Proceedings of the 9th International Conference on Structural Dynamics, EURODYN 2014
Porto, Portugal, 30 June - 2 July 2014
A. Cunha, E. Caetano, P. Ribeiro, G. Müller (eds.)
ISSN: 2311-9020; ISBN: 978-972-752-165-4

Bayesian Uncertainty Quantification and Propagation in Structural Dynamics Simulations

Costas Papadimitriou[1]
[1]Department of Mechanical Engineering, University of Thessaly, 38334 Volos, Greece
email: costasp@uth.gr

ABSTRACT: A computational framework for Bayesian uncertainty quantification and propagation (UQ+P) in structural dynamics is presented. High performance computing techniques are integrated with Bayesian techniques to efficiently handle large-order models of hundreds of thousands or millions degrees of freedom, and localized nonlinear actions activated during system operation. Fast and accurate component mode synthesis (CMS) techniques, consistent with the finite element model parameterization, are employed to achieve drastic reductions in the model order and the computational effort. Surrogate models are also used to substantially speed-up computations, avoiding full system re-analyses. Significant computational savings are achieved by adopting parallel computing algorithms to efficiently distribute the computations in available multi-core CPUs. Important issues related to the computational efficiency of the Bayesian asymptotic approximations versus the stochastic simulation algorithms for conventional or high performance computing environments are discussed. Implementation issues for Bayesian UQ+P for linear and nonlinear structural dynamics models using vibration measurements are emphasized.

KEY WORDS: Large-order models; Laplace asymptotics; MCMC; Component mode synthesis; Surrogate models; HPC.

1 INTRODUCTION

Bayesian inference is used for quantifying and calibrating uncertainty models in structural dynamics based on vibration measurements, as well as propagating these uncertainties in structural dynamics simulations for updating robust predictions of system performance, reliability and safety ([1], [2]). The Bayesian tools for identifying system and uncertainty models as well as performing robust prediction analyses are Laplace methods of asymptotic approximation [3] and stochastic simulation algorithms such as Markov Chain Monte Carlo (MCMC) [4-6], Transitional MCMC [7] and Delayed Rejection Adaptive Metropolis (DRAM) [8].

Both tools are used to represent the posterior distribution of the parameters of a model class introduced to simulate the behavior of the engineering system, as well as compute multidimensional integrals over high-dimensional spaces of the uncertain model parameters, manifested in the formulations for model class selection, robust predictions and model averaging [9]. The asymptotic approximations involve solving optimization problems as well as computing the Hessian of certain functions in a small number of points in the parameter space. The stochastic simulation tools involve generating samples for tracing and then populating the important uncertainty region in the parameter space, as well as evaluating integrals over high-dimensional spaces of the uncertain model parameters. These tools require a moderate to very large number of system re-analyses to be performed over the space of uncertain parameters. Consequently, the computational demands depend highly on the number of system analyses and the time required for performing a system analysis.

For complex models of engineering systems, one simulation may require a significant amount of time and the overall computational demands involved in the Bayesian tools may be substantial, or even excessive for stochastic simulation algorithms. The complexity and computational demands depend on the complexity of the model of the analyzed system as well as the number of uncertain parameters involved. Moreover it depends on the support of the posterior distribution of the model parameters, the size of the support in the multi-dimensional parameter space, the multimodality of the posterior PDF and the unidentifiability that may arise when the number of the data are not informative enough for the number of model parameters involved.

For FE models involving hundreds of thousands or even million degrees of freedom and localized nonlinear actions activated during system operation, the computational demands in the Bayesian framework may be excessive. The present work proposes methods for drastically reducing the computational demands at the system, algorithm and hardware levels involved in the implementation of Bayesian tools. At the system level, CMS techniques [10, 11] are integrated with Bayesian techniques to efficiently handle linear structural components of large-order models of hundreds of thousands or millions degrees of freedom and localized nonlinear actions activated during system operation. The dynamics of the linear components is represented by the dynamics of the lowest fixed-based interface modes and the interface constraints modes. At the level of the stochastic simulation algorithms, surrogate models [12] are adopted to drastically reduce the number of computationally expensive full model runs. At the computer hardware level, parallel computing algorithms [13] are proposed to efficiently distribute the computations in available multi-core CPUs. Important issues related to the computational efficiency of the asymptotic approximations versus the stochastic simulation algorithms for conventional or high performance computing (HPC) environments are discussed.

This study is organized as follows. The Bayesian framework for uncertainty quantification, calibration and propagation is reviewed in Section 2. The formulations using the asymptotic approximations and the stochastic simulation tools are presented in Sections 3 and 4, respectively. Methods to reduce the computational demands, such as model reduction techniques and surrogate models, are presented in Section 5. In Section 6 the efficient of computational tools for carrying out the Bayesian analysis in conventional (serial) and HPC (parallel) computing environments is investigated. Advantages and disadvantages of the asymptotic and stochastic simulation algorithms are investigated from the point of view of accuracy and computational efficiency. Implementation issues in linear and nonlinear structural dynamics models using vibration measurements are discussed in Section 7. Conclusions are summarized in Section 8.

2 BAYESIAN UNCERTAINTY QUANTIFICATION AND PROPAGATION

2.1 Formulation

Consider a parameterized class M_m of structural dynamics models used to predict various output quantities of interest of a system. Let $\underline{\theta}_m \in R^{N_m}$ be a set of parameters in this model class that need to be estimated using experimental data and $\underline{f}(\underline{\theta}_m \,|\, \mathsf{M}_m)$ be model predictions of output quantities of interest given a value of the parameter set $\underline{\theta}_m$.

The values of the model parameters $\underline{\theta}_m$ are considered to be uncertain. Probability distributions are convenient mathematical tools to quantify the uncertainty in these parameters. Specifically, the probability distribution of the parameter set $\underline{\theta}_m$ quantifies how plausible is each possible value of the model parameters. The user may assign a prior probability distribution $\pi_m(\underline{\theta}_m)$ to the model parameters to incorporate prior information on the values of the model parameters. The structural model and uncertainty propagation algorithms can be used to identify the uncertainty in the prediction of the output quantities of interest. However, the probability distribution $\pi_m(\underline{\theta}_m)$ is subjective based on previous knowledge and user experience.

2.2 Parameter Estimation

In Bayesian inference, the interest lies in updating the probability distribution of the model parameters $\underline{\theta}_m$ based on measurements and then propagate these uncertainties through the structural dynamics model to quantify the uncertainty in the output quantities of interest.

For this, let $D \equiv \underline{\hat{y}} = \{\hat{y}_r \in R^{N_0}, r = 1, \cdots, m\}$ be a set of observations available from experiments, where m is the number of observations. The Bayesian formulation starts by building a probabilistic model that characterizes the discrepancy between the model predictions $\underline{f}(\underline{\theta}_m \,|\, \mathsf{M}_m)$ obtained from a particular value of the model parameters $\underline{\theta}_m$ and the corresponding data $\underline{\hat{y}}$ that are available from experiments. This discrepancy always exists due to measurement and model errors. An error term $\underline{e}$ is introduced to denote this discrepancy. The observation data and the model predictions satisfies the predicton error equation

$$\underline{\hat{y}} = \underline{f}(\underline{\theta}_m \,|\, \mathsf{M}_m) + \underline{e} \tag{1}$$

A probabilistic structure for the prediction error needs to be defined in order to proceed with the Bayesian calibration. Let M_e be a family of probability model classes for the error term $\underline{e}$. This model class depend on a set of prediction error parameters $\underline{\theta}_e$ to be determined using the experimental data. Similarly to the structural model parameters $\underline{\theta}_m$, probability distribution $\pi_e(\underline{\theta}_e)$ is also assigned to quantify the possible values of the prediction error parameters $\underline{\theta}_e$.

The Bayesian approach [14] to model calibration is used for updating the values of the combined set $\underline{\theta} = (\underline{\theta}_m, \underline{\theta}_e)$ associated with the structural and the prediction error parameters. The parameters $\underline{\theta}_m$ and $\underline{\theta}_e$ can be considered to be independent with prior probability distribution for the combined set given by $\pi(\underline{\theta} \,|\, \mathsf{M}) = \pi_m(\underline{\theta}_m \,|\, \mathsf{M}_m)\ \pi_e(\underline{\theta}_e \,|\, \mathsf{M}_e)$, where $\mathsf{M} = \{\mathsf{M}_m, \mathsf{M}_e\}$ includes the structural and prediction error model classes. The updated distribution $p(\underline{\theta} \,|\, D, \mathsf{M})$ of the parameters $\underline{\theta}$, given the data D and the model class M, results from the application of the Bayes theorem

$$p(\underline{\theta} \,|\, D, \mathsf{M}) = \frac{p(D \,|\, \underline{\theta}, \mathsf{M})\ \pi(\underline{\theta} \,|\, \mathsf{M})}{p(D \,|\, \mathsf{M})} \tag{2}$$

where $p(D \,|\, \underline{\theta}, \mathsf{M})$ is the likelihood of observing the data from the model class and $p(D \,|\, \mathsf{M})$ is the evidence of the model, given by the multi-dimensional integral

$$p(D \,|\, \mathsf{M}) = \int_{\Theta} p(D \,|\, \underline{\theta}, \mathsf{M})\ \pi(\underline{\theta} \,|\, \mathsf{M})\ d\underline{\theta} \tag{3}$$

over the space of the uncertain model parameters.

The updated probability distribution of the model parameters depends on the selection of the prediction error $\underline{e}$. Invoking the maximum entropy principle, a normal distribution is a reasonable choice for the error since the normal distribution is the least informative among all distributions with specified the lowest two moments. Consequently, the prediction error is assumed to follow the normal distribution $\underline{e} \sim N(\underline{\mu}, \Sigma)$, where $\underline{\mu}$ is the mean and Σ is the covariance. The structure imposed on the mean vectors and the covariance matrices affect the uncertainty in the model parameter estimates. A diagonal matrix is a reasonable choice for the covariance matrix, that is, $\Sigma = diag(\sigma_r^2 \hat{y}_r^2)$, where the variance parameters σ_r^2 are unknown constants to be determined by the Bayesian calibration. Depending on the nature of the simulated

quantities of interest, alternative prediction error models can be used that group the variances into two or more groups, each one associated with the same variance parameter, thus reducing the number of prediction error model parameters. The usual case is to assume that σ_r^2 are the same so that the covariance is described by one parameter. A zero mean model error is usually assumed so that $\underline{\mu} = \underline{0}$. However, to take into account the bias in the model predictions of the various response quantities involved in $\underline{f}(\underline{\theta}_m \mid \mathsf{M}_m)$ and try to reconcile conflicting predictions, one could introduce a shift in the predictions by taking $\underline{\mu} \neq \underline{0}$. Recently, the effect of prediction error correlation has also been investigated and found to affect the results of the model calibration in structural dynamics when the sensors are closely located [15].

Using the prediction error equation (1), the measured quantities follow the normal distribution $\underline{\hat{y}} \sim N(\underline{f}(\underline{\theta}_m \mid D) + \underline{\mu}(\underline{\theta}_e), \Sigma(\underline{\theta}_e))$, where the explicit dependence of $\underline{\mu}(\underline{\theta}_e)$ and $\Sigma(\underline{\theta}_e)$ on $\underline{\theta}_e$ is introduced to point out that the mean and the covariance of the overall normal prediction error model depends only on the model prediction error parameters $\underline{\theta}_e$ and is independent of the structural parameters $\underline{\theta}_m$. Consequently, the likelihood $p(D \mid \underline{\theta}, \mathsf{M})$ of observing the data follows the multi-variable normal distribution given by

$$p(D \mid \underline{\theta}, \mathsf{M}) = \frac{|\Sigma(\underline{\theta}_e)|^{-1/2}}{(2\pi)^{m/2}} \exp\left[-\frac{1}{2} J(\underline{\theta}; \mathsf{M})\right] \tag{4}$$

where

$$J(\underline{\theta}; \mathsf{M}) = [\underline{\hat{y}} - \underline{f}(\underline{\theta}_m \mid \mathsf{M})]^T \Sigma^{-1}(\underline{\theta}_e)[\underline{\hat{y}} - \underline{f}(\underline{\theta}_m \mid \mathsf{M})] \tag{5}$$

In particular, the optimal value $\underline{\hat{\theta}}$ of the model parameters corresponds to the most probable value that is obtained by maximizing the posterior probability distribution $p(\underline{\theta} \mid D, \mathsf{M})$ or, equivalently, minimizing the function

$$\begin{aligned} g(\underline{\theta}; \mathsf{M}) &= -\ln p(\underline{\theta} \mid D, \mathsf{M}) \\ &= \frac{1}{2} J(\underline{\theta}; \mathsf{M}) + \frac{1}{2} |\Sigma(\underline{\theta}_e)| - \ln \pi(\underline{\theta} \mid \mathsf{M}) \end{aligned} \tag{6}$$

For the case for which analytical expressions for $\underline{f}(\underline{\theta}_m \mid \mathsf{M})$ are available, computationally efficient gradient-based optimization algorithms can be used to obtain the optimal value of the model parameters by minimizing the function $g(\underline{\theta}; \mathsf{M})$.

2.3 Model Selection

The Bayesian probabilistic framework can also be used to compare two or more competing model classes and select the optimal model class based on the available data. Consider a family $\mathsf{M}_{Fam} = \{\mathsf{M}_i, \ i = 1, \cdots, \kappa\}$, of κ alternative, competing, parameterized FE and prediction error model classes, and let $\underline{\theta}_i \in R^{N_{\theta_i}}$ be the free parameters of the model class M_i. The posterior probabilities $P(\mathsf{M}_i \mid D)$ of the various model classes given the data D is [16]

$$P(\mathsf{M}_i \mid D) = \frac{p(D \mid \mathsf{M}_i)\ P(\mathsf{M}_i)}{p(D \mid \mathsf{M}_{Fam})} \tag{7}$$

where $P(\mathsf{M}_i)$ is the prior probability and $p(D \mid \mathsf{M}_i)$ is the evidence of the model class M_i. The optimal model class M_{best} is selected as the one that maximizes $P(\mathsf{M}_i \mid D)$ given by (7). Model class selection is used to compare between alternative model classes and select the best model class (e.g. [17]) as well as for structural damage identification [18]

2.4 Uncertainty Propagation

Let q be an output quantity of interest in structural dynamics simulations. Posterior robust predictions of q are obtained by taking into account the updated uncertainties in the model parameters given the measurements D. Let $p(q \mid \underline{\theta}, \mathsf{M})$ be the conditional probability distribution of q given the values of the parameters. Using the total probability theorem, the posterior robust probability distribution $p(q \mid D, \mathsf{M})$ of q, taking into account the model M and the data D, is given by [1]

$$p(q \mid D, \mathsf{M}) = \int p(q \mid \underline{\theta}, \mathsf{M})\ p(\underline{\theta} \mid D, \mathsf{M})\ d\underline{\theta} \tag{8}$$

as an average of the conditional probability distribution $p(q \mid \underline{\theta}, \mathsf{M})$ weighting by the posterior probability distribution $p(\underline{\theta} \mid D, \mathsf{M})$ of the model parameters.

Let $G(q; \underline{\theta})$ be a function of a deterministic output quantity of interest $q(\underline{\theta})$. A posterior robust performance measure of the system given the data D is

$$E[G(q; \underline{\theta}) \mid D, \mathsf{M})] = \int G(q; \underline{\theta})\ p(\underline{\theta} \mid D, \mathsf{M})\ d\underline{\theta} \tag{9}$$

For $G(q; \underline{\theta}) = q(\underline{\theta})$ and $G(q; \underline{\theta}) = (q(\underline{\theta}) - E[q(\underline{\theta}) \mid D, \mathsf{M}])^2$, the measure (9) is the robust mean and the variance of the output quantity of interest q taking into account the model parameter uncertainties that are estimated by the data D.

3 ASYMPTOTIC APPROXIMATIONS

3.1 Posterior PDF, Model Selection and Robust Predictions

For large enough number of measured data, the posterior distribution of the model parameters in (2) can be asymptotically approximated by a Gaussian distribution [3]

$$p(\underline{\theta} \mid D,\mathsf{M}) \approx \frac{|\, h(\underline{\hat{\theta}})\,|^{1/2}}{(2\pi)^{N_\theta/2}} \exp\left[-\frac{1}{2}(\underline{\theta}-\underline{\hat{\theta}})^T h(\underline{\hat{\theta}})(\underline{\theta}-\underline{\hat{\theta}})\right] \quad (10)$$

centered at the most probable value $\underline{\hat{\theta}}$ of the model parameters with covariance matrix equal to the inverse of the Hessian $h(\theta) = \underline{\nabla}\underline{\nabla}^T g(\underline{\theta},\mathsf{M})$ of the function $g(\underline{\theta};\mathsf{M})$ in (6) evaluated at the most probable value $\underline{\hat{\theta}}$. This approximation is also known as the Bayesian central limit theorem. The asymptotic expression (10), although approximate, provides a good representation of the posterior PDF for a number of applications involving even a relatively small number of data. Given the Gaussian approximation (10), the marginal distributions of the parameters are readily obtained to be Gaussian distributions with means and variances equal to the individual means appearing in the mean vector $\underline{\hat{\theta}}$ and the variances appearing in the diagonal elements of the covariance matrix $h^{-1}(\underline{\hat{\theta}})$.

The asymptotic approximation (10) fails to provide an adequate representation of the posterior probability distribution in the case of multimodal distributions. To improve on the asymptotic approximation, one needs to identify all modes of the posterior PDF and take them into account in the asymptotic expression by considering a weighted contribution of each mode with weights based on the probability volume of the PDF in the neighborhood of each mode [3]. The weighted estimate is reasonable, provided that the modes are separable. For interacting modes or closely spaced modes this estimate is inaccurate due to overlapping of the regions of high probability volume involved in the interaction. Numerical implementation problems arise in multi-modal cases, associated with the inconvenience in estimating all modes of the distribution [19]. The asymptotic approximation fails to provide acceptable estimates for un-identifiable cases [20] manifested for relatively large number of model parameters in relation to the information contained in the data.

The results from the asymptotic estimate are also useful for efficiently populating the posterior PDF with samples generating from MCMC algorithms. For uni-modal posterior PDFs, the asymptotic estimate can be performed as a first step in a Bayesian analysis to obtain information and identify the importance region in the parameter space of high posterior probability volume. Then the mode of the distribution can be used as a starting point of a stochastic simulation algorithm for exploring the support of the posterior PDF, while the Hessian at the mode provides valuable information for selecting the proposal PDF in MCMC algorithms. For multi-modal posterior PDFs with disjoint supports, the information from an asymptotic approximation may be misleading since other important regions in the parameter space may be easily missed. As a result, the stochastic simulation algorithms starting from the mode provided by the asymptotic estimate will usually fail to adequately explore the parameter space and identify the domains with high probability volume.

For model selection, an asymptotic approximation [16, 21, 22] based on Laplace's method can also used to give an estimate of the evidence integral in (3) that appears in the model selection equation (7). Substituting this estimate in (7), the final asymptotic estimate for $P(\mathsf{M}_i \mid D)$ is given in the form

$$P(\mathsf{M}_i \mid D) = \left[p(D \mid \mathsf{M}_{Fam})\right]^{-1} \left(\sqrt{2\pi}\right)^{n_i} \frac{p(D \mid \underline{\hat{\theta}}_i,\mathsf{M}_i)\ \pi(\underline{\hat{\theta}}_i \mid \mathsf{M}_i)}{\det\left[h_i(\underline{\hat{\theta}}_i,\mathsf{M}_i)\right]} P(\mathsf{M}_i) \quad (11)$$

where $\underline{\hat{\theta}}_i$ is the most probable value of the parameters of the model class M_i and $h_i(\theta) = \underline{\nabla}\underline{\nabla}^T g_i(\underline{\theta},\mathsf{M})$ is the Hessian of the function $g_i(\underline{\theta};\mathsf{M}_i)$ given in (6) for the model class M_i. It should be noted that the asymptotic estimate for the probability of a model class M_i can readily be obtained given the most probable value and the Hessian of the particular mode. For the multi modal case the expression (11) can be generalized by adding the contributions from all modes.

For the robust prediction integrals such as (8) or (9), a similar asymptotic approximation can be applied to simplify the integrals. Specifically, substituting the posterior PDF $p(\underline{\theta} \mid D,\mathsf{M})$ from (2) into (9), one obtains that the robust prediction integral is given by [1]

$$E[G(q;\underline{\theta}) \mid D,\mathsf{M})] = \left[p(D \mid \mathsf{M})\right]^{-1} \int G(\underline{\theta};\mathsf{M})\ p(D \mid \underline{\theta},\mathsf{M})\ \pi(\underline{\theta} \mid \mathsf{M})\ d\underline{\theta} \quad (12)$$

Introducing the function

$$r_G(\underline{\theta};\mathsf{M}) = -\ln[G(\underline{\theta};\mathsf{M})\ p(D \mid \underline{\theta},\mathsf{M})\ \pi(\underline{\theta} \mid \mathsf{M})] \quad (13)$$

the integral in (12) takes the form of Laplace integral which can be approximated as before in the form:

$$\int \exp[-r_G(\underline{\theta})]\, d\underline{\theta} = \frac{\exp[-r_G(\underline{\tilde{\theta}})]\left[\sqrt{2\pi}\right]^m}{\det\left[H_G(\underline{\tilde{\theta}})\right]} \quad (14)$$

where $\underline{\tilde{\theta}}$ is the value of $\underline{\theta}$ that minimizes the function $r_G(\underline{\theta};\mathsf{M})$, and $H_G(\underline{\tilde{\theta}},\mathsf{M})$ is the Hessian of the function $r_G(\underline{\theta};\mathsf{M})$ evaluated at $\underline{\tilde{\theta}}$. Substituting in (12), using (11) to asymptotically approximate the term $p(D \mid \mathsf{M})$ and replacing $r_G(\underline{\theta})$ by (13), it can be readily derived that $E[G(q;\underline{\theta}) \mid D,\mathsf{M})]$ is given by the asymptotic approximation [23]

$$E[G(q) \mid D,\mathsf{M})] = G(\underline{\tilde{\theta}};\mathsf{M})\ \frac{p(D \mid \underline{\tilde{\theta}},\mathsf{M})}{p(D \mid \underline{\hat{\theta}},\mathsf{M})}\ \frac{\pi(\underline{\tilde{\theta}} \mid \mathsf{M})}{\pi(\underline{\hat{\theta}} \mid \mathsf{M})}\ \frac{\det\left[h(\underline{\hat{\theta}},\mathsf{M})\right]}{\det\left[H(\underline{\tilde{\theta}},\mathsf{M})\right]} \quad (15)$$

The error in the asymptotic estimate is of order N^{-2}. However, the asymptotic estimate requires solving two extra

optimization problems, one for the mean and one for the variance of $G(q;\underline{\theta})$. In general, one needs to carry out $2N_G$ extra optimization problems, where N_G is the number of output quantities of interest. Such optimization problems are independent and can be performed in parallel.

Similarly, the asymptotic approximation for the posterior robust probability distribution $p(q \mid D,\mathsf{M})$ of q is given by

$$p(q \mid D,\mathsf{M}) = p(q(\tilde{\underline{\theta}}(q));\mathsf{M})\ \frac{p(D \mid \tilde{\underline{\theta}},\mathsf{M})}{p(D \mid \hat{\underline{\theta}},\mathsf{M})}\ \frac{\pi(\tilde{\underline{\theta}}(q) \mid \mathsf{M})}{\pi(\hat{\underline{\theta}} \mid \mathsf{M})}\ \frac{\det\left[h(\hat{\underline{\theta}},\mathsf{M})\right]}{\det\left[H_p(\tilde{\underline{\theta}}(q),\mathsf{M})\right]} \tag{16}$$

where $\tilde{\underline{\theta}}(q)$ is the value of $\underline{\theta}$ that minimizes the function

$$r_p(\underline{\theta};\mathsf{M}) = -\ln[\,p(q \mid \underline{\theta},\mathsf{M})\ p(D \mid \underline{\theta},\mathsf{M})\ \pi(\underline{\theta} \mid \mathsf{M})] \tag{17}$$

and $H_p(\tilde{\underline{\theta}}(q),\mathsf{M})$ is the Hessian of the function $r_p(\underline{\theta};\mathsf{M})$ evaluated at $\tilde{\underline{\theta}}$. The estimate of the robust posterior probability distribution of q using (16) can be implemented efficiently in a parallel computer cluster, carrying out simultaneously the optimization problems for a range of q values.

3.2 Gradien- based Optimization Algorithms

The optimization problems that arise in the asymptotic approximations are solved using available single objective optimization algorithms. The optimization of $g(\underline{\theta};\mathsf{M})$ given in (6) and the optimization of $r_G(\underline{\theta};\mathsf{M})$ or $r_p(\underline{\theta};\mathsf{M})$ given in (13) or (17), respectively, with respect to $\underline{\theta}$ can readily be carried out numerically using any available algorithm for optimizing a nonlinear function of several variables. In particular, iterative gradient-based optimization algorithms can be conveniently used to achieve fast convergence to the optimum. However, to guarantee the convergence of the gradient-based algorithms for models involving a relatively large number of DOFs, the gradient of the objective function with respect to the parameter set $\underline{\theta}$ has to be estimated with sufficient accuracy. It has been observed that numerical algorithms such as finite difference methods for gradient evaluation do not converge due to the fact that the errors in the numerical estimation may provide the wrong directions in the search space, especially for intermediate parameter values in the vicinity of a local/global optimum. The remedy is to provide analytical expressions for the gradients of the objective function. This, however, requires the development of the analytical equations for the gradients of the response quantities of interest involved in the objective functions $g(\underline{\theta};\mathsf{M})$ and $r_G(\underline{\theta};\mathsf{M})$ which, for complex models of systems, might not be convenient or it may be impossible to accomplish for non-smooth systems.

Adjoint methods, if applicable for a system, provide a fast estimate of the gradients of the objective function with respect to all parameters, which is computationally very effective since it requires the solution of a single adjoint problem for finding the gradients, independently of the number of variables in the set $\underline{\theta}$. Example of adjoint methods for Bayesian parameter estimation can be found in [24] for linear structural dynamics applications of the Bayesian framework based on modal frequencies and mode shapes. In particular, for linear representation of the stiffness and mass matrices with respect to the model parameters, adjoint methods can be made model non-intrusive as presented in Section 7.1. For nonlinear models of structures, the adjoint techniques are model intrusive, requiring tedious algorithmic and software development that in most cases are not easily integrated within the commercial software packages. Selected examples of model intrusiveness includes the sensitivity formulation for hysteretic-type nonlinearities in structural dynamics and earthquake engineering [25, 26], and the adjoint formulation for certain classes of turbulence models in computational fluid dynamics applications [27].

Independent of the computer resources available, a drawback of the gradient-based optimization algorithms is that they may convergence to a local optimum, failing to estimate the global optimum for the cases where multiple local/global optima exist.

3.3 Stochastic Optimization Algorithms

Evolution strategies are more appropriate and effective to use in cases of multiple local/global optima. Evolution strategies are random search algorithms that explore better the parameter space for detecting the neighborhood of the global optimum, avoiding premature convergence to a local optimum. A disadvantage of evolution strategies is their slow convergence at the neighborhood of an optimum since they do not exploit the gradient information. However, evolutionary strategies are highly parallelizable so the time to solution in a HPC environment is often comparable to conventional gradient based optimization methods, with the extra advantages that evolutionary stategies will have a better chance of finding the global optimum. In addition, stochastic optimization algorithms do not require the evaluation of the gradient of the objective function with respect to the parameters. Thus, they are model non-intrusive since there is no need to formulate the adjoint problem. In some cases the adjoint formulation requires considerable algorithmic development time to set up the equations for the adjoint problem and implement this formulation in software. In other cases (e.g. contact and impact problems) the development of an adjoint formulation or analytical equations for the sensitivity of objective functions to parameters is not possible.

Stochastic optimization algorithms can be used with parallel computing environments to find the optimum for non-smooth functions or for models that an adjoint formulation is not possible to develop. Examples include hysteretic models of structural components, as well as problems involving contact and impact. In the absence of a HPC environment, the disadvantage of the stochastic optimization algorithms arises from the high number of system re-analyses which may make the computational effort excessive for real world problems for which a simulation may take minutes, hours or even days to complete.

The covariance matrix adaptation (CMA) algorithm [28] exhibits fast convergence properties among several classes of evolutionary algorithms, especially when searching for a single global optimum. In this work, a parallelized version of the CMA is used to solve the single-objective optimization problems arising in the asymptotic estimates.

The Hessian estimation required in Bayesian asymptotic approximations can be computed using the Romberg method [29]. This procedure is based on a number of system re-analyses at the neighborhood of the optimum, which can all be performed independently for problems involving either calibration or propagation, and are thus highly parallelizable.

Note that an alternative way for uncertainty propagation that can substantially expedite the propagation process as well as improve the accuracy of the estimates in a HPC environment is to draw samples from the asymptotic Gaussian posterior PDF and then provide a sampling estimate of the robust propagation integral. The sample generation from the Gaussian posterior PDF and the propagation to provide robust estimate of the uncertainties of a number of important quantities of interest are fully parallelized processes.

4 STOCHASTIC SIMULATION ALGORITHMS

Stochastic simulation algorithms (e.g. Markov Chain Monte Carlo (MCMC) [4-6] and Transitional MCMC [7]) are used to generate samples $\underline{\theta}^{(i)}$, $i=1,\cdots,N$, that populate the posterior pdf in (2). Among the stochastic simulation algorithms available, the transitional MCMC algorithm (TMCMC) [7] is one of the most promising algorithms for finding and populating with samples the importance region of interest of the posterior probability distribution, even in challenging unidentifiable cases and multi-modal posterior distributions. Approximate methods based on Kernels are then used to estimate marginal distributions of the parameters. In addition, the TMCMC method yields an estimate of the evidence in (3) of the model class M_i based on the samples generated by the algorithm.

Stochastic simulation methods can be conveniently used to estimate the multi-dimensional integrals (8) and (9) from the samples $\underline{\theta}^{(i)}$, $i=1,\cdots,N$, generated from the posterior probability distribution $p(\underline{\theta}\,|\,D,\mathsf{M})$. In this case, the integrals (8) and (9) can be approximated by the sample estimates

$$p(q\,|\,D,\mathsf{M}) \approx \frac{1}{N}\sum_{i=1}^{N} p(q\,|\,\theta^{(i)},\mathsf{M}) \tag{18}$$

$$E[G(q;\underline{\theta})\,|\,D,\mathsf{M})] \approx \frac{1}{N}\sum_{i=1}^{N} G(q;\underline{\theta}^{(i)}) \tag{19}$$

respectively. For $G(q)=q$ and $G(q)=(q-E[q\,|\,D,\mathsf{M}])^2$, the estimate (19) respectively simplify to the posterior (updated) robust mean of q

$$\overline{\mu}_q \equiv E[q(\underline{\theta})\,|\,D,\mathsf{M})] \approx \frac{1}{N}\sum_{i=1}^{N} q(\underline{\theta}^{(i)};\mathsf{M}) \tag{20}$$

and the posterior robust variance $\overline{\sigma}_q^2 = E[(q(\underline{\theta})-\overline{\mu}_q)^2\,|\,D,\mathsf{M})]$ of q given by

$$\overline{\sigma}_q^2 \approx \frac{1}{N}\sum_{i=1}^{N}[q(\underline{\theta}^{(i)};\mathsf{M})-\overline{\mu}]^2 \tag{21}$$

taking into account the model uncertainties.

4.1 Parallel TMCMC in HPC Environment

HPC techniques are used to reduce the computational time of TMCMC algorithm [7]at the computer hardware level. The TMCMC algorithm is very-well suited for parallel implementation in a computer cluster. Details of the parallel implementation are given in [13]. Specifically, a parallel implementation algorithm is activated at every stage of the TMCMC algorithm exploiting the large number of short, variable length, chains that need to be generated starting from the leader samples determined from the TMCMC algorithm at the particular stage. Static and dynamic scheduling schemes can be conveniently used to optimally distribute these chains in a multi-host configuration of complete heterogeneous computer workers. The static scheduling scheme distributes the chains in the workers using a weighted round-robin algorithm so that the number of likelihood evaluations is arranged to be the same for each computer worker. The static scheduling scheme is computational efficient when the computational time for a likelihood evaluation is the same independently of the location of sample in the parameter space as well as when surrogate estimates are not activated. The dynamic scheduling scheme is more general, ensuring a more efficient balancing of the loads per computer worker in the case of variable run time of likelihood function evaluations and unknown number of surrogates activated during estimation. Specifically, each worker is periodically interrogated at regular time intervals by the master computer about its availability and samples from TMCMC chains are submitted to the workers on a first come first serve basis to perform the likelihood function evaluations so that the idle time of the multiple workers is minimized.

It should be noted that uncertainty propagation using stochastic simulation algorithms is highly parallelizable. For infinite computing resources, the time to solution for making robust prediction of a number of response quantities of interest can be of the order of the time to solution for one simulation run. In addition, in contrast to asymptotic approximations, stochastic simulation algorithms are non-local methods capable of providing accurate representations for the posterior PDF and accurate robust predictions of quantities of interest.

5 MODEL REDUCTION TECHNIQUES AND SURROGATE MODELS

5.1 Component Mode Synthesis (CMS) Technique

Model reduction techniques can be applied at the system level to reduce the order of the model selected to simulate the behavior of the system. The objective is to obtain reduced models that run significantly faster than the original high-fidelity models, incorporating the important dynamics of the

system analyzed so that the simulations from the reduced model are sufficiently accurate.

In structural dynamics, dynamic reduction techniques have been integrated with Bayesian techniques to carry out system analyses in a significantly reduced space of generalized coordinates and thus efficiently handle large-order models of hundreds of thousands or millions degrees of freedom and localized nonlinear actions activated during system operation. Specifically, component mode synthesis (CMS) techniques [30-32] can be used to alleviate the computational burden associated with each model run in the re-analyses required in the asymptotic and stochastic simulation methods. CMS techniques divide the structure into components with mass and stiffness matrices that are reduced using fixed-interface and constrained modes. Dividing the structure into components and reducing the number of physical coordinates to a much smaller number of generalized coordinates certainly alleviates part of the computational effort. However, at each iteration or TMCMC sampling point one needs to re-compute the eigen-problem and the interface constrained modes for each component. This procedure is usually a very time consuming operation and computationally more expensive that solving directly the original matrices for the eigenvalues and the eigenvectors, due to the substantial computational overhead that arises at component level.

The main objective in methods involving re-analyses of models with varying properties is to avoid, to the extent possible, the re-computation of the eigenproperties at the component or system level. Such techniques have been incorporated in methods for uncertainty management in structural dynamics to efficiently handle the computational effort in system re-analyses that arise from FE model variations caused by variations in the values of the uncertain parameters [33-35]. In particular, perturbation techniques [36] provide accurate results locally for small variations of the model parameters about a reference structure. To improve the accuracy of the approximations for large variation of the model parameters, methods have been proposed to approximate the modes at the component or system level in terms of the modes of a family of structures corresponding to support points in the parameter space [33]. In [37], linear and quadratic interpolations of the structural mass and stiffness matrix and the matrix of eigenvectors at the component and/or system level using support points in the larger region in the parameter space have been proposed. Such methods have been successfully used for model updating of large-order models of structures [38], and for damage detection at component level [39]. These techniques proved to be quite effective in substantially reducing the computational demands in problems requiring system re-analyses.

Fast and accurate CMS techniques, consistent with the finite element (FE) model parameterization, have recently been proposed [10] to achieve drastic reductions in computational effort. Specifically, for certain parameterization schemes for which the mass and stiffness matrices of a component depend linearly on only one of the free model parameters to be updated, often encountered in FE model updating formulations, the full re-analyses of the component eigen-problems are avoided, reducing substantially the computational demands, without compromising the solution accuracy. The eigenproperties and the interface constrained modes at a structural component as a function of the model parameters can be computed inexpensively from the eigenproperties and the interface constrained modes that correspond to a nominal value of the model parameters.

Specifically let Δ_j be the set of structural components that depend on the j-th parameter θ_j. Consider the case for which the stiffness matrix of a component $s \in \Delta_j$ depends linearly on θ_j and the mass matrix is independent of θ_j, i.e. $K^{(s)} = \bar{K}^{(s)}\theta_j$ and $M^{(s)} = M_0^{(s)}$. It can be readily derived that the stiffness and mass matrices of the Craig-Bampton reduced system admits the representation

$$\hat{K}^{CB} = \hat{K}_0^{CB} + \sum_{j=1}^{N_\theta} \hat{K}_{,j}^{CB}\theta_j \tag{22}$$

and

$$\hat{M}^{CB} = \hat{M}_0^{CB} \tag{23}$$

It is important to note that the assembled matrices $\hat{K}_0^{CB}$, $\hat{K}_{,j}^{CB}$ and $\hat{M}_0^{CB}$ of the Craig-Bampton reduced system in the expansion (22) and (23) are independent of the values of $\underline{\theta}$. In order to save computational time, these constant matrices are computed and assembled once and, therefore, there is no need this computation to be repeated during the iterations involved in optimization or TMCMC sampling algorithms for model updating due to the changes in the values of the parameter vector $\underline{\theta}$. This saves substantial computational effort since it avoids (a) re-computing the fixed-interface and constrained modes for each component, and (b) assembling the reduced matrices from these components.

Further reduction in the generalized coordinates can be achieved by replacing the interface DOFs by a reduced number of constraint interface modes [40] formed by a reduced basis. Selecting the reduced basis to be constant, independent of $\underline{\theta}$, the formulation significantly simplifies. The reduced basis can be kept constant at each iteration involved in the optimization algorithm or updated every few iterations in order to improve convergence and maintain accuracy.

Following the formulation proposed in [10], the aforementioned framework can be extended to handle the case for which the component stiffness and mass matrices depend nonlinearly on a single parameter θ_j of the system parameter set $\underline{\theta}$. This is the case for which the stiffness and mass matrices of a component $s \in \Delta_j$ depend nonlinearly on θ_j, i.e. $K^{(s)} = \bar{K}^{(s)} f^{(s)}(\theta_j)$ and $M^{(s)} = \bar{M}^{(s)} g^{(s)}(\theta_j)$, where $f^{(s)}(\theta_j)$ and $g^{(s)}(\theta_j)$ are nonlinear functions of the parameter θ_j. The interface modes, the modal frequencies and the interface constrained modes of a component can readily be computed by the corresponding interface modes, modal

frequencies and interface constrained modes of the same component for a reference structural configuration corresponding to a particular nominal value of the parameter set $\underline{\theta}$ as well as the current value of the parameter θ_j. In the nonlinear case, a representation similar to (22) and (23) is no longer applicable and the reduced mass and stiffness matrices of the reduced structure should be re-assembled from the component mass and stiffness matrices for the new value of θ_j. This procedure also saves substantial computational effort since it avoids re-computing the fixed-interface and constrained modes for each component.

Applications of the parameterization consistent CMS techniques in structural identification and Bayesian UQ+P can be found in [10] for a high fidelity solid finite element model of a bridge. Reduction of three to four orders of magnitude in generalized DOFs were achieved for accurately estimating the lowest 20 modes of the bridge. The aforementioned CMS techniques have also been extended to reduce the models of linear structural components of nonlinear structures. Applications in civil infrastructure can be found in [11] for a base isolated building where the superstructure behaves linearly and the isolation is nonlinear. Drastic reductions in the time required to carry out the simulations of the nonlinear system with reduced linear components, within the Transitional MCMC, were observed, without compromising the solution accuracy. Model reduction techniques and parallelization were used in [38] to drastically reduce the computational time for a class of models for complex aerospace structures involving hundreds of thousands of DOFs. The model reduction techniques are essential in certain damage identification methods that are based on analyzing a large number of parameterized finite element model classes, each one monitoring a possible damage scenario in the structure, for selecting the model class and thus the damage scenario that best fits the measured data [18]. It has been demonstrated that the computational time for challenging model-based damage identification (detection and localization of damage) of structures modeled by finite element models with hundreds of thousands of DOFs can be drastically reduced by exploiting model reduction techniques and parallelization [41].

5.2 Surrogate Models

Surrogate models are used to reduce the computational time at the level of the algorithm. The objective is to avoid the full structural dynamics model runs at a sampling point in the parameters space by exploiting the function evaluations that are available at the neighbour (design) points in order to generate an approximate estimate. Surrogate models are well-suited to be used with the TMCMC method. Details of the implementation of surrogate models with TMCMC algorithm are given in [13]. Specifically, following [13], a kriging technique [12] is used to approximate the function evaluation at a new sampling point at a TMCMC stage using the function evaluations at neighbour points in the parameter space available from previous TMCMC stages. To ensure a high quality approximation, a surrogate estimate is accepted only if it satisfies certain conditions as follows.

The surrogate estimate is based on a user-defined number of support points which are in the neighbor of the surrogate point. The minimum number of support points depends on the dimension of the uncertain parameter space and the order of the kriging interpolation. The surrogate point belongs to the convex hull of the design points so that an interpolation is performed, while extrapolations are prohibited. The design points correspond to actual system simulations and not other surrogate estimates from previous stages, avoiding error propagation and subsequent deterioration of the surrogate quality. The design point are kept the same when generating the surrogate estimates within a chain of the TMCMC stage, avoiding discontinuities in the estimates of the sampling points in a chain caused by changing the design points. The surrogate estimate is checked whether its predicted value is within the lower 95% quantile of all the design point's likelihood values accounted so far. The purpose of the threshold is to prevent overshooting surrogate estimates as this will quickly lead to the breakdown of the sampling procedure due to the concentration of most points around this overshooting estimate. The surrogate estimate is accepted if the prediction error is smaller than a user specified tolerance value.

It has been demonstrated that the proposed adaptive kriging method can achieve up to one order of magnitude reduction in computational effort.

6 CONVENTIONAL VERSUS HIGH PERFORMANCE COMPUTING

The effectiveness of the Bayesian computing tools outlined in the previous sections, in terms of convenience and computational efficiency, depends on the computing environment available. Next, we will discuss the effectiveness for the case where the computations run in series in conventional multi-core desktop machines or in parallel in HPC environments involving a large number of cores.

It should be noted that the time to solution for representing the posterior PDF by either an approximate Gaussian distribution or by MCMC samples drawn from the posterior PDF, depends on the computational time required to perform the number of system re-analyses involved in the asymptotic approximations and stochastic simulation algorithms. The time to solution can be significantly reduced in HPC environments if the processes involved in these two classes of Bayesian tools are parallelizable. For the two classes of Bayesian computing tools, requiring gradient based or stochastic optimization algorithms as well as stochastic simulation algorithms, computational efficient schemes for conventional as well as high performance computing environments are next discussed.

6.1 Gradient-Based Algorithms

Gradient-based optimization algorithms are iterative algorithms for which the computations at an iteration depend on the computations involved in the previous iterations. So parallelization of the iterative scheme is not possible. Independently of the number of computer cores available, the time to solution equals to

$$t_{G,sol} = (t_f + t_a) N_{G,iter} \tag{24}$$

where t_f and t_a are the computational times required to solve the forward problem and the adjoint problem or the equations for computing the gradients of the objective function, respectively, and $N_{G,iter}$ is the number of iterations required in the optimization algorithm for convergence to the optimum.

For the case for which an analytical formulation for the gradients of the objective function with respect to the number of variables is available, then at each iteration these gradient evaluations can be performed in parallel for all variables involved. In this case, the number of parallel evaluations equals the number of variables. For gradients of objectives that can be efficiently evaluated through adjoint formulations available for the system analyzed, only one adjoint system has to be solved which requires the solution of the forward system. In such cases extra computer cores for parallelization are not needed at the level of the iterative optimization algorithm. However, for certain systems, parallelization may be needed at the level of the adjoint formulation as it is described in the Section 7.1 for uncertainty quantification of linear models based on identified modal characteristics. Finally, parallelization of the solution process is also possible at the model level using, for example, the CMS techniques to carry out the analysis in all linear structural components in parallel.

All methods for computing the Hessian at the optimum, required in asymptotic approximations, can be fully parallelized. Computations of the Hessian can be performed approximately using the finite difference method, requiring $N_\theta(N_\theta+1)$ model runs. Alternatively, depending on the type of the system, more accurate second-order adjoint methods for the system at hand can be developed which usually requires N_θ computations that can be performed in parallel. The solution of the N_θ adjoint problems can be parallelized to be executed simultaneously, yielding a time to solution equal to the one required for solving a single second-order adjoint problem. This means that the time to solution in a parallel environment increases by a factor of $(t_f+t_a+t_{aa})/(t_f+t_a)$, where t_{aa} is the solution time required for the second-order adjoint system. Thus, in HPC environment the adjoint system re-analyses does not add to the computational effort required for estimating the asymptotic Gaussian posterior PDF. For serial computations, the computing time for second-order adjoint system is $t_{aa}N_\theta$ which depends on the number N_θ of parameters and the solution time t_{aa} of the second-order adjoint system. This additional time is significant only if the number of parameters N_θ is of the order the number of iterations $N_{G,iter}$ in (24). Otherwise, the estimate in (24) controls the time to solution.

6.2 Evolutionary Algorithms

For evolutionary algorithms, such as the CMA algorithm [28] used in this work, the parallel computations can be performed at each generation simultaneously for the multiple analyses of the system for all number of population samples involved in each generation. In a parallel computing environment, the time to solution equals to

$$t_{ES,sol} = t_f N_{ES,gen} ceil(N_{ES,pop} / N_{ES,cores}) \tag{25}$$

where $N_{ES,gen}$ is the number of generations involved in the evolutionary algorithm, $N_{ES,pop}$ is the number of population at each generation and $N_{ES,cores}$ the number of computer cores available to run computations in parallel. For large number of available computer cores such that $N_{ES,cores} \geq N_{ES,pop}$, the time to solution is $t_{ES,sol} = t_f N_{ES,gen}$ which depends only on the number of generations $N_{ES,gen}$.

Comparing this simplified estimate to the time to solution for the gradient-based optimization method (assuming that $t_a \approx t_f$), it is clear that the ratio $\beta = N_{ES,gen}/(2N_{G,iter})$ is a measure of the computational efficiency of the stochastic optimization method in relation to the gradient-based optimization method. Values of this ratio close to one or less than one makes the stochastic optimization method the preferred method since the extra burden of formulating the sensitivities of response quantities with respect to the model parameters is avoided without significantly affecting the computation effort. Certainly, for values of β significantly higher than one, in selecting one method against another, one has to trade off the development time required to formulate and implement in software the gradients required in gradient based optimization methods. For computations in conventional machines, the time to solution is $t_f N_{ES,gen} N_{ES,pop}$ which may be excessive due to the very large number of function evaluations arising from the high number of population samples $N_{ES,pop}$.

6.3 Stochastic Simulation Algorithms

For single-chain MCMC stochastic simulation algorithms, parallelization of the system runs is not possible. The time to solution is

$$t_{MCMC,sol} = t_f N_{MCMC,samples} \tag{26}$$

where $N_{MCMC,samples}$ is the number of samples required in the MCMC algorithm to obtain an sufficiently accurate description of the posterior PDF. For the multi-chain MCMC algorithms, such as the TMCMC, the time to solution depends on the number of parallel chains and the number of computer cores available. For the highly-parallelized TMCMC algorithm [13], the time to solution for the case of infinite resources is given by

$$t^{parallel}_{TMCMC,sol} = t_f \sum_{j=0}^{m-1} \max_i (n^{TMCMC}_{j+1,i}) \tag{27}$$

where $n_{j+1,i}^{TMCMC}$ is the number of samples available for the i-th chain out of the $i=1,\ldots,\hat{N}_{j+1}^{TMCMC}$ chains generated at TMCMC stage $j+1$, with $\sum_{i=1}^{\hat{N}_{j+1}^{TMCMC}} n_{j+1,i}^{TMCMC} = N_{j+1}^{TMCMC}$ be the number of samples per TMCMC stage. Note that the ratio $\max_i(n_{j+1,i}^{TMCMC}) / N_{j+1}^{TMCMC} = \alpha_{j+1}$ of the largest number of samples over any chain in the stage $j+1$ over the total number of samples in the chain is a very small fraction of the number samples at a stage, i.e, $\alpha_{j+1} << 1$.

The total number of TMCMC samples for all stages is $\sum_{j=0}^{m-1} N_{j+1}^{TMCMC} = N_{samples}^{TMCMC}$. For computations carried out in a conventional desktop computer, the time to solution is

$$t_{TMCMC,sol}^{serial} = t_f \frac{N_{samples}^{TMCMC}}{N_{cores}^{desktop}} \tag{28}$$

where $N_{cores}^{desktop}$ is the number of desktop cores available. Usually, $N_{cores}^{desktop}$ is 8 for a conventional four-core multi-threaded computer. In the case of infinite computer resources, the effectiveness of the parallel computation defined as the ratio of the time to solution performed in parallel computing environment over the time to solution performed in serial computing environments is

$$\frac{t_{TMCMC,sol}^{paralel}}{t_{TMCMC,sol}^{serial}} = \frac{1}{m}\sum_{j=1}^{m-1} \alpha_{j+1} \tag{29}$$

the average value of $\alpha_{j+1} << 1$. For surrogate models, this number can be reduced considerably (by an order of magnitude), resulting in additional substantial reductions in the time to solution.

Based on the analyses in this section, it can be inferred that in the absence of a HPC environment, Bayesian asymptotic approximations using gradient-based optimization algorithms and adjoint methods are clearly the preferred tools for performing parameter estimation. For uncertainty propagation and estimation of simple measures of uncertainties such as mean and standard deviation for output quantities of interest, there might be significant computational effort involved when the number of response quantities of interest is large. Stochastic optimization algorithms should be avoided since they involve a substantially larger number of model re-analyses. For a HPC environment, it is best to implement the Bayesian asymptotic tools using stochastic optimization algorithms and the Romberg method for evaluation of the Hessian, since the time to solution is usually comparable to the one required for the gradient-based optimization algorithms and the often tedious development of the adjoint formulation is completely avoided. For certain problems where adjoint formulation are not possible, the stochastic simulation algorithms should be the preferred algorithm to use. Moreover, the availability of a HPC environment promote the use of parallelized stochastic simulation algorithms such as TMCMC since the time to solution is drastically reduced.

Model reduction methods significantly reduce the computational effort. Surrogate models mainly applicable with TMCMC reduce further the time to solution by one order of magnitude.

7 IMPLEMENTATION IN STRUCTURAL DYNAMICS

In structural dynamics the uncertainty quantification analysis and the formulation of the likelihood in (2) depends on the models used. For linear models one often employs as measurements the identified modal characteristics (modal frequencies, mode shapes and damping ratios) to quantify the uncertainty in structural model parameters. For nonlinear models one usually employs full response time history measurements or nonlinear frequency response spectra measurements. Details in the implementation of the Bayesian framework for the linear and nonlinear model cases are presented next separately for each model case and measurements available.

7.1 Bayesian UQ for Linear Modes

For linear models of structures the quantification of the uncertainties in the model parameters is often based on identified modal characteristics such as modal frequencies and mode shapes at the locations where sensors are placed. Details on the formulation of the likelihood in (2) can be found in a number of published papers (e.g. [15, 42-46]). The formulation often depends on the user postulation of the prediction errors in (1). Independently of the details in the formulation, the likelihood and the posterior of the parameters of a model, usually a finite element model, are functions of the modal frequencies and the mode shapes predicted by the finite element model. A posterior PDF evaluation for a particular value of the model parameters requires the solution of the eigenvalue problem related to stiffness and mass matrices of the structural model.

At the model level, model reductions techniques [10, 37] have been proposed to considerably reduce the size of the stiffness and mass matrices by several orders of magnitude. In particular, computational efficient model reduction techniques based on component mode synthesis have been developed to handle certain parameterization schemes for which the mass and stiffness matrices of a component depend linearly on only one of the free model parameters to be updated, often encountered in FE model updating formulations. In such schemes, it has been shown that the repeated solutions of the component eigen-problems are completely avoided, reducing substantially the computational demands, without compromising the solution accuracy [10]. The model reduction methods are applicable to both asymptotic and stochastic simulation tools used in Bayesian framework.

For Bayesian asymptotic approximations, first-order and second order adjoint techniques have been developed [24] using the Nelson's method [47] to efficient compute the required first and second order sensitivities in the optimization problems and the Hessian computations. An advantage of the Nelson method is that the gradient of the modal frequencies and the modeshape vector of a specific mode can be computed from only the value of the modal frequency and the

modeshape vector of the same mode, independently of the values of the modal frequencies and modeshape vectors of the rest of the modes. For structural model classes with large number of degrees of freedom and very few contributing modes, this representation of the gradients clearly presents significant computational advantages over methods that represent modeshape gradients as a weighted, usually arbitrarily truncated, sum of all system modeshape vectors [48]. The end result of the proposed adjoint method is the solution of as many linear systems of equations as the number of model predicted modes. The size of the linear systems equals the number of the DOFs of the structural model which adds to the computational burden. However, the linear system of equations is independent of each other and can be carried out in parallel, significantly accelerating the time to solution. The integration of model reduction techniques with the adjoint methods can be found in [10].

Adjoint methods are easily applicable with any commercial computational mechanics software package since the only information required is the representation of the mass and stiffnesss matrices with respect to the model parameters. For linear representation given by the expansion

$$K = K_0 + \sum_{i=1}^{N_\theta} K_i \theta_i \tag{30}$$

$$M = M_0 + \sum_{i=1}^{N_\theta} M_i \theta_i \tag{31}$$

the adjoint formulation becomes independent of the software package used to assemble the stiffness and mass matrices. The evaluation of the individual matrices appearing in the expansions (30) and (31) can be performed off-line using any software package. Then the adjoint methods can be formulated based on these matrices, avoiding model intrusion or integration with a computational mechanics software package.

For Bayesian asymptotic approximations using stochastic optimization algorithms, the model reduction methods can substantially reduce the computational effort. For Bayesian stochastic simulation tools, in addition to the model reduction method applied at the system level, surrogate estimates applied at the algorithmic TMCMC level can be used to reduce the number of full system re-analyses. Due to the large number of measured quantities (modal frequencies and modeshape compoents for all measured modes), surrogate estimates are conveniently used on the log posterior function instead of the individual modal frequencies and mode shape components.

7.2 Bayesian UQ for Nonlinear Models

The type of nonlinearities encountered in structural dynamics include hysteretic nonlinearities as well as nonlinearities arising from contact and impact between surfaces, as well as from nonlinear isolation devices such as nonlinear dampers in civil infrastructure and nonlinear suspension models in vehicles. In a number of structural dynamics cases, the nonlinearities are localized in isolated parts of a structure, while the rest of the structure behaves linearly. Such localized nonlinearities can be found in vehicles where the frame usually behaves linearly and the nonlinearities are activated at the suspension mainly due to the dampers. In civil engineering structures the nonlinearities are at some cases localized at the various structural elements (dampers, etc) introduced to isolate the structure during system operation.

For nonlinear models of structures the quantification of the uncertainties in the model parameters depends on the measured quantities that are available. Depending on the type of application, two types of measured quantities are usually available: full response time histories or frequency response functions. The likelihood formulation in (2) depends on the type of the measured quantities provided.

Details on the formulation of the likelihood for the case where full measured response time histories are available can be found in (e.g. [49-51]). The formulation often depends on the user postulation of the prediction errors in (1). The likelihood and the posterior of the parameters of a finite element model are functions of the response time histories predicted by the finite element model. Each posterior evaluation requires the integration of the nonlinear set of equation of motion of the structure.

The formulation of the likelihood for the case where nonlinear frequency response spectra are available can be found in [11, 52]. The likelihood and the posterior of the parameters of the nonlinear finite element model are functions of the frequency response spectra predicted by the finite element model. Each posterior evaluation requires the integration of the nonlinear set of equation of motion of the structure for as many different number of harmonic excitations as the number of frequency response spectra ordinates. This, however, increases substantially the computational effort.

At the model level, model reduction techniques based on CMS are readily applicable for special class of problems where the nonlinearities are localized at isolated parts of the structure. In such cases the structure can be decomposed into linear and nonlinear components and the dynamic behavior of the linear components be represented by reduced models. An implementation of such framework can be found in [11] where it is demonstrated that substantial reductions in the DOFs of the model can be achieved which eventually yield to reduction in computational effort for performing a simulation run without sacrificing the accuracy.

For Bayesian asymptotic approximations, analytical approximations of the gradients of objective functions are not readily available. The development time and software implementation may be substantial. For certain classes of hysteretic nonlinearities, formulations for the sensitivities of the response quantities to parameter uncertainties have been developed [25] and can be used within the Bayesian framework. However, it should be pointed out that such formulation are model intrusive and are not easily integrated to commercial computer software packages available for simulating nonlinear structural dynamics problems. For the model cases where adjoint techniques can be applied, the development time may be substantial. However, for a number of important nonlinear class of models (e.g. impact, hysteretic) or output quantities of interests (e.g. frequency response spectra), adjoint methods are not applicable. The absence of adjoint formulation may substantially increase the

computational cost and/or render gradient-based optimization algorithms unreliable for use with Bayesian asymptotic approximation tools. Stochastic optimization and stochastic simulations algorithms within a HPC environment are respectively the preferred algorithms to be used with Bayesian asymptotic and stochastic simulation tools.

At the algorithmic level, surrogate estimates are also applicable. For the case where the measurements are given as full response time histories, the surrogate estimates are applied to approximate the value of the log posterior PDF. For the case where the measurements consist of nonlinear frequency response spectra, it is more convenient computationally to apply the surrogate estimates for each spectral ordinate of the spectrum [52]. In addition, in the latter case, it should be pointed out that the frequency response spectral values can run in parallel, taking advantage of HPC environments to speed up computations.

8 CONCLUSIONS

Asymptotic approximations and stochastic simulation algorithms used in Bayesian tools for model uncertainty quantification and calibration, model selection and propagation requires a moderate to large number of finite element model simulation runs. For large order finite element models with hundred of thousands or even million DOFs and localized nonlinearities encountered in structural dynamics, the computational demands involved may be excessive, especially when a model simulation takes several minutes, hours or even days to complete. Drastic reductions in the time to solution are achieved by integrating model reduction techniques to substantially reduce the` order of high fidelity large order finite element models, surrogate models to reduce the number of full model simulations within certain classes of stochastic simulation algorithms such as TMCMC, and parallelization techniques to efficiently distribute the computations in available multi-core CPUs.

In the case where parallel computing facilities are not available, adjoint techniques integrated within Bayesian asymptotic tools provide an alternative feasible solution for large order finite element models that can reduce considerably the computational effort in iterative gradient-based optimization schemes. However, adjoint techniques are model intrusive and not applicable for certain class of models, while gradient-based optimization algorithms for Bayesian asymptotic analyses have certain disadvantages, ranging from inability of adequately exploring the parameter space, finding the global optimum, representing multimodal posterior distributions, treating challenging supports and unidentifiable cases. In parallel computing environments, stochastic optimization algorithms provide computationally feasible solution strategies for Bayesian asymptotic analyses. They do no require adjoint techniques and thus are model non-intrusive and applicable to any type of linear and nonlinear models. Also, they can explore better the parameter space with higher chances of finding the global optimum.

Parallelization is critical in Bayesian stochastic simulation tools used for uncertainty quantification and propagation of large-order finite element models. The time to solution for certain classes of parallelizable MCMC such as TMCMC is not prohibitively excessive and often is comparable to a certain degree to the time of solution observed in Bayesian asymptotic tools. However, only certain types of MCMC techniques, such as the TMCMC used in this study, are parallelizable. In order to manage uncertainties in modeling and predictions using high-fidelity large-order nonlinear finite element models involving millions of DOFs, future research efforts should concentrate in further developing highly parallelizable stochastic simulation algorithms to interface with general-purpose commercial computational structural dynamics software.

ACKNOWLEDGMENTS

This research has been implemented under the “ARISTEIA” Action of the “Operational Programme Education and Lifelong Learning” and was co-funded by the European Social Fund (ESF) and Greek National Resources. The author would like to dedicate this paper to the memory of Professor Gerhart I. Schueller. The developments presented in this work for the Bayesian computational tools, model reduction techniques, surrogate modeling and parallelization were inspired by a number of contributions published over the years in the area of reliability of dynamical systems under the leadership of Professor Gerhart Schueller.

REFERENCES

[1] C. Papadimitriou, J.L. Beck, and L.S. Katafygiotis, *Updating robust reliability using structural test data.* Probabilistic Engineering Mechanics, 2001. 16(2): p. 103-113.

[2] J.L. Beck, and S.K. Au, *Bayesian updating of structural models and reliability using Markov chain Monte Carlo simulation.* Journal of Engineering Mechanics-ASCE, 2002. 128(4): p. 380-391.

[3] J.L. Beck, and L.S. Katafygiotis, *Updating models and their uncertainties. I: Bayesian statistical framework.* Journal of Engineering Mechanics-ASCE, 1998. 124(4): p. 455-461.

[4] N. Metropolis, et al., *Equation of State Calculations by Fast Computing Machines.* J. Chem. Phys., 1953. 21(6): p. 1087--1092.

[5] W.K. Hastings, *Monte-Carlo Sampling Methods Using Markov Chains and Their Applications.* Biometrika, 1970. 57(1): p. 97-&.

[6] S.H. Cheung, and J.L. Beck, *Bayesian Model Updating Using Hybrid Monte Carlo Simulation with Application to Structural Dynamic Models with Many Uncertain Parameters.* Journal of Engineering Mechanics-ASCE, 2009. 135(4): p. 243-255.

[7] J.Y. Ching, and Y.C. Chen, *Transitional markov chain monte carlo method for Bayesian model updating, model class selection, and model averaging.* Journal of Engineering Mechanics-Asce, 2007. 133(7): p. 816-832.

[8] H. Haario et al., *DRAM: Efficient adaptive MCMC.* Stat. Comp., 2006. 16: p. 339-354.

[9] S.H. Cheung, and J.L. Beck, *Calculation of Posterior Probabilities for Bayesian Model Class Assessment and Averaging from Posterior Samples Based on Dynamic System Data.* Computer-Aided Civil and Infrastructure Engineering, 2010. 25(5): p. 304-321.

[10] C. Papadimitriou, and D.C. Papadioti, *Component mode synthesis techniques for finite element model updating.* Computers & Structures, 2013. 126: p. 15-28.

[11] H.A. Jensen et al., *Model-Reduction Techniques for Bayesian Finite Element Model Updating Using Dynamic Response Data.* Computers and Structures. submitted Jan. 2014.

[12] S.N. Lophaven, H.B. Nielsen, and J. Søndergaard, *DACE,A MATLAB Kriging Toolbox*, 2002, DTU: DK-2800 Kgs. Lyngby -- Denmark.

[13] P. Angelikopoulos, C. Papadimitriou, and P. Koumoutsakos, *Bayesian uncertainty quantification and propagation in molecular dynamics simulations: A high performance computing framework.* Journal of Chemical Physics, 2012. 137(14).

[14] J.L. Beck, *Bayesian system identification based on probability logic.* Structural Control & Health Monitoring, 2010. 17(7): p. 825-847.

[15] E. Simoen, C. Papadimitriou, and G. Lombaert, *On prediction error correlation in Bayesian model updating.* Journal of Sound and Vibration, 2013. 332(18): p. 4136-4152.

[16] K.V. Yuen, *Bayesian Methods for Structural Dynamics and Civil Engineering*. 2010, Singapore: John Wiley & Sons.
[17] M. Muto, and J.L. Beck, *Bayesian updating and model class selection for hysteretic structural models using stochastic simulation.* Journal of Vibration and Control, 2008. 14(1-2): p. 7-34.
[18] E. Ntotsios et al., *Bridge health monitoring system based on vibration measurements.* Bulletin of Earthquake Engineering, 2009. 7(2): p. 469-483.
[19] L.S. Katafygiotis, and J.L. Beck, *Updating models and their uncertainties. II: Model identifiability.* Journal of Engineering Mechanics-ASCE, 1998. 124(4): p. 463-467.
[20] L.S. Katafygiotis, and H.F. Lam, *Tangential-projection algorithm for manifold representation in unidentifiable model updating problems.* Earthquake Engineering & Structural Dynamics, 2002. 31(4): p. 791-812.
[21] J.L. Beck, and K.V. Yuen, *Model selection using response measurements: Bayesian probabilistic approach.* Journal of Engineering Mechanics-ASCE, 2004. 130(2): p. 192-203.
[22] C. Papadimitriou, and L.S. Katafygiotis, *A Bayesian Methodology for Structural Integrity and Reliability Assessment.* International Journal of Advanced Manufacturing Systems, 2001. 4(1): p. 93-100.
[23] L. Tierney, and J.B. Kadane, *Accurate Approximations for Posterior Moments and Marginal Densities.* Journal of the American Statistical Association, 1986. 81(393): p. 82-86.
[24] E. Ntotsios, and C. Papadimitriou, *Multi-objective optimization algorithms for finite element model updating.* Proceedings of Isma 2008: International Conference on Noise and Vibration Engineering, Vols. 1-8, 2008: p. 1895-1909.
[25] M. Barbato, A. Zona, and J.P. Conte, *Finite element response sensitivity analysis using three-field mixed formulation: General theory and application to frame structures.* International Journal for Numerical Methods in Engineering, 2007. 69(1): p. 114-161.
[26] M. Barbato, and J.P. Conte, *Finite element response sensitivity analysis: a comparison between force-based and displacement-based frame element models.* Computer Methods in Applied Mechanics and Engineering, 2005. 194(12-16): p. 1479-1512.
[27] D.I. Papadimitriou, and C. Papadimitriou. *Bayesian Estimation of Turbulence Model Parameters Using High-Order Sensitivity Analysis.* in *FEMTEC 2013 Conference*. 2013. Las Vegas, USA.
[28] N. Hansen, S.D. Muller, and P. Koumoutsakos, *Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES).* Evolutionary Computation, 2003. 11(1): p. 1-18.
[29] J.N. Lyness, and C.B. Moler, *Generalized Romberg Methods for Integrals of Derivatives.* Numerische Mathematik, 1969. 14(1): p. 1-&.
[30] R.R. Craig Jr., and M.C.C. Bampton, *Coupling of substructures for dynamic analysis.* AIAA Journal 1968. 6(7): p. 1313-1319.
[31] W. Hurty, *Dynamic analysis of structural systems using component modes.* AIAA Journal, 1965. 3(4): p. 678–685.
[32] R.R. Craig Jr., *Structural Dynamics - An Introduction to Computer Methods*. 1981, New York: John Wiley & Sons.
[33] E. Balmes, *Parametric families of reduced finite element models. Theory and applications.* Mechanical Systems and Signal Processing, 1996. 10(4): p. 381-394.
[34] H.J. Pradlwarter, G.I. Schueller, and G.S. Szekely, *Random eigenvalue problems for large systems.* Computers & Structures, 2002. 80(27-30): p. 2415-2424.
[35] L. Hinke et al., *Component mode synthesis as a framework for uncertainty analysis.* Journal of Sound and Vibration, 2009. 324(1-2): p. 161-178.
[36] B.R. Mace, and P.J. Shorter, *A local modal/perturbational method for estimating frequency response statistics of built-up structures with uncertain properties.* Journal of Sound and Vibration, 2001. 242(5): p. 793-811.
[37] B. Goller, H.J. Pradlwarter, and G.I. Schueller, *An interpolation scheme for the approximation of dynamical systems.* Computer Methods in Applied Mechanics and Engineering, 2011. 200(1-4): p. 414-423.
[38] B. Goller et al., *A stochastic model updating technique for complex aerospace structures.* Finite Elements in Analysis and Design, 2011. 47(7): p. 739-752.
[39] S.K. Hong et al., *Parametric reduced-order models for predicting the vibration response of complex structures with component damage and uncertainties.* Journal of Sound and Vibration, 2011. 330(6): p. 1091-1110.
[40] M.P. Castanier, Y.C. Tan, and C. Pierre, *Characteristic constraint modes for component mode synthesis.* Aiaa Journal, 2001. 39(6): p. 1182-1187.
[41] D.-C.Papadioti, and C. Papadimitriou. *Fast Bayesian Structural Damage Localization and Quantification using High Fidelity FE Models and CMS Techniques*. in *3rd International Symposium on Life-Cycle Civil Engineering*. 2012. Vienna, Austria.
[42] M.W. Vanik, J.L. Beck, and S.K. Au, *Bayesian probabilistic approach to structural health monitoring.* Journal of Engineering Mechanics-Asce, 2000. 126(7): p. 738-745.
[43] K.V. Yuen, J.L. Beck, and L.S. Katafygiotis, *Efficient model updating and health monitoring methodology using incomplete modal data without mode matching.* Structural Control & Health Monitoring, 2006. 13(1): p. 91-107.
[44] K. Christodoulou, and C. Papadimitriou, *Structural identification based on optimally weighted modal residuals.* Mechanical Systems and Signal Processing, 2007. 21(1): p. 4-23.
[45] E. Simoen et al., *Uncertainty Quantification in the Assessment of Progressive Damage in a 7-Story Full-Scale Building Slice.* Journal of Engineering Mechanics, 2013. 139(12): p. 1818-1830.
[46] B. Goller, J.L. Beck, and G.I. Schueller, *Evidence-Based Identification of Weighting Factors in Bayesian Model Updating Using Modal Data.* Journal of Engineering Mechanics-ASCE, 2012. 138(5): p. 430-440.
[47] R.B. Nelson, *Simplified Calculation of Eigenvector Derivatives.* AIAA Journal, 1976. 14(9): p. 1201-1205.
[48] R.L. Fox, and M.P. Kapoor, *Rate of change of eigenvalues and eigenvectors.* AIAA Journal, 1968. 6(12): p. 2426-2429.
[49] P. Metallidis et al., *Fault detection and optimal sensor location in vehicle suspensions.* Journal of Vibration and Control, 2003. 9(3-4): p. 337-359.
[50] P. Metallidis, I. Stavrakis, and S. Natsiavas, *Parametric identification and health monitoring of complex ground vehicle models.* Journal of Vibration and Control, 2008. 14(7): p. 1021-1036.
[51] H.A. Jensen et al., *The use of updated robust reliability measures in stochastic dynamical systems.* Computer Methods in Applied Mechanics and Engineering, 2013. 267: p. 293-317.
[52] S. Natsiavas et al. *Bayesian Uncertainty Quantification and Propagation in Nonlinear Structural Dynamics.* in *International Modal Analysis Conference (IMAC)*. 2013. California.