Sample regeneration algorithm for structural failure probability function estimation

Xiukai Yuan [a*], Shanglong Wang [a], Marcos A. Valdebenito [b], Matthias G.R. Faes [b], Michael Beer [c,d,e]

[a] *School of Aerospace Engineering, Xiamen University, Xiamen 361005, P. R. China*
[b] *Chair for Reliability Engineering, TU Dortmund University, Leonhard-Euler-Strasse 5, 44227 Dortmund, Germany.*
[c] *Institute for Risk and Reliability, Leibniz Universität Hannover, Callinstr. 34, Hannover, Germany*
[d] *Institute for Risk and Uncertainty, University of Liverpool, Peach Street, L69 7ZF Liverpool, United Kingdom*
[e] *International Joint Research Center for Resilient Infrastructure & International Joint Research Center for Engineering Reliability and Stochastic Mechanics, Tongji University, Shanghai 200092, China*

Abstract

An efficient strategy to approximate the failure probability function in structural reliability problems is proposed. The failure probability function (FPF) is defined as the failure probability of the structure expressed as a function of the design parameters, which in this study are considered to be distribution parameters of random variables representing uncertain model quantities. The task of determining the FPF is commonly numerically demanding since repeated reliability analyses are required. The proposed strategy is based on the concept of augmented reliability analysis, which only requires a single run of a simulation-based reliability method. This paper introduces a new sample regeneration algorithm that allows to generate the required failure samples of design parameters without any additional evaluation of the structural response. In this way, efficiency is further improved while ensuring high accuracy in the estimate of the FPF. To illustrate the efficiency and effectiveness of the method, case studies involving a turbine disk and an aircraft inner flap are included in this study.

Keywords: Reliability, Failure probability function, Regeneration algorithm, Bayesian theory, Maximum Entropy method

1. Introduction

Reliability-based analysis has become an appropriate and useful tool for structural design. Such an approach allows taking uncertainty explicitly into consideration, and it has been widely used and applied in many research areas [1, 2]. In reliability-based design optimization (RBDO), the failure probability (i.e., the complement of the reliability) of the target system under various design configurations usually needs to be evaluated. For example, the reliability constraint in a reliability-based design optimization problem usually requires many reliability analyses to be executed at various design values. The failure probability function (FPF) can be defined as the failure probability with respect to the set of design parameters, which can be expressed as follows [3]:

$$P_F(\boldsymbol{\theta}) = \int I_F(\boldsymbol{x}, \boldsymbol{\theta}_g) f\left(\boldsymbol{x}|\boldsymbol{\theta}_d\right) \mathrm{d}\boldsymbol{x} \tag{1}$$

where $\boldsymbol{\theta}$ is the vector of n_θ design parameters which can be decomposed as $\boldsymbol{\theta} = [\boldsymbol{\theta}_g; \boldsymbol{\theta}_d]$; $\boldsymbol{\theta}_g$ is the parameter vector that affects structural performance; $\boldsymbol{x}$ is the vector of basic random variables associated with the structural model; $\boldsymbol{\theta}_d$ is the parameter vector that affects the joint probability density function (PDF) $f(\boldsymbol{x}|\boldsymbol{\theta}_d)$ and usually contains distribution parameters such as mean value; and $I_F(\boldsymbol{x}, \boldsymbol{\theta}_g)$ is an indicator function which assumes the value 1 whenever a particular realization of the pair $(\boldsymbol{x}, \boldsymbol{\theta}_g)$ leads to an unacceptable structural behavior, otherwise $I_F(\boldsymbol{x}, \boldsymbol{\theta}_g) = 0$. If the FPF over the whole design space can be obtained beforehand, the RBDO problem can be transformed into an ordinary optimization problem, that is, it can be decoupled into a traditional optimization problem without the need for repeated reliability analyses [4]. However, analytic solutions for Eq. (1) are generally not available.

There is a vast number of contributions which address the failure probability estimation problem, such as first/second order reliability methods (FORM/SORM)[5, 6], Monte Carlo simulation [7], Importance Sampling [8, 9], Subset Simulation [10], and Line Sampling [11, 12] etc. However, in practical applications, it is difficult to obtain the failure probability as an explicit function of the design parameters $\boldsymbol{\theta}$ as it often demands an intractable computational cost. Even in case highly efficient reliability methods are used, the repeated evaluation of Eq. (1) for different realizations of $\boldsymbol{\theta}$ makes the calculation costly apart from near-trivial cases.

Various strategies for constructing an approximation of the FPF have been developed. One kind of strategy is to apply surrogate models to construct an approximation by selecting some predefined interpolation points in the space of the design parameters by means of an appropriate design-of-experiments (DOE) scheme. For example, Gasser [13] used a quadratic function with respect to the design parameters $\boldsymbol{\theta}$ to approximate the logarithm of FPF. The number of coefficients to be determined in such an approach is equal to $n_\theta + n_\theta\left(n_\theta + 1\right)/2$. As such, at least $n_\theta + n_\theta\left(n_\theta + 1\right)/2$ reliability analyses are needed to obtain all the coefficients of the quadratic surrogate model. Jensen [14] adopted a linear function to approximate the logarithm of FPF when handling a linear system subject to stochastic excitation, for which at least $n_\theta + 1$ reliability analyses are needed. There are also several other surrogate model methods that can be used to build the approximation of the FPF, for example, Kriging regression models [15, 16], Support Vector Machines [17, 18] or Polynomial Chaos Expansions [19, 20], which are widely used in reliability analysis to approximate the performance function [21, 22]. These techniques can also be used to build a surrogate model of the FPF, however, also in this case repeated evaluations of reliability are involved. The second kind of strategy is to solve FPF in an augmented space. Au [23] was the first to utilize Bayes' rule to include the solution of FPF in an augmented reliability problem, where the design parameters are treated as random variables with a predefined distribution. This allows to estimate the FPF in a single simulation run. Based on the augmented space idea, Ching [24, 25] adopted a pre-defined exponential functional form to approximate the FPF and applied the maximum entropy principle to estimate the corresponding coefficients. Taflanidis [26, 27, 28, 29] also adopted this idea and utilized Subset Simulation to solve the reliability-based optimization, sensitivity analysis and robust optimization. Also Liu [30] utilized the augmented space idea to carry out reliability-based design optimization, where the design space is partitioned iteratively for accurate approximation. The third strategy is a reweighting approach, which builds a local approximation of FPF based on the information of a single reliability analysis. This kind of method focuses on the problem where only the distribution parameter $\boldsymbol{\theta_d}$ is included in the analysis. In this context, Zou and Mahadevan [4] proposed a decoupling approach in which the FPF is expressed by a first-order Taylor series expansion based on the reliability sensitivity results. Yi et al. [31] proposed a new reliability optimization allocation for multifunction systems with multistate units based on goal-oriented (GO) methodology. Time-varying and high nonlinear performance brings a new challenge for the reliability-based robust design optimization. Thus, Yu et al. [32] proposed a multi-objective integrated framework for time-dependent reliability-based robust design optimization and the corresponding algorithms. Yuan [33] and Yuan and Lu [34] proposed a weighted approach, which expresses the FPF based on a set of samples which are generated in a single reliability analysis. Its efficiency depends on the simulation method used. Finally, Wei et al. [35, 36] also developed a non-intrusive imprecise stochastic simulation for uncertainty propagation that includes the approximation of FPF.

In this contribution, a sample regeneration (SR) scheme is proposed to further improve the efficiency of the 'augmented space' strategy for estimating the failure probability function. The proposed scheme is targeted to problems where the design variables correspond to distribution parameters $\boldsymbol{\theta_d}$. As such, it effectively combines the latter two classes of methods to approximate the FPF, as it encompasses an augmented space strategy with reweighting. The proposed approach first considers the problem in augmented space, where the estimation of the FPF is transformed to the calculation of an augmented failure probability, and a conditional PDF associated with the design parameters. Then, a regeneration strategy utilizing Bayesian theory is proposed in order to estimate the PDF of design parameters efficiently. Based on a few samples of the design parameters that cause system failure, the proposed strategy can generate more samples of the design parameters without requiring further structural analyses. Hence, it can efficiently generate an approximation of the FPF with a limited number of structural analysis. Note that the first author of this contribution recently proposed an alternative augmented space integral method for FPF estimation [37]. Although both the approach in [37] and the approach proposed in this work share the idea of augmented reliability, there are obvious differences between them. The augmented space integral method presented in [37] estimates the FPF by calculating a transformed integral. On the contrary, in this work, the FPF is calculated by means of sample regeneration, which is a novel concept that brings substantial advantages, as illustrated in the examples.

This paper is organized as follows. The definition of FPF problem of interest here and a brief review about the calculation of FPF based on augmented space are first provided in Section 2. Then, the proposed sample regeneration algorithm is developed in Section 3. At last, various examples are given to illustrate the performance of the proposed algorithm. The paper closes with conclusions.

2. Failure probability function and its calculation using an augmented space strategy

In this contribution, the failure probability function (FPF) that represents the functional relationship between the failure probability of a system and the distribution parameters of the basic random variables is considered. Mathematically, this is represented as:

$$P_F(\boldsymbol{\theta}) = \int I_F(\boldsymbol{x}) f\left(\boldsymbol{x}|\boldsymbol{\theta}\right) \mathrm{d}\boldsymbol{x} \tag{2}$$

where $\boldsymbol{\theta} = [\boldsymbol{\theta}_d] = [\theta_1, \ldots, \theta_{n_\theta}]$ indicates that the design parameters only refer to the distribution parameters of $\boldsymbol{x}$. In essence, this means that the design parameters represent, for example, the mean value or standard deviation of the random variable; $f(\boldsymbol{x}|\boldsymbol{\theta})$ is the probability density function (PDF) of the random variables $\boldsymbol{x}$, conditional on the distribution parameters $\boldsymbol{\theta}$, where it is assumed that the components of $\boldsymbol{x}$ are independent of each other; $I_F(\boldsymbol{x})$ is the indicator function, which is defined to be $I_F(\boldsymbol{x}) = 1$ if a system failure event F happens, and $I_F(\boldsymbol{x}) = 0$ otherwise. A system failure event is defined as $F = \{\boldsymbol{x} : g(\boldsymbol{x}) < 0\}$ where $g(\boldsymbol{x})$ is the performance function of the system under consideration. For simplicity, in the following, we just use $\boldsymbol{\theta}$ in place of $\boldsymbol{\theta}_d$.

The FPF can be expressed in an augmented space by virtue of Bayes' theorem, which allows to approximate the FPF by means of a single reliability analysis, and hence, avoid repeated reliability analyses. Following this approach, the design parameters are artificially considered as random variables with arbitrary probability density function $\varphi(\boldsymbol{\theta})$. In this context, the FPF $P_F(\boldsymbol{\theta})$ can be rewritten using Bayes' theorem as:

$$P_F(\boldsymbol{\theta}) = P(F|\boldsymbol{\theta}) = \frac{\varphi(\boldsymbol{\theta}|F)P(F)}{\varphi(\boldsymbol{\theta})} \tag{3}$$

where $\varphi(\boldsymbol{\theta})$ is the PDF associated with $\boldsymbol{\theta}$, whose support spans the associated feasible design space; $\varphi(\boldsymbol{\theta}|F)$ is the PDF of $\boldsymbol{\theta}$ conditional on the failure event, and $P(F)$ is the failure probability of the augmented reliability problem which is given by:

$$P(F) = \int\int I_F(\boldsymbol{x}) f(\boldsymbol{x}|\boldsymbol{\theta}) \varphi(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{x} \mathrm{d}\boldsymbol{\theta}. \tag{4}$$

According to the expression in Eq. (3), the FPF is represented by three components: $\varphi(\boldsymbol{\theta})$, $P(F)$ and $\varphi(\boldsymbol{\theta}|F)$. Among them, $\varphi(\boldsymbol{\theta})$ can be arbitrarily selected as long as its support spans the design space. It is important to note that in theory, different distributions for $\boldsymbol{\theta}$ do not affect the results of the FPF estimation since they purely serve as a tool to scan the design space. For example, either Normal or Uniform distributions can be considered [24]. However, one should be careful when using a Gaussian distribution, as it may assign negative values to quantities that are strictly positive due to physical reasons (e.g., plate thickness values). $P(F)$ can be estimated by using typically applied reliability analysis methods in augmented space, such as Monte Carlo Simulation or Subset Simulation [23].

Usually, the most challenging issue when estimating the FPF using an augmented space approach is the estimation of the conditional distribution $\varphi(\boldsymbol{\theta}|F)$. Generally, this conditional distribution cannot be derived analytically. Therefore, it is appealing to estimate this quantity using samples, which involves repeated evaluations of the performance function. This possibly entails a non-negligible numerical and computational cost, especially since typically many failure samples are required for an accurate estimation of $\varphi(\boldsymbol{\theta}|F)$. In [23, 24, 25], the failure samples are first selected and then $\varphi(\boldsymbol{\theta}|F)$ is estimated based on this set of samples by means of histograms or Maximum Entropy methods. However, the generation of sufficient samples in the failure domain can be challenging, especially when small failure probabilities are considered. In this contribution, a sample regeneration (SR) algorithm based on Bayesian theory is proposed to alleviate this problem, as it can generate more samples which belong to $\varphi(\boldsymbol{\theta}|F)$, based on just a few failure samples. Hence, no repeated evaluations of the performance function are needed, improving the efficiency of the estimation of the FPF significantly. This approach will be explained in detail in Section 3.

3. Proposed sample regeneration algorithm for efficient FPF estimation

3.1. Proposed sample regeneration algorithm

In Section 2, it is shown that the key issue for efficiently approximating the FPF is to obtain the conditional PDF of design parameter $\varphi(\boldsymbol{\theta}|F)$. However, this target PDF, $\varphi(\boldsymbol{\theta}|F)$, cannot be generally obtained in closed form as it is a posterior PDF conditional on the occurrence of failure. Therefore, it is usually approximated based on the generated samples of $\boldsymbol{\theta}$ in the failure domain using sample fitting methods, which is not trivial. Indeed, in order to generate failure samples of $\boldsymbol{\theta}$, it is usually required to generate samples $(\boldsymbol{x}, \boldsymbol{\theta})$ in the whole augmented space first and then, select the failure samples located in failure region F, which is typically performed using rejection sampling. That is, one has to generate the samples $(\boldsymbol{x}, \boldsymbol{\theta})$ according to the joint PDF $\varphi(\boldsymbol{x}, \boldsymbol{\theta}) = f(\boldsymbol{x}|\boldsymbol{\theta})\varphi(\boldsymbol{\theta})$, i.e., first get $\boldsymbol{\theta}$ sample from $\varphi(\boldsymbol{\theta})$, and then generate sample of $\boldsymbol{x}$ according to $f(\boldsymbol{x}|\boldsymbol{\theta})$. Then the failure samples of $(\boldsymbol{x}, \boldsymbol{\theta})$ are obtained by calculating the performance $g(\boldsymbol{x})$ for each sample and retaining the samples of $\boldsymbol{\theta}$ yielding a failure sample. This set of samples is distributed following $\varphi(\boldsymbol{\theta}|F)$. The simulation of samples can be performed using Monte Carlo simulation (MCS) or Subset Simulation in augmented space [10, 24]. The accuracy of the $\varphi(\boldsymbol{\theta}|F)$ estimate depends on the number of samples that are generated in the failure domain. In [24], a Markov chain simulation technique is used to generate additional samples of $(\boldsymbol{x}, \boldsymbol{\theta})$. However, this still involves the evaluation of the performance function, and thus entails a non-negligible numerical and computational cost.

In order to generate sufficient samples of $\boldsymbol{\theta}$ at reduced computational cost, a sample regeneration (SR) algorithm is introduced, which circumvents repeated evaluations of the performance function. The key of this proposed algorithm is to re-generate samples of $\boldsymbol{\theta}$ based on a set of pre-samples according to the utilizing of Bayesian theory.

When the PDF $\varphi(\boldsymbol{x}, \boldsymbol{\theta})$ is restricted to the failure region F instead of the augmented whole space, we have:

$$\varphi(\boldsymbol{x}, \boldsymbol{\theta}|F) = \varphi(\boldsymbol{\theta}|\boldsymbol{x}, F) f(\boldsymbol{x}|F) \tag{5}$$

where $f(\boldsymbol{x}|F)$ is the marginal PDF of $\boldsymbol{x}$ conditional on F and $\varphi(\boldsymbol{\theta}|\boldsymbol{x}, F)$ is a conditional PDF which can be further deduced based on Bayesian theory:

$$\varphi(\boldsymbol{\theta}|\boldsymbol{x}, F) = \frac{I_F(\boldsymbol{x})\varphi(\boldsymbol{\theta}|\boldsymbol{x})}{\int I_F(\boldsymbol{x})\varphi(\boldsymbol{\theta}|\boldsymbol{x})\mathrm{d}\boldsymbol{\theta}} = I_F(\boldsymbol{x})\varphi(\boldsymbol{\theta}|\boldsymbol{x}) \tag{6}$$

where $\varphi(\boldsymbol{\theta}|\boldsymbol{x})$ is the conditional PDF of $\boldsymbol{\theta}$ with respect $\boldsymbol{x}$.

Inspection of Eq. (6) reveals that in order to generate samples of $\boldsymbol{\theta}$ in the failure region F, we can just choose the failure sample of $\boldsymbol{x}$ which leads to $I_F(\boldsymbol{x}) = 1$ a priori, and then generate $\boldsymbol{\theta}$ according to $\varphi(\boldsymbol{\theta}|\boldsymbol{x})$. Meanwhile, according to Eq. (5), as the selected failure sample $\boldsymbol{x}$ is distributed as $f(\boldsymbol{x}|F)$, and then the re-generated sample $\boldsymbol{\theta}$ according to Eq. (6), then the united sample $(\boldsymbol{x}, \boldsymbol{\theta})$ will distributed as $\varphi(\boldsymbol{x}, \boldsymbol{\theta}|F)$. This means that the re-generated $\boldsymbol{\theta}$ in this way is also distributed as $\varphi(\boldsymbol{\theta}|F)$ which is a key term in FPF solution. Note that, instead of direct MCS or SS, Eq. (6) provides an efficient way to produce more samples of $\varphi(\boldsymbol{\theta}|F)$. As it is through re-sampling from conditional samples, it is called 'Sample Re-generation (SR)' algorithm in this paper. The advantage of the proposed SR is obvious, as it re-generate samples based on a few samples of $\boldsymbol{x}$, as such, no additional $(\boldsymbol{x}, \boldsymbol{\theta})$ samples are needed as a whole, and hence also no more evaluations of the performance function are required.

For illustration, the schematic diagram of the proposed Sample Regeneration algorithm is presented in Fig. 1. Let $(\boldsymbol{x}^{(j)}, \boldsymbol{\theta}^{(j)}) \in F$ be a failure sample in the augmented space shown by a dot in the figure. The corresponding $\boldsymbol{x}^{(j)}$ component is distributed as $f(\boldsymbol{x}|\boldsymbol{\theta}^{(j)}, F)$ which is denoted by an ellipse. Then, based on each pre-sample $(\boldsymbol{x}^{(j)}, \boldsymbol{\theta}^{(j)}) \in F$, more samples of $\boldsymbol{\theta} \in F$ can be generated by the proposed SR algorithm. Regarding the $\boldsymbol{\theta}^{(j)}$ part, the obtained $\varphi(\boldsymbol{\theta}|\boldsymbol{x}^{(j)}, F) = \varphi(\boldsymbol{\theta}|\boldsymbol{x}^{(j)})$(as $I_F(\boldsymbol{x}^{(j)}) = 1$) is used to generate a number of samples of $\boldsymbol{\theta}$, i.e., $\boldsymbol{\theta}_j^{(1)}, \ldots, \boldsymbol{\theta}_j^{(m)}, \ldots, \boldsymbol{\theta}_j^{(M)}$ (M is the number of re-sampling rounds or times based on a single sample of $\boldsymbol{x}^{(j)}$). Then, $\{\boldsymbol{\theta}_j^{(1)}, \ldots, \boldsymbol{\theta}_j^{(M)}\}$ are distributed as $\varphi(\boldsymbol{\theta}|\boldsymbol{x}^{(j)})$, Suppose that there is a generated set of samples in the failure region, say, $D = \{(\boldsymbol{x}^{(j)}, \boldsymbol{\theta}^{(j)}) : j = 1, \ldots, N_F\}$. If the $\boldsymbol{x}$ components $D_x = \{\boldsymbol{x}^{(j)} : j = 1, \ldots, N_F\}$ are selected, then a set of samples $D_\theta^{(R)} = \{\boldsymbol{\theta}_1^{(1)}, \ldots, \boldsymbol{\theta}_1^{(M)}, \ldots, \boldsymbol{\theta}_{N_F}^{(1)}, \ldots, \boldsymbol{\theta}_{N_F}^{(M)}\}$ can be re-generated based on D_x. According to Eq. (6), these samples are distributed as:

$$\hat{\varphi}\left(\boldsymbol{\theta}|D_x, F\right) = \frac{1}{N_F}\sum_{j=1}^{N_F} I\left(\boldsymbol{x}^{(j)}\right)\varphi\left(\boldsymbol{\theta}|\boldsymbol{x}^{(j)}\right) = \frac{1}{N_F}\sum_{j=1}^{N_F}\varphi\left(\boldsymbol{\theta}|\boldsymbol{x}^{(j)}\right) \tag{7}$$

While a single sample $\boldsymbol{x}^{(j)}$, is distributed as $f(\boldsymbol{x}|\boldsymbol{\theta}^{(j)}, F)$, a set of samples D_x are distributed as $f(\boldsymbol{x}|F)$ since D_x is collected from $D = \{(\boldsymbol{x}^{(j)}, \boldsymbol{\theta}^{(j)}) : j = 1, \ldots, N_F\}$. As such, we have:

$$E\left[\hat{\varphi}\left(\boldsymbol{\theta}|D_x, F\right)\right] = \frac{1}{N_F}\sum_{j=1}^{N_F} E\left[\varphi\left(\boldsymbol{\theta}|\boldsymbol{x}^{(j)}\right)\right] = \int I(\boldsymbol{x})\varphi(\boldsymbol{\theta}|\boldsymbol{x})f(\boldsymbol{x}|F)\mathrm{d}\boldsymbol{x} = \varphi(\boldsymbol{\theta}|F) \tag{8}$$

Inspection of Eq. (8) reveals that, $\hat{\varphi}(\boldsymbol{\theta}|D_x, F)$ is unbiased, that means that in case the size N_F of the set of samples is large enough, the samples re-generated by the proposed algorithm converge to the target distribution $\varphi(\boldsymbol{\theta})$.

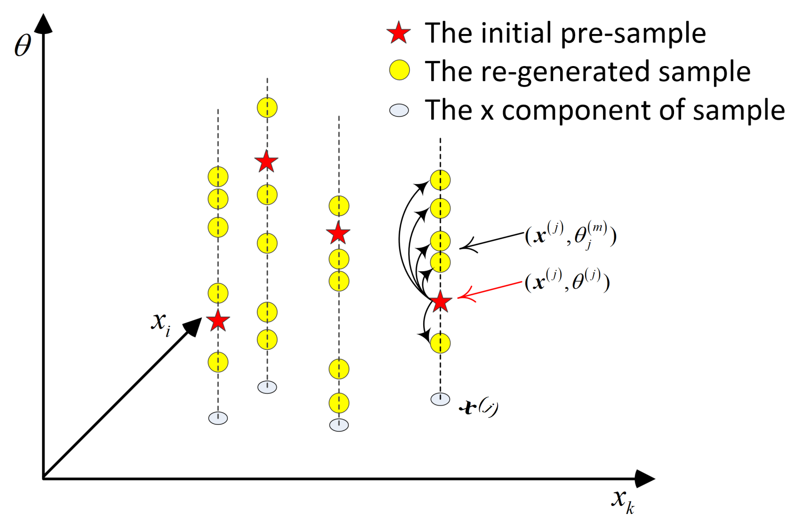


Figure 1: Schematic diagram of the proposed Sample Regeneration (SR) algorithm in augmented space (x_i, x_k, θ), where $(\boldsymbol{x}^{(j)}, \theta^{(j)}) = (x_i^{(j)}, x_k^{(j)}, \theta^{(j)})$ is an initial pre-sample, and $(\boldsymbol{x}^{(j)}, \theta_j^{(m)})$ is one of the regenerated samples (where $m = 1, \ldots, M$, for each $\boldsymbol{x}^{(j)}(j = 1, \ldots, N_F)$) through SR algorithm. The dotted line indicates that the re-generated samples by SR algorithm possess the same component $\boldsymbol{x}^{(j)}$.

3.2. Implementation strategy

3.2.1. Choice of prior distribution $\varphi(\boldsymbol{\theta})$

Note that the selection of appropriate distribution functions for $\boldsymbol{\theta}$ may not trivial. This subsection shows how to determine the $\varphi(\boldsymbol{\theta}|\boldsymbol{x})$ and generate samples from it.

For general cases, it is always possible to assume that θ_i is uniformly distributed over the design region, i.e., $\theta_i \sim U[\underline{\theta_i}, \bar{\theta}_i]$. Without particular preference for the region of the design parameters to be explored, a uniform distribution can be chosen for convenience and leads to appropriate estimates of the FPF [23]. Then, the posterior distribution $\varphi(\theta_i|x_i)$ can be obtained according to Bayesian theory as:

$$\varphi(\theta_i|x_i) = \frac{f(x_i|\theta_i)\varphi(\theta_i)}{f(x_i)} = \frac{1}{(\bar{\theta}_i - \underline{\theta}_i)f(x_i)} f(x_i|\theta_i) = \frac{1}{\Delta_i} f(x_i|\theta_i) \tag{9}$$

where $\Delta_i = (\bar{\theta}_i - \underline{\theta}_i)f(x_i)$ is a constant for given x_i. This quantity can be determined straightforwardly according to $f(x_i)$ or just determined by imposing the condition that the integral of density $\varphi(\theta_i|x_i)$ is equal to 1. Then we can re-generate the samples of θ_i according to Eq. (9). For simplicity, a Normal basic variable x_i is taken as an example to illustrate the proposed SR algorithm which is given in Appendix A.

It should be noted that there are several conjugate distribution families for certain distributions which can selected for $\varphi(\theta_i)$ in order to obtain a known PDF of $\varphi(\theta_i|x_i)$. A class of prior distribution is a conjugated family for certain distribution if the corresponding posterior distribution is in the same class. In general, for any sampling distribution, there is a natural family of prior distribution, i.e., the conjugated family [38]. Some typical conjugated families are listed in Table 1. Based on this information, we can choose the conjugate distribution for prior distribution $\varphi(\theta_i)$. In this way, the posterior distribution can be analytically obtained for $\varphi(\theta_i|x_i)$.

Table 1: Some typical conjugate distributions

Likelihood	Parameters	Prior distribution	Prior hyperparameters	Posterior hyperparameters
Normal with known σ^2	μ (mean)	Normal	μ_0, σ_0^2	$\left(\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{x}}{\sigma^2}\right) / \left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}\right), \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}$
Normal with known μ	σ^2(Variance)	Inverse gamma	a, p	$a + \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2, p + \frac{n}{2}$
Binomial	θ (probability of success)	Beta	p, q	$p + \sum_{i=1}^{n} x_i, q - \sum_{i=1}^{n} x_i + n$
Poisson	λ (rate)	Gamma	a, p	$a + n, p + \sum_{i=1}^{n} x_i$
Exponential	λ (rate)	Gamma	a, p	$a + \sum_{i=1}^{n} x_i, p + n$

3.2.2. Sampling from the posterior distribution

Once the expression of $\varphi(\boldsymbol{\theta}|\boldsymbol{x})$ is obtained, there are two ways to sample from it in order to re-generate the samples $D_\theta^{(R)} = \{\boldsymbol{\theta}_1^{(1)}, \ldots, \boldsymbol{\theta}_1^{(M)}, \ldots, \boldsymbol{\theta}_{N_F}^{(1)}, \ldots, \boldsymbol{\theta}_{N_F}^{(M)}\}$ based on the pre-samples D_x.

(1) **Direct sampling from** $\varphi(\boldsymbol{\theta}|\boldsymbol{x})$: If the posterior distribution $\varphi(\boldsymbol{\theta}|\boldsymbol{x})$ can be explicitly expressed, it usually can be used for sample generation. For example, suppose θ_i is the location parameter of the distribution $f(x_i|\theta_i)$, then according to Eq. (9), it has the same distribution kernel as that of x_i, even though it will be truncated. In this case, the conditional samples can be directly generated according to the explicit distribution function in Eq. (9).

(2) **Markov Chain Monte Carlo (MCMC) simulation by Metropolis-Hasting algorithm** : The Metropolis-Hasting algorithm [39] is a powerful tool to generate samples from a stochastic sequential process (Markov Chain) having the desired distribution as the stationary target distribution. This allows to generate samples from $\varphi(\boldsymbol{\theta}|\boldsymbol{x}, F)$ in case no closed-form solution exists.

In this contribution, MCMC is used to generate samples from the desired distribution $\varphi(\boldsymbol{\theta}|\boldsymbol{x}, F)$, as given in Eq. (6), since a closed-form solution is generally not available. Hereto, the stationary target distribution of Markov chain is selected as:

$$\pi(\boldsymbol{\theta}) = \varphi(\boldsymbol{\theta}|\boldsymbol{x}, F) = I_F(\boldsymbol{x})\varphi(\boldsymbol{\theta}|\boldsymbol{x}) \tag{10}$$

Then given a sample $\left(\boldsymbol{x}^{(j)}, \boldsymbol{\theta}^{(j)}\right)$ $(j = 1, \ldots, N_F)$, the Metropolis-Hasting algorithm computes the ratio r as:

$$r = \frac{\pi(\xi)}{\pi\left(\boldsymbol{\theta}_j^{(i)}\right)} = I_F\left(\boldsymbol{x}^{(j)}\right)\frac{\varphi\left(\boldsymbol{\xi}|\boldsymbol{x}^{(j)}\right)}{\varphi\left(\boldsymbol{\theta}_j^{(i)}|\boldsymbol{x}^{(j)}\right)} = \frac{\varphi\left(\boldsymbol{\xi}|\boldsymbol{x}^{(j)}\right)}{\varphi\left(\boldsymbol{\theta}_j^{(i)}|\boldsymbol{x}^{(j)}\right)} \tag{11}$$

where ξ is the candidate state, $\boldsymbol{\theta}_j^{(i)}$ is the i-th state of the Markov chain based on the initial point $\boldsymbol{\theta}_j^{(1)} = \boldsymbol{\theta}^{(j)}$. Substitution of Eq. (9) into Eq. (11) yields:

$$r = \frac{\varphi\left(\boldsymbol{x}^{(j)}|\boldsymbol{\xi}\right)}{\varphi\left(\boldsymbol{x}^{(j)}|\boldsymbol{\theta}_j^{(i)}\right)} \tag{12}$$

If $r > 1$, then ξ is accepted as the next state, otherwise ξ is accepted as the next state with probability r and $\boldsymbol{\theta}_j^{(i)}$ is accepted as the $(i{+}1)$-th state with the remaining probability $1 - r$, i.e., $\boldsymbol{\theta}_j^{(i+1)} = \boldsymbol{\theta}_j^{(i)}$, where a repeated state is obtained. Note that no burn-in issues are present in this case since the Markov chain starts with a given point $\boldsymbol{\theta}_j^{(1)} = \boldsymbol{\theta}^{(j)}$ that is distributed as the target distribution $\varphi(\boldsymbol{\theta}|\boldsymbol{x}, F)$. More details on Markov Chain Monte Carlo can be found in [39] and [10].

3.2.3. Maximum Entropy method for distribution fitting

After the samples $D_\theta^{(R)}$ have been regenerated according to the procedure described previously, the posterior distribution $\varphi(\boldsymbol{\theta}|F)$ can be estimated. In this work, the maximum entropy method is adopted, which is briefly described in the following [24]. For simplicity, consider a one dimensional parameter vector as example. The entropy of the corresponding PDF $f(\boldsymbol{\theta})$ is given by:

$$H = -\int_{-\infty}^{+\infty} f(\boldsymbol{\theta})\ln[f(\boldsymbol{\theta})]\mathrm{d}\boldsymbol{\theta} \tag{13}$$

The maximum entropy method for estimating $f(\boldsymbol{\theta})$ is stated as follows:

$$\begin{aligned} \text{Max} \quad & H = -\int_{-\infty}^{+\infty} f(\boldsymbol{\theta}) \ln[f(\boldsymbol{\theta})] \mathrm{d}\boldsymbol{\theta} \\ \text{s.t.} \quad & \int_{-\infty}^{+\infty} f(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} = 1 \\ & \int_{-\infty}^{+\infty} \theta_i^k f(\boldsymbol{\theta}) \mathrm{d}\theta_i = \mu_{\theta_i}^{(k)} \quad k = 1, 2, 3 \ldots \\ & \int_{-\infty}^{+\infty} \theta_i \theta_j f(\boldsymbol{\theta}) \mathrm{d}\theta_i \mathrm{d}\theta_j = \mu_{ij} \quad i, j = 1, 2, \ldots, n_\theta \end{aligned} \tag{14}$$

where $\mu_i^{(k)}$ is the sample mean of θ_i^k and μ_{ij} is the sample mean of $\theta_i \theta_j$; Solving Eq. (14), we can obtain the PDF $f(\boldsymbol{\theta})$ estimator as

$$\hat{f}(\boldsymbol{\theta}) = \exp\left[b_0 + \sum_{i=1}^{n_\theta} b_i \theta_i + \sum_{i \geq j = 1}^{n_\theta} b_{ij} \theta_i \theta_j + \cdots \right] \tag{15}$$

The details of Maximum Entropy method can be found in e.g., [40]. Note that in case one can consider different orders for the maximum entropy approximation in Eq. (15), e.g. first-order, second-order and third-order, which correspond to $k = 1, 2$ and 3 in Eq. (14), respectively.

In order to obtain a robust PDF approximation of $\varphi(\boldsymbol{\theta}|F)$, the samples are first transformed according to:

$$\boldsymbol{u} = T(\boldsymbol{\theta}) = \frac{\boldsymbol{\theta} - \underline{\boldsymbol{\theta}}}{\bar{\boldsymbol{\theta}} - \underline{\boldsymbol{\theta}}} \tag{16}$$

where $\boldsymbol{\theta} \in [\underline{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}}]$. using the Maximum Entropy method, the estimated PDF for these transformed samples can be obtained as:

$$\hat{f}_u(\boldsymbol{u}) = \exp\left[b_0 + \sum_{i=1}^{n_\theta} b_i u_i + \sum_{i \geq j = 1}^{n_\theta} b_{ij} u_i u_j + \cdots \right] \tag{17}$$

Then the estimate of $\varphi(\boldsymbol{\theta}|F)$ is obtained by transforming Eq. (17) to the original space of $\boldsymbol{\theta}$ by:

$$\begin{aligned} \hat{\varphi}(\boldsymbol{\theta}|F) &= \hat{f}_u\left[T(\boldsymbol{\theta})\right] \cdot |T(\boldsymbol{\theta})|_J \\ &= \exp\left[b_0 + \sum_{i=1}^{n_\theta} b_i \left(\frac{\theta_i - \underline{\theta}_i}{\bar{\theta}_i - \underline{\theta}_i} \right) + \sum_{i \geq j = 1}^{n_\theta} b_i b_j \left(\frac{\theta_i - \underline{\theta}_i}{\bar{\theta}_i - \underline{\theta}_i} \right) \left(\frac{\theta_j - \underline{\theta}_j}{\bar{\theta}_j - \underline{\theta}_j} \right) + \cdots \right] \cdot \prod_i^{n_\theta} \frac{1}{\bar{\theta}_i - \underline{\theta}_i} \end{aligned} \tag{18}$$

where $|\cdot|_J$ mean the Jacobian function determinant.

3.2.4. Estimation of FPF

Once the PDF $\varphi(\boldsymbol{\theta}|F)$ is estimated by the Maximum Entropy method given in Eq. (18), and in case that $\varphi(\boldsymbol{\theta})$ is selected to be a uniform distribution over $[\underline{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}}]$, then the FPF in Eq. (3) can be estimated by:

$$
\begin{aligned}
\hat{P}_F(\boldsymbol{\theta}) &= \frac{\hat{P}(F)}{\varphi(\boldsymbol{\theta})}\hat{\varphi}(\boldsymbol{\theta}|F) \\
&= \hat{P}(F)\exp\left[b_0 + \sum_{i=1}^{n_\theta} b_i\left(\frac{\theta_i - \underline{\theta}_i}{\bar{\theta}_i - \underline{\theta}_i}\right) + \sum_{i=1}^{n_\theta} b_i b_j \left(\frac{\theta_i - \underline{\theta}_i}{\bar{\theta}_i - \underline{\theta}_i}\right)\left(\frac{\theta_j - \underline{\theta}_j}{\bar{\theta}_j - \underline{\theta}_j}\right) + \cdots\right]
\end{aligned}
\tag{19}
$$

where $\hat{P}(F)$ is the estimator of the failure probability $P(F)$. In conclusion, it can as such be seen that the procedure yields an explicit expression of the FPF. Note that it is obtained by solving a single reliability problem and, most notably, with only a few failure samples due to the proposed SR algorithm.

3.3. Procedure of the proposed strategy

The procedure of the proposed strategy is summarized as follows, which is also presented schematically in Fig. 2:

(1) Select the prior distribution $\varphi(\boldsymbol{\theta})$.

If no additional information is available, a feasible choice is a uniform distribution within the support of the design parameters.

(2) Reliability analysis in augmented space.

Simulation-based reliability analysis (Monte Carlo Simulation or Subset Simulation) is carried out in augmented space. Then, the augmented failure probability $P(F)$ can be calculated and the failure samples $D_x = \{\boldsymbol{x}^{(j)} : j = 1, \ldots, N_F\}$ are obtained as pre-samples.

(3) Re-generate samples by using the proposed algorithm.

Based on the pre-samples set D_x, the proposed sample re-generation algorithm is used to generate more samples of $\boldsymbol{\theta}$, which are denoted as $D_\theta^{(R)}$.

(4) Estimate $\varphi(\boldsymbol{\theta}|F)$ using Maximum Entropy method.

Apply the Maximum Entropy method to obtain the estimator of $\varphi(\boldsymbol{\theta}|F)$ according to Eq. (18) based on the regenerated samples $D_\theta^{(R)}$.

(5) Obtain the FPF estimate.

After the conditional distribution $\varphi(\boldsymbol{\theta}|F)$ and augmented failure probability $P(F)$ are estimated, the FPF can be obtained by Eq. (19).

It should be stressed again that there is very little numerical cost involved in the proposed algorithm, as it does not involve any additional evaluation of the performance function (that is, no additional structural analyses) besides those required for solving the augmented reliability problem. Indeed, the approach only demands relatively simple numerical computations related to sample regeneration. Note that in the first step of the algorithm, the uniform distribution or the conjugated distribution can be selected.

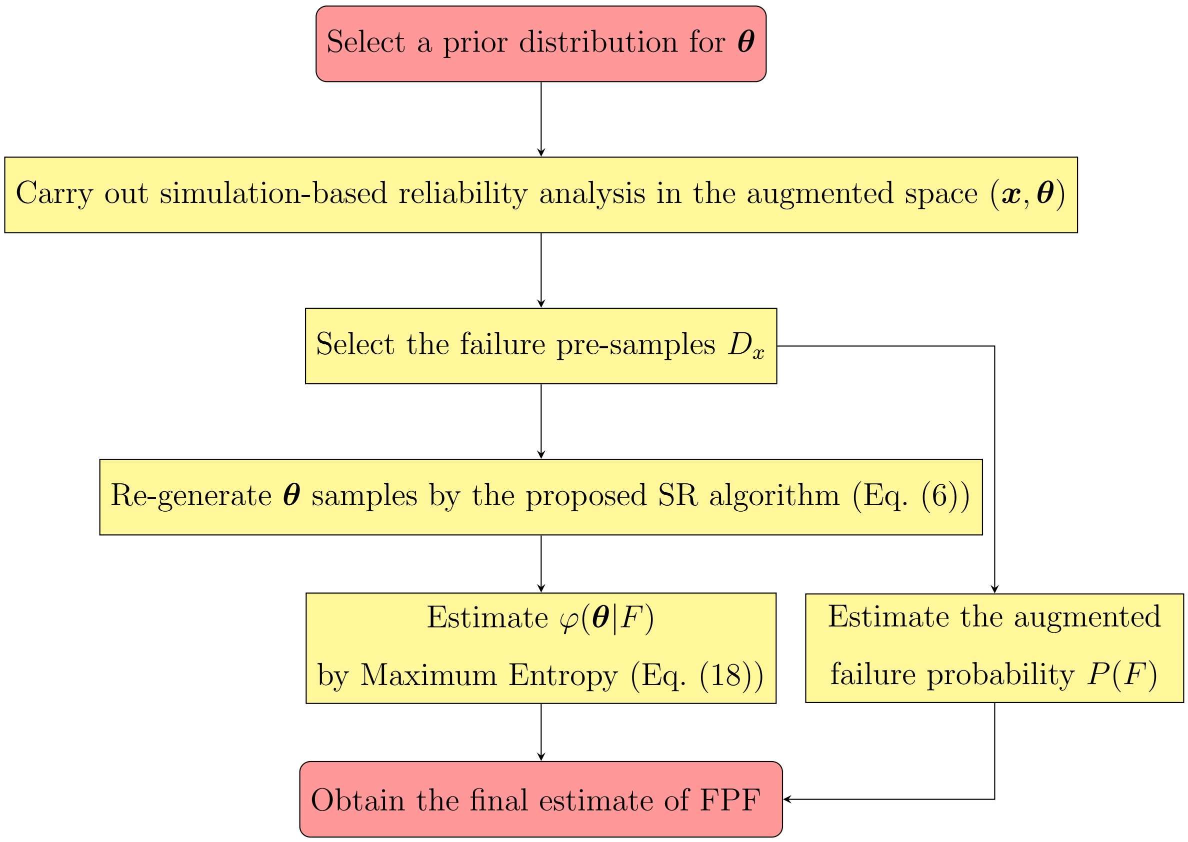


Figure 2: The procedure of the proposed approach for FPF estimation

4. Examples

In order to illustrate the effectiveness and accuracy of the proposed method, numerical and practical engineering examples are given. Meanwhile, various methods are also used for comparing the performance of the proposed approach; more precisely Direct Monte Carlo method and several surrogate modelling approaches. In the following, 'First order exponential RSM' refers to the method in which the FPF is constructed by the first order exponential response surface method [14]. 'Second order exponential RSM' refers to the method where the FPF is constructed by second order exponential response surface method [13]. 'Kriging' refers to the method the FPF is constructed by Kriging model. 'WMCS' refers to the 'Weighted Monte Carlo simulation' method [33]; 'WIS' refers to the 'Weighted Importance sampling' method [33], 'WSS' refers to the 'Weighted Subset simulation' method [33]. In the PDF fitting by the Maximum Entropy method given in Eq. (14), $k = 3$ is set for one design parameter case (Examples 1, 3, and 4), and $k = 2$ is set for two design parameters case in Example 2. These settings are determined after numerical tests which show that they lead to a good performance on accuracy.

4.1. Example 1: test example

A simple two-dimensional example is firstly presented to illustrate the proposed SR algorithm. In this example, the performance function is given by $g(\boldsymbol{x}) = 2.3 - x_1 - x_2$, where x_1 and x_2 are independent variables and the distribution information is given in Table 2. The design parameter θ is the mean value of variable x_1, and the design region is $\theta \in [0.5, 1.5]$.

Table 2: The distribution information of variables (Example 1)

Random variable	Mean	C.o.v.	Distribution
x_1	$\theta \in [0.5, 1.5]$	0.1	Gumbel
x_2	1	0.1	Normal

First suppose that θ is uniformly distributed over $[0.5, 1.5]$, MCS simulation is adopted to generate 10^4 samples of (x_1, x_2, θ) in the augmented space, and a total of $N_f = 1654$ failure samples (x_1, x_2, θ) are obtained. In order to show the effectiveness and advantage, the proposed algorithm does not carry out a new reliability analyses. Instead, for the sake of comparison, only a number of $N_{\text{pre}} = 50$ failure samples are selected from these $N_f = 1654$ failure samples. These selected samples are taken as the pre-samples of the proposed algorithm. Finally, $N_r = 1000$ samples of θ are re-generated based on the $\boldsymbol{x}$ part of the initial failure pre-samples through the proposed SR algorithm. It is emphasized that no additional evaluation of the performance function is involved in this step.

In order to show the performance of the proposed SR algorithm, the empirical CDFs of different sample sets are plotted Fig. 3. This figure shows the results obtained by the MCS samples(1654 failure samples); the original samples (50 samples) and the regenerated samples (1000 samples by the proposed SR algorithm (Metropolis algorithm)). The empirical CDF of the MCS failure samples (1654 samples) here is regarded as the 'exact' value. It can be seen from Fig. 3 that the empirical CDF of original samples is not smooth and accurate enough since the number of samples is too small. On the contrary, the empirical CDF result obtained by the proposed SR algorithm based on the same original samples is quite consistent with the 'exact' result. This shows that the

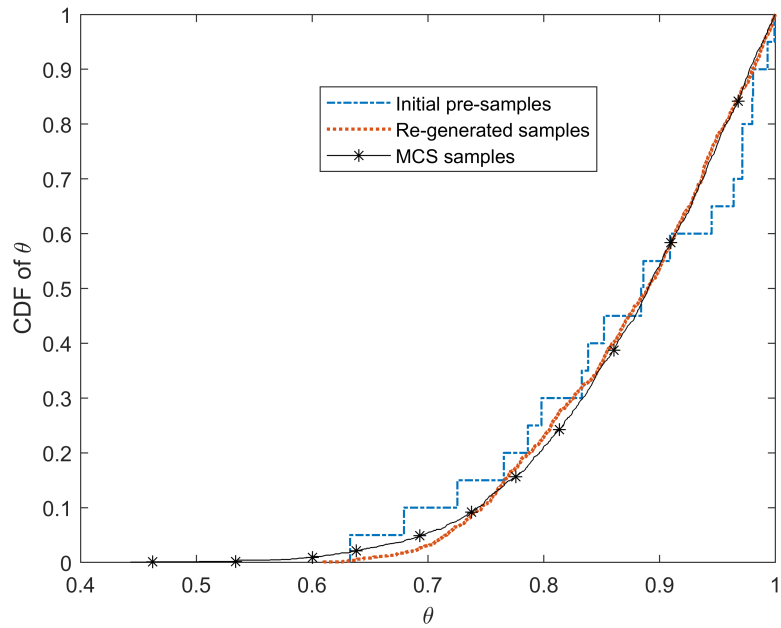


Figure 3: The empirical CDFs for the initial pre-samples and the re-generated samples by proposed SR algorithm (Example 1)

proposed algorithm is capable of determining the CDF of $\varphi(\theta|F)$ accurately based on a small set of samples. Hence, it seems that the proposed algorithm can extract more information of θ based on the pre-samples, which results in more accurate CDF estimate.

The FPF results obtained by different methods are shown in Fig. 4. In this figure, "Initial pre-samples" refers to the FPF result obtained based on Bayesian transformation in Eq. (3) and the CDF $\varphi(\theta|F)$ is estimated from the 50 failure pre-samples which is estimated by Maximum Entropy method. The proposed method is based on just 50 failure pre-samples, then we regenerate $N_r = 1000$ samples in order to estimate the CDF using the Maximum Entropy method, and at last the FPF obtained. The expression of FPF by the proposed SR algorithm (Metropolis algorithm) is given by

$$\hat{P}_F(\theta) = 0.1654 \exp\left[-17.2578 + 32.0817\left(\frac{\theta-0.5}{1}\right) - 5.1246\left(\frac{\theta-0.5}{1}\right)^2 - 8.2598\left(\frac{\theta-0.5}{1}\right)^3\right] \tag{20}$$

"MCS samples" refers to the FPF result obtained based on 1654 failure samples. Weighted IS (WIS) uses 1000 (excluding the design point solving cost) and Direct Monte Carlo simulation (denoted as "Direct MCS") is used to obtain the point-by-point failure probability values. Each failure probability point is calculated by one run of MCS with 10^6 samples, and these values are

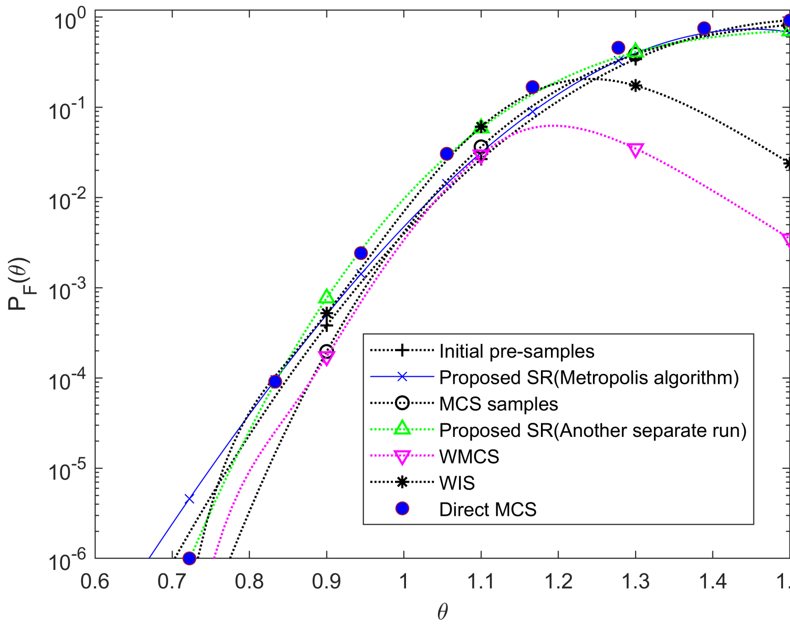


Figure 4: The results of FPF obtained by different methods (Example 1)

regarded as the exact FPF values. As shown in the Fig. 4, the FPF obtained by the proposed method has reached a high accuracy with 50 initial pre-samples and 20 regeneration rounds. The other methods cannot obtain satisfactory results, where a considerable error exists especially in the result based on only the initial pre-samples.

The comparison of different methods is also listed in Table 3. It can be seen that the proposed method only needs one reliability analysis, and the most important feature is that it only uses about 50 failure points, as few as possible, to obtain the FPF, which also means it will need less samples in one reliability analysis. The proposed method has obvious advantages in terms of computational efficiency.

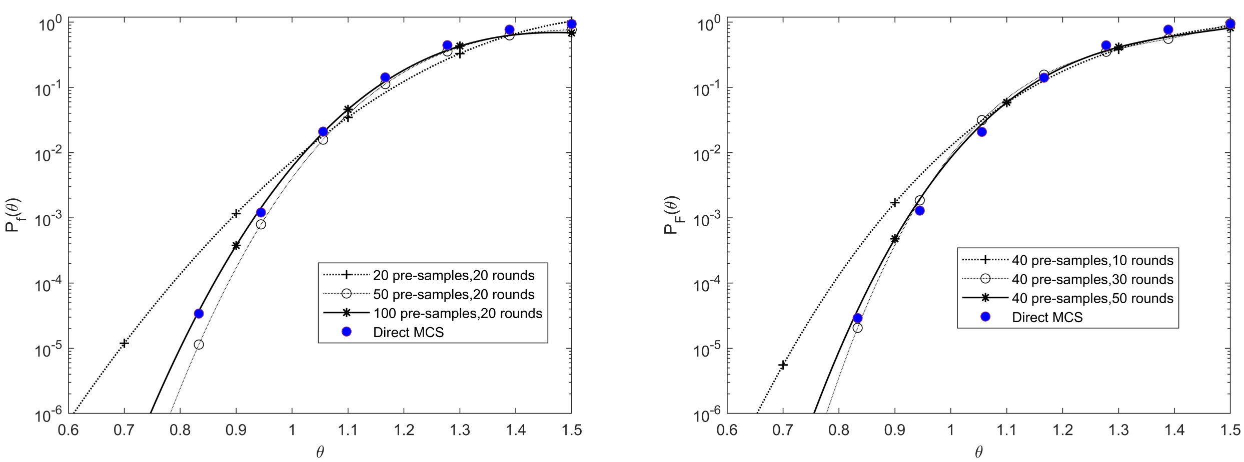
In addition, it is found that the accuracy of the FPF estimate obtained by the proposed method will be related with the number of initial pre-samples and regeneration rounds. Figure 5(a) shows the obtained FPF results when the number of initial pre-samples is 20, 50 and 100, respectively (the number of regeneration rounds is 50 for all cases). Definitely, the larger the number of initial pre-samples is, the higher the accuracy of the estimate is. However, more failure samples involve more computational cost. Meanwhile, increasing the number of regeneration rounds M will benefit the improvement of the approximation of CDF. However, the improvement has a limit, which means that when the number of re-generation rounds is already large enough, there is little improvement despite of increasing the number of the re-generation rounds, as the information that

Table 3: Computational cost of different methods (Example 1)

Methods	No. of samples	No. of failure samples
Initial pre-samples	–*	50
Proposed SR(Metropolis algorithm)	–	50
Proposed SR(Another separate run)	300	46
MCS samples	10^4	1654
WMCS	10^5	974
WIS	1000	515
Direct MCS	10^6	–

* Number of samples is not provided as the set of pre-samples is partly selected from MCS samples.

a certain number of pre-samples can provide is limited.



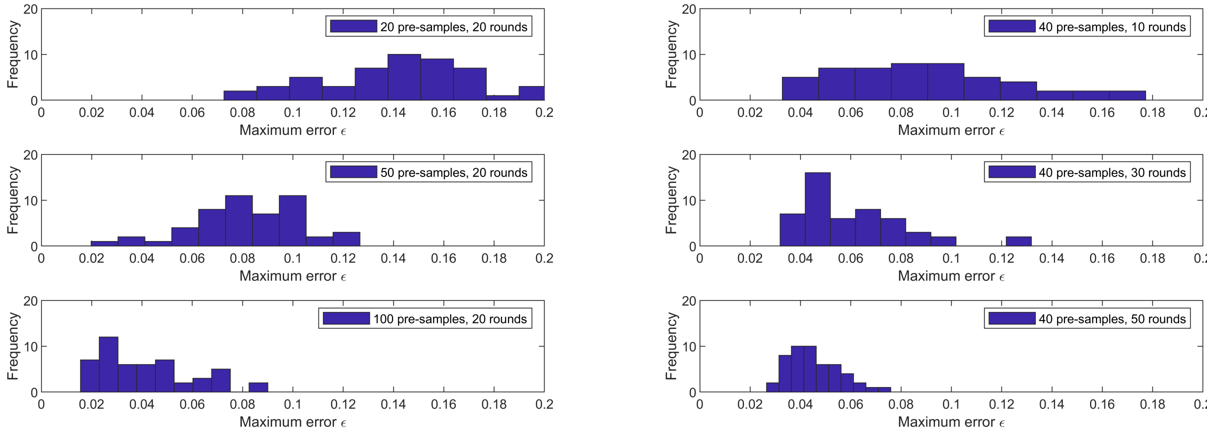
(a) Different pre-samples

(b) Different rounds

Figure 5: The FPF results obtained by different number of pre-samples and regeneration rounds (Example 1).

Fig. 5(b) shows the obtained FPF results when the numbers of regeneration rounds are 10, 30 and 50, respectively. The number of initial pre-samples is the same, $N_{\text{pre}} = 40$. It is shown that the results of the FPF are very consistent with the exact values in 30 or 50 rounds.

In order to see more clearly, Fig. 6 depicts histograms of the error of FPF results for the different number of pre-samples and rounds. The error ϵ is the maximum value of the differences between the obtained FPF and the exact value obtained by direct MCS. In order to eliminate the randomness, the simulation of the proposed algorithm is carried out 50 times for each case, and then the histogram is generated based on these data. The same conclusion can be drawn from



(a) Different pre-samples (b) Different rounds

Figure 6: The error of the FPF results obtained by different number of pre-samples and regeneration rounds (Example 1).

the figure that, as the number of pre-samples and re-generation rounds grows, the error becomes small.

4.2. Example 2: front axle

Front axle is an important component of an automobile that bears heavy loads [37]. An I-beam is often used in the design of front axle due to its high bend strength and light weight. As shown in Fig. 7, a critical component of the axle is located in the I-beam part with geometry variables a, b, t and h. To test the static strength of the front axle, the performance function can be expressed as

$$g(\boldsymbol{x}) = \sigma_s - \sqrt{\sigma^2(\boldsymbol{x}) + 3\tau^2(\boldsymbol{x})} \tag{21}$$

where $\boldsymbol{x} = [a, b, t, h, M, T]$ is the vector of random variables; σ_s is the limit-state stress associated with yielding, according to the material property of the front axle, the limit stress of yielding σ_s is 680MPa; the maximum normal stress and shear stress are $\sigma(\boldsymbol{x}) = M/W_x(\boldsymbol{x})$ and $\tau(\boldsymbol{x}) = T/W_\rho(\boldsymbol{x})$, where M and T are bending moment and torque, respectively, W_x and W_ρ are section factor and polar section factor, respectively, which are given as

$$W_x(\boldsymbol{x}) = \frac{a(h-2t)^3}{6h} + \frac{b}{6h}\left[h^3 - (h-2t)^3\right] \tag{22}$$

$$W_\rho(\boldsymbol{x}) = 0.8bt^2 + 0.4\left[a^3(h-2t)/t\right] \tag{23}$$

All variables are modeled as independent random variables with distribution parameters listed in Table 4. Note that all the variables are restricted to positive value dues to physical reasons,

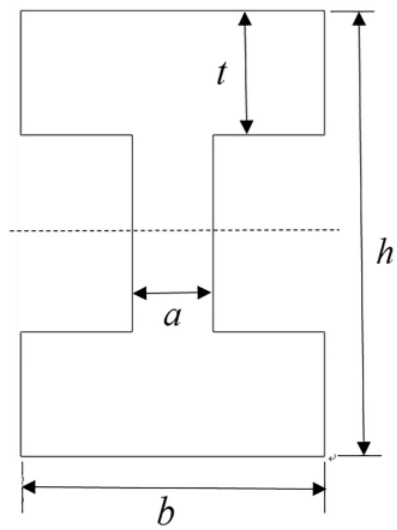


Figure 7: Diagram of automobile front axle

actually they are all truncated variables. The design parameters given in Table 4 include the mean values of the normal variables, and the design domains are $\theta_1 = \mu_a \in [10, 16]$mm and $\theta_2 = \mu_t \in [12, 18]$mm respectively.

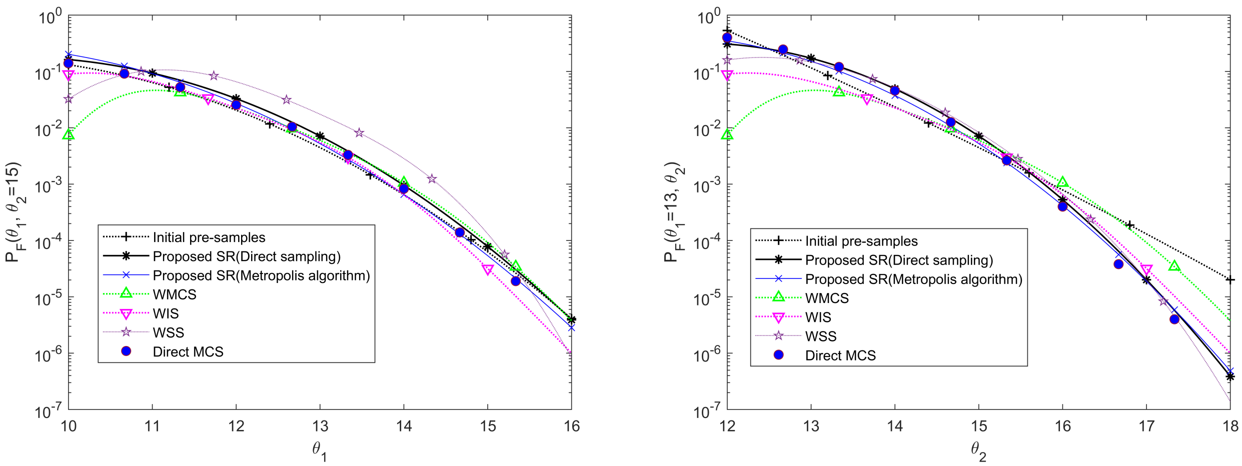
Table 4: The distribution information of the random variables of the front axle (Example 2)

Random variable	Mean	C.o.v.	Distribution
a (mm)	$\theta_1 = \mu_a$	0.05	Normal
t (mm)	$\theta_2 = \mu_t$	0.05	Normal
b (mm)	65	0.05	Normal
h (mm)	85	0.05	Normal
M (kN $\cdot$ m)	3.5	0.05	Normal
T (kN $\cdot$ m)	3.1	0.05	Normal

The proposed SR method and other different methods (WMCS, WIS and WSS) are applied to obtain the two-dimensional FPF. MCS simulation is first adopted to generate 500 samples in augmented space and a number of 55 failure samples are obtained. The proposed algorithm then re-generates 55×20 rounds $= 1100$ samples by utilizing direct sampling and the Metropolis-Hasting algorithm based on this pre-samples set, respectively. The obtained expression of FPF is:

$$\hat{P}_F(\theta_1,\theta_2) = \hat{P}(F)\exp\left[a_0 + \sum_{k=1}^{2} a_k\left(\frac{\theta_1-\underline{\theta}_1}{\bar{\theta}_1-\underline{\theta}_1}\right)^k + \sum_{k=1}^{2} b_k\left(\frac{\theta_2-\underline{\theta}_2}{\bar{\theta}_2-\underline{\theta}_2}\right)^k + c\left(\frac{\theta_1-\underline{\theta}_1}{\bar{\theta}_1-\underline{\theta}_1}\right)\left(\frac{\theta_2-\underline{\theta}_2}{\bar{\theta}_2-\underline{\theta}_2}\right)\right] \tag{24}$$

where the coefficients $a_i, b_i (i = 1, 2)$, and c for the proposed methods (SR(Direct sampling) and SR (Metropolis algorithm)) are shown in the Table 5.



(a) $P_F(\theta_1, \theta_2 = 15)$ (b) $P_F(\theta_1 = 13, \theta_2)$

Figure 8: The one dimensional FPF results obtained by different methods (Example 2) (a) $P_F(\theta_1, \theta_2 = 15)$; (b) $P_F(\theta_1 = 13, \theta_2)$. The corresponding computational cost is listed in Table 6.

The one dimensional FPF results, $P_F(\theta_1, \theta_2 = 15)$ and $P_F(\theta_1 = 13, \theta_2)$, by the proposed algorithm and other methods are shown in Fig. 8. It can be seen that the result based on solely the initial pre-samples has a big error. At the same time, the proposed method (regenerating based on the initial pre-samples) is still able to obtain a high precision.

The computational cost information of each method is listed in Table 6. Note that Weighted MCS (WMCS) uses 10^4 samples; Weighted IS (WIS) uses 500 (excluding the cost of solving for the design point); Weighted SS (WSS) uses 2000 (1000 for each level). Note that all the weighted approaches are carried out in the original space, and considerable error can be seen in the results of these methods. Through comparison, it can be found that in the two-dimensional normal example, the proposed SR algorithm has high efficiency and accuracy.

4.3. Example 3: turbine disk

This case study considers the disk of a turbine of a turbo-engine, of which the reliability is key to the safety of aeronautical transport vehicles. According to the well-known Mason-Coffin law to

Table 5: The values of coefficients of the FPF estimation in Eq. (24) (Example 2)

Coefficients	Proposed SR(Metropolis algorithm)	Proposed SR(Direct sampling)
a_0	1.6623	1.8686
a_1	3.0960	2.4325
a_2	3.5174	2.5541
b_1	−8.7622	−7.7262
b_2	−12.1310	−10.1725
c	−9.9302	−11.7398

Table 6: Computational cost of different methods (Example 2)

Methods	No. of samples	No. of failure samples
Initial pre-samples	–	55
Proposed SR(Direct sampling)	500	55
Proposed SR(Metropolis algorithm)	500	55
WMCS	10^4	60
WIS	500	114
WSS	1000×2	257
Direct MCS	10^6	–

consider the effect of mean stress and mean strain on the fatigue life of the disk N_l, the fatigue life can be computed as:

$$\frac{\Delta\varepsilon}{2} = \left(\frac{\sigma'_f - \sigma_m}{E}\right)(2N_l)^b + \left(\varepsilon'_f - \varepsilon_m\right)(2N_l)^c \tag{25}$$

where σ'_f is the fatigue strength coefficient; ε'_f is the fatigue ductility coefficient; ε_m is the mean strain; σ_m is the mean stress; b is the fatigue strength exponent of Basquin' law; c is the fatigue ductility exponent of Coffin's law; $\Delta\varepsilon_m$ is the strain range which $\Delta\varepsilon_m = \varepsilon_m/2$ under 0 takeoff-0 load cycle here; $E = 1.85 \times 10^5$MPa is Young's modulus. In this study, it is assumed that the actual life under a 0-takeoff-0 load cycle must exceed the required fatigue life. In this case, the performance function can be expressed as:

$$g(\boldsymbol{x}) = N_l\left(\sigma'_f, \varepsilon'_f, \sigma_m, \varepsilon_m, b, c\right) - N_{l0} \tag{26}$$

where N_{l0} is the required minimum service life and it is set as a constant $N_{l0} = 10^6$ cycles; N_l is the computed fatigue life under the 0 -takeoff- 0 load cycle. All the random variables are assumed to be normally distributed and the corresponding distribution information is given in Table 7 . The mean value of σ'_f is taken to be the design parameter, i.e., $\theta = \mu_{\sigma' f} \in [1400, 2300]$MPa.

Table 7: Distribution information of basic random variables of turbine disk (Example 3)

Random variable	Mean	C.o.v.	Distribution
σ_m(MPa)	1077.63	0.1	Normal
ε_m	0.0045497	0.1	Normal
σ'_f(MPa)	θ	0.1	Normal
ε'_f	0.0196	0.1	Normal
b	-0.096	-0.05	Normal
c	-0.41	-0.05	Normal

For implementing the augmented state approach, it is assumed that θ is uniformly distributed over $[1400, 2300]$. In this example, Subset simulation (MCMC) in augmented space is used to carry out the simulation. It uses 200 samples (100 for each level), and 171 failure samples are obtained. The proposed SR algorithm (Direct sampling) is used to regenerated 171×20 samples.

The FPF is obtained as

$$\hat{P}_F(\theta) = 0.0850 \exp\left[1.7237 - 4.8132\left(\frac{\theta - 1400}{900}\right) - 1.0211\left(\frac{\theta - 1400}{900}\right)^2 - 3.8165\left(\frac{\theta - 1400}{900}\right)^3\right] \tag{27}$$

The FPF results obtained by different methods are plotted in Fig. 9. The details of different methods are given in Table 8. From Fig. 9, it can be seen that the solution of FPF by the proposed method is the most accurate. For the other methods, the result based on the initial samples has larger error. First order and second order exponential RSM obtain results which have an acceptable error, but the corresponding computational cost is larger. Concerning the weighted approaches, only WIS obtains satisfactory results, but it uses 1000 sample which excludes the design point solving cost. The results of Weighted MCS (WMCS) with 4×10^4 samples and weighted SS (WSS) with 3000(1000 for each level) samples still have some errors. The proposed SR algorithm obtains the satisfied results with only 200 samples (100 for each level) which uses the least computational cost. High efficiency of the proposed algorithm still can be seen.

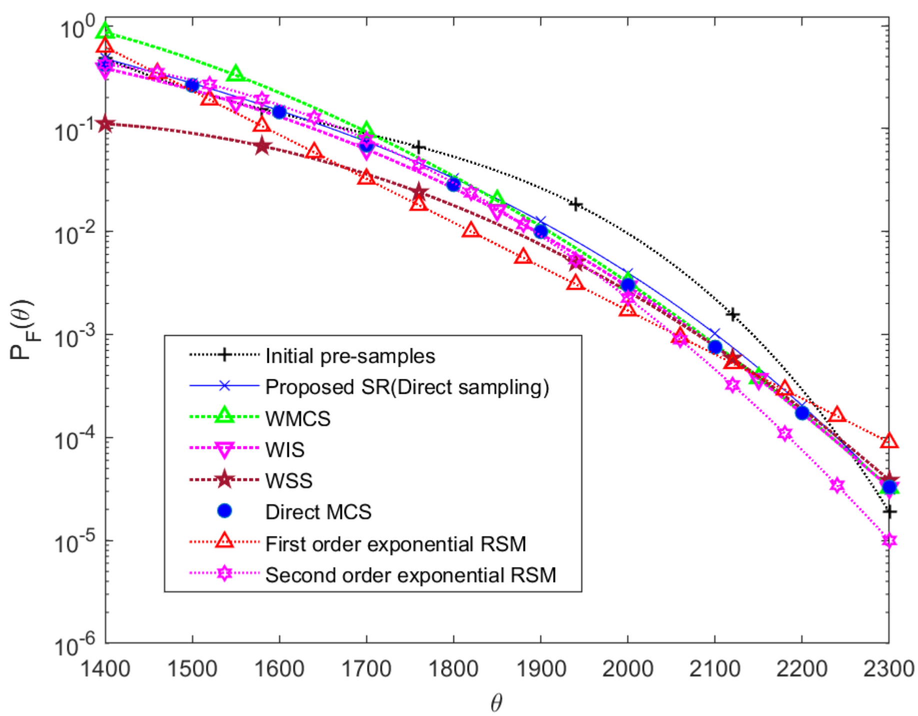


Figure 9: The FPF results obtained by different methods (Example 3), the corresponding computational cost is listed in Table 8.

Table 8: Computational cost of different methods (Example 3)

Methods	No. of samples	No. of failure samples
Original samples	–	171
Proposed SR(Direct sampling)	200	171
WMCS	4×10^4	87
WIS	1000	506
WSS	1000×3	187
RSM first	$10^5 \times 3$	–
RSM second	$10^5 \times 3$	–
Direct MCS	$10^6 \times 10$	–

4.4. Example 4: aircraft inner flap

This case study considers the case of an inner aircraft flap, subjected to an aerodynamic load which is transformed into a concentrated load F applied to the nodes of the finite element model (3274 elements in total), which is shown in Fig. 10. Failure is defined as the maximum displacement of all the nodes d_{max} exceeding an admissible maximal displacement $D_a = 34.1$ mm. The performance function is defined as:

$$g(\boldsymbol{x}) = D_a - d_{max}\left(t_1, t_2, t_3, t_4, A_1, A_2, E_1, G_1, E_2, G_2, F\right) \tag{28}$$

where t_1, t_2, t_3 and t_4 are the thickness values of four kinds of beams in the flap; A_1 and A_2 are the cross section areas of two beams; $E_1\,(E_2)$ and $G_1\,(G_2)$ are the elastic modulus and shear modulus, respectively; the instrumental random variable F represents the randomness of the load applied to the nodes, and for the load applied in the node i, the value is $F_i = (1 + F)F_{i0}$ where F_{i0} is a constant nominal value. All variables are mutual independent normal variables and the corresponding distribution information is given in Table 9.

In this case, it is found that the reliability of this structure is more sensitive to the thickness t_4 than other shape parameters, so the mean value of t_4 is taken as the design parameter, i.e., $\theta = \mu_{t_4} \in [1.3, 1.7]$ mm. And the FPF with respect to μ_{t_4} is investigated.

The proposed SR algorithm (Direct sampling) is applied. First, suppose $\theta = \mu_{t_4}$ is uniformly distributed over design domain $[1.3, 1.7]$ mm, MCMC simulation is adopted with 200 sample for each level, and 154 failure samples are obtained at the last level. Hereto, there failure samples are selected as the pre-samples and then 154×20 samples of θ are regenerated based on these pre-samples. The estimate of FPF obtained by the proposed method is:

$$\hat{P}_F(\theta) = 0.0770 \exp\left[1.6471 - 5.0394\left(\frac{\theta - 1.3}{0.4}\right) + 2.3538\left(\frac{\theta - 1.3}{0.4}\right)^2 - 6.2372\left(\frac{\theta - 1.3}{0.4}\right)^3\right] \tag{29}$$

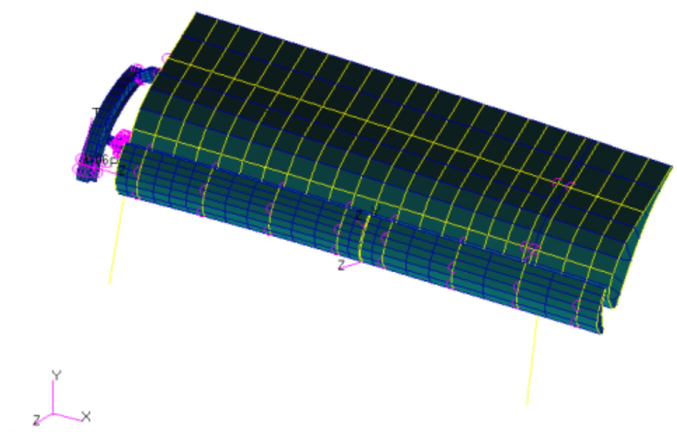


Figure 10: The finite element model of aircraft inner flap (Example 4)

This FPF by different methods are plotted in Fig. 11. Note that Direct MCS is carried out considering a Kriging surrogate model instead of the original limit state function in which finite element analysis is involved. It can be seen that FPF decreases as the mean value of t_4 increases, which is reasonable from a physical viewpoint. The FPF results based on pre-samples (the original 154 samples) has a big error. On the other hand, the proposed SR algorithm based on these pre-samples obtains satisfactory results. Thus the effectiveness of the proposed SR algorithm is shown. Table 10 summarizes the computation cost of the different methods. Although the proposed SR algorithm uses only one reliability analysis (i.e., MCMC simulation) with 200 samples, the solution of FPF still has very high precision. This shows that the accuracy and efficiency of the proposed approach.

Table 9: Distribution information of basic random variables of the inner flap (Example 4)

Random variable	Mean	C.o.v.
t_1(mm)	2	0.05
t_2(mm)	2	0.05
t_3(mm)	4	0.05
t_4(mm)	θ	0.05
A_1 (cm^2)	50	0.05
A_2 (cm^2)	150	0.05
E_1(MPa)	70380	0.05
G_1(MPa)	26458.6	0.05
E_2(MPa)	72450	0.05
G_2(MPa)	27236.8	0.05
F	0	0.05

Table 10: Computational cost of different methods (Example 4)

Methods	No. of samples	No. of failure samples
Original samples	–	154
Proposed SR(Direct sampling)	200×2	154
WMCS	5000	106
WIS	1000	490
WSS	1000×2	256
Direct MCS	$10^6 \times 10$	–

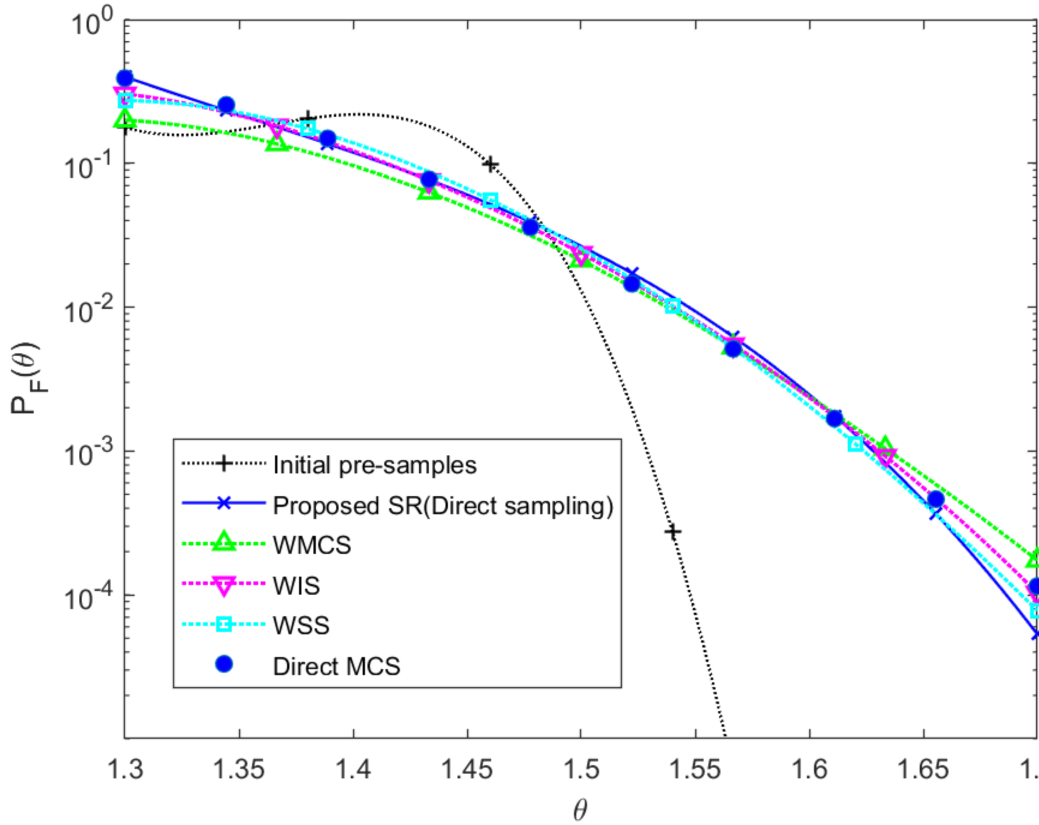


Figure 11: The results of FPF obtained by different methods(Example 4), the corresponding computational cost is given in Table 10.

5. Conclusions

In this paper, a methodology is presented for the estimation of the structural failure probability function. The method is based on a newly proposed regeneration algorithm that utilizes Bayesian theory. The most attractive features are that (1) it needs only one reliability analysis in augmented space, i.e., using classical MCS or Subset simulation; (2) only a limited number of failure samples is required, from which the proposed algorithm can generate more samples in the failure domain of the problem under consideration. Hence, no repeated limit state function evaluations are needed, which results a highly efficient, yet accurate, estimation of the FPF.

Numerical examples are given to test the proposed algorithm and illustrate its application. It is found that the proposed algorithm for FPF estimation owns remarkable advantages both in terms of accuracy and computational efficiency when compared to other methods. While these results are encouraging, it should be pointed out that the proposed algorithm is best suited for treating low-to-moderate dimensional cases. This is a natural consequence of the difficulties corresponding to sampling from a high-dimensional augmented space. This is well-known in the literature of augmented space methods. Future work will involve the application of the proposed method to the reliability-based optimization combined with decoupling approach, reliability analysis under non-probabilistic uncertainty and reliability sensitivity analysis.

CRediT authorship contribution statement

Xiukai Yuan: Conceptualization, Methodology, Software, Validation, Writing - original draft, Writing - review & editing, Formal analysis, Funding acquisition. Shanglong Wang: Software, Writing - original draft, Formal analysis. Matthias Faes: Writing - original draft, Writing - review & editing. Marcos A. Valdebenito: Writing - original draft, Writing - review & editing. Michael Beer: Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to acknowledge financial support from the Aeronautical Science Foundation of China (Grant No. ASFC-20170968002), National Science and Technology Major Project (Grant No. 2019-I-0006-0006), Special Project on the Integration of Industry, Education and Research of AECC (Grant No. HFZL2020CXY004 and HFZL2020CXY009).

Appendix A. The SR algorithm for the mean value of a Normal variable

Suppose the mean value of x_i is the design parameter θ. The conditional PDF of x_i is given by

$$f(x_i|\theta_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(x_i-\theta_i)^2}{2\sigma_i^2}\right] \tag{A.1}$$

where $x_i \sim N(\theta_i, \sigma_i^2)$; σ_i^2 is the known variance. In augmented space, assume θ_i is uniform distributed over the design region $[\underline{\theta}_i, \bar{\theta}_i]$, i.e., $\theta_i \sim U[\underline{\theta}_i, \bar{\theta}_i]$, the distribution $\varphi(\theta_i)$ is given as

$$\varphi(\theta_i) = \frac{1}{\bar{\theta}_i - \underline{\theta}_i}, \quad \theta_i \in [\underline{\theta}_i, \bar{\theta}_i] \tag{A.2}$$

Then the marginal distribution of x can be obtained as

$$f(x_i) = \int_{\underline{\theta}_i}^{\bar{\theta}_i} f(x_i|\theta_i)\varphi(\theta_i)\mathrm{d}\theta_i = \frac{1}{\bar{\theta}_i - \underline{\theta}_i}\left[\Phi\left(\frac{\bar{\theta}_i - x_i}{\sigma_i}\right) - \Phi\left(\frac{\underline{\theta}_i - x_i}{\sigma_i}\right)\right] \tag{A.3}$$

where $\Phi(\cdot)$ is the cumulative distribution function of normal standard variable. Next, based on Bayesian theory explained in Eq. (6), the posterior PDF of θ can be derived by substituting Eq. (A.1), Eq. (A.2) and Eq. (A.3) into Eq. (6):

$$\varphi(\theta_i|x_i) = \frac{f(x_i|\theta_i)\varphi(\theta_i)}{f(x_i)} = \frac{1}{\Delta_i}\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{(\theta_i - x_i)^2}{2\sigma_i^2}\right), \quad \theta_i \in [\underline{\theta}_i, \bar{\theta}_i] \tag{A.4}$$

where $\Delta_i = \Phi\left(\frac{\bar{\theta}_i - x_i}{\sigma_i}\right) - \Phi\left(\frac{\underline{\theta}_i - x_i}{\sigma_i}\right)$ is a constant for given x_i. It can be seen that the posterior PDF $\varphi(\theta_i|x_i)$ in Eq. (A.4) is actually a truncated PDF of Normal distribution for a given x_i. This means that if we have samples of x beforehand, we can generate samples of θ_i which follow $\varphi(\theta_i)$ by directly using the posterior PDF $\varphi(\theta_i|x_i)$ in Eq. (A.4). Further, the conditional distribution can be obtained according to Eq. (6)

$$\varphi(\theta_i|x_i, F) = I_F(\boldsymbol{x})\varphi(\theta_i|x_i) = \frac{1}{\Delta_i}\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{(\theta_i - x_i)^2}{2\sigma_i^2}\right), \quad \theta_i \in [\underline{\theta}_i, \bar{\theta}_i], \boldsymbol{x} \in F \tag{A.5}$$

References

[1] M. A. Valdebenito, G. I. Schuëller, A survey on approaches for reliability-based optimization, Structural and Multidisciplinary Optimization 42 (2010) 645–663.

[2] D. Jerez, H. Jensen, M. Beer, Reliability-based design optimization of structural systems under stochastic excitation: An overview, Mechanical Systems and Signal Processing 166 (2022) 108397.

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

[4] T. Zou, S. Mahadevan, A direct decoupling approach for efficient reliability-based design optimization, Structural and Multidisciplinary Optimization 31 (2006) 190–200.

[5] R. Rackwitz, B. Flessler, Structural reliability under combined random load sequences, Computers & Structures 9 (1978) 489–494.

[6] K. Breitung, Asymptotic approximations for probability integrals, Probabilistic Engineering Mechanics 4 (1989) 187–190.

[7] N. Metropolis, S. Ulam, The monte carlo method, Journal of the American statistical association 44 (1949) 335–341.

[8] R. Melchers, Importance sampling in structural systems, Structural safety 6 (1989) 3–10.

[9] M. A. Misraji, M. A. Valdebenito, H. A. Jensen, C. F. Mayorga, Application of directional importance sampling for estimation of first excursion probabilities of linear structural systems subject to stochastic gaussian loading, Mechanical Systems and Signal Processing 139 (2020) 106621.

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

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

[12] M. A. Valdebenito, P. Wei, J. Song, M. Beer, M. Broggi, Failure probability estimation of a class of series systems by multidomain line sampling, Reliability Engineering & System Safety 213 (2021) 107673.

[13] M. Gasser, G. I. Schuëller, Reliability-based optimization of structural systems, Mathematical Methods of Operations Research 46 (1997) 287–307.

[14] H. A. Jensen, Structural optimization of linear dynamical systems under stochastic excitation: a moving reliability database approach, Computer methods in applied mechanics and engineering 194 (2005) 1757–1778.

[15] D. R. Jones, M. Schonlau, W. J. Welch, Efficient global optimization of expensive black-box functions, Journal of Global optimization 13 (1998) 455–492.

[16] H. Jensen, D. Jerez, M. Valdebenito, An adaptive scheme for reliability-based global design optimization: A markov chain monte carlo approach, Mechanical Systems and Signal Processing 143 (2020) 106836.

[17] C. Cortes, V. Vapnik, Support-vector networks, Machine learning 20 (1995) 273–297.

[18] V. N. Vapnik, An overview of statistical learning theory, IEEE transactions on neural networks 10 (1999) 988–999.

[19] G. Blatman, B. Sudret, Sparse polynomial chaos expansions and adaptive stochastic finite elements using a regression approach, Comptes Rendus Mécanique 336 (2008) 518–523.

[20] Y. Zhou, Z. Lu, W. Yun, Active sparse polynomial chaos expansion for system reliability analysis, Reliability Engineering & System Safety 202 (2020) 107025.

[21] B. Echard, N. Gayton, M. Lemaire, AK-MCS: an active learning reliability method combining kriging and monte carlo simulation, Structural Safety 33 (2011) 145–154.

[22] H. Li, Z. Lü, Z. Yue, Support vector machine for structural reliability analysis, Applied Mathematics and Mechanics 27 (2006) 1295–1303.

[23] S. Au, Reliability-based design sensitivity by efficient simulation, Computers & structures 83 (2005) 1048–1061.

[24] J. Ching, Y.-H. Hsieh, Local estimation of failure probability function and its confidence interval with maximum entropy principle, Probabilistic Engineering Mechanics 22 (2007) 39–49.

[25] J. Ching, Y.-H. Hsieh, Approximate reliability-based optimization using a three-step approach based on subset simulation, Journal of Engineering Mechanics 133 (2007) 481–493.

[26] A. A. Taflanidis, J. L. Beck, An efficient framework for optimal robust stochastic system design using stochastic simulation, Computer Methods in Applied Mechanics and Engineering 198 (2008) 88–101.

[27] A. A. Taflanidis, J. L. Beck, Stochastic subset optimization for reliability optimization and sensitivity analysis in system design, Computers & Structures 87 (2009) 318–331.

[28] A. A. Taflanidis, J. L. Beck, Reliability-based design using two-stage stochastic optimization with a treatment of model prediction errors, Journal of Engineering Mechanics 136 (2010) 1460–1473.

[29] A. A. Taflanidis, Stochastic subset optimization incorporating moving least squares response surface methodologies for stochastic sampling, Advances in Engineering Software 44 (2012) 3–14.

[30] W.-S. Liu, S. H. Cheung, Reliability based design optimization with approximate failure probability function in partitioned design space, Reliability Engineering & System Safety 167 (2017) 602–611.

[31] X. Yi, B. Dhillon, J. Shi, H. Mu, P. Hou, Reliability optimization allocation method for multifunction systems with multistate units based on goal-oriented methodology, ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg 3 (2017).

[32] S. Yu, Z. Wang, Z. Wang, Time-dependent reliability-based robust design optimization using evolutionary algorithm, ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering 5 (2019) 020911.

[33] X. Yuan, Local estimation of failure probability function by weighted approach, Probabilistic Engineering Mechanics 34 (2013) 1–11.

[34] X. Yuan, Z. Lu, Efficient approach for reliability-based optimization based on weighted importance sampling approach, Reliability Engineering & System Safety 132 (2014) 107–114.

[35] P. Wei, J. Song, S. Bi, M. Broggi, M. Beer, Z. Lu, Z. Yue, Non-intrusive stochastic analysis with parameterized imprecise probability models: I. Performance estimation, Mechanical Systems and Signal Processing 124 (2019) 349–368.

[36] P. Wei, J. Song, S. Bi, M. Broggi, M. Beer, Z. Lu, Z. Yue, Non-intrusive stochastic analysis with parameterized imprecise probability models: II. Reliability and rare events analysis, Mechanical Systems and Signal Processing 126 (2019) 227–247.

[37] X. Yuan, S. Liu, M. Valdebenito, J. Gu, M. Beer, Efficient procedure for failure probability function estimation in augmented space, Structural Safety 92 (2021) 102104.

[38] G. Casella, R. L. Berger, Statistical inference, Cengage Learning, 2021.

[39] S.-K. Au, J. L. Beck, A new adaptive importance sampling scheme for reliability calculations, Structural safety 21 (1999) 135–158.

[40] D. Ormoneit, H. White, An efficient algorithm to compute maximum entropy densities, Econometric reviews 18 (1999) 127–140.