A Nonlinear Wavelet Density-based Importance Sampling for Reliability Analysis

Wei Wang
Professor, School of Civil Engineering, Harbin Institute of Technology, Harbin, China

Hongzhe Dai
Associate Professor, School of Civil Engineering, Harbin Institute of Technology, Harbin, China

ABSTRACT: Importance sampling is a commonly used variance reduction technique for estimating reliability of a structural system. The performance of importance sampling is critically dependent on the choice of the sampling density. For the commonly used adaptive importance sampling method, the construction of the sampling density relies on the kernel-based density estimation. However, the choice of the initial bandwidth of the local windows may heavily affect the accuracy of the kernel method, particularly when the number of samples is not very large. To overcome this difficulty, this study develops a new adaptive importance sampling method based on nonlinear wavelet thresholding density estimator. The method utilizes the adaptive Markov chain simulation to generate samples that can adaptively populate the important region. The importance sampling density is then constructed using nonparametric wavelet density to implement the importance sampling. The methods takes advantage of the attractive properties of the Daubechies' wavelet family (e.g., localization, various degrees of smoothness, and fast implementation) to provide good density estimations. Compared with the kernel density estimator, the nonlinear wavelet thresholding density estimator has a high degree of flexibility in terms of convergence rate and smoothness. Moreover, the choice of the initial parameters slightly affects the accuracy of the method. Two examples are given to demonstrate the proposed method.

1. INTRODUCTION

In the structural reliability theory, importance sampling method is widely used to calculate the probability of failure, expressed as

$$P_f = \int_{\mathbb{R}^d} I[g(\mathbf{x}) \le 0] \frac{f_{\mathbf{x}}(\mathbf{x})}{h(\mathbf{x})} h(\mathbf{x}) \mathrm{d}\mathbf{x} \qquad (1)$$

where $f_{\mathbf{x}}(\mathbf{x})$ is the joint probability density function of the d -dimensional random variables $\mathbf{X}=(X_1,\ldots,X_d)$, $I[g(\mathbf{x}) \le 0]$ is the indicator function for the failure event $g(\mathbf{x}) \le 0$, and $h(\boldsymbol{x})$ is the importance sampling density function(Balesdent et al. 2013). There exits an optimal importance sampling density function $h_{\mathrm{opt}}(\mathbf{x})$ (Melchers 1989).

$$h_{\mathrm{opt}}(\mathbf{x}) = I[g(\mathbf{x}) \le 0] f_{\mathbf{x}}(\mathbf{x}) / P_f \qquad (2)$$

However, the optimal importance sampling function is unknown as it involves the unknown P_f . Therefore nonparametric density estimation techniques are adopted to construct $h_{\mathrm{opt}}(\boldsymbol{x})$. This approach first generates samples that are distributed asymptotically according to the optimal importance sampling density. Then nonparametric density estimation techniques are used to construct the importance sampling density. Ang was the first one who simulated samples in the failure region by Monte Carlo sampling and used these samples to construct the importance sampling density based on kernel method (Ang et al. 1992). Au and Beck used Markov chain simulation to construct a Markov chain whose target distribution is the optimal importance sampling function. Based on the samples generated, the sampling density is constructed using the adaptive kernel

density estimation (Au and Beck 1999). Kurtz and Song proposed to construct the near-optimal importance sampling density by using Kullback–Leibler cross entropy coupled with a Gaussian mixture kernel (Kurtz and Song 2013). Dai established an adaptive importance sampling method on the basis of fast Gaussian transform kernel density (Dai et al. 2011). In sum, most of the current practice for constructing the importance sampling density relies on the kernel-based density estimation. And the choice of the kernel function and the initial bandwidth of the local windows may heavily affect the accuracy of the kernel method, particularly when the number of samples is not very large (Silverman. 1986).

As a nonparametric density estimation method with good localization and asymptotic property, the wavelet density estimation outperforms the kernel density estimator in terms of global approximation (Hardle et al. 1997). This paper presents a new wavelet thresholding density-based adaptive importance sampling scheme. The proposed method involves the generation of samples that can adaptively populate the important region by the adaptive Markov chain simulation, and the construction of importance sampling density by nonlinear wavelet thresholding density estimator, followed by importance sampling to evaluate P_f. Numerical examples are given to demonstrate the application and efficiency of the proposed method.

2. SAMPLE GENERATION: ADAPTIVE MARKOV CHAIN SIMULATION

Since the closed-form expression for the optimal importance sampling function is unknown, the direct Monte Carlo sampling procedure cannot be used to generate the samples. Therefore, the adaptive Markov chain simulation is proposed to generate the samples that cover the region of most interest (Metropolis et al. 1953).

Let $\pi()$ be the target distribution density function. Suppose that at time $t-1$ the states $(\mathbf{x}_0,\ldots,\mathbf{x}_{t-1})$ have been sampled, a candidate point $\mathbf{y}$ is then sampled from a proposal distribution $q(\cdot\,|\,\mathbf{x}_0,\ldots,\mathbf{x}_{t-1})$. The candidate point $\mathbf{y}$ is accepted with a probability

$$\alpha = \min\left(1, \frac{\pi(\mathbf{y})}{\pi(\mathbf{x}_{t-1})}\right), \tag{3}$$

and is rejected with the remaining probability $(1-\alpha)$, i.e., $\mathbf{x}_t = \mathbf{x}_{t-1}$.

The proposal distribution can be chosen as a multivariate normal distribution with a mean at the point $\mathbf{x}_{t-1}$ and a covariance matrix $\boldsymbol{C}_t$ given by

$$\boldsymbol{C}_t = S_d \mathrm{Cov}(\mathbf{x}_0,\ldots,\mathbf{x}_{t-1}) + S_d \varepsilon \mathrm{I}_d \ , \tag{4}$$

where the subscript d represents the dimension of the random variable, S_d is a scaling parameter which was suggested as $S_d = 2.4^2/d$ (Gelman AG 1996), ε is a positive constant that can be chosen very small, I_d denotes a d-dimensional identity matrix, and $\mathrm{Cov}(\mathbf{x}_0,\ldots,\mathbf{x}_{t-1})$ is the empirical covariance matrix for the samples $(\mathbf{x}_0,\ldots,\mathbf{x}_{t-1})$ (Haario et al. 2001). To start the adaptive Markov chain simulation, an arbitrary, strictly positive definite initial covariance matrix $\boldsymbol{C}_0$ is selected using prior knowledge. Note that $\boldsymbol{C}_{t+1}$ and $\bar{\mathbf{x}}_t$ satisfy the recursion equations:

$$\begin{aligned} \boldsymbol{C}_{t+1} &= \frac{t-1}{t}\boldsymbol{C}_t + \\ &\frac{S_d}{t}\left(\mathbf{t}\bar{\mathbf{x}}_{t-1}\bar{\mathbf{x}}_{t-1}^{\mathrm{T}} - (t+1)\bar{\mathbf{x}}_t\bar{\mathbf{x}}_t^{\mathrm{T}} + \mathbf{x}_t\mathbf{x}_t^{\mathrm{T}} + \varepsilon \mathrm{I}_d\right) \end{aligned} \tag{5}$$

and

$$\bar{\mathbf{x}}_t = \frac{t}{t+1}\bar{\mathbf{x}}_{t-1} + \frac{1}{t+1}\mathbf{x}_t \ . \tag{6}$$

where $\bar{\mathbf{x}}_k = 1/(k+1)\sum_{i=0}^{k}\mathbf{x}_i$.

The aforementioned adaptive Markov chain simulation lends itself to simulating a sample from $h_{opt}()$ by assuming $h_{opt}()$ is the target distribution of the adaptive Markov chain. Suppose that the Markov chain is in state $\mathbf{x}_{t-1}$, then a can-

didate point $\mathbf{y}$ is generated from the Gaussian proposal distribution with a mean of $\mathbf{x}_{t-1}$ and a covariance matrix $\boldsymbol{C}_t$, which is computed recursively from Eq.(5). The acceptance probability α is given by:

$$\alpha = \min\left(1, \frac{\mathbf{1}\left[g(\mathbf{y}) \le 0\right] f_{\mathbf{x}}(\mathbf{y})}{f_{\mathbf{x}}(\mathbf{x}_{t-1})}\right). \qquad (7)$$

Note that the evaluation of Eq.(7) only requires the ratio of the target distribution between consecutive states, and P_f is not needed.

The first sample in the failure domain can be generated as a point simulated according to $h_{opt}()$, or be assigned more efficiently based on engineering judgment (Au and Beck 1999). Although the adaptive Metropolis algorithm is non-Markovian, it has been shown that it has the right ergodic properties, and converges correctly to the target distribution. It has also been shown that the adaptive Metropolis algorithm often achieves a faster convergence than the classical Metropolis algorithm (Haario et al. 2001).

3. SAMPLING DENSITY ESTIMATION: WAVELET EXPANSION

3.1. Multiresolution wavelet analysis and Wavelet density estimation Methodology

Wavelets involve a new family of basis functions that can be used to approximate or express other functions. This section gives a brief introduction to the basic concept of multiresolution analysis and wavelet expansions. Detailed information can be found, for example, in (Chui and K. 1992; Daubechies 1992) .

We introduce wavelets using the multiresolution framework developed in (Mallat 1989). Let $\mathbb{R}$ and $\mathbb{Z}$ denote the set of real numbers and integers, respectively. $L_2(\mathbb{R})$ denotes the set of square-integrable one-dimensional function $f(x)$. The wavelet representation approximates any function $f \in L_2(\mathbb{R})$ by a sequence of functions f_j which are smoother than f and which can be characterized by their sampling on the lattice $2^{-j}\mathbb{Z}$. A multiresolution analysis (or approximation) of $L_2(\mathbb{R})$ consists of a nested sequence $\{V_j, j \in \mathbb{Z}\}$ of closed subspaces $L_2(\mathbb{R})$ and there also exists a function $\phi \in V_0$ such that the sequence $\{\phi(x-k), k \in \mathbb{Z}\}$ is an orthonormal basis of the space V_0. The function ϕ is the scaling function. Define

$$\phi_{j,k}(x) = 2^{j/2}\phi\left(2^j x - k\right), \qquad (8)$$

The variable j determines the amount of variable scaling or dilation, and k represents shift or translation. The family $\{\phi_{j,k}(x), k \in \mathbb{Z}\}$ spans the jth scale of the multiresolution analysis and forms an orthonormal basis for V_j .When the scale increases from j to $j+1$, the approximation at the finer level is obtained by adding some details about f. These details can be modeled at the scale j by the orthogonal complement of V_j in V_{j+1}. Let W_j be this orthogonal complement of V_j in V_{j+1}, we get another sequence $\{W_j, j \in \mathbb{Z}\}$ of closed mutually orthogonal subspace of $L_2(\mathbb{R})$.Let

$$\varphi_{j,k}(x) = 2^{j/2}\varphi\left(2^j x - k\right), \qquad (9)$$

then the family $\{\phi_{j,k}(x), k \in \mathbb{Z}\}$ is an orthogonal basis for W_j and is an orthonormal basis for $L_2(\mathbb{R})$. The spaces $\{W_j, j \in \mathbb{Z}\}$ decompose the function into its smooth and detail parts. Any function $f \in L_2(\mathbb{R})$ can be represented in a wavelet series as

$$f(x) = \sum_{k\in\mathbb{Z}} c_{j_0,k}\phi_{j_0,k}(x) + \sum_{j=j_0}^{\infty}\sum_{k\in\mathbb{Z}} d_{j,k}\varphi_{j,k}(x). \qquad (10)$$

The range of k will be discussed later in Section 3.3. The first part of Eq. (10) is the projection of f on the coarse approximation space

V_{j_0}, and the oscillating features are approximated in fine details by the second part. The scaling coefficients $c_{j,k}$ and wavelet coefficients $d_{j,k}$ are defined as

$$\begin{aligned} c_{j,k} &= \left\langle f, \phi_{j,k} \right\rangle = \int f(x)\phi_{j,k}(x)\,dx \\ d_{j,k} &= \left\langle f, \varphi_{j,k} \right\rangle = \int f(x)\varphi_{j,k}(x)\,dx \end{aligned} \tag{11}$$

Let $X_1,\ldots,X_N$ be independent identically distributed samples from X with a probability density function f. Using an orthogonal wavelet basis, the wavelet representation of f is given by Eq.(10) The scaling coefficients $c_{j,k}$ and wavelet coefficients $d_{j,k}$ can be estimated by

$$\begin{aligned} \hat{c}_{j,k} &= \frac{1}{N}\sum_{i=1}^{N}\phi_{j,k}(X_i), \\ \hat{d}_{j,k} &= \frac{1}{N}\sum_{i=1}^{N}\varphi_{j,k}(X_i). \end{aligned} \tag{12}$$

And one can estimate f by

$$\hat{f}_{j_1}(x) = \sum_k \hat{c}_{j_0,k}\phi_{j_0,k}(x) + \sum_{j_0}^{j_1}\sum_k \hat{d}_{j,k}\varphi_{j,k}(x), \tag{13}$$

where $j_1 \geq j_0$. Nonlinear wavelet thresholding density estimator can be yeilded by adopting thresholding shrinkage method (Donoho et al. 1996),

$$\hat{f}^*(x) = \sum_k \hat{c}_{j_0,k}\phi_{j_0,k}(x) + \sum_{j_0}^{j_1}\sum_k \hat{d}^*_{j,k}\varphi_{j,k}(x), \tag{14}$$

where $\hat{d}^*_{j,k}$ denotes the wavelet coefficients after thresholding operation, and the soft thresholding function can be chosen as follows(Walter and Shen 1994)

$$d^*_{j,k} = \operatorname{sign}\left(\hat{d}_{j,k}\right)\left(\hat{d}_{j,k} - \lambda\right)_+ \tag{15}$$

where $x_+ = \max(x,0)$, and thresholding value λ is given by(Mallat 1989)

$$\lambda = (0.6 \sim 0.8)\max_{j,k}\left|\hat{d}_{j,k}\right| \tag{16}$$

3.2. Multivariate wavelet density estimation

The one-dimensional multiresolution analysis described above can be readily extended to higher dimensional case using the tensor product of one dimensional multiresolution analysis (Vidakovic 1999). The resulting d-dimensional multiresolution analysis corresponds to one d-variate scaling function

$$\phi(x_1,\ldots,x_d) = \prod_{i=1}^{d}\phi_{(i)}(x_i), \tag{17}$$

and d-dimensional wavelet functions $\varphi^{(l)}(x)$

$$\varphi^{(l)}(x_1,\ldots,x_d) = \prod_{i=1}^{d}\xi_{(i)}(x_i), \tag{18}$$

with $\xi = \phi$ or φ, but not all $\xi = \phi$.Any function $f \in L_2(\mathbb{R})$ can be represented as

$$f(\mathbf{x}) = \sum_{\mathbf{k}} c_{j_0,\mathbf{k}}\phi_{j_0,\mathbf{k}}(\mathbf{x}) + \sum_{j=j_0}^{\infty}\sum_{\mathbf{k}}\sum_{l=1}^{2^d-1} d^{(l)}_{j,\mathbf{k}}\varphi^{(l)}_{j,\mathbf{k}}(\mathbf{x}), \tag{19}$$

where $\mathbf{x} = (x_1,\ldots,x_d)$, $\mathbf{k} = (k_1,\ldots,k_d) \in \mathbb{Z}^d$,and

$$\begin{aligned} \phi_{j_0,\mathbf{k}}(\mathbf{x}) &= 2^{j_0 d/2}\prod_{i=1}^{d}\phi_{(i)}\left(2^{j_0}x_i - k_i\right), \\ \varphi_{j,\mathbf{k}}(\mathbf{x}) &= 2^{jd/2}\prod_{i=1}^{d}\xi_{(i)}\left(2^{j}x_i - k_i\right), \end{aligned} \tag{20}$$

with $\xi = \phi$ or φ, but not all $\xi = \phi$.

Multivariate wavelet density estimation is straightforward generalization of the aforementioned univariate density estimation. Let $\left\{\mathbf{X}^{(i)} : i = 1,\ldots,N\right\}$ be a sample of d-dimensional random variables $\mathbf{X}$ with probability density function f, and $\mathbf{X}^{(i)} = \left(X_1^{(i)},\ldots,X_d^{(i)}\right)$, the wavelet estimator of f is

$$\hat{f}(\mathbf{x}) = \sum_{\mathbf{k}} \hat{c}_{j_0,\mathbf{k}} \phi_{j_0,\mathbf{k}}(\mathbf{x}) + \sum_{j=j_0}^{j_1} \sum_{\mathbf{k}} \sum_{l=1}^{2^d-1} \hat{d}_{j,\mathbf{k}}^{(l)} \varphi_{j,\mathbf{k}}^{(l)}(\mathbf{x}), \tag{21}$$

where $\mathbf{x} = (x_1, \ldots, x_d)$, $\mathbf{k} = (k_1, \ldots, k_d)$, and $\phi_{j_0,\mathbf{k}}$ and $\varphi_{j_0,\mathbf{k}}^{(l)}$ are given by Eq.(20). The empirical scaling coefficient and wavelet coefficient are computed as

$$\begin{aligned} \hat{c}_{j,k} &= N^{-1} \sum_{i=1}^{N} \phi_{j,k}\left(\mathbf{X}^{(i)}\right) \\ &= N^{-1} \sum_{i=1}^{N} 2^{jd/2} \prod_{m=1}^{d} \phi_{(m)}\left(2^j X_m^{(i)} - k_m\right) \\ \hat{d}_{j,k}^{(l)} &= N^{-1} \sum_{i=1}^{N} \phi_{j,k}^{(l)}\left(\mathbf{X}^{(i)}\right) \\ &= N^{-1} \sum_{i=1}^{N} 2^{jd/2} \prod_{m=1}^{d} \xi_{(m)}\left(2^j X_m^{(i)} - k_m\right) \end{aligned} \tag{22}$$

where $\xi = \phi$ or φ, but not all $\xi = \phi$. Note that the d-dimensional scaling function is the product of the scaling functions on each dimension. Similar threshold shrinkage operation as Eq.(15) can be applied to wavelet coefficients $\hat{d}_{j,k}^{(l)}$ to obtain the nonlinear wavelet density estimator.

3.3. Computation of the wavelet estimator

There are several practical considerations when using wavelets for density estimation: (1) selecting wavelet family, and (2) finding the range of j_1 and k for $\phi_{j,k}(x)$ and $\varphi_{j,k}(x)$ in Eq.(10). j_1 can be determined by the scalogram of the density f,

$$\prod_{f}(j) = \sum_{k} \left| \langle f, \varphi_{j,k} \rangle \right|^2 . \tag{23}$$

The scalogram describes the energy distribution at various scale j of the density f. The scalogram can be obtained by empirical scalogram

$$\prod(j) = \sum_{l=1}^{2^d-1} \sum_{k} \left| \hat{d}_{j,k}^{(l)} \right|^2 \tag{24}$$

The optimal level j_1 is the smaller one at which the energy distribution of the density f at two successive scales increases exponentially.

Another issue is to determine the range of the translation index k for both of the scaling function and the wavelet function. For Db#q, the support of $\phi_{j,k}(x) = 2^{j/2}\phi\left(2^j x - k\right)$ is $[0, 2q-1]$, therefore,

$$\frac{k}{2^j} \le x \le \frac{2q-1+k}{2^j} . \tag{25}$$

Assume that the range of sample is the interval $[a,b]$, one can calculate the values of k for which the support of corresponding functions $\phi_{j,k}$ intersects $[a,b]$. Therefore, the range of k is

$$\lceil a2^j \rceil - 2q + 1 \le k \le \lfloor b2^j \rfloor, \tag{26}$$

in which $\lceil y \rceil$ denotes the smallest integer which is larger than or equal to y and $\lfloor y \rfloor$ is the largest integer which does not exceed y. For the multivariate wavelet density estimator, if the range of the data is $[a_1,b_1] \times \cdots \times [a_d,b_d]$, the range of $\mathbf{k} = (k_1, \ldots, k_d)$ is

$$\lceil a_i 2^j \rceil - 2q + 1 \le k_i \le \lfloor b_i 2^j \rfloor . \tag{27}$$

Likewise, the range of $\mathbf{k} = (k_1, \ldots, k_d)$ for the wavelet functions is

$$\lceil a_i 2^j \rceil - q \le k_i \le \lfloor b_i 2^j \rfloor + q + 1 \tag{28}$$

4. NONLINEAR WAVELET THRESHOLDING DENSITY-BASED ADAPTIVE IMPORTANCE SAMPLING

The procedure of the proposed methodology can be summarized as follows.

Step 1: Generate N samples, whose target distribution is the $h_{\text{opt}}(\boldsymbol{x})$, using the adaptive Markov chain simulation procedure presented in Section2.

Step 2: Based on the samples generated in Step 1, construct the wavelet sampling density $\hat{f}(x)$ in Eqs.(10) or (21) by nonlinear wavelet thresholding density estimator.

Step 3: Use the nonlinear wavelet thresholding density estimator constructed in Step 2 as the importance sampling density, and perform importance sampling simulation.

Assuming that N Markov chain samples are generated in Step 1 and M samples are used in the importance sampling process (Step 3), the total number of function evaluations of the limit state is $N+M$. For most reliability analysis of structures of practical interest, majority of the computational cost is expended on the multiple evaluations of the limit state function. The CPU time needed for the wavelet density estimation can be negligible in comparison with that of performing multiple limit state analyses. Therefore, the total number of function calls of limit state is used as the measure of the computational cost.

5. EXAMPLES

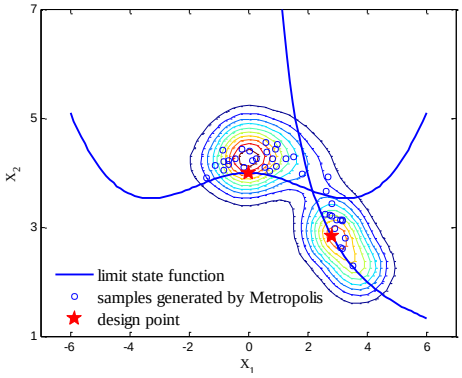
Two examples from literature were selected to demonstrate the proposed method. The multivariate normal distribution was used as the proposal distribution for the adaptive Markov chain simulations. The Daubechies wavelet with 4 vanishing moments (Db#4) was used in the wavelet density estimation. The efficiency of the wavelet density estimation is examined through comparison with the classical kernel density-based importance sampling method. In the following discussion, Wavelet-based IS denotes the proposed method, while Kernel-based IS represents the importance sampling using the kernel density estimation.

5.1. Example 1: a series system

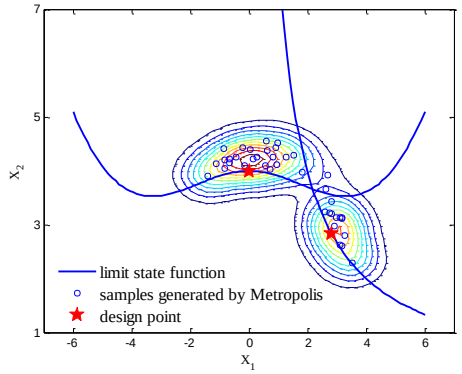
The first example is a series system with two branches,(Au and Beck 1999; Dubourg et al. 2013)

$$g(\mathbf{x}) = \min\begin{Bmatrix} 3 - X_2 + \exp\left(-\frac{X_1^2}{10}\right) + \left(\frac{X_1}{5}\right)^4 \\ 8 - X_1 X_2 \end{Bmatrix}$$

where X_1 and X_2 are independent standard normal variables. The system has two design points, $(0,4)$ and $(2.83,2.83)$, respectively.



(a) Kernel sampling density



(b) Wavelet sampling density

Fig. 1 Importance sampling density for Example 1

Fig. 1(a) and (b) plot the contours of the sampling density constructed by the nonlinear wavelet and kernel method, respectively, using 300 samples generated by the adaptive Markov chain simulation. The generated samples are clustered around the region that contributes most to the probability of failure. Using these points, both methods can construct the sampling density that reveals the import regions near the limit state surface. This example suggests that the wavelet thresholding density can be used as an alternative

to the kernel density for importance sampling. Since the constructed wavelet density has a shape similar to the optimal sampling density, a failure region of almost any shape can be handled without performing preliminary component reliability analysis for each limit state.

Table 1 compares the failure probabilities from Wavelet-based IS and Kernel-based IS. In both methods, 300 samples ($N=300$) were generated by the adaptive Markov chain simulation, and another 500 samples ($M=500$) were used in the subsequent importance sampling simulation. Therefore, the computational costs of the two methods are comparable. The 'exact' probability of failure was found to be 6.81×10^{-5} using 10^6 Monte Carlo simulations. It can be seen that the Wavelet-based IS is more accurate than the Kernel-based IS, due to high degree of flexibility and adaptivity of the nonlinear wavelet thrsholding density estimator and the dependence on the choice of initial parameters of the kernel-based IS.

Table 1 Reliability results of example 1

Methods	No. samples	P_f	Error(%)
Monte Carlo	10^6	6.81×10^{-5}	-
wavelet	N=300, M=500	7.41×10^{-5}	8.81
kernel	N=300, M=500	7.74×10^{-5}	13.66

Table 2 Reliability results for different choices of N

No. samples	Wavelet-based IS		Kernel-based IS	
	P_f	Error(%)	P_f	Error(%)
N=100,M=500	5.39×10^{-5}	20.85	8.55×10^{-5}	25.55
N=200,M=500	5.94×10^{-5}	12.78	5.48×10^{-5}	19.53
N=300,M=500	7.41×10^{-5}	8.81	7.74×10^{-5}	13.66
N=400,M=500	7.26×10^{-5}	6.61	7.60×10^{-5}	11.60
N=500,M=500	7.19×10^{-5}	5.58	6.24×10^{-5}	8.37

The results of two importance sampling methods with different values of N , while keeping that $M=500$ unchanged, are tabulated in Table 2. It can be seen that the proposed method is more accurate than the kernel method, particularly when the number of adaptive Markov chain samples is relatively small, which can be attributed to the fast convergence rate of the nonlinear wavelet thresholding density estimator. With the increasing number of the Markov chain samples (i.e., $N=500$), the performance of the kernel method is improved. This suggests that the kernel method has a relatively high demand to the number of pre-samples, resulting in the computational inefficiency.

5.2. Example 2:Noisy limit state function

A system with noisy limit state function was chosen to investigate the robustness of the proposed method (Kurtz and Song 2013). The limit state function is defined by

$$g(\mathbf{x})=b-X_2-\kappa(X_1-e)^2+0.001\sum_{i=1}^{2}\sin(100X_i)$$

where X_i are independent standard normal variables. Parameters b,κ and e are constants, with $b=5$, $\kappa=5$ and $e=0.1$. This limit state corresponds to two design points and the noisy term makes it difficult to search the design points accurately.

Table 3 compares the failure probabilities from Wavelet-based IS and Kernel-based IS. Results show that the proposed method is relatively insensitive to the noisy term and the relative error is about half of the traditional Kernel method, demonstrating the noisy reduction effect of the nonlinear thresholding rules.

Table 3 Reliability results of example 2

Methods	No. samples	P_f	Error(%)
Monte Carlo	3.6×10^5	3.07×10^{-3}	-
wavelet	N=500,M=500	3.33×10^{-3}	8.47
kernel	N=500,M=500	2.51×10^{-3}	18.24

Comparison between the nonlinear Wavelet-based IS and the Kernel-based IS with different combination N and M ($N+M=500$)is given in Table 4. It is shown that the nonlinear Wavelet-based IS is more accurate than the kernel method, particularly when the number of adaptive Markov chain samples is relatively

small, illustrating the advantage of the efficiency of the wavelet thresholding estimator. It's should be noted that too small M (number of IS samples) may reduce the accuracy of the proposed method, even if the constructed sampling density is approaching to the optimal sampling density.

Table 4 Reliability results for different choices of N and M

No. samples	Wavelet-based IS		Kernel-based IS	
	P_f	error(%)	P_f	error(%)
N=100,M=400	3.96×10^{-3}	28.99	4.46×10^{-3}	45.82
N=200,M=300	3.83×10^{-3}	24.76	1.98×10^{-3}	35.55
N=300,M=200	2.44×10^{-3}	20.52	2.06×10^{-3}	32.90
N=400,M=100	1.87×10^{-3}	39.09	4.12×10^{-3}	34.20

6. CONCLUSIONS

A new adaptive importance sampling method has been developed using adaptive Markov chain simulation and nonlinear wavelet density estimation. Wavelet density estimator has remarkable advantages, including different degrees of smoothness, localization, and fast implementation; it may approximate the near-optimal sampling density more efficiently than the classical kernel method. Two examples have been analyzed to demonstrate the efficiency and accuracy of the proposed method. In both examples, the proposed method gave good results with reasonable computational effort. The error is less than or about 10% when 1000 samples (including both Markov chain samples and subsequent importance sampling samples) were used. It was also found that the proposed method is more accurate than the conventional kernel density-based importance sampling method.

7. REFERENCES

Ang, G. L., Ang, A. H. S., and Tang, W. H. (1992). "Optimal Importance-Sampling Density Estimator." *Journal of Engineering Mechanics*, 118(6), 1146-1163.

Au, S. K., and Beck, J. L. (1999). "A new adaptive importance sampling scheme for reliability calculations." *Structural Safety*, 21(2), 135 - 158.

Balesdent, M., Morio, J., and Marzat, J. (2013). "Kriging-based adaptive Importance Sampling algorithms for rare event estimation." *Structural Safety*, 44, 1-10.

Chui, and K., C. (1992). *An introduction to wavelets*, Academic Press.

Dai, H., Zhao, W., and Wang, W. (2011). "An efficient adaptive importance samping method for structural reliability analysis." *Lixue Xuebao/Chinese Journal of Theoretical and Applied Mechanics*, 43(6), 1133-1140.

Daubechies, I. (1992). *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., and Picard, D. (1996). "Density estimation by wavelet thresholding." *The Annals of Statistics*, 24(2), 508-539.

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

Gelman AG, R. G. G. W. (1996). "Efficient Metropolis jumping rules.", Oxford University Press, Oxford, 599 – 608.

Haario, H., Saksman, E., and Tamminen, J. (2001). "An Adaptive Metropolis Algorithm." *Bernoulli*, 7(2), 223-242.

Hardle, W., Kerkyacharian, G., Picard, D., and Tsybakov, A. B. (1997). *Wavelets, Approximation And Statistical Application*, Chapman and Hall, Berlin-Paris.

Kurtz, N., and Song, J. (2013). "Cross-entropy-based adaptive importance sampling using Gaussian mixture." *Structural Safety*, 42, 35-44.

Mallat, S. G. (1989). "A theory for multiresolution signal decomposition: the wavelet representation." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7), 674-693.

Mallat, S. G. (1989). "Multiresolution Approximations and Wavelet Orthonormal Bases of $L^2(R)$." *Transactions of the American Mathematical Society*, 315(1), 69-87.

Melchers, R. E. (1989). "Importance sampling in structural systems." *Structural Safety*, 6(1), 3 - 10.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). "Equation of State Calculations by Fast Computing Machines." *Journal of Chemical Physics*(No.6), 1087-1092.

Silverman., B. W. (1986). "Density estimation for statistics and data analysis.", Chapman and Hall/CRC, 175.

Vidakovic, B. (1999). *Statistical modeling by wavelets*, John Wiley & Sons, Inc.

Walter, G. G., and Shen, X. (1994). *Wavelets and other orthogonal systems*, CRC Press, Boca Raton.