}_k^*$ is selected as

$$\mathcal{X}_k^* = \underset{\mathcal{X}_k^+ \in \mathbb{X}}{\arg\min}\, \mathrm{SUR}^{(1)}\left(\mathcal{X}_k^+\right). \tag{F.2}$$

Since both $\sigma^2_{\widehat{P}_{f,n}}$ and $\mathrm{SUR}^{(1)}\left(\mathcal{X}_k^+\right)$ are expressed as double integrals of $\Phi_2(\cdot;\cdot,\cdot)$, they are too computationally demanding to use in practice.

To this end, the $\mathrm{SUR}^{(2)}(\cdot)$ is defined from the following residual uncertainty measure of failure probability

$$\mathrm{IBV}_n = \mathbb{E}_{\boldsymbol{X}}\left[\sigma^2_{\widehat{\mathbf{1}}_n}(\boldsymbol{x})\right] = \int_{\mathbb{X}} \Phi\left(-\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right) \Phi\left(-\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right) f_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}, \tag{F.3}$$

which is called the integrated Bernoulli variance and acts as the upper bound of $\sigma^2_{\widehat{P}_{f,n}}$, according to Cauchy-Schwarz inequality. Notably, the IBV_n is only a single integral of $\Phi(\cdot)$ and is much cheaper than $\sigma^2_{\widehat{P}_{f,n}}$.

Then, by measuring the potential reduction of IBV_n brought by $\mathcal{X}_k^+$, the $\mathrm{SUR}^{(2)}\left(\mathcal{X}_k^+\right)$ is finally expressed as [32]

$$\mathrm{SUR}^{(2)}\left(\mathcal{X}_k^+\right) = \int_{\mathbb{X}} \Phi_2\left(\begin{bmatrix} a(\boldsymbol{x}) \\ -a(\boldsymbol{x}) \end{bmatrix}; \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} c(\boldsymbol{x}) & 1-c(\boldsymbol{x}) \\ 1-c(\boldsymbol{x}) & c(\boldsymbol{x}) \end{bmatrix}\right) f_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}, \tag{F.4}$$

where $a(\boldsymbol{x}) = \frac{\mu_n(\boldsymbol{x})}{\sigma_{n+k}(\boldsymbol{x})}$ and $c(\boldsymbol{x}) = \frac{\sigma_n^2(\boldsymbol{x})}{\sigma_{n+k}^2(\boldsymbol{x})}$.

Then, the $\mathcal{X}_k^*$ is selected as

$$\mathcal{X}_k^* = \underset{\mathcal{X}_k^+ \in \mathbb{X}}{\arg\min}\, \mathrm{SUR}^{(2)}\left(\mathcal{X}_k^+\right). \tag{F.5}$$

Clearly, in comparison with $\mathrm{SUR}^{(1)}(\cdot)$, the $\mathrm{SUR}^{(2)}(\cdot)$ is only a single integral. However, due to the presence of $\Phi_2(\cdot;\cdot,\cdot)$, carrying out the minimization in Eq. (F.5) is still very expensive.

References

[1] B. Ellingwood, M. Maes, F. Michael Bartlett, A. T. Beck, C. Caprani, A. Der Kiureghian, L. Dueñas-Osorio, N. Galvão, R. Gilbert, J. Li, J. Matos, Y. Mori, I. Papaioannou, R. Parades, D. Straub, B. Sudret, Development of methods of structural reliability, Structural Safety 113 (2025). doi:10.1016/j.strusafe.2024.102474.

[2] A. M. Hasofer, N. C. Lind, Exact and invariant second-moment code format., ASCE Journal of Engineering Mechanics 100 (1) (1974) 111–121. doi:10.1061/JMCEA3.0001848.

[3] K. Breitung, Asymptotic approximations for multinormal integrals, Journal of Engineering Mechanics 110 (3) (1984) 357–366. doi:10.1061/(ASCE)0733-9399(1984)110:3(357).

[4] S. Engelund, R. Rackwitz, A benchmark study on importance sampling techniques in structural reliability, Structural Safety 12 (4) (1993) 255–276. doi:10.1016/0167-4730(93)90056-7.

[5] S.-K. Au, J. Beck, Estimation of small failure probabilities in high dimensions by subset simulation, Probabilistic Engineering Mechanics 16 (4) (2001) 263–277. doi:10.1016/S0266-8920(01)00019-4.

[6] I. Papaioannou, D. Straub, Combination line sampling for structural reliability analysis, Structural Safety 88 (2021). doi:10.1016/j.strusafe.2020.102025.

[7] Y.-G. Zhao, Z.-H. Lu, Structural reliability: approaches from perspectives of statistical moments, John Wiley & Sons, 2021.

[8] T. Zhou, Y. Peng, Adaptive bayesian quadrature based statistical moments estimation for structural reliability analysis, Reliability Engineering and System Safety 198 (2020). doi:10.1016/j.ress.2020.106902.

[9] J. Li, J. Chen, The principle of preservation of probability and the generalized density evolution equation, Structural Safety 30 (1) (2008) 65–77. doi:10.1016/j.strusafe.2006.08.001.

[10] T. Zhou, Y. Peng, A two-stage point selection strategy for probability density evolution method-based reliability analysis, Structural and Multidisciplinary Optimization 65 (5) (2022). doi:10.1007/s00158-022-03244-7.

[11] X. Li, G. Chen, H. Cui, D. Yang, Direct probability integral method for static and dynamic reliability analysis of structures with complicated performance functions, Computer Methods in Applied Mechanics and Engineering 374 (2021). doi:10.1016/j.cma.2020.113583.

[12] R. Teixeira, M. Nogal, A. O'Connor, Adaptive approaches in metamodel-based reliability analysis: A review, Structural Safety 89 (2021). doi:10.1016/j.strusafe.2020.102019.

[13] M. Moustapha, S. Marelli, B. Sudret, Active learning for structural reliability: Survey, general framework and benchmark, Structural Safety 96 (2022). doi:10.1016/j.strusafe.2021.102174.

[14] B. Bichon, M. Eldred, L. Swiler, S. Mahadevan, J. McFarland, Efficient global reliability analysis for nonlinear implicit performance functions, AIAA Journal 46 (10) (2008) 2459–2468. doi:10.2514/1.34321.

[15] B. Echard, N. Gayton, M. Lemaire, Ak-mcs: An active learning reliability method combining kriging and monte carlo simulation, Structural Safety 33 (2) (2011) 145–154. doi:10.1016/j.strusafe.2011.01.002.

[16] C. Lataniotis, S. Marelli, B. Sudret, The gaussian process modeling module in uqlab, Journal of Soft Computing in Civil Engineering 2 (3) (2018) 91–116. doi:10.22115/SCCE.2018.129323.1062.

[17] D. Wang, D. Zhang, Y. Meng, M. Yang, C. Meng, X. Han, Q. Li, Ak-hrn: An efficient adaptive kriging-based n-hypersphere rings method for structural reliability analysis, Computer Methods in Applied Mechanics and Engineering 414 (2023). doi:10.1016/j.cma.2023.116146.
[18] D. Xiu, G. Em Karniadakis, The wiener-askey polynomial chaos for stochastic differential equations, SIAM Journal on Scientific Computing 24 (2) (2003) 619–644. doi:10.1137/S1064827501387826.
[19] T. Zhou, Y. Peng, An active-learning reliability method based on support vector regression and cross validation, Computers and Structures 276 (2023). doi:10.1016/j.compstruc.2022.106943.
[20] T. Zhou, Y. Peng, J. Li, An efficient reliability method combining adaptive global metamodel and probability density evolution method, Mechanical Systems and Signal Processing 131 (2019) 592–616. doi:10.1016/j.ymssp.2019.06.009.
[21] D. Zhang, N. Zhang, N. Ye, J. Fang, X. Han, Hybrid learning algorithm of radial basis function networks for reliability analysis, IEEE Transactions on Reliability 70 (3) (2021) 887 – 900. doi:10.1109/TR.2020.3001232.
[22] T. Zhou, Y. Peng, Ensemble of metamodels-assisted probability density evolution method for structural reliability analysis, Reliability Engineering and System Safety 228 (2022). doi:10.1016/j.ress.2022.108778.
[23] G. Li, T. Wang, Z. Chen, J. He, X. Wang, X. Du, Rbik-ss: A parallel adaptive structural reliability analysis method for rare failure events, Reliability Engineering and System Safety 239 (2023). doi:10.1016/j.ress.2023.109513.
[24] T. Zhou, Y. Peng, A new active-learning function for adaptive polynomial-chaos kriging probability density evolution method, Applied Mathematical Modelling 106 (2022) 86–99. doi:10.1016/j.apm.2022.01.030.
[25] T. Zhou, Y. Peng, Reliability analysis using adaptive polynomial-chaos kriging and probability density evolution method, Reliability Engineering and System Safety 220 (2022). doi:10.1016/j.ress.2021.108283.
[26] V. Dubourg, B. Sudret, J.-M. Bourinet, Reliability-based design optimization using kriging surrogates and subset simulation, Structural and Multidisciplinary Optimization 44 (5) (2011) 673 – 690. doi:10.1007/s00158-011-0653-8.
[27] C. Chevalier, V. Picheny, D. Ginsbourger, Kriginv: An efficient and user-friendly implementation of batch-sequential inversion strategies based on kriging, Computational Statistics and Data Analysis 71 (2014) 1021 – 1034. doi:10.1016/j.csda.2013.03.008.
[28] R. Schöbi, B. Sudret, S. Marelli, Rare event estimation using polynomial-chaos kriging, ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering 3 (2) (2017). doi:10.1061/AJRUA6.0000870.
[29] J. Bect, D. Ginsbourger, L. Li, V. Picheny, E. Vazquez, Sequential design of computer experiments for the estimation of a probability of failure, Statistics and Computing 22 (3) (2012) 773–793. doi:10.1007/s11222-011-9241-4.
[30] C. Dang, P. Wei, M. Faes, M. Valdebenito, M. Beer, Parallel adaptive bayesian quadrature for rare event estimation, Reliability Engineering and System Safety 225 (2022). doi:10.1016/j.ress.2022.108621.
[31] C. Agrell, K. Dahl, Sequential bayesian optimal experimental design for structural reliability analysis, Statistics and Computing 31 (3) (2021). doi:10.1007/s11222-021-10000-2.
[32] C. Chevalier, D. Ginsbourger, J. Bect, E. Vazquez, V. Picheny, Y. Richet, Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set, Technometrics 56 (4) (2014) 455–465. doi:10.1080/00401706.2013.860918.
[33] C. Dang, M. A. Valdebenito, M. G. Faes, P. Wei, M. Beer, Structural reliability analysis: A bayesian perspective, Structural Safety 99 (2022). doi:10.1016/j.strusafe.2022.102259.
[34] C. Dang, M. G. Faes, M. A. Valdebenito, P. Wei, M. Beer, Partially bayesian active learning cubature for structural reliability analysis with extremely small failure probabilities, Computer Methods in Applied Mechanics and Engineering 422 (2024). doi:10.1016/j.cma.2024.116828.
[35] T. Zhou, T. Guo, Y. Dong, Y. Peng, Polynomial chaos kriging-based structural reliability analysis via the expected margin volume reduction, Computers and Structures 287 (2023). doi:10.1016/j.compstruc.2023.107117.
[36] M. Li, S. Shen, V. Barzegar, M. Sadoughi, C. Hu, S. Laflamme, Kriging-based reliability analysis considering predictive uncertainty reduction, Structural and Multidisciplinary Optimization 63 (6) (2021) 2721–2737. doi:10.1007/s00158-020-02831-w.
[37] P. Wei, Y. Zheng, J. Fu, Y. Xu, W. Gao, An expected integrated error reduction function for accelerating bayesian active learning of failure probability, Reliability Engineering and System Safety 231 (2023). doi:10.1016/j.ress.2022.108971.
[38] C. Peng, C. Chen, T. Guo, W. Xu, Ak-seur: An adaptive kriging-based learning function for structural reliability analysis through sample-based expected uncertainty reduction, Structural Safety 106 (2024). doi:10.1016/j.strusafe.2023.102384.
[39] T. Zhou, T. Guo, C. Dang, M. Beer, Bayesian reinforcement learning reliability analysis, Computer Methods in Applied Mechanics and Engineering 424 (2024). doi:10.1016/j.cma.2024.116902.
[40] T. Zhou, T. Guo, Y. Dong, F. Yang, D. M. Frangopol, Look-ahead active learning reliability analysis based on stepwise margin reduction, Reliability Engineering and System Safety 243 (2024). doi:10.1016/j.ress.2023.109830.
[41] C. Duhamel, C. Helbert, M. Munoz Zuniga, C. Prieur, D. Sinoquet, A sur version of the bichon criterion for excursion set estimation, Statistics and Computing 33 (2) (2023). doi:10.1007/s11222-023-10208-4.
[42] J. Zhang, A. A. Taflanidis, Adaptive kriging stochastic sampling and density approximation and its application to rare-event estimation, ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering 4 (3) (2018) 04018021.
[43] Z. Chen, G. Li, J. He, Z. Yang, J. Wang, A new parallel adaptive structural reliability analysis method based on importance sampling and k-medoids clustering, Reliability Engineering and System Safety 218 (2022). doi:10.1016/j.ress.2021.108124.
[44] Z. Wen, H. Pei, H. Liu, Z. Yue, A sequential kriging reliability analysis method with characteristics of adaptive sampling regions and parallelizability, Reliability Engineering and System Safety 153 (2016) 170 – 179. doi:10.1016/j.ress.2016.05.002.
[45] F. Yang, R. Kang, Q. Liu, C. Shen, R. Du, F. Zhang, A new active learning method for reliability analysis based on local optimization and adaptive parallelization strategy, Probabilistic Engineering Mechanics 75 (2024). doi:10.1016/j.probengmech.2023.103572.
[46] Y. Meng, D. Zhang, B. Shi, D. Wang, F. Wang, An active learning kriging model with approximating parallel strategy for structural reliability analysis, Reliability Engineering and System Safety 247 (2024). doi:10.1016/j.ress.2024.110098.
[47] D. Zhan, J. Qian, Y. Cheng, Pseudo expected improvement criterion for parallel ego algorithm, Journal of Global Optimization 68 (3) (2017) 641 – 662. doi:10.1007/s10898-016-0484-7.
[48] Z. Zhao, Z.-H. Lu, Y.-G. Zhao, P-ak-mcs: Parallel ak-mcs method for structural reliability analysis, Probabilistic Engineering Mechanics 75 (2024). doi:10.1016/j.probengmech.2023.103573.
[49] C. Dang, M. Beer, Semi-bayesian active learning quadrature for estimating extremely low failure probabilities, Reliability Engineering and System Safety 246 (2024). doi:10.1016/j.ress.2024.110052.
[50] F. A. C. Viana, R. T. Haftka, L. T. Watson, Efficient global optimization algorithm assisted by multiple surrogate techniques, Journal of Global Optimization 56 (2) (2013) 669 – 689. doi:10.1007/s10898-012-9892-5.
[51] C. Chevalier, D. Ginsbourger, X. Emery, Corrected kriging update formulae for batch-sequential data assimilation, Lecture Notes in Earth System Sciences 0 (2014) 119–122. doi:10.1007/978-3-642-32408-6_29.

[52] T. Zhou, X. Zhu, T. Guo, Y. Dong, M. Beer, Multi-point bayesian active learning reliability analysis, Structural Safety 114 (2025). doi:10.1016/j.strusafe.2024.102557.
[53] V. J. Romero, J. V. Burkardt, M. D. Gunzburger, J. S. Peterson, Comparison of pure and "latinized" centroidal voronoi tessellation against various other statistical sampling methods, Reliability Engineering and System Safety 91 (10-11) (2006) 1266 – 1280. doi:10.1016/j.ress.2005.11.023.
[54] J. Wang, Z. Cao, G. Xu, J. Yang, A. Kareem, An adaptive kriging method based on k-means clustering and sampling in n-ball for structural reliability analysis, Engineering Computations (Swansea, Wales) 40 (2) (2023) 378 – 410. doi:10.1108/EC-12-2021-0705.
[55] A. Hlobilova, S. Marelli, B. Sudret, Benchmark case datasets - truss model, Zenodo (2024). doi:https://doi.org/10.5281/zenodo.12699396.
[56] F. McKenna, M. Scott, G. Fenves, Nonlinear finite-element analysis software architecture using object composition, Journal of Computing in Civil Engineering 24 (1) (2010) 95–107. doi:10.1061/(ASCE)CP.1943-5487.0000002.
[57] G. Xu, T. Guo, A. qun Li, T. Zhou, C. Shuang, Seismic performance of steel frame structures with novel self-centering beams: Shaking-table tests and numerical analysis, Journal of Structural Engineering 151 (3) (2025) 04025002. doi:10.1061/JSENDH.STENG-13516.
[58] S. Rezaeian, A. Der Kiureghian, A stochastic ground motion model with separable temporal and spectral nonstationarities, Earthquake Engineering and Structural Dynamics 37 (13) (2008) 1565 – 1584. doi:10.1002/eqe.831.
[59] S. Rezaeian, A. D. Kiureghian, Simulation of synthetic ground motions for specified earthquake and site characteristics, Earthquake Engineering and Structural Dynamics 39 (10) (2010) 1155 – 1180. doi:10.1002/eqe.997.
[60] C. Nardin, O. S. Bursi, F. Paolacci, A. Pavese, G. Quinci, Experimental performance of a multi-storey braced frame structure with non-structural industrial components subjected to synthetic ground motions, Earthquake Engineering and Structural Dynamics 51 (9) (2022) 2113 – 2136. doi:10.1002/eqe.3656.
[61] D. Zhang, P. Zhou, C. Jiang, M. Yang, X. Han, Q. Li, A stochastic process discretization method combing active learning kriging model for efficient time-variant reliability analysis, Computer Methods in Applied Mechanics and Engineering 384 (2021). doi:10.1016/j.cma.2021.113990.
[62] T. Zhou, Y. Peng, Active learning and active subspace enhancement for pdem-based high-dimensional reliability analysis, Structural Safety 88 (2021). doi:10.1016/j.strusafe.2020.102026.
[63] T. Zhou, T. Guo, C. Dang, L. Jia, Y. Dong, Parallel active learning reliability analysis: A multi-point look-ahead paradigm, Computer Methods in Applied Mechanics and Engineering 434 (2025) 117524. doi:https://doi.org/10.1016/j.cma.2024.117524.
[64] Z. Drezner, Computation of the bivariate normal integral, Mathematics of Computation 32 (141) (1978) 277 – 279. doi:10.1090/S0025-5718-1978-0461849-9.