Proceedings of DETC'01
ASME 2001 Design Engineering Technical Conference
and Computers and Information in Engineering Conference
Pittsburgh, PA, September 9-12, 2001

DETC2001/VIB-21398

BAYESIAN UPDATING OF NONLINEAR MODEL PREDICTIONS USING MARKOV CHAIN MONTE CARLO SIMULATION

James L. Beck*
Division of Engineering and Applied Science
California Institute of Technology
Pasadena, California, 91125
Email: jimbeck@caltech.edu

S. K. Au
K.-V. Yuen
Division of Engineering and Applied Science
California Institute of Technology
Pasadena, California, 91125
Email: siu@caltech.edu,kvyuen@caltech.edu

ABSTRACT

The usual practice in system identification is to use system data to identify one model from a set of possible models and then to use this model for predicting system behavior. In contrast, the present robust predictive approach rigorously combines the predictions of all the possible models, appropriately weighted by their updated probabilities based on the data. This Bayesian system identification approach is applied to update the robust reliability of a dynamical system based on its measured response time histories. A Markov chain simulation method based on the Metropolis-Hastings algorithm and an adaptive scheme is proposed to evaluate the robust reliability integrals. An example for updating the reliability of a Duffing oscillator is given to illustrate the proposed method.

INTRODUCTION

Let $\mathcal{D}$ denote some dynamic data obtained from a dynamical system to give information on the model parameters $\theta = [\theta_1, \theta_2, \ldots, \theta_n]$ that define a model within a set of possible models for the system specified by $\mathcal{M}$. Here, $\mathcal{M}$ specifies the modeling assumptions used in the analysis (both structural and probabilistic). In Bayesian system identification, $\mathcal{D}$ modifies the knowledge about the relative plausibilities of the different models in $\mathcal{M}$, that is, the different values that the model parameters θ may assume (Beck and Katafygiotis, 1998). These plausibilities are measured by the updated probability density function (PDF), $p_{\mathcal{D}}(\theta) = p(\theta|\mathcal{D}, \mathcal{M})$, obtained using Bayes' Theorem:

$$\begin{aligned} p_{\mathcal{D}}(\theta) &= p(\mathcal{D}|\theta, \mathcal{M}) p_0(\theta) / p(\mathcal{D}|\mathcal{M}) \\ &= c\, p(\mathcal{D}|\theta, \mathcal{M}) p_0(\theta) \end{aligned} \tag{1}$$

where $c^{-1} = p(\mathcal{D}|\mathcal{M}) = \int p(\mathcal{D}|\theta, \mathcal{M}) p_0(\theta) d\theta$ is a normalizing constant, and $p_0(\theta) = p(\theta|\mathcal{M})$ is the 'prior' PDF specified by $\mathcal{M}$ which reflects the knowledge about θ in the absence of the data $\mathcal{D}$. The term $p(\mathcal{D}|\theta, \mathcal{M})$ is the probability of obtaining the data $\mathcal{D}$ for a given set of model parameters θ. It is formulated by assuming a probability model for the prediction error relating the model output with model parameters θ to the actual system output $\mathcal{D}$ (Beck and Katafygiotis, 1998).

Better response predictions can be made by utilizing the information in the data $\mathcal{D}$ through the use of the updated PDF $p_{\mathcal{D}}$. In particular, if $P_F(\theta) = P(F|\theta, \mathcal{M})$ denotes the probability of failure of the dynamical system (under some stochastic description of its environment) which is predicted using the model specified by θ, then the initial robust failure probability

$$R_0 = P(F|\mathcal{M}) = \int P_F(\theta) p_0(\theta) d\theta \tag{2}$$

may be updated to give

$$R_{\mathcal{D}} = P(F|\mathcal{M}, \mathcal{D}) = \int P_F(\theta) p_{\mathcal{D}}(\theta) d\theta \tag{3}$$

*Address all correspondence to this author.

where the Theorem of Total Probability is used (Papadimitriou et al., 2001). Note that R_0 and $R_{\mathcal{D}}$ are said to provide a robust reliability for the system because they incorporate modeling uncertainties. Also, $R_{\mathcal{D}}$ not only incorporates knowledge about θ from $\mathcal{M}$ but also updated information from $\mathcal{D}$.

The evaluation of $R_{\mathcal{D}}$ in (3) is difficult since the dimension of the parameter space is usually too large for numerical integration. Also, $p_{\mathcal{D}}$ is known only up to a multiplicative constant because the normalizing constant c is itself difficult to evaluate. In terms of quantities whose values can be computed explicitly for a given θ, $R_{\mathcal{D}}$ can be expressed as (leaving dependence on $\mathcal{M}$ implicit):

$$R_{\mathcal{D}} = \frac{\int P_F(\theta) p(\mathcal{D}|\theta) p_0(\theta) d\theta}{\int p(\mathcal{D}|\theta) p_0(\theta) d\theta} \tag{4}$$

and so its evaluation involves two multi-dimensional integrals. Additional difficulties come from the nature of the updated PDF $p_{\mathcal{D}}$ which is usually concentrated in a small volume of the parameter space. In practical applications, the variation of $p_{\mathcal{D}}$ in the parameter space is more dominant than that of the response quantity $P_F(\theta)$, and so methods for evaluating $R_{\mathcal{D}}$ are differentiated according to the shape of $p_{\mathcal{D}}$, which depends on the information that the available data $\mathcal{D}$ produces on the model parameters θ.

The characteristics of $p_{\mathcal{D}}$ have been studied for model identification of structures using as dynamic data the measured time histories of response and excitation (Beck and Katafygiotis, 1998; Katafygiotis and Beck, 1998; Katafygiotis et al., 1998), but the results can be applied in general. For a large amount of data (e.g., the number of data points M in the measured time histories is large), $p_{\mathcal{D}}$ is concentrated in the neighborhood $\mathcal{N}$ of a lower dimensional manifold $\mathcal{S}$ in the parameter space, on which $p(\mathcal{D}|\theta)$ is globally maximized. The thickness of $\mathcal{N}$ around $\mathcal{S}$ is of the order of $\varepsilon = 1/\sqrt{M}$. The characterization of $p_{\mathcal{D}}$ can be made according to the dimension of $\mathcal{S}$, which depends on the amount of data available compared to the information to be extracted from it. In the 'identifiable' case where the number of model parameters is less than or equal to the number of 'effective constraints' from the data $\mathcal{D}$, the dimension of $\mathcal{S}$ is zero, that is, the updated PDF is concentrated in the close neighborhood $\mathcal{N}$ of a finite number of isolated points, referred to as 'optimal parameter' points. In this case, $p_{\mathcal{D}}$ can be well approximated by a weighted sum of Gaussian PDFs with spread of $O(\varepsilon)$ centered at the optimal parameter points (Beck and Katafygiotis, 1998). Consequently, $R_{\mathcal{D}}$ can be approximated as a weighted sum of values of $P_F(\theta)$ evaluated at the optimal parameter points, and the problem of finding $R_{\mathcal{D}}$ in the identifiable case is reduced to finding the optimal parameter points and their associated probability weights. This leads to a non-convex global optimization problem which is not trivial to solve. There may be multiple optimal parameter points (Katafygiotis and Beck, 1998), so 'global' optimization algorithms should be used to find the optimal parameter points (Törn and Zilinskas, 1989; Yang and Beck, 1998). The computational effort is often much greater than is needed in a local optimization problem, especially when the dimension of the parameter space is not small and the number of optimal parameters is unknown, as is usually the case.

In the 'unidentifiable' case where the number of model parameters is larger than the number of effective data constraints, the dimension of the manifold $\mathcal{S}$ is greater than zero, and there exists a continuum of optimal parameter points lying on $\mathcal{S}$ which give the same global maximum value of $p(\mathcal{D}|\theta)$. Deterministic search methods for computing $R_{\mathcal{D}}$ in (4) have been developed (Katafygiotis et al., 1998), which discretize the manifold $\mathcal{S}$ by a finite number of representative points and approximate $p_{\mathcal{D}}$ as a discrete probability mass distributed among the representative points. The main computational effort is spent on locating the representative points on $\mathcal{S}$, which requires a series of local optimizations in the parameter space. The representative points have to be located over the region of $\mathcal{S}$ where the prior PDF p_0 is significant so that the contributions from different parts of $\mathcal{S}$ and hence $\mathcal{N}$ are accounted for. Consequently, the complexity and computational effort are expected to grow in a similar manner to that of numerical integration as the extent and dimension of the manifold increases.

This paper presents a Markov chain simulation method to evaluate the robust reliability integral without the need for optimization to find the manifold $\mathcal{S}$. It is based on the Metropolis-Hastings (MH) algorithm and an adaptive scheme similar to that of simulated annealing to gain information about the manifold in a gradual manner. By carrying out a series of Markov chain simulations with limiting stationary distributions equal to a sequence of intermediate PDFs, the region $\mathcal{N}$ of significant probability density of $p_{\mathcal{D}}$ is gradually portrayed. The Markov chain samples can be used for statistical averaging to estimate the robust reliability integrals.

DIFFICULTIES IN EVALUATING $R_{\mathcal{D}}$ BY SIMULATION

Equations (3) and (4) suggest two ways of evaluating $R_{\mathcal{D}}$ by simulation. The former suggests estimating $R_{\mathcal{D}}$ as the average of $P_F(\theta)$ over samples simulated from $p_{\mathcal{D}}$, while the latter indicates that the integrals in the numerator and denominator could be estimated individually and then combined to give an estimate for $R_{\mathcal{D}}$. Since $p_{\mathcal{D}}$ is only known up to a multiplicative constant and in general a method for simulating independent samples from $p_{\mathcal{D}}$ is not available, the first option of using (3) is not feasible for existing methods such as Monte Carlo simulation (MCS) or importance sampling (Rubinstein, 1981). Therefore, consider the application of these existing methods to evaluate $R_{\mathcal{D}}$ based on (4). Using MCS, the numerator is estimated as the average of $P_F(\theta)p(\mathcal{D}|\theta)$ over samples drawn from the prior PDF p_0 (assuming a method for simulating samples from p_0 is available). The resulting estimate, however, is very likely to be biased, since $p(\mathcal{D}|\theta)$ is concentrated in the small neighborhood $\mathcal{N}$ of thick-

ness $O(\varepsilon)$, and hence the chance of generating a sample from p_0 which lies in $\mathcal{N}$ is extremely small. Similar difficulties will be encountered in evaluating the denominator. Using importance sampling, it is necessary to choose a sampling density that is concentrated in $\mathcal{N}$, otherwise similar problems to those in MCS will be encountered. However, this is extremely difficult since information about the manifold $\mathcal{S}$ where the probability density is concentrated is not available.

Next, consider again evaluating $R_{\mathcal{D}}$ based on (3). Although a method for generating independent samples according to $p_{\mathcal{D}}$ is generally not available, it is noted that Markov chain simulation, in particular, the MH (Metropolis-Hastings) algorithm, offers a feasible way to simulate samples according to an arbitrary distribution, at the expense of introducing dependence about the samples. However, direct application of the MH algorithm to simulate Markov chain samples according to $p_{\mathcal{D}}$ is not feasible due to the small region $\mathcal{N}$ of probability concentration of $p_{\mathcal{D}}$. Nevertheless, as our proposed method is built on the MH algorithm, we first discuss its implementation.

METROPOLIS-HASTINGS ALGORITHM

The Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970) is a simple procedure to simulate samples according to an arbitrary PDF where the target PDF need only be known up to a scaling constant. It was originally developed by Metropolis and his co-workers and later generalized by Hastings. Its potential use for solving reliability problems has been recently demonstrated by (Au and Beck, 1999) and (Au and Beck, 2000).

In the MH method, samples are simulated as the states of a special Markov chain whose limiting stationary distribution is equal to the target PDF. In other words, the PDF of the Markov chain sample θ_k simulated at the k-th Markov step tends to the target PDF as $k \to \infty$. The Markov chain samples, which are dependent in general, can be used for statistical averaging as if they were independent, although with some reduction of efficiency in the estimator.

Let $p^*(\xi|\theta)$ be a chosen PDF, called the 'proposal PDF', which is a PDF for ξ that depends on θ. The role of p^* will become clear shortly. For convenience in notation, let $q(\theta) = p(\mathcal{D}|\theta)p_0(\theta) = c^{-1}p_{\mathcal{D}}(\theta)$. Note that the value of q can be computed readily for a given θ, while the same is not true for $p_{\mathcal{D}}$. The MH algorithm to simulate Markov chain samples $\{\theta_1, \theta_2, \ldots, \theta_{N_s}\}$ with limiting stationary distribution equal to the target PDF $p_{\mathcal{D}}$ is described as follows. To start the Markov chain, let θ_1 be a point chosen either deterministically or simulated according to some PDF which approximates $p_{\mathcal{D}}$. In general, to simulate the next sample θ_{k+1} from the current sample θ_k, $k = 1, 2, \ldots, N_s - 1$, first simulate a 'candidate state' ξ from the proposal PDF $p^*(\xi|\theta_k)$. Compute the ratio

$$r(\xi, \theta_k) = [q(\xi)p^*(\theta_k|\xi)]/[q(\theta_k)p^*(\xi|\theta_k)] \tag{5}$$

Then, accept the candidate state ξ with probability $\min\{1, r\}$ and reject with the remaining probability $1 - \min\{1, r\}$. If accepted, the candidate state will be taken as the next state of the Markov chain, i.e., $\theta_{k+1} = \xi$. Otherwise, the current state is taken as the next state, i.e., $\theta_{k+1} = \theta_k$. The process is repeated until N_s Markov chain samples have been simulated.

We now show that the next sample θ_{k+1} will be distributed as $p_{\mathcal{D}}$ if the current sample θ_k is. From the Chapman-Kolmogorov equation (Doob, 1953), the PDF of the next sample is given by

$$p(\theta_{k+1}) = \int p(\theta_{k+1}|\theta_k)p(\theta_k)d\theta_k \tag{6}$$

where $p(\theta_{k+1}|\theta_k)$ is the transition PDF governing the probabilistic properties of the Markov chain. From the MH algorithm, the transition PDF is given by, for $\theta_{k+1} \neq \theta_k$:

$$p(\theta_{k+1}|\theta_k) = p^*(\theta_{k+1}|\theta_k)\min\{1, r(\theta_{k+1}, \theta_k)\} \tag{7}$$

Using (5) and (7), along with the identity $\min\{1, a/b\}b = \min\{1, b/a\}a$ for any positive numbers a and b, and the fact that q differs from $p_{\mathcal{D}}$ only by a normalizing constant, one can readily show the following 'reversibility' condition:

$$p(\theta_{k+1}|\theta_k)p_{\mathcal{D}}(\theta_k) = p(\theta_k|\theta_{k+1})p_{\mathcal{D}}(\theta_{k+1}) \tag{8}$$

Note that (8) is trivial for $\theta_{k+1} = \theta_k$. Assuming the current sample θ_k is distributed as $p_{\mathcal{D}}$, i.e., $p(\theta_k) = p_{\mathcal{D}}(\theta_k)$, and using the reversibility condition in (8), $p(\theta_{k+1})$ in (6) becomes $p(\theta_{k+1}) = \int p(\theta_k|\theta_{k+1})p_{\mathcal{D}}(\theta_{k+1})d\theta_k = p_{\mathcal{D}}(\theta_{k+1})\int p(\theta_k|\theta_{k+1})d\theta_k = p_{\mathcal{D}}(\theta_{k+1})$ since $\int p(\theta_k|\theta_{k+1})d\theta_k = 1$. This means that if the current sample θ_k is distributed as the target PDF $p_{\mathcal{D}}$, then so is the next sample θ_{k+1}, and hence $p_{\mathcal{D}}$ is the stationary PDF of the Markov chain.

In the actual implementation, the Markov chain is started with the initial state θ_1 simulated from a PDF different from $p_{\mathcal{D}}$, so the Markov chain is in a transient state and its samples will not be distributed exactly as $p_{\mathcal{D}}$. Under the assumption of ergodicity, however, the Markov chain will converge to the stationary state, and so the PDF of θ_k will tend to $p_{\mathcal{D}}$ as $k \to \infty$. With a finite sample size used in the actual implementation, ergodicity often becomes an issue of whether the Markov chain samples can populate sufficiently well the region of significant probability of the target PDF $p_{\mathcal{D}}$. See (Au and Beck, 1999) and (Au and Beck, 2000) for a more detailed discussion of ergodicity in applying the MH method to reliability problems.

Using the Markov chain samples $\{\theta_1, \theta_2, \ldots, \theta_{N_s}\}$, $R_{\mathcal{D}}$ is estimated as the average $\tilde{R}_{\mathcal{D}}$ of $P_F(\theta)$ over the samples, which is the same as the usual MCS estimator, except that the samples are simulated from a Markov chain instead of being independent

and identically distributed (i.i.d.). Nevertheless, the estimator $\tilde{R}_{\mathcal{D}}$ has similar statistical properties as those of MCS estimators (see later). In order to reduce the initial transient effect of the Markov chain on the estimate, the first few samples (say 10) are often not used to compute the estimate $\tilde{R}_{\mathcal{D}}$. Unless otherwise stated, the Markov chain samples $\{\theta_1, \theta_2, \ldots, \theta_{N_s}\}$ used for computing the estimate $\tilde{R}_{\mathcal{D}}$ are those simulated after the initial transient stage.

The proposal PDF p^* affects the distribution of the candidate state ξ given the current state, and consequently the convergence rate of the estimator $\tilde{R}_{\mathcal{D}}$ to $R_{\mathcal{D}}$. If the candidate state is rejected too often because small values of r in (5) are encountered during simulation, the Markov chain will consist of many repeated samples. As a result, the correlation among samples will be increased, slowing down the convergence of $\tilde{R}_{\mathcal{D}}$. To understand how p^* affects the acceptance rate of the candidate state, first note that if $p^*(\xi|\theta) = p_{\mathcal{D}}(\xi)$, then according to (5), $r = 1$ and hence the candidate state is always accepted. In this case, the MH algorithm reduces to a standard Monte Carlo procedure with i.i.d. samples simulated from $p_{\mathcal{D}}$. Although this choice is not possible, it indicates that if p^* is chosen to be 'non-adaptive', i.e., $p^*(\xi|\theta) = p^*(\xi)$, then the closer that p^* is to the target PDF, the better the acceptance rate of the candidate state, and the faster the convergence.

Choosing a non-adaptive proposal PDF p^* means that the information from the current sample is not used to explore the important region of significant probability density of the target PDF during simulation. When information about the important region is not available, constructing a non-adaptive p^* so that the candidate state simulated from it will lie in $\mathcal{N}$ is similar to constructing an importance sampling density concentrated in $\mathcal{N}$, and is thus very difficult. It is therefore more desirable to choose an 'adaptive' p^* which depends on the current sample. One popular choice is to have p^* localized and symmetric, i.e., $p^*(\xi|\theta) = p^*(\theta|\xi)$, which corresponds to the one used in the original algorithm (Metropolis et al., 1953). In this case, p^* can be interpreted as a PDF localized at the current sample. The Markov chain simulation process can then be viewed as a 'local random walk' in which the region of probability concentration of the target PDF is adaptively explored.

Direct application of the MH algorithm to simulate samples according to the target PDF $p_{\mathcal{D}}$ is not feasible, however, due to the problems arising from the small region $\mathcal{N}$ of probability concentration of $p_{\mathcal{D}}$, as described for the other methods discussed in the last section. In particular, it is difficult to choose the proposal PDF p^* so that the acceptance rate of the candidate state is not too small while at the same time the Markov chain samples effectively explore $\mathcal{N}$. To see this, first note that it is not possible to choose a non-adaptive p^* which can generate samples lying in $\mathcal{N}$, since the information about the manifold is not available. Thus, consider choosing an adaptive p^*, such as a symmetric one in the original Metropolis algorithm. If the Markov chain is started in a region not near $\mathcal{N}$, then the chance of generating a candidate state from p^* that visits $\mathcal{N}$ is extremely small, and most of the candidate states will be rejected. As a result, most of the Markov chain samples are repeated, and clearly they cannot be used to estimate $R_{\mathcal{D}}$. On the other hand, if the Markov chain is started in $\mathcal{N}$, then in order that the simulated candidate state remain in $\mathcal{N}$ and have a high probability of being accepted, the spread of p^* around the current sample has to be $O(\varepsilon)$, because the direction along which $\mathcal{N}$ extends around the current sample is not known. But this means the candidate state will be very close to the current sample, and as a result the Markov chain samples will not efficiently explore $\mathcal{N}$. In all these cases, the region visited by the Markov chain samples will be small compared to $\mathcal{N}$, leading to significant bias in the estimate for $R_{\mathcal{D}}$.

PROPOSED METHOD

The problems encountered in applying the simulation methods discussed in the previous sections, including the MH algorithm, arise from the fact that the updated PDF $p_{\mathcal{D}}$ is concentrated in a small neighborhood $\mathcal{N}$ of the manifold $\mathcal{S}$. The process of adapting samples to $\mathcal{N}$ in the MH algorithm is inhibited by the small scale ε of the thickness of $\mathcal{N}$ compared to the size of the proposal PDF required to cover $\mathcal{N}$ with an affordable number of samples. This suggests that direct adaptation using a proposal PDF which varies with a vastly different length scale from that of the target PDF will not be effective. In view of this, we introduce a sequence of intermediate PDFs which bridge the gap in length scale between the prior PDF p_0 and the target updated PDF $p_{\mathcal{D}}$. By using the MH algorithm utilizing successively the information from the previous adapted intermediate PDFs, the region $\mathcal{N}$ can be populated by the Markov chain samples in a gradual manner. Conceptually, this is similar to 'simulated annealing' (Fishman, 1996).

Let $p_1, p_2, \ldots, p_m$ be a chosen sequence of PDFs converging to $p_{\mathcal{D}}$ $(= p_m)$ so that their region of significant probability content gradually diminishes to that of $p_{\mathcal{D}}$. For example, p_i may be chosen as the updated PDF from Bayes' Theorem based on an increasing amount of data, i.e., $p_i = p_{\mathcal{D}_i}$, where $\mathcal{D}_1 \subset \cdots \subset \mathcal{D}_m = \mathcal{D}$. Thus, starting with the prior PDF p_0 as the proposal PDF, the MH algorithm is carried out to simulate samples $\{\theta_1^{(1)}, \theta_2^{(1)}, \ldots, \theta_{N_s}^{(1)}\}$ with target PDF p_1. A kernel sampling density k_1 is constructed to approximate p_1 (Silverman, 1986; Au and Beck, 1999):

$$k_1(\theta) = \sum_{k=1}^{N_s} w_k \phi(\theta; \theta_k^{(1)}, C_k) \tag{9}$$

where $\phi(\theta; \theta_k^{(1)}, C_k)$ is the multi-dimensional Gaussian PDF evaluated at θ with mean $\theta_k^{(1)}$ and covariance matrix C_k; and the w_k are the probability weights associated with the Gaussian PDFs,

which satisfy $w_k \geq 0$ and $\sum_{k=1}^{N_s} w_k = 1$. The choice of w_k and C_k is completely defined by the samples $\{\theta_k^{(1)}\}$, and so is k_1 (Au and Beck, 1999). Since k_1 is a weighted sum of Gaussian PDFs, the kernel marginal PDF for a particular component θ_j of θ can be obtained by analytically integrating (9). Similarly, the kernel marginal cumulative distribution function (CDF) of θ_j can be readily computed in terms of a weighted sum of Gaussian CDFs. More importantly, independent samples distributed as k_1 can be simulated readily, which are approximately distributed as p_1 and so lie in the region of significant probability of p_1.

To proceed, k_1 is used as the proposal PDF for simulating Markov chain samples $\{\theta_1^{(2)}, \theta_2^{(2)}, \ldots, \theta_{N_s}^{(2)}\}$ with target PDF p_2. These samples are then used to construct the kernel sampling density k_2 as in (9), which gives an approximation to p_2. In general, the kernel sampling density k_i (which approximates p_i) is constructed using the Markov chain samples at the i-th simulation level which is then used as the proposal PDF for simulating Markov chain samples for the next level with target PDF p_{i+1}. This is continued until the m-th simulation level, where Markov chain samples for the target updated PDF $p_{\mathcal{D}} = p_m$ are simulated.

Let R_i be the expectation of $P_F(\theta)$ when θ is distributed as p_i. It is estimated using the Markov chain samples $\{\theta_1^{(i)}, \theta_2^{(i)}, \ldots, \theta_{N_s}^{(i)}\}$ by

$$R_i \approx \tilde{R}_i = \frac{1}{N_s} \sum_{k=1}^{N_s} P_F(\theta_k^{(i)}) \tag{10}$$

Since p_i converges to $p_{\mathcal{D}}$ as the simulation level i increases to m, R_i converges to $R_{\mathcal{D}}$ defined by (3).

The proposed strategy makes use of the fact that if a non-adaptive proposed PDF is similar in shape to the target PDF, then the acceptance rate of the candidate state will not be small, and the MH algorithm will be effective in generating samples populating the important region of the target PDF. Thus, instead of using an adaptive proposal PDF within each simulation level, the adaptation is done from one simulation level to the next, where the latest kernel sampling density transfers the information about the important region from the current level to the next level.

The success of the proposed strategy relies on effective application of the MH algorithm at each simulation level, which requires that p_{i-1} (which is approximated by k_{i-1}) varies with a similar length scale to p_i for $i = 1, 2, \ldots, m$. The choice of the sequence $\{p_i\}$ is thus important to the success of the proposed method. A good choice for $\{p_i\}$ is $p_{i+1} \sim p_i^2$ (up to a normalizing constant). In this case, the size of the region of probability concentration for p_i is roughly $1/\sqrt{i}$ that of p_0. When the data $\mathcal{D}$ consists of measured response time histories (Beck and Katafygiotis, 1998), this schedule can be achieved by choosing $p_i = p_{\mathcal{D}_i}$ where the duration of time history data is doubled in successive simulation levels from $\mathcal{D}_i$ to $\mathcal{D}_{i+1}$. On the other hand, when the updated PDF with data $\mathcal{D}$ is of the form $p_{\mathcal{D}}(\theta) = c p_0(\theta) \exp(-J(\theta)/\varepsilon^2)$ (Yuen and Beck, 2001), where $J(\theta)$ is a measure-of-fit function between data and model, and ε is a measure of the size of the prediction error, then the sequence $\{p_i\}$ can be obtained by $p_i = c_i \exp(-J(\theta)/\varepsilon_i^2)$ where $\varepsilon_i^2 = 2^{m-i}\varepsilon^2$, $i = 1, 2, \ldots, m$, with $2^m \approx \varepsilon^{-2}$.

STATISTICAL PROPERTIES OF ESTIMATOR

The statistical properties of the estimator $\tilde{R}_i$ in (10) are presented in this section, assuming the Markov chain generated according to the MH algorithm at each simulation level is ergodic.

In spite of the fact that $\tilde{R}_i$ is computed using dependent samples from a Markov chain, it still has the usual properties of MCS estimators using i.i.d. samples (Doob, 1953). For example, $\tilde{R}_i$ converges to R_i with probability 1 as $N_s \to \infty$ (Strong Law of Large Numbers), and under similar conditions as those for Monte Carlo estimators, $\tilde{R}_i$ is Normally distributed as $N_s \to \infty$ (Central Limit Theorem). If the Markov chain is started with the initial state $\theta_1^{(i)}$ distributed as the target PDF p_i, then the Markov chain is stationary, and $\tilde{R}_i$ is unbiased. Otherwise, $\tilde{R}_i$ is only asymptotically unbiased, although the bias decays exponentially with the number of Markov steps. For a fixed proposal PDF for the i-th simulation level, assuming that the Markov chain has settled into its stationary state, it can be shown that the coefficient of variation (c.o.v.) of $\tilde{R}_i$ is:

$$\delta_i^2 = (1 + \gamma_i)\Delta_i^2 / N_s \tag{11}$$

where Δ_i is the c.o.v. of $P_F(\theta)$ when θ is distributed as p_i; γ_i is a correlation factor:

$$\gamma_i = 2 \sum_{k=1}^{N_s - 1} (1 - \frac{k}{N_s}) \rho_i(k) \tag{12}$$

and $\rho_i(k)$ is the correlation coefficient between $P_F(\theta)$ evaluated at Markov chain samples separated by k Markov steps. By estimating the correlation sequence $\{\rho_i(k)\}$ from the Markov chain samples, γ_i in (12), and hence δ_i in (11), can be estimated in a single simulation run.

The term Δ_i^2/N_s in (11) is the familiar term for the square of c.o.v. in MCS with N_s independent samples. The c.o.v. of $\tilde{R}_i$ is thus equivalent to the c.o.v. in MCS with an effective number of independent samples $N_s/(1+\gamma_i)$. The efficiency of the estimator using correlated samples of a Markov chain ($\gamma_i > 0$) is reduced compared to the case when the samples are uncorrelated ($\gamma_i = 0$), and smaller values of γ_i imply higher efficiency.

The result for δ_i in (11) is derived assuming the proposal PDF is fixed in independent simulation runs. According to the proposed methodology, however, the proposal PDF is chosen

as the kernel sampling density k_i constructed using the Markov chain samples from the previous simulation level (except for the first simulation level where the prior PDF is used as the proposal PDF), and so k_i is different for each independent simulation run. This gives rise to additional variability in $\tilde{R}_i$, and the actual c.o.v. of $\tilde{R}_i$ will be greater than that given by (11). It can be argued that this additional variability is of the order of the bias in $\tilde{R}_i$, and hence is often small. The numerical example shows that the c.o.v. predicted by (11) is quite close to the actual c.o.v., and hence it can be used for assessing the variability of $\tilde{R}_i$.

NUMERICAL APPLICATIONS

The proposed methodology is applied to update the reliability of a Duffing oscillator using simulated response measurements. The excitation f is assumed to be unknown and is modeled by a white noise random process with zero mean and spectral intensity $S_o = 0.01m^2\sec^{-3}$. The equation of motion of the system is given by:

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) + \mu x^3(t) = f(t) \tag{13}$$

where $m = 1kg$ is the mass, $c = 0.1kg\sec$ is the damping coefficient and $k = 4.0\text{N/m}$ and $\mu = 1.0\text{N/m}^3$ are the first and third order stiffness of the system. To define a set of possible models for the system, the stiffnesses are parameterized as $k = \bar{k}\theta_1$ and $\mu = \bar{\mu}\theta_2$, where θ_1 and θ_2 are stiffness parameters whose values are assumed to be uncertain, and $\bar{k} = 4.0$ N/m, $\bar{\mu} = 1.0$ N/m^3 are the nominal values for the first and third order stiffness, respectively. Response displacement measurements are taken for 32 sec with a sampling interval of 0.1 sec.

The updated PDF is obtained using a Bayesian probabilistic approach presented in Yuen and Beck (2001). A summary of this approach is given as follows:

Assume response data $\mathcal{D}$ is available at N discrete time instants, i.e., $\mathcal{D} = \{x(n\Delta t), n = 0,1,\ldots,N-1\}$. The estimator of the spectral density of x, $S_{x,N}(\omega_k), k = 0,1,\ldots,N_1$, is given by (Katafygiotis and Yuen, 2001):

$$S_{x,N}(\omega_k) = \frac{\Delta t}{2\pi N}\left|\sum_{n=0}^{N-1}\exp(-i\omega_k n\Delta t)x(n\Delta t)\right|^2 \tag{14}$$

where N_1 is equal to the integer part of $N/2$. For a given set of model parameters θ, the expected value of the spectral density estimator $E[S_{x,N}(\omega_k|\theta)]$ is given by:

$$E[S_{x,N}(\omega_k|\theta)] = \frac{\Delta t}{2\pi N}\sum_{n=0}^{N-1}\gamma_n R_x(n\Delta t|\theta)\cos(n\omega_k\Delta t) \tag{15}$$

where $\gamma_0 = N$ and $\gamma_n = 2(N-n), n \geq 1$; R_x is the autocorrelation function for the response x. In the case of a Duffing oscillator, it is given to good accuracy by the solution of the following second order ordinary differential equation with constant coefficients:

$$mR_x''(\tau|\theta) + cR_x'(\tau|\theta) + (\bar{k}\theta_1 + 3\sigma_x^2(\theta)\bar{\mu}\theta_2)R_x(\tau|\theta) = 0 \tag{16}$$

with initial conditions: $R_x(0|\theta) = \sigma_x^2(\theta)$ and $R_x'(0|\theta) = 0$.

The updated PDF for the stiffness parameters given data $\mathcal{D}$ is given by (Yuen and Beck, 2001):

$$p_{\mathcal{D}}(\theta) = cp_0(\theta)\exp(-J(\theta)/\varepsilon^2) \tag{17}$$

where c is a normalizing constant, $p_0(\theta)$ is a prior PDF for the stiffness parameters and $J(\theta)$ is given by:

$$J(\theta) = \sum_{k=0}^{N_1}\left[ln(E[S_{x,N}(\omega_k|\theta)]) + \frac{S_{x,N}(\omega_k)}{E[S_{x,N}(\omega_k|\theta)]}\right] \tag{18}$$

The prior PDF p_0 for θ_1 and θ_2 is chosen to be an independent uniform distribution from 0 to 3. That is, $p(\theta_1,\theta_2) = \frac{1}{9}$, if $\theta_1,\theta_2 \in (0,3)$ and $p(\theta_1,\theta_2) = 0$, otherwise. Note that the truncation at $\theta_2 = 3$ is arbitrary and the "ridge" of high probability content would extend much further if a broader prior PDF was chosen.

The updated PDF and the corresponding updated robust reliability with decreasing prediction error levels $\varepsilon_i^2 = 1/2^{i-1}$ for successive simulation levels $i = 1,2,\ldots,7$ are investigated. The sequence of intermediate PDFs $\{p_i\}$ is constructed by successively substituting the sequence of values $\varepsilon_i^2 = 1/2^{i-1}$ into (17).

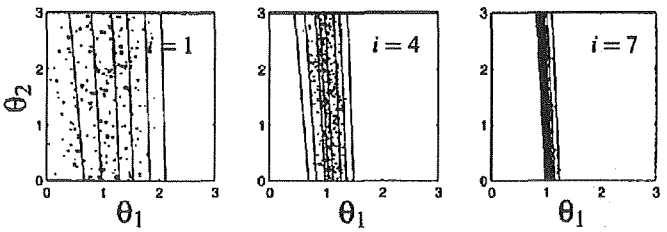

Figure 1. Markov chain samples

Figure 1 shows the Markov chain samples at simulation levels $i = 1,4,7$, corresponding to $\varepsilon^2 = 1, 1/8, 1/64$ in (17). For each simulation level, after the first 10 Markov chain samples are ignored, $N_s = 500$ Markov chain samples are simulated, which are shown with dots in Figure 1. Note that the Markov chain samples are not all distinct. To show the population of samples

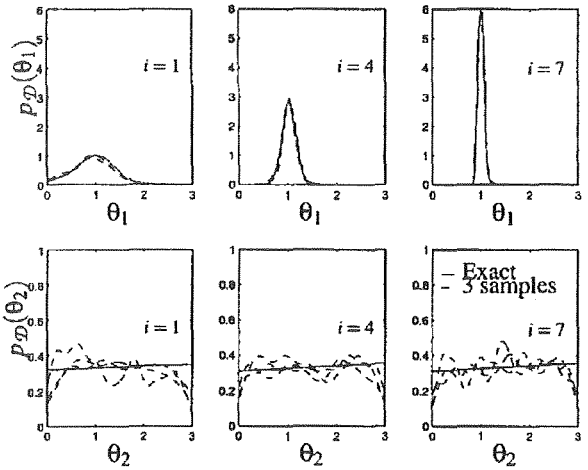

Figure 2. Prior and Updated Marginal PDFs for θ_1 and θ_2

consistently, the area of the dots are shown proportional to the number of samples at each location. The number of distinct samples are 335, 361 and 331 for simulation levels $i = 1,4$ and 7, respectively. The contour lines for each updated PDF are plotted at levels $0.01, 0.1, 0.5$ and 0.9 relative to the peak value of the updated PDF in each figure. It can be seen that the area enclosed by the outermost contour (0.01), and hence the region of probability concentration of the updated PDF, diminishes as ε decreases. The samples populate well the important region of the updated PDFs. More importantly, the density of the population is consistent with the exact updated PDF.

The marginal kernel PDFs for θ_1 and θ_2 constructed from the Markov chain samples are shown in Figure 2. The results from three independent simulation runs are shown with dashed lines. Note that the estimated marginal PDFs can be readily obtained from the joint kernel PDF without numerical integration. For comparison purposes, the exact marginal PDFs obtained by numerical integration are also plotted with solid lines in Figure 2. The results can be considered acceptable if fine detail is not required. The results for CDFs, although not shown here, exhibit a better match with the exact results, as the spurious "noise" in the PDFs are filtered out by integration in the CDFs. The errors in the marginal PDFs come from two sources. The first source is that the finite number of Markov chain samples are not distributed exactly as the target PDF, which introduces bias in the estimates. The second is due to the approximate nature of the kernel PDF using a finite number of samples. The dependent nature of the Markov chain samples in general does not bias the kernel PDF, but it slows down the convergence of the kernel PDF compared to the case when the samples are independent.

Note that the marginal PDF for θ_2 is almost a uniform distribution, implying that the set of data contains very little information for identifying the third order stiffness of the oscillator, i.e., θ_2 is essentially unidentifiable. The reason is that the response is of mild nonlinearity so it contains very little information about the nonlinear term. One way to reduce the uncertainty is to obtain another set of measurement with a higher level of S_o. Note also from Figure 1 that θ_1 and θ_2 are highly correlated, i.e., given one set of dynamic data, there is a range of θ_1 and θ_2 values which give similar values for the equivalent linear stiffness and hence similar values for the posterior PDF.

The reliability of the oscillator when it is subjected to an uncertain excitation is updated. The excitation is modeled by stationary Gaussian white noise with spectral intensity $S_o = 0.04m^2\,\mathrm{sec}^{-3}$. The updated robust failure probability that the stationary response of the oscillator will exceed the threshold level $b = 1.0m$ within a duration of $T = 20\,\mathrm{sec}$ is studied:

$$P(F|\mathcal{D}) = \int P_F(\theta) p_{\mathcal{D}}(\theta) d\theta \tag{19}$$

where $P_F(\theta) = P(F|\theta)$ is the failure probability for given model parameters θ. Assuming that the out-crossing events follow a Poisson process, $P_F(\theta)$ is approximated by $P_F(\theta) = 1 - \exp[-2\nu(\theta)T]$ (Soong and Grigoriu, 1993), where $\nu(\theta)$ is the up-crossing rate:

$$\nu(\theta) = \frac{\sigma_{\dot{x}}(\theta)}{2\pi\sigma_x(\theta)} \exp\left[-\frac{b^2}{2\sigma_x^2(\theta)}\right] \tag{20}$$

In (20), $\sigma_x(\theta)$ and $\sigma_{\dot{x}}(\theta)$ are the stationary standard deviations of the displacement and velocity response, x and $\dot{x}$, respectively, which can be approximated using equivalent linearization (Lutes and Sarkani, 1997):

$$\begin{aligned} \sigma_x(\theta) &= \sqrt{\frac{\bar{k}\theta_1}{6\bar{\mu}\theta_2}\left(\sqrt{1+\frac{12\bar{\mu}\theta_2\pi S_o}{c\bar{k}^2\theta_1^2}}-1\right)} \\ \sigma_{\dot{x}}(\theta) &= \sqrt{\frac{\pi S_o}{mc}} \end{aligned} \tag{21}$$

The estimates $\tilde{P}(F|\mathcal{D})$ for the robust failure probabilities computed based on (10) using the Markov chain samples are shown in Figure 3(a). Three sample estimates, corresponding to the same simulation runs in Figure 2, are shown with circles. The exact results obtained by numerical integration are shown with solid lines. In Figure 3(a), the simulation level $i = 0$ refers to the case when no data is available and the robust failure probabilities are computed based on the prior PDF p_0 only, that is, by (2). To investigate the bias of the simulation results, the average of the estimates over 50 independent simulation runs are computed and shown with dashed lines in Figure 3(a). The 50-average results (dashed lines) almost overlap with the exact results (solid

lines), showing that the bias from the initial transient stage of the simulated Markov chains is negligible.

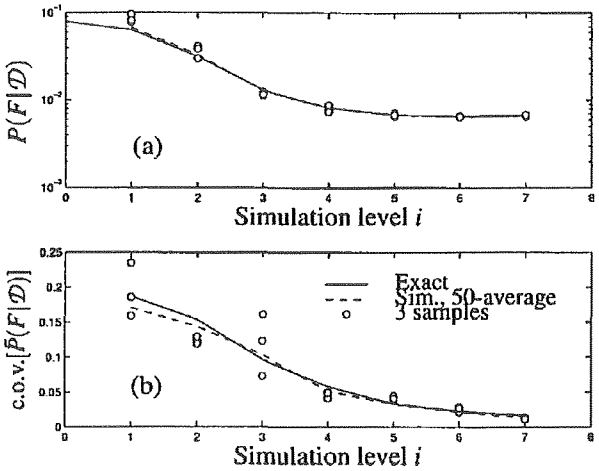

Figure 3. Updated Failure Probabilities $P(F|\mathcal{D})$

The c.o.v. of the robust failure probability estimates computed from the 50 independent runs are computed and shown in Figure 3(b) with solid lines. The c.o.v. estimates based on (11), which does not include the variability of the proposal PDF, are also computed and the results for the three simulation runs in Figure 2 are shown with circles. The average of the c.o.v. estimates based on (11) over 50 simulation runs are shown with dashed lines. From Figure 3(b), it can be seen that the actual c.o.v. (solid lines) are quite close to those predicted by (11) (dashed lines), showing that the latter is useful for assessing the variability of the robust reliability estimates.

CONCLUSIONS

The proposed simulation approach for Bayesian system identification gradually adapts to the region of probability concentration of the updated PDF for the model parameters through successive simulation levels. Its success relies on a good choice of the intermediate PDFs and their approximation by kernel PDFs. Future research may focus on developing better intermediate PDFs as well as enhancing the performance of the kernel PDF approximations.

ACKNOWLEDGMENT

This paper is based upon work partly supported by the Pacific Earthquake Engineering Research Center under National Science Foundation Cooperative Agreement No. CMS-9701568. This support is gratefully acknowledged.

REFERENCES

Au, S. K. and Beck, J. L., *A New Adaptive Importance Sampling Scheme*, Structural Safety, 21, pp. 135-158, 1999.

Au, S. K. and Beck, J. L., *Subset Simulation: A New Approach to Calculating Small Failure Probabilities*, Proc International Conf. on Monte Carlo Simulation, Monte Carlo, 2000.

Beck, J. L. and Katafygiotis, L. S., *Updating models and Their Uncertainties – Bayesian Statistical Framework*, ASCE Journal of Engineering Mechanics, 124(4), pp. 455-461, 1998.

Doob, J. L., *Stochastic Processes*, John Wiley & Sons, Inc., New York, N.Y., 1953.

Fishman, G. S., *Monte Carlo: Concepts, Algorithms, and Applications*, Springer Series in Operations Research, Springer-Verlag, New York, 1996.

Hastings, W. K., *Monte Carlo Sampling Methods Using Markov Chains and Their Applications*, Biometrika, 57, pp. 97-109, 1970.

Katafygiotis, L. S. and Beck, J. L., *Updating Models and Their Uncertainties – Model Identifiability*, ASCE Journal of Engineering Mechanics, 124(4), pp. 463-467, 1998.

Katafygiotis, L. S., Papadimitriou, C. and Lam, H. F., *A Probabilistic Approach to Structural Model Updating*, Soil Dynamics and Earthquake Engineering, 17(7-8), pp. 495-507, 1998.

Katafygiotis, L. S. and Yuen, K. V., *Bayesian Spectral Density Approach for Modal Updating Using Ambient Data*, Earthquake Engineering and Structural Dynamics, 30(8), pp. 1103-1123, 2001.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N. and Teller, A. H., *Equations of State Calculations by Fast Computing Machines*, Journal of Chemical Physics, 21(6), pp. 1087-1092, 1953.

Papadimitriou, C., Beck, J. L. and Katafygiotis, L. S., *Updating Robust Reliability Using Structural Test Data*, Probabilistic Engineering Mechanics, 16(2), pp.103-113, 2001.

Lutes, L. D. and Sarkani, S., *Stochastic Analysis of Structural and Mechanical Vibrations*, Prentice Hall, 1997.

Rubinstein, R. Y., *Simulation and the Monte-Carlo Method*, John Wiley & Sons, Inc., New York, N.Y., 1981.

Silverman, B. W., *Density Estimators*, Chapman and Hall, New York, 1986.

Soong, T. T. and Grigoriu, M., *Random Vibration of Mechanical and Structural Systems*, Prentice Hall, Inc., 1993.

Törn, A. and Zilinskas, A., *Global Optimization*, Lecture Notes in Computer Science, vol.350, Springer-Verlag, 1989.

Yang, C. M. and Beck, J. L., *Generalized Trajectory Methods for Finding Multiple Extrema and Roots of Functions*, Journal of Optimization Theory and Applications, 97(4), pp. 211-227, 1998.

Yuen, K. V. and Beck, J. L., *Probabilistic System Identification for Nonlinear Systems With Uncertain Input*, ICOSSAR'01, Newport Beach, California, USA, 2001.

Proceedings of the

2001 ASME DESIGN ENGINEERING TECHNICAL CONFERENCES AND COMPUTERS AND INFORMATION IN ENGINEERING CONFERENCE

VOLUME 6

18TH BIENNIAL CONFERENCE ON MECHANICAL VIBRATION AND NOISE

PART A

Organized by
D. T. MOOK
VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY

B. BALACHANDRAN
UNIVERSITY OF MARYLAND

THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS
Three Park Avenue / New York, N.Y. 10016