Journal of Computational Physics 230 (2011) 8683–8697

Contents lists available at SciVerse ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp

An efficient surrogate-based method for computing rare failure probability

Jing Li [a], Jinglai Li [b], Dongbin Xiu [a,*]

[a] Department of Mathematics, Purdue University, West Lafayette, IN 47907, United States
[b] Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139, United States

ARTICLE INFO

Article history:
Received 31 December 2010
Received in revised form 13 July 2011
Accepted 11 August 2011
Available online 28 August 2011

Keywords:
Rare events
Failure probability
Importance sampling
Cross-entropy

ABSTRACT

In this paper, we present an efficient numerical method for evaluating rare failure probability. The method is based on a recently developed surrogate-based method from Li and Xiu [J. Li, D. Xiu, Evaluation of failure probability via surrogate models, J. Comput. Phys. 229 (2010) 8966–8980] for failure probability computation. The method by Li and Xiu is of hybrid nature, in the sense that samples of both the surrogate model and the true physical model are used, and its efficiency gain relies on using only very few samples of the true model. Here we extend the capability of the method to rare probability computation by using the idea of importance sampling (IS). In particular, we employ cross-entropy (CE) method, which is an effective method to determine the biasing distribution in IS. We demonstrate that, by combining with the CE method, a surrogate-based IS algorithm can be constructed and is highly efficient for rare failure probability computation—it incurs much reduced simulation efforts compared to the traditional CE-IS method. In many cases, the new method is capable of capturing failure probability as small as $10^{-12} \sim 10^{-6}$ with only several hundreds samples.

1. Introduction

Uncertainties in specifying material properties, geometric parameters, boundary conditions and applied loadings are unavoidable in describing real-life engineering systems. Traditionally, this has been catered for in an *ad hoc* way through the use of safety factors at the design stage. Such an approach is becoming less satisfactory in today's competitive design environment, for example, in minimum weight design of aircraft structures. Therefore, accurate evaluation of failure probability of a given system is becoming increasingly important and a fundamental problem in many fields such as risk management, structural design, reliability based optimization, etc.

Since in most cases the function separating the safe and failure domains is not known explicitly and can be highly irregular in high dimensional spaces, the standard integration rules such as Gauss quadrature, cubature, sparse grids, etc., are not directly applicable. The most commonly used method is Monte Carlo simulation (MCS), which requires one to simulate, also known as to sample, the underlying system repetitively. Though straightforward to implement, MCS can be highly time-consuming when the underlying system is complex, for each sample requires a full-scale numerical simulation of the system. To reduce the computational effort, many alternative non-sampling based methods have been developed, such as FORM/SORM (first-order/second-order reliability method) [13,9,6,12,20,27,26], RSM (response surface method) [29,8,3,21,10,11,19], etc. These methods usually incur much less simulation cost, compared to MCS, at the expense of reduced accuracy.

For systems with rare failure probability, which in this paper is defined as failure probability less than 10^{-5}, the problem becomes drastically more difficult. To this end, one needs to employ methods with sufficiently high accuracy so that the rare

* Corresponding author.
E-mail addresses: li176@math.purdue.edu (J. Li), jinglai@mit.edu (J. Li), dxiu@purdue.edu (D. Xiu).

0021-9991/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2011.08.008

failure probability in the small tails of the distribution can be captured. And to this date none of the non-sampling methods can be highly effective in a general setting, due to their lack of high accuracy. Probably the most reliable method is based on MCS. To this end, the computational cost becomes much higher. As a rule-of-thumb, one usually needs about 10 samples in the failure domain to reliably estimate the failure probability. For rare probability this translates to prohibitively large number of samples. For example, for probability of $O(10^{-6})$, it is not uncommon to use $O(10^7)$, or even $O(10^8)$, number of samples for good accuracy.

To reduce the number of samples, one of the most widely used methods is importance sampling (IS). (See, for example, [22,28,4,25]). In IS, one seeks to sample the random variables from a different distribution, called *biasing distribution*, rather than the original one. The biasing distribution is constructed in such a way that more samples will land in the failure domain and thus results in (much) less total number of samples. The use of the biasing distribution, essentially a change of measure approach, is incorporated in the sampling estimate by adjusting the weight of each sample to ensure that the final estimate is unbiased. A successfully constructed IS method can significantly increase the efficiency of MCS. For the aforementioned example, for probability of $O(10^{-6})$, a carefully constructed IS method can reduce the total number of samples by several orders, e.g., $O(10^4)$, a drastic reduction from the $O(10^7)$ samples by the brute force MCS.

For practical engineering systems, computer simulations can be extremely time consuming. In many cases, one can only afford very limited number of simulations—nothing beyond a few hundreds. In this case, even the most effective IS method is not applicable. And one often has to resort to some highly problem dependent and/or *ad hoc* approaches to estimate the rare probability and then use safety factor to mitigate the impacts caused by the inaccuracy. It should be noted that the idea of combining IS with techniques such as FORM/SORM has been pursued and resulted in several interesting strategies. Also there are other types of approaches that do not employ IS, e.g., sequential Monte Carlo method ([15]), etc. These methods have their own specific strength and limitations, and will not be discussed in this paper. For detailed discussions, see, for example, [7,20,27,26,15].

The purpose of this paper is to present a new algorithm for computing rare probability, with significantly enhanced efficiency so that it can capture rare probability of less than 10^{-5} with (in many cases) a few hundreds of samples. A key feature of the method is that it is in a general setting and does not require approximations, transformations, or manipulations of the underlying systems. The method relies on a recently developed method by Li and Xiu [17], where a fundamental flaw in the traditional RSM was identified and an improvement – a surrogate-based hybrid method – was proposed. The method utilizes samples of the response surface, which hereafter will be referred to as *surrogate*, in the majority of the probability space and samples of the real system only in the region surrounding the failure mode. It achieves high accuracy in the failure probability estimation while incurring much reduced number of samples of the real system. In this paper the hybrid method from [17] is extended to the case of rare failure probability by incorporating the idea of importance sampling. In particular, the cross-entropy (CE) method for IS is adopted. The CE method ([24,5]) is a relatively new Monte Carlo technique for both estimation and optimization. In the estimation setting, the CE method provides an adaptive way to find a good biasing distribution for quite general problems. The biasing distribution is obtained by an optimization procedure that minimizes its distance, measured by cross-entropy between two distributions, from the *optimal* biasing distribution, which exists in theory but is unavailable in practice. Furthermore, an adaptive multilevel iterative algorithm is available to effectively compute the biasing distribution. In this paper we present a new algorithm based by combining the ideas of both the CE method and the hybrid method. The surrogate model is utilized in both the CE optimization step and the final IS integral evaluation. And the result is a highly efficient algorithm where the simulation cost is significantly reduced further, compared to the direct CE-IS method. In many test cases, the new algorithm is able to resolve failure probability as low as 10^{-6} with only several hundreds samples.

It should be emphasized that the purpose of this paper is *not* to compare different IS methods or other types of methods for rare probability. This is impossible to do owing to the large amount of existing methods. Rather, the paper presents an approach of enhancing certain IS strategies, by using the surrogate-based hybrid idea from [17], and demonstrates that the new method is rigorous, efficient, and more importantly, easy to implement. We restrict our discussion to the CE method, and leave the potential extension to other IS strategies to future research.

Throughout this paper, we assume that a surrogate model is available to use. The surrogate model can be constructed by simulations or given by physics laws or literature. Therefore we do not count the construction of the surrogate as part of the computational cost. Also, the proposed method does not seek to improve the accuracy of the surrogate, which is impossible to do if the surrogate is obtained from literature or physics laws. It should also be noted that improving the accuracy of the surrogate may not be useful at all [17]. In fact, the method does not require the surrogate model to be of high accuracy.

The rest of this paper is organized as the following. After presenting the formulation of rare failure probability computation in Section 2, we briefly review the key ingredients of the present method. These include importance sampling, cross-entropy method, and the hybrid method of [17]. The details of the new method are presented in Section 4, where two numerical algorithms are presented. Numerical examples are presented in Section 5 to demonstrate the effectiveness of the new algorithms.

2. Problem setup

In this paper we consider the problem of failure probability evaluation. In a general setting, the problem can be described as the following. Let $Z = (Z_1, \ldots, Z_{n_Z})$ be a n_Z-dimensional random vector with distribution function $F_Z(z) = \text{Prob}(Z \leqslant z)$, where $z \in \mathbb{R}^{n_Z}$ is a real vector. Here we restrict the discussion to the case of continuous distribution, where a probability density function (PDF) $q(z)$ exists in the sense that $dF_Z(z) = q(z)dz$. We then consider the following integral

$$P_f = \text{Prob}(Z \in \Omega_f) = \int_{\Omega_f} q(z)dz = \int \mathbb{I}_{\Omega_f}(z)q(z)dz, \tag{2.1}$$

where $\mathbb{I}$ is the characteristic function satisfying

$$\mathbb{I}_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A, \end{cases} \tag{2.2}$$

and Ω_f is the failure domain defined as

$$\Omega_f \triangleq \{Z : g(Z) < 0\}. \tag{2.3}$$

Here $g(z)$ is a (scalar) limit state function, also called performance function, that defines the failure domain Ω_f. More precisely, the domain where $g < 0$ is the failure domain, whereas the domain of $g \geqslant 0$ is the safe domain. This is a general definition for failure probability, which is used widely in many disciplines involving reliability analysis and risk management. From now on we will use the shorthanded notation of $\{g(Z) < 0\}$ to stand for the set (2.3) and (2.1) will be frequently written as

$$P_f = \int \mathbb{I}_{\{g(z)<0\}}(z)q(z)dz, \tag{2.4}$$

unless confusion arises.

In this paper, we will consider the case of *rare failure probability*, which refers to $P_f < 10^{-5}$. (Note in different application areas, different measures of "rare" are used.)

It is obvious that the limit function g is important, because it separates the safe and failure domain. The difficulty is that in almost all practical cases, the limit state function is not known explicitly. Instead, usually the random vector Z represents uncertain inputs to a complex stochastic system, and the limit state function g depends on the solution of the stochastic system, often in a nonlinear and/or implicit manner. Hence $g(Z)$ can only be evaluated by solving the underlying stochastic system which can be time consuming. Here in this paper we will not explicitly define the underlying stochastic system.

3. Preliminaries

This section is devoted to the introduction of the core components of the new method. These include importance sampling (IS) method, particularly cross-entropy (CE) method, and the surrogate-based hybrid method for failure probability computation from [17].

3.1. Importance sampling

The most reliable and straightforward approach for (2.4) is Monte Carlo sampling (MCS), where random samples are generated according to the distribution of Z and those land in the failure region are counted to estimate the failure probability. Let $z^{(i)} \in \mathbb{R}^{n_Z},\ i = 1, \ldots, M$, be a set of samples drawn from the distribution $q(z)$, then the MCS estimate of the failure probability (2.4) is

$$P_f^{mc} = \frac{1}{M}\sum_{i=1}^{M} \mathbb{I}_{\{g(z)<0\}}(z^{(i)}). \tag{3.1}$$

Although easy to implement, the MCS approach can be costly in practice, for each sample point requires a full-scale simulation of the underlying stochastic system. And usually a large number of samples is required to obtain an accurate estimate of the failure probability. This is especially challenging for the rare failure probability considered in this paper. For failure probability less than 10^{-5}, the required number of samples would be so large that it is beyond the reach of most practical complex systems, for which one can usually afford at most a few hundreds simulations.

An effective way to simulate rare probability is importance sampling (IS). The idea is that, instead of sampling (2.4) using the density $q(z)$ directly, one introduces a biasing distribution $p(z)$ and rewrites (2.4) as

$$P_f = \int \mathbb{I}_{\{g<0\}}(z)\frac{q(z)}{p(z)}p(z)dz = \int \mathbb{I}_{\{g<0\}}(z)W(z)p(z)dz, \tag{3.2}$$

where $W(z) = q(z)/p(z)$ is the likelihood ratio. One then conducts MCS estimation for (3.2) by drawing samples $\{z^{(i)}\}$ from the distribution $p(z)$ and obtains

$$P_f^{is} = \frac{1}{M}\sum_{i=1}^{M} \mathbb{I}_{\{g(z)<0\}}(z^{(i)})W(z^{(i)}). \tag{3.3}$$

The key in designing an effective IS method is to construct a "good" biasing distribution $p(z)$ so that more samples will land in the failure domain. (The contributions of the samples will be adjusted by the likelihood ratio.) The successful design of the biasing distribution is, however, not a trivial task. And this is where different IS methods differ from each other. See, for example, [22,28,4,25]. It should be noted that the "optimal" biasing distribution exists,

$$p^*(z) = \frac{1}{P_f}\mathbb{I}_{\{g(z)<0\}}(z)q(z). \tag{3.4}$$

It will result in zero variance in (3.2) and therefore require only a single sample to evaluate the integral. The problem is that it depends on the unknown P_f and thus can not be evaluated.

3.2. Cross-entropy method

There exist several different approaches for IS, some of which are problem dependent. Here we employ the cross-entropy (CE) method, which was motivated by an adaptive algorithm for estimating probabilities of rare events in complex stochastic networks ([23]). It is a quite general IS strategy and has been under intense development since its introduction. For general overviews of the method, see [24,5]. The key ingredient of the CE method is the use of cross-entropy, also known as Kullback–Leibler divergence (KLD) [16], between two probability distributions to optimally determine the biasing distribution in (3.2). For two distributions p_1 and p_2, the CE, or KLD, takes the following form,

$$\mathcal{D}(p_1, p_2) = \mathbb{E}_{p_1}\left[\ln\frac{p_1(z)}{p_2(z)}\right] = \int p_1(z)\ln\frac{p_1(z)}{p_2(z)}dz. \tag{3.5}$$

It is easy to verify that $\mathcal{D}(p_1, p_2) \geqslant 0$, where equality is achieved when $p_1 = p_2$ a.e.

In the CE method, the biasing distribution $p(z)$ in (3.2) is determined by requiring $p(z)$ to be "close" to p^*, the optimal and yet unavailable biasing distribution (3.4). The CE distance from p^* to p can be written in two parts:

$$\mathcal{D}(p^*, p) = \int p^*(z)\ln p^*(z)dz - \int p^*(z)\ln p(z)dz. \tag{3.6}$$

Minimizing (3.6) is equivalent to maximize the second integral on the right-hand-side. Often it is possible to search for the biasing distribution from a parameterized family of distribution $\{p(z;v)\}_{v\in V}$, where V is the parameter set. By using (3.4), the optimization problem (3.6) becomes solving for

$$v = \text{argmax}_v \int \mathbb{I}_{\{g(z)<0\}}(z)\ln(p(z;v))q(z)dz.$$

However, the same difficulty caused by the rareness of $\{g < 0\}$ in term of the distribution $q(z)$ persists. To circumvent the difficulty, the idea of IS is again utilized. By adopting another biasing distribution $p(z; w)$ we obtain the following optimization problem

$$v = \text{argmax}_v \int \mathbb{I}_{\{g(z)<0\}}(z)W(z;w)\ln(p(z;v))p(z;w)dz, \tag{3.7}$$

where $W(z; w) = q(z)/p(z; w)$ is the likelihood ratio. In practice, the stochastic counterpart of (3.7) is usually employed. Let $\{z^{(i)}\}_{i=1}^M$ be samples drawn from the distribution $p(z; w)$, we solve

$$\hat{v} = \text{argmax}_v \frac{1}{M}\sum_{i=1}^{M} \mathbb{I}_{\{g(z)<0\}}(z^{(i)})W(z^{(i)};w)\ln(p(z^{(i)};v)). \tag{3.8}$$

This optimization problem can be effectively solved by a multilevel iterative method, where one generates a sequence of reference parameters v_k, $k = 0, 1, 2, \ldots$ and a sequence of decreasing levels $\gamma_k > 0$. These sequences are then used to define intermediate failure domains $\mathbb{I}_{\{g<\gamma_k\}}$ under distribution $p(z; v_k)$, instead of $p(z; w)$ as in (3.7), whose probability is not rare. The iteration terminates when the level γ_k drops to zero, which is the original failure domain, and the converged v_k defines the biasing distribution $p(z; v)$ in the IS integral (3.2).

The CE method is quite general, so long as one can choose a proper parameterized family for the choice of the biasing distribution. It should be noted that implementation of non-parametric CE method has been investigated [14,2]. Also, many studies have been conducted on CE methods, and we will not engage in more discussions here.

3.3. Hybrid method

The hybrid method from [17] utilizes surrogate models for the limit state function to enhance the performance of MCS. The basic requirement is the availability of a surrogate model $\tilde{g}$ that approximates the original limit state function g. The

surrogate can be constructed either by numerical simulations or by physical laws. Very often surrogate models take the form of (multi-dimensional) polynomials of Z. In general, we require that the surrogate is an approximation in L^p norm. That is,

$$\epsilon_g = \|g(Z) - \tilde{g}(Z)\|_{L^p} = \left(\int |g(z) - \tilde{g}(z)|^p q(z) dz \right)^{1/p}, \quad p \geqslant 1, \tag{3.9}$$

exists. Note that the existence of this error is a mathematical necessity to ensure the rigorousness of the approach. In practice the knowledge of this error is not required. For a detailed discussion, see [17].

Given the surrogate $\tilde{g}$, instead of evaluating the failure probability (2.4) directly, one can evaluate

$$\widetilde{P}_f = \int \mathbb{I}_{\{\tilde{g}(z)<0\}}(z) q(z) dz, \tag{3.10}$$

where the surrogate $\tilde{g}$ is used to approximate the failure domain. The motivation is that (3.10) is easier to approximate by MCS. That is, let $\{z^{(i)}\}_{i=1}^M$ be samples drawn from the distribution $q(z)$, then

$$\widetilde{P}_f^{mc} = \frac{1}{M} \sum_{i=1}^{M} \mathbb{I}_{\{\tilde{g}(z)<0\}}(z^{(i)}), \tag{3.11}$$

is an approximation of (3.10) and easy to carry out, because sampling $\tilde{g}$ requires only repetitive evaluations of the explicitly known functional form of the surrogate and does not incur any simulation effort of the underlying system. Hence one can employ extremely large number of samples to conduct the estimate.

This is essentially the idea behind response surface method (RSM), which has been under development for a long time. Its accuracy and robustness have not been consistent, and the reason was largely associated with the lack of high accuracy in most surrogate models. And efforts have been made to improve the performance of RSM ([1,21,11,18]). However, in the work of [17], it was shown that such a straightforward sampling of the surrogate is fundamentally flawed and may lead to erroneous results, no matter how accurate the surrogate $\tilde{g}$ is. To address the flaw, the hybrid method was proposed in [17], where the direct surrogate sampling result is "corrected" by using samples of the true limit state function g. The key of the method is to replace the samples that are "close" to $\tilde{g} = 0$ by g so that most of the samples of the surrogate $\tilde{g}$ are kept. To this end, the hybrid method seeks to estimate the failure probability via the following two integrals,

$$\begin{aligned} P_\delta &= \int \mathbb{I}_{\{\tilde{g}<-\delta\}}(z) q(z) dz, \\ Q_\delta &= \int \mathbb{I}_{\{-\delta\leqslant\tilde{g}\leqslant\delta\}\cap\{g<0\}}(z) q(z) dz, \end{aligned} \tag{3.12}$$

where $\delta \geqslant 0$ is a (small) real number, and the hybrid estimate is the summation of the two integrals

$$P_f^h = P_\delta + Q_\delta. \tag{3.13}$$

It was proved in [17] that with properly chosen parameter δ, the error of hybrid estimate $\left|P_f^h - P_f\right|$ can be controlled by any prescribed accuracy threshold, and the choice of δ is directly related to the L^p norm ($p \geqslant 1$) of the error of the surrogate $\tilde{g}$.

To facilitate the following discussion, we discuss the hybrid estimate in a slightly more general setting by considering a failure domain $\{g(Z) < \gamma\}$, where γ is a real parameter.

For a given limit state function g, its surrogate model $\tilde{g}$, and a real parameter γ, we define

$$\mathrm{Prob}^{(h)}(\gamma; g, \tilde{g}) \triangleq \mathrm{Prob}(\{\tilde{g} < \gamma - \delta\}) + \mathrm{Prob}(\{|\tilde{g} - \gamma| \leqslant \delta\} \cap \{g < \gamma\}). \tag{3.14}$$

Or, equivalently, we write it in the following form,

$$\int \mathbb{I}^{(h)}_{\{\tilde{g}<\gamma\}}(z) q(z) dz \triangleq \int \mathbb{I}_{\{\tilde{g}<\gamma-\delta\}}(z) q(z) dz + \int \mathbb{I}_{\{|\tilde{g}-\gamma|\leqslant\delta\}\cap\{g<\gamma\}}(z) q(z) dz, \tag{3.15}$$

where the indicator function $\mathbb{I}^{(h)}_{\{\tilde{g}<\gamma\}}$ is implicitly defined as

$$\mathbb{I}^{(h)}_{\{\tilde{g}<\gamma\}}(z) \triangleq \mathbb{I}_{\{\tilde{g}<\gamma-\delta\}}(z) + \mathbb{I}_{\{|\tilde{g}-\gamma|\leqslant\delta\}\cap\{g<\gamma\}}(z). \tag{3.16}$$

The stochastic sampling counterpart of (3.14) is as the following. Let $\{z^{(i)}\}_{i=1}^M$ be samples drawn from the distribution $q(z)$, then

$$\widehat{\mathrm{Prob}}^{(h)}(\gamma; g, \tilde{g}) \triangleq \frac{1}{M} \sum_{i=1}^{M} \mathbb{I}^{(h)}_{\{\tilde{g}<\gamma\}}(z^{(i)}) = \frac{1}{M} \sum_{i=1}^{M} (\mathbb{I}_{\{\tilde{g}<\gamma-\delta\}}(z^{(i)}) + \mathbb{I}_{\{|\tilde{g}-\gamma|\leqslant\delta\}}(z^{(i)}) \cdot \mathbb{I}_{\{g<\gamma\}}(z^{(i)})). \tag{3.17}$$

Note the hybrid failure probability estimate (3.13) is a special case of $\gamma = 0$, i.e.,

$$P_f^h = \mathrm{Prob}^{(h)}(0; g, \tilde{g}) = \int \mathbb{I}^{(h)}_{\{\tilde{g}<0\}}(z) q(z) dz. \tag{3.18}$$

And the MCS estimate of the hybrid integral is

$$\widehat{P}_f^h = \widehat{\mathrm{Prob}}^{(h)}(0; g, \tilde{g}) = \frac{1}{M}\sum_{i=1}^{M} \mathbb{I}^{(h)}_{\{\tilde{g}<0\}}(z^{(i)}). \tag{3.19}$$

To effectively apply the method in practice and eliminate the explicit specification of the parameter δ, an efficient iterative algorithm was proposed in [17]. Let δM be an integer, much smaller than total number of samples M, as the "step size" of the iteration.

- **Initialization**: Estimate the failure probability using the surrogate model via (3.11), and sort $\{|\tilde{g}(z^{(i)})|\}_{i=1}^{M}$ in ascending order.
- **Iteration**: At kth iteration ($k > 0$), do the following.
  - Identify the $(k-1)*\delta M + 1$ to $k*\delta M$ samples in the sorted sequence of $|\tilde{g}|$, re-evaluate these samples by g, and update the probability estimate.
  - Repeat the iteration till the probability estimate is converged.

This is a prediction–correction type of algorithm, where the estimate by the surrogate model $\tilde{g}$ plays the role of the prediction, and the δM samples of $\tilde{g}$ that are "closest" to zero are updated by g to "correct" the failure probability estimate until a convergence is established. More details can be found in [17], where the efficiency gain of the method was demonstrated by several examples.

4. Surrogate-based importance sampling: the main algorithms

In this section we present the main algorithm for computing rare failure probability using surrogate model. The current method utilizes the CE optimization to determine a biasing distribution for the IS integral. And the idea of the hybrid method from [17] is employed to enhance the performance of both the CE optimization and the IS integration. A direct algorithm is presented first, to illustrate the idea of the approach. An iterative algorithm, easy to implement and much more practical, is then presented.

4.1. Direct algorithm

The major components for the current method are the CE optimization procedure to determine a good biasing distribution, and the IS procedure to evaluate the failure integral. We now present details of both procedures and explain how the usage of a surrogate model can improve the efficiency of the traditional CE–IS method.

4.1.1. Surrogate-based cross-entropy optimization

To improve the performance of the CE optimization problem (3.7), we conduct the CE optimization on the surrogate $\tilde{g}$ instead of the true limit state function g. However, to avoid the potentially severe loss of accuracy, pointed out by Li and Xiu [17], we employ the hybrid idea by using information of g when $\tilde{g}$ is close to zero. That is, we seek to solve the following surrogate-based CE optimization problem

$$\tilde{v} = \mathrm{argmax}_v \int \mathbb{I}^{(h)}_{\{\tilde{g}<0\}}(z) W(z; w) \ln(p(z; v)) p(z; w) dz, \tag{4.1}$$

where again $W(z; w) = q(z)/p(z; w)$ is the likelihood ratio and $\mathbb{I}^{(h)}_{\{\tilde{g}<0\}}$ is defined in (3.16). To effectively solve the optimization problem, we employ the multilevel adaptive iteration from [24,5], with the difference in that the algorithm now operates on both the surrogate $\tilde{g}$ and the true model g. The details of the algorithm are as follows. Let $k \geqslant 0$ be the iteration count and $\{p(z; v)\}_{v\in V}$ be the parameterized family of distributions from which the biasing distribution will be determined.

- **Initialization:**
  - Set $k = 0$, and choose the initial parameters v_0 in the family of distributions of $\{p(z; v)\}$ and set $\tilde{v}_0 = v_0$.
  - Choose a not-so-small real number ρ, say $10^{-2} \leqslant \rho \leqslant 10^{-1}$, and let γ_0 be the smallest real number such that, for Z with distribution $p(z; \tilde{v}_0)$,

$$\mathrm{Prob}^{(h)}(\gamma_0; g, \tilde{g}) \geqslant \rho, \tag{4.2}$$

where the operator $\mathrm{Prob}^{(h)}$ is defined in (3.14).

- **Iteration:** At kth iteration, do the following:
  - *Updating of* γ_k. For a fixed $\tilde{v}_{k-1}$, find the smallest real number γ_k such that, for Z with distribution $p(z; \tilde{v}_{k-1})$,

$$\mathrm{Prob}^{(h)}(\gamma_k; g, \tilde{g}) \geqslant \rho. \tag{4.3}$$

  - *Updating of* $\tilde{v}_k$. For fixed $\tilde{v}_{k-1}$ and γ_k, derive $\tilde{v}_k$ from

$$\tilde{v}_k = \text{argmax}_v \int \mathbb{I}^{(h)}_{\{\tilde{g}<\gamma_k\}}(z) W(z;\tilde{v}_{k-1}) \ln(p(z;v)) p(z;\tilde{v}_{k-1}) dz. \tag{4.4}$$

The iteration terminates at certain iteration count $K > 0$, when the γ_K reaches the level of zero, which is the original critical value of failure. This implies that the distribution $p(z;\tilde{v}_K)$ can now generate not-too-rare samples, with probability at least ρ, and hence can be used as a good biasing distribution. Upon convergence we set $\tilde{v} = \tilde{v}_K$ as the solution of (4.1) and the biasing distribution as $p(z) = p(z;\tilde{v})$.

In practice, the stochastic counterpart of the CE optimization is often more useful, and it takes the following form.

$$\hat{\tilde{v}} = \text{argmax}_v \frac{1}{M} \sum_{i=1}^{M} \mathbb{I}^{(h)}_{\{\tilde{g}<0\}}(z^{(i)}) W(z^{(i)};w) \ln(p(z^{(i)};v)), \tag{4.5}$$

where $\{z^{(i)}\}_{i=1}^M$ are samples drawn from the distribution $p(z;w)$, and the sampling with respect to $\mathbb{I}^{(h)}_{\{\tilde{g}<0\}}$ is defined in (3.17).

The multilevel adaptive iteration algorithm used in the traditional CE optimization can now be adapted to the current stochastic formulation.

- **Initialization:**
  - Set $k = 0$, and choose the initial parameters v_0 in the family of distributions of $\{p(z;v)\}$ and set $\hat{\tilde{v}}_0 = v_0$.
  - Choose a not-so-small real number ρ, say $10^{-2} \leqslant \rho \leqslant 10^{-1}$. Let $\left\{z_i^{(0)}\right\}_{i=1}^M$ be samples drawn from the distribution $p(z;\hat{\tilde{v}}_0)$ and find $\hat{\gamma}_0$ such that

$$\widehat{\text{Prob}}^{(h)}(\hat{\gamma}_0; g, \tilde{g}) \geqslant \rho, \tag{4.6}$$

where the operator $\widehat{\text{Prob}}^{(h)}$ is defined in (3.17).

- **Iteration:** At kth iteration, do the following:
  - *Updating of* $\hat{\gamma}_k$. For fixed $\hat{\tilde{v}}_{k-1}$, draw random samples $\{z^{(i)}\}_{i=1}^M$ from the distribution $p(z;\hat{\tilde{v}}_{k-1})$, and find $\hat{\gamma}_k$ such that

$$\widehat{\text{Prob}}^{(h)}(\hat{\gamma}_k; g, \tilde{g}) \geqslant \rho, \tag{4.7}$$

  - *Updating of* $\hat{\tilde{v}}_k$. For fixed $\hat{\tilde{v}}_{k-1}$ and $\hat{\gamma}_k$, derive $\hat{\tilde{v}}_k$ from

$$\hat{\tilde{v}}_k = \text{argmax}_v \frac{1}{M} \sum_{i=1}^{M} \mathbb{I}^{(h)}_{\{\tilde{g}<\hat{\gamma}_k\}}(z^{(i)}) W(z^{(i)}, \hat{\tilde{v}}_{k-1}) \ln(p(z^{(i)};v)), \tag{4.8}$$

where $\{z^{(i)}\}_{i=1}^M$ are the same samples drawn from the distribution $p(z;\hat{\tilde{v}}_{k-1})$. Obviously, this is the stochastic version of (4.4).

The iteration terminates at certain iteration count $K > 0$, when the $\hat{\gamma}_K$ reaches the level of zero. Upon convergence we set $\hat{\tilde{v}} = \hat{\tilde{v}}_K$ as the solution of (4.5), and the biasing distribution as $p(z) = p(z;\hat{\tilde{v}})$.

4.1.2. Surrogate-based importance sampling

Once the optimal parameter $\tilde{v}$ is determined from the surrogate-based CE optimization (4.1), or $\hat{\tilde{v}}$ from its stochastic counterpart (4.5), the biasing distribution $p(z)$ is determined, via either $p(z;\tilde{v})$ or $p(z;\hat{\tilde{v}})$, respectively. We then proceed to estimate the IS integral (3.2). Here we employ the surrogate-based hybrid method by Li and Xiu [17] and compute the rare failure probability by

$$\widetilde{P}_f = \int \mathbb{I}^{(h)}_{\{\tilde{g}<0\}}(z) W(z) p(z) dz, \tag{4.9}$$

where $W(z) = q(z)/p(z)$ is the likelihood ratio. The convergence of this estimate to the true rare probability (3.2) is a trivial extension of the theorem in [17].

Let $\{z^{(i)}\}_{i=1}^M$ be samples from the distribution $p(z)$. Then the stochastic version of (4.9) is

$$\hat{\tilde{P}}_f = \frac{1}{M} \sum_{i=1}^{M} \mathbb{I}^{(h)}_{\{\tilde{g}<0\}}(z^{(i)}) W(z^{(i)}). \tag{4.10}$$

This estimate converges to (4.9) as one increases the sample size M.

4.2. Iterative algorithm

The aforementioned method for both the CE optimization and the IS integral utilizes the hybrid sampling idea of [17]. A major requirement is the specification of the parameter δ, which is used to separate the probability space into two sub-domains. The specification of δ is directly related to the L^p error of the surrogate $\tilde{g}$, and in many practical simulations will not be available. Here we extend the iterative algorithm from [17] to the current CE-IS method. The advantage of the iterative algo-

rithm is that it does not require the explicit knowledge of the parameter δ. Since the iterative algorithm is based on sampling, we will focus on the stochastic version of the CE optimization (4.5) and the IS computation (3.3).

4.2.1. Surrogate-based cross-entropy optimization

Let us consider the stochastic version of the CE optimization (4.5). The same multilevel procedure presented for (4.5) will be retained, except the probability conditions (4.6) and (4.7) and the sampling in the optimization (4.8) are to be computed by the iterative algorithm.

4.2.1.1. Computation of (4.6). We first present the iterative algorithm for (4.6). (Note that (4.7) is the same as (4.6), except with a different parameter value γ_k.) Let $j \geqslant 0$ be an iteration count, $\delta M \ll M$ be an integer denoting the "step size", and $\eta \geqslant 0$ be a small number for the stopping criterion, for the following iteration.

- **Initialization:**
  - Set $j = 0$, $M^{(j)} = 0$, and $S_z^{(j)} = \emptyset$.
  - Generate $\{z^{(i)}\}_{i=1}^{M}$, samples drawn from the distribution $p(z; \hat{\tilde{v}}_0)$. Denote S_z as the sample set.
  - Compute the *surrogate* realizations of the samples to obtain the set $G^{(j)} = \{\tilde{g}(z^{(i)})\}_{i=1}^{M} = \{\tilde{g}(z),\ z \in S_z\}$.
  - Approximate $\hat{\gamma}_0$ by the ρ-quantile, i.e.,

$$\hat{\gamma}_0^{(j)} = G_{\lceil \rho M \rceil}^{(j)}, \tag{4.11}$$

where $\lceil \cdot \rceil$ is the ceiling function.
  - Sort $\{|\tilde{g}(z^{(i)}) - \hat{\gamma}_0|\}_{i=1}^{M}$ in ascending order.
- **Iteration:** At jth iteration ($j > 0$), do the following.
  - Identify the $(M^{(j)} + 1)$ to $(M^{(j)} + \delta M)$ elements in the sorted sequence of $|\tilde{g} - \hat{\gamma}_0|$ and their corresponding sample points z. Denote $\delta S_z^{(j)}$ the set for these samples.
  - Evaluate the original limit state function g at the sample points in the set $\delta S_z^{(j)}$, and let $S_z^{(j)} = S_z^{(j-1)} \cup \delta S_z^{(j)}$.
  - Let $G^{(j)} = \left\{g(z), z \in S_z^{(j)}\right\} \cup \left\{\tilde{g}(z),\ z \in S_z \setminus S_z^{(j)}\right\}$, and find the ρ-quantile so that

$$\hat{\gamma}_0^{(j)} = G_{\lceil \rho M \rceil}^{(j)}, \tag{4.12}$$

  - If $\left|\hat{\gamma}_0^{(j)} - \hat{\gamma}_0^{(j-1)}\right| \leqslant \eta$, exit; if not, let $j \leftarrow j + 1$, $M^{(j)} = M^{(j-1)} + \delta M$, and repeat the iteration.

Note in this procedure it is natural to set $\eta = 0$, since all operations are on samples with finite size. Also, even though the procedure is of iterative nature, the same sequence of random samples are used throughout the iterations.

4.2.1.2. Computation of (4.8). The problem of (4.8) is an optimization problem. To this end, one can adopt any standard optimization algorithm that has been well tested for the standard CE optimization. Hence we will not focus on the particular choice of the optimization algorithm for (4.8). Rather, we focus on the implementation of the sampling part of (4.8), using an iterative algorithm. We now consider the computation of

$$R(\hat{\gamma}_k; v) \triangleq \frac{1}{M} \sum_{i=1}^{M} \mathbb{I}_{\{\tilde{g} < \hat{\gamma}_k\}}^{(h)}(z^{(i)}) U(z^{(i)}; v), \tag{4.13}$$

where $U(z^{(i)};\ v) = W(z^{(i)}) \ln(p(z^{(i)};\ v))$ and $\{z^{(i)}\}_{i=1}^{M}$ are samples drawn from $p(z; \hat{\tilde{v}}_{k-1})$. By using these notations, (4.8) can be written as

$$\hat{\tilde{v}}_k = \operatorname{argmax}_v R(\hat{\gamma}_k; v).$$

The iterative algorithm for computing (4.13) is as follows. Once again, let $j \geqslant 0$ be iteration count, $\delta M \ll M$ be an integer denoting the "step size", and $\eta \geqslant 0$ be a small number for the stopping criterion.

- **Initialization**:
  - Set $j = 0$, $M^{(j)} = 0$.
  - Estimate (4.13) using the surrogate model $\tilde{g}$. That is, for the samples $\{z^{(i)}\}_{i=1}^{M}$, let

$$R^{(j)} = \frac{1}{M} \sum_{i=1}^{M} \mathbb{I}_{\{\tilde{g} < \hat{\gamma}_k\}}(z^{(i)}) U(z^{(i)}; v). \tag{4.14}$$

  - Sort $\{|\tilde{g}(z^{(i)}) - \hat{\gamma}_k|\}_{i=1}^{M}$ in ascending order.
- **Iteration**: At jth iteration ($j > 0$), do the following.
  - Identify the $(M^{(j)} + 1)$ to $(M^{(j)} + \delta M)$ elements in the sorted sequence of $|\tilde{g} - \hat{\gamma}_k|$ and their corresponding sample points z. Denote $\delta S_z^{(j)}$ the set for these samples.

- Evaluate the original limit state function g at the sample points in the set $\delta S_z^{(j)}$.
- Update the estimate for R using the values of g on $\delta S_z^{(j)}$.

$$R^{(j)} = R^{(j-1)} + \frac{1}{M}\sum_{z^{(i)}\in\delta S_z^{(j)}} [-\mathbb{I}_{\{\tilde{g}<\hat{\gamma}_k\}}(z^{(i)}) + \mathbb{I}_{\{g<\hat{\gamma}_k\}}(z^{(i)})]U(z^{(i)};\nu). \tag{4.15}$$

- If $|R^{(j)} - R^{(j-1)}| \leqslant \eta$, exit; if not, let $j \leftarrow j+1$, $M^{(j)} = M^{(j-1)} + \delta M$, and repeat the iteration.

Once again, as a feature of the hybrid sampling method, the same sequence of the random samples are used throughout this iterative algorithm. However, we shall emphasize that when this iterative algorithm is implemented in a chosen optimization algorithm for (4.8), the random samples cannot be re-used during the optimization iteration steps. Each step of the optimization would require a new sequence of random samples to conduct the iterative computation of (4.13).

4.2.2. Surrogate-based importance sampling

Let us now focus on the hybrid sampling estimate (4.10), where $\{z^{(i)}\}_{i=1}^M$ are the samples generated from the biasing distribution $p(z)$ determined by the CE optimization step. The iterative algorithm for the IS will be similar to that of (4.13). In fact, it is essentially a special case of (4.13) with $\hat{\gamma}_k = 0$. Again, let $j \geqslant 0$ be the iteration count, δM be an integer (much) smaller than M for the "step size", and $\eta \geqslant 0$ a small number for the stopping criterion.

- **Initialization**:
  - Set $j = 0$, $M^{(j)} = 0$.
  - Estimate the failure probability (3.2) using the surrogate model $\tilde{g}$. That is, let $\{z^{(i)}\}_{i=1}^M$ be samples drawn from the distribution $p(z)$, for $j = 0$, let

$$P_f^{(j)} = \frac{1}{M}\sum_{i=1}^{M} \mathbb{I}_{\{\tilde{g}<0\}}(z^{(i)})W(z^{(i)}). \tag{4.16}$$

  - Sort $\{|\tilde{g}(z^{(i)})|\}_{i=1}^M$ in ascending order.
- **Iteration**: At jth iteration ($j > 0$), do the following.
  - Identify the $(M^{(j)}+1)$ to $(M^{(j)}+\delta M)$ elements in the sorted sequence of $|\tilde{g}|$ and their corresponding sample points z. Denote $\delta S_z^{(j)}$ the set for these samples.
  - Evaluate the original limit state function g at the sample points in the set $\delta S_z^{(j)}$.
  - Update the failure probability estimate using the values of g on $\delta S_z^{(j)}$.

$$P_f^{(j)} = P_f^{(j-1)} + \frac{1}{M}\sum_{z^{(i)}\in\delta S_z^{(j)}} [-\mathbb{I}_{\{\tilde{g}<0\}}(z^{(i)}) + \mathbb{I}_{\{g<0\}}(z^{(i)})]W(z^{(i)}). \tag{4.17}$$

  - If $\left|P_f^{(j)} - P_f^{(j-1)}\right| \leqslant \eta$, exit; if not, let $j \leftarrow j+1$, $M^{(j)} = M^{(j-1)} + \delta M$, and repeat the iteration.

This is essentially the same algorithm presented in [17], with the difference in the introduction of the likelihood function in the current IS formulation. For more detailed discussion of the algorithm, see [17].

5. Numerical examples

In this section we provide several numerical examples to demonstrate the performance of the new algorithm. For benchmarking purpose, in all examples the evaluation of the true limit state function g is trivial so that we can obtain "numerical exact solution", which is termed as reference solution P_f^r. These reference solutions are obtained by using the traditional CE-IS method, and we consider the following cases.

- "Rare" failure probability with $P_f \sim 10^{-6}$. Here we employ 5000 samples in the CE optimization procedure and 10,000 samples in the final IS estimation. In all the examples the CE optimization converges in three steps. This implies that the simulation costs to obtain the reference solutions are 25,000 samples.
- "Very rare" failure probability with $P_f \sim 10^{-9}$. Here we employ 50,000 samples in the CE optimization procedure and 100,000 samples in the final IS estimation. In all the examples the CE optimization converges in four steps. This implies that the simulation costs to obtain the reference solutions are 300,000 samples.
- "Extremely rare" failure probability with $P_f \sim 10^{-12}$. Here we employ 500,000 samples in the CE optimization procedure and 1,000,000 samples in the final IS estimation. In all examples the CE optimization converges in four steps. This implies that the simulation cost to obtain the reference solutions are 3,000,000 samples.

Obviously, the adjectives "very" and "extremely" are used here only to distinguish the three cases on a relative scale. Extensive numerical tests are conducted to verify that the choices of the sampling size result in failure probability estimates P_f^r that are accurate for at least two significant digits for the first two cases. In the last case of extremely rare probability, our

reference solutions are accurate for at least one significant digit. (More accurate estimation can be obtained by increasing the size of sampling, which is hard to do in the last case. This is, however, not the purpose of our tests. In many practical cases, results with two accurate significant digits are quite adequate.) In all of the examples, it is sufficient to use Gaussian distribution as the parameterized family of distributions in the CE step.

To examine the properties of the new surrogate-based CE-IS algorithm, we employ the same random sequences used in obtaining the reference solutions. By doing so, the new algorithm could in principle re-produce exactly the same results by matching every digits of the reference solutions. We define the relatively error as

$$\text{Error} = \frac{\left|P_f^h - P_f^r\right|}{P_f^r}, \tag{5.1}$$

where P_f^h is our estimate by the hybrid method and P_f^r is the reference solution. Note our examination of error is somewhat different from the traditional way, where the mean and variance of the estimates are usually presented. Since the hybrid method here can be considered as a direct improvement over the traditional CE-IS, we seek to recover precisely the same estimates of CE-IS by using the same sampling sequences. The solution statistics are properties of the CE-IS. They have been studied extensively and are not the focus of this paper.

The simulation costs of the new algorithm will be reported as the total number of samples of the true limit state function g in the entire procedure.

5.1. Ordinary differential equation

We first consider a simple random ODE

$$\frac{du}{dt} = -Zu, \quad u(0) = u_0, \tag{5.2}$$

where the decay rate coefficient Z is a random variable with a prescribed probability distribution $q(z)$. The exact solution of (5.2) is $u(t, Z) = u_0 e^{-Zt}$. Suppose the failure probability is $P_f = \text{Prob}[g(u(t, Z)) < 0]$, where the limit state function g is, for a prescribed value u_d,

$$g(u(t,Z)) = u_d - u(t,Z). \tag{5.3}$$

We first fix $u_0 = 1$, $t = 1$, and assume $Z \sim \mathcal{N}(0,1)$, a Gaussian random variable with mean value of 0 and standard deviation of 1. Let $u_d = 100$, and the reference solution of the failure probability is $P_f^r = 2.06059 \times 10^{-6}$, obtained by the traditional CE-IS algorithm with a total of 25,000 samples.

For the surrogate model $\tilde{g}$, we employ the generalized polynomial chaos (gPC) approximation g_n, where n is the order of the gPC expansion, obtained by solving (5.2) via stochastic Galerkin approach. (For details of the gPC methods, see [31,30].

The results for this rare probability case are tabulated in Table 5.1, with different orders (n) of the gPC expansions. In all cases, the CE optimization converged in at most four steps. The final failure probability estimates are all highly accurate, with almost all errors less than 1%. It can be seen that when the gPC model has reasonable accuracy ($n \geqslant 3$), the number of samples is reduced drastically. Starting with $n \geqslant 7$, only a few hundreds of samples are required to produce extremely accurate estimate of the failure probability. In fact, all simulations with $n \geqslant 11$ can produce the reference solution *exactly*, i.e., with no errors, by using only 500 samples.

The results for the second case, "very rare" probability, are tabulated in Table 5.2. The reference solution $P_f^r = 1.04149 \times 10^{-9}$, which is obtained by fixing $u_d = 400$ and using the traditional CE-IS with a total of 300,000 samples. It is clear that all hybrid estimates are highly accurate—all errors are less than 1%. And again, when the gPC model has reasonable accuracy ($n \geqslant 3$), the number of samples is reduced drastically. For all simulations with $n \geqslant 11$, the current method can produce the extremely rare reference solution *almost exactly* by using only 600 samples.

For the "extremely rare probability" case, we fix $u_d = 1000$ and obtain the reference solution of $P_f = 2.45457 \times 10^{-12}$ with a total of 3,000,000 samples by CE-IS. The results obtained by the new hybrid algorithm are tabulated in Table 5.3. For this case, the failure is so exceedingly small that the lower-order gPC surrogate models with $n \leqslant 3$ lack of sufficient accuracy

Table 5.1
Example 1 with $P_f^r = 2.06059 \times 10^{-6}$. Simulation results by the present algorithm, where n is the gPC expansion order, "CE steps" is the number of iteration steps in the CE optimization, and # is the total number of samples of the true limit state function g required by the present algorithm. All simulations with gPC orders $n > 11$ have zero error with 400 samples.

	n						
	2	3	4	5	7	9	11
CE steps	4	4	3	3	3	3	3
$P_f \times 10^6$	2.04952	2.05621	2.02532	2.09455	2.05687	2.06962	2.06059
Error	0.54%	0.21%	1.71%	1.65%	0.18%	0.44%	0
#	12,200	12,200	5500	2300	600	500	500

Table 5.2
Example 1 with $P_f^r = 1.04149 \times 10^{-9}$. Simulation results by the present algorithm, where n is the gPC expansion order and # is the total number of samples of the true limit state function g required by the present algorithm. All simulations with gPC orders $n > 11$ have error less than 0.01% with 600 samples.

	n						
	2	3	4	5	7	9	11,...
$P_f \times 10^9$	1.04337	1.04297	1.04477	1.05121	1.03488	1.04507	1.04088
Error	0.18%	0.14%	0.31%	0.93%	0.63%	0.34%	<0.05%
#	114,000	111,500	84,800	66,900	15,200	2700	600

Table 5.3
Example 1 with $P_f^r = 2.45457 \times 10^{-12}$. Simulation results by the present algorithm, where n is the gPC expansion order and # is the total number of samples of the true limit state function g required by the present algorithm. All simulations with gPC orders $n \leqslant 3$ are not able to detect failure.

	n						
	4	5	7	9	11	13	15
$P_f \times 10^{12}$	0.913832	1.10337	2.46308	2.46734	2.45654	2.45351	2.45462
Error	62.8%	55.0%	0.35%	0.52%	0.080%	0.043%	0.002%
#	1,081,900	960,800	287,600	60,600	9100	1400	600

to capture the failure. Starting with moderately low order $n = 4$, the failure probability can be captured but with modest simulation saving and low accuracy. The situation dramatically improves when moderately high-order gPC surrogates are used. From order of $n \geqslant 7$ the hybrid method produces virtually no errors and with drastically reduced number of samples. Even with $n = 7$, the number of samples is less than 10%, and at $n = 9$ it is about 2%. Once again, when the surrogate has high order accuracy, e.g., $n = 15$, the hybrid method produces extremely accurate result with only 600 samples.

5.2. Multivariate benchmark

We now consider a multivariate benchmark problem, also considered in [17]. The failure function is defined as

$$g(X) = X_1 + 2X_2 + 2X_3 + X_4 - 5X_5 - 5X_6 + 0.001 \sum_{i=1}^{6} \sin(100X_i), \tag{5.4}$$

where $X_i \sim LN(\mu_i, \sigma_i)$ for $i = 1, 2, \ldots, 6$, are independent log-normal random variables. Let $\{Z_i\}_{i=1}^6$ be i.i.d. $\mathcal{N}(0,1)$ Gaussian random variables. We employ Hermite polynomials to approximate $\{X_i\}_{i=1}^6$, i.e.,

$$X_i \approx X_{n,i}(Z_i) = \sum_{k=0}^{n} c_{i,k} H_k(Z_i), \quad c_{i,k} = e^{\mu_i + \frac{\sigma_i^2}{2}} \frac{\sigma_i^k}{k!}. \tag{5.5}$$

We then construct our gPC surrogate model as

$$\tilde{g}_n(Z) = X_{n,1} + 2X_{n,2} + 2X_{n,3} + X_{n,4} - 5X_{n,5} - 5X_{n,6}. \tag{5.6}$$

A direct approximation of the nonlinear terms involving the sine functions is avoided because of the small coefficient. By doing so the surrogate $\tilde{g}_n$ does not converge to g and will possess a finite error of $O(10^{-3})$ even if n goes to infinity.

The failure probability is defined as $P_f = \text{Prob}(g(Z) < 0)$. We first consider the rare failure probability case. By setting $\mu_1 = 120$, $\sigma_1 = 12$; $\mu_2 = 120$, $\sigma_2 = 12$; $\mu_3 = 120$, $\sigma_3 = 12$; $\mu_4 = 120$, $\sigma_4 = 12$; $\mu_5 = 50$, $\sigma_5 = 5$; $\mu_6 = 40$, $\sigma_6 = 6$, we obtain the reference solution $P_f^r = 2.0250 \times 10^{-6}$, computed by the traditional CE-IS with a total of 25,000 samples.

We perform the surrogate-based CE-IS simulation using $\tilde{g}_n$ of different orders of n. The results are summarized in Table 5.4. In all cases the CE optimization of the present algorithm converged in three steps. Again we obtain highly accurate

Table 5.4
Example 2 with $P_f^r = 2.0250 \times 10^{-6}$. Simulation results by the present algorithm, where n is the gPC expansion order, "CE steps" is the number of iteration steps in the CE optimization, and # is the total number of samples of the true limit state function g required by the present algorithm. All simulations with orders $n \geqslant 4$ have zero error with 400 samples.

	n			
	1	2	3	4,5,...
CE steps	3	3	3	3
$P_f \times 10^6$	1.7115	2.0752	2.0250	2.0250
Error	15.5%	2.48%	0	0
#	5200	1700	600	400

estimates by the present algorithm, with *much reduced* number of samples. For all surrogate models with $n \geqslant 4$, the present algorithm can exactly reproduce the reference solution with only 400 samples. We emphasize again that here the surrogate models $\tilde{g}_n$ at higher n do not converge to g.

Next we consider the "very rare" failure probability case, by defining $P_f = \text{Prob}(g(Z) < -80)$. The reference solution is now $P_f^r = 4.58108 \times 10^{-9}$, which is obtained by CE-IS with a total of 300,000 samples. The results by the hybrid method are tabulated in Table 5.5. We notice that at the lowest order of $n = 1$, the result is off by more than 50%. This should not be surprising because at such a low order, the surrogate is a linear approximation of the true limit state function and is not sufficient to capture the extremely small tail behavior of the limit state function. However, starting even at very low order of second order, the accuracy of the hybrid estimate drastically improves. From the third order and so on, the hybrid algorithm can reproduce the reference solution with extremely high accuracy but with only about 500 samples. This efficiency gain, compared with the 300,000 samples required to obtained the reference solution, is notable.

We then consider the "extremely rare" case, by defining $P_f = \text{Prob}(g(Z) < -160)$. The reference solution is now $P_f^r = 6.33919 \times 10^{-12}$, obtained by CE-IS with a total of 3,000,000 samples. The results by the hybrid method are tabulated in Table 5.6. The performance here is in fact better than the same extremely small case in Example 1. The hybrid method produces highly accurate results at almost all orders, and with zero error starting at $n = 6$ with just a few hundreds samples.

5.3. System of ordinary differential equations

We now consider the following ODE system:

$$\begin{aligned} \dot{y_1} &= -(Z_1 + Z_2)y_1, \\ \dot{y_2} &= Z_4 y_3 + Z_5 y_4, \\ \dot{y_3} &= Z_1 y_1 + Z_3 y_4 - Z_3 y_3 - Z_4 y_3, \\ \dot{y_4} &= Z_2 y_1 - Z_3 y_4 + Z_3 y_3 - Z_5 y_4, \end{aligned} \tag{5.7}$$

with initial condition

$$(y_1, y_2, y_3, y_4) = (1, 0, 0, 0), \quad t = 0.$$

Here each component of $Z = (Z_1, \ldots, Z_5)$ is a Gaussian random variable with distribution $\mathcal{N}(1, 0.1^2)$. The surrogate model is obtained by solving the system of equations by gPC based stochastic Galerkin method using Hermite polynomial as basis. (See [30] for details).

For the rare probability case, the limit state function is defined as $g(Z) = 0.5 - y_2(t = 1)$. The reference solution obtained by the traditional CE-IS method is $P_f^r = 2.96944 \times 10^{-6}$, where 25,000 samples were employed. We examine the performance of the hybrid algorithm with different orders of expansion, $n = 2, 3, 4, 5$. In all cases the surrogate-based CE optimization converged in three iterations, and produced highly accurate failure probability estimates with just a few hundreds samples. The results are tabulated in Table 5.7.

We now consider the "very rare" failure probability case. By letting $g(Z) = 0.523 - y_2(t = 1)$, the reference solution of the failure probability is $P_f^r = 8.77715 \times 10^{-9}$, obtained by CE-IS with 300,000 samples. The results obtained by the iterative hybrid method are tabulated in Table 5.8. Once again, we observe that even with a surrogate model at relatively low order, the hybrid method can produce extremely accurate results at significantly lower computational cost.

Finally we consider the "extremely rare" failure probability case. By letting $g(Z) = 0.55 - y_2(t = 1)$, the reference solution of the failure probability is $P_f^r = 1.43037 \times 10^{-12}$, obtained by CE-IS with 3,000,000 samples. The results obtained by the iterative hybrid method are tabulated in Table 5.9. Similar to the very rare case, in this even rarer case the performance of the hybrid algorithm is notable. It can produce extremely accurate results at significantly reduced computational cost.

5.4. Cantilever beam example

We now consider a cantilever beam problem as illustrated in Fig. 5.1, with width w, height t, length L, and subject to transverse load Y and horizontal load X. This is a well adopted benchmark problem in engineering, where failure is determined by the following limit state function

Table 5.5
Example 2 with $P_f^r = 4.58108 \times 10^{-9}$. Simulation results by the present algorithm, where n is the gPC expansion order and # is the total number of samples of the true limit state function g required by the present algorithm. All simulations with orders $n \geqslant 6$ have zero error with 500 samples.

	n					
	1	2	3	4	5	6,...
$P_f \times 10^9$	1.95476	4.2667	4.5550	4.57994	4.58122	4.58108
Error	57.33%	6.86%	0.57%	0.02%	0.03%	0
#	59,000	18,500	600	1300	500	500

Table 5.6
Example 2 with $P_f^r = 6.33919 \times 10^{-12}$. Simulation results by the present algorithm, where n is the gPC expansion order and # is the total number of samples of the true limit state function g required by the present algorithm. All simulations with orders $n \geqslant 7$ have zero error with 500 samples.

	n					
	2	3	4	5	6	7,...
$P_f \times 10^{12}$	5.57878	6.20678	6.31494	6.33602	6.33919	6.33919
Error	12.0%	2.09%	0.38%	0.05%	0	0
#	345,800	67,300	11,900	2400	800	500

Table 5.7
Example 3 with $P_f^r = 2.96944 \times 10^{-6}$. Simulation results by the present algorithm, where n is the gPC expansion order, "CE steps" is the number of iteration steps in the CE optimization, and # is the total number of samples of the true limit state function g required by the present algorithm.

	n			
	2	3	4	5
CE steps	3	3	3	3
$P_f \times 10^6$	2.93785	2.97165	2.96944	2.96944
Error	1.06%	0.07%	0	0
#	1200	600	500	400

Table 5.8
Example 3 with $P_f^r = 8.77715 \times 10^{-9}$. Simulation results by the present algorithm, where n is the gPC expansion order and # is the total number of samples of the true limit state function g required by the present algorithm.

	n			
	2	3	4	5
$P_f \times 10^9$	8.78170	8.76889	8.77935	8.77427
Error	0.05%	0.09%	0.03%	0.03%
#	11,000	3000	1100	700

Table 5.9
Example 3 with $P_f^r = 1.43037 \times 10^{-12}$. Simulation results by the present algorithm, where n is the gPC expansion order and # is the total number of samples of the true limit state function g required by the present algorithm.

	n			
	2	3	4	5
$P_f \times 10^{12}$	1.44190	1.43147	1.43003	1.43045
Error	0.81%	0.08%	0.02%	0.01%
#	125,100	42,500	8100	1700

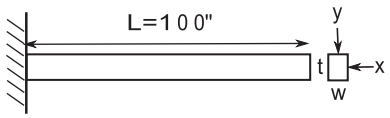

Fig. 5.1. Example 4. Schematic illustration of a cantilever beam subject to horizontal and vertical loads.

$$g = D_o - \frac{4L^3}{Ewt}\sqrt{\left(\frac{Y}{t^2}\right)^2 + \left(\frac{X}{w^2}\right)^2}.$$

For the case considered here, the parameter setting is $w = 2.6535$, $t = 3.9792$, $L = 100$, with the elastic modulus $E \sim \mathcal{N}(29 \times 10^6,\ 25 \times 10^{12})$. The external loads are $X \sim \mathcal{N}(700, 100^2)$ and $Y = 500$. With the parameter $D_o = 6$, the failure probability is on the order of $O(10^{-6})$. The reference solution obtained by the standard CE-IS simulation $P_f^r = 4.97899 \times 10^{-6}$, where a total of 25,000 samples are used. For the surrogate-based approach, we employ multivariate Hermite gPC expansion to approximate g, with different orders (n) of the polynomial expansion. The expansion coefficients are determined by discrete projection method using sparse grids as quadrature rule. (Details of this approach can be found in

Table 5.10
Example 4 with $P_f^r = 4.97899 \times 10^{-6}$. Simulation results by the present algorithm, where n is the gPC expansion order, "CE steps" is the number of iteration steps in the CE optimization, and # is the total number of samples of the true limit state function g required by the present algorithm.

	n				
	3	5	7	11	15
CE steps	3	3	3	3	3
$P_f \times 10^6$	5.1044	4.9419	4.9684	5.0069	4.9802
Error	2.52%	0.74%	0.21%	0.56%	0.02%
#	10,600	2300	900	900	700

[30, Ch. 7]) We conducted simulations of the orders from $n = 1$ to $n = 15$ to thoroughly examine the performance of the new algorithm. Some of the representative results are tabulated in Table 5.10. Again we observe a significant reduction of the number of samples in the current algorithm. Note the number of samples appear to be larger, compared to the previous examples. This is largely because of the nature of the problem. It is quite well recognized that this example is relatively difficulty to simulate, and the traditional response surface method performs poorly. However, here the new algorithm can produce accurate solutions with much smaller number of samples.

Unlike in the previous examples, in this example we did not pursue even smaller failure probability. The reason is because this particular problem in fact does not possess infinitely long tails, due to the square root operators. Therefore it is highly non-trivial to construct a test case, with properly chosen parameter values, so that its failure probability is 10^{-9} or less.

6. Summary

An efficient method for computing rare failure probability is presented in this paper. The method combines the idea of surrogate-based hybrid method in [17] with importance sampling (IS), particularly with the cross-entropy (CE) method. The resulting algorithm is an iterative scheme and can be easily implemented. Furthermore, it does not require approximations or manipulations of the underlying problem, and thus is a quite general approach. In several numerical tests, the method demonstrates significant efficiency gain—only a few hundreds samples are sufficient to capture failure probability as small as 10^{-6}, 10^{-9} or even 10^{-12}. Future research will involve numerical analysis of its properties, more extensive tests to examine its performance for complex systems, and investigation of its combination with other techniques for rare failure probability computations.

Acknowledgment

This work is in part supported by AFOSR, DOE, NNSA, NSF.

References

[1] S.K. Au, J. Beck, Estimation of small failure probabilities in high dimensions by subset simulation, Prob. Eng. Mech. 16 (2001) 263–277.
[2] Z.I. Botev, D.P. Kroese, T. Taimre, Generalized cross-entropy methods with applications to rare-event simulation and optimization, Simulation 83 (11) (2007) 785–806.
[3] C.G. Bucher, U. Bourgund, A fast and efficient response surface approach for structural reliability problems, Struct. Saf. 7 (1990) 57–66.
[4] J.A. Bucklew, Introduction to Rare Event Simulation, Springer-Verlag, 2004.
[5] P.-T. de Boer, D.P. Kroese, S. Mannor, R.Y. Rubinstein, A tutorial on cross-entropy method, Ann. Oper. Res. 134 (2005) 19–67.
[6] A. Der Kiureghian, H.Z. Lin, S.J. Hwang, Second-order reliability approximations, J. Eng. Mech. 113 (8) (1987) 1208–1225.
[7] S. Engelund, R. Rackwitz, A benchmark study on importance sampling techniques in structural reliability, Struct. Saf. 12 (1993) 255–276.
[8] L. Faravelli, Response surface approach for reliability analysis, J. Eng. Mech. 115 (12) (1989) 2763–2781.
[9] B. Fiessler, H.-J. Neumann, R. Rackwitz, Quadratic limit states in structural reliability, J. Eng. Mech. 105 (4) (1979) 661–676.
[10] N. Gayton, J.M. Bourinet, M. Lemaire, CQ2RS: a new statistical approach to the response surface method for reliability analysis, Struct. Saf. 25 (2003) 99–121.
[11] S. Gupta, C.S. Manohar, An improved response surface method for the determination of failure probability and importance measure, Struct. Saf. 26 (2004) 123–139.
[12] M. Hohenbichler, S. Gollwitzer, W. Kruse, R. Rackwitz, New light on first- and second-order reliability methods, Struct. Saf. 4 (1987) 267–284.
[13] M. Hohenbichler, R. Rackwitz, First-order concepts in system reliability, Struct. Saf. 1 (1983) 177–188.
[14] T. Homem-de-Mello, A study on the cross-entropy method for rare-event probability estimation, Inform. J. Comput. 19 (3) (2007) 381–394.
[15] A.M. Johansen, P. Del Moral, A. Doucet, Sequential Monte Carlo samplers for rare events, in: Proceedings of the 6th International Workshop on Rare Event Simulation, Bamberg, Germany, 2006, pp. 256–267.
[16] S. Kullback, R.A. Leibler, On information and sufficiency, Ann. Math. Stat. 22 (1) (1951) 79–86.
[17] J. Li, D. Xiu, Evaluation of failure probability via surrogate models, J. Comput. Phys. 229 (2010) 8966–8980.
[18] M. Paffrath, U. Wever, Adapted polynomial chaos expansion for failure detection, J. Comput. Phys. 226 (1) (2007) 263–281.
[19] R. Pulch, Polynomial chaos for the computation of failure probabilities in periodic problems, in: J. Roos, L. Costa (Eds.), Scientific Computing in Electrical Engineering SCEE 2008, 2010.
[20] R. Rackwitz, Reliability analysis – a review and some perspectives, Struct. Saf. 23 (2001) 365–395.
[21] M.R. Rajashekhar, B.R. Ellingwood, A new look at the response surface approach for reliability analysis, Struct. Saf. 12 (1993) 205–220.
[22] B.D. Ripley, Stochastic Simulation, John Wiley & Sons, New York, 1987.
[23] R.Y. Rubinstein, Optimization of computer simulation models with rare events, Eur. J. Oper. Res. 99 (1997) 89–112.
[24] R.Y. Rubinstein, D.P. Kroese, The Cross-entropy Method, Springer Science + Business Media, Inc., New York, NY, 2004.

[25] R.Y. Rubinstein, D.P. Kroese, Simulation and the Monte Carlo Method, second ed., Wiley-Interscience, 2007.
[26] G.I. Schueller, H.J. Pradlwarter, P.S. Koutsourelakis, A critical appraisal of reliability estimation procedures for high dimensions, Prob. Eng. Mech. 19 (2004) 463–474.
[27] G.I. Schueller, R. Stix, A critical appraisal of methods to determine failure probabilities, Struct. Saf. 4 (1987) 293–309.
[28] R. Srinivasan, Importance Sampling: Applications in Communications and Detection, Springer-Verlag, 2002.
[29] F.S. Wong, Slope reliability and response surface method, J. Geol. Eng. 111 (1985) 32–53.
[30] D. Xiu, Numerical Methods for Stochastic Computations, Princeton University Press, Princeton, New Jersey, 2010.
[31] D. Xiu, G.E. Karniadakis, The Wiener-Askey polynomial chaos for stochastic differential equations, SIAM J. Sci. Comput. 24 (2) (2002) 619–644.