An Empirical Fitting-Adaptive Approach to Importance Sampling in Reliability Analysis

by Zheng Yunlong*, Member Yukio Fujimoto**, Member Mitsumasa Iwata**, Member

Summary

This paper presents a fitting-adaptive approach to importance sampling, in which a fitting density function is introduced to calculate the failure probability instead of the importance sampling density function. The approach improves the accuracy of the importance sampling technique when the sample size is relatively small. Therefore, it is very useful to problems that have high reliability and consume much CPU time in a single Monte Carlo run. Furthermore, the adaptive operation is performed if the fitting density function is not satisfactory. When the design point is used, the proposed approach yields even better results. Numerical examples demonstrate that a significant reduction in variance of the failure probability can be achieved by use of this approach.

Nomenclature

- $X$: basic random vector
- $Y$: random vector for importance sampling
- $Z$: standard normal random vector
- $\Phi(.)$: standard normal distribution function
- $\phi(.)$: standard normal density function
- $F_X(X)$: cumulative distribution function for $X$
- $f_X(X)$: probability density function for $X$
- $g_Y(Y)$: importance sampling density function (i. s. d. f.)
- $h_Y(Y)$: fitting-adaptive density function (f. a. d. f.)
- $s(Z)$: failure surface function in $Z$-space
- $X^*$: design point in $X$-space
- $Z^*$: design point in $Z$-space
- $\beta$: safety index
- $P_f$: failure probability
- $P_{M}$: failure probability by Monte Carlo simulation
- $P_{IS}$: failure probability by importance sampling
- $P_{FA}$: failure probability by fitting-adaptive approach
- $\Omega$: entire domain
- $\Omega_s$: safety domain
- $\Omega_f$: failure domain
- $N$: sample size
- $M$: number of random variables
- $M_o$: number of failure modes

1. Introduction

In reliability analysis, the Monte Carlo Simulation (MCS) is often employed when analytical solution is not obtainable. However, the computational effort involved in MCS sometimes becomes exorbitant because of the enormous sample size, especially for those problems which take a lot of CPU time in each Monte Carlo run. In order to improve the efficiency and simultaneously keep the accuracy, a variety of variance reduction techniques have been proposed, among which the Importance Sampling (IS) technique is generally recognized as the most efficient)\(^{-7}\)-\(^{-11}\). Although many kinds of IS techniques have been read in the literature, they may basically be classified into three categories, viz. (1) IS using design point, (2) IS based on the information obtained from the direct MCS and (3) IS through adaptive procedures. The first one is the combination of IS and First-/Second-Order Reliability Method\(^{(3,7)}\), but this is only possible when the performance function, or the failure surface function, can be given explicitly. The second one needs the information of the direct MCS within the failure domain\(^7\). However, this is rather difficult for low failure probability problems. The third one refines the solution step by step by using the 'knowledge' just learnt from the problem under consideration\(^{10,19}\). This sort of method generally gives good results if the number of random variables is not large\(^a\).

To date there have appeared a great number of papers on the application of the IS technique to reliability analysis. However, most studies are concerned with the formation of the ideal importance sampling density function (i. s. d. f.), but little attention has been paid to the discrepancy between the i. s. d. f. and the one simulated by a finite number of sample points. Ang et al. (1989)\(^{15}\) presented a Kernel Method to estimate the
optimal i. s. d. f. based on the samples within failure domain provided by the direct MCS. The i. s. d. f. thus formed meets the zero-variance condition to higher degree and therefore yields good results. Theoretically this method is very attractive, but for a high reliability problem, as is often the case in engineering, it is very difficult to produce as many failure samples as needed to construct such a kernel function. For instance, if \( Z \) follows normal distribution, then some 5626 samples are necessary for a crude MCS to produce 128 samples within the failure domain \( \mathcal{Q}_i(Z : Z - 2 < 0) \).

The so-called Adaptive Sampling procedure (C. G. Bucher, 1988) starts from an initial sampling density in order to obtain failure samples, and the mean value and the variance of these failure samples are calculated from which a new sampling density is determined. This sampling density is used as the i. s. d. f. to generate the next set of samples and from the results the mean value and the variance are updated again\(^{(6,19)}\). In this procedure, only the mean value and the variance of the previous simulation are used, while the generated samples themselves are not used. This is indeed a waste of information. Moreover, as pointed out by A. Karamchandani (1989), the variance determined by failure samples tends to be very small. Therefore, the adapted sampling density sometimes fails to move to the important part of some of the variables. In this case the resulting estimates significantly underestimate the correct failure probability although the estimated COV is small\(^{(8)}\).

For a realistic structure such as a platform or a ship’s hull, it is not an easy task to generate thousands of samples, especially when the nonlinear behavior is involved. Therefore, it is necessary to develop a method that efficiently uses the samples obtainable within a reasonable amount of computation.

In this paper, a fitting-adaptive approach is proposed to improve the accuracy and efficiency of the importance sampling technique. In this approach a fitting density function is used for calculating the failure probability so as to make full use of the samples generated from the selected i. s. d. f. Moreover, the fitting density function can be improved by an adaptive procedure. The key point of the proposed approach is to exploit thoroughly the samples obtained, and hence make an improvement on the IS technique in solving problems that are very time-consuming.

When the failure surface function can be expressed explicitly, the so-called design point can be readily found by means of Level 2 methods. It is generally agreed that IS is very powerful when the design point is utilized. Even in this case the fitting-adaptive approach still yields a substantial improvement on the IS technique.

As to the structural reliability analysis, the randomness of both the loads and the strength is taken into consideration. Because the COV values of strength parameters are usually small, the Neumann Expansion (NE) method is employed to speed up the structural analysis in every Monte Carlo run. The formulation of NE is given in Appendix A, and the details can be found in Refs\(^{(5,4)}\).

2. Fitting-Adaptive Approach

As opposed to the conventional IS technique in which the i. s. d. f. is to be simulated by generating sufficient samples according to this very density function, in the fitting-adaptive approach a fitting function, named fitting-adaptive density function (f. a. d. f.), is introduced to properly fit the samples drawn from the selected i. s. d. f. This approach is described below after a brief presentation of the IS technique.

Let \( X = (x_1, x_2, \ldots, x_m) \) denote the basic random vector consisting of all the random variables considered and \( f_S(X) \) denote the joint probability density function of \( X \), then the failure probability \( P_f \) is expressed as the multi-dimensional integral of \( f_S(X) \), i.e.

\[
P_f = \int_{\mathcal{Q}_f} f_S(X) dX = \int_{\mathcal{Q}_f} I(X) f_S(X) dX
\]

where \( \mathcal{Q}_f \) and \( \mathcal{Q} \) denote the failure domain and the entire domain respectively, \( I(X) \) is an indicator defined by the following expression.

\[
I(X) = \begin{cases} 1, & \text{if } X \in \mathcal{Q}_f \\ 0, & \text{otherwise} \end{cases}
\]

The direct MCS of \( P_f \) can be formulated in terms of sample-mean\(^{(4)}\), i.e.

\[
P_{f_{\text{MS}}} = \frac{1}{N} \sum_{i=1}^{N} I(X_i)
\]

where \( N \) is the sample size, \( X_i (i=1, 2, \ldots, N) \) is a sample taken from the density function \( f_s(X) \).

Assume that \( Y = (y_1, y_2, \ldots, y_m) \in \mathcal{Q} \) is another random vector, and is distributed with the joint probability density function \( g_{v_1}(Y) \), then, eq. (1) can be rewritten as

\[
P_f = \int_{\mathcal{Q}_f} I(Y) f_s(Y) g_{v_1}(Y) dY
\]

where \( g_{v_1}(Y) \) is the so-called importance sampling density function.

Similarly, the importance sampling MCS of eq. (4) can be obtained by means of sample-mean, i.e.

\[
P_{f_{\text{MS}}} = \frac{1}{N} \sum_{i=1}^{N} I(Y_i) f_s(Y_i) g_{v_1}(Y_i)
\]

where \( Y_i \) is a sample generated according to \( g_{v_1}(Y) \).

The transformation of eq. (1) into eq. (4) only contains a mathematical manipulation of introducing a function \( g_{v_1}(Y) \), so eq. (4) will result in correct answer as long as the integrand is well-behaved. As for the simulation formula, eq. (5), the selection of i. s. d. f. \( g_{v_1}(Y) \) has a critical bearing on both the efficiency and the accuracy of the simulation. The variance of \( P_{f_{\text{MS}}} \) is proved\(^{(4)}\) to be
What follows is the description of the procedure of within the same domain have different values of a sense, to the moment of that area. Different locations domain. Whereas $P_f$ in eq. (5) should be compared, in one-dimensional case, $P_f$ can be interpreted as the area merely depends upon which domain $X$, falls in. In

renders the rigorous determination of eq. (7) useless. However, eq. (7) contains $P_f$ which is unknown. This would be enough to estimate the failure probability. Therefore, the best choice of $g_Y( Y)$ should be

which can reduce the variance of $P_f$ to zero. Could an

the fitting-adaptive approach.

Then, the sample-mean of eq. (8) is of the form

and store $Y_i$ and $I(Y_i)$.

(3) Fit the sample points $Y_i(i=1, 2, \cdots, N)$ with a proper f. a. d. f. $h_Y(Y)$. (4) If the f. a. d. f. is not satisfactory, then adapt it as described in the subsequent chapter or reselect the i. s. d. f. $g_Y(Y)$, go to step (2) and proceed with the remaining samples. Otherwise, go to step (5).

(5) Compute the failure probability by eq. (9).

It is clear at this point that, when the sample size is enlarged for adaptive operation, the samples already generated are still useful. This saves much computational time at the cost of additional storage for storing $Y$ and $I(Y)$.

The crux of the fitting-adaptive approach lies in the formation of the $h$-function. The simplest way is to fit the sample points with a function of the same type as the i. s. d. f. used. For example, if the i. s. d. f. of $Y$, $g_Y(Y)$, is chosen as the normal distribution $N(Y_i)$, then the f. a. d. f. $h_Y(Y)$ can be taken as $N(Y_i)$, with $Y_i$ and $(\sigma_i)^2$ being the sample-mean and sample-variance. Unfortunately, this kind of f. a. d. f. has little, if any, improvement on importance sampling. In the present paper the histogram is employed as f. a. d. f. because it is straightforward and fulfills his purpose. Other kinds of f. a. d. f.'s may as well be used in further studies.

If an i. s. d. f. fails to produce enough failure samples, it has to be moved towards the failure domain to generate 'new' samples. Even in such a case the 'old' samples may still be used, because a histogram is capable of covering sample points that come from different distributions provided the density functions are not too far away. This is a distinctive feature of the fitting-adaptive approach.

3. How to Form the Histogram

After the i. s. d. f. is determined, the goodness of an IS depends on the sample size. The solution gets improved as the sample size increases. On the other hand, if the sample size is also fixed, the goodness of an IS depends on how effectively the generated samples are made use of, i. e. how well the f. a. d. f. behaves. For simplicity, the discussion on the histogram is confined to univariate problems in this chapter.

Let $W$ denote the width of each window, the histogram may be expressed by

in which $y_i$ is a sample point, $y_e$ is the coordinate of the mid-point of the window where point $y$ belongs to, and $K(.)$ is defined as follows.

Figure 1 illustrates the normal density curves and the associated histograms drawn to the same scale. The histograms are plotted with 30 windows within the interval $[-3\sigma, +3\sigma]$. In Fig.1(a) the histogram is formed with 400 sample points. It can be seen that the histogram turns out different from the corresponding
normal density curve, and has an empty window at point $y_2$. At point $y_1$, the ratios $\frac{f_x(y_1)}{g_y(y_1)} = \frac{a}{b}$ and $\frac{f_x(y_1)}{h_y(y_1)} = \frac{a}{c}$ are quite different. Because $a/c$ reflects the obtained samples more accurately than $a/b$ does, the FA estimator $P_m$ has smaller variance than the IS estimator $P_f$. The histogram in Fig. 1(b) is formed using 800 sample points. It is closer to the normal density curve than that in Fig. 1(a), but it still has an empty window at $y_4(y_4 = y_2)$. In addition, it clearly deviates from the normal density curve at point $y_3$. The histogram shown in Fig. 1(c) is formed using 10000 sample points, it is by far better.

In order to get better results, the following two conditions should be taken into account in forming a histogram.

1. The histogram should narrowly cover all the sample points, that is, its lower and upper tails should, respectively, reach the minimum and maximum values of all sample points.

2. There must not be empty windows in the histogram in order not to lose information.

When the design point is used, condition (1) may be relaxed. In fact, the solution becomes a little better if some extreme sample points, which are thought of as spurious noise, are removed. If empty windows emerge they must be eliminated either by increasing the sample size or by adjusting the window width. The algorithm to adapt the histogram is briefly presented below.

Let $N_1$, $N_2$ and $\Delta N$ denote the minimum, maximum and increment of sample size respectively, $M_1$, $M_2$ and $\Delta M$ denote minimum, maximum and increment of the number of windows respectively, $A$ and $B$ denote the lower and upper bounds of the histogram respectively. If there exist empty windows in the formed histogram, the number of windows is first decreased step by step. If there are still empty windows after the number of windows has reached its minimum, then the sample size is increased step by step. If all the empty windows can not be eliminated in this way, some information is fed back. Figure 2 shows the flow chart of the algorithm for histogram adaptation.

4. Design Point

In Level 2 reliability analysis, for a known failure-mode the failure surface may be defined in terms of the basic random vector $X$. 
An Empirical Fitting-Adaptive Approach to Importance Sampling in Reliability Analysis

(12)

where $f(X)$ is the so-called performance function, it has positive or non-positive values if $X$ is in safety domain $\Omega_s$ or failure domain $\Omega_f$.

\[
f(X) = \begin{cases} 
0, & X \in \Omega_s \\
< 0, & X \in \Omega_f
\end{cases}
\]

(13)

Variables $X$ are first of all transformed into uncorrelated and standardized normal variables $Z$.

(14)

In the case that all the variables are mutually independent and $X_i$ is distributed with $F_{x_i}(X_i)$, eq. (14) becomes

\[
s(Z) = f(T^{-1}(Z))
\]

(16)

The safety index $\beta$ is defined as the shortest distance from the $Z$-origin to the failure surface $s(Z) = 0$.

(17)

Then, the failure probability can be expressed as

\[
P_f = \Phi(-\beta)
\]

(18)

If point $Z''$ satisfies eq. (17), then point $X'' = T^{-1}(Z'')$ is called the design point, which has highest probability density. The design point can be found in several ways, such as the First-Order Second-Moment (FOSM) method, the optimization procedure, etc.

The principal idea of the FOSM method is to expand the failure surface at $Z''$ in $Z$-space in Taylor series, retaining only the first-order terms. The safety index $\beta$ is iteratively solved (see Appendix B). References quoted give further information.

For a mono-modal failure problem, the calculated design point can be directly used for IS. That is, the i.s.d.f. is chosen such that its mean value coincides with the design point.

\[
g^*(Y) = g_i(Y)
\]

(19)

For a multi-modal failure problem, say the one which has $M_o$ failure modes, special treatment is necessary. One way to do this is to find the joint failure probabilities of each two failure modes concurrently taking place, i.e.

\[
\sum_{i=1}^{M_o} \sum_{j=1}^{M_o} P_{ij} = P_{ij} = P_{ij}(\text{Mode } i \text{ and Mode } j)
\]

(20)

and then calculate the upper and lower bounds of the failure probability. Fairly good solution can be obtained if a narrow bound formula is used. However, this method requires $M_o \times (M_o + 1)/2$ times of simulations, so it may become intractable if $M_o$ is comparatively large.

Another way, known as the ‘Composition Method’, is to compose a unique i.s.d.f. that is a probability mixture of density functions of all failure modes. This method is employed in the present study and is described in the following. Firstly, choose an i.s.d.f. $g(Y)$ for each of the failure modes according to the respective design points in the same way as in the mono-modal case; Secondly, assign a weighting factor $p_i(p_i > 0 \text{ and } \sum p_i = 1)$ to $g_i(Y)$ in reference to the contribution of the failure mode to the total probability of failure; Thirdly, sum up $g_i(Y)$ to form the unique i.s.d.f.

\[
g(Y) = \sum_{i=1}^{M_o} p_i g_i(Y)
\]

in which

\[
p_i = \Phi(-\beta_i)\sum_{i=1}^{M_o} p_i \Phi(-\beta_i)
\]

(21)

(22)

5. Numerical Examples

Several examples are presented in the following to demonstrate the feasibility of the fitting-adaptive approach in the application to one-dimensional and multi-dimensional problems. Mono-modal failure and multi-modal failure are considered. All the computations are performed on HITAC computer (M680H). Random numbers are generated by calling the computer library routines. In all the examples presented, the IS and the FA are carried out under identical conditions so that comparison may be made.

5.1 Mathematical Problems

Example 1: One-dimensional mono-modal case. Consider respectively the following three cases of failure mode, where $Z$ follows $N(0, 1)$ distribution.

Case 1: $f(Z) = 2Z$

Case 2: $f(Z) = 3Z$

Case 3: $f(Z) = 4Z$

For each case, the ‘seed’ for generating random numbers is initialized once for all, i.e. after each simulation the seed is saved to be used in the next simulation. In this way the effect of initial data may be reflected. In order to examine the influence of different locations of the i.s.d.f., the same distribution $N(3, 1)$ is used as the i.s.d.f. for all these three cases. Its expected value is situated at the design point of Case 2, in the failure domain of Case 1, and in the safety domain of Case 3.

Figures 3 through 5 show the failure probabilities of the three cases. In general, the fitting-adaptive approach behaves much better than the importance sampling technique. In Case 1 the failure probability is relatively high, so the direct MCS applies, exhibiting larger variance (Fig. 3). In Case 2 the design point is utilized, so the IS is expected to yield good results. However, even in such a case the FA still improves the accuracy of IS (Fig. 4). Fig. 5 highlights the applicability of the fitting-adaptive approach to low probability problems.

Example 2: Two-dimensional multi-modal case with design points utilized. Consider the following two failure modes simultaneously

Mode 1: $f(X_1, X_2) = 1(X_1^2 + X_2)$

Mode 2: $f(X_1, X_2) = 1(X_1^2 - X_2)$

where $X_i$ and $X_j$ are independent and distributed with $N(0, 0.1)$ and $N(0, 0.01)$, respectively. Because the performance functions are known, the safety indexes and the design points can be readily
found: $\beta_1=4.657, X^T=(0.184, 0.966); \beta_2=5.148, X^T=(0.202, -0.959)$. They are used to form the multi-modal i. s. d. f. Fig. 6 shows the failure probability with respect to the sample size. The 'exact' solution in the figure is computed by IS with large sample size: $N=500000$. These results are summarized in Table 1. It is well known that the IS technique is very efficient when combined with the design points. Still, the FA approach converges much faster.

5.2 Structural Problems

Structural analysis is performed in the same way as in Ref.\textsuperscript{11}. In addition, the proportion that each failure pattern occurs is also found. For strength parameters, the Young's modulus, the yielding stress and the plastic modulus are treated as random variables because they are significant to the probability of failure\textsuperscript{6}. Their COV values are all 0.05. All load parameters are treated as random variables except for the topside weights of Example 2. COV values for random loads are all 0.30. Both load variables and strength variables are assumed normally distributed and mutually independent. In such a case, the i. s. d. f. can be reduced into an expression only depending on one parameter $V$. $V$ is defined as $V=(Y-X)/\sigma_X$ (Fig. 7). It was concluded in Ref.\textsuperscript{11} that IS can be applied to strength variables, load variables or both. In this paper, the IS is applied to load variables. The 'exact' solution in the figures refers to the result by IS with sample size $N=500000$.

The structural failure is defined as the occurrence of one or more plastic hinges due to the combined load effect of axial force, shearing force and bending moment\textsuperscript{6}.

Example 1: Figure 8 shows a portal frame structure to which two load cases are respectively applied. Numerical data are listed in Table 2. $V=2.0$ for Load case 1, and $V=3.25$ for Load case 2.

![Fig. 3 Failure probability of Case 1](image)

![Fig. 4 Failure probability of Case 2](image)

![Fig. 5 Failure probability of Case 3](image)

![Fig. 6 Failure probability of multi-modal failure](image)

![Fig. 7 Schematic representation of parameter $V$](image)

Table 1 Failure probability of multi-modal failure

<table>
<thead>
<tr>
<th>Sample size</th>
<th>$P_fH$</th>
<th>$P_fI$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>$2.47\times10^{-6}$</td>
<td>$2.73\times10^{-6}$</td>
</tr>
<tr>
<td>2000</td>
<td>$2.31\times10^{-6}$</td>
<td>$2.47\times10^{-6}$</td>
</tr>
<tr>
<td>3000</td>
<td>$2.31\times10^{-6}$</td>
<td>$2.41\times10^{-6}$</td>
</tr>
<tr>
<td>500000</td>
<td>2.32\times10^{-6}</td>
<td></td>
</tr>
</tbody>
</table>

$\Phi(-\beta_1)=1.60\times10^{-6}$ $\Phi(-\beta_2)=1.32\times10^{-7}$
An Empirical Fitting-Adaptive Approach to Importance Sampling in Reliability Analysis

Figure 9(a) and Figure 9(b) show the failure probabilities versus sample sizes of Load case 1 and Load case 2, respectively. It can be seen that the FA approach exhibits noticeable superiority to the IS technique. For high probability problem (Fig. 9(a)), FA yields very good result with hundreds of samples. Both FA and IS are feasible when thousands of samples are used. For low probability problem (Fig. 9(b)), however, IS converges very slowly. The result by IS is not yet satisfactory when the sample size is 4000.

Example 2: Figure 10 shows a simplified plane structure, which can be thought of as the residual strength of a platform after braces on the second floor and the third floor have been subjected to brittle failure. The numerical data are given in Table 3. For this structure, $V = 2.5$. The failure probability versus sample size is shown in Fig. 11. It can be observed that FA improves IS markedly. Moreover, a total of thirteen failure patterns are identified and Fig. 12 shows the first four patterns. Note that the four major patterns take a proportion of 94%.

6. Conclusions

(1) In the proposed approach, a fitting-adaptive density function is introduced to calculate the failure probability instead of the importance sampling density function, while the importance sampling density function is only used for generating sample points. This approach is very effective when the sample size is small, it is therefore useful for time-consuming problems.

(2) The fitting-adaptive density function allows...
the use of samples from different distributions. Therefore, after the importance sampling density function is shifted for adaptive operation, the generated samples may still be used to save computational time.

(3) When the design point is used, the proposed approach is also of value to variance reduction. For multi-modal failure problems, the so-called Composition Method is feasible.

(4) When the two conditions stated in Chapter 3 are satisfied, the histogram gives very good results. Further studies on the fitting-adaptive density function would be beneficial. For example, the histogram with variable-width windows(6) may offer advantage.

Acknowledgement

The authors wish to express their deep thanks to Prof. Hiroshi Itagaki, Yokohama National University, for his helpful suggestion and discussion on this paper.

References

An Empirical Fitting-Adaptive Approach to Importance Sampling in Reliability Analysis


Appendix A: Formulation of Neumann Expansion

The governing equation of FEM involving random parameters of both strength and loads may be written as:

$$([\bar{K}]+[\Delta K])\{d\}=[\bar{R}]+[\Delta R]$$

(A 1)

where $[\bar{K}]$, $[\bar{R}]$ and $\{d\}$ are global stiffness matrix, right-hand vector and displacement vector, respectively. The over a letter and the 4 before a letter refer to the mean value and the deviated part, respectively.

By premultiplying $[\bar{K}]^{-1}$ on both sides of eq. (A 1), the following equation is derived.

$$([I]-[P])\{d\}=[d]_0$$

in which

$$[P]=-[\bar{K}]^{-1}[\Delta K]$$

(A 2)

$$[d]_0=[d]_0+(\bar{d})_0$$

(A 3)

or can be expanded in Neumann series as follows5).

$$[d]=([I]+[P]+[P]^2+\cdots)[d]_0$$

(A 4)

or in recursive form

$$(d)=[d]+([d]+([d]+\cdots))$$

(A 5)

Eq. (A 4) or eq. (A 5) converges if all the eigenvalues $\lambda$, $\lambda$, $\cdots$ satisfy the following condition.

$$\rho=\max(|\lambda|)<1, i=1, 2, \cdots$$

(A 6)

For efficient execution of eq. (A 5), the Cholesky decomposition may be used. Substituting eq. (A 3) into eq. (A 5), we have

$$[\bar{K}]\{d\}_0=([I]+[P]+[P]^2+\cdots)[d]_0$$

(A 7)

Let

$$[\bar{K}]=[L][D][L]’$$

(A 8)

then eq. (A 7) can be solved in three steps.

$$[L]\{y\}=-[\Delta K]\{d\}$$

$$[D]\{z\}=[y]$$

$$[L]’\{d\}_0=\{z\}, i=0, 1, 2, \cdots$$

(A 9)

Eq. (A 8) is solved only once, and is repeatedly used for computing each order of $\{d\}$ of every structural sample.

Appendix B: Iteration Algorithm for FOSM

Let $a^T=(a_1, a_2, \cdots, a_M)$ be the unit vector from the Z-origin, where $M$ is the number of variables, then the iteration algorithm can be formulated as follows.

$$s(z_i)\alpha|=0$$

$$a_i=\frac{-s\alpha_i}{\sum_{i=1}^{M} s\alpha_i}$$

(B 1)

It is readily seen from the above that FOSM requires the availability of explicit failure function, and preferably the function is differentiable.

Consider the two-dimensional failure surface function used in aforementioned example (Section 5.1).

$$s(z_i, z_j)=-1-(z_i^2+z_j^2)$$

(B 2)

In this case eq. (15) in Chapter 4 becomes

$$z_i=(x_i-\bar{X}_i)/\sigma_i$$

$$z_j=(x_j-\bar{X}_j)/\sigma_j$$

(B 3)

Substituting eq. (B 3) into eq. (B 2) leads to the failure function defined in Z-space.

$$s(z_i, z_j)=-1-(z_i^2+z_j^2+2\bar{X}_i\sigma_i z_i+z_j^2+\bar{X}_j^2+X_j^2)$$

(B 4)

Then, the iteration formulae are derived as follows:

$$\alpha_1=(2\sigma^2_i/\beta_0+(2\bar{X}_i/\sigma_i))/R$$

$$\alpha_2=\sigma_i/R$$

$$\beta_{i,j}=(1-\bar{X}_i^2-z_j^2)/(\sigma_i^2/\beta_0+2\bar{X}_i\sigma_i+\sigma_i^2)$$

(B 5)

where $R$ is determined by the condition

$$\alpha^T\alpha=1$$

(B 6)

Starting from an initial vector $v_0=(a', \beta_0)^T=(a_1, a_2, \beta_0)$ a succession of $v_i$'s can be iteratively found using eqs. (B 5) and (B 6). The iteration terminates if $\|v_{i+1}-v_i\|<\epsilon$

(B 7)

where $\|\|$ denotes the norm and $\epsilon$ is a small prescribed constant.

Finally, the design point is calculated by

$$Z^*=\beta\alpha$$

(B 9)

Appendix C: Likelihood Analysis

To show the superiority of the histogram $h_1(y)$ over the density $g_1(y)$, likelihood analysis is performed (see Chapter 3). The Log-likelihood is defined as follows.

$$L_g=\sum_{i=1}^{N} \log(g_y(y_i))$$

$$L_h=\sum_{i=1}^{N} \log(h_y(y_i))$$

(C 1)

where $y_i$ is a sample point and $N$ the sample size.

Let $g_1(y)$ be a standard normal density and $h_1(y)$ be the histogram formed from the samples drawn from $g_1(y)$. For a certain sample size $N$, the ‘seed’ is initialized only once, and the calculation of $L_g$ and $L_h$ is repeated for 100 times. The mean values of $L_g$ and $L_h$ with respect to different sample sizes are given in Table C1. The absolute values of $L_h$ are smaller than those of $L_g$, and the difference is larger when $N$ is small. This means that the histogram fits the samples better than the original density $g_1(y)$.

<table>
<thead>
<tr>
<th>N</th>
<th>100</th>
<th>400</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{L}_g$</td>
<td>-142</td>
<td>-566</td>
<td>-1416</td>
</tr>
<tr>
<td>$\bar{L}_h$</td>
<td>-112</td>
<td>-529</td>
<td>-1369</td>
</tr>
</tbody>
</table>

Table C1: Mean value of Log-Likelihood