Adaptive variant of the BUS approach to Bayesian updating

W. Betz, I. Papaioannou, D. Straub
Engineering Risk Analysis Group, Technische Universität München, Arcisstr. 21, 80290 München,
email: wolfgang.betz@tum.de, iason.papaioannou@tum.de, straub@tum.de

ABSTRACT: The BUS approach is a recently proposed method [1] for Bayesian updating of models with structural reliability techniques. Especially for high-dimensional problems, the combination of BUS with subset simulation [2] is shown to be efficient in drawing samples from the posterior distribution. The BUS approach can be considered an extension of rejection sampling, where a standard uniform random variable is added to the space of random variables. Each generated sample from this extended random variable space is accepted if the sample of the uniform random variable is smaller than the likelihood function scaled by a constant c. The constant c has to be selected such that it's reciprocal is not smaller than the maximum of the likelihood function. For $1/c$ considerably larger than the maximum of the likelihood function, the efficiency of the approach decreases. However, in many cases the maximum of the likelihood function is not known in advance. In this contribution, we propose a technique for adaptively selecting the parameter c. This causes the rejection/acceptance criterion to change throughout the simulation. The proposed approach is compared to the TMCMC method proposed by Ching and Chen in [3] by means of a numerical example. We show that the proposed approach maintains the efficiency of BUS for problems with many random variables.

KEY WORDS: Bayesian updating; Structural reliability; TMCMC.

1 INTRODUCTION

Bayesian analysis provides a consistent framework to reduce uncertainties in existing models through new information. The uncertainties in the model are expressed by means of input parameters that are regarded as uncertain. The new information is then used to update our prior belief about the parameters of the model to a resulting posterior belief. If the posterior model cannot be derived analytically, samples of the posterior have to be generated numerically.

Markov chain Monte Carlo (MCMC) methods constitute a popular class of methods to sample from the posterior distribution [4,5]. One problem with MCMC methods is that the samples used after an initial burn-in phase may not have reached the stationary distribution of the Markov chain [6]. Another problem is that the MCMC algorithm can usually not be applied efficiently for problems with many uncertain parameters. Some MCMC algorithms [7,8] can cope with such problems, they require however additional evaluations of the likelihood function or its gradient for each sample. The burn-in problem is efficiently tackled for problems with just a few uncertain parameters by the TMCMC method [3], which belongs to the class of sequential particle filter methods [9].

An approach that does not suffer a burn-in problem and can potentially cope with high-dimensional problems (i.e., problems with many uncertain parameters) was recently proposed by Straub and Papaioannou [1]: Bayesian updating with structural reliability methods, termed BUS. The downside of BUS is that prior to the analysis a constant c has to be selected, where c^{-1} should not be smaller than the maximum value that the likelihood function can take. However, in many cases the maximum of the likelihood function is not known in advance. If c^{-1} is chosen considerably larger than the maximum of the likelihood function, the efficiency of the approach decreases.

In this contribution, we propose a technique for adaptively selecting the constant c. The adaptive variant of the BUS approach is compared to the TMCMC method by means of three numerical examples.

2 BAYESIAN MODEL UPDATING

Let $\mathbf{d}$ denote the new information that becomes available in the form of measurements or observations. Furthermore, let $\mathbf{X}$ be the vector of model parameters that are considered uncertain in the analysis. The likelihood of observing $\mathbf{d}$ given the parameter set $\mathbf{x}$ is expressed as $L(\mathbf{x}|\mathbf{d})$. The learning process is formalized through Bayes' theorem as:

$$f(\mathbf{x}|\mathbf{d}) = c_{\mathrm{E}}^{-1}\, L(\mathbf{x}|\mathbf{d})\, f(\mathbf{x}) \tag{1}$$

Where probability density function (PDF) $f(\mathbf{x})$ represents our prior belief of the distribution of $\mathbf{X}$, $f(\mathbf{x}|\mathbf{d})$ is the resulting posterior distribution, and c_{E} acts as a scaling constant:

$$c_{\mathrm{E}} = \int_{\mathbf{X}} L(\mathbf{x}|\mathbf{d})\, f(\mathbf{x})\, \mathrm{d}\mathbf{x} \tag{2}$$

If $\int_{\mathbf{d}} L(\mathbf{x}|\mathbf{d})\, \mathrm{d}\mathbf{d} = 1$, the constant c_{E} is equivalent to the evidence of the assumed model class, i.e., $c_{\mathrm{E}} = f(\mathbf{d})$ [10]. The evidence is a measure for the likelihood of a model class. In case of multiple model classes, knowledge of the evidence allows us to evaluate the posterior plausibilities of the individual model classes. The plausibilities are required for Bayesian model class selection and Bayesian model averaging [11,12]. Therefore, it is of advantage if a method for Bayesian updating does not only return samples from the posterior distribution, but returns an estimate of the evidence as well.

2.1 Bayesian Updating with structural reliability methods

In the BUS approach [1], the updating problem is interpreted as a structural reliability problem. As a consequence, methods that were originally developed for structural reliability analysis can be applied to represent the posterior distribution.

Let P be a standard uniform random variable defined on the interval [0,1]. Combining P with the uncertain parameter vector $\mathbf{X}$ to be learned gives the augmented outcome space $[\mathbf{X}, P]$. Note that for the joint probability density function of $[\mathbf{X}, P]$ the following relation holds: $f(\mathbf{x}, p) = f(\mathbf{x}) \cdot f(p) = f(\mathbf{x})$, since $f(p) = 1$; therefore, $f(\mathbf{x}, p)$ and $f(\mathbf{x})$ will be used interchangeably. Furthermore, let the domain Ω be defined as:

$$\Omega = \{p \leq c \cdot L(\mathbf{x}|\mathbf{d})\} \tag{3}$$

where c is a positive constant such that $c \cdot L(\mathbf{x}|\mathbf{d}) \leq 1$ for all $\mathbf{x}$. Straub and Papaioannou [1] showed that

$$\int_{p\in\Omega} f(\mathbf{x})\, \mathrm{d}p = f(\mathbf{x}) \cdot c \cdot L(\mathbf{x}|\mathbf{d}) \tag{4}$$

and, consequently, Eq. (1) can be written as:

$$f(\mathbf{x}|\mathbf{d}) = \frac{\int_{p\in\Omega} f(\mathbf{x})\, \mathrm{d}p}{\int_{[\mathbf{x},p]\in\Omega} f(\mathbf{x})\, \mathrm{d}p\, \mathrm{d}\mathbf{x}} \tag{5}$$

The denominator in Eq. (5) constitutes a structural reliability problem with limit-state function [13]:

$$g(\mathbf{x}, p) = p - c \cdot L(\mathbf{x}|\mathbf{d}) \tag{6}$$

The limit-state function is defined such that it is $g(\mathbf{x}, p) \leq 0$ if outcome $[\mathbf{x}, p]$ is inside the domain Ω; and $g(\mathbf{x}, p) > 0$ if it is outside Ω. If the reliability problem is solved via a sampling based approach, the samples generated from the prior $f(\mathbf{x}, p)$ that fall into the domain Ω will be distributed according to the posterior $f(\mathbf{x}|\mathbf{d})$ [1]. Let p_a denote the probability that a sample from $f(\mathbf{x}, p)$ will fall into Ω; i.e., $p_\mathrm{a} = \Pr[\, g(\mathbf{X}, P) \leq 0]$. The scaling constant c_E, defined in Eq. (2), is linked to p_a through:

$$c_\mathrm{E} = p_\mathrm{a}/c \tag{7}$$

The simplest application of this idea is the rejection sampling algorithm [14,1] for drawing K samples from the posterior distribution:

1. Set counter $k = 1$.
2. Propose a sample $[\mathbf{x}_{(k)}, p_{(k)}]$:
   a. Draw $\mathbf{x}_{(k)}$ from the prior distribution $f(\mathbf{x})$.
   b. Draw $p_{(k)}$ from the uniform distribution defined on [0,1].
3. If $g(\mathbf{x}_{(k)}, p_{(k)}) \leq 0$:
   a. Accept the proposed sample $[\mathbf{x}_{(k)}, p_{(k)}]$.
   b. Increase the counter $k = k + 1$.
4. Go to step 2 as long as $k < K$.

The above algorithm is equivalent to applying Monte Carlo simulation for solving a structural reliability problem with limit-state function $g(\mathbf{x}, p)$. The simulation continues until K failures are observed. On average, the procedure needs to be repeated K/p_a times, where p_a is the probability of accepting a proposed sample. The probability p_a can be estimated as the number of accepted samples divided by the total number of samples. If the posterior distribution does not match the prior distribution well, p_a becomes small and renders the algorithm inefficient. However, as was pointed out in [1], other structural reliability methods can be used instead of the simple rejection sampling algorithm.

2.2 BUS with subset simulation

In this paper, subset simulation (SuS) is applied to perform the reliability analysis. SuS was proposed by Au and Beck in [2] and is an adaptive Monte Carlo method that is efficient for estimating small probabilities in high dimensional problems. The advantage of BUS with subset simulation is that the same MCMC methods that render subset simulation efficient for high-dimensional problems can be applied. With SuS, the domain Ω is expressed as the intersection of M intermediate nested domains Z_i, where $Z_0 \subset Z_1 \subset \cdots \subset Z_M = \Omega$. The domains Z_i are defined as the sets $\{g(\mathbf{x}, p) \leq b_i\}$, where $b_0 = \infty > b_1 > b_2 > \cdots > b_M = 0$ holds. Samples of $Z_i,\ i \in \{0, \ldots, M\}$ are denoted $[\mathbf{x}_{(i,k)}, p_{(i,k)}],\ k \in \{1, \ldots, N_\mathrm{s}\}$.

SuS starts with drawing N_s random samples $[\mathbf{x}_{(0,k)}, p_{(0,k)}]$ from the prior distribution $f(\mathbf{x})$. Note that the samples are realizations from $[\mathbf{X}, P]$ conditional on the domain Z_0. The scalar b_{i+1} is picked as the p_t-percentile of the set $\{g(\mathbf{x}_{(i,k)}, p_{(i,k)})\}_{k=1}^{N_s}$, where p_t is usually chosen equal to 10% in each step. If the p_t-percentile is negative, b_{i+1} is set to zero and $p_{\mathrm{t},i+1}$ is set to the fraction of negative values in the list, otherwise $p_{\mathrm{t},i+1} = p_\mathrm{t}$. Consequently, $p_{\mathrm{t},i+1} \cdot N_\mathrm{s}$ realizations conditional on Z_i are also in the domain Z_{i+1}. In order to obtain N_s realizations in the domain Z_{i+1}, Markov chain Monte Carlo (MCMC) methods are used to generate $(1 - p_{\mathrm{t},i+1}) \cdot N_\mathrm{s}$ additional samples of Z_{i+1}, where the $p_{\mathrm{t},i+1} \cdot N_\mathrm{s}$ original realizations of Z_{i+1} are used as seed values for the Markov chains. The described iterative process is continued until $b_i = b_M = 0$. Note that samples from the domain Z_M follow the posterior distribution. Let K denote the number of samples to be generated from the posterior distribution. If $K > N_\mathrm{s}$, the MCMC sampling at the Mth iteration step is continued until K (and not only N_s) realizations from the domain Z_M are available. The acceptance probability can be estimated as $p_\mathrm{a} = \prod_{i=1}^{M} p_{\mathrm{t},i}$. For a more detailed description of the sampling procedure, the reader is referred to Straub and Papaioannou [1].

2.3 Adaptive variant of the BUS approach

An advantage of the BUS approach is that several methods from structural reliability can be readily applied to perform the Bayesian analysis. If it is combined with SuS as described above, it inherits the advantage of SuS that it is applicable in high dimensions. The only difference compared to solving a structural reliability problem is that the coordinates of the samples that fall into the failure domain have to be stored, because they are samples of the posterior distribution.

However, the required choice of the constant c might be a problem, because the true maximum that the likelihood can take, denoted $\hat{L}$, is not always known in advance. On the one hand, choosing the constant c too conservative, i.e., $c < \hat{L}^{-1}$, will decrease the efficiency of the method, since the acceptance probability p_a that has to be estimated decreases with a decreasing c: $p_\mathrm{a} = c_\mathrm{E} \cdot c$. On the other hand, if c is selected too large, the obtained samples might not follow the

posterior distribution. In the remainder of this section, we propose a procedure to learn the parameter c adaptively. The proposed procedure is based on the combination of BUS with subset simulation (see section 2.2). However, the implementation of subset simulation has to be modified, compared to its traditional application for reliability analysis.

Just as for BUS with subset simulation, the domain Ω is expressed as the intersection of M intermediate nested domains Z_i, where $Z_0 \subset Z_1 \subset \cdots \subset Z_M = \Omega$. Let $g_i(\mathbf{x}, p)$ be intermediate limit-state functions defined as:

$$g_i(\mathbf{x}, p) = p - L(\mathbf{x}|\mathbf{d}) \cdot c_i \tag{8}$$

The constants $c_i \in \mathbb{R},\ i \in \{0,1,\dots,M\}$ are set equal to the reciprocal of the maximum value of the likelihood of the samples obtained in all previous steps. This ensures that $c_i \leq c_{i-1}$, i.e. $c_0 \geq c_1 \geq \cdots \geq c_M = c$. We now distinguish two kinds of intermediate domains, Z_i and Z_i^*. The former are nested, whereas the latter are not necessarily. The domains Z_i^* are defined through $g_i(\mathbf{x}, p)$ as $Z_i^* = \{g_i(\mathbf{x}, p) \leq a_i\}$, where $a_i \in [0, \infty]$, with $a_0 = \infty$ and $a_M = 0$. The domains Z_i are defined as $Z_i = Z_i^* \cap Z_{i-1}$, which ensures that they are nested. The constants a_i defining Z_i^* are chosen such that a fraction p_t of the samples in Z_{i-1} fall also in the domain Z_i^*. For $c_i = c_{i-1}$, the condition $a_i \leq a_{i-1}$ holds. However, if $c_i > c_{i-1}$, a_i might be larger than a_{i-1}. In this case, the domain Z_i^* might not be entirely contained in Z_{i-1}^*. In the last step, with $a_i = 0$, it is $Z_i = Z_i^*$, since $Z_i \subseteq Z_{i-1}$, for any $c_i \geq c_{i-1}$. Note that for c_0 any value can be selected, but in case $c_0 \leq 1/\hat{L}$ it is $c_0 = c_1 = \ldots = c_M = c$, and the procedure presented in the following will reduce to the one described in section 2.2. If the value of $\hat{L}$ is not known, it is suggested to set $c_0 = \infty$.

In the following, samples of Z_i are denoted $\left[\mathbf{x}_{(i,k)}, p_{(i,k)}\right],\ k \in \{1,\dots,N_\mathrm{s}\}$. The procedure starts with drawing $N_0 = N_\mathrm{s}$ random samples $\left[\mathbf{x}_{(0,k)}, p_{(0,k)}\right]$ from the prior distribution $f(\mathbf{x})$. These samples are realizations conditional on the domain Z_0. The constant c_{i+1} is picked as $c_{i+1} = \min\left(c_i, 1/\max\left(\left\{L(\mathbf{x}_{(j,k)}|\mathbf{d})\right\}_{k=1}^{N_i}\right)\right)$, where $L(\mathbf{x}_{(j,k)}|\mathbf{d})$ is the likelihood value that is associated with sample $\mathbf{x}_{(i,k)}$. The scalar a_{i+1} is picked as the p_t-percentile of the set $\left\{g_{i+1}\left(\mathbf{x}_{(i,k)}, p_{(i,k)}\right)\right\}_{k=1}^{N_i}$, where p_t is usually chosen equal to 10% in each step. If a_{i+1} is negative, a_{i+1} is set to zero and $p_{\mathrm{t},i+1}$ is set to the fraction of negative values in the list, otherwise $p_{\mathrm{t},i+1} = p_\mathrm{t}$. All samples for which $g_{i+1}\left(\mathbf{x}_{(i,k)}, p_{(i,k)}\right) \leq a_{i+1}$ holds are also realizations from the domain Z_{i+1}. These samples are used as seeds in the MCMC method to generate additional samples, so that in total N_{i+1} realizations from Z_{i+1} are available, where $N_{i+1} = K$ if $\min\{a_j\}_{j=0}^{i+1} = 0$ and $N_{i+1} = N_\mathrm{s}$ otherwise. The described iterative process is continued until $a_{i+1} = 0 \wedge c_{i+1} = c_i$. The c_E from Eq. (2) can be estimated as $c_\mathrm{E} = (1/c_M) \cdot \prod_{i=1}^{M} p_{\mathrm{t},i}$.

Note that $c = c_M$ will be larger than $1/\hat{L}$, because c is equal to the reciprocal of the largest likelihood value that was observed in the simulation. However, we conjecture that the samples $\left\{\mathbf{x}_{(M,k)}\right\}_{k=1}^{K}$ will follow the posterior distribution for most practical situations. The validity and performance of the method is demonstrated by numerical examples.

2.4 TMCMC

In this section, we briefly summarize the transitional Markov chain Monte Carlo (TMCMC) method [3], since we will use it as a reference to compare the performance of the adaptive BUS in the numerical examples. TMCMC belongs to the class of sequential particle filter methods [9]. The idea is to start with samples from the prior distribution, and in subsequent steps, gradually transform the shape of the sampling distribution such that it approaches the posterior distribution. For this purpose, Eq. (1) is transformed as:

$$f_j(\mathbf{x}) \propto f(\mathbf{x}) \cdot L(\mathbf{x}|\mathbf{d})^{q_j} \tag{9}$$

where $j = 0, \dots, m$ denotes the stage number, and the $q_j \in [0,1]$ are chosen such that $q_0 = 0 < q_1 < \cdots < q_m = 1$. Consequently, for $j = 0$, $f_0(\mathbf{x})$ is equal to the prior distribution $f(\mathbf{x})$; and for $j = m$, $f_m(\mathbf{x})$ matches the posterior distribution $f(\mathbf{x}|\mathbf{d})$.

Ching an Chen [3] proposed to select q_{i+1} such that the coefficient of variation of $L(\mathbf{x}|\mathbf{d})^{q_{j+1}-q_j}$ becomes v_t, with $v_\mathrm{t} = 100\%$. Let $\mathbf{x}_{(j,k)}$ with $k \in N_\mathrm{s}$ denote the kth sample from distribution $f_j(\mathbf{x})$ at stage j, where N_s is the number of samples generated at each stage. Furthermore, let $c_{\mathrm{v},(j)}(q)$ with $q \in]q_j, 1]$ denote the coefficient of variation of the set $\left\{L(\mathbf{x}_{(j,k)}|\mathbf{d})^{q-q_j}\right\}_{k=1}^{N_\mathrm{s}}$ at stage j, where $L(\mathbf{x}_{(j,k)}|\mathbf{d})$ is the likelihood value that is associated with sample $\mathbf{x}_{(j,k)}$. Based on the set of samples $\left\{\mathbf{x}_{(j,k)}\right\}_{k=1}^{N_\mathrm{s}}$, the value of q_{i+1} can be determined as:

$$q_{i+1} = \arg\min_q \left(c_{\mathrm{v},(j)}(q) - q_t\right) \tag{10}$$

In the initial stage (i.e., for $j = 0$), N_s samples $\mathbf{x}_{(0,k)}$ are drawn from the prior distribution. Samples $\mathbf{x}_{(j+1,k)}$ of distribution $f_{j+1}(\mathbf{x})$ can be obtained from the set of samples $\left\{\mathbf{x}_{(j,k)}\right\}_{k=1}^{N_\mathrm{s}}$ by the following procedure:

1. For each $k = 1, \dots, N_\mathrm{s}$ compute a weighting coefficient $w_{(j,k)}$ as:

$$w_{(j,k)} = \left(L(\mathbf{x}_{(j,k)}|\mathbf{d})\right)^{q-q_j} \tag{11}$$

2. Compute the mean of the weighting coefficients:

$$S_j = \frac{1}{N_s} \sum_{k=1}^{N_\mathrm{s}} w_{(j,k)} \tag{12}$$

3. Perform a resampling of $\left\{\mathbf{x}_{(j,k)}\right\}_{k=1}^{N_\mathrm{s}}$ according to the weights $w_{(j,k)}$, i.e. draw N_S samples from the discrete distribution in which each value $\mathbf{x}_{(j,k)}$ has probability $w_{(j,k)}$. The samples obtained through the resampling step are denoted $\mathbf{x}'_{(j+1,k)}$ and follow distribution $f_{j+1}(\mathbf{x})$.
4. For each sample $\mathbf{x}'_{(j+1,k)},\ k = 1, \dots, N_\mathrm{s}$ perform a single MCMC step, where the sample $\mathbf{x}'_{(j+1,k)}$ acts as seed value of the chain. The obtained sample is denoted $\mathbf{x}_{(j+1,k)}$.

The procedure is repeated until $q_j = q_m = 1$. Let K bet the total number of posterior samples that we want to obtain. If $K > N_\mathrm{s}$ and $q_m = 1$, the MCMC sampling is continued until

K realizations are avialable. The evidence introduced in Eq. (2) can be estimated as $c_{\mathrm{E}} = \prod_{i=0}^{m-1} S_j$.

Note that steps (3) and (4) are taken from the algorithm proposed in [9], and differ from the original TMCMC algorithm proposed in [3], wherein steps (3) and (4) are performed as: *1.* initialize the samples $\mathbf{x}'_{(j+1,k)}$, $k = 1, \dots, N_{\mathrm{s}}$ as $\mathbf{x}'_{(j+1,k)} = \mathbf{x}_{(j,k)}$ and set $m = 1$, *2.* select an index i from $1, \dots, N_{\mathrm{s}}$ with probability proportional to the weights $w_{(j,k)}$, *3.* perform a single MCMC step with the sample $\mathbf{x}'_{(j+1,i)}$ as seed and overwrite $\mathbf{x}'_{(j+1,i)}$ with the obtained sample, *4.* set $\mathbf{x}_{(j+1,m)} = \mathbf{x}'_{(j+1,i)}$, *5.* set $m = m + 1$ and continue with step *2* as long as $m \leq N_{\mathrm{s}}$. We use the modified TMCMC procedure, because it reduces the bias and the variance in the obtained posterior estimate considerably.

3 NUMERICAL EXAMPLES

In the following we will compare the adaptive variant of the BUS approach (referred to as aBUS) with BUS with subset simulation, and with the variant of the TMCMC method described in the previous section by means of three examples. The number of posterior samples generated by each method is 1000. The number of samples in the intermediate levels is also set to 1000. For both BUS and aBUS, the threshold probability p_{t} is set to 10%. The constant c in BUS is set equal to $1/\hat{L}$, and c_0 in aBUS is ∞. The target coefficient of variation in TMCMC is $v_{\mathrm{t}} = 100\%$.

The component-wise Metropolis-Hastings algorithm was applied as MCMC method in BUS and aBUS. The standard deviation of the Gaussian proposal distribution was modified adaptively such that the acceptance rate of the individual chains is close to 0.44 (compare [15]). For TMCMC the Metropolis-Hastings algorithm is applied. As proposal distribution the multivariate Gaussian distribution is used, where the covariance matrix is estimated from the set of current samples [3]. Ching and Chen [3] suggest to scale the covariance matrix of the proposal distribution with a constant factor of 0.2^2. Contrary to that, in the following studies we scale the covariance matrix of the proposal distribution such that the acceptance rate of the MCMC chains is close to 0.44, because it reduces the bias and the variance in the obtained posterior estimate considerably.

3.1 Example 1 – unimodal distribution

Let $\mathbf{X}$ be a d-dimensional random vector. The components $\{X_i : i = 0, \dots, d\}$ of $\mathbf{X}$ are independent and have identical distributions. The prior distribution of $\mathbf{X}$ is $f(\mathbf{x}) = \prod_{i=1}^{d} \varphi(x_i)$, where φ is the probability density function (PDF) of the univariate standard normal distribution. The likelihood of the problem is defined as $L(\mathbf{x}|\mathbf{d}) = \prod_{i=1}^{d} \frac{1}{\sigma_l} \varphi\left(\frac{x_i - \mu_l(d)}{\sigma_l}\right)$, where $\sigma_l = 0.2$ and $\mu_l(d)$ is selected such that the evidence $c_{\mathrm{E}} = 10^{-9}$ is independent of the dimension d. The performance of the different methods is investigated for increasing d. Since both the prior and the likelihood are Gaussian, the problem at hand has an analytical solution. The mean of the likelihood that typically represents the measured quantity can be computed as:

$$\mu_l(d) = \sqrt{-2(1+\sigma_l{}^2) \cdot \ln\left[{c_{\mathrm{E}}}^{1/d} \cdot \sqrt{2\pi \cdot \sqrt{1+\sigma_l{}^2}}\right]}$$

The mean and standard deviation of a posterior x_i is: $\mu''_X(d) = \frac{\mu_l(d)}{[\sigma_l{}^2(1+1/\sigma_l{}^2)]}$, $\sigma''_X(d) = \sqrt{\frac{1}{1+1/\sigma_l{}^2}}$. The maximum value the likelihood can take is $\hat{L}(d) = [\varphi({\sigma_l}^{-1})/\sigma_l]^d$.

In a first study, the estimated evidence $\hat{c}_{\mathrm{E}}$ is compared to the true evidence c_{E}. Let $\mu_{\hat{c}_{\mathrm{E}}}$ denote the mean of $\hat{c}_{\mathrm{E}}$. In Table 1 the ratio $\mu_{\hat{c}_{\mathrm{E}}}/c_{\mathrm{E}}$ is listed, where the value in brackets is the coefficient of variation (C.o.V.) of the estimate, denoted $\delta_{\hat{c}_{\mathrm{E}}}$. The quantities $\mu_{\hat{c}_{\mathrm{E}}}$ and $\delta_{\hat{c}_{\mathrm{E}}}$ were estimated by a 1000 runs of the updating problem. Looking at the results listed in Table 1, we see that the estimated evidence is biased for TMCMC, BUS and aBUS. For all three methods, the bias for $d = 1$ is larger than the bias for $d = 2$. Independent of the dimension d, the bias of the TMCMC estimate is clearly larger than the one of BUS and aBUS. For $d \geq 2$, the bias of the TMCMC and BUS estimate increases with an increasing d, where the bias increases faster for TMCMC than for BUS. For BUS, the dependency of the bias on the dimension can be explained through the acceptance probability p_{a}: since the evidence is kept constant and $\hat{L}$ increases as d increases, the acceptance probability $p_{\mathrm{a}} = c_{\mathrm{E}} \cdot c$ decreases, where $c = 1/\hat{L}$ for BUS and this example. For smaller acceptance probabilities, more subset steps are required in BUS. This increases the correlation of the posterior samples and, thus, the bias in the estimate of the evidence [2]. We believe that the relatively strong bias in the TMCMC estimate for this example is due to the learning of the covariance structure of the intermediate MCMC proposal distributions from the corresponding intermediate samples. We observed that if a similar technique to obtain the proposal distribution is applied in the MCMC step of the BUS approach, the bias in the estimate of the evidence will increase considerably for this example. Comparing BUS and aBUS, BUS performs slightly better than aBUS for $d < 10$; whereas for $d > 10$ it is the other way round. The observation that aBUS performs slightly better than BUS for higher dimensions can be explained by comparing the acceptance probabilities of the BUS and the aBUS problem: The last column in Table 1 lists the mean of the largest observed likelihood value in aBUS divided by $\hat{L}$. As the dimension increases, this ratio decreases considerably and, thus, the acceptance probability $p_{\mathrm{a}} = c_{\mathrm{E}} \cdot c$ that has to be computed in aBUS is larger than the one in BUS – which explains the smaller bias in the aBUS estimate of the evidence.

In a second study, we look at the number of model runs required to draw 1000 samples from the posterior distribution. The mean number of model runs required to draw 1000 samples from the posterior is listed in Table 2. The values were obtained as the average of performing the updating problem 1000 times. For this example, the number of model runs required depends on the dimension d for the three investigated methods. For BUS and aBUS this is a direct consequence of the decrease in the acceptance probability for increasing d: As the acceptance probability decreases, the number of intermediate subset steps increases on average. For TMCMC, the mean of N_{eval} increases up to $d = 10$, and

decreases for $d > 7$. However, we believe that the observed decrease might, at least partially, be a side effect of the strong bias observed in Table 1. For $d \leq 10$, BUS and aBUS need a smaller number of model runs than TMCMC. aBUS performs slightly better than BUS, where the difference increases for increasing d. Again, this is a consequence of the acceptance probability that has to be computed.

Table 1. Statistical properties of the estimated evidence for increasing d (Example 1).

	$\frac{\mu_{\hat{c}_E}}{c_E}$ $(\delta_{\hat{c}_E})$			$\frac{E[1/c]}{\hat{L}}$ $(\frac{\sqrt{Var[1/c]}}{E[1/c]})$
d	TMCMC	BUS	aBUS	aBUS
1	3.06 (9.8)	0.81 (4.8)	2.91 (10)	1.0 (1‰)
2	0.69 (4.6)	1.05 (2.5)	1.74 (2.5)	0.99 (9‰)
3	0.35 (3.0)	1.10 (2.8)	1.68 (1.9)	0.96 (3%)
5	0.24 (7.0)	1.04 (1.7)	1.33 (1.5)	0.83 (0.11)
7	0.09 (8.1)	1.23 (1.8)	1.30 (1.5)	0.65 (0.21)
10	0.11 (8.1)	1.13 (1.1)	1.14 (0.9)	0.39 (0.40)
15	0.50 (8.0)	1.14 (0.9)	1.06 (0.9)	0.14 (0.74)
20	72.5 (9.3)	1.14 (0.9)	1.06 (1.0)	0.05 (1.13)

Table 2. Mean number of model runs required to obtain samples of the posterior with 1000 samples per intermediate level (Example 1).

d	TMCMC	BUS	aBUS
1	$11.9 \cdot 10^3$	$9.47 \cdot 10^3$	$8.98 \cdot 10^3$
2	$12.0 \cdot 10^3$	$10.0 \cdot 10^3$	$9.68 \cdot 10^3$
3	$12.7 \cdot 10^3$	$10.3 \cdot 10^3$	$10.1 \cdot 10^3$
5	$13.0 \cdot 10^3$	$10.8 \cdot 10^3$	$10.7 \cdot 10^3$
7	$13.4 \cdot 10^3$	$11.3 \cdot 10^3$	$11.1 \cdot 10^3$
10	$13.4 \cdot 10^3$	$12.1 \cdot 10^3$	$11.8 \cdot 10^3$
15	$12.8 \cdot 10^3$	$13.4 \cdot 10^3$	$12.7 \cdot 10^3$
20	$12.3 \cdot 10^3$	$14.7 \cdot 10^3$	$13.6 \cdot 10^3$

Next, the statistical properties of posterior samples of the random variable X_1 are compared with the corresponding analytical solution. Let $\Theta = \left\{x_1^{(1)}, x_1^{(2)}, \dots, x_1^{(1000)}\right\}$ denote a set of 1000 posterior samples of X_1. Furthermore, let the sample mean and sample variance of the samples in Θ be $\bar{x}$ and s^2, respectively. $\bar{x}$ and s^2 are themselves random variables whose theoretical mean is equivalent to the posterior mean and variance, i.e., $\mu_{\bar{x}} = E[\bar{x}] = \mu_X''$ and $\mu_{s^2} = E[s^2] = (\sigma_X'')^2$. If the samples in Θ were independent, the theoretical standard deviation of $\bar{x}$ and s^2 is $\sigma_{\bar{x}} = \sigma_X''/\sqrt{1000}$ and $\sigma_{s^2} = (\sigma_X'')^2\sqrt{2/999}$, respectively. Numerical estimates of $\mu_{\bar{x}}$, $\sigma_{\bar{x}}$, μ_{s^2}, σ_{s^2} were obtained by means of performing the updating problem 1000 times and are denoted $\hat{\mu}_{\bar{x}}$, $\hat{\sigma}_{\bar{x}}$, $\hat{\mu}_{s^2}$, $\hat{\sigma}_{s^2}$. The estimates $\hat{\mu}_{\bar{x}}$ and $\hat{\sigma}_{\bar{x}}$ are compared to their corresponding reference values in Table 3. The closer the ratio $\hat{\mu}_{\bar{x}}/\mu_{\bar{x}}$ is to 1.0, the smaller is the bias in the estimate of the posterior mean of θ_1. For BUS and aBUS the bias is negligible and independent of the dimension d. For TMCMC, the bias increases with d; for $d \geq 10$ the bias is larger than 5%. The ratio $\hat{\sigma}_{\bar{x}}/\sigma_{\bar{x}}$ that is also presented in Table 3 can be regarded as an indirect measure of the correlation of the samples in Θ. A ratio of one indicates that samples in Θ are uncorrelated; the larger the ratio is compared to the optimum of one, the higher the correlation of the samples in Θ. For TMCMC, the ratio $\hat{\sigma}_{\bar{x}}/\sigma_{\bar{x}}$ is considerably larger than the one of BUS and aBUS, whereas the performance of BUS and aBUS is similar. The bias in the estimate of the posterior variance can be assessed by means of the ratio $\hat{\mu}_{s^2}/\mu_{s^2}$ listed in Table 4, where the bias is zero if the ratio is equal to one. For all three investigated methods, the bias in the estimate of the posterior variance $\hat{\mu}_{s^2}$ is larger than the bias in the estimate of the posterior mean $\hat{\mu}_{\bar{x}}$ and increases with increasing d. For $d \leq 7$, the bias in TMCMC and aBUS is comparable and smaller than the bias in BUS. For $d \geq 10$, BUS has the smallest bias, and TMCMC the largest. The ratio $\hat{\sigma}_{s^2}/\sigma_{s^2}$ (Table 4) shows how much the standard deviation $\hat{\sigma}_{s^2}$ of the estimator s^2 deviates from the reference value σ_{s^2}. The behavior of the three methods is similar with one exception: for $2 \leq d \leq 7$, TMCMC has the smallest $\hat{\sigma}_{s^2}/\sigma_{s^2}$.

Table 3. Statistical properties of the estimated posterior mean of x_1 for increasing d (Example 1).

	$\frac{\hat{\mu}_{\bar{x}}}{\mu_{\bar{x}}}$			$\frac{\hat{\sigma}_{\bar{x}}}{\sigma_{\bar{x}}}$		
d	TMCMC	BUS	aBUS	TMCMC	BUS	aBUS
1	1.0	0.98	0.99	33.6	13.4	11.8
2	0.99	0.99	1.00	21.6	8.7	7.7
3	0.99	0.99	1.00	18.4	7.3	6.5
5	0.98	0.99	1.00	10.9	6.4	5.6
7	0.96	0.99	1.00	14.5	6.0	5.7
10	0.93	1.00	1.00	21.5	6.3	6.6
15	0.88	1.00	1.00	24.0	7.3	7.9
20	0.88	1.00	1.00	20.3	8.4	9.2

Table 4. Statistical properties of the estimated posterior variance of x_1 for increasing d (Example 1).

	$\frac{\hat{\mu}_{s^2}}{\mu_{s^2}}$			$\frac{\hat{\sigma}_{s^2}}{\sigma_{s^2}}$		
d	TMCMC	BUS	aBUS	TMCMC	BUS	aBUS
1	0.99	1.02	0.98	5.9	5.4	5.3
2	0.98	1.14	1.01	2.6	5.4	5.1
3	0.97	1.15	1.00	2.7	5.5	5.2
5	0.97	1.08	0.98	3.5	5.4	5.2
7	0.96	1.04	0.94	4.2	5.3	5.5
10	0.90	0.98	0.93	5.2	5.4	5.7
15	0.74	0.93	0.90	5.6	5.6	6.0
20	0.62	0.90	0.88	4.9	5.7	6.5

Let Θ_i denote $i = 1, \dots, 1000$ sets, each obtained from a different run of the updating problem. Each set Θ_i contains 1000 posterior samples of X_1. Furthermore, let $\Theta_U = \bigcup_{i=1}^{1000} \Theta_i$ denote the union of all sets Θ_i. The variance of samples in the set Θ_U is denoted s_U^2. The expected value of s_U^2, referred to as $\mu_{s_U^2}$, is equal to the posterior variance $(\sigma_X'')^2$. Let $\hat{\mu}_{s_U^2}$ be an estimate of $\mu_{s_U^2}$, obtained with samples from either TMCMC, BUS or aBUS. Note that even if both the estimates $\hat{\mu}_{\bar{x}}$ and $\hat{\mu}_{s^2}$ are unbiased, the estimate $\hat{\mu}_{s_U^2}$ might be, nevertheless, biased. If $\hat{\mu}_{\bar{x}}$ and $\hat{\mu}_{s^2}$ are unbiased, the bias in $\hat{\mu}_{s_U^2}$ is mainly influenced by the difference of $\hat{\sigma}_{\bar{x}}$ from its theoretical

optimum. The ratio $\hat{\mu}_{s_U^2}/\mu_{s_U^2}$ is listed in Table 5 for different d. The best performance in terms of this quantity has aBUS followed by BUS. One exception is TMCMC for $d = 20$, whose estimate seems to be good. However, since the estimates $\hat{\mu}_{\bar{x}}$ and $\hat{\mu}_{s^2}$ are clearly biased in TMCMC for $d = 20$, the ratio $\hat{\mu}_{s_U^2}/\mu_{s_U^2}$ loses its explanatory power.

Table 5. Variance in a set of 10^6 posterior samples of θ_1, obtained from 10^3 runs of the updating problem with 10^3 samples generated in each run, divided by the reference variance of the posterior x_1 (Example 1).

d	TMCMC	BUS	aBUS
1	2.12	1.20	1.12
2	1.45	1.24	1.06
3	1.31	1.19	1.04
5	1.09	1.15	1.00
7	1.17	1.07	1.01
10	1.36	1.03	0.96
15	1.32	0.99	0.97
20	1.03	0.95	0.95

In a next study we assess the quality of the approximated posterior distribution of the largest and smallest coefficient in **x**. Let $x_{\max}$ denote the largest coefficient of the vector **x**, i.e. $x_{\max} = \max\{x_1, \dots, x_d\}$; and let $x_{\min}$ denote the smallest coordinate of **x**, i.e. $x_{\min} = \min\{x_1, \dots, x_d\}$. The range r is defined as $r = x_{\max} - x_{\min}$. The posterior PDF of $X_{\max}$ is defined as:

$$f''_{X_{\max}}(x) = d \cdot \frac{1}{\sigma''_X}\varphi\left(\frac{\theta-\mu''_X}{\sigma''_X}\right) \cdot \left[\Phi\left(\frac{\theta-\mu''_X}{\sigma''_X}\right)\right]^{d-1}.$$

The posterior PDF of $X_{\min}$ is defined as:

$$f''_{X_{\min}}(x) = d \cdot \frac{1}{\sigma''_X}\varphi\left(\frac{x-\mu''_X}{\sigma''_X}\right) \cdot \left[1 - \Phi\left(\frac{x-\mu''_X}{\sigma''_X}\right)\right]^{d-1}.$$

Reference values for the mean and standard deviation of $X_{\max}$, denoted $\mu''_{X_{\max}}$, $\sigma''_{X_{\max}}$, can be computed from $f''_{X_{\max}}$, and a reference value for the mean of r, denoted μ''_r, can be computed from $f''_{X_{\max}}$ and $f''_{X_{\min}}$. Numerical estimates of these quantities, denoted $\hat{\mu}''_{X_{\max}}$, $\hat{\sigma}''_{X_{\max}}$, $\hat{\mu}''_r$, obtained from 1000 runs of the updating problem, are compared to the reference values in Table 6 and Table 7. The bias in the estimate $\hat{\mu}''_{X_{\max}}$ is negligible for BUS and aBUS. For TMCMC, the bias is negligible for $d \leq 5$, however, it increases with an increase in d. Similar to $\hat{\sigma}_{\bar{x}}/\sigma_{\bar{x}}$, the ratio is an indirect indicator for how much the numerically obtained posterior samples of $X_{\max}$ are correlated. The larger this ratio the less efficient is the numerical updating technique. Looking at the values listed in Table 6, BUS and aBUS are comparable to each other, and TMCMC is clearly less efficient. In Table 7 the ratio $\hat{\mu}''_r/\mu''_r$ and the coefficient of variation of the estimate $\hat{\mu}''_r$ is listed. aBUS gives the estimate that has on average the smallest bias. The estimate of BUS has a small bias for $d \geq 10$. The estimate of TMCMC has a small bias for $d \leq 5$ and for $d = 20$, however, the coefficient of variation of the estimate is larger than the one obtained with BUS or aBUS for $d \leq 5$.

Table 6. Statistical properties of the posterior $x_{\max}$ for increasing d (Example 1).

	$\frac{\hat{\mu}''_{X_{\max}}}{\mu''_{X_{\max}}}$			$\frac{\hat{\sigma}''_{X_{\max}}}{\sigma''_{X_{\max}}}$		
d	TMCMC	BUS	aBUS	TMCMC	BUS	aBUS
1	1.0	0.98	0.99	33.6	13.4	11.8
2	1.0	0.99	1.00	17.8	8.1	6.7
3	0.99	0.99	1.00	13.9	5.3	4.4
5	0.99	1.00	1.00	8.6	4.3	4.1
7	0.98	1.00	1.00	9.1	5.1	5.5
10	0.96	1.00	1.00	14.2	6.8	7.1
15	0.95	1.00	0.99	15.9	8.1	9.2
20	0.94	0.99	0.99	13.8	10.2	11.1

Table 7. Statistical properties of the posterior range r for increasing d (Example 1).

	$\frac{\hat{\mu}''_r}{\mu''_r}$ $(\frac{\hat{\sigma}''_r}{\hat{\mu}''_r})$		
d	TMCMC	BUS	aBUS
2	1.01 (26%)	1.15 (13%)	1.03 (14%)
3	1.02 (23%)	1.12 (12%)	1.02 (12%)
5	1.04 (18%)	1.08 (11%)	1.00 (11%)
7	1.07 (10%)	1.05 (10%)	0.98 (11%)
10	1.15 (11%)	1.01 (10%)	0.99 (10%)
15	1.14 (10%)	1.00 (9%)	0.98 (10%)
20	1.02 (8%)	0.98 (9%)	0.98 (10%)

3.2 Example 2 – sum of random variables

In this example, we investigate the performance of the methods for large d, i.e. for d up to 5000. For this example, only the statistics of the evidence and the mean number of model runs are analyzed.

Let **X** be a d-dimensional random vector. The components $\{X_i : i = 0, \dots, d\}$ of **X** are independent and have identical distributions. The prior distribution of **X** is $f(\mathbf{x}) = \prod_{i=1}^{d}\varphi(x_i)$, where φ is the PDF of the univariate standard normal distribution. Furthermore, let $h(\mathbf{x})$ be a function of **x** that is defined as $h(\mathbf{x}) = \frac{1}{\sqrt{d}}\sum_{i=1}^{d} x_i$. Consequently, the prior distribution of $h(\mathbf{x})$ is standard normal. The likelihood of the problem is defined through $h(\mathbf{x})$ as $L(\mathbf{x}|\mathbf{d}) = \frac{1}{\sigma_\varepsilon}\varphi\left(\frac{h(\mathbf{x})-\mu_\varepsilon}{\sigma_\varepsilon}\right)$, where $\sigma_\varepsilon = 0.2$ and $\mu_\varepsilon = 4$. The problem at hand has an analytical solution: The evidence c_{E} does not depend on the dimension d of the problem and can be evaluated as $c_{\mathrm{E}} = \frac{1}{\sqrt{1+(\sigma_\varepsilon)^2}}\varphi\left(\frac{\mu_\varepsilon}{\sqrt{1+(\sigma_\varepsilon)^2}}\right) = 1.785 \cdot 10^{-4}$.

First, the estimated evidence $\hat{c}_{\mathrm{E}}$ is compared to the true evidence c_{E}. Let $\mu_{\hat{c}_{\mathrm{E}}}$ denote the mean of $\hat{c}_{\mathrm{E}}$. In Table 8 the ratio $\mu_{\hat{c}_{\mathrm{E}}}/c_{\mathrm{E}}$ is listed, where the value in brackets is the coefficient of variation (C.o.V.) of the estimate, denoted $\delta_{\hat{c}_{\mathrm{E}}}$. The quantities $\mu_{\hat{c}_{\mathrm{E}}}$ and $\delta_{\hat{c}_{\mathrm{E}}}$ were estimated by 1000 runs of the updating problem. The estimates of BUS and aBUS have, on average, a bias of 4%, where the bias is independent of the dimension. The estimate of TMCMC is unbiased for $d = 1$. However, the bias in TMCMC increases rapidly with d. The

coefficient of variation of the estimate does not depend on the dimension for BUS and aBUS if $d \geq 10$, and increases with an increase in the dimension for TMCMC. Note that for TMCMC with 1000 samples per level, the problem can only be solved for $d < 1000$: The proposal distribution in the MCMC step is estimated from the covariance structure of the samples – which is not positive definite if the number of random variables is larger than the number of samples.

In Table 9 the mean number of model runs required to draw 1000 samples from the posterior distribution is listed. The values were obtained as the average of performing the updating problem 1000 times. The number of model runs is not influenced by the dimension for all three methods. Both BUS and aBUS require on average 40% fewer model evaluations than TMCMC.

For this example BUS and aBUS clearly outperform TMCMC: they solve the problem with a smaller number of model runs and the bias of the estimated evidence does not depend on the dimension. The fact that neither BUS nor aBUS are influenced by the dimension is due to the application of the component-wise Metropolis Hastings algorithm in the MCMC step. This sampler would require additional likelihood function evaluations for application within TMCMC, for which the original Metropolis Hastings algorithm is used.

Table 8. Statistical properties of the estimated evidence for increasing d (Example 2).

	$\frac{\mu_{\hat{c}_E}}{c_E}$ $(\delta_{\hat{c}_E})$		
d	TMCMC	BUS	aBUS
1	1.00 (0.39)	1.05 (0.71)	1.04 (0.66)
2	0.97 (0.59)	1.04 (0.55)	1.06 (0.54)
3	0.91 (0.72)	1.01 (0.50)	1.05 (0.48)
5	0.76 (0.96)	1.05 (0.58)	1.04 (0.44)
10	0.44 (1.26)	1.04 (0.44)	1.04 (0.42)
20	0.20 (1.89)	1.04 (0.42)	1.03 (0.41)
100	0.13 (3.06)	1.02 (0.43)	1.06 (0.41)
500	0.11 (3.88)	1.04 (0.42)	1.05 (0.42)
5000		1.03 (0.44)	1.05 (0.40)

Table 9. Mean number of model runs required to obtain samples of the posterior with 1000 samples per intermediate level (Example 2).

d	TMCMC	BUS	aBUS
1	$8.00 \cdot 10^3$	$4.47 \cdot 10^3$	$4.46 \cdot 10^3$
2	$7.97 \cdot 10^3$	$4.70 \cdot 10^3$	$4.68 \cdot 10^3$
3	$7.95 \cdot 10^3$	$4.78 \cdot 10^3$	$4.75 \cdot 10^3$
5	$8.00 \cdot 10^3$	$4.79 \cdot 10^3$	$4.79 \cdot 10^3$
10	$8.20 \cdot 10^3$	$4.79 \cdot 10^3$	$4.78 \cdot 10^3$
20	$8.34 \cdot 10^3$	$4.78 \cdot 10^3$	$4.78 \cdot 10^3$
100	$8.02 \cdot 10^3$	$4.80 \cdot 10^3$	$4.77 \cdot 10^3$
500	$7.42 \cdot 10^3$	$4.78 \cdot 10^3$	$4.77 \cdot 10^3$
5000		$4.80 \cdot 10^3$	$4.78 \cdot 10^3$

3.3 Example 3 – 2DOF dynamic problem

The response of the two-story linear structure shown in Fig. 1 to a narrow banded ground acceleration is investigated. This example was originally discussed in [3], where the performance of the TMCMC method is assessed.

The ground acceleration and the acceleration of the roof are measured every 0.02 seconds for a period of one second, where the measured roof acceleration is contaminated with white Gaussian noise. The variance of the noise σ^2, the stiffness parameters k_1, k_2, and the damping ratio ξ of the two modes are assumed uncertain. The aim of the analysis is to update our belief about the uncertain parameters conditioned on the measurements. Our prior belief is: k_1, k_2 are uniformly distributed on the interval $[0; 3000]$, ξ is uniform on $[0.01; 0.05]$, and σ^2 is equally likely between $[0; 1]$. The true parameter values used to generate the measured roof acceleration are: $k_1 = k_2 = 1000$, $\xi = 0.03$, and $\sigma^2 = 0.2$. The masses are assumed known: $m_1 = m_2 = 1$.

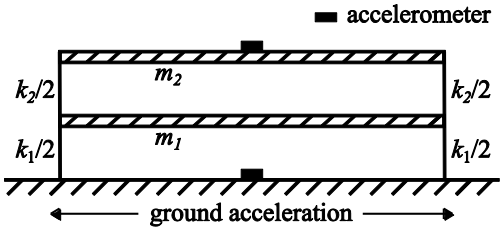

Figure 1. 2DOF system investigated in Example 3.

In this example, the maximum that the likelihood can take depends on the measured data. Therefore, only the performance of aBUS and TMCMC is assessed, and BUS is not considered. The performance of the two investigated methods is compared to a reference solution that was obtained numerically through rejection sampling, i.e. the obtained samples of the posterior are truly independent. The statistical properties of the reference solution were computed by running the updating problem more than 3000 times, and the statistical properties of the solutions were obtained with aBUS and TMCMC by 1000 updating runs.

First, we assess how many model runs aBUS and TMCMC require to draw 1000 samples from the posterior distribution: aBUS needs on average $3.9 \cdot 10^3$ samples and TMCMC needs $6.0 \cdot 10^3$ samples. Consequently, with aBUS the problem is solved approximately 35% faster than with TMCMC.

The averaged estimate of the evidence is $2.87 \cdot 10^{-16}$ for aBUS, $2.79 \cdot 10^{-16}$ for TMCMC, and $2.81 \cdot 10^{-16}$ for the reference solution. Thus, the TMCMC estimate is slightly less biased than the aBUS estimate, however, the bias in both estimates is rather small. The coefficient of variation (C.o.V.) of the reference solution is approximately 3%, the C.oV. of aBUS and TMCMC is 37% and 29%, respectively. Thus, both aBUS and TMCMC have, due to correlated posterior samples, a considerable larger C.o.V. than the reference solution. TMCMC has a smaller C.o.V. than aBUS because its solution is based on more model runs.

Next, the average means of k_1 and k_2 are investigated by means of 1000 runs of the updating problem. aBUS gives $1.36 \cdot 10^3$ for k_1 and $1.13 \cdot 10^3$ for k_2, and TMCMC gives

$1.35 \cdot 10^3$ for k_1 and $1.13 \cdot 10^3$ for k_2. The reference solution is very close to the estimates of aBUS and TMCMC: $1.37 \cdot 10^3$ for k_1 and $1.12 \cdot 10^3$ for k_2. Both methods exhibit a larger deviation for the C.o.V. of the estimate of the mean: for k_1 the reference solution is 1.4%, whereas aBUS gives 15% and TMCMC 13%. The average standard deviation of k_1 is $5.5 \cdot 10^2$ for aBUS, $5.6 \cdot 10^2$ for TMCMC, and $6.1 \cdot 10^2$in case of the reference solution. Consequently, the bias in the standard deviation is larger than the bias in the mean, however, still smaller than 10% for both methods. The C.o.V. of the estimate of the standard deviation of 1000 posterior samples of k_1 is 26% for aBUS, 22% for TMCMC, and only 2% in case of the reference solution.

In a last study we assess the estimated mean maximum interstory drift of the building during the simulated excitation. The average mean estimated with aBUS is $3.61 \cdot 10^{-3}$ with a C.o.V. of 17%, and the average TMCMC estimate is $3.59 \cdot 10^{-3}$ with 15% C.o.V. The average mean of the reference solution is $5.57 \cdot 10^{-3}$ with a C.o.V. of 0.4%. Thus, both aBUS and TMCMC underestimate the maximum interstory drift on average by approximately 40%. The C.o.V. of the estimate is again considerably larger than the reference C.o.V. due to correlated posterior samples.

In summary, aBUS and TMCMC behave similar for this example, with one exception: aBUS solves the problem with fewer model runs than TMCMC.

CONCLUSION

The recently proposed BUS approach (Bayesian updating with structural reliability methods) is modified by adaptively learning the constant c, which reflects the reciprocal of the maximum of the likelihood function. It is shown that the performance of the proposed variant of BUS is similar to the one of the original BUS algorithm proposed in [1]. Both methods clearly outperform the TMCMC method for problems with many random variables. One drawback of TMCMC is that, contrary to the BUS variants, the Metropolis-Hastings algorithm cannot be performed efficiently in a component-wise manner. This renders TMCMC inefficient in case of high-dimensional problems.

The advantage of the proposed adaptive BUS variant over the original one is that no upper bound for the maximum likelihood has to be specified a-priori. However, while we showed that the adaptive variant is working well for practical problems, a theoretical analysis of the results obtained with the algorithm is still needed.

ACKNOWLEDGMENTS

With the support of the Technische Universität München - Institute for Advanced Study, funded by the German Excellence Initiative.

REFERENCES

[1] D. Straub and I. Papaioannou, Bayesian updating with structural reliability methods. *Journal of Engineering Mechanics*, ASCE, under review.

[2] S. K. Au, J. L. Beck. Estimation of small probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, 16(4): 262-277, 2001.

[3] J. Ching and Y. Chen, Transitional Markov Chain Monte Carlo Method for Bayesian Model Updating, Model Class Selection, and Model Averaging. *Journal of Engineering Mechanics*, 133: 816-832, 2007.

[4] W. R. Gilks, S. Richardson and D. J. Spiegelhalter, *Markov chain Monte Carlo in practice*. Vol. 2. CRC press, 1996.

[5] A. Gelman, J. B. Carlin, H. S. Stern and D. B. Rubin, *Bayesian data analysis*. CRC press, 2003.

[6] M. Plummer, N. Best, K. Cowles and K. Vines, CODA: Convergence diagnosis and output analysis for MCMC. *R news*, 6(1): 7–11, 2006.

[7] H. Haario, E. Saksman, J. Tamminen. Componentwise adaptation for high dimensional MCMC. *Computational Statistics*, 20: 265-273, 2005.

[8] G. O. Roberts, R. L. Tweedie. Exponential convergence of Langevin diffusions and their discrete approximations. *Bernoulli*, 2: 341-363, 1996.

[9] N. Chopin, A sequential particle filter method for static models. Biometrica, 89(3): 539-552, 2002.

[10] J. L. Beck and K.-V. Yuen, Model Selection Using Response Measurements : Bayesian Probabilistic Approach. *Journal of Engineering Mechanics*, 130: 192-203, 2004.

[11] D.J.C. MacKay, Bayesian interpolation. *Neural Computation*, 4: 415-447, 1992.

[12] L. Wassermann, Bayesian model selection and model averaging. *Journal of Mathematical Psychology*, 44: 92-107, 2000.

[13] D. Straub, Reliability updating with equality information. *Probabilistic Engineering Mechanics*, 26(2): 254–258, 2011.

[14] A. F. Smith and A. E. Gelfand, Bayesian statistics without tears: a sampling-resampling perspective. *The American Statistician*, 46(2): 84–88, 1992.

[15] I. Papaioannou, W. Betz, K. Zwirglmaier, D. Straub, MCMC algorithms for subset simulation. Manuscript, Engineering Risk Analysis Group, Technische Universität München.