LIF: A new Kriging based learning function and its application to structural reliability analysis

Zhili Sun\textsuperscript{a}, Jian Wang\textsuperscript{a,\*}, Rui Li\textsuperscript{a}, Cao Tong\textsuperscript{b}

\textsuperscript{a} School of Mechanical Engineering & Automation, Northeastern University, Shenyang 110819, China
\textsuperscript{b} Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China

\textbf{ABSTRACT}

The main task of structural reliability analysis is to estimate failure probability of a studied structure taking randomness of input variables into account. To consider structural behavior practically, numerical models become more and more complicated and time-consuming, which increases the difficulty of reliability analysis. Therefore, sequential strategies of design of experiment (DoE) are raised. In this research, a new learning function, named least improvement function (LIF), is proposed to update DoE of Kriging based reliability analysis method. LIF values how much the accuracy of estimated failure probability will be improved if adding a given point into DoE. It takes both statistical information provided by the Kriging model and the joint probability density function of input variables into account, which is the most important difference from the existing learning functions. Maximum point of LIF is approximately determined with Markov Chain Monte Carlo(MCMC) simulation. A new reliability analysis method is developed based on the Kriging model, in which LIF, MCMC and Monte Carlo(MC) simulation are employed. Three examples are analyzed. Results show that LIF and the new method proposed in this research are very efficient when dealing with nonlinear performance function, small probability, complicated limit state and engineering problems with high dimension.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

For a given mechanical structure with \(M\)-dimension input random variable \(\mathbf{X} = [X_1, X_2, \ldots X_M]^T\), its performance function \(G(\mathbf{X})\) divides the variable space into two domains, i.e. the safe domain \(G(\mathbf{X}) > 0\) and the failure domain \(G(\mathbf{X}) \leq 0\). The boundary of them is called the limit state \(G(\mathbf{X}) = 0\). The failure probability of the structure is defined as

\[ P_f = \int_{\{G(\mathbf{X}) \leq 0\}} f(\mathbf{X})d\mathbf{X} \]  

where \(f(\mathbf{X})\) is the joint probability density function of \(\mathbf{X}\) and \(\{G(\mathbf{X}) \leq 0\}\) is a failure indicator function. The fundamental task of structural reliability analysis is to perform the integration approximately. In engineering, the performance function of a given mechanical structure is nonlinear and almost impossible to get its explicit formulation generally. A numerical method is usually used to obtain the structural output. Therefore, several structural reliability analysis methods are developed to estimate the probability of failure.

Monte Carlo simulation (MCS) \cite{1}, estimating \(P_f\) with failure rate of random samples, is the most robust method. However, it needs to call the performance function (or numerical model) too many times. Some variance reduction techniques, e.g. line sampling (LS) \cite{2,3}, subset simulation (SS) \cite{4,5,6}, and importance sampling (IS) \cite{6,8,9,10}, are proposed to reduce the number of calls to the performance function and relieve the pressure of computation. At the same time, to consider structural behavior practically and improve their accuracy, numerical models are becoming more and more complicated, which increases the difficulty of reliability analysis. These variance reduction techniques cannot satisfy the computational requirements either. Compared with MCS and its improved methods, the first and second order reliability method (FORM and SORM) perform structural reliability analysis with much fewer samples. They are based on the knowledge of design point and Taylor expansion that ignores higher order terms, therefore their accuracy can hardly be guaranteed especially when dealing with complex nonlinear structures.

Within recent years, different kinds of meta-models \cite{3,11} are proposed to replace the target performance function. Their emphasis is to fit the limit state well for only the sign of a predicted structural response influences the estimate of \(P_f\) directly. Among meta-models, polynomial response surface \cite{11,12,13,14,15,16} is the most widely used one in engineering. The number of terms for a polynomial response surface method grows dramatically as polynomial
order and the dimension of input variable increase especially when cross terms are taken into account. Sparse polynomial
[15,18,19] is adopted to overcome this situation. Kang et al. [20], Roussouly et al. [15], Blatman and Sudret [18] raise methods for
determining significant terms of a sparse polynomial, and apply
their methods to structural reliability analysis. These methods
obtain a compromise between complexity and nonlinearity of
meta-models. Compared with polynomial models, machine
learning methods (neural network [21,22], support vector ma-
chine (SVM) [23–25], Multi-Layer Perceptrons [26] and so on) are
more suitable to matching performance function with highly
nonlinear input-output relationships. Kriging-based methods
[27,28] are also widely used in reliability analysis [9,29–34] and
Global Optimization [35,36]. As an exact interpolation technique,
the Kriging model, including a regression part and a stochastic
process, combines fitting of a polynomial and correlation analysis
of residual error. It not only predicts the structural response of a
point but also provides the local uncertainty measure (the so-
called Kriging variance). Therefore, the Kriging model is employed
in this research.

Besides meta-model formulation, the design of experiments
(DoE) also has huge influence on the convergence rate and the
accuracy of reliability analysis [12,15,24,29,37,38]. An excellent
strategy of DoE leads the process of reliability analysis converging
quickly and provides high accuracy at the same time [31,34,38–
40]. Various kinds of sequential DoE based on the Kriging model
have been proposed to improve the accuracy of structural relia-
bility analysis and reduce the number of calls to the real perform-
ance function [8,29,32,38,41,42]. Kriging based sequential DoE
has drawn more and more attention because it is often active and
can update itself by adding new sample point based on the sta-
tistical information provided by the Kriging model. Despite the
differences of strategies of DoE, their main idea and steps can be
summarized as follows:

1. According to the statistical information provided by the Kriging
model, define a learning function which can reflect how close a
point is to the limit state \( G(x) = 0 \), the potential of a point
crossing the predicted limit state \( G(x) = 0 \) or some other
properties.

2. Minimize (or maximize) the learning function in a random
sample set or the whole variable space, and find the minimum
(or maximum) point.

3. The minimum (or maximum) point is the best next one. And
add it to the DoE.

The learning function plays a critical role during the DoE up-
dating. An inefficient learning function may slow down the pro-
gress of reliability analysis, or even worse. Many learning functions
have been proposed from different perspectives. Expected fea-
sibility function (EFF), developed by Bichon et al. [29], searches
for points in the vicinity of the limit state over the variable space.
The maximum point of EFF is added into the DoE step by step so
that \( G(x) = 0 \) converges to \( G(x) = 0 \). EFF roughly estimates the failure
probability very quickly. The learning function \( U \), proposed by
Echard et al. [34], focuses on the probability of misclassification
made by the Kriging model on the sign of \( G(x) \). Just like discussed
in [41], \( U \) gives more weight to points in the vicinity of the pre-
dicted limit state rather than the Kriging variance, which is its
main difference from EFF. Lv et al. [41] and Yang et al. [42,43]
present two new learning functions which are named \( H \) and ex-
pected risk function (ERF) respectively. \( H \) is based on information
entropy, and its principle is similar to EFF’s. Just like the expected
improvement function (EIF) [35] values how much the objective
function would be improved at a point than the current optimum,
ERF can identify the risk that the sign of a point is wrongly
predicted.

In this research, a new learning function named least im-
provement function (LIF) is proposed and applied to structural
reliability analysis. According to the statistical information
provided by the Kriging model, the probability of making a mistake
on the sign of any point in the variable space can be measured, based
on which the uncertainty of predicted failure probability \( \hat{P}_f \)
is defined. LIF is constructed to approximately measure how much
the uncertainty of \( \hat{P}_f \) will be improved if adding a given point into
DoE. The main difference between LIF and other learning functions
is that it takes both Kriging statistical information and the prob-
bability density function into account. LIF searches the point that
influences the accuracy of \( \hat{P}_f \) most rather than the one whose sign
is uncertain most. The optimal point of LIF is determined with
MCMC simulation. An active learning reliability method is raised
by employing the Kriging model and LIF. MCS is adopted to esti-
mate \( \hat{P}_f \) and its uncertainty.

The remainder of this paper is organized as follows. Section 2
introduces the Kriging model and MC simulation briefly. Section 3
presents the least improvement function in detail and analyzes it,
which is followed by the proposed reliability analysis method in
section 4. Three examples are employed in Section 5 to illustrate
the efficiency of LIF and the proposed method. Section 6 is the
conclusion.

## 2. Kriging model for structural reliability analysis

The Kriging model, a nonlinear interpolation meta-model de-
veloped for geostatistics by Matheron [27], consists of two parts,
i.e. a linear regression model and a random function. It is supposed
that the performance function \( G(x) \) can be denoted as

\[
G(x) = \sum_{h=1}^{H} \beta_h g_h(x) + z(x) = \mathbf{g}^T(x)\beta + z(x)
\]  

(2)

The first term of Eq. (2) is a realization of a regression function.
\( g_h(x) \ (h = 1, 2, \ldots, p) \) is the basis regression function and its order
is set to be one in this research. The second term of Eq. (2) is a
Gaussian stochastic process whose mean is zero and the var-
ciance of random process \( z(x) \) is

\[
\text{Cov}(z(x_m), z(x_j)) = \sigma^2 \mathbf{R}(x_m, x_j; \theta)
\]

where \( \sigma^2 \) is the variance of the Gaussian process, and \( \mathbf{R}(x_m, x_j; \theta) \) is the correlation coefficient between \( z(x_m) \) and \( z(x_j) \) with parameter
\( \theta \). The most widely used correlation function is the Gaussian cor-
rrelation function, whose form is

\[
\mathbf{R}(x_m, x_j; \theta) = \prod_{m=1}^{M} \exp\left[-\theta_m (x_m - x_j)^2\right]
\]

(4)

where \( x_m \) is nth component of \( x \).

Given an initial DoE \( \mathbf{X}_{\text{doe}} = [x_1, x_2, \ldots, x_N] \), and the correspon-
ding structural response \( \mathbf{Y}_{\text{doe}} = [y_1, y_2, \ldots, y_N] \), the unknown param-
eters \( \beta, \sigma^2, \theta \) in Eqs. (2)-(4) can be estimated by maximizing the
likelihood function,

\[
\max L(\theta) = -\left( N \ln(\sigma^2) + \ln\left| \mathbf{R}_l \right| \right)
\]

(5)

where
\[
\delta^2 = \frac{1}{N_0} (Y - G\beta)^T R^{-1} (Y - G\beta) = (G^T R^{-1} C)^{-1} G^T R^{-1} Y \gamma \\
= [g(x_1), g(x_2), \ldots, g(x_{N_0})]^T R = (R(x, x_j; \theta))_{N_0 \times N_0}
\]

Let \( \hat{\beta}, \sigma^2, \hat{\theta} \) denote the optimum estimation of \( \beta, \sigma^2, \theta \) respectively gotten from Eq. (5). Now consider the linear combination predictor,

\[
\hat{G}(x) = c^T(x) Y
\]

The error of Eq. (6) is

\[
\hat{G}(x) - G(x) = c^T(x) Y - G(x) = c^T(x) Z - z + (G^T c(x) - g(x))^T \hat{\beta}
\]

where

\[
Z = (z_1, z_2, \ldots, z_{N_0})^T
\]

To minimize the mean square error (MSE) with unbiased estimation of \( y \), the result is

\[
\mu_c(x) = \hat{G}(x) = \hat{g}(x)^T \hat{\beta} + \Gamma(x)^T \gamma
\]

\[
\sigma_c^2(x) = \sigma^2 \left[ 1 + \Gamma(x)^T \left( G^T R^{-1} G \right)^{-1} \Gamma(x) - \Gamma^T(x) R^{-1} \Gamma(x) \right]
\]

where

\[
\gamma = R^{-1} (Y - G\hat{\beta}) \Gamma(x) = \left[ R(x, x_1; \hat{\theta}), \ldots, R(x, x_{N_0}; \hat{\theta}) \right] \Gamma(x) = G^T R^{-1} \Gamma(x) - g(x)
\]

Reference [44] gives the detailed derivation of Eqs. (7) and (8). According to the Gaussian process regression theory, the structural response of \( x \) is subject to a normal distribution,

\[
G(x) \sim N(\mu_c(x), \sigma_c^2(x))
\]

The estimation of the failure probability of the structural model mentioned in Section 1 is

\[
\hat{P}_f = \int I_{G(x) < 0} f(x) dx
\]

It is very time consuming to perform the multiple integration of Eq. (10) directly when \( M > 3 \). Therefore, MCS and some variance reduction techniques (IS, LS and so on) are often adopted to compute \( \hat{P}_f \) approximately. As the most robust method, MCS is used in this research. According to the law of large numbers,

\[
\hat{P}_f \approx \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} I_{G(x_{MC;i}) < 0}
\]

where \( N_{MC} \) is the number of random samples and \( x_{MC,i} (i = 1, 2, \ldots, N_{MC}) \) is a sequence of i.i.d. random samples drawn from distribution \( f(x) \). So \( I_{G(x_{MC;i}) < 0} \times \hat{P}_f \) is a sequence of i.i.d. random samples. The coefficient of variation of \( \hat{P}_f \) is calculated as

\[
\delta_{MC} = \sqrt{\text{var}(\hat{P}_f)} = \frac{1 - \hat{P}_f}{\sqrt{N_{MC} \hat{P}_f}}
\]

3. Least improvement function

In this section, a new learning function named least improvement function (LIF) is proposed. First of all, some widely referenced learning functions are introduced briefly. Then the uncertainty of \( \hat{P}_f \) is measured based on the statistical information provided by the Kriging model, following which the new learning function LIF is constructed and analyzed.

3.1. Some widely used learning functions

EFF, developed by Bichon et al. [29], searches for points in the vicinity of the limit state over the variable space, which is defined as

\[
\text{EFF}(x) = \int_{G(x) < 0} \left( \int_{G(x) - G(\hat{x})} f_{G(x)}(G(x)) dG(x) \right) dx
\]

where \( G(x) \) is set to be 0 in structural reliability analysis and \( e \) is proportional to \( \sigma_c(x) \). The value of \( e \) and the integral result of Eq. (13) are detailed in [29]. It is obvious that EFF becomes larger when \( x \) is near the predicted limit state or \( \hat{G}(x) \) has more uncertainty. It can be treated as an expectation that point \( x \) is located in the domain of \( -\epsilon < G(x) < \epsilon \).

The learning function \( U \), proposed by Echard et al. [34], focuses on the probability of misclassification made by the Kriging model on the sign of \( G(x) \). \( U \) is constructed as

\[
U(x) = \frac{\left| \mu_c(x) \right|}{\sigma_c(x)}
\]

According to Eq. (9), \( \phi(\cdot - U(x)) \) is the probability of point \( x \) crossing \( \hat{G}(x) = 0 \) and deteriorating the accuracy of \( \hat{P}_f \).

Eqs. (15) and (16) are two learning functions present by Lv et al. [41] and Yang et al. [43,42] respectively.

\[
\text{UF}(x) = \left| \int_x \phi(\cdot - U(x)) f_{G(x)}(G(x)) dG(x) \right|
\]

\[
\text{ERF}(x) = \left\{ \begin{array}{ll}
\int_x^{\infty} G(x) f_{G(x)}(G(x)) dG(x) \hat{G}(x) < 0 \\
\int_0^{-\infty} G(x) f_{G(x)}(G(x)) dG(x) \hat{G}(x) \geq 0
\end{array} \right.
\]

3.2. The least improvement function

Comparing Eq. (1) with Eq. (10), it is easy to conclude that the sign of \( \hat{G}(x) \) affects the accuracy of \( \hat{P}_f \) directly, and \( f(x) \) weights how important \( x \) is to the accuracy of \( \hat{P}_f \). The probability of making a wrong prediction about the sign of \( G(x) \) is \( \phi(\cdot - U(x)) \) (Eq. (9)). So the uncertainty or accuracy of \( \hat{P}_f \) can be measured as

\[
\text{UF}(x) = \int \phi(\cdot - U(x)) f(x) dx
\]

which is named as the uncertainty function \( \text{UF}(\hat{P}_f) \). It is obvious that as UF converges to 0, the estimated failure probability \( \hat{P}_f \) tends to \( P_f \):

\[
\lim_{\text{UF} \to 0} \hat{P}_f = P_f
\]

Taking [8,34] as reference, the sign of \( G(x) \) is seen as confirmed if \( U(x) > 2 \). The domain of \( \{x | U(x) > 2\} \) contributes little to UF, therefore Eq. (17) can be approximated as

\[
\text{UF} \approx \int_{U(x) \leq 2} \phi(\cdot - U(x)) f(x) dx
\]

The principle of the new learning function proposed in this research is to diminish UF by adding new point into DoE and updating the Kriging model sequentially. To approximately value how much a given point \( x_0 \) in the variable space but not included
in the current DoE can diminish UF, two hypotheses are formed.

1. Add \( x_0 \) into the current DoE with its performance function value \( G(x_0) \) and reconstruct the Kriging model. It is worthy to emphasize that \( G(x_0) \) here refers to the normal distribution variable defined by Eq. (9). The updated \( U \) and \( UF \) are denoted as \( U_0 \) and \( UF_0 \) respectively. If \( G(x_0) \) is not exactly equal to 0, there is a positive constant \( e_0 \) for the studied structure and the current Kriging model.

\[
\text{If } |x - x_0| < r(x_0), \text{ then } U_0(x) > 2, \text{ where } r(x_0) = e_0|G(x)|
\]

2. As a result of adding \( x_0 \) into the current DoE with \( G(x_0) \), the uncertainty of the sign for points out of the spherical neighborhood of \( x_0 \) does not become worse as a whole, which is to say

\[
\int_{(x - x_0)^2 < r^2(x_0)} \phi(-U_0(x))f(x)dx \geq \int_{(x - x_0)^2 < r^2(x_0)} \phi(-U_0(x))f(x)dx
\]

(19)

The first hypothesis can be interpreted as the way that benefitting from adding \( x_0 \) into the current DoE and reconstructing the Kriging model, the sign of \( G(x) \) is almost sure if \( x \) is located in the spherical neighborhood of \( x_0 \). And the radius of the spherical neighborhood is proportional to \( |G(x)| \). It is not always true in mathematics, but it is strict and can be proved in theory if \( |G(x)| > \delta_0 \), where \( \delta_0 \) is a fixed positive number. The second one indicates that the uncertainty of the sign for points out of the spherical neighborhood of \( x_0 \) does not become worse in general if its real response is calculated. It is reason able in engineering and theory, although authors cannot prove it for now.

According to the hypotheses above and Eqs. (17)-(19), UF diminishes as a result of \( x_0 \) added into DoE, which is constructed as

\[
UF - UF_0 = \int \phi(-U(x))f(x)dx - \int \phi(-U_0(x))f(x)dx
\]

\[
= \int_{(x - x_0)^2 < r^2(x_0)} \phi(-U_0(x))f(x)dx - \int_{(x - x_0)^2 < r^2(x_0)} \phi(-U_0(x))f(x)dx + \int_{(x - x_0)^2 < r^2(x_0)} \phi(-U(x))f(x)dx - \int_{(x - x_0)^2 < r^2(x_0)} \phi(-U(x))f(x)dx
\]

\[
\geq \int_{(x - x_0)^2 < r^2(x_0)} \phi(-U(x))f(x)dx
\]

(20)

The integrand of the last term in Eq. (20) is approximately treated as a constant function because the domain of integration is not large and the integrand is continuous. Therefore,

\[
\int_{(x - x_0)^2 < r^2(x_0)} \phi(-U(x))f(x)dx \approx k_0 \phi(-U(x_0))r^M(x_0)
\]

(21)

where \( k_0 \) is related to the dimension of basic variable(M). According to the first hypothesis, \( r(x_0) \) is proportional to \( |G(x)| \). Then,

\[
UF - UF_0 \geq k_0 \phi(-U(x_0))r^M(x_0) = k_0 \phi(-U(x_0))r^M(x_0)
\]

(22)

Taking statistical information Eq. (9) provided by the Kriging model into account, the expectation improvement of \( (UF - UF_0) \) is

\[
(UF - UF_0) = k_0 \phi^M(-U(x_0))f(x_0)E(G(x_0))
\]

(23)

For a given studied structure and a Kriging model, \( k_0 \) and \( e_0 \) are constant and have no relationship with \( x_0 \), so they have no influence on the maximum point of the right side of the inequality sign in Eq. (23). The lower limit of \( E(UF - UF_0) \) is proportional to \( \phi(-U(x_0))f(x_0)E(G(x_0)) \). Therefore, this research proposes it as a new learning function:

\[
LIF(x_0) = \phi(-U(x_0))f(x_0)E(G(x_0))
\]

(24)

LIF is the abbreviation of least improvement function. One can understand Eq. (24) in an intuitive way. If point \( x_0 \) is added to the current DoE, one can almost confirm the signs of all points near \( x_0 \). In another word, a spherical neighborhood is confirmed as a part of the safe domain or the failure domain after DoE includes \( x_0 \). The volume of the spherical neighborhood is proportional to \( IG(x_0) \), and \( \phi(-U(x_0))f(x_0)IG(x_0) \) approximately quantifies how important a point in the spherical neighborhood is to the current UF. \( \phi(-U(x_0))f(x_0)IG(x_0) \) is approximately proportional to the decrease of UF because of \( x_0 \). As \( G(x_0) \) is a normal distribution variable before calling to the performance function to calculate it, one can only measure the expectation of \( \phi(-U(x_0))f(x_0)IG(x_0) \), which is the LIF(x) defined in Eq. (24).

According to Eq. (9), \( G(x) \) can be treated as a normal distribution variable, and its mean value and variance are \( \mu_0(x) \) and \( \sigma_0(x) \) respectively. If \( M \) is an even number, the last term of Eq. (24) which is a mathematical expectation of \( IG(x_0) \) is calculated as,

\[
E(IG(x_0)) = \int \frac{1}{\sqrt{2\pi}\sigma_0(x)} \exp \left( \frac{-(G(x_0) - u_0(x_0))^2}{2\sigma_0^2(x)} \right) dG(x_0)
\]

\[
= \int_{-\infty}^{+\infty} \left( \frac{1}{\sqrt{2\pi}\sigma_0(x)} \exp \left( \left( G(x_0) - u_0(x_0) \right)^2 / 2\sigma_0^2(x) \right) \right) dt
\]

\[
= 2\sigma_0^2(x) \int_{-\infty}^{+\infty} \left( \frac{1}{\sqrt{2\pi}\sigma_0(x)} \exp \left( \left( G(x_0) - u_0(x_0) \right)^2 / 2\sigma_0^2(x) \right) \right) dt
\]

\[
= 2\sigma_0^2(x) \int_{-\infty}^{+\infty} \left( \frac{1}{\sqrt{2\pi}\sigma_0(x)} \exp \left( \left( G(x_0) - u_0(x_0) \right)^2 / 2\sigma_0^2(x) \right) \right) dt
\]

(25)

Otherwise, if \( M \) is an uneven number, \( E(IG(x_0)) \) is calculated as

\[
E(IG(x_0)) = \int \left( \frac{1}{\sqrt{2\pi}\sigma_0(x)} \exp \left( \left( G(x_0) - u_0(x_0) \right)^2 / 2\sigma_0^2(x) \right) \right) dG(x_0)
\]

\[
= 2\sigma_0^2(x) \int_{-\infty}^{+\infty} \left( \frac{1}{\sqrt{2\pi}\sigma_0(x)} \exp \left( \left( G(x_0) - u_0(x_0) \right)^2 / 2\sigma_0^2(x) \right) \right) dt
\]

\[
= 2\sigma_0^2(x) \int_{-\infty}^{+\infty} \left( \frac{1}{\sqrt{2\pi}\sigma_0(x)} \exp \left( \left( G(x_0) - u_0(x_0) \right)^2 / 2\sigma_0^2(x) \right) \right) dt
\]

(26)
where
\[
\ell = \frac{G(X_0) - U_c(X_0)}{\sigma_c(X_0)}
\]
\[
\int_{-\infty}^{+\infty} \frac{t^n \exp(-t^2/2) dt}{\sqrt{2\pi}} = \left( \frac{m-1}{2} \right) \int_{-\infty}^{+\infty} t^{m-2} \exp(-t^2/2) dt
\]
(27)
\[
\int_{-\infty}^{+\infty} \frac{U_c(X_0)}{\sigma_c(X_0)} t^n \exp(-t^2/2) dt = \exp\left( -\frac{U_c(X_0)^2}{2\sigma_c^2(X_0)} \right)
\]
(28)
\[
\int_{-\infty}^{+\infty} \frac{U_c(X_0)}{\sigma_c(X_0)} \exp(-t^2/2) dt = \sqrt{2\pi} \left( 1 - \Phi\left( -\frac{U_c(X_0)}{\sigma_c(X_0)} \right) \right)
\]
(29)

For any \( x \) in the variable space, LIF(\( x \)) can be obtained by Eqs. (24)-(29):
\[
\phi(-U(x))f(x) = \left\{ \begin{array}{ll}
\phi(-U(x))f(x) & M \text{ is even} \\
\phi(-U(x))f(x) & M \text{ is uneven}
\end{array} \right.
\]
\[
LIF(x) = \left\{ \begin{array}{ll}
\phi(-U(x))f(x) & M \text{ is even} \\
\phi(-U(x))f(x) & M \text{ is uneven}
\end{array} \right. + \sum_{m=1}^{M/2} C_{2m}^{M} \frac{U_c(M-2m)(x)}{\sigma_c^2(M-2m)} (2m-1)!
\]
(30)
The larger LIF(\( x \)) is, the more \( x \) can diminish UF. Then we can add the maximum point of LIF into the DoE and reconstruct the corresponding Kriging model. As a result the uncertainty of \( \hat{P}_f \) converges to zero. Repeat the above process until a certain condition of convergence is satisfied.

Considering an analytical example with 2-dimension input variable [15,20,45], its performance function is
\[
G(x) = \exp(0.4x_1 + 7) - \exp(0.3x_2 + 5) - 200
\]
(31)

It is assumed that the input random variable \( \mathbf{X} = [X_1, X_2]^T \) is subject to standard bivariate normal distribution and \( X_1 \) and \( X_2 \) are independent. The limit state is derived as
\[
x_t = 2.5 \left[ \ln(\exp(0.3x_2 + 5) + 200) - 7 \right]
\]
(32)

Generate 20 random samples with Latin hypercube sampling (LHS), which is the initial DoE. Fig. 1 illustrates the LIF. This 3-D plot announces that LIF may have more than one local maximum point for a given Kriging model, and points far away from the predicted limit state are with LIF values close to zero generally. The latter coincides with the conclusion that the domain \( U(x) < 2 \) contributes most of the UF. Therefore, to search the maximum point of LIF(\( x \)) or determine the best next point, one just needs to search the domain \( U(x) < 2 \) and avoid local maximum points as far as possible. The MCMC method performs the generation of conditional samples efficiently. This characteristic is suitable for generating random points from the conditional distribution \( f(x|U(x) < 2) \). The main reason that the MCMC simulation is adopted rather than an optimization algorithm is that the result from the latter depends on the initial value seriously. The procedure of determining the best point is summarized as follows:

Step 1: Generate \( N_{\text{uf}} \) conditional random points with the MCMC method from the conditional distribution \( f(x|U(x) < 2) \).

Step 2: Determine the best next point among \( \{x_{\text{uf},n} \mid n = 1, 2, \ldots, N_{\text{uf}}\} \). Compute LIF(\( x_{\text{uf},n} \)) according to Eq. (30). And the point that maximizes LIF is defined as the best next point.
\[
x_{\text{best}} = \arg\max\{\text{LIF}(x_{\text{uf},n}) \mid n = 1, 2, \ldots, N_{\text{uf}}\}
\]

4. An adaptive reliability analysis method

In this section, a new adaptive reliability analysis method is proposed based on the Kriging model. In the method, candidate points of the best next one are refreshed with the MCMC method every iteration, and LIF developed in Section 3 is adopted to determine the best next one among \( N_{\text{uf}} \) candidates. The main steps of the proposed method are summarized as follows:

Step 1: Produce initial DoE with LHS method and call the performance function to evaluate the structural response of points in the initial DoE. It is supposed that the input vector \( \mathbf{X} \) is subject to multivariate normal distribution with zero mean and unit variance, which is reasonable in engineering because most of random vectors can be transformed to multivariate normal distribution exactly or approximately. The upper and lower bounds of LHS are set to be 5 and \( -5 \).
respectively. The number of points in the initial DoE is illustrated in detail in Section 5.

Step 2: Construct the Kriging model based on the current DoE according to the theory in Section 2. DACE [28,46–48], a toolbox about Kriging in MATLAB, is used to build the Kriging model and predict performance values and their variances in the variable space.

Step 3: Estimate $\hat{P}_f$ and UF with MCS. The number of random samples ($N_{MC}$) is set to make sure the coefficient of variation of $\hat{P}_f$ no larger than 0.03, which is to say

![Graphs of $\hat{P}_f$ and UF/$\hat{P}_f$ with $N = 6$. Imaginary lines in (a) and (b) correspond to the referenced probability of failure ($4.416 \times 10^{-3}$) and condition of convergence (0.05) respectively.](image_url)
The estimation of UF is denoted as

\[
\Phi \approx -\left( x_{\text{N}} U \right)
\]

Therefore, the condition of convergence in this research is set as

\[
\delta_{\text{MC}} \leq 0.03
\]

\[
\text{UF} \geq \frac{1}{N_{\text{MC}}} \sum_{n=1}^{N_{\text{MC}}} \Phi(-U(x_n))
\]

Step 4: Decide whether \( \hat{\theta}_1 \) and UF satisfy the condition of convergence. As analyzed in Section 3, \( \hat{\theta}_1 \) converges to \( \theta_1 \) when UF is small enough. According to the definition of UF (Eq. (17)), it can be easily proved that UF is a upper limit of the absolute error of \( \hat{\theta}_1 \) relative to \( \theta_1 \) in the sense of mathematic expectation.

\[
E(\hat{\theta}_1 - \theta_1) \leq \text{UF}
\]

Therefore, the condition of convergence in this research is set as

\[
\text{UF} / \hat{\theta}_1 \leq 0.05
\]

Step 5: If \( \hat{\theta}_1 \) and UF satisfy Eq. (35), end the iteration, and \( \hat{\theta}_1 \) is the estimation of the studied structure. Otherwise continue the iterative process.

Step 6: Determine the best next point. As mentioned in Section 4, MCMC simulation is adopted to determine the maximum point of LIF approximately. Generate \( N_{\text{UF}} \) random samples with MCMC method in the domain of integration of Eq. (18). And the maximum point of LIF for \( \{ x_{\text{UF}}, n = 1, 2, ..., N_{\text{UF}} \} \) is treated as the best next point. Add the best next point into DoE and evaluate its structural response. Return to step 2.

5. Academic validation

In this section, two analytical examples and a truss structure are employed to validate the efficiency of LIF and the proposed
reliability analysis method.

5.1. Example 1: a series system with two-dimension input variable

The first analytical example already adopted in [49,5,34] is a series system including four branches. Its performance function is defined as

\[
G(x) = \min \left\{ 3 + 0.1(x_1 - x_2)^2 - (x_1 + x_2)/\sqrt{2} ; \right. \\
\left. 3 + 0.1(x_1 - x_2)^2 + (x_1 + x_2)/\sqrt{2} ; \right. \\
\left. (x_1 - x_2) + 6/\sqrt{2} ; \right. \\
\left. (x_2 - x_1) + 6/\sqrt{2} \right\}
\]

where the input variable \( X = [X_1, X_2]^T \) is subject to standard normal distribution and \( X_1 \) and \( X_2 \) are mutually independent.

The proposed method described in Section 4 is applied to this example. All the results are listed in Table 1. To demonstrate the efficiency of LIF and the method, results from [34] are also summarized. The information provided in Table 1 includes the number of calls to the performance function (\( N_{\text{call}} \)), the estimation of failure probability (\( \hat{P}_f \)) and its corresponding coefficient of variation (\( \delta_{\hat{P}_f} \)), the relative error (\( \Delta_{\hat{P}_f} \)) of \( \hat{P}_f \) compared with reference failure probability.

Fig. 2 shows the procedure that predicted limit state converges to the real one (\( N_0 = 6 \)). From Fig. 2, the initial predicted limit state does not tally with the real, which is the same to the initial \( \hat{P}_f \) (Fig. 3a) and \( UF/\hat{P}_f \) (Fig. 3b). As the number of DoE grows, things change to much better (Figs. 2b, c and 3). Fig. 3a indicates that the proposed method can roughly estimate \( P_f \) with about 20 points, whose accuracy is enough for engineering. Other points are used to make \( \hat{P}_f \) and UF satisfy the condition of convergence (Eq. (35)). When the process converges, the predicted limit state is close to the real in the area of interest. Although they do not coincide with each other in the whole space, the accuracy of \( \hat{P}_f \) is enough.

Fig. 4 compares different methods in terms of the procedure of predicted limit states converging to the real, in which AK-MCS+U,
AK-MCS+EFF and the proposed method are contained. The number of initial DoE is set to be 6 for all methods, and subgraphs in a same row holds the same number of DoE. Fig. 4 clarifies that the proposed method is able to identify the area of importance quickly and approximate the target limit state roughly, which distinguishes it from other methods. One can also realize that the accuracy of methods mentioned in Fig. 4 is no much difference if there are enough points in DoE (subgraphs of the third row). To test how much the number of initial DoE can influence the process of convergence, different values of \( N_0 \) are adopted.

Fig. 6. Comparisons among AK-MCS+U, AK-MCS+LIF and the proposed method in terms of predicted limit states with the same number of DoE points for example 2. \( N_{\text{out}} \) is set to be 100, 200, and 300 respectively.

Fig. 7. Comparisons of different methods in terms of \( \hat{P}_f \) and \( U_{P_f} / \hat{P}_f \) for example 2.
Fig. 8. Comparison between AK-MCS + LIF and the proposed method with $N_0 = 6$. The first two sub graphs, (a) and (b), reflect the convergence procedure of AK-MCS + LIF, and the latter two, (c) and (d), correspond to the proposed method.

Fig. 9. Graphs of $\hat{P}_f$ and $U_F/\hat{P}_f$ with $N_0 = 20$ in (a), (b) and $N_0 = 40$ in (c), (d).
example and both methods are run 3 times with \( N_0 = 6 \) to test their stability. AK-MCS + LIF means that the method raised by [34] is adopted but LIF is employed as the learning function. AK-MCS + LIF here is to demonstrate the outperformance of LIF in the framework of AK-MCS and compare the proposed method to AK-MCS with the same learning function. Results are summarized in Table 2. It is easily noticed that numbers of calls to the performance function for AK-MCS + LIF and the proposed method are lower. Fig. 6 lists the comparisons among predicted limit states from AK-MCS + U, AK-MCS + LIF and the proposed method. Fig. 7 compares AK-MCS + U, AK-MCS + EFF, AK-MCS + LIF and the proposed method, in which \( 2 \times 10^5 \) given random points are employed. Figs. 7(a) and 7(b) show the procedures of \( \hat{P}_L \) converging to \( P_f \) and the decrease of \( UF/\hat{P}_L \) respectively.

Both Figs. 6 and 7 validate the efficiency of LIF developed in this research. Compared with U and EFF, LIF is more adept in recognizing points that are helpful to improve the accuracy of \( \hat{P}_L \) or the Kriging model, which benefits the rough estimate of \( P_f \). AK-MCS + LIF and the proposed method are very similar to each other in the sense of accuracy. From Fig. 8 and Table 2, it is easily noticed that to satisfy the condition of convergence defined by Eq. (35) the proposed method needs more points than AK-MCS + LIF in general and \( \hat{P}_L \) from the latter is very steady. The main difference between them is that the former determines the best next point from a given sample of points along the procedure if the initial random points are enough, while the proposed method generates random candidates with MCMC method every iteration. Fig. 9 shows the graphs of predicted failure probability and the corresponding \( UF/\hat{P}_L \) with \( N_0 \) set to be 20 and 40, whose results are from the proposed method and also listed in Table 2.

5.3. Example 3: truss structure with 10 dimensions

This truss structure has already been studied in [15,18]. As shown in Fig. 10, it contains 11 horizontal bars and 12 sloping ones. \( A_1, E_1 \) denote horizontal bars cross section and Youngs moduli respectively, while \( A_2, E_2 \) denote sloping bars’. 6 loads, from \( P_1 \) to \( P_6 \), are applied on nodes of horizontal bars. These ten random variables are independent, and their distribution information is listed in Table 3.

The deflection of node E in Fig. 10 denoted by \( \mathbf{s}(x) \), is the response of the structure for it is larger than any other nodes. The threshold of \( \mathbf{s}(x) \) is 0.14 m in accordance with [15,18]. Hence the performance function of the studied structure is

\[
G(x) = 0.14 - |\mathbf{s}(x)| \tag{38}
\]

According to [18], the referential failure probability ( \( P_f^{\text{REF}} \)) is \( 3.45 \times 10^{-5} \), which is from IS with 500,000 simulations.

Apply the proposed method to the truss structure, and run it several times with different numbers of initial DoE. Table 4 summarizes the results. Fig. 11 presents graphs of \( \hat{P}_L \) and \( UF/\hat{P}_L \). From Fig. 11, \( \hat{P}_L \) converges to \( P_f \) with desirable accuracy quickly. However, does not satisfy the condition of convergence until the number of DoE is about 150. Similar with Fig. 7, Fig. 12 shows the comparison among AK-MCS + U, AK-MCS + EFF, AK-MCS + LIF and the proposed method. One can notice that methods with LIF are efficient. AK-MCS + LIF and the proposed method make little distinction. However, it takes more than 80 h to perform AK-MCS + LIF, while the proposed method needs less than 10 h. AK-MCS + U and AK-MCS + EFF are approximately equal to AK-MCS + LIF in the sense of running time. The main reason is that the failure of the studied structure is a rare event ( \( 3.45 \times 10^{-5} \)). The number of random points for AK-MCS methods is larger than 40 million. As the scale

---

**Table 3** Distribution information of the truss structure.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Distribution</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_1 - P_6 (N) )</td>
<td>Gumbel</td>
<td>( 5 \times 10^4 )</td>
<td>( 7.5 \times 10^3 )</td>
</tr>
<tr>
<td>( A_1 (m^2) )</td>
<td>Lognormal</td>
<td>( 2 \times 10^{-3} )</td>
<td>( 2 \times 10^{-4} )</td>
</tr>
<tr>
<td>( A_2 (m^2) )</td>
<td>Lognormal</td>
<td>( 1 \times 10^{-1} )</td>
<td>( 1 \times 10^{-4} )</td>
</tr>
<tr>
<td>( E_1 (Pa) )</td>
<td>Lognormal</td>
<td>( 2.1 \times 10^{11} )</td>
<td>( 2.1 \times 10^{10} )</td>
</tr>
<tr>
<td>( E_2 (Pa) )</td>
<td>Lognormal</td>
<td>( 2.1 \times 10^{11} )</td>
<td>( 2.1 \times 10^{10} )</td>
</tr>
</tbody>
</table>

---

**Table 4** Results of example 3.

<table>
<thead>
<tr>
<th>Method</th>
<th>( N_{\text{eff}} )</th>
<th>( \hat{P}_L (10^{-5}) )</th>
<th>( \Delta_{P_f} (%) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS</td>
<td>( 5 \times 10^5 )</td>
<td>3.45</td>
<td>-</td>
</tr>
<tr>
<td>AK-MCS + U</td>
<td>&gt;200</td>
<td>0.330</td>
<td>-2.46</td>
</tr>
<tr>
<td>AK-MCS + EFF</td>
<td>&gt;200</td>
<td>3.27</td>
<td>-2.47</td>
</tr>
<tr>
<td>AK-MCS + LIF</td>
<td>11+110</td>
<td>3.31</td>
<td>-2.51</td>
</tr>
<tr>
<td>The proposed method</td>
<td>15+130</td>
<td>3.49</td>
<td>-2.40</td>
</tr>
<tr>
<td></td>
<td>15+120</td>
<td>3.48</td>
<td>-2.45</td>
</tr>
<tr>
<td></td>
<td>30+130</td>
<td>3.50</td>
<td>-2.06</td>
</tr>
<tr>
<td></td>
<td>30+140</td>
<td>3.41</td>
<td>-2.45</td>
</tr>
<tr>
<td></td>
<td>50+120</td>
<td>3.55</td>
<td>-2.31</td>
</tr>
<tr>
<td></td>
<td>50+100</td>
<td>3.39</td>
<td>-2.04</td>
</tr>
</tbody>
</table>

---

According to Fig. 5a and c, increasing the number of initial DoE improves the quality of initial \( \hat{P}_L \), which results from good initial predicted limit state. However it does not fasten the convergence obviously. On the contrary, it may negatively affect the number of calls to the performance function (Table 1). The comparison of different number of initial DoE is also detailed in the following examples.

5.2. Example 2: modified Rastrigin function

This two-dimension analytical example already analyzed in [9,34,49] consists of a modified Rastrigin function. The performance function is

\[
G(x) = 10 - \sum_{m=1}^{2} \left( x_m^2 - 5 \cos(2 \pi x_m) \right) \tag{37}
\]

The input variable \( \mathbf{x} = [x_1, x_2]^T \) is subject to standard normal distribution and \( X_1 \) and \( X_2 \) are mutually independent. The limit state of Eq. (37) is much more complicated than the one in Section 5.1 (Fig. 2). In this research, this example is employed to demonstrate the efficiency of LIF and the proposed method and illustrate the different between some other Kriging based active methods and the proposed one.

AK-MCS + LIF and the proposed method are applied to this
of DoE grows, these methods are much more time-consuming than the proposed method to determine a best next point.

6. Conclusion

In this research, a new learning function named least improvement function and a reliability analysis method are proposed. Compared with existing learning functions, LIF takes the joint probability density function of input variables into account and is used to approximately indicate how much the accuracy of $P_f$ will be improved in the sense of mathematical expectation if adding a given point into the DoE. It obtains a compromise among the Kriging variance, estimated response and joint probability density function. Based on the Kriging model and LIF, a new reliability analysis method is developed. In the method, MCMC is used to generate conditional random points as candidates of the best next one. The proposed learning function LIF quantifies how much a given point $x$ can improve the accuracy of the Kriging model and determines the best next one. Three examples are analyzed. Results show that LIF and the proposed method are very efficient when dealing with nonlinear performance function, small probability, complicated limit state and engineering problems with high dimension. By analyzing example 2 and example 3, the conclusion is reached that AK-MCS + LIF and the proposed method perform similarly in the sense of accuracy and the number of calls to the target performance function. The proposed method is more efficient in determining the best next point when dealing with
problems with a high dimension input vector and a rare event to failure. From Tables 1, 2 and 4, larger number of initial DoE is likely to result in larger number of calls to performance function for the proposed learning function and method, especially when the performance function is high nonlinear or with many input variables. The condition of convergence in the proposed method is not religious in theory. Integrating all results into account, it is obvious that the condition of convergence defined by Eq. (35) is too rigorous. As mentioned in Section 3.1, UF measures the accuracy of \( \hat{\beta} \) or its corresponding Kriging model. Actually, UF is an upper limit of the error of \( \hat{\beta} \) relative to \( \beta \) in the sense of mathematical expectation. It is a difficult task to estimate how much UF is larger than the real error of \( \hat{\beta} \), which is various with problems and Kriging models. Therefore, the proposed condition of convergence behaves differently for the studied examples. The accuracy measurement of \( \hat{\beta} \) or the Kriging model benefits not only the timely termination of reliability analysis procedure but also the improvement of the Kriging model or the efficient DoE strategy. Finding a reasonable condition of convergence or a more accurate measurement of \( \hat{\beta} \) for Kriging based reliability analysis method is an important part of our future work.

Acknowledgment

The study was funded by National Science and Technology Major Project of China (Grant No. 2013ZX04011-011) and the Fundamental Research Funds for the Central Universities of China (Grant No. N140306004). Their financial supports are gratefully acknowledged.

Analysis of the uncertainty of Eq. (34)

To simplify the description, a new symbol \( U_{\text{MC}} \) is introduced:

\[
\text{var}(U_{\text{MC}}) = \frac{1}{N_{\text{MC}}} \text{var} \left( \sum_{n=1}^{N_{\text{MC}}} \phi(-U(x_n)) \right) = \frac{1}{N_{\text{MC}}} \text{var} \left( \phi(-U(x)) \right) \\
\approx \frac{1}{N_{\text{MC}}} \sum_{n=1}^{N_{\text{MC}}} \phi(-U(x_n)) - \left( \frac{1}{N_{\text{MC}}} \sum_{n=1}^{N_{\text{MC}}} \phi(-U(x_n)) \right)^2 \\
< \frac{1}{2N_{\text{MC}}} \sum_{n=1}^{N_{\text{MC}}} \phi(-U(x_n))
\]

Hence,

\[
\sqrt{\text{var}(U_{\text{MC}})} < \frac{1}{\sqrt{2N_{\text{MC}}}} \sum_{n=1}^{N_{\text{MC}}} \phi(-U(x_n))
\]

The number of random points for MCS \( (N_{\text{MC}}) \) satisfies Eq. (33):

\[
\frac{1 - \frac{\hat{\beta}}{N_{\text{MC}} \hat{\beta}}} {N_{\text{MC}} \hat{\beta}} \leq 0.03
\]

Therefore,

\[ N_{\text{MC}} \hat{\beta} \geq 1 - \frac{\hat{\beta}}{0.03^2} \approx 1111 \]

If \( U_{\text{MC}} \) meets the condition of convergence defined by Eq. (35), then,

\[
U_{\text{MC}} \leq 1.65 \frac{\sqrt{\text{var}(U_{\text{MC}})}}{\sqrt{N_{\text{MC}}} \hat{\beta}} = \sum_{n=1}^{N_{\text{MC}}} \phi(-U(x_n)) \leq 0.05
\]

According to the central limit theorem,

\[
P(\frac{U}{\hat{\beta}} < 0.058) \geq 0.95
\]

Integrating Eqs. (39)–(41),

\[
1.65 \sqrt{\text{var}(U_{\text{MC}})} = 1.65 \frac{\sqrt{\text{var}(U_{\text{MC}})}}{\sqrt{2N_{\text{MC}}} \hat{\beta}} = \frac{1.65 \sqrt{0.05 N_{\text{MC}}} \hat{\beta}}{\sqrt{2N_{\text{MC}}} \hat{\beta}} < 0.0078
\]

Then,

\[
P\left( \frac{U}{\hat{\beta}} < 0.058 \right) \geq 0.95
\]

It can be concluded that the real value of UF is less than 0.058, with confidence coefficient of 95% at least if the estimate of UF \( (U_{\text{MC}}) \) satisfies Eq. (35).

References
