Uncertainty Quantification in Vibration-based Structural Health Monitoring using Bayesian Statistics

by

Binbin Li

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Civil and Environmental Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Armen Der Kiureghian, Chair
Professor Steven D. Glaser
Professor David R. Brillinger

Fall 2016

Uncertainty Quantification in Vibration-based Structural Health Monitoring using Bayesian Statistics

Abstract

Uncertainty Quantification in Vibration-based Structural Health Monitoring using Bayesian Statistics

by

Binbin Li

Doctor of Philosophy in Engineering – Civil and Environmental Engineering

University of California, Berkeley

Professor Armen Der Kiureghian, Chair

Although great advancements have been made in structural health monitoring (SHM) for civil structures since 1990s, a lack of accurate and reliable techniques to interpret measured data still challenges the whole community. A common consensus is that raw data cannot directly tell the damage, only features-extracted data can. One major difficulty is that there is too much noise in the SHM data such that the "signal," even if it conveys the damage information, is buried in the noise. Under the particular circumstance of low signal-to-noise ratio, uncertainty quantification is an invaluable step to determine the influence of uncertainties on predicted values. Armed with Bayesian statistics, this dissertation is devoted to the uncertainty quantification in vibration-based SHM.

A robust optimal sensor placement (OSP) for operational modal analysis is proposed based on the maximum expected utility theory. First, a probabilistic model for OSP considering model uncertainty, load uncertainty and measurement error is introduced, which turns out to be a linear Bayesian normal model. The maximum expected utility theory is then applied with this model by considering utility functions based on three principles: Shannon information, quadratic loss, and K-L divergence. The covariance of modal responses is theoretically derived, and its nearest Kronecker product approximation is developed for fast evaluation of the utility functions. As demonstration and validation examples, sensor placements in a 16-degrees-of-freedom shear-type building, and in Guangzhou (China) TV Tower excited by ground motion and wind load are considered. The results show that, when accounting for prior information, the optimal placement configuration of displacement meter, velocimeter and accelerometer do not have to be congruent, and mixed sensor placement becomes possible. Prior information has less influence on accelerometer

placement than on the other sensors, justifying the commonly used mode-shape-based accelerometer placement. The magnitude of input to noise ratio has a great influence on the optimal configuration of sensors, and it connects the kinetic energy-based and Fisher information-based sensor placement approaches.

The uncertainty quantification in the operational modal analysis (OMA) is investigated, where the structural excitations are not directly measured but modeled by band-limited white noise processes. We start with the state-space representation of the dynamical system. By assigning probability distributions to the error terms and specifying prior distributions for the unknown parameters, a probabilistic model, belonging to the conjugate-exponential model, is formally constructed for OMA. The expectation-maximization and the variational Bayes algorithms and the Gibbs sampler are employed to infer the modal parameters from the measured structural responses. For the purpose of restraining the accumulated numerical error in the forward-backward inference, a robust implementation strategy is developed based on square-root filtering and Cholesky decomposition. The proposed approaches are illustrated by application to an example mass-spring system, a laboratory shear-type building model, and the One Rincon Hill Tower in San Francisco. It is observed that the modal frequencies and mode shapes can be identified with small uncertainties comparing to those of identified damping ratios. In addition, the coefficient of variation of the estimated frequency is approximately equal to the standard deviation of the estimated damping ratio in the same mode.

The last problem we consider is the uncertainty quantification in finite element model updating (FEMU) using the measured incomplete and noisy modal data. Based on the generalized eigenvalue decomposition of the stiffness and mass matrices and the assumptions on the error models, a Bayesian probabilistic model for FEMU is formulated, which can incorporate the time-variability, measurement error and model parameter uncertainty. In order to obtain the posterior distributions of the stiffness parameters, a Metropolis-within-Gibbs (MwG) sampler is introduced and a robust implementation strategy is provided as well. The performance of the proposed Bayesian method is illustrated through two examples: a numerical 8-DoF mass-spring system and an experimental 6-story shear-type building. The examples show that the designed MwG sampler accurately recovers the posterior distributions of the stiffness parameters. The posterior variance highly depends on the number of data sets, and correlations between the stiffness parameters represent their physical dependence. It is recommended to use a sufficiently complex model so as to fully explain the measured modal data and include as many modes as possible in estimation to get a more representative model.

Contents

List of Figures

List of Tables

Acknowledgments

For 23 years in studying, it has been my dream to earn the Ph.D. degree. Through hard work, pain and sacrifices, this has been my target. And now I have made it. At this moment, my strongest emotion is deep gratitude to everyone who supported me to make this dream come true.

I would like to express my sincerest gratitude to my research advisor Professor Armen Der Kiureghian for his continuous support of my Ph.D. study, for his patience, motivation, and immense knowledge. While his encouragement gave me the great freedom to explore new research ideas, his thoughtful insight has always refined and enriched them. Being a mentor, a professor, and a researcher, he has set an ideal for me to follow during my whole career. Besides my advisor, I would like to thank the rest of my dissertation committee: Professor Steven D. Glaser and Professor David R. Brillinger, for their insightful comments and encouragement.

Next, I want to show my deep appreciation to the UC Berkeley community. It has been a great honor to study in such a great place, immersed in the world-leading teaching and research environment. The four years spent here has totally shaped my personality and will always motivate me in my future life. I should thank my friends at Berkeley (in alphabetical order): Guoqing, Iris, James, Marco, Mayssa, Qingsong, Qingkai, Shanshan, Xin, Zhishuai and Zhilong, who made my school life full of happiness. In the same vein, sincere thanks also go to my friends at American University of Armenia, where I stayed for ten months, especially to Dr. Armen Mkrtchyan and Professor Alex Thomasian.

Finally, I owe the deepest debt of gratitude to my family and my parents for their unconditional love and endless support. There are both tears and joys during this period: my father unfortunately passed away and my daughter was born. My warmest thanks go to my parents for their love and their sacrifices in raising and nurturing me. Special appreciation and affection go to my wife, Cui. The dissertation writing process covered the pregnancy period of my wife and I am in debt to her and our daughter Cecilia for not being able to take perfect care of them during this important period of life.

Chapter 1
Introduction

1.1 Motivation

Civil infrastructures are the foundation that connects the nation's businesses, communities, and people, driving our economy and sustaining and improving our quality of life. Due to natural aging and natural/man-made hazards, infrastructure systems deteriorate with time and accumulate damage throughout their service life. Unless these systems are properly maintained the deterioration may result in reduced safety and service quality and increased operational costs, and may lead to catastrophic failures with devastating environmental, social and economic consequences.

The American Society of Civil Engineers (ASCE) [1] rated the condition of America's infrastructure as near failing D+ in 2013. Although, the bridge system earned a grade of C+, one in nine bridges were rated as structurally deficient, meaning the bridge has a significant defect that requires reduced weight or speed limits. It also estimated that $3.6 trillion is needed over a five-year period to bring the nation's infrastructure to a good condition. The need to upgrade the nation's aging and deteriorating civil infrastructure with constrained budgets poses an enormous challenge. Meanwhile, with increased awareness about the extent of deficiencies of existing infrastructures, the US National Academy of Engineering has identified "Restore and Improve Urban Infrastructure" as one of the 14 grand challenges for engineering in the 21st century [2].

It is essential to retrofit and reconstruct existing deteriorating infrastructure systems, but considering the budget constraint, a more rational solution is to seek maintenance and inspection alternatives [3] that minimize economic and social costs while maximizing the operational life span. Current inspection practices rely heavily on expert knowledge through visual inspection [4]. The possible drawbacks are: (1) The expert knowledge is subjective so that different engineers may yield vastly different judgments on the same structure; (2) Invisible damages, like internal cracks and rebar corrosion in concrete, cannot be detected; (3) The inspection process is time-consuming, and the continuous inspection is impossible. As for the maintenance plan, a schedule-based maintenance strategy is extensively used today [4]. This approach requires that critical components are serviced or replaced at predefined times, regardless of the true condition of the structure. Apparently, neither the

inspection nor the maintenance strategy is satisfactory. With the development of sensor and communication technologies, structural health monitoring (SHM) provides a new and improved way for real-time inspection and a condition-based maintenance of civil infrastructures.

SHM refers to the process of equipping a structure with sensors, then extracting useful features from measurements for the purpose of diagnosis, prognosis and structural health management. For long-term SHM, the output of this process is periodically updated to quantify the capability of the structure to perform its intended function despite inevitable ageing and degradation resulting from operations and environmental exposure [5]. After a disastrous event such as an earthquake, blast, or hurricane, SHM could be used for rapid condition screening, to provide, in near-real time, information about the performance of the structure during the event and about the subsequent integrity, and to prescribe appropriate rehabilitation schemes to increase the resilience of the structure.

The main task of SHM is to diagnose structural damage, which is defined as intentional or unintentional changes to the material and/or geometric properties of the structural system, including changes to the boundary conditions and system connectivity, which adversely affect the current or future performance of the system [6]. The damage diagnosis answers the questions about the existence, location, type and extent of damage [7]. When implemented appropriately, SHM is able to detect the onset of damage, even invisible, at the earliest time, thus, reducing the necessity for redundancies and system down-time due to debilitating damage.

Compared with damage diagnosis, prognosis is a more difficult problem. It aims at predicting the useful life remaining in a structural system that has experienced damage. The existence of damage does not mean the structure is unsafe. Unlike aerospace and mechanical structures, civil structures are more redundant and thus more robust to damage. An accurate physical model is generally required in order to quantify the effect of the damage on the integrity of the whole structure. A successful damage prognosis requires measurements of the current condition and prediction of the possible deterioration, when subjected to future loading, which involves uncertainties. An alternative goal might be to estimate how long the structure can continue to safely perform in its anticipated environment [8].

Structural health management can be defined as the process of making appropriate decisions about operation and maintenance actions based on the damage diagnosis and prognosis. SHM allows the current time-based maintenance approaches to evolve into condition-based maintenance philosophies, which dictates that the maintenance should only be performed when certain indicators show signs of decreasing performance or upcoming failure. Through damage diagnosis and prognosis, SHM can provide a prompt warning such that corrective actions can be taken before the damage or degradation evolves into critical level. Through this

process, maintenance and repair costs decrease, while life-safety increases.

1.2 Objective and scope

SHM has been extensively developed since 1990s and many systems have been implemented for buildings [9,10] and bridges [11,12,13] all over the world. Although great achievements have been made, a lack of accurate and reliable techniques to interpret measured data still challenges the whole community. A common consensus is that raw data cannot directly tell the damage, only features-extracted data can. One major difficulty is that there is too much noise in the SHM data of civil structures. First, civil structures themselves are subject to intrinsic variability: material properties and construction processes may have large deviations from the initial design. Second, external loads are largely uncertain: the wind load, ground motion, traffic and environmental effects (temperature, humidity, etc.) are generally uncontrollable and can only be represented in statistical terms. Third, the measurement error of sensory system is ubiquitous, including the sensor error, the data transmission error, the transducer error and the computer error. Fourth, model error is unavoidable for both physical and data-driven models. For example, linear time-invariant behavior is usually assumed in SHM, but some level of nonlinearity and time-variance is often present, even under normal operational conditions. Fifth, other errors, such as parameter and statistical uncertainty and human error [14] are inevitably present. The accumulated effect of all these errors is a poor signal-to-noise ratio (SNR): the "signal," even if it conveys the damage information, is buried in the noise [15].

It is a difficult task to successfully apply SHM to civil structures considering the noise condition. Thorough planning, a systematic methodology, and a careful implementation are crucial for an accurate and reliable SHM application. Under the particular circumstance of low SNR, uncertainty quantification is an invaluable step to minimize the influence of uncertainties. Armed with Bayesian statistics, this dissertation is devoted to the uncertainty quantification in vibration-based SHM, which is aimed at inferring the damage state of a structure using measurements of the dynamic response to environmental loads.

Although many other methods exist for uncertainty quantification, including evidence theory [16], interval analysis [17] or fuzzy sets [18], the Bayesian statistics [19] is a perfect choice for SHM of civil structures. First, Bayesian statistics makes it possible to absorb the prior information, like an expert's knowledge, into the statistical analysis by means of the prior distribution. Unlike mass-produced aerospace and mechanical systems, civil structures are uniquely designed and constructed. Furthermore, failure of civil structures is a rare event, rendering past experience in modeling the structure and analysis of the damage important. Second,

all variables, including unknown parameters, are treated as random variables in the Bayesian paradigm, thus yielding a unified framework to deal with all kinds of uncertainties, especially parameter uncertainty. Third, Bayesian statistics takes advantage of the conditional distribution to model the dependencies among variables, equipping the method with the capability to handle complex models. Fourth, it is natural to incorporate decision theory within the Bayesian framework, thus providing a solid mathematical background for health management. In this dissertation, we apply the Bayesian statistical approach to provide the foundation for an in-depth understanding and further development of vibration-based SHM.

The remainder of this chapter reviews some key aspects in vibration-based SHM and Bayesian statistics. Section 1.3 reviews the use of dynamic analysis in SHM, mainly focusing on optimal sensor placement, operational modal analysis and finite element model updating. In section 1.4, we make a short introduction to Bayesian statistics, especially to posterior approximation methods, including the Laplace method, variational Bayes and the Markov chain Monte Carlo sampling. Important terminologies used in Bayesian inference are listed in Appendix A. Finally, Section 1.5 briefly summarizes the remaining chapters of this dissertation.

1.3 Vibration-based SHM

Vibration-based SHM focuses on the dynamic response of the structure. The basic premise is that damage can significantly alter the stiffness, mass or energy dissipation properties of the structure, which, in turn, alter the measured dynamic responses [20]. Therefore, by identifying the structural properties from the measured data, which in essence is an inverse problem, the potential damage can be detected, localized or even quantified.

One appealing feature of vibration-based SHM over other SHM techniques is the global nature of the dynamic characteristics. Utilization of the global signatures such as natural frequencies and mode shapes leads to the monitoring of the entire structural system, not just one structural component. This means a large civil structure can be effectively monitored with a relatively small set of sensors [21].

The most widely used model assumption in vibration-based SHM is that the structure is linear and time-invariant (LTI). The equation of motion of such a system under a stochastic external force and a base motion is represented as

$$\boldsymbol{M}\ddot{\boldsymbol{u}}(t) + \boldsymbol{C}_d\dot{\boldsymbol{u}}(t) + \boldsymbol{K}\boldsymbol{u}(t) = \boldsymbol{P}\boldsymbol{f}(t) - \boldsymbol{M}\boldsymbol{l}\ddot{\boldsymbol{u}}_g(t) \tag{1.1}$$

where $\boldsymbol{M} \in \mathbb{R}^{N_d \times N_d}$, $\boldsymbol{C}_d \in \mathbb{R}^{N_d \times N_d}$ and $\boldsymbol{K} \in \mathbb{R}^{N_d \times N_d}$ are the mass, damping and stiffness matrices, respectively, where N_d is the number of degrees of freedom of the

system; $\boldsymbol{u}(t) \in \mathbb{R}^{N_d}$, $\dot{\boldsymbol{u}}(t) \in \mathbb{R}^{N_d}$ and $\ddot{\boldsymbol{u}}(t) \in \mathbb{R}^{N_d}$ are the nodal displacement, velocity and acceleration responses relative to the ground, respectively, with $\boldsymbol{u}(0) = \boldsymbol{u}_0$ and $\dot{\boldsymbol{u}}(0) = \dot{\boldsymbol{u}}_0$ being the initial relative displacement and velocity vectors; $\boldsymbol{f}(t) \in \mathbb{R}^{N_f}$ is the external force vector; $\boldsymbol{P} \in \mathbb{R}^{N_d \times N_f}$ is the load coefficient matrix representing the spatial influence of the external force; $\ddot{\boldsymbol{u}}_g(t) \in \mathbb{R}^{N_g}$ is the ground acceleration vector, and $\boldsymbol{l} \in \mathbb{R}^{N_d \times N_g}$ is the corresponding influence matrix.

If classical damping [22] is further assumed, the equations of motion can be decoupled by introducing the transformation $\boldsymbol{u}(t) = \boldsymbol{\Phi}\boldsymbol{q}(t)$ to yield the modal form

$$\begin{gathered} \ddot{q}_i(t) + 2\xi_i\omega_i\dot{q}_i(t) + \omega_i^2 q_i(t) = \boldsymbol{\gamma}_i^e \boldsymbol{f}(t) + \boldsymbol{\gamma}_i^g \ddot{\boldsymbol{u}}_g(t) \\ q_i(0) = \boldsymbol{\phi}_i^{\mathrm{T}} \boldsymbol{M}\boldsymbol{u}_0/M_i,\ \dot{q}_i(0) = \boldsymbol{\phi}_i^{\mathrm{T}} \boldsymbol{M}\dot{\boldsymbol{u}}_0/M_i \end{gathered} \tag{1.2}$$

for $i = 1, \dots, N_d$, where

$$\begin{gathered} \boldsymbol{q}(t) = \left[q_1(t), q_2(t), \dots, q_{N_d}(t)\right]^{\mathrm{T}} \in \mathbb{R}^{N_d} \\ \boldsymbol{\Phi} = \left[\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \dots, \boldsymbol{\phi}_{N_d}\right] \in \mathbb{R}^{N_d \times N_d} \end{gathered} \tag{1.3}$$

are, respectively, the modal displacement vector and the mode shape matrix. In addition,

$$\begin{gathered} M_i = \boldsymbol{\phi}_i^{\mathrm{T}} \boldsymbol{M} \boldsymbol{\phi}_i \\ \omega_i = \sqrt{\boldsymbol{\phi}_i^{\mathrm{T}} \boldsymbol{K} \boldsymbol{\phi}_i / M_i} \\ \xi_i = \boldsymbol{\phi}_i^{\mathrm{T}} \boldsymbol{C} \boldsymbol{\phi}_i / (2M_i\omega_i) \\ \boldsymbol{\gamma}_i^e = \boldsymbol{\phi}_i^{\mathrm{T}} \boldsymbol{P} / M_i \\ \boldsymbol{\gamma}_i^g = \boldsymbol{\phi}_i^{\mathrm{T}} \boldsymbol{M} \boldsymbol{l} / M_i \end{gathered} \tag{1.4}$$

are the ith modal mass, natural frequency, damping ratio and participation factors for external force and base motion, respectively. Moreover, it is easy to show that ω_i and $\boldsymbol{\phi}_i$ satisfy the following eigen-equation:

$$\boldsymbol{K}\boldsymbol{\phi}_i = \omega_i^2 \boldsymbol{M}\boldsymbol{\phi}_i \tag{1.5}$$

Based on the above three equations, numerous algorithms have been proposed to identify the potential damage, such as the changes of natural frequency and mode shapes, mode shape curvature, changes in stiffness matrix, etc. The reader may refer to review articles [15,23,24] for further information. In this dissertation, we focus on more fundamental problems of how to optimally place sensors for modal

identification, and then how to identify the modal parameters and, subsequently, the model parameters. To start, a short literature review is conducted in the following three subsections.

1.3.1 Optimal sensor placement

In order to identify modal parameters (natural frequencies, damping ratios and mode shapes) from measured structural responses, a sensor system needs to be designed in advance. Although with the advent of sensor technology the cost of sensors is rapidly decreasing, the affordable number of sensors is still an issue in long-span bridges and super-tall buildings. When a limited number of sensors are used, inevitably uncertainties arise in the identified parameters. Many researchers [25,26] have shown that optimal sensor placement (OSP) is of crucial importance in order to improve the accuracy and precision of modal parameter identification.

OSP can be formally defined as an optimization problem consisting of performance criteria and computational issues. Among various formulations of performance criteria, information-theory based approaches have attracted most attention. Kammer [25] developed the effective independence (EI) method for modal identification, which tends to maximize the determinant of the Fisher information matrix (FIM). Udwadia [26] proposed the Fisher information criterion for OSP in parameter identification, in which the optimal configuration corresponds to that maximizing the trace of the FIM. Fisher information was also introduced by Borguet & Léonard [27] in the field of engine health monitoring, where the weighted sum of the condition number, trace and determinant of the FIM was selected as the performance criterion. Papadimitriou [28] introduced the concept of information entropy for the purpose of minimizing the uncertainty in the model parameter estimation; the effect of prediction error correlation was further examined recently [29]. In addition, Trendafilova, et al. [30] employed mutual entropy to select sensor locations to produce independent measurements. Li & Ou [31] derived the expected Kullback-Leibler divergence criterion to deploy sensors, which is identical to the weighted sum of the determinant and trace of the FIM. A comprehensive discussion and comparison of information theory-based OSP can be found in [32]. Aside from these information-theory based approaches, many other methods are proposed based on modal kinetic energy [33,34], model reduction [35] and observability [36]. The literature on performance criteria for OSP is so extensive that we cannot list all references. The interested reader may refer to [37] for additional references.

Structural responses are distinct under different input loads; thus, the optimal sensor configuration generally depends on the load case. Li [38] proposed a load-dependent sensor placement method considering both the load and structural response, and showed improved identification performance. Brehm [39] determined

the optimal locations of a reference sensor under white-noise excitation and multiple impulse excitations. The nominal model of the structure is invariably biased, so that a nominal model-based approach cannot provide a robust design. Vinot [40] introduced a test planning procedure based on info-gap decision theory to optimize the worst possible performance for all realizations of model parameters. Castro-Triguero [41] examined the influence of model parameters on OSP using Monte Carlo simulation. The Bayesian approach has also been employed to tackle these problems. Heredia-Zavoni [42] used the expected Bayesian loss function to deploy sensors for parameter identification under seismic load considering uncertainties in stiffness. Yuan et al. [43] proposed an information entropy-based OSP methodology for modal identification considering an uncertain excitation, wherein the model parameter uncertainties were investigated by Monte Carlo sampling. Flynn [44] placed sensors in order to minimize the expected total presence of either type I or type II errors during the damage detection process using guided ultrasonic waves.

Even though great progress has been made in optimal sensor placement, many aspects of the problem remain unresolved. For example, mode-shape-based sensor placement approaches cannot incorporate prior information about natural frequencies and damping ratios. Secondly, in the formulation of FIM, independence of measured responses is assumed. This assumption obviously is not appropriate because different responses of a structure to the same excitation are naturally correlated. Thirdly, an OSP design based on a nominal model of the structure can be over-optimistic or even misleading in certain cases. Fourthly, most methods consider deployment of displacement meter, velocimeter and accelerometer to have the same effect for modal identification, but intuition suggests that this may not be true. Furthermore, although Bayesian approaches have been used [42-44], they have only employed diffuse priors, which essentially degenerates the problem into the traditional FIM-based approach.

1.3.2 Operational modal analysis

Modal parameters, i.e. modal frequencies, damping ratios and mode shapes, are characteristic properties of linear structural models. These modal parameters and their derivations reflect the structure mass, stiffness and damping properties that depend on the condition of the structure. Changes in modal properties may indicate changes in the structure leading to the possibility that such changes can be used to detect, possibly locate, and even quantify the potential damage.

A controlled dynamic test usually yields accurate estimates of the modal parameters, but it is seldom applicable to civil structures due to their large size, high load capacity and noisy operational conditions. Therefore, operational modal analysis (OMA), which utilizes stochastic dynamic response, has become the primary modal testing method in civil engineering [45]. The basic assumption in OMA is that the

sources of excitation are broad-band stochastic processes adequately modeled by band-limited white noise. Since the main sources of excitation in civil structures are wind, traffic, ground tremor and low-magnitude earthquakes, this assumption is generally satisfied.

Many identification methods have been proposed for OMA, but most provide only point estimates. These include NExT-type methods [46], stochastic subspace identification (SSI) [47], the prediction error method [48], FDD-type methods [49] and output-only LSCF-type methods [50].

Since the SNR cannot be directly controlled and various uncertainties are present in different stages of OMA, more recent developments in OMA have addressed the statistical properties of the OMA-derived modal parameters. The impact of uncertainties on operational modal identification has been studied in Ref. [51], where it is shown that both aleatory and epistemic uncertainties have significant influences on OMA. To quantify the associated uncertainties in OMA, many researchers have improved some of the deterministic algorithms. Lam & Mevel [52] derived the confidence interval for the Eigensystem-Realization-Algorithm based on perturbation analysis; Reynders et al. [53] and Dohler et al. [54] developed confidence intervals for subspace identification; and Vu & Thomas [55] obtained variances of modal parameters from a vector autoregression model.

The statistical learning method provides a more appropriate approach to extract uncertainty information. Related approaches include maximum likelihood estimation (MLE) [56,57] and the Bayesian approach [58-61]. In MLE, the confidence interval cannot be derived directly and only an asymptotic approximation of the distribution can be derived based on the law of large numbers. In the Bayesian approach, the modal parameters are regarded as random variables, so that their posterior distributions can be determined given the measured data and modeling assumptions. Given the difficulty in directly computing the posterior distribution, both Yuen & Katafygiotis [58,59] and Au [60,61] applied the Laplace method yielding a multivariate normal distribution for modal parameters. However, as discussed in Section 1.2.1, the Laplace method may not be suitable because the natural frequencies and damping ratios must be positive for a stable structure. Most existing Bayesian formulations of OMA are confined to the frequency-domain [59-61], where the Fourier transform first implemented before the OMA is performed. In this light, it is helpful to propose new Bayesian methods in the time domain.

1.3.3 Finite element model updating

An initial finite element model (FEM) is often a poor representation of a real structure, particularly as related to in-field structural dynamic response. This is due to the presence of unavoidable errors, such as improper modeling of boundary conditions

and joints, incorrect value of material properties, and oversimplified modeling assumptions (e.g. classical damping). Measured dynamic responses generally provide a better reflection of how the structure behaves than predictions from the initial FEM. Finite element model updating (FEMU) seeks to determine the set of most plausible parameter values to reproduce, as closely as possible, the measured dynamic responses. Under the assumption that localized structural damage results in a local reduction of stiffness, the updated FEM can be used as an indication to quantify the location and extent of damage by comparing with a baseline FEM corresponding to the intact structure.

Depending on the measurements used, the approaches for FEMU can be categorized into two categories: time-domain methods and frequency-domain methods. Time-domain methods directly utilize the recorded time histories of accelerations, displacements or velocities, while frequency-domain methods employ identified modal parameters, such as natural frequencies, mode shapes and damping ratios. In this dissertation, we will focus on the latter method to adopt the so called two-stage FEMU strategy [62]. The relevant literature is briefly reviewed below.

In the early stage of FEMU study, most research pursued an optimal point estimate of the unknown parameters in the FEM to match the measured data. There are two dominant methods falling into this category: matrix modification methods [63,64] and sensitivity-based methods [65,66]. Matrix modification methods compute a closed-form solution for system matrices by solving a set of matrix equations. This method generally is computationally efficient because no iteration is needed, and it can reproduce the measured data almost exactly. The main drawbacks are that updated mass and stiffness matrices may not be positive (semi)definite and it is usually hard to interpret the updated system matrices. Sensitivity-based methods try to minimize a goodness-of-fit function, which contains the differences between the measurements and the predictions of the FEM, by adjusting a pre-selected set of physical parameters. The optimal solution is obtained in an iterative procedure by using sensitivity-based optimization methods such as the Newton's method [67]. The updated mass and stiffness matrices automatically retain the fundamental properties of the original FE model such as positivity, symmetry and sparse patterns, and an immediate physical interpretation of the updated results can be acquired.

As an inverse problem, the FEMU is prone to ill-posedness and ill-conditioning, meaning that the existence, uniqueness and stability (with respect to small errors) of a solution cannot be guaranteed [68]. Since measurement and modeling uncertainties are always present, it is important and indispensable to explicitly treat them in the FEMU. In this regard, stochastic model updating has gained an increasing attention recently. The random matrix-based method [69], the covariance matrix adjustment method [70], the perturbation method [71,72], and a fuzzy set-based method [73] are all good examples to deal with the uncertainties in the FEMU. Bayesian statistics is

another class of uncertainty quantification methods that has been widely applied in the FEMU after the pioneering work by Beck and his colleagues [74-78]. For example, Vanik et al. [74] first formulated a FEMU framework using Bayesian statistics in frequency domain and solved it via the Laplace method. In order to achieve a better approximation of the posterior, Ching et al. applied the expectation-maximization algorithm [75], the Gibbs sampler [76] and the transitional Markov chain Monte Carlo [77] in the FEMU problem. Besides, Yuen et al. [78] further modified this framework to explicitly control the model error; Yan & Katafygiotis [79] considered the scenario of multiple setups; Sun & Büyüköztürk [80] proposed a similar approach through model reduction. Although significant progress has been made, full-fledged applications of Bayesian approaches in the FEMU are still in their infancy. First, the original formulation is imperfect and cannot propagate the various uncertainties correctly. As argued by Behmanesh et al. [81], the formulation in [74] vastly underestimated the uncertainty in the updated FEM. Second, the approximation of the posterior distribution in the FEMU is still computationally demanding, especially for the Monte Carlo method.

1.4 Bayesian statistics

There are generally two interpretations of probability [82]: frequentist and Bayesian. The frequentist interprets the probability of an event as the limit of the relative frequency with which the event occurs, in repeated trials under identical conditions. Although the frequentist view is straightforward and useful, it has limitations. For example, if we want to determine the probability that the strength of a piece of material is less than a given threshold, we have to imagine many samples of the same material, even though our interest is in the specific piece of the material. Furthermore, we cannot subject the same piece of material to repeated destructive tests to obtain the frequency of the event of interest. Alternatively, the Bayesian interpretation regards the probability as a measure of the degree of belief of the individual assessing the uncertainty of a particular event on a $[0,1]$ scale. With this interpretation, the probability of strength of the piece of material being less than the given threshold becomes meaningful, though it may vary from person to person.

Based on the Bayesian approach, a Bayesian statistician assigns probability distributions to all relevant unknown quantities reflecting his/her knowledge of their values. This is the most distinct feature of the approach over all other methods. Because the unknown parameters are treated as random variables, it provides a general, coherent methodology for statistical analysis by solely working with probability distributions. For example, statistical inference becomes the task to infer

the posterior distribution of a variable, given the observed data and model assumptions. Note that the treatment of unknown parameters as random variables in Bayesian paradigm is not a description of their variability but a description of the uncertainty about their true values [83].

Let us start with a general setting in Bayesian statistics: models $\boldsymbol{m}$, their parameters $\boldsymbol{\theta}$ and measurements $\boldsymbol{y}$. Assuming a prior distribution $\pi(\boldsymbol{m})$ over the models and a prior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{m})$ over the parameters, by Bayes' rule, the posterior over models $\boldsymbol{m}$ is given by

$$p(\boldsymbol{m}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{m})\pi(\boldsymbol{m})}{f(\boldsymbol{y})} \tag{1.6}$$

The posterior $p(\boldsymbol{m}|\boldsymbol{y})$ is used in model selection; the most probable model is the one that maximizes $p(\boldsymbol{m}|\boldsymbol{y})$. For a given model, we can also derive the posterior distribution over the parameters

$$p(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{m}) = \frac{f(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{m})\pi(\boldsymbol{\theta}|\boldsymbol{m})}{f(\boldsymbol{y}|\boldsymbol{m})} \tag{1.7}$$

which allows us to quantify the uncertainty in parameters after observing the data. The first term in the numerator in the above equation, $f(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{m})$, is called the likelihood; it measures the support provided by the data $\boldsymbol{y}$ for each possible set of values of the parameters $\boldsymbol{\theta}$. The denominator $f(\boldsymbol{y}|\boldsymbol{m})$ is called the marginal likelihood or Type-II likelihood or evidence. It is an important quantity in Bayesian learning, for computing quantities such as Bayes factors (the ratio of two marginal likelihoods [84]), or the partition function in statistical mechanics [85]. From Eqn. (1.7), the marginal likelihood can be computed as

$$f(\boldsymbol{y}|\boldsymbol{m}) = \int f(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{m})\pi(\boldsymbol{\theta}|\boldsymbol{m})d\boldsymbol{\theta} \tag{1.8}$$

Generally speaking, the marginal likelihood is difficult to compute because it involves integrating over all parameters, which in many cases yields a high-dimensional integral for which most simple approximation methods fail.

We can also compute the density of new measurements $\boldsymbol{y}'$ by averaging over both the models and their parameters,

$$f(\boldsymbol{y}'|\boldsymbol{y}) = \iint f(\boldsymbol{y}'|\boldsymbol{\theta},\boldsymbol{m},\boldsymbol{y})\, p(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{m})p(\boldsymbol{m}|\boldsymbol{y})d\boldsymbol{\theta}d\boldsymbol{m} \tag{1.9}$$

This is known as the predictive distribution.

Hereafter, we ignore the model symbol $\boldsymbol{m}$ in the problem formulation, since we do not consider the problem of model selection in the dissertation.

In Bayesian statistics, the posterior of unknown parameters captures all the information inferred from the data and the prior. This distribution can then be used to make optimal predictions or decisions, or to select between models. However, for

almost all interesting applications, posterior distributions are analytically intractable, so that we have to fall back on approximation methods. In the next three subsections, three approximation methods used in this dissertation are briefly reviewed.

1.4.1 The Laplace method

The simplest approximation to the posterior distribution is to use a point estimate, such as the maximum a posteriori (MAP) parameter estimate,

$$\boldsymbol{\theta}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{\theta}} f(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{1.10}$$

which corresponds to the mode of the posterior. The MAP estimate works well in many problems, but it cannot provide any uncertainty information about the estimator. To fill this gap, the Laplace method [86] is extended by assuming a local normal approximation with the mean $\boldsymbol{\theta}_{\mathrm{MAP}}$ and the covariance matrix as the negative of the inverse of the Hessian matrix $\boldsymbol{H}(\boldsymbol{\theta}_{\mathrm{MAP}})$, which is computed by

$$\boldsymbol{H}(\boldsymbol{\theta}_{\mathrm{MAP}}) = \left.\frac{\partial^2 \ln[f(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\mathrm{MAP}}} \tag{1.11}$$

The justification for this approximation is based on the asymptotic normality of the posterior under some regularity conditions, when the number of data tends to infinity [87]. The Laplace method is simple and elegant, all one needs is the point of the maximum and the curvature at that location. One does not need to explore the whole posterior distribution since the assumed normal distribution is completely defined by the mean vector and the covariance matrix.

The Laplace method also has several shortcomings. The normal approximation represents the posterior poorly for small data sets and it is not suitable for bounded, constrained or positive parameters. Moreover, the posterior may have multiple modes, while the Laplace can only capture one of them, leading to a poor posterior approximation. Finally, the Hessian matrix may be hard to compute, especially when the number of unknown parameters is large. In addition, the Hessian tends to be singular if the parameters are not fully identifiable.

1.4.2 The Expectation-Maximization Method

The Laplace method is based on the MAP parameter estimate, which itself may be a troublesome problem, especially for the case of latent-variable models (See Appendix A). The expectation-maximization (EM) algorithm [88] provides a systematic and tractable approach for the problem of maximum likelihood estimation (MLE) or MAP in statistical models with latent variables and missing values.

Let $\boldsymbol{x}$ and $\boldsymbol{y}$ respectively denote the latent and observed random variables, and $\boldsymbol{\theta}$ be the unknown parameters. In order to get the MAP estimate of $\boldsymbol{\theta}$, we need to solve the following optimization problem:

$$\max_{\boldsymbol{\theta}} \log[f(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})] = \max_{\boldsymbol{\theta}} \log\left[\int f(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta})d\boldsymbol{x}\right] \tag{1.12}$$

Starting with an arbitrary value $\boldsymbol{\theta}^{(0)}$, the EM algorithm seeks to solve Eqn. (1.12) by iteratively applying the following two steps:

(1) Expectation step (E step): Calculate the expectation of the logarithm of the unnormalized posterior with respect to the conditional distribution of $\boldsymbol{x}$ given $\boldsymbol{y}$ and the current parameter values $\boldsymbol{\theta}^{(t)}$

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \int f\left(\boldsymbol{x}\middle|\boldsymbol{y},\boldsymbol{\theta}^{(t)}\right)\log[f(\boldsymbol{x},\boldsymbol{y}|\,\boldsymbol{\theta})\pi(\boldsymbol{\theta})]\,d\boldsymbol{x} \tag{1.13}$$

(2) Maximization step (M step): Find the parameter values that maximizes the above quantity

$$\boldsymbol{\theta}^{(t+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) \tag{1.14}$$

The preceding steps are repeatedly implemented until convergence is achieved. It has been proven that the EM algorithm monotonically increases the unnormalized posterior to a local mode with a linear convergence rate [89].

The EM algorithm takes advantage of the model structure to decouple the optimization problem into more manageable pieces. This divide-and-conquer strategy gives a conceptual clarity and simplicity of the algorithm. It is particularly useful when the likelihood is an exponential family (See Appendix A): the E step becomes the sum of expectations of sufficient statistics, and the M step involves maximizing a linear function. In such a case, it is usually possible to derive a closed-form update for each step [90].

One drawback of the EM algorithm is that it does not have a built-in procedure for producing an estimate of the covariance matrix of the parameter estimates, although some methodologies have been proposed to remove this disadvantage, such as the Supplemented EM algorithm [91]. The other problem associated with the EM algorithm is the possible slow converge even in some seemingly innocuous problems and in problems where there is too much "incomplete information" [92]. For a deeper discussion about the EM algorithm, the reader is referred to [93].

1.4.3 Variational Bayes

The variational Bayes (VB) approach [94,95] provides another way to approximate the posterior distribution of unknown parameters in the latent variable model using

optimization. The basic idea is to find a surrogate distribution from a predetermined family $\mathbb{Q}$ to minimize the distance from the true posterior distribution in the sense of the Kullback-Leibler (KL) divergence measure [96].

Let us denote $\boldsymbol{x}$ and $\boldsymbol{y}$ are respectively the latent and the observed random variables, and $\boldsymbol{\theta}$ is the unknown parameter. The VB approach uses the following surrogate distribution to approximate the posterior:

$$\begin{aligned} q^*(\boldsymbol{x},\boldsymbol{\theta}) &= \arg\min_{q(\boldsymbol{x},\boldsymbol{\theta})} \mathrm{KL}\left[q(\boldsymbol{x},\boldsymbol{\theta})||p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})\right] \\ &= \arg\min_{q(\boldsymbol{x},\boldsymbol{\theta})} \iint q(\boldsymbol{x},\boldsymbol{\theta})\, ln\frac{q(\boldsymbol{x},\boldsymbol{\theta})}{p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})} d\boldsymbol{\theta} d\boldsymbol{x} \end{aligned} \tag{1.15}$$

The VB thus turns the inference problem into an optimization problem, and the reach of the family $\mathbb{Q}$ manages the complexity of this optimization. One of the key ideas behind variational inference is to choose $\mathbb{Q}$ to be flexible enough to capture a distribution close to $p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})$, but simple enough for efficient optimization.

It is usually difficult to evaluate the KL divergence in Eqn. (1.15), so we use an alternative method to minimize the divergence without its direct evaluation. The log marginal likelihood of the observed variables $\boldsymbol{y}$ is given by

$$\begin{aligned} \ln p(\boldsymbol{y}) &= \iint q(\boldsymbol{x},\boldsymbol{\theta}) \ln p(\boldsymbol{y})\, d\boldsymbol{\theta} d\boldsymbol{x} \\ &= \iint q(\boldsymbol{x},\boldsymbol{\theta}) \ln \frac{f(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\theta})}{p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})} d\boldsymbol{\theta} d\boldsymbol{x} \\ &= \iint q(\boldsymbol{x},\boldsymbol{\theta}) \ln \frac{q(\boldsymbol{x},\boldsymbol{\theta})}{p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})} d\boldsymbol{\theta} d\boldsymbol{x} + \iint q(\boldsymbol{x},\boldsymbol{\theta}) \ln \frac{f(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\theta})}{q(\boldsymbol{x},\boldsymbol{\theta})} d\boldsymbol{\theta} d\boldsymbol{x} \\ &\triangleq \mathrm{KL}[q(\boldsymbol{x},\boldsymbol{\theta})||p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})] + \mathrm{F}[q(\boldsymbol{x},\boldsymbol{\theta})] \end{aligned} \tag{1.16}$$

where $\mathrm{F}[q(\boldsymbol{x},\boldsymbol{\theta})]$, known as "free energy"[95], is the last integral in the penultimate line in Eqn. (1.16). Since the log-marginal likelihood must remain constant for a given model, minimizing the KL divergence is equivalent to maximizing the free energy $\mathrm{F}[q(\boldsymbol{x},\boldsymbol{\theta})]$. On the other hand, since the KL divergence is non-negative [112], the free energy works as a lower bound of the log-marginal likelihood; therefore, we can regard the VB as maximizing the marginal likelihood by increasing its lower bound.

To complete specification of the VB algorithm, the distribution family $\mathbb{Q}$ should be predetermined. Here, we choose the mean-field variational family [97], where the unknown variables are divided into mutually independent groups and each governed by a distinct factor in the variational distribution, i.e.

$$q(\boldsymbol{x},\boldsymbol{\theta}) = q(\boldsymbol{x})q(\boldsymbol{\theta}) \tag{1.17}$$

Although this assumption may seem drastic, one may think of it as replacing the stochastic dependence between $\boldsymbol{\theta}$ and $\boldsymbol{x}$ by deterministic dependencies between their relevant moments [94]. The advantage of ignoring how variations in $\boldsymbol{x}$ induce variations in $\boldsymbol{\theta}$ (and vice-versa) is that we can obtain analytical approximations to their posterior distributions. If the dependence between the latent variables and parameters is low, the above approximation yields a good result. Following this assumption and assigning independent prior distributions $\pi(\boldsymbol{\theta},\boldsymbol{x}) = \pi(\boldsymbol{x})\pi(\boldsymbol{\theta})$, we have

$$\begin{aligned}
\mathrm{F}[q(\boldsymbol{\theta},\boldsymbol{x})] &\approx \mathrm{F}[q(\boldsymbol{\theta}), q(\boldsymbol{x})] \\
&= \iint q(\boldsymbol{\theta})q(\boldsymbol{x}) \ln f(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{x})\, d\boldsymbol{\theta} d\boldsymbol{x} - \int q(\boldsymbol{\theta}) \ln \frac{q(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta})} d\boldsymbol{\theta} \\
&\quad - \int q(\boldsymbol{x}) \ln \frac{q(\boldsymbol{x})}{\pi(\boldsymbol{x})} d\boldsymbol{x} \\
&= \langle \ln p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{x}) \rangle - \mathrm{KL}[q(\boldsymbol{\theta})||\pi(\boldsymbol{\theta})] - \mathrm{KL}[q(\boldsymbol{x})||\pi(\boldsymbol{x})]
\end{aligned} \tag{1.18}$$

where '$\langle\cdot\rangle$' denotes the expectation operator. Thus, in order to maximize the free energy $\mathrm{F}[q_{\boldsymbol{\theta}}(\boldsymbol{\theta}), q_{\boldsymbol{X}}(\boldsymbol{x})]$, we have to reduce the KL divergences between the surrogate and prior distributions and simultaneously increase the expected log conditional likelihood, thus reaching a balance between fitting the data and obeying the prior. The free energy is a functional of the surrogate distributions $q_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ and $q_{\boldsymbol{X}}(\boldsymbol{x})$. We can iteratively maximize it with respect to the surrogate distributions, which is essentially a coordinate ascent method in the function space. The following VB expectation-maximization (VBEM)* algorithm describes the steps in the updating formula for VB inference, which are illustrated in Figure 1.1.

(1) VB Expectation (VBE): Compute the surrogate distribution of latent variables

$$q^{(t+1)}(\boldsymbol{x}) = Z_{\boldsymbol{X}}^{-1}\pi(\boldsymbol{x}) \exp\left[\int q^{(t)}(\boldsymbol{\theta}) \ln p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{x})\, d\boldsymbol{\theta}\right] \tag{1.19}$$

(2) VB Maximization (VBM): Compute the surrogate distribution of unknown parameters

$$q^{(t+1)}(\boldsymbol{\theta}) = Z_{\boldsymbol{\theta}}^{-1}\pi(\boldsymbol{\theta}) \exp\left[\int q^{(t)}(\boldsymbol{x})\, ln\, p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{x})\, d\boldsymbol{x}\right] \tag{1.20}$$

where $Z_{\boldsymbol{X}}$ and $Z_{\boldsymbol{\theta}}$ are normalizing constants. It is noted that each step increases $\mathrm{F}[q_{\boldsymbol{\theta}}(\boldsymbol{\theta}), q_{\boldsymbol{X}}(\boldsymbol{x})]$ monotonically and the sequence converges to a local maximum.

* The algorithm is named due to its similarity to the EM algorithm.

The above theorem can be proved using variational calculus, which explains its name. A general proof of the above theorem can be found in [94].

In addition to the monotonic convergence, for the exponential family of distributions, Wang & Titterington [98] have shown that the VB-EM algorithm is statistically consistent, i.e. it converges to the true value in an asymptotic sense, provided the starting distributions are sufficiently close to the true solutions. The drawbacks of the method are: (a) it is impossible to represent multimodal posteriors because we approximate the posterior by only one approximating distribution; (b) the method is sensitive to the starting point, so that one must provide a good initial guess; and (c) the variability in the posterior distribution tends to be underestimated [99] in the mean-field variational family.

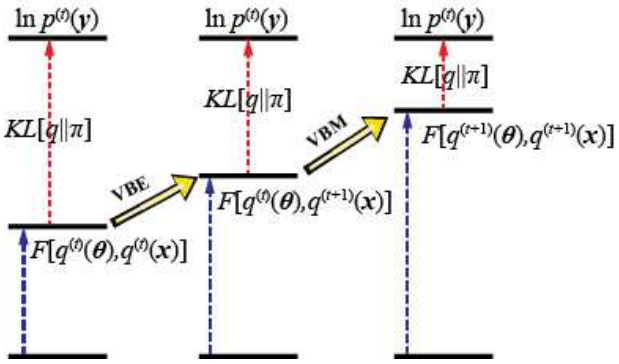


Figure 1.1 Variational Bayes EM.

1.4.4 Markov chain Monte Carlo

The Markov chain Monte Carlo (MCMC) method has been the dominant paradigm to approximate the posterior distribution in Bayesian statistics for over 50 years. First, we construct a Markov chain on $\boldsymbol{\theta}$ whose stationary distribution is the posterior $p(\boldsymbol{\theta}|\mathbf{y})$. Then, we sample from the chain for a long time to collect independent samples from the stationary distribution. Finally, we approximate the posterior with an empirical estimate constructed from the collected samples.

The MCMC method has evolved into an indispensable tool to the modern Bayesian statistician. Landmark developments include the Metropolis-Hastings (M-H) algorithm [100,101], the Gibbs sampler [102] and its application to Bayesian statistics [103]. The MCMC algorithms are under active investigation, and they have been widely studied, extended, and applied; see [104] for an extensive discussion. Here, we make a quick review of the M-H algorithm and Gibbs sampler, because they are used in this dissertation.

The M-H algorithm is given in terms of a proposal distribution $q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ conditioned on the current sample $\boldsymbol{\theta}^{(t)}$. In order to construct a Markov chain with a

stationary distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$, the M-H algorithm iterates the following steps:

(1) Let $\boldsymbol{\theta}^{(t)}$ be the current sample, and generate a candidate $\boldsymbol{\theta}^*$ using $q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$;

(2) Calculate the acceptance probability:

$$\mathcal{A}\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^*\right) = \min\left\{1, \quad \frac{p(\boldsymbol{\theta}^*|\boldsymbol{y})q(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t+1)})}{p\left(\boldsymbol{\theta}^{(t)}\middle|\boldsymbol{y}\right)q(\boldsymbol{\theta}^{(t+1)}|\boldsymbol{\theta}^{(t)})}\right\} \tag{1.21}$$

Set $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^*$ if $\mathcal{A}\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^*\right) > u$, in which u is a random number with uniform distribution over the unit interval $(0,1)$; otherwise, set $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)}$.

It is fairly easy to prove that the samples generated by the M-H algorithm mimic samples drawn from the posterior asymptotically [104], so that it constructs a valid Markov chain. The M-H algorithm can be arbitrarily initialized, because theoretically, it is insensitive to the starting point, but a good initialization, such as the mode, can accelerate its convergence. The specific choice of the proposal distribution can have a marked effect on the performance of the algorithm. A good practice is to adaptively tune the proposal to achieve an optimal acceptance rate [105].

The Gibbs sampler is another popular MCMC method. The basic idea is to generate posterior samples by sweeping through each variable (or block of variables) to sample from its conditional distribution with the remaining variables fixed to their current values. The Gibbs sampler is particularly well-adapted to latent variable models, because the conditional distributions $p(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{y})$ and $p(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})$ belong to some standard types of distributions in the conjugate-exponential family, so that they can be efficiently sampled.

Starting with an arbitrary parameter set $\boldsymbol{\theta}^{(0)}$, the Gibbs sampler performs the following steps to sample the latent variable model with unknown parameters $\boldsymbol{\theta}$, latent variables $\boldsymbol{x}$ and observed variables $\boldsymbol{y}$:

(1) Sample $\boldsymbol{x}^{(t+1)} \sim p(\boldsymbol{x}|\boldsymbol{\theta}^{(t)},\boldsymbol{y})$;

(2) Sample $\boldsymbol{\theta}^{(t+1)} \sim p(\boldsymbol{\theta}|\boldsymbol{x}^{(t+1)},\boldsymbol{y})$.

The Gibbs sampler is simple, and the proof of its validity is trivial [106]. In fact, the Gibbs sampler can be viewed as a special M-H algorithm, which takes each conditional distribution as the proposal distribution, resulting in the acceptance probability $\mathcal{A}\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^*\right) \equiv 1$. Meanwhile, it is also possible to introduce the M-H steps into the Gibbs sampler if the conditional distribution cannot be directly sampled, to form the so called Metropolis-within-Gibbs (MwG) sampler [104].

The MCMC method has a wide applicability and it is easy to implement. In fact, the normalization factor of the posterior is not needed. However, compared with an algorithm that directly generates independent samples, the MCMC method has a number of disadvantages. First, the samples are correlated, which decreases the statistical efficiency. If we want a set of independent samples, a 'thinning' procedure

is usually implemented, i.e. discarding all but every kth samples. Second, the Markov chain eventually converges to the stationary distribution, but the initial samples may follow a different distribution and it is not easy to diagnose the convergence. In practice, a "burn-in" procedure is applied, that is to discard the first hundreds or thousands of samples. As for the diagnostics, multiple runs of a Markov chain with different starting points are usually used to check the convergence.

1.5 Summary of the remaining chapters

Chapter 2 introduces the principle of maximum expected utility (MEU) for the problem of OSP. A probabilistic model for OSP is proposed based on the modal form of equation of motion, and it turns out to be a Bayesian linear model. According to different design objectives, three utility functions, including the quadratic loss, Shannon information and the K-L divergence, are derived within the MEU framework. In order to improve the computational efficiency, we first derive the closed-form solution of covariance of modal responses under the banded white-noise excitation, and apply the nearest Kronecker product approximation, eigenvalue and Cholesky decompositions for the evaluation of utility functions. Finally, the sensor placements of a shear-type building and in the Guangzhou Tower are utilized to demonstrate the proposed approach.

Chapter 3 presents the application of Bayesian statistics for the OMA in the time-domain. First, a formulation of the OMA is developed using a state-space model of the structure. By making appropriate assumptions on the error models and the prior distributions of unknown parameters, a statistical model is finally constructed, which is a latent variable model in the conjugate-exponential family. Thereafter, the EM algorithm, the VB and the Gibbs sampler are employed sequentially to approximate the posterior distributions of the unknown parameters. Moreover, robust and efficient implementation strategies of all three algorithms are provided based on the square-root filtering and Cholesky decomposition. In the section of empirical study, a mass-spring numerical model, a laboratory shear-type building model and a high-rise building are used to show the performance of the proposed algorithms.

Chapter 4 investigates the application of Bayesian statistic to the problem of the FEMU. Based on the eigen-equation, a physical model for the FEMU is proposed considering the possible time-variance of the model parameters. The physical model is then transformed into a statistical model with latent variables via assigning probability distributions to the errors. Due to the complexity of the FEMU problem, an MwG sampler is designed to approximate the posterior distributions of the unknown stiffness parameters, and a robust implementation is provided as well. The

FEMU for a mass-spring model and a shear-type building model illustrate the validity of the developed method.

Chapter 5 provides a summary of the dissertation, the major findings of the study and the possible future work.

Appendix A

A.1 Latent-variable model

Due to lack of sensors or missing data, not all variables of interest can be directly measured in SHM. If we want to include these variables in the statistical analysis, then a latent variable model must be used. More specifically, a latent or hidden variable is a variable that is not directly observable but affects the observed or manifest variables; a latent-variable model is a statistical model with latent variables.

The difference of the latent-variable model with models that in which all variables are observed lies in the formulation of the likelihood. Specifically, the latent variable must be integrated out, because the likelihood needs to only bridge the observations and unknown parameters. Let denote $\boldsymbol{x}$ as the latent variables, $\boldsymbol{y}$ as the observed variables and $\boldsymbol{\theta}$ as the unknown parameters. Then

$$f(\boldsymbol{y}|\boldsymbol{\theta}) = \int f(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta})\, d\boldsymbol{x} \tag{1.22}$$

The term $f(\boldsymbol{y}|\boldsymbol{\theta})$ is called the incomplete likelihood, and the integrand $f(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta})$ is called the complete likelihood.

In this dissertation, the OMA and FEMU are both formulated as latent variable models. By introducing the latent variable, the inference for unknown parameters is simplified. But there is no free lunch: we must infer the latent variables first, resulting in a nonconvex optimization problem.

A.2 Conjugate-exponential family

Specification of parameter priors $\pi(\boldsymbol{\theta})$ is obviously a key element of the Bayesian machinery. Three general approaches exist in the literature [83]: subjective, objective, and empirical. The subjective Bayesian attempts to encapsulate prior knowledge as fully as possible in the form of previous experimental data or expert knowledge. It is often difficult to articulate qualitative experience or beliefs in mathematical form, but there exist one convenient and analytically favorable class of subjective priors: the conjugate priors in the exponential family, which is the choice in this dissertation.

Considering the latent-variable model in Figure 1.2, the conjugate-exponential

family satisfies the following conditions [94]:

Condition 1. The complete likelihood is of the exponential family:

$$f(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta}) = g(\boldsymbol{\theta})h(\boldsymbol{x}, \boldsymbol{y}) \exp[\boldsymbol{\phi}(\boldsymbol{\theta})^{\mathrm{T}}\boldsymbol{u}(\boldsymbol{x}, \boldsymbol{y})] \tag{1.23}$$

where $\boldsymbol{\phi}(\boldsymbol{\theta})$ is the vector of natural parameters, $\boldsymbol{u}$ and h are the functions that define the exponential family, and g is a normalisation constant.

Condition 2. The parameter prior is conjugate to the complete likelihood:

$$\pi(\boldsymbol{\theta}; \eta, \boldsymbol{\nu}) = c(\eta, \boldsymbol{\nu})g(\boldsymbol{\theta})^{\eta} \exp[\boldsymbol{\phi}(\boldsymbol{\theta})^{\mathrm{T}}\boldsymbol{\nu})] \tag{1.24}$$

where η and $\boldsymbol{\nu}$ are hyperparameters of the prior, and c is a normalisation constant.

Combining Condition 1 and Condition 2, it is easy to see that the posterior of the unknown parameter $\boldsymbol{\theta}$, which is proportional to the product of Eqns. (1.23) and (1.24) has the same parametric form as its prior. The hyperparameters of a conjugate prior can be interpreted as the number (η) and values ($\boldsymbol{\nu}$) of pseudo-observations under the corresponding likelihood. Because of tractability in computing the posterior and the interpretability of the prior, all statistical models in this dissertation are formulated within the conjugate-exponential family.

A.3 Bayesian network

Statistical modeling problems often involve large numbers of interacting random variables and it is often convenient to express the dependencies between these variables graphically. The Bayesian network [107] or directed acyclic graph is one of the probabilistic graphical models that represent a set of random variables and their conditional dependencies.

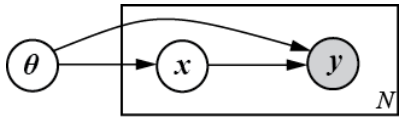


Figure 1.2 Bayesian network for a latent variable model.

As illustrated in Figure 1.2 for a latent-variable model, Bayesian networks are comprised of nodes and directed edges. Each node represents a random variable (or vector of random variables), which can be the observed variables (shaded node $\boldsymbol{y}$), the latent variables (node $\boldsymbol{x}$) or the unknown parameters (node $\boldsymbol{\theta}$). The directed edges represent conditional dependencies; nodes that are not connected (there is no path from one of the variables to the other) represent variables that are conditionally independent of each other. The plate notation is applied in Figure1 to indicate repeated variables $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$, where N shown in the corner of the plate indicates

the number of repetitions. Figure 1.2 shows a scenario, where the repeated variables $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ are independent and identically distributed (i.i.d.) given the parameter $\boldsymbol{\theta}$. In the case, where $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ are correlated with $\boldsymbol{x}_{i+1}$ and $\boldsymbol{y}_{i+1}$ given $\boldsymbol{\theta}$, the dynamic Bayesian networks can be used. This is the case for the state-space model (SSM) in Chapter 3.

A conditional probability distribution is associated with each node, given its parent nodes. Taking advantage of the graphical representation, efficient algorithms have been developed to perform inference and learning, e.g. the junction tree algorithm for discrete distributions [107]. Since we will only deal with continuous distributions in the dissertation, such algorithms are not applicable, but all of the models constructed in this dissertation will be accompanied with graphical model descriptions for visual description of the model and variable dependencies.

Chapter 2
Optimal sensor placement[†]

2.1 Introduction

In this chapter, we focus on the optimal sensor placement (OSP), that is an inevitable problem in SHM. Within the framework of vibration-based SHM, we place the sensors for optimal operational modal analysis (OMA), where only structural responses under operating conditions are measured and the unmeasured force is modeled as a broad-band random process, in most cases a zero-mean, band-limited white-noise (BWN). To account for the uncertainties in the unmeasured force, the unknown structural model, as well as the ubiquitous measurement error, it is indispensable to employ a probabilistic model for the OSP. Therefore, we formulate the OSP as a decision problem under uncertainty, and the maximum expected utility (MEU) theory and a Bayesian linear model are applied for a robust sensor placement.

We first develop a probabilistic framework for sensor placement incorporating model, load and measurement uncertainties in Section 2.2. It turns out that the problem of OSP corresponds to a Bayesian linear model. In Section 2.3, the principle of MEU is applied to find the optimal sensor configuration, and different design objectives are encoded into three utility functions. The computational issues are addressed in Section 2.4, including calculation of the covariance matrix of modal responses under BWN and fast evaluation of the utility functions. Optimal sensor placements of a shear-type mass-spring system and in the Guangzhou TV Tower are presented in Section 2.5 in order to demonstrate the proposed approach, based on which conclusions are provided in Section 2.6.

2.2 Problem formulation

2.2.1 The physical model

The equation of motion of a discrete, linear, and time-invariant dynamical system with N_d degrees of freedom (DOFs) under stochastic external force and ground motion is described by

[†] This chapter has been published in Mechanical Systems and Signal Processing, 2016, **75**:155-175.

$$\boldsymbol{M}\ddot{\boldsymbol{u}}(t) + \boldsymbol{C}_d\dot{\boldsymbol{u}}(t) + \boldsymbol{K}\boldsymbol{u}(t) = \boldsymbol{P}\boldsymbol{f}(t) - \boldsymbol{M}\boldsymbol{l}\ddot{\boldsymbol{u}}_g(t) \tag{2.1}$$

where $\boldsymbol{M} \in \mathbb{R}^{N_d \times N_d}$, $\boldsymbol{C}_d \in \mathbb{R}^{N_d \times N_d}$ and $\boldsymbol{K} \in \mathbb{R}^{N_d \times N_d}$ are the mass, damping and stiffness matrices, respectively; $\boldsymbol{u}(t) \in \mathbb{R}^{N_d}$, $\dot{\boldsymbol{u}}(t) \in \mathbb{R}^{N_d}$ and $\ddot{\boldsymbol{u}}(t) \in \mathbb{R}^{N_d}$ are the nodal displacement, velocity and acceleration responses relative to the ground motion, respectively, with $\boldsymbol{u}(0) = \boldsymbol{u}_0$ and $\dot{\boldsymbol{u}}(0) = \dot{\boldsymbol{u}}_0$ being the initial relative displacement and velocity vectors; $\boldsymbol{f}(t) \in \mathbb{R}^{N_f}$ is the external force vector; $\boldsymbol{P} \in \mathbb{R}^{N_d \times N_f}$ is the load coefficient matrix representing the spatial influence of the external force; $\ddot{\boldsymbol{u}}_g(t) \in \mathbb{R}^{N_g}$ is the ground acceleration vector, and $\boldsymbol{l} \in \mathbb{R}^{N_d \times N_g}$ is the corresponding influence matrix.

If the classical damping [22] is further assumed, the equations of motion can be decoupled by introducing the transformation $\boldsymbol{u}(t) = \boldsymbol{\Phi}\boldsymbol{q}(t)$ to yield the modal form

$$\begin{gathered} \ddot{q}_i(t) + 2\xi_i\omega_i\dot{q}_i(t) + \omega_i^2 q_i(t) = \boldsymbol{\gamma}_i^e\boldsymbol{f}(t) + \boldsymbol{\gamma}_i^g\ddot{\boldsymbol{u}}_g(t) \\ q_i(0) = \boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{M}\boldsymbol{u}_0/M_i,\ \dot{q}_i(0) = \boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{M}\dot{\boldsymbol{u}}_0/M_i \end{gathered} \tag{2.2}$$

for $i = 1, \dots, N_d$, where

$$\begin{gathered} \boldsymbol{q}(t) = \left[q_1(t), q_2(t), \dots, q_{N_d}(t)\right]^{\mathrm{T}} \in \mathbb{R}^{N_d} \\ \boldsymbol{\Phi} = \left[\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \dots, \boldsymbol{\phi}_{N_d}\right] \in \mathbb{R}^{N_d \times N_d} \end{gathered} \tag{2.3}$$

are, respectively, the modal displacement vector and the mode shape matrix. In addition,

$$\begin{gathered} M_i = \boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{M}\boldsymbol{\phi}_i \\ \omega_i = \sqrt{\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{K}\boldsymbol{\phi}_i/M_i} \\ \xi_i = \boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{C}\boldsymbol{\phi}_i/(2M_i\omega_i) \\ \boldsymbol{\gamma}_i^e = \boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{P}/M_i \\ \boldsymbol{\gamma}_i^g = \boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{M}\boldsymbol{l}/M_i \end{gathered} \tag{2.4}$$

are the ith modal mass, natural frequency, damping ratio and participation factors for external force and base motion, respectively.

Let there be n sensors and $\boldsymbol{S} \in \mathbb{R}^{n \times N_d}$ denote the selection matrix for the placement of sensors such that $\ddot{\boldsymbol{u}}_{ms} = \boldsymbol{S}\ddot{\boldsymbol{u}}$ represents the measured accelerations. In monitoring applications, usually only a few well-excited modes are observable. Let m denote the number of such modes and $\boldsymbol{S}_q \in \mathbb{R}^{m \times N_d}$ denote the selection matrix

so that $\boldsymbol{S}_q\boldsymbol{q}$ represents the coordinates and $\boldsymbol{\Phi}\boldsymbol{S}_q^{\mathrm{T}}$ represents the eigenvectors of the selected modes. The product $\boldsymbol{\Phi}\boldsymbol{S}_q^{\mathrm{T}}\boldsymbol{S}_q\ddot{\boldsymbol{q}}$ then gives the contributions of the selected modes to the acceleration response. Furthermore, responses are recorded in discrete time and measurement errors are ubiquitous. Letting T denote the sampling period, one can write for the measured accelerations at time $t = kT,\ k = 0, 1, \dots, N$,

$$\ddot{\boldsymbol{u}}_{ms}[k] = \boldsymbol{S}(\boldsymbol{\Phi}\boldsymbol{S}_q^{\mathrm{T}}\boldsymbol{S}_q\ddot{\boldsymbol{q}}[k] + \boldsymbol{l}\ddot{\boldsymbol{u}}_g[k] + \boldsymbol{w}[k]) \tag{2.5}$$

where the error term $\boldsymbol{w}[k]$ represents the joint contribution of the unobservable modes and the measurement noise.

Both $\boldsymbol{S}$ and $\boldsymbol{S}_q$ consist of only ones and zeroes so that $\boldsymbol{S}^{\mathrm{T}}\boldsymbol{S}$ and $\boldsymbol{S}_q^{\mathrm{T}}\boldsymbol{S}_q$ are diagonal matrices with unity in their ith diagonal element only if the ith DOF or ith mode is selected, and zero otherwise. A similar formulation can be used to represent measured displacement or velocity, or even a mixture of measurements. Here, we show the formulation for measured acceleration because it is the most commonly measured quantity used for modal identification. The measured responses are absolute quantities, and it is for that reason that the ground acceleration $\ddot{\boldsymbol{u}}_g$ appears in the equation. Because in most cases there will be an accelerograph to record the ground motion, we assume it to be known with some measurement uncertainty and reformulate Eqn. (2.5)to read

$$\ddot{\boldsymbol{u}}_{rs}[k] = \ddot{\boldsymbol{u}}_{ms}[k] - \boldsymbol{S}\boldsymbol{l}\ddot{\boldsymbol{u}}_g[k] = \boldsymbol{S}\boldsymbol{\Phi}\boldsymbol{S}_q^{\mathrm{T}}\boldsymbol{S}_q\ddot{\boldsymbol{q}}[k] + \boldsymbol{S}\boldsymbol{w}[k] \tag{2.6}$$

where $\ddot{\boldsymbol{u}}_{rs}$ is now the "measured" relative acceleration. Another issue is that both $\ddot{\boldsymbol{q}}[k]$ and $\ddot{\boldsymbol{u}}_{rs}[k]$ are temporally correlated processes, hence we have to consider all N measurement instants simultaneously for identification of modal response. For the sake of convenience, we introduce the following notations:

$$\begin{aligned} \boldsymbol{Y}_{rs} &= \left[\ddot{\boldsymbol{u}}_{rs}[1]; \dots; \ddot{\boldsymbol{u}}_{rs}[N]\right] \\ \boldsymbol{X}_s &= \boldsymbol{I}_N \otimes (\boldsymbol{S}\boldsymbol{\Phi}\boldsymbol{S}_q^{\mathrm{T}}) \\ \boldsymbol{\Theta}_s &= \left[\boldsymbol{I}_N \otimes \boldsymbol{S}_q\right]\left[\ddot{\boldsymbol{q}}[1]; \dots; \ddot{\boldsymbol{q}}[N]\right] \\ \boldsymbol{W}_s &= \left[\boldsymbol{I}_N \otimes \boldsymbol{S}\right]\left[\boldsymbol{w}[1]; \dots; \boldsymbol{w}[N]\right] \end{aligned} \tag{2.7}$$

in which $\boldsymbol{I}_N \in \mathbb{R}^{N\times N}$ is the identity matrix and "$\otimes$" stands for the Kronecker product. Eqn. (3.7) then becomes

$$\boldsymbol{Y}_{rs} = \boldsymbol{X}_s\boldsymbol{\Theta}_s + \boldsymbol{W}_s \tag{2.8}$$

As mentioned earlier, the primary goal of sensor placement for OMA is to identify the modal parameters. However, the measured structural responses are nonlinearly related to the modal parameters, making the problem of directly

optimizing sensor locations for estimating modal parameters exceedingly difficult. On the other hand, as shown in Eqn. (3.10), a linear relation exits between the measured structural responses and the modal responses $\boldsymbol{\Theta}_s$. Since $\boldsymbol{\Theta}_s$ contains all the information necessary for identifying modal frequencies and damping ratios, and the mode shapes can be estimated from Eqn. (3.10), e.g., by the method of least squares, once $\boldsymbol{\Theta}_s$ are known, the parameter $\boldsymbol{\Theta}_s$ can be used as a surrogate for modal parameters in OSP [25]. Although the optimal sensor locations for estimating $\boldsymbol{\Theta}_s$ are not necessarily optimal for estimating the modal parameters, the approach at least provides a suboptimal solution. This is the approach employed in this paper.

2.2.2 The probabilistic model

In OMA, the structural load is unknown and modeled as a broad-band random process, in most cases as a zero-mean, band-limited white-noise (BWN) process. As a consequence, the modal responses are also random processes. Considering the ubiquitous measurement error as well, it is necessary to describe the OSP as a probabilistic model. The uncertainties in OMA mainly come from the uncertain structural model, the unmeasured structural load and the measurement error. Therefore, the terms $\boldsymbol{Y}_{rs}$, $\boldsymbol{X}_s$, $\boldsymbol{\Theta}_s$ and $\boldsymbol{W}_s$ in Eqn. (3.10) are all random variables.

The measurement error $\boldsymbol{W}_s$ represents the uncertainties in the measurement system, which may include sensor error, transmission error, transducer error and computer error. From the principle of maximum entropy [108], it is modeled by a zero-mean Gaussian white-noise process with covariance matrix $\boldsymbol{\Sigma}$, i.e. $\boldsymbol{W}_s \sim MVN(0, \boldsymbol{I}_N \otimes \boldsymbol{\Sigma}_s)$ where $\boldsymbol{\Sigma}_s = \boldsymbol{S}\boldsymbol{\Sigma}\boldsymbol{S}^{\mathrm{T}}$.Strictly speaking, the measurement error can be spatially correlated, but for the ease of mathematical treatment, they are usually treated as independent and identically distributed random variables. Here, we will employ this approximation to assume $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_n$. The variance σ^2 is also introduced as a random variable to allow a full Bayesian treatment.

With the input load modeled as a zero-mean, Gaussian BWN, $\boldsymbol{\Theta}_s$ follows a zero-mean multivariate normal distribution $MVN(0, \boldsymbol{\Gamma}_s)$ with $\boldsymbol{\Gamma}_s = (\boldsymbol{I}_N \otimes \boldsymbol{S}_q)\boldsymbol{\Gamma}(\boldsymbol{I}_N \otimes \boldsymbol{S}_q^{\mathrm{T}})$. The covariance matrix $\boldsymbol{\Gamma} \in \mathbb{R}^{N_d N \times N_d N}$ can be determined once the power spectrum density (PSD) of the load and the structural model are known. In fact, only the bandwidth ω_c and amplitude P_0 are in need because of the BWN assumption on the load. Since both the load and measurement error are unmeasurable, it is not necessary to distinguish their individual variance, only the input-to-noise ratio (INR) $\omega_c P_0/\sigma^2$ is sufficient. We assume the INR to follow a uniform distribution $[a, b]$ to represent an inherent large uncertainty. As for the bandwidth, the usual assumption is that it includes at least all the modes of interest. Measurements on site can be used to guide the specification of these parameters in practice. Note that covariance matrix $\boldsymbol{\Gamma}$ a matrix random variable, whose uncertainty inherits from the

uncertain load and structural model.

The random variable $\boldsymbol{X}_s$ represents the uncertainty in our characterization of the target structure. A finite element model is suitable to provide this information, in which the nominal value from the design drawings serves as the mean and the covariance matrix of input parameters is specified by expert knowledge to represent the confidence in the nominal model. Since $\boldsymbol{X}_s$ only contains the mode shapes, we can alternatively randomize the mode shape matrix, which perhaps is more efficient because repeated eigenvalue analysis is avoided. We choose to use a finite element model due to its intuitiveness, although more computation is involved.

With $\boldsymbol{X}_s$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ considered to be random variables, Eqn. (3.10) describes a linear normal Bayesian model, which can be conveniently represented by a Bayesian network model shown in Figure 2.1.

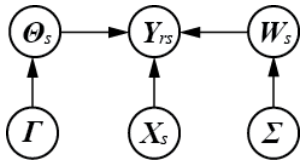


Figure 2.1 Bayesian network for the OSP.

In this model, the random variable $\boldsymbol{Y}_{rs}$ given $\boldsymbol{X}_s$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ follows the multivariate normal distribution. Furthermore, we have

$$\begin{aligned} \mathrm{E}[\boldsymbol{Y}_{rs}|\boldsymbol{X}_s,\boldsymbol{\Gamma},\boldsymbol{\Sigma}] &= \mathrm{E}[\boldsymbol{X}_s\boldsymbol{\Theta}_s + \boldsymbol{W}_s] = \boldsymbol{0} \\ \mathrm{E}[\boldsymbol{\Theta}_s\boldsymbol{Y}_{rs}^{\mathrm{T}}|\boldsymbol{X}_s,\boldsymbol{\Gamma},\boldsymbol{\Sigma}] &= \mathrm{E}[\boldsymbol{\Theta}_s\boldsymbol{\Theta}_s^{\mathrm{T}}|\boldsymbol{\Gamma}]\boldsymbol{X}_s^{\mathrm{T}} = \boldsymbol{\Gamma}_s\boldsymbol{X}_s^{\mathrm{T}} \\ \mathrm{Cov}[\boldsymbol{Y}_{rs}|\boldsymbol{X}_s,\boldsymbol{\Gamma},\boldsymbol{\Sigma}] &= \boldsymbol{X}_s\boldsymbol{\Gamma}_s\boldsymbol{X}_s^{\mathrm{T}} + \boldsymbol{I}_N \otimes \boldsymbol{\Sigma}_s \end{aligned} \tag{2.9}$$

The above equations imply that the joint distribution of $\boldsymbol{\Theta}_s$ and $\boldsymbol{Y}_{rs}$ given $\boldsymbol{X}_s$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ is a multivariate normal distribution with zero-mean and covariance

$$\begin{bmatrix} \boldsymbol{\Gamma}_s & \boldsymbol{\Gamma}_s\boldsymbol{X}_s^{\mathrm{T}} \\ \boldsymbol{X}_s\boldsymbol{\Gamma}_s^{\mathrm{T}} & \boldsymbol{X}_s\boldsymbol{\Gamma}_s\boldsymbol{X}_s^{\mathrm{T}} + \boldsymbol{I}_N \otimes \boldsymbol{\Sigma}_s \end{bmatrix} \tag{2.10}$$

Therefore, given $\boldsymbol{Y}_{rs}$, $\boldsymbol{X}_s$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ the posterior distribution of $\boldsymbol{\Theta}_s$ is multivariate normal with mean and covariance:

$$\begin{aligned} &\mathrm{E}[\boldsymbol{\Theta}_s|\boldsymbol{Y}_{rs},\boldsymbol{X}_s,\boldsymbol{\Gamma},\boldsymbol{\Sigma}] = \boldsymbol{\Gamma}_s\boldsymbol{X}_s^{\mathrm{T}}[\boldsymbol{X}_s\boldsymbol{\Gamma}_s\boldsymbol{X}_s^{\mathrm{T}} + \boldsymbol{I}_N \otimes \boldsymbol{\Sigma}_s]^{-1}\boldsymbol{Y}_{rs} \\ &= [\boldsymbol{\Gamma}_s^{-1} + \boldsymbol{X}_s^{\mathrm{T}}(\boldsymbol{I}_N \otimes \boldsymbol{\Sigma}_s^{-1})\boldsymbol{X}_s]\boldsymbol{X}_s^{\mathrm{T}}(\boldsymbol{I}_N \otimes \boldsymbol{\Sigma}_s^{-1})\boldsymbol{Y}_{rs} \end{aligned} \tag{2.11}$$

$$\begin{aligned} &\mathrm{Cov}[\boldsymbol{\Theta}_s|\boldsymbol{Y}_{rs},\boldsymbol{X}_s,\boldsymbol{\Gamma},\boldsymbol{\Sigma}] = \boldsymbol{\Gamma}_s - \boldsymbol{\Gamma}_s\boldsymbol{X}_s^{\mathrm{T}}[\boldsymbol{X}_s\boldsymbol{\Gamma}_s\boldsymbol{X}_s^{\mathrm{T}} + \boldsymbol{I}_N \otimes \boldsymbol{\Sigma}_s]^{-1}\boldsymbol{X}_s\boldsymbol{\Gamma}_s^{\mathrm{T}} \\ &= [\boldsymbol{\Gamma}_s^{-1} + \boldsymbol{X}_s^{\mathrm{T}}(\boldsymbol{I}_N \otimes \boldsymbol{\Sigma}_s^{-1})\boldsymbol{X}_s]^{-1} = [\boldsymbol{\Gamma}_s^{-1} + \boldsymbol{I}_N \otimes \boldsymbol{F}_s]^{-1} \end{aligned} \tag{2.12}$$

where the matrix inversion lemma [109] is used in both equations. Here, we have

introduced the Fisher information matrix

$$\boldsymbol{F}_s = \left(\boldsymbol{S\Phi S}_q^{\mathrm{T}}\right)^{\mathrm{T}} \boldsymbol{\Sigma}_s^{-1} (\boldsymbol{S\Phi S}_q^{\mathrm{T}}) \tag{2.13}$$

Eqn. (2.11) provides a minimum mean-square-error estimate of $\boldsymbol{\Theta}_s$, and the covariance in Eqn. (2.12) incorporates the uncertainty inherent in the modal responses, the measurement noise and the uncertain mode shapes because of the randomness in $\boldsymbol{\Theta}_s$, $\boldsymbol{\Sigma}$ and $\boldsymbol{X}_s$, respectively. Note that random variables $\boldsymbol{X}_s$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ depend on the structural properties and sensor quality, and that all of them remain unknown in the design stage. In order to achieve a robust sensor deployment, we need to integrate out all these uncertainties. Unfortunately, there is no closed form solution for the required high-dimension integrals, so we compute them using Monte Carlo integration, as described in Section 2.4.2.

2.3 Maximum expected utility based OSP

2.3.1 Methodology

As we have seen, in the probabilistic model for sensor placement there are uncertainties in the load, in the model parameters, and in the measurements. Consequently, the OSP is essentially a problem of making a decision under uncertainty. The Von Neumann–Morgenstern utility theorem [110] justifies the expected utility hypothesis, i.e. that the optimal decision is the one that provides maximum expected utility (MEU). Thus, MEU provides a solid mathematical criterion for solving the OSP problem.

The decision alternatives in OSP mainly include the number and location of sensors and choice of modes of interest. Let the utility function $U(d, \boldsymbol{\Theta}_s)$ encode the consequences (costs) of choosing decision alternative d from a decision space $\mathcal{D}$ when the value of unknown parameters is $\boldsymbol{\Theta}_s$. According to MEU, the optimal decision is given by

$$U(d^*) = \max_{d\in\mathcal{D}} \mathrm{E}_{\boldsymbol{\Theta}_s}[U(d, \boldsymbol{\Theta}_s)] = \max_{d\in\mathcal{D}} \int_{\mathcal{P}} U(d, \boldsymbol{\Theta}_s) p(\boldsymbol{\Theta}_s) d\boldsymbol{\Theta}_s \tag{2.14}$$

where $p(\boldsymbol{\Theta}_s)$ is the posterior distribution of $\boldsymbol{\Theta}_s$ and $\mathrm{E}_{\boldsymbol{\Theta}_s}[\cdot]$ denotes the expectation with respect to the distribution of $\boldsymbol{\Theta}_s$. The expectation with respect to the marginal distribution of $\boldsymbol{\Theta}_s$ accounts for the uncertainty in $\boldsymbol{\Theta}_s$; however, in the design stage, only the pre-posterior distribution of $\boldsymbol{\Theta}_s$ given $\boldsymbol{Y}_{rs}$, $\boldsymbol{X}_s$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ is available. Therefore, we have to consider all the possible outcomes of the measured responses $\boldsymbol{Y}_{rs}$, mode shapes $\boldsymbol{X}_s$, structure/load relevant term $\boldsymbol{\Gamma}$, and measurement noise $\boldsymbol{\Sigma}$. Eqn. (2.14), thus, becomes

$$U(d^*) = \max_{d\in\mathcal{D}} \mathrm{E}_{\boldsymbol{Y}_{rs}\boldsymbol{X}_s\boldsymbol{\Gamma}\boldsymbol{\Sigma}}\left[\mathrm{E}_{\boldsymbol{\theta}_s|\boldsymbol{Y}_{rs}\boldsymbol{X}_s\boldsymbol{\Gamma}\boldsymbol{\Sigma}}[U(d,\boldsymbol{\theta}_s)]\right] \tag{2.15}$$

If the independence between $\boldsymbol{X}_s$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ is assumed, we have the joint distribution of $\boldsymbol{Y}_{rs}$, $\boldsymbol{X}_s$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ as

$$p(\boldsymbol{Y}_{rs},\boldsymbol{X}_s,\boldsymbol{\Gamma},\boldsymbol{\Sigma}) = p(\boldsymbol{Y}_{rs}|\boldsymbol{X}_s,\boldsymbol{\Gamma},\boldsymbol{\Sigma})d\boldsymbol{Y}_{rs}p(\boldsymbol{X}_s)p(\boldsymbol{\Gamma})p(\boldsymbol{\Sigma}) \tag{2.16}$$

Hence, the marginal distribution of $\boldsymbol{\theta}_s$ is obtained by integrating out the uncertainties in $\boldsymbol{Y}_{rs}$, $\boldsymbol{X}_s$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$, thereby providing a robust sensor design against unknown structure, load and sensor properties.

2.3.2 Choice of utility functions

Selecting a utility function tailored to the goals of the design problem is an important step. For the present problem OSP for modal identification, we know that only lower modes of vibration are well excited. So, in practice, typically a few of the lower modes are selected as the targets for identification. The number of sensors is also constrained by the cost of the sensory system. In this paper, we consider a simplified sensor placement problem: find the optimal sensor locations given the number of available sensors n and target modes $\boldsymbol{\Phi}\boldsymbol{S}_q$. Accordingly, three utility functions based on the quadratic loss, Shannon information, and Kullback-Leibler (K-L) divergence are investigated.

Quadratic Loss

Quadratic Loss corresponds to the widely used mean square error (MSE) criterion in operational modal identification. This utility function is appropriate when random phenomena are Gaussian distributed. Specifically, the expected utility function is expressed as:

$$\begin{aligned} U_1(d) &= \mathrm{E}_{\boldsymbol{Y}_{rs}\boldsymbol{X}_s\boldsymbol{\Gamma}\boldsymbol{\Sigma}}\left[\mathrm{E}_{\boldsymbol{\theta}_s|\boldsymbol{Y}_{rs}\boldsymbol{X}_s\boldsymbol{\Gamma}\boldsymbol{\Sigma}}[-(\boldsymbol{\theta}_s-\overline{\boldsymbol{\theta}}_s)^{\mathrm{T}}(\boldsymbol{\theta}_s-\overline{\boldsymbol{\theta}}_s)]\right] \\ &= -\mathrm{E}_{\boldsymbol{X}_s\boldsymbol{\Gamma}\boldsymbol{\Sigma}}[\mathrm{tr}\,\{[\boldsymbol{\Gamma}_s^{-1}+\boldsymbol{I}_N\otimes\boldsymbol{F}_s]^{-1}\})] \end{aligned} \tag{2.17}$$

where $\overline{\boldsymbol{\theta}}_s$ is the posterior mean of random variables $\boldsymbol{\theta}_s$. This criterion has been widely used in Bayesian experimental design, see [42,111].

Shannon Information

Shannon information [112] is a measure of the amount of variability associated with a random variable; it considers the probability density function instead of only the second order statistics. Since most real-life problems are governed by nonlinear

equations and the random phenomenon may be quite far from being normally distributed, it may be more reasonable to use the Shannon information as the cost function. Negative Shannon information has been used by Bernardo [113] as a utility function in Bayesian experimental design and recommended by Papadimitriou [28] for OSP. For our application, this utility function has the following form:

$$\begin{aligned} U_2(d) &= \mathrm{E}_{\boldsymbol{Y}_{rs}\boldsymbol{X}_s\boldsymbol{\Gamma}\boldsymbol{\Sigma}}\left[\mathrm{E}_{\boldsymbol{\theta}_s|\boldsymbol{Y}_{rs}\boldsymbol{X}_s\boldsymbol{\Gamma}\boldsymbol{\Sigma}}[\ln p(\boldsymbol{\theta}_s|\boldsymbol{Y}_{rs},\boldsymbol{X}_s,\boldsymbol{\Gamma},\boldsymbol{\Sigma})]\right] \\ &= \mathrm{E}_{\boldsymbol{X}_s\boldsymbol{\Gamma}\boldsymbol{\Sigma}}[\ln\det[\boldsymbol{\Gamma}_s^{-1}+\boldsymbol{I}_N\otimes\boldsymbol{F}_s])] \end{aligned} \tag{2.18}$$

where the final result is obtained due to the joint normal distribution of $(\boldsymbol{\theta}_s,\boldsymbol{Y}_{rs})$ for given values of $\boldsymbol{X}_s,\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$.

There are two other ways to interpret the above selection of the utility function via the K-L divergence and mutual information. Rewrite Eqn. (11) as follows:

$$U_2(d) \propto \mathrm{E}_{\boldsymbol{Y}_{rs}\boldsymbol{X}_s\boldsymbol{\Gamma}\boldsymbol{\Sigma}}\left[\mathrm{E}_{\boldsymbol{\theta}_s|\boldsymbol{Y}_{rs}\boldsymbol{X}_s\boldsymbol{\Gamma}\boldsymbol{\Sigma}}\left[\ln\frac{p(\boldsymbol{\theta}_s|\boldsymbol{Y}_{rs},\boldsymbol{X}_s,\boldsymbol{\Gamma},\boldsymbol{\Sigma})}{\pi(\boldsymbol{\theta}_s)}\right]\right] \tag{2.19}$$

$$\propto \mathrm{E}_{\boldsymbol{X}_s\boldsymbol{\Gamma}\boldsymbol{\Sigma}}\left[\mathrm{E}_{\boldsymbol{\theta}_s\boldsymbol{Y}_{rs}|\boldsymbol{X}_s\boldsymbol{\Gamma}\boldsymbol{\Sigma}}\left[\ln\frac{p(\boldsymbol{\theta}_s,\boldsymbol{Y}_{rs}|\boldsymbol{X}_s,\boldsymbol{\Gamma},\boldsymbol{\Sigma})}{\pi(\boldsymbol{\theta}_s)p(\boldsymbol{Y}_{rs}|\boldsymbol{X}_s,\boldsymbol{\Gamma},\boldsymbol{\Sigma})}\right]\right] \tag{2.20}$$

Eqn. (2.19) is the expected K-L divergence [112] between the posterior and prior distributions of parameters $\boldsymbol{\theta}_s$. $U_2(d)$ and the K-L divergence achieve their maxima under the same decision d because the prior distribution $\pi(\boldsymbol{\theta}_s)$ does not depend on d. K-L divergence is a measure of distance between two probability distributions, and here, maximizing it implies extracting more information from the data other than the prior. Eqn. (2.20) expresses the expected mutual information [112] of parameters $\boldsymbol{\theta}_s$ and data $\boldsymbol{Y}_{rs}$ conditioned on $\boldsymbol{X}_s$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$. Intuitively, mutual information measures the information that $\boldsymbol{\theta}_s$ and $\boldsymbol{Y}_{rs}$ share; it measures the amount of uncertainty reduction in one set of variable when the other set is known. Hence, maximizing the expected gain in the Shannon information is equivalent to maximizing the dependence of $\boldsymbol{\theta}_s$ on data $\boldsymbol{Y}_{rs}$. Besides these interpretations, Eqn. (2.18) can also be derived from other utility functions, see [114,115].

K-L Divergence

As described above, the K-L divergence is a measure of distance between two probability distributions. Li [31] has introduced the K-L divergence criterion for OSP. Unlike the preceding utility functions that were focused on OMA, the proposition here is more test-oriented: the perfect information on modal parameters $\boldsymbol{\theta}_s$ can be extracted by placing sensors everywhere in the structure; so the optimal configuration for a given set of $n < N_d$ sensors should provide information that is as close as possible to that of full sensor set. Contrary to the value of information [116] that

quantifies the reduction in expected loss (negative utility) due to the availability of the information, K-L divergence measures the reduction in utility by removal of sensors from the full configuration. Since for the N_d sensor case $\boldsymbol{S} = \boldsymbol{I}_{N_d}$ and $\boldsymbol{S}_q$ doesn't change, hereafter the subscript s on all symbols for this case are removed except for $\boldsymbol{\Theta}_s$ and $\boldsymbol{\Gamma}_s$.

Let $p(\boldsymbol{\Theta}_s|\boldsymbol{Y}_{rs}, \boldsymbol{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})$ denote the posterior distribution of $\boldsymbol{\Theta}_s$ from n sensors and $p(\boldsymbol{\Theta}_s|\boldsymbol{Y}_r, \boldsymbol{X}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})$ denote the posterior distribution of $\boldsymbol{\Theta}_s$ from N_d sensors. Selecting the negative K-L divergence between $p(\boldsymbol{\Theta}_s|\boldsymbol{Y}_{rs}, \boldsymbol{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})$ and $p(\boldsymbol{\Theta}_s|\boldsymbol{Y}_r, \boldsymbol{X}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})$ as the utility function, we have

$$
\begin{aligned}
&U_3(d) \\
&= \mathrm{E}_{\boldsymbol{Y}_r \boldsymbol{X} \boldsymbol{\Gamma} \boldsymbol{\Sigma}} \left[\mathrm{E}_{\boldsymbol{\Theta}_s|\boldsymbol{Y}_{rs}\boldsymbol{X}_s\boldsymbol{\Gamma}\boldsymbol{\Sigma}} \left[\ln \frac{p(\boldsymbol{\Theta}_s|\boldsymbol{Y}_{rs}, \boldsymbol{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})}{p(\boldsymbol{\Theta}_s|\boldsymbol{Y}_r, \boldsymbol{X}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})} \right] \right] \\
&= -\mathrm{E}_{\boldsymbol{X}\boldsymbol{\Gamma}\boldsymbol{\Sigma}}[0.5 \ln \det[\boldsymbol{\Gamma}_s^{-1} + \boldsymbol{I}_N \otimes \boldsymbol{F}_s] + \mathrm{tr}\,\{[\boldsymbol{\Gamma}_s^{-1} + \boldsymbol{I}_N \otimes \boldsymbol{F}_s]^{-1}[\boldsymbol{\Gamma}_s^{-1} + \boldsymbol{I}_N \otimes \boldsymbol{F}]\}]
\end{aligned}
\tag{2.21}
$$

The derivation of this utility function is not as straightforward as the previous two; it is described in Appendix B.1. This utility function is related to the preceding two utility functions: the first term inside the large curly bracket is one half of $U_2(d)$ and the second term can be regarded as a normalized version of $U_1(d)$. As can be seen in the examples below, the second term dominates the utility function $U_3(d)$.

Besides the motivation behind the K-L divergence, there is another meaningful interpretation. Recall $\boldsymbol{Y}_r = [\boldsymbol{Y}_{rs}^{\mathrm{T}}, \boldsymbol{Y}_{rn}^{\mathrm{T}}]^{\mathrm{T}}$, where $\boldsymbol{Y}_{rn}$ denotes the response at the non-selected DOFs. Eqn. (2.21) can be equivalently written as

$$
\begin{aligned}
&U_3(d) \\
&= \mathrm{E}_{\boldsymbol{Y}_{rs}\boldsymbol{X}\boldsymbol{\Gamma}\boldsymbol{\Sigma}} \left[\mathrm{E}_{\boldsymbol{\Theta}_s|\boldsymbol{Y}_{rs}\boldsymbol{X}\boldsymbol{\Gamma}\boldsymbol{\Sigma}} \left[\mathrm{E}_{\boldsymbol{Y}_{rn}|\boldsymbol{Y}_{rs}\boldsymbol{X}\boldsymbol{\Gamma}\boldsymbol{\Sigma}} \left[\ln \frac{p(\boldsymbol{\Theta}_s|\boldsymbol{Y}_{rs}, \boldsymbol{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})\, p(\boldsymbol{Y}_{rn}|\boldsymbol{Y}_{rs}, \boldsymbol{X}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})}{p(\boldsymbol{\Theta}_s, \boldsymbol{Y}_{rn}|\boldsymbol{Y}_{rs}, \boldsymbol{X}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})} \right] \right] \right]
\end{aligned}
\tag{2.22}
$$

in which the term inside the outmost square bracket is the K-L divergence of the product of conditional distributions $p(\boldsymbol{\Theta}_s|\boldsymbol{Y}_{rs}, \boldsymbol{X}_s, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})$ and $p(\boldsymbol{Y}_{rn}|\boldsymbol{Y}_{rs}, \boldsymbol{X}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})$ and the joint distribution $p(\boldsymbol{\Theta}_s, \boldsymbol{Y}_{rn}|\boldsymbol{Y}_{rs}, \boldsymbol{X}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})$. Maximizing utility $U_3(d)$ means minimizing the above K-L divergence, i.e. making the joint distribution of $\boldsymbol{\Theta}_s$ and $\boldsymbol{Y}_{rn}$ given $\boldsymbol{Y}_{rs}$ approximately equal to their product. Indeed, this assures conditional independence of $\boldsymbol{\Theta}_s$ and $\boldsymbol{Y}_{rn}$, given $\boldsymbol{Y}_{rs}$. Therefore, this criterion makes the modal parameters and the unmeasured responses as independent as possible, given the measured responses.

2.4 Computational issues

2.4.1 Modal Responses under BWN Excitation

Measurements of structural responses, such as displacement, velocity and acceleration, are used for modal identification. For computational efficiency, closed form solutions of the covariance matrix of these responses to a BWN excitation are desired. Crandall and Mark [36] derived the variance of the displacement response under BWN excitation. However, to our knowledge, closed form solutions of the variances and covariances of the velocity and acceleration responses under BWN excitation are not available. Here, we derive these results by use of partial fraction integration. A similar approach was used by Der Kiureghian [117] for a stationary excitation and Harichandran [118] for a propagating excitation.

Assume each pair of the elements of the forces $\boldsymbol{f}$ or $\ddot{\boldsymbol{u}}_g$ in the modal equation of motion in Eqn. (2.2) are totally incoherent and each is modeled as a BWN with a PSD magnitude of $P_{0,k}$ and bandwidth of $\omega_{c,k}$. The cross-PSD matrix of the input forces then has the elements

$$P_{kl} = \begin{cases} \Phi_{0,k}\delta_{kl}, & if\ -\omega_{c,k} \le \omega \le \omega_{c,k} \\ 0, & elsewhere \end{cases} ; k,l = 1,2,\dots,N_f \tag{2.23}$$

in which δ_{kl} is the Kronecker delta. Under this input PSD, the covariances of the ith and the jth modal responses can be calculated by

$$\Gamma_{q_i^{(e)} q_j^{(f)}}(\tau) = (-1)^f I^{e+f} \sum_{k=1}^{N_f} \gamma_{ik}\gamma_{jk}\Phi_{0,k} \sum_{l=1}^{4} Res(\Omega_l) \int_{-\omega_{c,k}}^{\omega_{c,k}} \frac{\omega^{e+f} e^{I\omega\tau}}{\omega - \Omega_l} d\omega \tag{2.24}$$

for $e, f = \{0,1,2,3,4\}$, where $q_i^{(e)}$ represents the eth derivative of the ith modal displacement with respect to time, γ_{ik} stands for the ith modal participation factor with respect to the kth BWN excitation, $I = \sqrt{-1}$ is the imaginary unit, Ω_l and $Res(\Omega_l)$ are the lth pole and corresponding residual of $Q(\omega) = H_i(\omega)H_j^*(\omega)$, in which $H_i(\omega)$ is the frequency-response function (FRF) of mode i, and an asterisk denotes the complex conjugate. The detailed derivation and the closed form of the integral in Eqn. (16) are given in Appendix B.2.

2.4.2 Calculation of utility function

Recall that we need to evaluate high dimensional integrals over $\boldsymbol{X}_s$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ for calculation of the utility functions $U_1(d)$, $U_2(d)$ and $U_3(d)$. Obtaining closed-form solutions of these integrals or using numerical integration is impractical. Hence, we select to use the Monte Carlo sampling (MCS) technique to obtain an approximate

solution. MCS is relatively simple for our problem because independent priors are assumed for random variables representing the excitation, the structural parameters and the measurement error. For each realization of structural parameters, the mode shapes $\boldsymbol{X}_s$ are computed.

Let N_s denote the sample size. The utility functions are computed as follows:

$$U_1(d) \cong -1/N_s \sum_{i=1}^{N_s} \text{tr } \{\left[\boldsymbol{\Gamma}_{s,i}^{-1} + \boldsymbol{I}_N \otimes \boldsymbol{F}_{s,i}\right]^{-1}\} \tag{2.25}$$

$$U_2(d) \cong 1/N_s \sum_{i=1}^{N_s} \ln \det\left[\boldsymbol{\Gamma}_{s,i}^{-1} + \boldsymbol{I}_N \otimes \boldsymbol{F}_{s,i}\right] \tag{2.26}$$

$$U_3(d) \cong -1/N_s \sum_{i=1}^{N_s} \left\{0.5 \ln \det\left[\boldsymbol{\Gamma}_{s,i}^{-1} + \boldsymbol{I}_N \otimes \boldsymbol{F}_{s,i}\right] + \text{tr}\left[\boldsymbol{\Gamma}_{s,i}^{-1} + \boldsymbol{I}_N \otimes \boldsymbol{F}_{s,i}\right]^{-1}\left[\boldsymbol{\Gamma}_{s,i}^{-1} + \boldsymbol{I}_N \otimes \boldsymbol{F}_i\right]\right\} \tag{2.27}$$

where the subscript i on each term indicates the ith simulated value. Another concern arises from the high dimension of $\boldsymbol{\Gamma}_s$, which equals the product of the sampling frequency f_s, time duration T_d, and the number of target modes m. If this dimension is too high, the inverse and determinant operations on $\boldsymbol{\Gamma}_s$ will be unfeasible. If $\boldsymbol{\Gamma}$ is n by n, the memory required to store $\boldsymbol{\Gamma}_s$ is $O(n^2)$ and the number of floating-point operations (Flops) required for matrix inverse and determinant calculations is about $O(n^3)$. Therefore, for large n, it is necessary to develop an approximate solution method. Considering the block Toeplitz structure of the covariance matrix $\boldsymbol{\Gamma}_s$, there are two possible strategies available in the literature: the nearest block circulant matrix approximation [119] and the nearest Kronecker product approximation [120]. These approximations reduce the number of operations from $O(n^3)$ to $O(n^2 \log n)$ and $O(n_1^3 + n_2^3)$, respectively, where $n_1 + n_2 = n$. Because of the possibility of high reduction in the number of operations and its appealing algebra [121], we choose the nearest Kronecker product approximation. Analysis of the error in this approximation can be found in [122]. For the sake of completeness, a detailed treatment of this method is reported in Appendix B.3, where it is shown that we only need to evaluate a block diagonal matrix multiplication and get its trace in evaluating $U_1(d)$ and $U_3(d)$; as for $U_2(d)$, only a summation over diagonal elements of a diagonal matrix is needed. Among the three, evaluation of $U_2(d)$ is most efficient.

2.5 Empirical studies

2.5.1 Shear-type Building

The first example investigated is a 16-DOF shear-type building simplified as a mass-spring system, shown in Figure 2.2. The floor masses $m_1 \sim m_{16}$ are modeled as joint lognormally distributed random variables with different logarithmic means λ_m but identical coefficient of variation (COV) δ_m and cross-correlation coefficients $\rho_{\ln m}$. The story stiffnesses are modeled in a similar way. Modal damping ratios are assumed to be identically and independently lognormally distributed with a larger COV. The building is subjected to base motion and wind load, both of which are modeled as independent BWN processes with the same PSD magnitude of Φ_0 and bandwidth, $\omega_c/2\pi = 5\,Hz$. Elements of the load coefficient vector of the wind load are assumed to be independent normal random variables with mean $P = 10^6 \times 0.616(0.1z)^{0.44}$[123] and 5% C.O.V, in which $z = kh$, $k = 0, \ldots, 16$, and $h = 3$m is the floor height. The same INR $2\Phi_0\omega_c/\sigma^2$ is applied for the measurement of acceleration, velocity and displacement, and is modeled by a uniform distribution bounded between 1 and 100. This quantity is dimensionless for measurement of acceleration, but has units of sec^{-2} and sec^{-4} for velocity and displacement, respectively. Although acceleration is usually measured more accurately than displacement and velocity, for the sake of simplicity in this example values of INR for all three are assumed to be in the same level. All these variables are defined in Table 2.1. We also set the sampling frequency as $f_s = 50\,Hz$, length of duration as $T_d = 60s$ and MCS sampling size as $N_s = 1000$. Through a number of tests, we determined that a sample of size of $N_s = 1000$ is sufficient to ensure convergence of the estimates based on MCS within a COV smaller than 0.01.

The lowest four modes of the structure are of interest. The histograms of the natural frequencies for the 1000 realizations are shown in the upper part of Figure 2.3. The mean mode shapes (black lines) and 2 standard deviation intervals are shown in the bottom part. The distributions of the natural frequencies appear to be lognormal. This is because the floor masses and story stiffnesses were assigned joint lognormal distributions. The fundamental frequency has a sharper distribution than the other frequencies. Similarly, the variability in the mode shapes increases with the order number. The uncertainties in mode shapes influence the OSP through $\boldsymbol{X}_s$ shown in Eqn. (2.8), but the effects of modal frequencies on OSP are not straightforward. Modal frequencies influence the unknown modal response $\boldsymbol{\theta}_s$ and, therefore, the mean and covariance matrix of $\boldsymbol{\theta}_s$. Also, recall from Eqn. (2.24) that computation of the prior covariance matrix $\boldsymbol{\Gamma}$ depends on modal frequencies. Therefore, the influence of uncertainties in modal frequencies on the OSP is through $\boldsymbol{\Gamma}$.

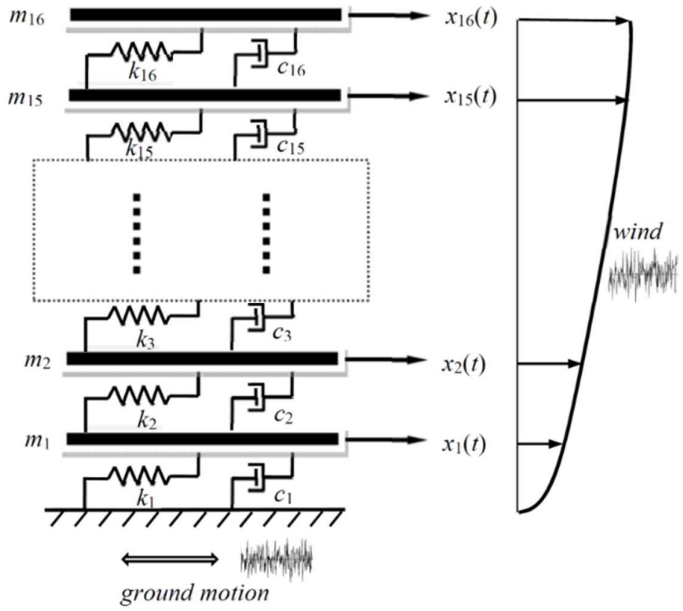


Figure 2.2 16-DOF shear-type building.

Table 2.1 Probabilistic model for example structure

Variable	Distribution	Mean	COV	Correlation
$m_1 \sim m_4$	Joint Lognormal	5×10^5	0.05	$\rho_{\ln m} = 0.6$
$m_5 \sim m_8$		4×10^5		
$m_9 \sim m_{12}$		3×10^5		
$m_{13} \sim m_{16}$		2×10^5		
$k_1 \sim k_4$	Joint Lognormal	1.6×10^8	0.1	$\rho_{\ln k} = 0.6$
$k_5 \sim k_8$		1.5×10^8		
$k_9 \sim k_{12}$		1.4×10^8		
$k_{13} \sim k_{16}$		1.3×10^8		
$\xi_1 \sim \xi_{16}$ (%)	Joint Lognormal	2	0.3	Independent
$2\Phi_0\omega_c/\sigma^2 \mid \omega_c$	Uniform(1,100)	50.5	0.515	Independent

The auto-covariance function (ACF) of the 1st modal acceleration and the cross-covariance function of the 1st and 2nd modal accelerations for one MCS realization are illustrated in Figure 2.4. Results obtained from the closed-form solution and from numerical integration match, verifying the closed-form solution. Another distinct observation is the large magnitude of the variance relative to the Fisher information $\boldsymbol{F}_s$ which is less than of order 10^2 in the example, implying large uncertainty in the modal acceleration responses. Given the large variance, it is expected that the inverse of the covariance matrix $\boldsymbol{\Gamma}$ will not make a significant

contribution to the selected utility functions when acceleration measurements are made.

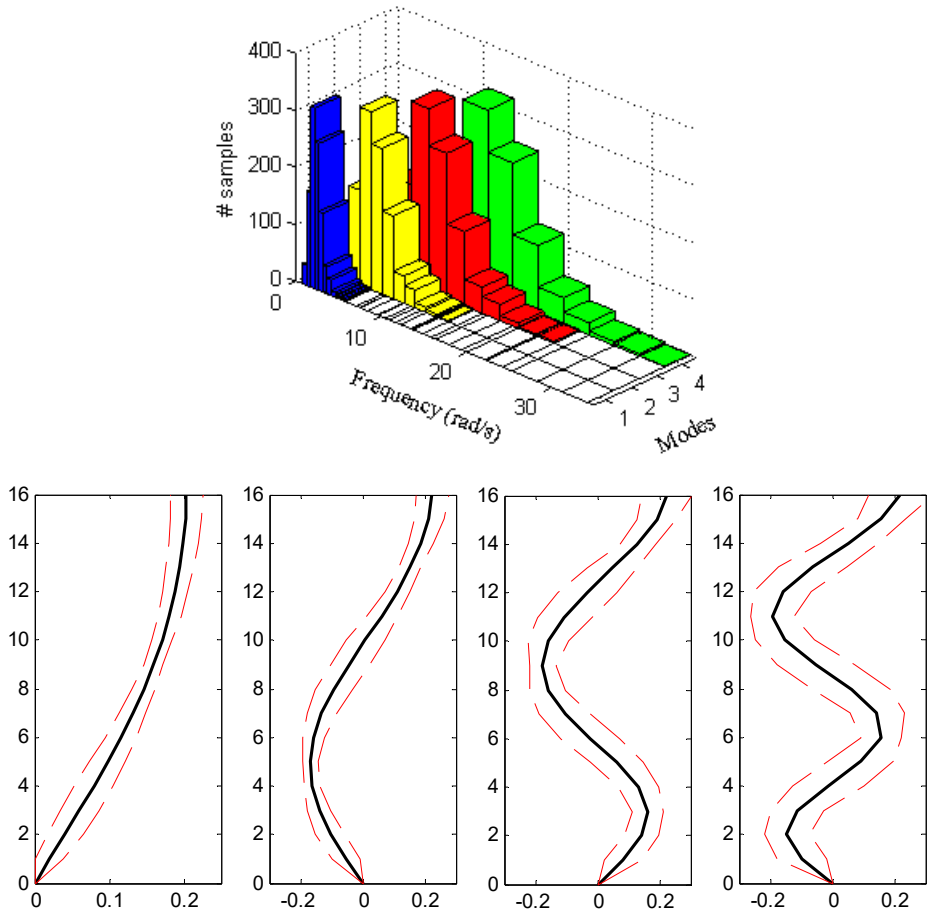


Figure 2.3 Variation of Modal Parameters.

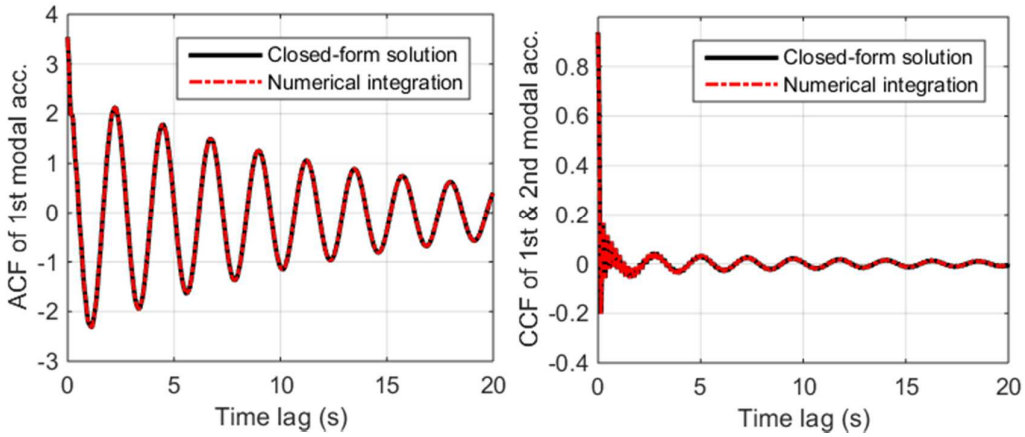


Figure 2.4 ACF and CCF for one MCS realization.

Non-Bayesian Placement of Sensors

In a non-Bayesian approach, i.e., without prior information, $\boldsymbol{\Gamma}^{-1} = \mathbf{0}$ and there is no difference in deploying displacement meter, velocimeter or accelerometer for modal identification. As a consequence, different types of sensors will occupy the same locations in a design scenario of mixed sensor placement. Therefore, for comparison, we first deploy sensors based on the nominal model and the probabilistic model considering the uncertainty in mode shapes only. Using the backward sequential method [28], the optimal configuration of sensors for the three utility functions is shown in Figure 2.5 and Figure 2.6. Because the Fisher information matrix becomes singular if the number of sensors is less than the number of modes, the placement starts from four sensors. The distinction between Figure 2.5 and Figure 2.6 lies in the consideration of the uncertainty in mode shapes in the latter. Comparing the two sets of figures, we see that the two methods give almost identical configurations for small number of sensors, but discrepancy between the results occurs for moderate number of sensors. The similarity within these configurations can be partly explained by the unbiasedness assumption on the nominal model, but it is hard to find a more reasonable model assumption in the design stage. The quadratic loss utility function $U_1(d)$ provides the same configuration for both scenarios, and this can be attributed to the linearity of the trace and mean operations; while the utility function $U_3(d)$ based on the K-L divergence is the most sensitive to the uncertainty in mode shapes, because it includes the combined nonlinearity of determinant and normalized trace terms.

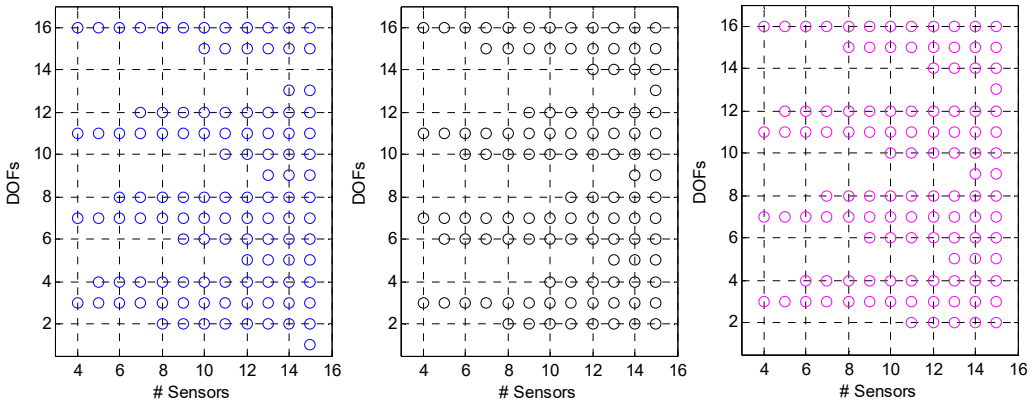


Figure 2.5 Configuration of sensors based on nominal model

Left to right: $U_1(d)$, $U_2(d)$, $U_3(d)$.

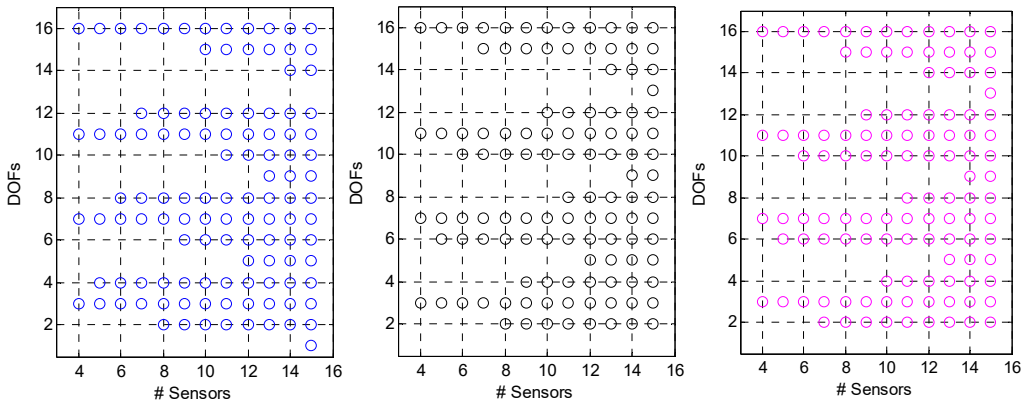


Figure 2.6 Configuration of sensors based on probabilistic model with uncertain mode shapes

Left to Right: $U_1(d)$, $U_2(d)$, $U_3(d)$

Placement of Displacement Meters

Next, we apply the proposed Bayesian method for displacement meters with the assumed prior distributions. The results are shown in Figure 2.7. Note that with this method we can deploy fewer sensors than the number of modes. This is because the covariance matrix, here consisting of not only the Fisher information matrix but also the prior covariance matrix, is not singular. Comparing with Figure 2.6, we see that the optimal configurations of sensors are influenced by the prior information. For example, more sensors are placed near the top of the structure and DOF 4 is now more preferred than DOF 3. The optimal sensor configurations are quite different under the three utility functions: the Shannon information is more influenced by prior information and tends to cluster sensors on the top, while the K-L divergence and quadratic loss give more similar configurations and distribute sensors more evenly. This is because both the latter utility functions involve similar trace terms, to which the FIM contributes more in this example.

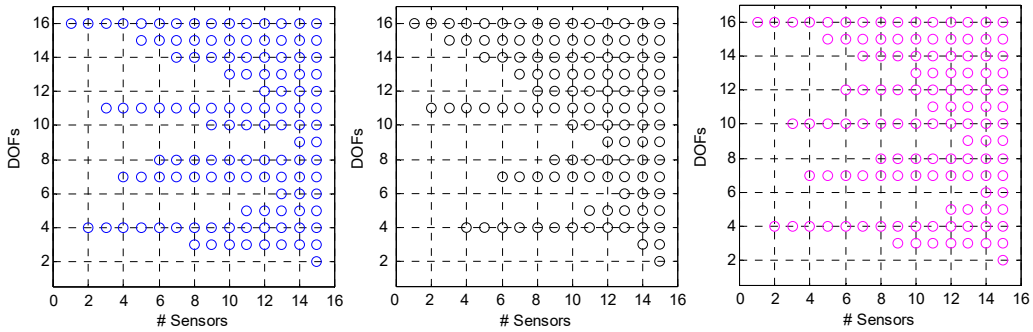


Figure 2.7 Configuration of displacement meters.

Left to Right: $U_1(d)$, $U_2(d)$, $U_3(d)$.

In order to check how the INR influences the optimal configuration of sensors, the configurations of displacement meters based on $U_2(d)$ are shown in Figure 2.8 under different values of INR. When the INR is extremely low (INR = 1), $\boldsymbol{\Gamma}_s^{-1}$ contributes more to the determinant, and all sensors cluster in the top of the building where the kinetic energy is higher. When the INR is high, the influence of $\boldsymbol{\Gamma}_s^{-1}$ diminishes and the FIM becomes dominant so that distributions of sensors are more even. Similar results are observed for the other two utility functions, but not so dramatically. Thus, the INR setting connects the kinetic energy-based and the FIM-based approaches, and a balanced configuration can be achieved by setting a reasonable INR.

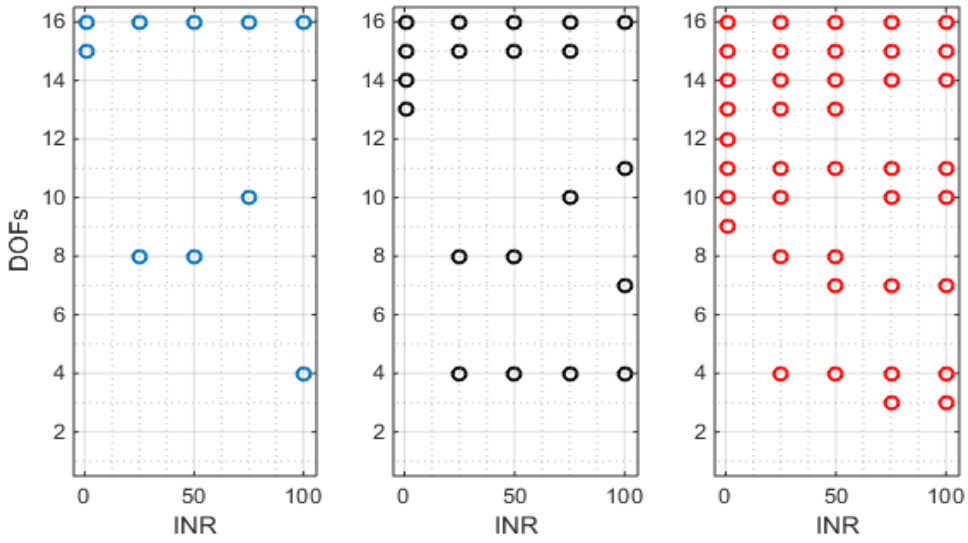


Figure 2.8 Relation between configurations of sensors and INR.

Placement of Velocimeters

Figure 2.9 shows the placement of velocimeters incorporating the prior covariance of modal velocity responses. Comparing with Figure 2.6 and Figure 2.9, we see that the prior information changes the configuration of velocimeters, but less severely than for the case of displacement meters. The reason is that the modal velocity has larger uncertainty than that of modal displacement, so that the inverse of the prior covariance matrix $\boldsymbol{\Gamma}_s^{-1}$ contributes less in determining the optimal locations. Optimal configurations of sensors differ for different utility functions, but DOFs 3, 6 or 7, 11 and 16 are selected in all configurations. Again, the K-L divergence yields a more even distribution of velocimeters.

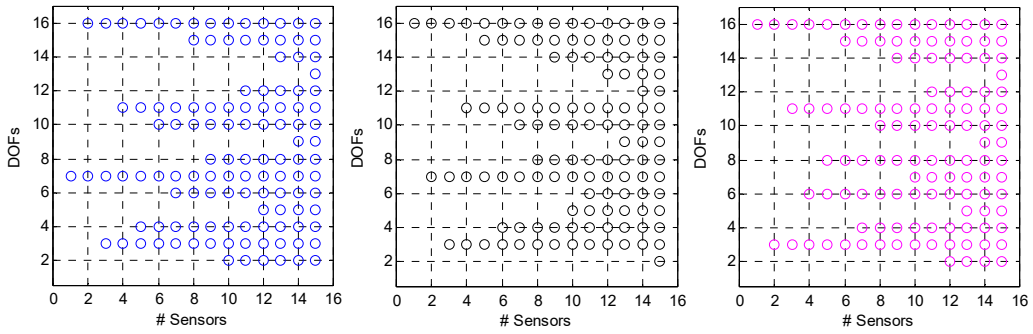


Figure 2.9 Configuration of velocimeters.
Left to Right: $U_1(d)$, $U_2(d)$, $U_3(d)$.

Placement of Accelerometers

Figure 2.10 shows the placement of accelerometers, while incorporating the prior information. The results for 4 or more sensors are almost identical to those in Figure 2.6. This is because the prior information does not contribute much to the utility function, as explained in the preceding section. Mathematically, this is because the inverse of $\boldsymbol{\Gamma}_S$ is negligible compared to the other terms in the utility functions. This observation justifies the use of non-Bayesian methods in the placement of accelerometers when the INR is high. Different utility functions give distinct configurations, but in all cases clusters occur around DOFs 3, 6 or 7, 11 and 16.

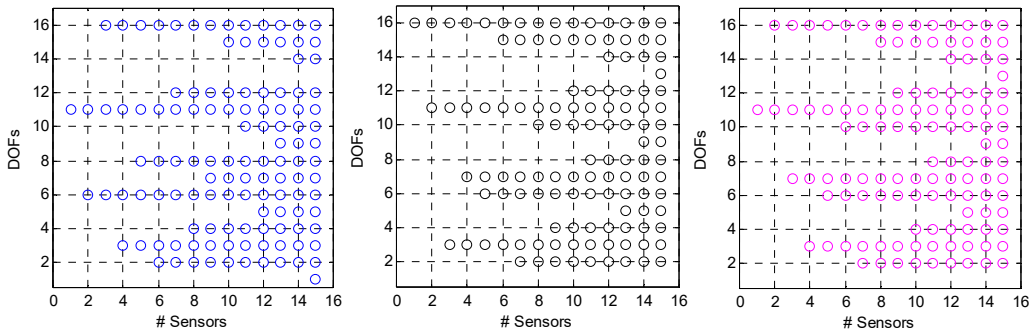


Figure 2.10 Configuration of accelerometers.
Left to Right: $U_1(d)$, $U_2(d)$, $U_3(d)$.

Mixed Placement of Displacement Meters and Accelerometers

Next, we consider a mixture of displacement meters and accelerometers for modal identification. The non-Bayesian method does not distinguish between different types of sensors, hence it cannot solve this problem. We consider a simple example of placing one displacement meter and three accelerometers to identify four modes. More general cases, e.g. optimizing the type and position of a fixed number of sensors,

can be solved in our framework; however, computational difficulties may be encountered in the optimization scheme and more work is needed before such a general problem can be solved. For our simple example, solutions obtained by cross entropy optimization [31] are listed in Table 2.2 for the three utility functions. The utility function based on Shannon information results in the displacement sensor and one accelerometer occupying the same location. This could be due to our neglecting the correlations between measurement errors, which has a larger influence in the case of the utility function based on Shannon information. The collocated displacement sensor and accelerometer may provide redundant information in identifying mode shapes, but this collocation may be desirable in practice in order to accurately identify the frequencies and damping ratios of lower modes from the low-frequency displacement data and the frequencies and damping ratios of higher modes from the high-frequency acceleration data, even though our algorithm cannot explicitly account for these aspects. The other two utility functions give identical configurations: the displacement sensor is placed at the top of the building, where the signal-to-noise ratio is highest, and the accelerometers are placed evenly to improve the identifiability of the mode shapes.

Table 2.2 Sensor location in mixed scenario.

Utility functions	Locations	
	Displacement meter	Accelerometer
$U_1(d)$	16	[5, 11,16]
$U_2(d)$	16	[4,9,13]
$U_3(d)$	16	[4,9,13]

2.5.2 Guangzhou TV Tower

The Guangzhou TV Tower (Figure 2.11), located in Guangzhou, China, is a super tall structure consisting of a 454 meter main tower and a 156 meter antenna mast. It has been developed as a benchmark problem for structural health monitoring [124], with optimal sensor placement being Task III in the program. Researchers [125] have used this structure as a test bed for investigating sensor placement algorithms. Here, we use this structure to illustrate the applicability of the proposed method.

A reduced finite element model (FEM) of the structure can be downloaded from http://www.cse.polyu.edu.hk/benchmark/. This FEM includes 27 elements for the main tower and 10 elements for the antenna mast. Nodes are numbered from 1 at the fixed base to 38 at the free top end. In order to obtain a robust design, the structural and force parameters are first randomized. Since the nominal element stiffness matrices provided are not positive semidefinite, the way used in Example 1 to generate the structural stiffness cannot guarantee its positive definiteness. In this paper,

we apply a two-step generating method: firstly, generate stiffness matrices modeled by multivariate normal distribution with means presented by the nominal values and COV 0.05; secondly, calculate the mean and covariance of the generated stiffness matrices and use these parameters to generate stiffness matrices following the Wishart distribution [126]. Element masses are modeled by independent lognormal distributions with means represented by the nominal values and a COV of 0.02. Classical damping is assumed, and modal damping ratios are assigned the same distribution as in the first example. The input force model is identical with that in the first example, and both the ground motion and wind load are considered. The other parameters, i.e., the sampling frequency, the time duration, and the number of Monte Carlo samples are identical to those assumed for the first example.

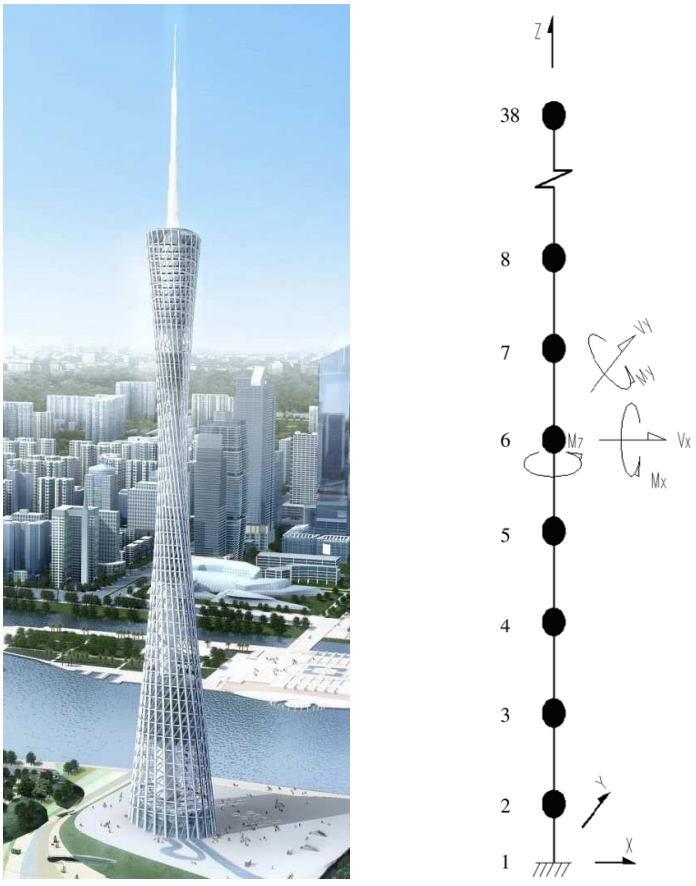


Figure 2.11 Guangzhou TV Tower: Overview and reduced FE model.

The Guangzhou TV Tower is a typical primary-secondary structure. The main tower is made of a reinforced concrete interior tube and steel external tube, while the antenna mast is made of steel. Relative to the main tower, the antenna mast is lighter and more flexible. As a result, the mode shapes of the combined system are expected

to have much larger amplitudes for DOFs associated with the antenna mast than those associated with the main tower. Furthermore, it is expected that the antenna mast will have relatively insignificant influence on the dynamic response of the tower. In the following, we consider two cases for sensor placement: one is to retain the antenna DOFs as possible locations for sensors, in which case we obtain the OSP solution to identify the modes of the combined system, and the other is to disallow placement of sensors at the DOFs of the antenna, in which case the OSP solution aims at identifying the dominant modes of the tower. Because rotational DOFs are difficult to measure, only translational DOFs in X and Y directions are considered for possible sensor installation. Furthermore, only the placement of accelerometers is considered since that is the most commonly used sensor for modal identification.

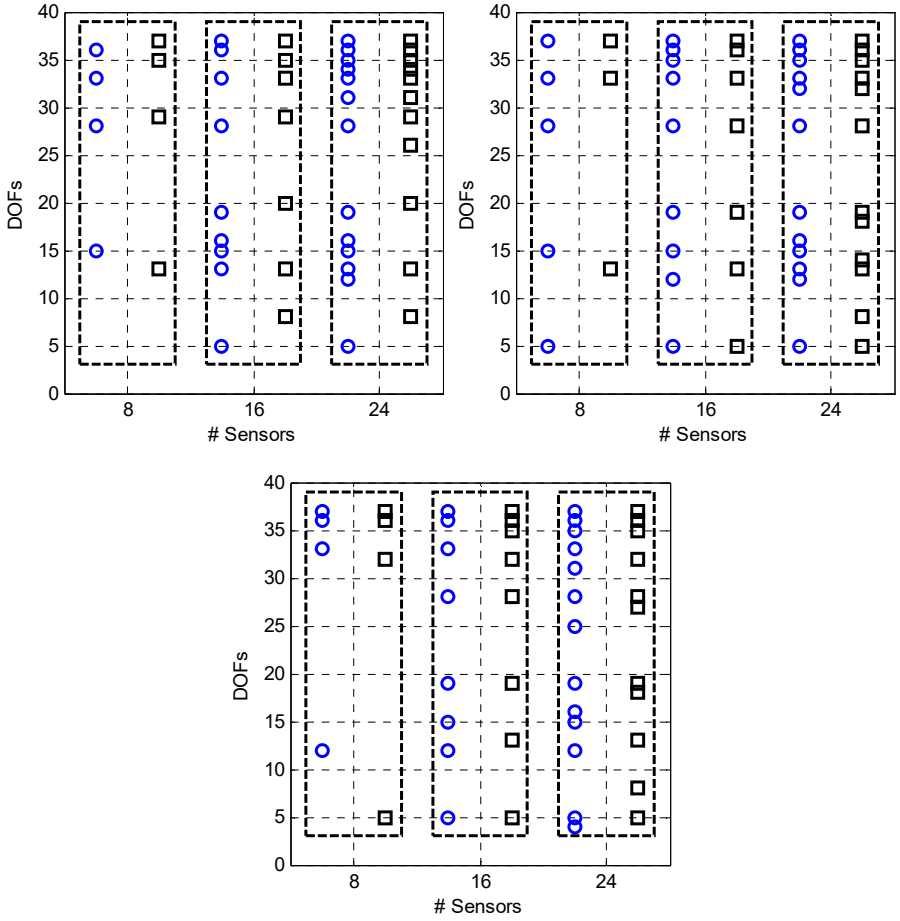


Figure 2.12 Configuration of accelerometers.
Left to Right: $U_1(d)$, $U_2(d)$, $U_3(d)$; ○: X direction; □: Y direction

Main Tower with Antenna Mast

In this case, the first 15 modes are selected as the modes of interest. The sensor configurations under different utility functions are shown in Figure 2.12. It is seen that more than half the sensors are clustered in the antenna mast, showing its importance in modeling the dynamical behavior of the combined system. Because for each mode shape the amplitudes at DOFs associated with the antenna mast are much larger than those of the main tower, these locations give a relatively large FIM. By checking the inverse of the covariance matrix of the modal accelerations, it is found that the magnitude of this matrix is much smaller than that of the FIM, so that the FIM dominates the optimization. In addition, due to the asymmetry of the structure, the configurations of accelerometers are not identical in the X and Y directions; indeed, even the number of accelerometers in many cases is not the same. There are significant differences in the configurations of sensors under different utility functions, but it is hard to say which one is better unless more evaluation criteria are included.

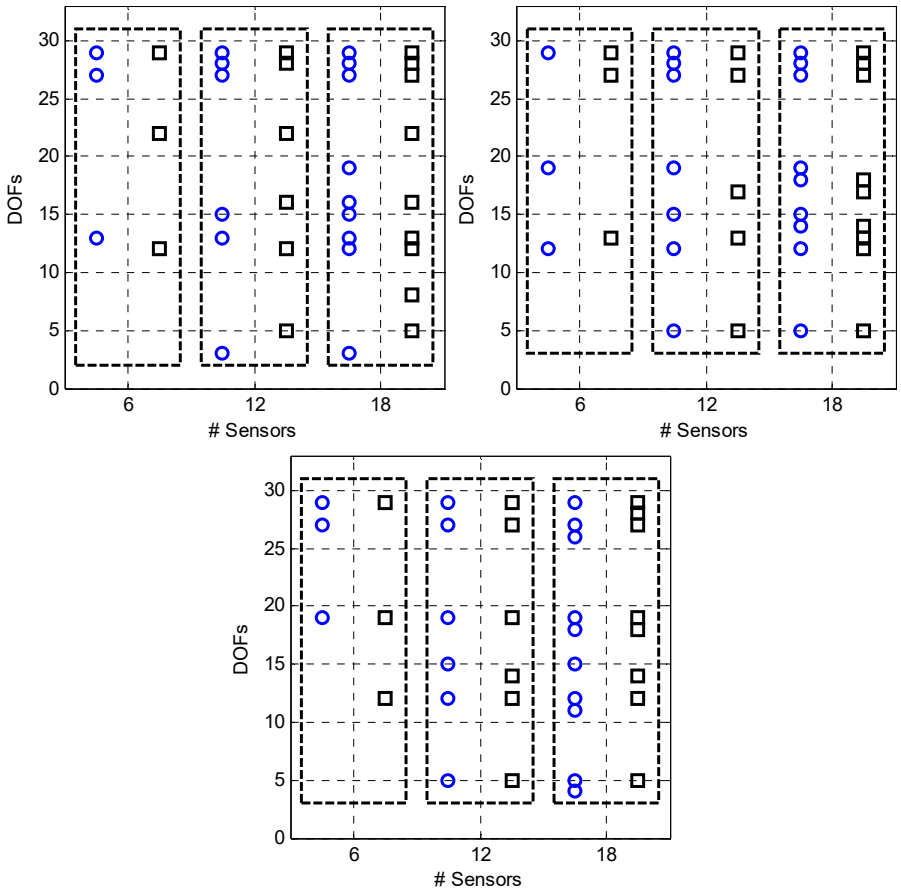


Figure 2.13 Configuration of accelerometers without antenna mast

Left to Right: $U_1(d)$, $U_2(d)$, $U_3(d)$; ○: X direction; □: Y direction

Main Tower without Antenna Mast

If the engineer is more concerned about the modal properties of the main tower, which are more complex and uncertain than those of the antenna mast, then it is advisable to exclude the antenna mast from candidate sensor locations. Therefore, we now consider placing sensors only on the main tower.

There are 11 vibration modes of the main tower in the first 15 modes of the combined system, and they are selected as the target modes. It is noted that a FEM of only the main tower was used for OSP in [125], but this way of modeling is not appropriate because antenna mast influences the mode shapes of main tower. In fact, a different structure is modeled if the antenna mast is totally ignored. Here, the FEM of the full structure is used, but only the DOFs corresponding to the main tower are selected as the potential locations of sensors. The optimal configurations of sensors are listed in Figure 2.13 for 6, 12 and 18 accelerometers. Some similarities exist between the results for this case and the case considering the antenna mast for the tower DOFs. In particular, the number and locations of accelerometers are not identical in the X and Y directions because of the asymmetry of the structure. However, some configurations appear to be more regular in this case, e.g., the placement of 12 accelerometers based on the K-L Divergence utility function. An interesting observation based upon all examples in the paper is that the sparsity of the configuration depends on the relation between the number of available sensors n and the number of modes of interest m: the configuration does not cluster together if n is not much larger than m; otherwise clusters may occur when n increases.

The relation between values of the utility functions and the number of sensors is shown in Figure 2.14. It can be seen that the utility functions monotonically increase with the number of sensors. The rate of increase is particularly high when the number of sensors is smaller than the number of modes of interest. This implies that the number of sensors to be placed should be at least equal to or greater than the number of modes to be identified. When the antenna mast is removed as a potential sensor location, significant drops occur in all utility function values. Obviously, optimal configurations without the antenna mast are inferior to those considering all the DOFs. When the number of sensors is small, the values of utility function $U_1(d)$ with or without the antenna mast are only slightly different. This is because $\boldsymbol{\Gamma}_s^{-1}$ dominates the trace operation for this utility function, not the FIM term where the difference between the two scenarios lies. Utility function $U_3(d)$ shows the largest drop in value when removing the antenna mast as a potential sensor location. This is because all available DOFs are used as the full set configuration in the process of computing $U_3(d)$.

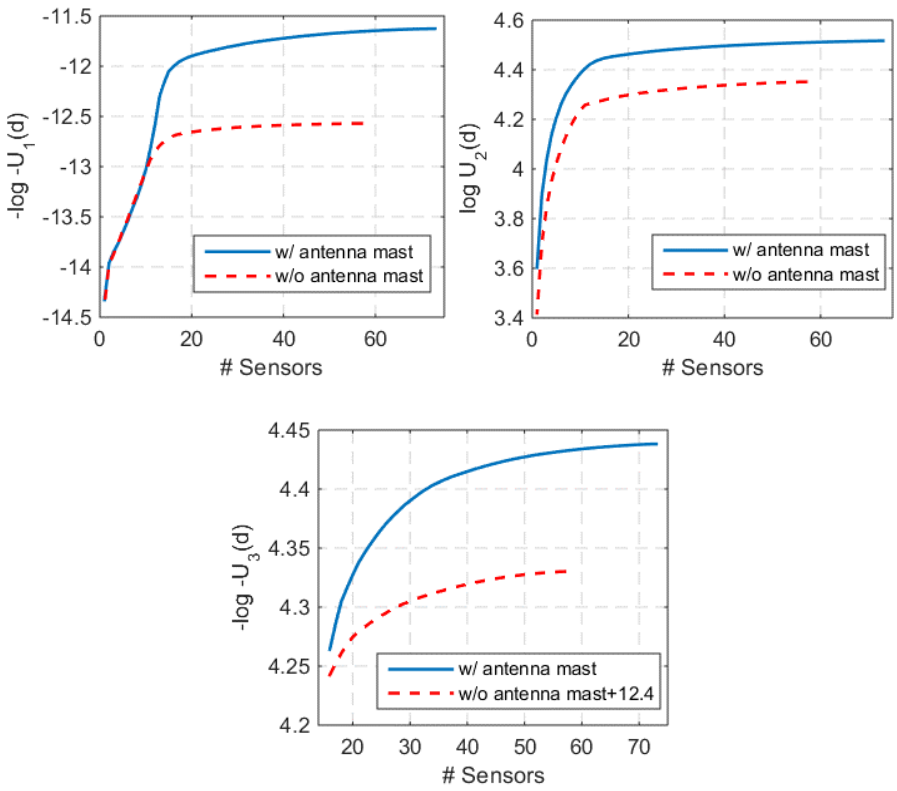


Figure 2.14 Relations between utilities and numbers of sensors.
Left to Right: $U_1(d)$, $U_2(d)$, $U_3(d)$

2.6 Conclusions

A Bayesian probabilistic model for sensor placement considering model and load uncertainty as well as the measurement error is formulated. The maximum expected utility theory is employed to select optimal sensor locations for modal response estimation, and this optimal configuration works as a suboptimum for the OSP in OMA. Three utility functions are considered, each based on a different motivation. The covariance of modal responses is theoretically derived, and its nearest Kronecker product approximation is developed for fast evaluation of the utility functions. Conclusions derived from two cases studies are as follows:

(1) The OSP is essentially a decision problem under uncertainty. The proposed framework consisting of Bayesian statistics and the maximum expected utility criterion provides a powerful tool to address all relevant uncertainties in the sensor placement problem and to incorporate all available prior information. By specifying a suitable utility function, a robust optimal configuration of sensors can be obtained.

(2) When accounting for prior information, the optimal placement configuration of displacement meter, velocimeter and accelerometer can be different; indeed, mixed sensor placement becomes possible. Since the prior covariance matrix of modal acceleration has a large magnitude, the placement of accelerometer is less influenced by the prior information. This justifies the commonly used mode-shape-based accelerometer placement, since the prior information of modal acceleration does not strongly influence the final result.

(3) The magnitude of INR has a great influence on the optimal configuration of sensors. If it is small, the magnitude of prior covariance of modal responses is also small; the result is that sensors are placed at positions with high kinetic energy; if the magnitude of the INR is large, the prior information could be ignored, and the configuration is fully determined by the FIM. Thus, the magnitude of the INR determines the balance between the kinetic energy and Fisher information.

(4) Model uncertainties considered in the example applications do not significantly influence the optimal configuration of sensors, especially when the number of sensors available is small. This observation is partly due to the fact that the nominal structural model is assumed to be unbiased; the case of a biased model is left for future study.

(5) Three utility functions are introduced based on quadratic loss, Shannon information and K-L divergence. As expected, different utility functions yield different configurations of sensors. The utility function based on Shannon information is most influenced by the prior information, and its computation is the most efficient; the K-L divergence based utility function gives relatively even distribution of sensors. As for how to select the utility function, more evaluating criteria are needed.

(6) The number of sensors placed should be at least equal to or more than the number of modes of interest. In this paper we assume both numbers are predetermined. In fact, these two parameters can be also incorporated into the maximum expected utility framework by specifying suitable utility functions.

Appendix B

B.1 Derivation of K-L divergence as utility function

In this appendix, we derive the expression for the utility function when employing the expected K-L divergence.

Following the notations in Section 2.2, the following identities hold:

$$\boldsymbol{Y}_{rs} = (\boldsymbol{I}_N \otimes \boldsymbol{S})\boldsymbol{Y}_r;\ \ \boldsymbol{X}_s = (\boldsymbol{I}_N \otimes \boldsymbol{S})\boldsymbol{X}; \tag{2.28}$$

$$\boldsymbol{W}_s = (\boldsymbol{I}_N \otimes \boldsymbol{S})\boldsymbol{W};\ \boldsymbol{\Sigma}_s = \boldsymbol{S}\boldsymbol{\Sigma}\boldsymbol{S}^{\mathrm{T}}$$

Similar to Eqns. (2.11) and (2.12), we obtain the posterior expectation and covariance of $\boldsymbol{\Theta}_s$ given $\boldsymbol{Y}_r$, $\boldsymbol{X}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ as

$$\begin{aligned}&\mathrm{E}[\boldsymbol{\Theta}_s|\boldsymbol{Y}_r,\boldsymbol{X},\boldsymbol{\Gamma},\boldsymbol{\Sigma}] = \boldsymbol{\Gamma}\boldsymbol{X}^{\mathrm{T}}[\boldsymbol{X}\boldsymbol{\Gamma}_s\boldsymbol{X}^{\mathrm{T}} + \boldsymbol{I}_N \otimes \boldsymbol{\Sigma}]^{-1}\boldsymbol{Y}_r \\ &= [\boldsymbol{\Gamma}_s^{-1} + \boldsymbol{X}^{\mathrm{T}}(\boldsymbol{I}_N \otimes \boldsymbol{\Sigma}^{-1})\boldsymbol{X}]\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{I}_N \otimes \boldsymbol{\Sigma}^{-1})\boldsymbol{Y}_r\end{aligned} \tag{2.29}$$

$$\begin{aligned}&\mathrm{Cov}[\boldsymbol{\Theta}_s|\boldsymbol{Y}_r,\boldsymbol{X},\boldsymbol{\Gamma},\boldsymbol{\Sigma}] = \boldsymbol{\Gamma} - \boldsymbol{\Gamma}\boldsymbol{X}^{\mathrm{T}}[\boldsymbol{X}\boldsymbol{\Gamma}\boldsymbol{X}^{\mathrm{T}} + \boldsymbol{I}_N \otimes \boldsymbol{\Sigma}]^{-1}\boldsymbol{X}\boldsymbol{\Gamma}^{\mathrm{T}} \\ &= [\boldsymbol{\Gamma}^{-1} + \boldsymbol{X}^{\mathrm{T}}(\boldsymbol{I}_N \otimes \boldsymbol{\Sigma}^{-1})\boldsymbol{X}]^{-1}\end{aligned} \tag{2.30}$$

Since posterior distributions of $\boldsymbol{\Theta}_s$ for both cases follow multivariate normal distribution with means and covariances as given in Eqns. (2.11) and (2.12) and the above, the K-L divergence between them is given by (see [112]):

$$\begin{aligned}\mathrm{KL}[\boldsymbol{\Theta}_s^{(s)}||\boldsymbol{\Theta}_s] = \frac{1}{2}\Big[&\ln\frac{\det\boldsymbol{\Sigma}_{\boldsymbol{\Theta}}}{\det\boldsymbol{\Sigma}_{\boldsymbol{\Theta}}^{(s)}} + \mathrm{tr}\left[\boldsymbol{\Sigma}_{\boldsymbol{\Theta}}^{(s)}\left(\boldsymbol{\Sigma}_{\boldsymbol{\Theta}}^{-1} - \boldsymbol{\Sigma}_{\boldsymbol{\Theta}}^{(s)-\mathbf{1}}\right)\right] + (\boldsymbol{\mu}_{\boldsymbol{\Theta}}^{(s)} - \\ &\boldsymbol{\mu}_{\boldsymbol{\Theta}})^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{\Theta}}^{-1}(\boldsymbol{\mu}_{\boldsymbol{\Theta}}^{(s)} - \boldsymbol{\mu}_{\boldsymbol{\Theta}})\Big]\end{aligned} \tag{2.31}$$

When inserting the preceding equation into Eqn. (2.21), the first two terms do not involve $\boldsymbol{Y}_r$; therefore, we only consider the third term, and its expectation with respect to the conditional distribution of $\boldsymbol{Y}_r$ for given $\boldsymbol{X}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ and derive it as follows:

$$\begin{aligned}&\mathrm{E}\Big[(\boldsymbol{\mu}_{\boldsymbol{\Theta}}^{(s)} - \boldsymbol{\mu}_{\boldsymbol{\Theta}})^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{\Theta}}^{-1}(\boldsymbol{\mu}_{\boldsymbol{\Theta}}^{(s)} - \boldsymbol{\mu}_{\boldsymbol{\Theta}})\Big] = \mathrm{tr}\left\{\mathrm{E}\Big[(\boldsymbol{\mu}_{\boldsymbol{\Theta}}^{(s)} - \boldsymbol{\mu}_{\boldsymbol{\Theta}})(\boldsymbol{\mu}_{\boldsymbol{\Theta}}^{(s)} - \boldsymbol{\mu}_{\boldsymbol{\Theta}})^{\mathrm{T}}\Big]\boldsymbol{\Sigma}_{\boldsymbol{\Theta}}^{-1}\right\} \\ &= \mathrm{tr}\left\{\mathrm{E}\Big[\boldsymbol{\mu}_{\boldsymbol{\Theta}}^{(s)}\boldsymbol{\mu}_{\boldsymbol{\Theta}}^{(s)\mathrm{T}}\Big]\boldsymbol{\Sigma}_{\boldsymbol{\Theta}}^{-1}\right\} - 2\mathrm{tr}\left\{\mathrm{E}\Big[\boldsymbol{\mu}_{\boldsymbol{\Theta}}^{(s)}\boldsymbol{\mu}_{\boldsymbol{\Theta}}^{\mathrm{T}}\Big]\boldsymbol{\Sigma}_{\boldsymbol{\Theta}}^{-1}\right\} + \mathrm{tr}\{\mathrm{E}\big[\boldsymbol{\mu}_{\boldsymbol{\Theta}}\boldsymbol{\mu}_{\boldsymbol{\Theta}}^{\mathrm{T}}\big]\boldsymbol{\Sigma}_{\boldsymbol{\Theta}}^{-1}\} \\ &= \mathrm{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{\Theta}}^{(s)}\boldsymbol{\Sigma}_{\boldsymbol{\Theta}}^{-1}\right) - m\end{aligned} \tag{2.32}$$

in which,

$$\begin{aligned}&\mathrm{tr}\left\{\mathrm{E}\Big[\boldsymbol{\mu}_{\boldsymbol{\Theta}}^{(s)}\boldsymbol{\mu}_{\boldsymbol{\Theta}}^{(s)\mathrm{T}}\Big]\boldsymbol{\Sigma}_{\boldsymbol{\Theta}}^{-1}\right\} \\ &= \mathrm{tr}\{\boldsymbol{\Gamma}_s\boldsymbol{X}_s^{\mathrm{T}}[\boldsymbol{X}_s\boldsymbol{\Gamma}_s\boldsymbol{X}_s^{\mathrm{T}} + \boldsymbol{I}_N \otimes \boldsymbol{\Sigma}_s]^{-1}\mathrm{E}[\boldsymbol{Y}_{rs}\boldsymbol{Y}_{rs}^{\mathbf{T}}][\boldsymbol{X}_s\boldsymbol{\Gamma}_s\boldsymbol{X}_s^{\mathrm{T}} + \boldsymbol{I}_N \otimes \boldsymbol{\Sigma}_s]^{-1}\boldsymbol{X}_s\boldsymbol{\Gamma}_s\boldsymbol{\Sigma}_{\boldsymbol{\Theta}}^{-1}\} \\ &= \mathrm{tr}\{\boldsymbol{\Gamma}_s\boldsymbol{X}_s^{\mathrm{T}}[\boldsymbol{X}_s\boldsymbol{\Gamma}_s\boldsymbol{X}_s^{\mathrm{T}} + \boldsymbol{I}_N \otimes \boldsymbol{\Sigma}_s]^{-1}\boldsymbol{X}_s\boldsymbol{\Gamma}_s\boldsymbol{\Sigma}_{\boldsymbol{\Theta}}^{-1}\} = \mathrm{tr}\Big[(\boldsymbol{\Gamma}_s - \boldsymbol{\Sigma}_{\boldsymbol{\Theta}}^{(s)})\boldsymbol{\Sigma}_{\boldsymbol{\Theta}}^{-1}\Big]\end{aligned} \tag{2.33}$$

$$\mathrm{tr}\left\{E\Big[\boldsymbol{\mu}_{\boldsymbol{\Theta}}^{(s)}\boldsymbol{\mu}_{\boldsymbol{\Theta}}^{\mathrm{T}}\Big]\boldsymbol{\Sigma}_{\boldsymbol{\Theta}}^{-1}\right\} \tag{2.34}$$

$$= \mathrm{tr}\{\boldsymbol{\Gamma}_s \boldsymbol{X}_s^{\mathrm{T}}[\boldsymbol{X}_s \boldsymbol{\Gamma}_s \boldsymbol{X}_s^{\mathrm{T}} + \boldsymbol{I}_N \otimes \boldsymbol{\Sigma}_s]^{-1} \mathrm{E}[\boldsymbol{Y}_{rs} \boldsymbol{Y}_r^{\mathrm{T}}][\boldsymbol{X}\boldsymbol{\Gamma}_s \boldsymbol{X}^{\mathrm{T}} + \boldsymbol{I}_N \otimes \boldsymbol{\Sigma}]^{-1} \boldsymbol{X}\boldsymbol{\Gamma}_s \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\}$$
$$= \mathrm{tr}\{\boldsymbol{\Gamma}_s \boldsymbol{X}_s^{\mathrm{T}}[\boldsymbol{X}_s \boldsymbol{\Gamma}_s \boldsymbol{X}_s^{\mathrm{T}} + \boldsymbol{I}_N \otimes \boldsymbol{\Sigma}_s]^{-1} (\boldsymbol{I}_N \otimes \boldsymbol{S}) \mathrm{E}[\boldsymbol{Y}_r \boldsymbol{Y}_r^{\mathrm{T}}][\boldsymbol{X}\boldsymbol{\Gamma}_s \boldsymbol{X}^{\mathrm{T}} + \boldsymbol{I}_N \otimes \boldsymbol{\Sigma}]^{-1} \boldsymbol{X}\boldsymbol{\Gamma}_s \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\}$$
$$= \mathrm{tr}\{\boldsymbol{\Gamma}_s \boldsymbol{X}_s^{\mathrm{T}}[\boldsymbol{X}_s \boldsymbol{\Gamma}_s \boldsymbol{X}_s^{\mathrm{T}} + \boldsymbol{I}_N \otimes \boldsymbol{\Sigma}_s]^{-1} (\boldsymbol{I}_N \otimes \boldsymbol{S}) \boldsymbol{X}\boldsymbol{\Gamma}_s \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\}$$
$$= \mathrm{tr}\{\boldsymbol{\Gamma}_s \boldsymbol{X}_s^{\mathrm{T}}[\boldsymbol{X}_s \boldsymbol{\Gamma}_s \boldsymbol{X}_s^{\mathrm{T}} + \boldsymbol{I}_N \otimes \boldsymbol{\Sigma}_s]^{-1} \boldsymbol{X}_s \boldsymbol{\Gamma}_s \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\} = \mathrm{tr}\left[(\boldsymbol{\Gamma}_s - \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(s)}) \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\right]$$

$$\begin{aligned} &\mathrm{tr}\{E\left[\boldsymbol{\mu}_{\boldsymbol{\theta}} \boldsymbol{\mu}_{\boldsymbol{\theta}}^{\mathrm{T}}\right] \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\} \\ &= \mathrm{tr}\{\boldsymbol{\Gamma}_s \boldsymbol{X}^{\mathrm{T}}[\boldsymbol{X}\boldsymbol{\Gamma}_s \boldsymbol{X}^{\mathrm{T}} + \boldsymbol{I}_N \otimes \boldsymbol{\Sigma}]^{-1} \mathrm{E}[\boldsymbol{Y}_r \boldsymbol{Y}_r^{\mathrm{T}}][\boldsymbol{X}\boldsymbol{\Gamma}_s \boldsymbol{X}^{\mathrm{T}} + \boldsymbol{I}_N \otimes \boldsymbol{\Sigma}]^{-1} \boldsymbol{X}\boldsymbol{\Gamma}_s \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\} \\ &= \mathrm{tr}\{\boldsymbol{\Gamma}_s \boldsymbol{X}^{\mathrm{T}}[\boldsymbol{X}\boldsymbol{\Gamma}_s \boldsymbol{X}^{\mathrm{T}} + \boldsymbol{I}_N \otimes \boldsymbol{\Sigma}]^{-1} \boldsymbol{X}\boldsymbol{\Gamma}_s \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\} = \mathrm{tr}[(\boldsymbol{\Gamma}_s - \boldsymbol{\Sigma}_{\boldsymbol{\theta}}) \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}] \end{aligned} \tag{2.35}$$

Substituting all these expressions into Eqn. (2.21) gives

$$\begin{aligned} &U_3(d) \\ &= \mathrm{E}_{\boldsymbol{X}\boldsymbol{\Gamma}\boldsymbol{\Sigma}}[\tfrac{1}{2} \ln \det \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(s)} - \mathrm{tr}(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(s)} \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1})] \\ &= -\mathrm{E}_{\boldsymbol{X}\boldsymbol{\Gamma}\boldsymbol{\Sigma}}[0.5 \ln \det[\boldsymbol{\Gamma}_s^{-1} + \boldsymbol{I}_N \otimes \boldsymbol{F}_s] + \mathrm{tr}\,\{[\boldsymbol{\Gamma}_s^{-1} + \boldsymbol{I}_N \otimes \boldsymbol{F}_s]^{-1}[\boldsymbol{\Gamma}_s^{-1} + \boldsymbol{I}_N \otimes \boldsymbol{F}]\}] \end{aligned} \tag{2.36}$$

B.2 Computing the covariance matrix of modal responses

Under the input PSD in Eqn.(15), the covariances of the ith and the jth modal responses can be calculated by inverse Fourier transform of the corresponding cross-PSD functions, i.e.,

$$\Gamma_{q_i^{(e)} q_j^{(f)}}(\tau) = (-1)^f I^{e+f} \sum_{k=1}^{N_f} \int_{-\omega_{c,k}}^{\omega_{c,k}} \gamma_{ik} \gamma_{jk} \Phi_{0,k} H_i(\omega) H_j^*(\omega) \omega^{e+f} e^{I\omega\tau} d\omega \tag{2.37}$$

for $e, f = \{0,1,2,3,4\}$, where

$$H_i(\omega) = \frac{1}{\omega_i^2 - \omega^2 + 2I\xi_i \omega_i \omega};\ H_j^*(\omega) = \frac{1}{\omega_j^2 - \omega^2 - 2I\xi_j \omega_j \omega} \tag{2.38}$$

In order to evaluate this integral, we first make the partial fraction decomposition of $Q(\omega) = H_i(\omega) H_j^*(\omega)$ based on its poles and residues. For $\xi_i < 1$, this function has four poles so that using partial fractions decomposition one can write

$$Q(\omega) = \sum_{l=1}^{4} \frac{Res(\Omega_l)}{\omega - \Omega_l} \tag{2.39}$$

where $Res(\Omega_l)$ is the residue corresponding to pole Ω_l and has the following form:

$$\begin{cases} \Omega_1 = \omega_i\sqrt{1-\xi_i^2} + I\xi_i\omega_i; \quad Res(\Omega_1) = \frac{1}{(\Omega_1-\Omega_2)(\Omega_1-\Omega_3)(\Omega_1-\Omega_4)} \\ \Omega_2 = -\omega_i\sqrt{1-\xi_i^2} + I\xi_i\omega_i; \; Res(\Omega_2) = \frac{1}{(\Omega_2-\Omega_1)(\Omega_2-\Omega_3)(\Omega_2-\Omega_4)} \\ \Omega_3 = \omega_j\sqrt{1-\xi_j^2} - I\xi_j\omega_j; \quad Res(\Omega_3) = \frac{1}{(\Omega_3-\Omega_1)(\Omega_3-\Omega_2)(\Omega_3-\Omega_4)} \\ \Omega_4 = -\omega_j\sqrt{1-\xi_j^2} - I\xi_j\omega_j; \; Res(\Omega_4) = \frac{1}{(\Omega_4-\Omega_1)(\Omega_4-\Omega_2)(\Omega_4-\Omega_3)} \end{cases} \tag{2.40}$$

Then, substituting the partial fraction decomposition of $Q(\omega)$ gives Eqn. (2.24), and the needed integral has a closed form solution as follows:

If $\tau = 0$,

$$\begin{aligned} \int_{-\omega_{c,k}}^{\omega_{c,k}} \frac{\omega^n e^{I\omega\tau}}{\omega-\Omega_l} d\omega &= \int_{-\omega_{c,k}}^{\omega_{c,k}} \frac{\omega^n}{\omega-\Omega_l} d\omega \\ &= \Omega_l^n \ln\frac{\omega_{c,k}-\Omega_l}{-\omega_{c,k}-\Omega_l} + \sum_{m=1}^{n} \frac{\Omega_l^{n-m}}{m}\left[\omega_{c,k}^m - (-1)^m\omega_{c,k}^m\right] \end{aligned} \tag{2.41}$$

If $\tau \neq 0$,

$$\begin{aligned} \int_{-\omega_{c,k}}^{\omega_{c,k}} \frac{\omega^n e^{I\omega\tau}}{\omega-\Omega_l} d\omega = \Omega_l^n e^{I\Omega_l\tau}\left\{Ei\left[I\tau\left(\omega_{c,k}-\Omega_l\right)\right] - Ei\left[-I\tau\left(\omega_{c,k}+\Omega_l\right)\right]\right\} + \\ \sum_{m=1}^{n} \Omega_l^{m-1} \int_{-\omega_{c,k}}^{\omega_{c,k}} \omega^{n-m} e^{I\omega\tau} d\omega \end{aligned} \tag{2.42}$$

where

$$\begin{aligned} \int_{-\omega_{c,k}}^{\omega_{c,k}} \omega^{n-m} e^{I\omega\tau} d\omega = -I\sum_{p=0}^{n-m} \frac{(n-m)!}{(n-m-p)!} \frac{e^{Ip\pi/2}}{\tau^{p+1}}\left[\omega_{c,k}^{n-m-p} e^{I\tau\omega_{c,k}} - \right. \\ \left. (-1)^{n-m-p}\omega_{c,k}^{n-m-p} e^{-I\tau\omega_{c,k}}\right] \end{aligned} \tag{2.43}$$

In the above, $\ln(z)$ denotes the multivalued complex logarithmic function and $Ei(z)$ is the exponential integral function

$$Ei(z) = \int_{-\infty}^{z} \frac{e^t}{t} dt \tag{2.44}$$

for any complex z not equal to zero and not on the positive real axis. Note that $Ei(z)$ is a multivalued function whose Cauchy principal value jumps by $2\pi i$ when the negative real axis is crossed. Accordingly, $2\pi i$ must be deducted from the difference between the Cauchy principal value of $Ei\left[I\tau\left(\omega_{c,k}-\Omega_l\right)\right]$ and

$Ei\left[-I\tau\left(\omega_{c,k}+\Omega_l\right)\right]$, if $-I\tau\left(\omega_{c,k}+\Omega_l\right)$ locates in the 3th quadrant and $I\tau\left(\omega_{c,k}-\Omega_l\right)$ in the 2nd quadrant. It is easy to show that $I\tau\left(\omega_{c,k}-\Omega_l\right)$ and $-I\tau(\omega_c+\Omega_l)$ share the same real part, and that the imaginary part of $-I\tau(\omega_c+\Omega_l)$ is always negative while the imaginary part of $I\tau\left(\omega_{c,k}-\Omega_l\right)$ is positive. Therefore, this adjustment needs to be made for two of the four poles.

B.3 Efficient evaluation of utility functions

Instead of approximating the covariance matrix, we first try to find the nearest Kronecker product of the correlation matrix, because it usually gives better numerical stability. Assume we have $\boldsymbol{\Gamma}_s = \boldsymbol{DRD}$, in which $\boldsymbol{\Gamma} \in \mathbb{R}^{k\times k}$ is the covariance matrix, $\boldsymbol{R} \in \mathbb{R}^{k\times k}$ is the correlation matrix, and $\boldsymbol{D} \in \mathbb{R}^{k\times k}$ is a diagonal matrix consisting of standard deviations. Recalling that $\boldsymbol{\Gamma}$ is a block Toeplitz matrix, we can express $\boldsymbol{D} = \boldsymbol{I}_N \otimes \boldsymbol{D}_0$ where $\boldsymbol{D}_0 \in \mathbb{R}^{m\times m}$ is composed of the standard deviations of modal responses. Suppose we have the nearest Kronecker product approximation $\boldsymbol{R} \cong \boldsymbol{A} \otimes \boldsymbol{B}$, where $\boldsymbol{A} \in \mathbb{R}^{k_1\times k_1}$, $\boldsymbol{B} \in \mathbb{R}^{k_2\times k_2}$, and $k = k_1 k_2$, then

$$\boldsymbol{\Gamma}_s = (\boldsymbol{I}_N \otimes \boldsymbol{D}_0)(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{I}_N \otimes \boldsymbol{D}_0) = \boldsymbol{A} \otimes [(\boldsymbol{I}_{N/k_1} \otimes \boldsymbol{D}_0)\boldsymbol{B}(\boldsymbol{I}_{N/k_1} \otimes \boldsymbol{D}_0)] \quad (2.45)$$

Perform the following Cholesky decomposition:

$$\boldsymbol{A} = \boldsymbol{L_A}\boldsymbol{L_A}^{\mathrm{T}};\ [(\boldsymbol{I}_{N/k_1} \otimes \boldsymbol{D}_0)\boldsymbol{B}(\boldsymbol{I}_{N/k_1} \otimes \boldsymbol{D}_0)] = \boldsymbol{L_B}\boldsymbol{L_B}^{\mathrm{T}} \quad (2.46)$$

Then, we have the Cholesky decomposition of the covariance matrix:

$$\boldsymbol{\Gamma}_s = (\boldsymbol{L_A} \otimes \boldsymbol{L_B})(\boldsymbol{L_A} \otimes \boldsymbol{L_B})^{\mathrm{T}} \quad (2.47)$$

The posterior covariance is computed as follows:

$$\begin{aligned}
&[\boldsymbol{\Gamma}_s^{-1} + \boldsymbol{I}_N \otimes \boldsymbol{F}_s]^{-1} \\
&= [(\boldsymbol{L_A} \otimes \boldsymbol{L_B})^{-\mathrm{T}}(\boldsymbol{L_A} \otimes \boldsymbol{L_B})^{-1} + \boldsymbol{I}_N \otimes \boldsymbol{F}_s]^{-1} \\
&= (\boldsymbol{L_A} \otimes \boldsymbol{L_B})[\boldsymbol{I}_k + (\boldsymbol{L_A} \otimes \boldsymbol{L_B})^{\mathrm{T}}(\boldsymbol{I}_N \otimes \boldsymbol{F}_s)(\boldsymbol{L_A} \otimes \boldsymbol{L_B})]^{-1}(\boldsymbol{L_A} \otimes \boldsymbol{L_B})^{\mathrm{T}} \\
&= (\boldsymbol{L_A} \otimes \boldsymbol{L_B})\left[\boldsymbol{I}_k + \left(\boldsymbol{L_A}^{\mathrm{T}}\boldsymbol{L_A}\right) \otimes [\boldsymbol{L_B}^{\mathrm{T}}(\boldsymbol{I}_{N/k_1} \otimes \boldsymbol{F}_s)\boldsymbol{L_B}]\right]^{-1}(\boldsymbol{L_A} \otimes \boldsymbol{L_B})^{\mathrm{T}}
\end{aligned} \quad (2.48)$$

Now, perform the eigenvalue decompositions

$$\boldsymbol{L_A}^{\mathrm{T}}\boldsymbol{L_A} = \boldsymbol{U\Lambda U}^{\mathrm{T}};\ \boldsymbol{L_B}^{\mathrm{T}}(\boldsymbol{I}_{N/k_1} \otimes \boldsymbol{F_s})\boldsymbol{L_B} = \boldsymbol{V\Psi V}^{\mathrm{T}} \quad (2.49)$$

in which $\boldsymbol{U} \in \mathbb{R}^{k_1\times k_1}$ and $\boldsymbol{V} \in \mathbb{R}^{k_2\times k_2}$ are orthogonal matrices, $\boldsymbol{\Lambda} \in \mathbb{R}^{k_1\times k_1}$ and $\boldsymbol{\Psi} \in \mathbb{R}^{k_2\times k_2}$ are eigenvalue diagonal matrices. Substituting these decompositions into

the above equation gives

$$\begin{aligned}
&[\boldsymbol{\Gamma}_s^{-1} + \boldsymbol{I}_N \otimes \boldsymbol{F}_s]^{-1} \\
&= (\boldsymbol{L_A} \otimes \boldsymbol{L_B})[\boldsymbol{I}_k + (\boldsymbol{U} \otimes \boldsymbol{V})(\boldsymbol{\Lambda} \otimes \boldsymbol{\Psi})(\boldsymbol{U}^{\mathrm{T}} \otimes \boldsymbol{V}^{\mathrm{T}})]^{-1}(\boldsymbol{L_A} \otimes \boldsymbol{L_B})^{\mathrm{T}} \\
&= (\boldsymbol{L_A} \otimes \boldsymbol{L_B})(\boldsymbol{U} \otimes \boldsymbol{V})[(\boldsymbol{U}^{\mathrm{T}} \otimes \boldsymbol{V}^{\mathrm{T}})(\boldsymbol{U} \otimes \boldsymbol{V}) + \boldsymbol{\Lambda} \otimes \boldsymbol{\Psi}]^{-1}(\boldsymbol{U}^{\mathrm{T}} \otimes \boldsymbol{V}^{\mathrm{T}})(\boldsymbol{L_A} \otimes \\
&\quad \boldsymbol{L_B})^{\mathrm{T}} \\
&= (\boldsymbol{L_A U} \otimes \boldsymbol{L_B V})[\boldsymbol{I}_k + \boldsymbol{\Lambda} \otimes \boldsymbol{\Psi}]^{-1}(\boldsymbol{L_A U} \otimes \boldsymbol{L_B V})^{\mathrm{T}}
\end{aligned} \tag{2.50}$$

It is easy to see that the matrix in the bracket is diagonal so that the calculation of its inverse becomes trivial. Therefore, the utility function $U_2(d)$ can be easily evaluated by

$$\begin{aligned}
&\ln\det[\boldsymbol{\Gamma}_s^{-1} + \boldsymbol{I}_N \otimes \boldsymbol{F}_s] \\
&= -2\ln\det(\boldsymbol{L_A} \otimes \boldsymbol{L_B}) + \ln\det(\boldsymbol{I}_k + \boldsymbol{\Lambda} \otimes \boldsymbol{\Psi}) \\
&\propto \ln\det(\boldsymbol{I}_k + \boldsymbol{\Lambda} \otimes \boldsymbol{\Psi})
\end{aligned} \tag{2.51}$$

For the utility function $U_1(d)$, we have

$$\begin{aligned}
&\mathrm{tr}\,\{[\boldsymbol{\Gamma}_s^{-1} + \boldsymbol{I}_N \otimes \boldsymbol{F}_s]^{-1}\} \\
&= \mathrm{tr}\,\{(\boldsymbol{L_A U} \otimes \boldsymbol{L_B V})^{\mathrm{T}}(\boldsymbol{L_A U} \otimes \boldsymbol{L_B V})[\boldsymbol{I}_k + \boldsymbol{\Lambda} \otimes \boldsymbol{\Psi}]^{-1}\} \\
&= \mathrm{tr}\left\{[(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{L}_{\boldsymbol{A}}^{\mathrm{T}}\boldsymbol{L_A U}) \otimes (\boldsymbol{V}^{\mathrm{T}}\boldsymbol{L}_{\boldsymbol{B}}^{\mathrm{T}}\boldsymbol{L_B V})][\boldsymbol{I}_k + \boldsymbol{\Lambda} \otimes \boldsymbol{\Psi}]^{-1}\right\} \\
&= \mathrm{tr}\left\{[\boldsymbol{\Lambda} \otimes (\boldsymbol{V}^{\mathrm{T}}\boldsymbol{L}_{\boldsymbol{B}}^{\mathrm{T}}\boldsymbol{L_B V})][\boldsymbol{I}_k + \boldsymbol{\Lambda} \otimes \boldsymbol{\Psi}]^{-1}\right\}
\end{aligned} \tag{2.52}$$

Since matrices in the brackets are block diagonal and diagonal matrix individually, their multiplication only involves diagonal blocks, which is easily calculated.

Now, let us consider the trace term in utility function $U_3(d)$.

$$\begin{aligned}
&\mathrm{tr}\,\{[\boldsymbol{\Gamma}_s^{-1} + \boldsymbol{I}_N \otimes \boldsymbol{F}_s]^{-1}[\boldsymbol{\Gamma}_s^{-1} + \boldsymbol{I}_N \otimes \boldsymbol{F}]\} \\
&= \mathrm{tr}\{(\boldsymbol{L_A U} \otimes \boldsymbol{L_B V})[\boldsymbol{I}_k + \boldsymbol{\Lambda} \otimes \boldsymbol{\Psi}]^{-1}(\boldsymbol{L_A U} \otimes \boldsymbol{L_B V})^{\mathrm{T}}(\boldsymbol{L_A U} \otimes \boldsymbol{L_B V'})^{-\mathrm{T}}[\boldsymbol{I}_k + \\
&\quad \boldsymbol{\Lambda} \otimes \boldsymbol{\Psi'}](\boldsymbol{L_A U} \otimes \boldsymbol{L_B V'})^{-1}\} \\
&= \mathrm{tr}\left\{(\boldsymbol{I}_{k_1} \otimes \boldsymbol{V'}^{\mathrm{T}}\boldsymbol{V})[\boldsymbol{I}_k + \boldsymbol{\Lambda} \otimes \boldsymbol{\Psi}]^{-1}(\boldsymbol{I}_{k_1} \otimes \boldsymbol{V}^{\mathrm{T}}\boldsymbol{V'})[\boldsymbol{I}_k + \boldsymbol{\Lambda} \otimes \boldsymbol{\Psi'}]\right\}
\end{aligned} \tag{2.53}$$

Here, we have employed the eigenvalue decomposition

$$\boldsymbol{L_B}^{\mathrm{T}}(\boldsymbol{I}_{N/k_1} \otimes \boldsymbol{F})\boldsymbol{L_B} = \boldsymbol{V'\Psi'V'}^{\mathrm{T}} \tag{2.54}$$

Similar to the utility function $U_2(d)$, each multiplicator is block diagonal or diagonal. Therefore, we can also efficiently evaluate $U_3(d)$.

Chapter 3
Operational modal analysis‡

3.1 Introduction

This chapter concerns the uncertainty quantification in the operational modal analysis (OMA) using Bayesian statistics in the time-domain. In the OMA, the modal parameters, e.g. natural frequencies, damping ratios and mode shapes, are identified from the structural responses without knowing the sources of excitations, which are assumed to be broad-band stochastic processes adequately modeled by band-limited white noises. Changes in these modal parameters and their derivations are the most widely used indicators of structural damage [6,11,12]. Because of the low signal-to-noise ratio (SNR) in OMA, the variability of identified modal parameters is much larger than those identified from controlled experiments. Therefore, it is important to quantify the associated uncertainties for reliable damage diagnosis and prognosis.

We start with formulation of the OMA in Section 3.2, from a physical model to a probabilistic model considering potential sources of uncertainty. The physical model is based on a state-space representation of the equation of motion, and the modal parameters can be extracted from the state transition and observation matrices. By assigning probability distributions to the error terms and specifying prior distributions for the unknown parameters, a probabilistic model for the OMA is formally constructed. This probabilistic model belongs to the conjugate-exponential family. Subsequently, in Sections 3.3-3.5, we apply the EM algorithm, the variational Bayes and the Gibbs sampler to infer the modal parameters from the measured structural responses. These algorithms are first derived for this particular probabilistic model; a robust implementation strategy is then introduced based on square-root filtering and Cholesky decomposition. Section 3.6 illustrates the application of the proposed algorithms in the OMA by three examples: a mass-spring numerical model, a laboratory shear-type building model and the One Rincon Hill Tower in San Francisco. Section 3.7 presents a summary of the results of the chapter.

‡‡ Part of this chapter has been accepted for publication in Mechanical Systems and Signal Processing, 2016.

3.2 Problem formulation

3.2.1 The physical model

For a discretized, linear, time-invariant dynamical system with N_d degrees of freedom (DoFs), the equation of motion under an external force and a base motion is represented as

$$\boldsymbol{M}\ddot{\boldsymbol{u}}(t) + \boldsymbol{C}_d\dot{\boldsymbol{u}}(t) + \boldsymbol{K}\boldsymbol{u}(t) = \boldsymbol{P}\boldsymbol{f}(t) - \boldsymbol{M}\boldsymbol{l}\ddot{\boldsymbol{u}}_g(t) \tag{3.1}$$

where $\boldsymbol{M} \in \mathbb{R}^{N_d \times N_d}$, $\boldsymbol{C}_d \in \mathbb{R}^{N_d \times N_d}$ and $\boldsymbol{K} \in \mathbb{R}^{N_d \times N_d}$ are the mass, damping and stiffness matrices, respectively; $\boldsymbol{u}(t) \in \mathbb{R}^{N_d}$, $\dot{\boldsymbol{u}}(t) \in \mathbb{R}^{N_d}$ and $\ddot{\boldsymbol{u}}(t) \in \mathbb{R}^{N_d}$ are the nodal displacement, velocity and acceleration responses relative to the ground, respectively, with $\boldsymbol{u}(0) = \boldsymbol{u}_0$ and $\dot{\boldsymbol{u}}(0) = \dot{\boldsymbol{u}}_0$ being the initial relative displacement and velocity vectors; $\boldsymbol{f}(t) \in \mathbb{R}^{N_f}$ is the external force vector; $\boldsymbol{P} \in \mathbb{R}^{N_d \times N_f}$ is the load coefficient matrix representing the spatial influence of the external force; $\ddot{\boldsymbol{u}}_g(t) \in \mathbb{R}^{N_g}$ is the ground acceleration vector, and $\boldsymbol{l} \in \mathbb{R}^{N_d \times N_g}$ is the corresponding influence matrix. Note that we are considering both an external force and a base motion, because under operating conditions the structure is usually subject to the effects of wind, traffic, ground tremor and small magnitude earthquakes. The equation of motion in Eqn. (3.1) is equivalent to the following continuous-time state-space model (SSM):

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}_c\boldsymbol{x}(t) + \boldsymbol{B}_{cf}\boldsymbol{f}(t) - \boldsymbol{B}_{cg}\ddot{\boldsymbol{u}}_g(t) \tag{3.2}$$

with

$$\begin{aligned} \boldsymbol{x}(t) &= \begin{bmatrix} \boldsymbol{u}(t) \\ \dot{\boldsymbol{u}}(t) \end{bmatrix} \in \mathbb{R}^{N_s} \\ \boldsymbol{A}_c &= \begin{bmatrix} \boldsymbol{0}_{N_d} & \boldsymbol{I}_{N_d} \\ -\boldsymbol{M}^{-1}\boldsymbol{K} & -\boldsymbol{M}^{-1}\boldsymbol{C}_d \end{bmatrix} \in \mathbb{R}^{N_s \times N_s} \\ \boldsymbol{B}_{cf} &= \begin{bmatrix} \boldsymbol{0}_{N_d \times N_f} \\ \boldsymbol{M}^{-1}\boldsymbol{P} \end{bmatrix} \in \mathbb{R}^{N_s} \\ \boldsymbol{B}_{cg} &= \begin{bmatrix} \boldsymbol{0}_{N_d \times N_g} \\ \boldsymbol{l} \end{bmatrix} \in \mathbb{R}^{N_s} \end{aligned} \tag{3.3}$$

where $\boldsymbol{0}_{N_d}$ and $\boldsymbol{I}_{N_d}$ represent N_d-by-N_d zero and identity matrices, respectively. $N_s = 2N_d$ is usually called the model order.

In the above SSM, the state variable $\boldsymbol{x}(t)$ includes the relative displacements and the velocities at all DoFs, but in practice only a few of these quantities can be directly measured due to limited instrumentation. Furthermore, one usually measures

the absolute displacement and velocity, not those relative to the ground. As a consequence, $\boldsymbol{x}(t)$ is a latent variable, and the observed variable is expressed through an observation equation

$$\boldsymbol{y}(t) = \boldsymbol{C}\boldsymbol{x}(t) + \boldsymbol{D}f(t) \tag{3.4}$$

where the observed variables $\boldsymbol{y}(t) \in \mathbb{R}^{N_o}$ can be any quantities of interest that directly relate to the nodal displacements or velocities. In practice, nodal accelerations are easy to measure with high resolution. In that case, $\boldsymbol{C} = \boldsymbol{S}_o[-\boldsymbol{M}^{-1}\boldsymbol{K} \quad -\boldsymbol{M}^{-1}\boldsymbol{C}_d]$ and $\boldsymbol{D} = \boldsymbol{S}_o\boldsymbol{M}^{-1}\boldsymbol{P}$, where $\boldsymbol{S}_o \in \mathbb{R}^{N_o \times N_s}$ is a selection matrix that defines the $N_o (N_o \leq N_s)$ DOFs of the structure at which measurements are made.

In reality, the data available is in a discrete form; therefore, it is necessary to convert Eqn. (3.2) into a discrete model through sampling. The detailed procedure is summarized as follows:

(1) Multiply Eqn. (3.2) by the matrix exponential $e^{-\boldsymbol{A}_c t}$ [§]:

$$\frac{d}{dt}[e^{-\boldsymbol{A}_c t}\boldsymbol{x}(t)] = e^{-\boldsymbol{A}_c t}\boldsymbol{B}_{cf}\boldsymbol{f}(t) - e^{-\boldsymbol{A}_c t}\boldsymbol{B}_{cg}\ddot{\boldsymbol{u}}_g(t) \tag{3.5}$$

(2) Integrate both sides of the above equation within time interval $[kT_s, (k+1)T_s]$:

$$\boldsymbol{x}_{k+1} = e^{\boldsymbol{A}_c T_s}\boldsymbol{x}_k + \int_{kT_s}^{(k+1)T_s} e^{\boldsymbol{A}_c[(k+1)T_s - t]}[\boldsymbol{B}_{cf}\boldsymbol{f}(t) - \boldsymbol{B}_{cg}\ddot{\boldsymbol{u}}_g(t)]dt \tag{3.6}$$

where T_s is the sampling period and $\boldsymbol{x}_k \equiv \boldsymbol{x}(kT_s)$.

(3) If we assume that the excitations $\boldsymbol{f}(t)$ and $\ddot{\boldsymbol{u}}_g(t)$ are constant (zero-order-hold assumption [127]), i.e. $\boldsymbol{f}(t) = \boldsymbol{f}_k$ and $\ddot{\boldsymbol{u}}_g(t) = \ddot{\boldsymbol{u}}_{g,k}$, within the interval $[kT_s, (k+1)T_s]$, then Eqn. (3.6) can be simplified as

$$\boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}_f\boldsymbol{f}_k - \boldsymbol{B}_g\ddot{\boldsymbol{u}}_{g,k} \tag{3.7}$$

where we have

$$\boldsymbol{A} = e^{\boldsymbol{A}_c T_s}, \ \boldsymbol{B}_f = \left(\boldsymbol{A} - I_{N_s}\right)\boldsymbol{A}_c^{-1}\boldsymbol{B}_{cf}, \ \boldsymbol{B}_g = \left(\boldsymbol{A} - I_{N_s}\right)\boldsymbol{A}_c^{-1}\boldsymbol{B}_{cg} \tag{3.8}$$

Correspondingly, the discretized observation equation becomes

$$\boldsymbol{y}_k = \boldsymbol{C}\boldsymbol{x}_k + \boldsymbol{D}\boldsymbol{f}_k \tag{3.9}$$

Note that the above simplification depends on the zero-order-hold assumption, which generally holds if the sampling period T_s is sufficiently small.

[§] $e^{-\boldsymbol{A}t} = [e^{\boldsymbol{A}t}]^{-1}$, and $e^{\boldsymbol{A}t} = \boldsymbol{\Psi}e^{\boldsymbol{\Lambda}t}\boldsymbol{\Psi}^{-1}$, where $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$ come from the eigenvalue decomposition $\boldsymbol{A} = \boldsymbol{\Psi}\boldsymbol{\Lambda}\boldsymbol{\Psi}^{-1}$, and $e^{\boldsymbol{\Lambda}t} = diag(e^{\lambda_1 t}, e^{\lambda_2 t}, \dots, e^{\lambda_{N_s} t})$

Eqns. (3.7) and (3.9) lay the foundation for deterministic computation of the dynamic structural responses. However, as discussed in Chapter 1, the system is always subject to various kinds of errors: Eqn. (3.7) cannot exactly predict the structural behavior due to the existence of model error, and Eqn. (3.9) should account for measurement error that is invariably present. In addition, the stochastic force $\boldsymbol{f}_k$ and base motion $\ddot{\boldsymbol{u}}_{g,k}$ are not explicitly measured in the OMA, so that they can only be modeled as random processes. Here, we model these uncertainties by replacing the combined effects of $\boldsymbol{f}_k$, $\ddot{\boldsymbol{u}}_{g,k}$ and the model and measurement errors in Eqns. (3.7) and (3.9) with two noise terms $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$, resulting in the stochastic SSM:

$$\begin{bmatrix} \boldsymbol{x}_{k+1} \\ \boldsymbol{y}_k \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix} \boldsymbol{x}_k + \begin{bmatrix} \boldsymbol{w}_k \\ \boldsymbol{v}_k \end{bmatrix} \tag{3.10}$$

More specifically, $\boldsymbol{w}_k$ represents the effects of the unknown external forces $\boldsymbol{f}_k$, ground accelerations $\ddot{\boldsymbol{u}}_{g,k}$, and the model error, while $\boldsymbol{v}_k$ stands for the joint effects of the external forces and measurement error.

Since our objective in the OMA is to identify modal parameters, it is necessary to connect the modal parameters with the state transition matrix $\boldsymbol{A}$ and observation matrix $\boldsymbol{C}$ in the stochastic SSM shown in Eqn. (3.10). In fact, we can extract the modal parameters by the following steps [45]:

(1) Apply the eigenvalue decomposition on $\boldsymbol{A}$ such that $\boldsymbol{A} = \boldsymbol{\Psi}\boldsymbol{\Lambda}\boldsymbol{\Psi}^{-1}$;

(2) Since the eigen-solutions occur in complex conjugate pairs, partition the eigenvalue and eigenvector matrices as

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\lambda} & \boldsymbol{0} \\ \boldsymbol{0} & \bar{\boldsymbol{\lambda}} \end{bmatrix}, \quad \boldsymbol{\Psi} = [\boldsymbol{\psi} \quad \bar{\boldsymbol{\psi}}] \tag{3.11}$$

where a superposed bar indicates the complex conjugate.

(3) Extract the i-th modal frequency and modal damping ratio as

$$\lambda_{ci} = \ln \lambda_i / T_s, \qquad f_i = |\lambda_{ci}|/2\pi, \quad \xi_i = -\mathrm{Re}(\lambda_{ci})/|\lambda_{ci}| \tag{3.12}$$

where λ_i is the i-th diagonal element of $\boldsymbol{\lambda}$.

(4) Calculate the i-th mode shape as:

$$\boldsymbol{\phi}_i = \boldsymbol{C}\boldsymbol{\psi}_i \tag{3.13}$$

where $\boldsymbol{\psi}_i$ is the i-th column of eigenvector matrix $\boldsymbol{\psi}$.

The obtained mode shape $\boldsymbol{\phi}_i$ in Eqn. (3.13) is a complex vector, but the real mode shape is more meaningful to the engineers. To obtain a real mode shape, considering that the mode shape can be arbitrarily normalized, it is possible to introduce a complex normalizing constant z_i such that

$$\min_{z_i} \|\mathrm{Im}(z_i \boldsymbol{C}\boldsymbol{\psi}_i)\|_2 \quad \text{s.t. } \|\mathrm{Re}(z_i \boldsymbol{C}\boldsymbol{\psi}_i)\|_2 = 1 \tag{3.14}$$

where '$\|\cdot\|_2$' stands for the L_2 vector norm. That is the normalized mode shape has a minimum imaginary part and a unit real part in terms of L_2 vector norm. The real mode shape is then obtained by remaining only the real part after normalization. Interestingly, a closed-form solution of Eqn. (3.14) exists based on the Lagrange multiplier method [67] and the eigenvalue decomposition. To extract the i-th real mode shape, the detailed procedures are listed below:

1) Define the matrix

$$\boldsymbol{\varphi}_i = [\mathrm{Re}(\boldsymbol{\psi}_i) \quad \mathrm{Im}(\boldsymbol{\psi}_i)] \tag{3.15}$$

where 'Im' and 'Re' denote the imaginary and real part of a complex vector, respectively

2) Take the generalized eigenvalue decomposition:

$$\boldsymbol{P}_1 \boldsymbol{\varphi}_i^{\mathrm{T}} \boldsymbol{C}^{\mathrm{T}} \boldsymbol{C} \boldsymbol{\varphi}_i \boldsymbol{P}_1 \hat{\boldsymbol{z}} = \lambda_{min}\, \boldsymbol{P}_2 \boldsymbol{\varphi}_i^{\mathrm{T}} \boldsymbol{C}^{\mathrm{T}} \boldsymbol{C} \boldsymbol{\varphi}_i \boldsymbol{P}_2 \hat{\boldsymbol{z}} \tag{3.16}$$

where $\boldsymbol{P}_1 = [0 \quad 1; 1 \quad 0]$ and $\boldsymbol{P}_2 = [1 \quad 0; 0 \quad -1]$; λ_{min} is the smallest eigenvalue and $\hat{\boldsymbol{z}}$ is the corresponding eigenvector;

3) Normalize the obtained eigenvector

$$\boldsymbol{z} = \hat{\boldsymbol{z}} / \sqrt{\hat{\boldsymbol{z}}^{\mathrm{T}} \boldsymbol{P}_2 \boldsymbol{\varphi}_i^{\mathrm{T}} \boldsymbol{C}^{\mathrm{T}} \boldsymbol{C} \boldsymbol{\varphi}_i \boldsymbol{P}_2 \hat{\boldsymbol{z}}} \tag{3.17}$$

4) Then, the optimal normalizing constant reads

$$z_i = z_1 + z_2 \sqrt{-1} \tag{3.18}$$

where z_1 and z_2 are the first and second elements of $\boldsymbol{z}$. The i-th real mode shape is

$$\boldsymbol{\phi}_i = \boldsymbol{C} \boldsymbol{\varphi}_i \boldsymbol{P}_2 \boldsymbol{z} \tag{3.19}$$

In addition, it is easy to show that $\min\|\mathrm{Im}(z_i \boldsymbol{C}\boldsymbol{\psi}_i)\|_2 = \lambda_{min}$, and we will use it as an index to distinguish spurious modes by presuming that the physical mode shapes tend to be real vectors.

3.2.2 The probabilistic model

In this section, we transform the stochastic SSM in Eqn. (3.10) into a probabilistic model by assigning probability distributions to the noise terms $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ and

specifying prior distributions for the unknown parameters.

Following the basic assumption in the OMA that the unmeasured excitations can be well modeled by band-limited Gaussian white-noise processes, i.e. that their power spectrum density functions are flat over the frequency range relevant to the structural response, the joint distribution of $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ is assumed to be the multivariate normal with zero-mean and unknown covariance matrix $\boldsymbol{\Sigma} = [\boldsymbol{Q},\ \boldsymbol{S};\ \ \boldsymbol{S}^{\mathrm{T}},\ \boldsymbol{R}]$ for $k = 1,2,\dots,N$. This assumption can be justified by the principle of maximum entropy [108]. Then, the probability density function (PDF) of the joint distribution of $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ is given by

$$f(\boldsymbol{w}_k, \boldsymbol{v}_k|\boldsymbol{\Sigma}) = |2\pi\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}[\boldsymbol{w}_k^{\mathrm{T}} \quad \boldsymbol{v}_k^{\mathrm{T}}]\boldsymbol{\Sigma}^{-1}\begin{bmatrix}\boldsymbol{w}_k \\ \boldsymbol{v}_k\end{bmatrix}\right\} \tag{3.20}$$

where '$|\cdot|$' denotes the absolute of the matrix determinant. As a consequence, the joint distribution of $\boldsymbol{z}_k = [\boldsymbol{x}_{k+1}^{\mathrm{T}} \quad \boldsymbol{y}_k^{\mathrm{T}}]^{\mathrm{T}}$ given $\boldsymbol{x}_k$ and $\boldsymbol{\Sigma}$ is also a multivariate normal with the PDF

$$f(\boldsymbol{z}_k|\boldsymbol{x}_k, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) = |2\pi\boldsymbol{\Sigma}|^{-1/2} \exp\left[-\frac{1}{2}(\boldsymbol{z}_k - \boldsymbol{\Gamma}\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}_k - \boldsymbol{\Gamma}\boldsymbol{x}_k)\right] \tag{3.21}$$

where we have defined $\boldsymbol{\Gamma} = [\boldsymbol{A}^{\mathrm{T}} \quad \boldsymbol{C}^{\mathrm{T}}]^{\mathrm{T}}$.

Given the likelihood function in Eqn. (3.21), we choose a multivariate normal distribution as the prior of the initial response $\boldsymbol{x}_1$ and a matrix normal inverse Wishart distribution [128] as the joint prior distribution of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ so that the formulated probabilistic model belongs to the conjugate-exponential family. To be more specific, their PDFs are listed below

$$f(\boldsymbol{x}_1; \boldsymbol{\mu}_{1,0}, \boldsymbol{P}_{1,0}) = \left|2\pi\boldsymbol{P}_{1,0}\right|^{-1/2} \exp\left[-\frac{1}{2}\left(\boldsymbol{x}_1 - \boldsymbol{\mu}_{1,0}\right)^{\mathrm{T}}\boldsymbol{P}_{1,0}^{-1}\left(\boldsymbol{x}_1 - \boldsymbol{\mu}_{1,0}\right)\right] \tag{3.22}$$

$$\pi(\boldsymbol{\Sigma}; d_0, \boldsymbol{D}_0) = \frac{|\boldsymbol{D}_0|^{d_0/2}|\boldsymbol{\Sigma}|^{-(d_0+N_s+N_o+1)/2}}{2^{d_0(N_s+N_o)/2}\Gamma_{N_s+N_o}(d_0/2)} \exp\left[-\frac{1}{2}\mathrm{tr}(\boldsymbol{D}_0\boldsymbol{\Sigma}^{-1})\right] \tag{3.23}$$

$$\pi(\boldsymbol{\Gamma}|\boldsymbol{\Sigma}; \boldsymbol{\Gamma}_0, \boldsymbol{\Pi}_0) = \frac{\exp\left\{-\frac{1}{2}\mathrm{tr}[\boldsymbol{\Pi}_0^{-1}(\boldsymbol{\Gamma} - \boldsymbol{\Gamma}_0)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Gamma} - \boldsymbol{\Gamma}_0)]\right\}}{|2\pi\boldsymbol{\Sigma}|^{N_s/2}|2\pi\boldsymbol{\Pi}_0|^{(N_s+N_o)/2}} \tag{3.24}$$

where $\Gamma_{N_s+N_o}(\cdot)$ represents the multivariate gamma function and $\mathrm{tr}(\cdot)$ denotes the matrix trace; the mean $\boldsymbol{\mu}_{1,0} \in \mathbb{R}^{N_s}$ and the covariance matrix $\boldsymbol{P}_{1,0} \in \mathbb{R}^{N_s \times N_s}$ are the hyper-parameters in the multivariate normal distribution; the degree of freedom d_0 and scale matrix $\boldsymbol{D}_0 \in \mathbb{R}^{(N_s+N_o)\times(N_s+N_o)}$ are the hyper-parameters in the inverse Wishart distribution; the mean $\boldsymbol{\Gamma}_0 \in \mathbb{R}^{(N_s+N_o)\times N_s}$ and the right-covariance matrix $\boldsymbol{\Pi}_0 \in \mathbb{R}^{N_s \times N_s}$ are hyper-parameters in the matrix normal distribution. All these hyper-parameters are selected such that the priors are as noninformative as possible.

Note that matrices $\boldsymbol{A}$ and $\boldsymbol{C}$ in Eqn. (3.10) share common terms, leading them to be statistically dependent; therefore, $\boldsymbol{\Gamma}$ as a whole is considered a matrix random variable. In addition, since both $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ include the effects of the unmeasured external force and ground accelerations, they should be statistically correlated, yielding a full covariance matrix $\boldsymbol{\Sigma}$. These aspects differentiate this study from all previous studies [94,129].

Collecting all the observed variables $\boldsymbol{y}_{1:N} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_N\}$, the latent variables $\boldsymbol{x}_{1:N+1} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_{N+1}\}$, and the unknown parameters $\boldsymbol{\theta} = \{\boldsymbol{\Gamma}, \boldsymbol{\Sigma}\}$ as well as their associated PDFs gives the probabilistic model for the OMA, which is shown as a Bayesian network in Figure 3.1.

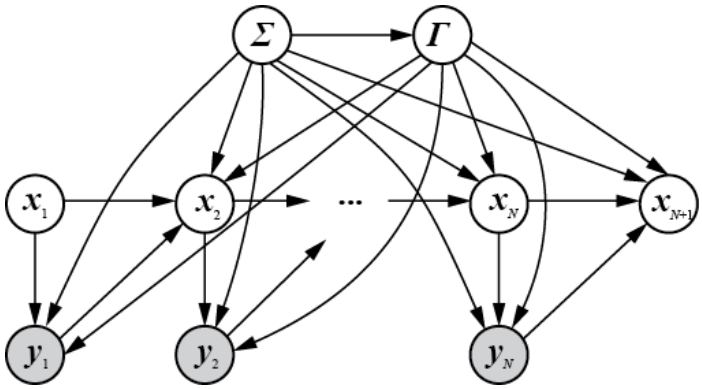


Figure 3.1 Bayesian network for the OMA.

The above Bayesian network is equivalent to the joint distribution given by

$$f(\boldsymbol{z}_{1:N}, \boldsymbol{x}_1, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) = \pi(\boldsymbol{\Sigma})\pi(\boldsymbol{\Gamma}|\boldsymbol{\Sigma})f(\boldsymbol{x}_1)\prod_{k=1}^{N} f(\boldsymbol{z}_k|\boldsymbol{x}_k, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) \tag{3.25}$$

where the PDFs in the right hand side are given in Eqns. (3.20)-(3.24). Here, we have exploited the conditional Markov property that can also be read in Figure 3.1, i.e.

$$f(\boldsymbol{z}_k|\boldsymbol{x}_{1:k}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) = f(\boldsymbol{z}_k|\boldsymbol{x}_k, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) \tag{3.26}$$

Given the probabilistic model shown in (3.25), once the structural responses $\boldsymbol{y}_{1:N}$ are recorded, it is possible to approximate the posterior of unknown parameters $\boldsymbol{\theta}$ using the techniques introduced in Section 1.2. In the following three sections, we will successively apply the EM algorithm, VB and Gibbs sampler to tackle the problem of OMA.

3.3 Expectation maximization

This section will focus on the maximum a posterior (MAP) estimation of the modal parameters using the EM algorithm. First, the basic updating procedures are derived for the probabilistic model shown in (3.25). Then, a robust implementation strategy is introduced to overcome the numerical error, arising from the repeated implementation of the forward-backward inference, based on the square-root filtering technique [130].

3.3.1 EM algorithm derivation

To derive the EM algorithm, we follow the procedures discussed in Section 1.2.1, i.e. we write down the complete log likelihood, take the expectation, and maximize this expected complete log likelihood with respect to each unknown parameter.

The E step

It is easy to find the log complete-data likelihood by substituting each PDF into Eqn. (3.25) and taking the logarithm of the function. For convenience, we multiply the resulting function by -2 to obtain:

$$\begin{aligned}
&\mathcal{L}(\boldsymbol{\Gamma}, \boldsymbol{\Sigma}, \boldsymbol{x}_{1:N+1}) \\
&= (d_0 + N + 2N_s + N_o + 1)\log|\boldsymbol{\Sigma}| + \mathrm{tr}(\boldsymbol{D}_0\boldsymbol{\Sigma}^{-1}) \\
&\quad + \textstyle\sum_{k=1}^{N}(\boldsymbol{z}_k - \boldsymbol{\Gamma}\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}_k - \boldsymbol{\Gamma}\boldsymbol{x}_k) \\
&\quad + \big(\boldsymbol{x}_1 - \boldsymbol{\mu}_{1,0}\big)^{\mathrm{T}}\boldsymbol{P}_{1,0}^{-1}\big(\boldsymbol{x}_1 - \boldsymbol{\mu}_{1,0}\big) + \mathrm{tr}[\boldsymbol{\Pi}_0^{-1}(\boldsymbol{\Gamma} - \boldsymbol{\Gamma}_0)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Gamma} - \boldsymbol{\Gamma}_0)] + c_0
\end{aligned} \tag{3.27}$$

where c_0 is a constant that does not depend on the unknown parameters. Next, we derive the conditional expectation of the log complete-data likelihood to obtain the function

$$\begin{aligned}
&Q\big(\boldsymbol{\Gamma}, \boldsymbol{\Sigma}\big|\boldsymbol{\Gamma}^{(t)}, \boldsymbol{\Sigma}^{(t)}\big) \\
&= \int f\big(\boldsymbol{x}_{1:N+1}\big|\boldsymbol{y}_{1:N}, \boldsymbol{\Gamma}^{(t)}, \boldsymbol{\Sigma}^{(t)}\big)\mathcal{L}(\boldsymbol{\Gamma}, \boldsymbol{\Sigma}, \boldsymbol{x}_{1:N+1})\, d\boldsymbol{x} \\
&= (d_0 + N + 2N_s + N_o + 1)\log|\boldsymbol{\Sigma}| + \mathrm{tr}(\boldsymbol{D}_0\boldsymbol{\Sigma}^{-1}) \\
&\quad + \mathrm{tr}\big\{\textstyle\sum_{k=1}^{N}\langle\boldsymbol{z}_k\boldsymbol{z}_k^{\mathrm{T}}\rangle\,\boldsymbol{\Sigma}^{-1}\big\} - 2\mathrm{tr}\big\{\textstyle\sum_{k=1}^{N}\langle\boldsymbol{z}_k\boldsymbol{x}_k^{\mathrm{T}}\rangle\,\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\big\} \\
&\quad + \mathrm{tr}\big\{\textstyle\sum_{k=1}^{N}\langle\boldsymbol{x}_k\boldsymbol{x}_k^{\mathrm{T}}\rangle\,\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}\big\} + \mathrm{tr}[\boldsymbol{\Pi}_0^{-1}(\boldsymbol{\Gamma} - \boldsymbol{\Gamma}_0)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Gamma} - \boldsymbol{\Gamma}_0)] + c_1
\end{aligned} \tag{3.28}$$

where c_1 is a constant that does not depend on the unknown parameters and the

dependence of the expectation terms on the current estimated parameters is not explicitly written.

The M step

In the M step, we need to maximize $Q\left(\boldsymbol{\Gamma}, \boldsymbol{\Sigma} \middle| \boldsymbol{\Gamma}^{(t)}, \boldsymbol{\Sigma}^{(t)}\right)$ to obtain the updated parameters. This is done as follows:

Taking the derivative of $Q\left(\boldsymbol{\Gamma}, \boldsymbol{\Sigma} \middle| \boldsymbol{\Gamma}^{(t)}, \boldsymbol{\Sigma}^{(t)}\right)$ with respect to $\boldsymbol{\Gamma}$ yields:

$$\begin{aligned} &\frac{\partial}{\partial \boldsymbol{\Gamma}} Q\left(\boldsymbol{\Gamma}, \boldsymbol{\Sigma} \middle| \boldsymbol{\Gamma}^{(t)}, \boldsymbol{\Sigma}^{(t)}\right) \\ &= -2\boldsymbol{\Sigma}^{-1} \sum_{k=1}^{N} \langle \boldsymbol{z}_k \boldsymbol{x}_k^{\mathrm{T}} \rangle + 2\boldsymbol{\Sigma}^{-1} \boldsymbol{\Gamma} \sum_{k=1}^{N} \langle \boldsymbol{x}_k \boldsymbol{x}_k^{\mathrm{T}} \rangle + 2\boldsymbol{\Sigma}^{-1} (\boldsymbol{\Gamma} - \boldsymbol{\Gamma}_0) \boldsymbol{\Pi}_0^{-1} \end{aligned} \tag{3.29}$$

Setting it equal to zero yields the updated $\boldsymbol{\Gamma}$:

$$\boldsymbol{\Gamma}^{(t+1)} = \left[\boldsymbol{\Gamma}_0 \boldsymbol{\Pi}_0^{-1} + \sum_{k=1}^{N} \langle \boldsymbol{z}_k \boldsymbol{x}_k^{\mathrm{T}} \rangle\right] \left[\boldsymbol{\Pi}_0^{-1} + \sum_{k=1}^{N} \langle \boldsymbol{x}_k \boldsymbol{x}_k^{\mathrm{T}} \rangle\right]^{-1} \tag{3.30}$$

Similarly, we get the updated $\boldsymbol{\Sigma}$ as

$$\begin{aligned} \boldsymbol{\Sigma}^{(t+1)} &= (d_0 + N + 2N_s + N_o + 1) \Big\{ \boldsymbol{D}_0 + \sum_{k=1}^{N} \langle \boldsymbol{z}_k \boldsymbol{z}_k^{\mathrm{T}} \rangle + \boldsymbol{\Gamma}_0 \boldsymbol{\Pi}_0^{-1} \boldsymbol{\Gamma}_0^{\mathrm{T}} \\ &- \left[\boldsymbol{\Gamma}_0 \boldsymbol{\Pi}_0^{-1} + \sum_{k=1}^{N} \langle \boldsymbol{z}_k \boldsymbol{x}_k^{\mathrm{T}} \rangle\right] \left[\boldsymbol{\Pi}_0^{-1} + \sum_{k=1}^{N} \langle \boldsymbol{x}_k \boldsymbol{x}_k^{\mathrm{T}} \rangle\right]^{-1} \left[\boldsymbol{\Gamma}_0 \boldsymbol{\Pi}_0^{-1} + \sum_{k=1}^{N} \langle \boldsymbol{z}_k \boldsymbol{x}_k^{\mathrm{T}} \rangle\right]^{\mathrm{T}} \Big\} \end{aligned} \tag{3.31}$$

The other issue arising is that we have to select an appropriate value to initiate the iteration process in order to maximize the Q function. This is crucial for the EM algorithm, because it is a local optimization algorithm in nature. Here, we choose the parameter values identified by the SSI algorithm [47], which, as a method of moments, usually gives good point estimates of modal parameters without iteration. Once the MAP solution for $\boldsymbol{\Gamma}$ is obtained, we can follow the steps introduced in Section 3.2.1 to deterministically compute the modal parameters as the MAP estimates.

3.3.2 Inference for latent variables

Now, let us turn our attention to the computation of the expectation terms in Eqn.(3.28), i.e. $\langle \boldsymbol{z}_k \boldsymbol{z}_k^{\mathrm{T}} \rangle$, $\langle \boldsymbol{x}_k \boldsymbol{x}_k^{\mathrm{T}} \rangle$ and $\langle \boldsymbol{z}_k \boldsymbol{x}_k^{\mathrm{T}} \rangle$. This is not a trivial task considering the chain-like structure of the probabilistic model in Figure 3.1. Following basic probability rules, we have

$$\langle \boldsymbol{x}_k \boldsymbol{x}_k^{\mathrm{T}} \rangle = \boldsymbol{P}_{k|N} + \boldsymbol{x}_{k|N} \boldsymbol{x}_{k|N}^{\mathrm{T}} \tag{3.32}$$

$$\langle \boldsymbol{z}_k \boldsymbol{x}_k^{\mathrm{T}} \rangle = \begin{bmatrix} \boldsymbol{P}_{k+1,k|N} + \boldsymbol{x}_{k+1|N} \boldsymbol{x}_{k|N}^{\mathrm{T}} \\ \boldsymbol{y}_k \boldsymbol{x}_{k|N}^{\mathrm{T}} \end{bmatrix} \tag{3.33}$$

$$\langle \boldsymbol{z}_k \boldsymbol{z}_k^{\mathrm{T}} \rangle = \begin{bmatrix} \boldsymbol{P}_{k+1|N} + \boldsymbol{x}_{k+1|N} \boldsymbol{x}_{k+1|N}^{\mathrm{T}} & \boldsymbol{x}_{k+1|N} \boldsymbol{y}_k^{\mathrm{T}} \\ \boldsymbol{y}_k \boldsymbol{x}_{k+1|N}^{\mathrm{T}} & \boldsymbol{y}_k \boldsymbol{y}_k^{\mathrm{T}} \end{bmatrix} \tag{3.34}$$

where $\boldsymbol{x}_{k|N}$, $\boldsymbol{P}_{k|N}$ and $\boldsymbol{P}_{k+1,k|N}$ denote the conditional mean of $\boldsymbol{x}_k$, the conditional covariance of $\boldsymbol{x}_k$ and the conditional covariance of $\boldsymbol{x}_{k+1}$ and $\boldsymbol{x}_k$, given $\boldsymbol{y}_{1:N}$. We apply the forward-backward inference algorithm [19], as shown in Figure 3.2, to compute these terms.

---

1) Initialization

Define $\begin{bmatrix} \boldsymbol{A}^{(t)} \\ \boldsymbol{C}^{(t)} \end{bmatrix} = \boldsymbol{\Gamma}^{(t)}$, $\begin{bmatrix} \boldsymbol{Q}^{(t)} & \boldsymbol{S}^{(t)} \\ \boldsymbol{S}^{(t)\mathrm{T}} & \boldsymbol{R}^{(t)} \end{bmatrix} = \boldsymbol{\Sigma}^{(t)}$,

$\widetilde{\boldsymbol{A}}^{(t)} = \boldsymbol{A}^{(t)} - \boldsymbol{S}^{(t)} \boldsymbol{R}^{(t)^{-1}} \boldsymbol{C}^{(t)}$, $\widetilde{\boldsymbol{Q}}^{(t)} = \boldsymbol{Q}^{(t)} - \boldsymbol{S}^{(t)} \boldsymbol{R}^{(t)^{-1}} \boldsymbol{S}^{(t)\mathrm{T}}$

and set $\boldsymbol{x}_{1|0} = \boldsymbol{\mu}_{1,0}$ and $\boldsymbol{P}_{1|0} = \boldsymbol{P}_{1,0}$

2) Forward inference

For $k = 1$ to N

Measurement update

$\boldsymbol{K}_k = \boldsymbol{P}_{k|k-1} \boldsymbol{C}^{(t)\mathrm{T}} \left[\boldsymbol{R}^{(t)} + \boldsymbol{C}^{(t)} \boldsymbol{P}_{k|k-1} \boldsymbol{C}^{(t)\mathrm{T}} \right]^{-1}$

$\boldsymbol{x}_{k|k} = \boldsymbol{x}_{k|k-1} + \boldsymbol{K}_k (\boldsymbol{y}_k - \boldsymbol{C}^{(t)} \boldsymbol{x}_{k|k-1})$

$\boldsymbol{P}_{k|k} = \boldsymbol{P}_{k|k-1} - \boldsymbol{K}_k \boldsymbol{C}^{(t)} \boldsymbol{P}_{k|k-1}$

Time update

$\boldsymbol{x}_{k+1|k} = \widetilde{\boldsymbol{A}}^{(t)} \boldsymbol{x}_{k|k} + \boldsymbol{S}^{(t)} \boldsymbol{R}^{(t)^{-1}} \boldsymbol{y}_k$

$\boldsymbol{P}_{k+1|k} = \widetilde{\boldsymbol{Q}}^{(t)} + \widetilde{\boldsymbol{A}}^{(t)} \boldsymbol{P}_{k|k} \widetilde{\boldsymbol{A}}^{(t)\mathrm{T}}$

End For

3) Backward inference

For $k = N$ to 1

$\boldsymbol{L}_k = \boldsymbol{P}_{k|k} \widetilde{\boldsymbol{A}}^{(t)\mathrm{T}} \boldsymbol{P}_{k+1|k}^{-1}$

$\boldsymbol{x}_{k|N} = \boldsymbol{x}_{k|k} + \boldsymbol{L}_k (\boldsymbol{x}_{k+1|N} - \boldsymbol{x}_{k+1|k})$

$\boldsymbol{P}_{k|N} = \boldsymbol{P}_{k|k} + \boldsymbol{L}_k (\boldsymbol{P}_{k+1|N} - \boldsymbol{P}_{k+1|k}) \boldsymbol{L}_k^{\mathrm{T}}$

$\boldsymbol{P}_{k+1,k|N} = \boldsymbol{P}_{k+1|N} \boldsymbol{L}_k^{\mathrm{T}}$

End For

Figure 3.2 Forward-backward inference.

3.3.3 Robust implementation

The last two sections show the basic procedure to infer the latent variables and parameters; a naïve implementations directly following the updating steps may suffer a serious numerical error. Since the matrices $\boldsymbol{P}_{k|k}$, $\boldsymbol{P}_{k|k-1}$, $\boldsymbol{P}_{k|N}$ and $\boldsymbol{\Sigma}$ must be kept symmetric and positive semi-definite at all iterations, a robust implementation of

the scheme is essential. Here, we apply the square-root filtering strategy [130] to deal with this problem.

The robust E step

This step involves transforming the forward-backward inference algorithm in Figure 3.2 to a robust implementation by repeatedly using the QR decomposition.

First, for computing $\boldsymbol{P}_{k|k}$ and $\boldsymbol{x}_{k|k}$, we start with $\boldsymbol{x}_{1|0}$ and $\boldsymbol{P}_{1|0}^{1/2}$ where $\boldsymbol{P}_{1|0} = \boldsymbol{P}_{1|0}^{\mathrm{T}/2}\boldsymbol{P}_{1|0}^{1/2}$ and take the following QR decomposition

$$\begin{bmatrix} \boldsymbol{R}^{(t)^{1/2}} & \boldsymbol{0}_{N_o\times N_s} & -\boldsymbol{R}^{(t)^{-\mathrm{T}/2}}\boldsymbol{y}_k \\ \boldsymbol{P}_{k|k-1}^{1/2}\boldsymbol{C}^{(t)^{\mathrm{T}}} & \boldsymbol{P}_{k|k-1}^{1/2} & \boldsymbol{P}_{k|k-1}^{-\mathrm{T}/2}\boldsymbol{x}_{k|k-1} \end{bmatrix} = \mathcal{Q}\mathcal{R} = \mathcal{Q}\begin{bmatrix} \mathcal{R}_{11} & \mathcal{R}_{12} & \mathcal{R}_{13} \\ \boldsymbol{0}_{N_s\times N_o} & \mathcal{R}_{22} & \mathcal{R}_{23} \end{bmatrix} \tag{3.35}$$

where $\mathcal{R}$ is portioned conformally to the left-hand side of Eqn. (3.35). Exploiting the unitary nature of $\mathcal{Q}$, and multiplying this equation on the left by its transpose, provides

$$\begin{bmatrix} \boldsymbol{R}^{(t)} + \boldsymbol{C}^{(t)}\boldsymbol{P}_{k|k-1}\boldsymbol{C}^{(t)^{\mathrm{T}}} & \boldsymbol{C}^{(t)}\boldsymbol{P}_{k|k-1} & \boldsymbol{C}^{(t)}\boldsymbol{x}_{k|k-1} - \boldsymbol{y}_k \\ \cdots & \boldsymbol{P}_{k|k-1} & \boldsymbol{x}_{k|k-1} \\ \cdots & \cdots & \cdots \end{bmatrix} = \begin{bmatrix} \mathcal{R}_{11}^{\mathrm{T}}\mathcal{R}_{11} & \mathcal{R}_{11}^{\mathrm{T}}\mathcal{R}_{12} & \mathcal{R}_{11}^{\mathrm{T}}\mathcal{R}_{13} \\ \cdots & \mathcal{R}_{12}^{\mathrm{T}}\mathcal{R}_{12} + \mathcal{R}_{22}^{\mathrm{T}}\mathcal{R}_{22} & \mathcal{R}_{12}^{\mathrm{T}}\mathcal{R}_{13} + \mathcal{R}_{22}^{\mathrm{T}}\mathcal{R}_{23} \\ \cdot\cdot & \cdots & \cdots \end{bmatrix} \tag{3.36}$$

where we have ignored terms of no interest. Equating the corresponding submatrices gives

$$\boldsymbol{K}_k = \mathcal{R}_{12}^{\mathrm{T}}\mathcal{R}_{11}[\mathcal{R}_{11}^{\mathrm{T}}\mathcal{R}_{11}]^{-1} = \mathcal{R}_{12}^{\mathrm{T}}\mathcal{R}_{11}^{-\mathrm{T}} \tag{3.37}$$

$$\boldsymbol{P}_{k|k} = \mathcal{R}_{12}^{\mathrm{T}}\mathcal{R}_{12} + \mathcal{R}_{22}^{\mathrm{T}}\mathcal{R}_{22} - \mathcal{R}_{12}^{\mathrm{T}}\mathcal{R}_{11}^{-\mathrm{T}}\mathcal{R}_{11}^{\mathrm{T}}\mathcal{R}_{12} = \mathcal{R}_{22}^{\mathrm{T}}\mathcal{R}_{22} \tag{3.38}$$

$$\boldsymbol{x}_{k|k} = \mathcal{R}_{12}^{\mathrm{T}}\mathcal{R}_{13} + \mathcal{R}_{22}^{\mathrm{T}}\mathcal{R}_{23} - \mathcal{R}_{12}^{\mathrm{T}}\mathcal{R}_{11}^{-\mathrm{T}}\mathcal{R}_{11}^{\mathrm{T}}\mathcal{R}_{13} = \mathcal{R}_{22}^{\mathrm{T}}\mathcal{R}_{23} \tag{3.39}$$

From the above equations, we only use $\mathcal{R}_{22}$ and $\mathcal{R}_{23}$ in the updating; thus there is no need to store the whole matrix.

Similarly, for the time update, we have

$$\begin{bmatrix} \widetilde{\boldsymbol{Q}}^{(t)^{1/2}} & \boldsymbol{0}_{N_s\times N_s} & \widetilde{\boldsymbol{Q}}^{(t)^{-\mathrm{T}/2}}\boldsymbol{S}^{(t)}\boldsymbol{R}^{(t)^{-1}}\boldsymbol{y}_k \\ \boldsymbol{P}_{k|k}^{1/2}\widetilde{\boldsymbol{A}}^{(t)^{\mathrm{T}}} & \boldsymbol{P}_{k|k}^{1/2} & \boldsymbol{P}_{k|k}^{-\mathrm{T}/2}\boldsymbol{x}_{k|k} \end{bmatrix} = \bar{\mathcal{Q}}\begin{bmatrix} \bar{\mathcal{R}}_{11} & \bar{\mathcal{R}}_{12} & \bar{\mathcal{R}}_{13} \\ \boldsymbol{0}_{N_s\times N_s} & \bar{\mathcal{R}}_{22} & \bar{\mathcal{R}}_{23} \end{bmatrix} \tag{3.40}$$

which yields

$$\boldsymbol{L}_k = \bar{\mathcal{R}}_{12}^{\mathrm{T}} \bar{\mathcal{R}}_{11}^{-\mathrm{T}} \tag{3.41}$$

$$\boldsymbol{P}_{k+1|k} = \bar{\mathcal{R}}_{11}^{\mathrm{T}} \bar{\mathcal{R}}_{11} \tag{3.42}$$

$$\boldsymbol{x}_{k+1|k} = \bar{\mathcal{R}}_{11}^{\mathrm{T}} \bar{\mathcal{R}}_{13} \tag{3.43}$$

As for the backward inference, a similar procedure is proposed as

$$\begin{bmatrix} \widetilde{\boldsymbol{Q}}^{(t)^{1/2}} & \boldsymbol{0}_{N_s \times N_s} & \widetilde{\boldsymbol{Q}}^{(t)^{-\mathrm{T}/2}} \boldsymbol{S}^{(t)} \boldsymbol{R}^{(t)^{-1}} \boldsymbol{y}_k \\ \boldsymbol{P}_{k|k}^{1/2} \widetilde{\boldsymbol{A}}^{(t)^{\mathrm{T}}} & \boldsymbol{P}_{k|k}^{1/2} & \boldsymbol{P}_{k|k}^{-\mathrm{T}/2} \boldsymbol{x}_{k|k} \\ \boldsymbol{0}_{N_s \times N_s} & \boldsymbol{P}_{k+1|N}^{1/2} \boldsymbol{L}_k^{\mathrm{T}} & \boldsymbol{P}_{k+1|N}^{-\mathrm{T}/2} \boldsymbol{x}_{k+1|N} \end{bmatrix} = \tilde{\mathcal{Q}} \begin{bmatrix} \tilde{\mathcal{R}}_{11} & \tilde{\mathcal{R}}_{12} & \tilde{\mathcal{R}}_{13} \\ \boldsymbol{0}_{N_s \times N_s} & \tilde{\mathcal{R}}_{22} & \tilde{\mathcal{R}}_{23} \\ \boldsymbol{0}_{N_s \times N_s} & \boldsymbol{0}_{N_s \times N_s} & \tilde{\mathcal{R}}_{33} \end{bmatrix} \tag{3.44}$$

As a consequence, we have

$$\boldsymbol{P}_{k|N} = \tilde{\mathcal{R}}_{22}^{\mathrm{T}} \tilde{\mathcal{R}}_{22} \tag{3.45}$$

$$\boldsymbol{x}_{k|N} = \tilde{\mathcal{R}}_{22}^{\mathrm{T}} \tilde{\mathcal{R}}_{23} \tag{3.46}$$

The robust M step

In the M step, the unknown parameters $\boldsymbol{\Gamma}^{(t+1)}$ and $\boldsymbol{\Sigma}^{(t+1)}$ are computed. Here, $\boldsymbol{\Sigma}^{(t+1)}$ must be kept symmetric and positive semi-definite at all iterations. For this purpose, the Cholesky decomposition is applied to simultaneously compute both matrices

$$\begin{bmatrix} \boldsymbol{H}^{(t)} & \boldsymbol{J}^{(t)^{\mathrm{T}}} \\ \boldsymbol{J}^{(t)} & \boldsymbol{K}^{(t)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{L}_{11} & \boldsymbol{0} \\ \boldsymbol{L}_{21} & \boldsymbol{L}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{L}_{11} & \boldsymbol{0} \\ \boldsymbol{L}_{21} & \boldsymbol{L}_{22} \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} \boldsymbol{L}_{11} \boldsymbol{L}_{11}^{\mathrm{T}} & \boldsymbol{L}_{11} \boldsymbol{L}_{21}^{\mathrm{T}} \\ \boldsymbol{L}_{21} \boldsymbol{L}_{11}^{\mathrm{T}} & \boldsymbol{L}_{21} \boldsymbol{L}_{21}^{\mathrm{T}} + \boldsymbol{L}_{22} \boldsymbol{L}_{22}^{\mathrm{T}} \end{bmatrix} \tag{3.47}$$

where we have defined

$$\boldsymbol{H}^{(t)} = \boldsymbol{\Pi}_0^{-1} + \sum_{k=1}^{N} \langle \boldsymbol{x}_k \boldsymbol{x}_k^{\mathrm{T}} \rangle \tag{3.48}$$

$$\boldsymbol{J}^{(t)} = \boldsymbol{\Gamma}_0 \boldsymbol{\Pi}_0^{-1} + \sum_{k=1}^{N} \langle \boldsymbol{z}_k \boldsymbol{x}_k^{\mathrm{T}} \rangle \tag{3.49}$$

$$\boldsymbol{K}^{(t)} = \boldsymbol{D}_0 + \sum_{k=1}^{N} \langle \boldsymbol{z}_k \boldsymbol{z}_k^{\mathrm{T}} \rangle + \boldsymbol{\Gamma}_0 \boldsymbol{\Pi}_0^{-1} \boldsymbol{\Gamma}_0^{\mathrm{T}} \tag{3.50}$$

By equating the sub-matrices, we have

$$\boldsymbol{\Gamma}^{(t+1)} = \boldsymbol{L}_{21} \boldsymbol{L}_{11}^{-1} \tag{3.51}$$

$$\boldsymbol{\Sigma}^{(t+1)} = \boldsymbol{L}_{22} \boldsymbol{L}_{22}^{\mathrm{T}} / (d_0 + N + 2N_s + N_o + 1) \tag{3.52}$$

Therefore, Eqn. (3.51) provides an efficient way to compute $\boldsymbol{\Gamma}^{(t+1)}$, while Eqn. (3.52) guarantees the symmetry and positive semi-definiteness of the covariance matrix. As a summary, the proposed robust implementation is shown in Figure 3.3.

1) Initialization

Set $\langle \boldsymbol{x}_{1|0} \rangle = \boldsymbol{\mu}_{1,0}$ and $\boldsymbol{P}_{1|0}^{1/2} = \boldsymbol{P}_{1,0}^{1/2}$

2) Robust E step

2.1) Forward inference

For $k = 1$ to N

Measurement update

$$\mathcal{R} = qr\left(\begin{bmatrix} \boldsymbol{R}^{(t)^{1/2}} & \boldsymbol{0}_{N_o \times N_s} & -\boldsymbol{R}^{(t)^{-\mathrm{T}/2}}\boldsymbol{y}_k \\ \boldsymbol{P}_{k|k-1}^{1/2}\boldsymbol{C}^{(t)^{\mathrm{T}}} & \boldsymbol{P}_{k|k-1}^{1/2} & \boldsymbol{P}_{k|k-1}^{-\mathrm{T}/2}\boldsymbol{x}_{k|k-1} \end{bmatrix}\right)$$

$\boldsymbol{x}_{k|k} = \mathcal{R}_{22}^{\mathrm{T}}\mathcal{R}_{23}$

$\boldsymbol{P}_{k|k} = \mathcal{R}_{22}^{\mathrm{T}}\mathcal{R}_{22}$

Time update

$$\bar{\mathcal{R}} = qr\left(\begin{bmatrix} \widetilde{\boldsymbol{Q}}^{1/2} & \boldsymbol{0}_{N_s \times N_s} & -\widetilde{\boldsymbol{Q}}^{-\mathrm{T}/2}\boldsymbol{S}^{(t)}\boldsymbol{R}^{(t)^{-1}}\boldsymbol{y}_k \\ \boldsymbol{P}_{k|k}^{1/2}\widetilde{\boldsymbol{A}}^{(t)^{\mathrm{T}}} & \boldsymbol{P}_{k|k}^{1/2} & \boldsymbol{P}_{k|k}^{-\mathrm{T}/2}\boldsymbol{x}_{k|k} \end{bmatrix}\right)$$

$\boldsymbol{L}_k = \bar{\mathcal{R}}_{12}^{\mathrm{T}}\bar{\mathcal{R}}_{11}^{-\mathrm{T}}$

$\boldsymbol{P}_{k+1|k} = \bar{\mathcal{R}}_{11}^{\mathrm{T}}\bar{\mathcal{R}}_{11}$

$\boldsymbol{x}_{k+1|k} = \bar{\mathcal{R}}_{11}^{\mathrm{T}}\bar{\mathcal{R}}_{13}$

End For

2.2) Backward inference

For $k = N$ to 1

$$\tilde{\mathcal{R}} = qr\left(\begin{bmatrix} \widetilde{\boldsymbol{Q}}^{1/2} & \boldsymbol{0}_{N_s \times N_s} & \widetilde{\boldsymbol{Q}}^{-\mathrm{T}/2}\boldsymbol{S}^{(t)}\boldsymbol{R}^{(t)^{-1}}\boldsymbol{y}_k \\ \boldsymbol{P}_{k|k}^{1/2}\widetilde{\boldsymbol{A}}^{(t)^{\mathrm{T}}} & \boldsymbol{P}_{k|k}^{1/2} & \boldsymbol{P}_{k|k}^{-\mathrm{T}/2}\boldsymbol{x}_{k|k} \\ \boldsymbol{0}_{N_s \times N_s} & \boldsymbol{P}_{k+1|k}^{1/2}\boldsymbol{L}_k^{\mathrm{T}} & \boldsymbol{P}_{k+1|k}^{-\mathrm{T}/2}\boldsymbol{x}_{k+1|N} \end{bmatrix}\right)$$

$\boldsymbol{x}_{k|N} = \tilde{\mathcal{R}}_{22}^{\mathrm{T}}\tilde{\mathcal{R}}_{23}$

$\boldsymbol{P}_{k|N} = \tilde{\mathcal{R}}_{22}^{\mathrm{T}}\tilde{\mathcal{R}}_{22}$

$\boldsymbol{P}_{k+1,k|N} = \boldsymbol{P}_{k+1|N}\boldsymbol{L}_k^{\mathrm{T}}$

End For

3) Robust M step

Compute matrix $\boldsymbol{H}^{(t)}$, $\boldsymbol{J}^{(t)}$, $\boldsymbol{K}^{(t)}$ defined in Eqns. (3.48)-(3.50)

$$L = \mathrm{chol}\left(\begin{bmatrix} \boldsymbol{H}^{(t)} & \boldsymbol{J}^{(t)^{\mathrm{T}}} \\ \boldsymbol{J}^{(t)} & \boldsymbol{K}^{(t)} \end{bmatrix}\right)$$

$\boldsymbol{\Gamma}^{(t+1)} = \boldsymbol{L}_{21}\boldsymbol{L}_{11}^{-1}$

$\boldsymbol{\Sigma}^{(t+1)} = \boldsymbol{L}_{22}\boldsymbol{L}_{22}^{\mathrm{T}}/(d_0 + N + 2N_s + N_o + 1)$

Figure 3.3 Robust EM algorithm for the stochastic SSM.

3.4 Variational Bayes

In the preceding section, we described the MAP estimates of the modal parameters. But that approach is not a full Bayesian analysis because the posterior distributions are not available. Instead of point estimation, this section introduces the variational Bayes (VB) approach to approximate the posterior distribution of modal parameters. Although several researchers investigated the VB for the linear SSM [94,129,131], this is the first time that a fully correlated stochastic SSM (Figure 3.1) is being considered. Specifically, out SSM involves correlated matrices $\boldsymbol{A}$ and $\boldsymbol{C}$ and correlated stochastic processes $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$, which give rise to a more complicated model. However, by virtue of the conjugate priors, the proposed VB algorithm has been the most elegant version so far.

3.4.1 Variational Bayes derivation

The VB algorithm tries to approximate the posteriors of unknown parameters and latent variables using surrogate distributions by minimizing their K-L divergence. Following the introduction in Section 1.2.2, we assume statistical independence between the sets of unknown parameters and latent variables in the surrogate distribution, but dependence within unknown parameter sets are considered to alleviate the side effect of this simplification, i.e.

$$q(\boldsymbol{\Gamma}, \boldsymbol{\Sigma}, \boldsymbol{x}_{1:N+1}) = q(\boldsymbol{\Gamma}, \boldsymbol{\Sigma}) q(\boldsymbol{x}_{1:N+1}) \tag{3.53}$$

is assumed. Given the above assumption and the probabilistic model shown in Figure 3.1, the free energy function gives

$$\begin{aligned} &\mathrm{F}[q(\boldsymbol{\Gamma}, \boldsymbol{\Sigma}), q(\boldsymbol{x}_{1:N+1})] \\ &= \iiint q(\boldsymbol{\Gamma}, \boldsymbol{\Sigma}) q(\boldsymbol{x}_{1:N+1}) \sum_{k=1}^{N} \ln f(\boldsymbol{z}_k | \boldsymbol{x}_k, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) \, d\boldsymbol{x}_k d\boldsymbol{\Gamma} d\boldsymbol{\Sigma} \\ &\quad - \iint q(\boldsymbol{\Gamma}, \boldsymbol{\Sigma}) \ln \frac{q(\boldsymbol{\Gamma}, \boldsymbol{\Sigma})}{\pi(\boldsymbol{\Gamma}, \boldsymbol{\Sigma})} d\boldsymbol{\Gamma} d\boldsymbol{\Sigma} - \int q(\boldsymbol{x}_{1:N+1}) \ln \frac{q(\boldsymbol{x}_{1:N+1})}{\pi(\boldsymbol{x}_1)} d\boldsymbol{x}_{1:N+1} \end{aligned} \tag{3.54}$$

Then, the VB approximation for the stochastic SSM can be derived based on the variational calculus as follows.

The VBE step

Take the functional derivative of $\mathrm{F}[q(\boldsymbol{\Gamma}, \boldsymbol{\Sigma}), q(\boldsymbol{x}_{1:N+1})]$ with respect to $q(\boldsymbol{x}_{1:N+1})$:

$$\frac{\partial}{\partial q(\boldsymbol{x}_{1:N+1})}\mathrm{F}[q(\boldsymbol{\Gamma},\boldsymbol{\Sigma}),q(\boldsymbol{x}_{1:N+1})] \tag{3.55}$$
$$=\sum_{k=1}^{N}\iint q(\boldsymbol{\Gamma},\boldsymbol{\Sigma})\ln f(\boldsymbol{z}_k|\boldsymbol{x}_k,\boldsymbol{\Gamma},\boldsymbol{\Sigma})\,d\boldsymbol{\Gamma}d\boldsymbol{\Sigma}-\ln q(\boldsymbol{x}_{1:N+1})+\ln\pi(\boldsymbol{x}_1)-1$$

Now, equating Eqn. (3.55) to zero yields

$$q^{(t+1)}(\boldsymbol{x}_{1:N+1})\propto\pi(\boldsymbol{x}_1)\exp\left\{-\frac{1}{2}\sum_{k=1}^{N}\langle(\boldsymbol{z}_k-\boldsymbol{\Gamma}\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}_k-\boldsymbol{\Gamma}\boldsymbol{x}_k)\rangle\right\} \tag{3.56}$$

For the expectation term, we have

$$\langle(\boldsymbol{z}_k-\boldsymbol{\Gamma}\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}_k-\boldsymbol{\Gamma}\boldsymbol{x}_k)\rangle$$
$$=d^{(t)}\left(\boldsymbol{z}_k-\boldsymbol{\Gamma}^{(t)}\boldsymbol{x}_k\right)^{\mathrm{T}}{\boldsymbol{D}^{(t)}}^{-1}\left(\boldsymbol{z}_k-\boldsymbol{\Gamma}^{(t)}\boldsymbol{x}_k\right) \tag{3.57}$$
$$+\boldsymbol{x}_k^{\mathrm{T}}\left\langle\left(\boldsymbol{\Gamma}-\boldsymbol{\Gamma}^{(t)}\right)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Gamma}-\boldsymbol{\Gamma}^{(t)}\right)\right\rangle\boldsymbol{x}_k$$

The (i,j) element of the expectation term in above equation is equal to

$$\left\langle\left(\boldsymbol{\Gamma}_i-\boldsymbol{\Gamma}_i^{(t)}\right)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Gamma}_j-\boldsymbol{\Gamma}_j^{(t)}\right)\right\rangle$$
$$=\mathrm{tr}\left[\left\langle\left(\boldsymbol{\Gamma}_j-\boldsymbol{\Gamma}_j^{(t)}\right)\left(\boldsymbol{\Gamma}_i-\boldsymbol{\Gamma}_i^{(t)}\right)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\right\rangle\right] \tag{3.58}$$
$$=\mathrm{tr}\left[\left(\boldsymbol{\Pi}_{ji}^{(t)}\boldsymbol{\Sigma}\right)\boldsymbol{\Sigma}^{-1}\right]=(N_s+N_o)\boldsymbol{\Pi}_{ji}^{(t)}$$

Thus,

$$\langle(\boldsymbol{z}_k-\boldsymbol{\Gamma}\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}_k-\boldsymbol{\Gamma}\boldsymbol{x}_k)\rangle=(N_s+N_o)\boldsymbol{\Pi}^{(t)} \tag{3.59}$$

Substituting Eqns. (3.57) and (3.59) into Eqn. (3.56) yields the surrogate distribution

$$q^{(t+1)}(\boldsymbol{x}_{1:N+1})$$
$$=Z^{-1}\exp\left\{-\tfrac{1}{2}\left[\left(\boldsymbol{x}_1-\boldsymbol{\mu}_{1,0}\right)^{\mathrm{T}}\boldsymbol{P}_{1,0}^{-1}\left(\boldsymbol{x}_1-\boldsymbol{\mu}_{1,0}\right)\right.\right. \tag{3.60}$$
$$\left.\left.+d^{(t)}\left(\boldsymbol{z}_k-\boldsymbol{\Gamma}^{(t)}\boldsymbol{x}_k\right)^{\mathrm{T}}{\boldsymbol{D}^{(t)}}^{-1}\left(\boldsymbol{z}_k-\boldsymbol{\Gamma}^{(t)}\boldsymbol{x}_k\right)+(N_s+N_o)\sum_{k=1}^{N}\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{\Pi}^{(t)}\boldsymbol{x}_k\right]\right\}$$

where Z is the normalizing constant. Obviously, the surrogate joint distribution of

$\boldsymbol{x}_{1:N+1}$ is multivariate normal because of its quadratic form in $\boldsymbol{x}_{1:N+1}$. Due to the existence of the last term in the bracket, we cannot directly apply the standard forward-backward inference algorithm introduced in Section 3.3.2 to calculate the first two conditional moments given $\boldsymbol{y}_{1:N}$. A trick [131] can be used here based on the mean-fluctuation decomposition so that the standard algorithm becomes applicable through augmentation. In detail, we need to make the following transformations before the inference:

$$\begin{aligned}&(N_s+N_o)\boldsymbol{\Pi}^{(t)}=\boldsymbol{U}^{(t)\mathrm{T}}\boldsymbol{U}^{(t)},\ \tilde{\boldsymbol{y}}_k=\begin{bmatrix}\boldsymbol{y}_k\\ \boldsymbol{0}_{N_s}\end{bmatrix},\ \tilde{\boldsymbol{C}}^{(t)}=\begin{bmatrix}\boldsymbol{C}^{(t)}\\ \boldsymbol{U}^{(t)}\end{bmatrix},\\ &\tilde{\boldsymbol{R}}^{(t)}=\begin{bmatrix}\boldsymbol{R}^{(t)} & \boldsymbol{0}_{N_d\times N_s}\\ \boldsymbol{0}_{N_s\times N_d} & \boldsymbol{I}_{N_s}\end{bmatrix},\ \tilde{\boldsymbol{S}}^{(t)}=\begin{bmatrix}\boldsymbol{S}^{(t)} & \boldsymbol{I}_{N_s}\end{bmatrix}\end{aligned} \tag{3.61}$$

where notations from Figure 3.2 are used. Once these transformations are done, the square-root filter strategy developed in Section 3.3.3 can be applied in the implementation.

The VBM step

Since conjugate priors are used for $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$, their posterior distributions will have the same form and we only need to update the distribution parameters. Taking the functional derivative of $\mathrm{F}\left[q_{\boldsymbol{\theta}}(\boldsymbol{\theta}), q_{\boldsymbol{X}_{1:N+1}}(\boldsymbol{x}_{1:N+1})\right]$ with respect to $q_{\boldsymbol{\Gamma\Sigma}}(\boldsymbol{\gamma},\boldsymbol{\sigma})$ yields

$$\begin{aligned}&\frac{\partial}{\partial q(\boldsymbol{\Gamma},\boldsymbol{\Sigma})}\mathrm{F}[q(\boldsymbol{\Gamma},\boldsymbol{\Sigma}),q(\boldsymbol{x}_{1:N+1})]\\ &=\sum_{k=1}^{N}\iint q(\boldsymbol{x}_{k:k+1})\ln f(\boldsymbol{z}_k|\boldsymbol{x}_k,\boldsymbol{\Gamma},\boldsymbol{\Sigma})\,d\boldsymbol{x}_{k:k+1}-\ln\frac{q(\boldsymbol{\gamma},\boldsymbol{\sigma})}{\pi(\boldsymbol{\gamma},\boldsymbol{\sigma})}-1\end{aligned} \tag{3.62}$$

Setting Eqn. (3.62) to be zero gives

$$\begin{aligned}&q(\boldsymbol{\Gamma},\boldsymbol{\Sigma})\\ &\propto\pi(\boldsymbol{\gamma},\boldsymbol{\sigma})\exp[\textstyle\sum_{k=1}^{N}\langle\ln f(\boldsymbol{z}_k|\boldsymbol{x}_k,\boldsymbol{\Gamma},\boldsymbol{\Sigma})\rangle]\\ &\propto|\boldsymbol{\Sigma}|^{-\frac{d_0+N+N_S+2N_O+1}{2}}\exp\Big\{-\frac{1}{2}\mathrm{tr}\big[\boldsymbol{D}_0\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Pi}_0^{-1}(\boldsymbol{\Gamma}-\boldsymbol{\Gamma}_0)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Gamma}-\boldsymbol{\Gamma}_0)+\\ &\langle(\boldsymbol{z}_k-\boldsymbol{\Gamma}\boldsymbol{x}_k)(\boldsymbol{z}_k-\boldsymbol{\Gamma}\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\rangle\big]\Big\}\\ &\propto|\boldsymbol{\Sigma}|^{-\frac{d^{(t+1)}+N_S+2N_O+1}{2}}\exp\Big\{-\frac{1}{2}\mathrm{tr}\Big[\boldsymbol{D}^{(t+1)}\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Pi}^{(t+1)^{-1}}\big(\boldsymbol{\Gamma}-\\ &\boldsymbol{\Gamma}^{(t+1)}\big)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\big(\boldsymbol{\Gamma}-\boldsymbol{\Gamma}^{(t+1)}\big)\Big]\Big\}\end{aligned} \tag{3.63}$$

where we have defined

$$\boldsymbol{\Pi}^{(t+1)} = \left[\boldsymbol{\Pi}_0^{-1} + \sum_{k=1}^{N}\langle \boldsymbol{x}_k \boldsymbol{x}_k^{\mathrm{T}}\rangle\right]^{-1} \tag{3.64}$$

$$\boldsymbol{\Gamma}^{(t+1)} = \left[\boldsymbol{\Gamma}_0\boldsymbol{\Pi}_0^{-1} + \sum_{k=1}^{N}\langle \boldsymbol{z}_k \boldsymbol{x}_k^{\mathrm{T}}\rangle\right]\boldsymbol{\Pi}^{(t+1)} \tag{3.65}$$

$$d^{(t+1)} = d_0 + N \tag{3.66}$$

$$\boldsymbol{D}^{(t+1)} = \boldsymbol{D}_0 + \sum_{k=1}^{N}\langle \boldsymbol{z}_k \boldsymbol{z}_k^{\mathrm{T}}\rangle + \boldsymbol{\Gamma}_0\boldsymbol{\Pi}_0^{-1}\boldsymbol{\Gamma}_0^{\mathrm{T}} - \boldsymbol{\Gamma}^{(t+1)}{\boldsymbol{\Pi}^{(t+1)}}^{-1}{\boldsymbol{\Gamma}^{(t+1)}}^{\mathrm{T}} \tag{3.67}$$

Comparing the above equations with the M step in the EM algorithm, a similar robust computation strategy can be developed using the Cholesky decomposition, but it is omitted here for brevity.

3.4.2 Posterior distribution of modal parameters

Although the surrogate distribution $q(\boldsymbol{\Gamma}, \boldsymbol{\Sigma})$ can be obtained by iteratively updating the VB algorithm, the distributions of modal parameters cannot directly be extracted. Since $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ are matrix normal-inverse Wishart distributed, the marginal distribution of $\boldsymbol{\Gamma}$ will be a matrix variate-*t* distribution [128]. However, there is no closed-form solution for the distribution of the modal parameters because of the eigenvalue decomposition involved in transforming $\boldsymbol{\Gamma}$ to the modal parameters. In the literature, the perturbation method [53] and asymptotic analysis [132] have been proposed to approximate these distributions. Here, we choose the first-order Taylor series expansion at the posterior mean value $\boldsymbol{\Gamma}^*$ to approximate the posterior joint distribution of the modal parameters. This approximation is expected to provide a good estimation, because this joint distribution is generally unimodal and there are small coefficients of variation for modal parameters. Since the modal frequencies and damping ratios must be positive for stable structures, the first-order Taylor expansion is operated on their logrithms

$$\begin{bmatrix} \ln f_i(\boldsymbol{\Gamma}) \\ \ln \xi_i(\boldsymbol{\Gamma}) \\ \boldsymbol{\phi}_i(\boldsymbol{\Gamma}) \end{bmatrix} \approx \begin{bmatrix} \ln f_i(\boldsymbol{\Gamma}^*) \\ \ln \xi_i(\boldsymbol{\Gamma}^*) \\ \boldsymbol{\phi}_i(\boldsymbol{\Gamma}^*) \end{bmatrix} + \left[\begin{matrix} \dfrac{\partial \ln f_i(\boldsymbol{\Gamma})}{\partial \mathrm{vec}(\boldsymbol{\Gamma})} \\ \dfrac{\partial \ln \xi_i(\boldsymbol{\Gamma})}{\partial \mathrm{vec}(\boldsymbol{\Gamma})} \\ \dfrac{\partial \boldsymbol{\phi}_i(\boldsymbol{\Gamma})}{\partial \mathrm{vec}(\boldsymbol{\Gamma})} \end{matrix}\right]_{\boldsymbol{\Gamma}=\boldsymbol{\Gamma}^*} [\mathrm{vec}(\boldsymbol{\Gamma}) - \mathrm{vec}(\boldsymbol{\Gamma}^*)] \tag{3.68}$$

where '$\mathrm{vec}(\cdot)$' means stacking the columns of the matrix into a column vector, and the partial derivatives with respect to this vector are defined to be row vectors.

Each element in the Jacobian matrix in Eqn. (3.68) can be calculated using the chain rule as follows:

$$\frac{\partial \ln f_i(\boldsymbol{\Gamma})}{\partial a_{mn}} = \frac{\partial \ln f_i(\boldsymbol{\Gamma})}{\partial f_i(\boldsymbol{\Gamma})}\frac{\partial f_i(\lambda_{ci})}{\partial \lambda_{ci}}\frac{\partial \lambda_{ci}}{\partial \lambda_i}\frac{\partial \lambda_i}{\partial a_{mn}}$$

$$= \frac{1}{f_i}\begin{bmatrix}\frac{\mathrm{Re}(\lambda_{ci})}{2\pi|\lambda_{ci}|} & \frac{\mathrm{Im}(\lambda_{ci})}{2\pi|\lambda_{ci}|}\end{bmatrix}\begin{bmatrix}\frac{\mathrm{Re}(\lambda_i)}{T_s|\lambda_i|^2} & \frac{\mathrm{Im}(\lambda_i)}{T_s|\lambda_i|^2} \\ -\frac{\mathrm{Im}(\lambda_i)}{T_s|\lambda_i|^2} & \frac{\mathrm{Re}(\lambda_i)}{T_s|\lambda_i|^2}\end{bmatrix}\begin{bmatrix}\frac{\partial \mathrm{Re}(\lambda_i)}{\partial a_{mn}} \\ \frac{\partial \mathrm{Im}(\lambda_i)}{\partial a_{mn}}\end{bmatrix} \tag{3.69}$$

$$\frac{\partial \ln \xi_i(\boldsymbol{\Gamma})}{\partial a_{mn}} = \frac{\partial \ln \xi_i(\boldsymbol{\Gamma})}{\partial \xi_i(\boldsymbol{\Gamma})}\frac{\partial \xi_i(\lambda_{ci})}{\partial \lambda_{ci}}\frac{\partial \lambda_{ci}}{\partial \lambda_i}\frac{\partial \lambda_i}{\partial a_{mn}}$$

$$= \frac{1}{\xi_i}\begin{bmatrix}-\frac{\mathrm{Im}(\lambda_{ci})^2}{|\lambda_{ci}|^3} & \frac{\mathrm{Re}(\lambda_{ci})\mathrm{Im}(\lambda_{ci})}{|\lambda_{ci}|^3}\end{bmatrix}\begin{bmatrix}\frac{\mathrm{Re}(\lambda_i)}{T_s|\lambda_i|^2} & \frac{\mathrm{Im}(\lambda_i)}{T_s|\lambda_i|^2} \\ -\frac{\mathrm{Im}(\lambda_i)}{T_s|\lambda_i|^2} & \frac{\mathrm{Re}(\lambda_i)}{T_s|\lambda_i|^2}\end{bmatrix}\begin{bmatrix}\frac{\partial \mathrm{Re}(\lambda_i)}{\partial a_{mn}} \\ \frac{\partial \mathrm{Im}(\lambda_i)}{\partial a_{mn}}\end{bmatrix} \tag{3.70}$$

$$\frac{\partial \boldsymbol{\phi}_i(\boldsymbol{\Gamma})}{\partial a_{mn}} = \boldsymbol{C}\begin{bmatrix}\frac{\partial \mathrm{Re}(\boldsymbol{\psi}_i)}{\partial a_{mn}} & \frac{\partial \mathrm{Im}(\boldsymbol{\psi}_i)}{\partial a_{mn}}\end{bmatrix}\boldsymbol{P}_2\boldsymbol{z} + \boldsymbol{C}\boldsymbol{\varphi}_i\boldsymbol{P}_2\frac{\partial \boldsymbol{z}}{\partial a_{mn}} \tag{3.71}$$

$$\frac{\partial \ln f_i(\boldsymbol{\Gamma})}{\partial c_{mn}} = \frac{\partial \ln \xi_i(\boldsymbol{\Gamma})}{\partial c_{mn}} = 0, \qquad \frac{\partial \boldsymbol{\phi}_i(\boldsymbol{\Gamma})}{\partial c_{mn}} = \boldsymbol{e}_m\boldsymbol{e}_n^{\mathrm{T}}\boldsymbol{\varphi}_i\boldsymbol{P}_2\boldsymbol{z} + \boldsymbol{C}\boldsymbol{\varphi}_i\boldsymbol{P}_2\frac{\partial \boldsymbol{z}}{\partial c_{mn}} \tag{3.72}$$

Computation of the partial derivatives of eigenvalues and eigenvectors is described in Appendix C. Similarly, we can compute the first-order Taylor expansion for all identified modes and then formulate them as a linear equation

$$\boldsymbol{\Xi}(\boldsymbol{\Gamma}) \approx \boldsymbol{\Xi}(\boldsymbol{\Gamma}^*) + \boldsymbol{J}(\boldsymbol{\Gamma})|_{\boldsymbol{\Gamma}=\boldsymbol{\Gamma}^*}[\mathrm{vec}(\boldsymbol{\Gamma}) - \mathrm{vec}(\boldsymbol{\Gamma}^*)] \tag{3.73}$$

where $\boldsymbol{\Xi}(\boldsymbol{\Gamma}) = [\ln f_1(\boldsymbol{\Gamma}) \quad \ln \xi_1(\boldsymbol{\Gamma}) \quad \boldsymbol{\phi}_1^{\mathrm{T}}(\boldsymbol{\Gamma}) \quad \ldots \quad \ln f_M(\boldsymbol{\Gamma}) \quad \ln \xi_M(\boldsymbol{\Gamma}) \quad \boldsymbol{\phi}_M^{\mathrm{T}}(\boldsymbol{\Gamma})]^{\mathrm{T}}$ and the Jacobian matrix $\boldsymbol{J}(\boldsymbol{\Gamma}) = [\boldsymbol{J}_1^{\mathrm{T}}(\boldsymbol{\Gamma}) \quad \ldots \quad \boldsymbol{J}_M^{\mathrm{T}}(\boldsymbol{\Gamma})]^{\mathrm{T}}$ with each submatrix being of the form shown in Eqn. (3.68).

Because the matrix variate-*t* distribution converges to a matrix normal distribution when the number of degrees of freedom tends to infinity [128], the distribution of $\boldsymbol{\Gamma}$ can be well approximated by a matrix normal when the sequence of data is long. Equivalently, $\mathrm{vec}(\boldsymbol{\Gamma})$ is approximately multivariate normal distributed with mean $\mathrm{vec}(\boldsymbol{\Gamma}^*)$ and covariance matrix $\boldsymbol{\Pi}^* \otimes \boldsymbol{D}^*/(d^* - N_s - N_o - 1)$, where $\boldsymbol{\Gamma}^*$, $\boldsymbol{\Pi}^*$, $\boldsymbol{D}^*$ and d^* are the parameters upon convergence, and '$\otimes$' stands for the Kronecker product. As a result, the parameter $\boldsymbol{\Xi}(\boldsymbol{\Gamma})$ will approximately follow a multivariate normal distribution with mean $\boldsymbol{\Xi}(\boldsymbol{\Gamma}^*)$ and covariance matrix $\boldsymbol{J}(\boldsymbol{\mu}_{\boldsymbol{\Gamma}}^*)(\boldsymbol{\Pi}^* \otimes \boldsymbol{D}^*)\boldsymbol{J}(\boldsymbol{\mu}_{\boldsymbol{\Gamma}}^*)^{\mathrm{T}}/(d^* - N_s - N_o - 1)$, yielding lognormal distributions for the posterior of the modal frequencies and damping ratios and multivariate normal distributions for the posteriors of mode shapes.

3.5 Gibbs sampler

The VB is efficient in approximating the posterior of modal parameters, but it is generally hard to evaluate the approximation error introduced by the mean-field variational family. On the contrary, the Gibbs sampler can provide an arbitrarily accurate approximation of the posterior at the expense of a much longer computation time. In this section, we will derive the Gibbs sampler for the OMA and use it as the baseline for the comparison of different approaches in the empirical study.

3.5.1 Derivation of the Gibbs sampler

Let us start with the joint distribution of the unknown parameters $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$, the latent variables $\boldsymbol{x}_{1:N+1}$, and the measurements $\boldsymbol{y}_{1:N}$

$$f(\boldsymbol{y}_{1:N}, \boldsymbol{x}_{1:N+1}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) = \pi(\boldsymbol{\Sigma})\pi(\boldsymbol{\Gamma}|\boldsymbol{\Sigma})f(\boldsymbol{x}_1)\prod_{k=1}^{N} f(\boldsymbol{y}_k, \boldsymbol{x}_{k+1}|\boldsymbol{x}_k, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) \tag{3.74}$$

In the Gibbs sampler, given the measurements $\boldsymbol{y}_{1:N}$, we need to derive the conditional distributions $p(\boldsymbol{x}_{1:N+1}|\boldsymbol{y}_{1:N}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})$ and $p(\boldsymbol{\Gamma}, \boldsymbol{\Sigma}|\boldsymbol{x}_{1:N+1}, \boldsymbol{y}_{1:N})$, and then iteratively sample from them, i.e.

(1) Given $\boldsymbol{\Gamma}^{(t)}$ and $\boldsymbol{\Sigma}^{(t)}$ and the measurement $\boldsymbol{y}_{1:N}$, generate a sample of latent variables according to

$$\boldsymbol{x}_{1:N+1}^{(t)} \sim p\left(\boldsymbol{x}_{1:N+1}|\boldsymbol{y}_{1:N}, \boldsymbol{\Gamma}^{(t)}, \boldsymbol{\Sigma}^{(t)}\right) \tag{3.75}$$

(2) Given $\boldsymbol{x}_{1:N+1}^{(t)}$, generate a sample of unknown parameters according to

$$\boldsymbol{\Gamma}^{(t+1)}, \boldsymbol{\Sigma}^{(t+1)} \sim p\left(\boldsymbol{\Gamma}, \boldsymbol{\Sigma}|\boldsymbol{x}_{1:N+1}^{(t)}, \boldsymbol{y}_{1:N}\right) \tag{3.76}$$

(3) Calculate modal parameters $f_i^{(t+1)}$, $\xi_i^{(t+1)}$ and $\boldsymbol{\phi}_i^{(t+1)}$ from $\boldsymbol{\Gamma}^{(t+1)}$ based on the procedure introduced in Section 3.2.1.

Sampling latent variables

Considering the joint distribution in Eqn. (3.74), the conditional distribution of latent variables is

$$\begin{aligned} &p\left(\boldsymbol{x}_{1:N+1}|\boldsymbol{y}_{1:N}, \boldsymbol{\Gamma}^{(t)}, \boldsymbol{\Sigma}^{(t)}\right) \\ &\propto \pi(\boldsymbol{x}_1)\exp\left\{-\frac{1}{2}\sum_{k=1}^{N}\left(\boldsymbol{z}_k - \boldsymbol{\Gamma}^{(t)}\boldsymbol{x}_k\right)^{\mathrm{T}}{\boldsymbol{\Sigma}^{(t)}}^{-1}\left(\boldsymbol{z}_k - \boldsymbol{\Gamma}^{(t)}\boldsymbol{x}_k\right)\right\} \end{aligned} \tag{3.77}$$

which is a multivariate normal distribution, but directly sampling from it is not feasible in this model because the computation of matrix inversions involved can be extremely demanding if N is large. As an alternative, we employ the forward-filtering-backward-sampling algorithm [133] to take advantage of the conditional Markov property.

The conditional distribution of latent variables can be written as

$$\begin{aligned} & p\left(\boldsymbol{x}_{1:N+1}|\boldsymbol{y}_{1:N}, \boldsymbol{\Gamma}^{(t)}, \boldsymbol{\Sigma}^{(t)}\right) \\ &= p\left(\boldsymbol{x}_{N+1}|\boldsymbol{y}_{1:N}, \boldsymbol{\Gamma}^{(t)}, \boldsymbol{\Sigma}^{(t)}\right) \prod_{k=1}^{N} p\left(\boldsymbol{x}_k|\boldsymbol{x}_{k+1:N+1}, \boldsymbol{y}_{1:N}, \boldsymbol{\Gamma}^{(t)}, \boldsymbol{\Sigma}^{(t)}\right) \\ &= p\left(\boldsymbol{x}_{N+1}|\boldsymbol{y}_{1:N}, \boldsymbol{\Gamma}^{(t)}, \boldsymbol{\Sigma}^{(t)}\right) \prod_{k=1}^{N} p\left(\boldsymbol{x}_k|\boldsymbol{x}_{k+1}, \boldsymbol{y}_{1:N}, \boldsymbol{\Gamma}^{(t)}, \boldsymbol{\Sigma}^{(t)}\right) \end{aligned} \tag{3.78}$$

where we have used the conditional Markov property in the second line. This factorization highlights the fact that we can sample from $p\left(\boldsymbol{x}_{1:N+1}|\boldsymbol{y}_{1:N}, \boldsymbol{\Gamma}^{(t)}, \boldsymbol{\Sigma}^{(t)}\right)$ by using a backward sampling strategy: first, draw a sample $\boldsymbol{x}_{N+1}$ from $p\left(\boldsymbol{x}_{N+1}|\boldsymbol{y}_{1:N}, \boldsymbol{\Gamma}^{(t)}, \boldsymbol{\Sigma}^{(t)}\right)$; then, conditioned on $\boldsymbol{x}_{k+1}$ draw from the conditional density $p\left(\boldsymbol{x}_k|\boldsymbol{x}_{k+1}, \boldsymbol{y}_{1:N}, \boldsymbol{\Gamma}^{(t)}, \boldsymbol{\Sigma}^{(t)}\right)$ and continue in this fashion until $k = 1$.

This backward sampling strategy decomposes the original problem into sub-problems of sampling from $p\left(\boldsymbol{x}_{N+1}|\boldsymbol{y}_{1:N}, \boldsymbol{\Gamma}^{(t)}, \boldsymbol{\Sigma}^{(t)}\right)$ and $p\left(\boldsymbol{x}_k|\boldsymbol{x}_{k+1}, \boldsymbol{y}_{1:N}, \boldsymbol{\Gamma}^{(t)}, \boldsymbol{\Sigma}^{(t)}\right)$, which are much simpler. In particular, $p\left(\boldsymbol{x}_{N+1}|\boldsymbol{y}_{1:N}, \boldsymbol{\Gamma}^{(t)}, \boldsymbol{\Sigma}^{(t)}\right)$ is the last-step distribution of the forward inference algorithm in Section 3.3.2, so that it is directly available. As for the PDF $p\left(\boldsymbol{x}_k|\boldsymbol{x}_{k+1}, \boldsymbol{y}_{1:N}, \boldsymbol{\Gamma}^{(t)}, \boldsymbol{\Sigma}^{(t)}\right)$, we can derive it from the conditional joint distribution of $\boldsymbol{x}_k$ and $\boldsymbol{x}_{k+1}$ given $\boldsymbol{y}_{1:N}$, $\boldsymbol{\Gamma}^{(t)}$ and $\boldsymbol{\Sigma}^{(t)}$, which is a multivariate normal distribution with the following mean and covariance matrix

$$\begin{bmatrix} \boldsymbol{x}_{k|N} \\ \boldsymbol{x}_{k+1|N} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_{k|k} + \boldsymbol{L}_k(\boldsymbol{x}_{k+1|N} - \boldsymbol{x}_{k+1|k}) \\ \boldsymbol{x}_{k+1|N} \end{bmatrix} \tag{3.79}$$

$$\boldsymbol{\Sigma}_{k,k+1|N} = \begin{bmatrix} \boldsymbol{P}_{k|k} + \boldsymbol{L}_k\left(\boldsymbol{P}_{k+1|N} - \boldsymbol{P}_{k+1|k}\right)\boldsymbol{L}_k^{\mathrm{T}} & \boldsymbol{L}_k\boldsymbol{P}_{k+1|N}^{\mathrm{T}} \\ \boldsymbol{P}_{k+1|N}\boldsymbol{L}_k^{\mathrm{T}} & \boldsymbol{P}_{k+1|N} \end{bmatrix} \tag{3.80}$$

where $\boldsymbol{L}_k$ is a gain matrix calculated in the forward filtering step as shown in Figure 3.2. Using the property of the multivariate normal distribution, we know that the conditional distribution of $\boldsymbol{x}_k$ given $\boldsymbol{x}_{k+1}^{(t)}$, $\boldsymbol{y}_{1:N}$, $\boldsymbol{\Gamma}^{(t)}$ and $\boldsymbol{\Sigma}^{(t)}$ is still multivariate normal, and the mean and covariance can be calculated as

$$\boldsymbol{x}_{k|k+1,N} = \boldsymbol{x}_{k|k} + \boldsymbol{L}_k\left(\boldsymbol{x}_{k+1}^{(t)} - \boldsymbol{x}_{k+1|k}\right) \tag{3.81}$$

$$\boldsymbol{P}_{k|k+1,N} = \boldsymbol{P}_{k|k} - \boldsymbol{L}_k\boldsymbol{P}_{k+1|k}\boldsymbol{L}_k^{\mathrm{T}} \tag{3.82}$$

Therefore, we can successfully sample $\boldsymbol{x}_k$ from this multivariate normal distribution by running only the forward inference. Interestingly, although we use the results from the backward inference in the derivation, we do not need to compute them because all terms in the backward inference cancel out. This is due to the following Markov property which we do not exploit in the derivation:

$$p\left(\boldsymbol{x}_k|\boldsymbol{x}_{k+1},\boldsymbol{y}_{1:N},\boldsymbol{\Gamma}^{(t)},\boldsymbol{\Sigma}^{(t)}\right)=p\left(\boldsymbol{x}_k|\boldsymbol{x}_{k+1},\boldsymbol{y}_{1:k},\boldsymbol{\Gamma}^{(t)},\boldsymbol{\Sigma}^{(t)}\right) \tag{3.83}$$

Sampling unknown parameters

Sampling the unknown parameters is a relatively easy task because the conjugate prior is applied in this model, so that the posterior is again the matrix normal inverse Wishart distribution. From the joint distribution in Eqn. (3.74), we have

$$\begin{aligned}
&p\left(\boldsymbol{\Gamma},\boldsymbol{\Sigma}|\boldsymbol{x}_{1:N+1}^{(t)},\boldsymbol{y}_{1:N}\right)\\
&\propto \pi(\boldsymbol{\gamma},\boldsymbol{\sigma})\exp\left[\sum_{k=1}^{N}\ln f\left(\boldsymbol{x}_{k+1}^{(t)},\boldsymbol{y}_k|\boldsymbol{x}_k^{(t)},\boldsymbol{\Gamma},\boldsymbol{\Sigma}\right)\right]\\
&\propto |\boldsymbol{\Sigma}|^{-\frac{d_0+N+N_S+2N_O+1}{2}}\exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{D}_0\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Pi}_0^{-1}(\boldsymbol{\Gamma}-\boldsymbol{\Gamma}_0)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Gamma}-\boldsymbol{\Gamma}_0)+\right.\right.\\
&\left.\left.\left(\boldsymbol{z}_k^{(t)}-\boldsymbol{\Gamma}\boldsymbol{x}_k^{(t)}\right)\left(\boldsymbol{z}_k^{(t)}-\boldsymbol{\Gamma}\boldsymbol{x}_k^{(t)}\right)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\right]\right\}\\
&\propto |\boldsymbol{\Sigma}|^{-\frac{d+N_S+2N_O+1}{2}}\exp\left\{-\frac{1}{2}\mathrm{tr}[\boldsymbol{D}\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Pi}^{-1}(\boldsymbol{\Gamma}-\boldsymbol{\mu}_\Gamma)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Gamma}-\boldsymbol{\mu}_\Gamma)]\right\}
\end{aligned} \tag{3.84}$$

where

$$\boldsymbol{\Pi}=\left[\boldsymbol{\Pi}_0^{-1}+\sum_{k=1}^{N}\boldsymbol{x}_k^{(t)}\boldsymbol{x}_k^{(t)^{\mathrm{T}}}\right]^{-1} \tag{3.85}$$

$$\boldsymbol{\mu}_\Gamma=\left[\boldsymbol{\Gamma}_0\boldsymbol{\Pi}_0^{-1}+\sum_{k=1}^{N}\boldsymbol{z}_k^{(t)}\boldsymbol{x}_k^{(t)^{\mathrm{T}}}\right]\boldsymbol{\Pi}^{(t+1)} \tag{3.86}$$

$$d=d_0+N \tag{3.87}$$

$$\boldsymbol{D}=\boldsymbol{D}_0+\sum_{k=1}^{N}\boldsymbol{z}_k^{(t)}\boldsymbol{z}_k^{(t)^{\mathrm{T}}}+\boldsymbol{\Gamma}_0\boldsymbol{\Pi}_0^{-1}\boldsymbol{\Gamma}_0^{\mathrm{T}}-\boldsymbol{\Gamma}^{(t+1)}\boldsymbol{\Pi}^{(t+1)^{-1}}\boldsymbol{\Gamma}^{(t+1)^{\mathrm{T}}} \tag{3.88}$$

Once these hyper-parameters are known, it is straightforward to sample from the standard distributions. We first sample $\boldsymbol{\Sigma}^{(t+1)}$ from the inverse Wishart distribution with d and $\boldsymbol{D}$, then conditioned on $\boldsymbol{\Sigma}^{(t+1)}$, we sample $\boldsymbol{\Gamma}^{(t+1)}$ from the matrix normal distribution with mean $\boldsymbol{\mu}_\Gamma$ and left and right covariance matrices $\boldsymbol{\Sigma}^{(t+1)}$, and $\boldsymbol{\Pi}$, respectively, which is equivalent to sampling from the multivariate normal distribution with mean $\mathrm{vec}(\boldsymbol{\mu}_\Gamma)$ and covariance $\boldsymbol{\Pi}\otimes\boldsymbol{\Sigma}^{(t+1)}$.

3.5.2 Robust implementation

As with the EM and VB algorithms, it is essential to develop a robust implementation of the Gibbs sampler to overcome the accumulated numerical error and guarantee the semi-definiteness of the covariance matrices.

For the latent variable sampling, we just copy the robust forward inference developed in Section 3.3.3. The only difference lies in the time update where more terms are used for sampling. Rewrite the Eqn. (3.40) here

$$\begin{bmatrix} \widetilde{\boldsymbol{Q}}^{(t)^{1/2}} & \boldsymbol{0}_{N_s\times N_s} & \widetilde{\boldsymbol{Q}}^{(t)^{-\mathrm{T}/2}}\boldsymbol{S}^{(t)}\boldsymbol{R}^{(t)^{-1}}\boldsymbol{y}_k \\ \boldsymbol{P}_{k|k}^{1/2}\widetilde{\boldsymbol{A}}^{(t)^{\mathrm{T}}} & \boldsymbol{P}_{k|k}^{1/2} & \boldsymbol{P}_{k|k}^{-\mathrm{T}/2}\boldsymbol{x}_{k|k} \end{bmatrix} = \bar{Q}\begin{bmatrix} \bar{\mathcal{R}}_{11} & \bar{\mathcal{R}}_{12} & \bar{\mathcal{R}}_{13} \\ \boldsymbol{0}_{N_s\times N_s} & \bar{\mathcal{R}}_{22} & \bar{\mathcal{R}}_{23} \end{bmatrix} \tag{3.89}$$

Exploiting the unitary nature of $\bar{Q}$, and multiplying this equation on the left by its transpose results in

$$\begin{bmatrix} \widetilde{\boldsymbol{Q}}^{(t)} + \widetilde{\boldsymbol{A}}^{(t)}\boldsymbol{P}_{k|k-1}\widetilde{\boldsymbol{A}}^{(t)^{\mathrm{T}}} & \widetilde{\boldsymbol{A}}^{(t)}\boldsymbol{P}_{k|k} & \widetilde{\boldsymbol{A}}^{(t)}\boldsymbol{x}_{k|k} + \boldsymbol{S}^{(t)}\boldsymbol{R}^{(t)^{-1}}\boldsymbol{y}_k \\ \cdots & \boldsymbol{P}_{k|k} & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix}$$
$$= \begin{bmatrix} \bar{\mathcal{R}}_{11}^{\mathrm{T}}\bar{\mathcal{R}}_{11} & \bar{\mathcal{R}}_{11}^{\mathrm{T}}\bar{\mathcal{R}}_{12} & \bar{\mathcal{R}}_{11}^{\mathrm{T}}\bar{\mathcal{R}}_{13} \\ \cdots & \bar{\mathcal{R}}_{12}^{\mathrm{T}}\bar{\mathcal{R}}_{12} + \bar{\mathcal{R}}_{22}^{\mathrm{T}}\bar{\mathcal{R}}_{22} & \cdots \\ \cdot\cdot & \cdots & \cdots \end{bmatrix} \tag{3.90}$$

Equating the sub-matrices yields

$$\boldsymbol{L}_k = \bar{\mathcal{R}}_{12}^{\mathrm{T}}\bar{\mathcal{R}}_{11}^{-\mathrm{T}} \tag{3.91}$$

$$\boldsymbol{P}_{k+1|k} = \bar{\mathcal{R}}_{11}^{\mathrm{T}}\bar{\mathcal{R}}_{11} \tag{3.92}$$

$$\boldsymbol{P}_{k|k} = \bar{\mathcal{R}}_{12}^{\mathrm{T}}\bar{\mathcal{R}}_{12} + \bar{\mathcal{R}}_{22}^{\mathrm{T}}\bar{\mathcal{R}}_{22} \tag{3.93}$$

$$\boldsymbol{x}_{k+1|k} = \bar{\mathcal{R}}_{11}^{\mathrm{T}}\bar{\mathcal{R}}_{13} \tag{3.94}$$

Substituting these equations into Eqns. (3.81) and (3.82) gives

$$\boldsymbol{x}_{k|k+1,N} = \boldsymbol{x}_{k|k} + \bar{\mathcal{R}}_{12}^{\mathrm{T}}\bar{\mathcal{R}}_{11}^{-\mathrm{T}}\left(\boldsymbol{x}_{k+1}^{(t)} - \bar{\mathcal{R}}_{11}^{\mathrm{T}}\bar{\mathcal{R}}_{13}\right) \tag{3.95}$$

$$\boldsymbol{P}_{k|k+1,N} = \bar{\mathcal{R}}_{12}^{\mathrm{T}}\bar{\mathcal{R}}_{12} + \bar{\mathcal{R}}_{22}^{\mathrm{T}}\bar{\mathcal{R}}_{22} - \bar{\mathcal{R}}_{12}^{\mathrm{T}}\bar{\mathcal{R}}_{11}^{-\mathrm{T}}\bar{\mathcal{R}}_{11}^{\mathrm{T}}\bar{\mathcal{R}}_{11}\bar{\mathcal{R}}_{11}^{-1}\bar{\mathcal{R}}_{12} = \bar{\mathcal{R}}_{22}^{\mathrm{T}}\bar{\mathcal{R}}_{22} \tag{3.96}$$

Therefore, we can sample $\boldsymbol{x}_k^{(t+1)}$ by first sampling a N_s-dimension standard normal variable $\boldsymbol{n}_k \in \mathbb{R}^{N_s}$, and then set

$$\boldsymbol{x}_k^{(t+1)} = \boldsymbol{x}_{k|k+1,N} + \bar{\mathcal{R}}_{22}^{\mathrm{T}}\boldsymbol{n}_k \tag{3.97}$$

1) Initialization

Set $\langle \boldsymbol{x}_{1|0} \rangle = \boldsymbol{\mu}_{1,0}$ and $\boldsymbol{P}_{1|0}^{1/2} = \boldsymbol{P}_{1,0}^{1/2}$

For $t = 1$ to T

2) Robust latent variables sampling

For $k = 1$ to N

Measurement update

$$\mathcal{R} = qr\left(\begin{bmatrix} \boldsymbol{R}^{(t)^{1/2}} & \boldsymbol{0}_{N_o \times N_s} & -\boldsymbol{R}^{(t)^{-\mathrm{T}/2}} \boldsymbol{y}_k \\ \boldsymbol{P}_{k|k-1}^{1/2} \boldsymbol{C}^{(t)^{\mathrm{T}}} & \boldsymbol{P}_{k|k-1}^{1/2} & \boldsymbol{P}_{k|k-1}^{-\mathrm{T}/2} \boldsymbol{x}_{k|k-1} \end{bmatrix}\right)$$

$$\boldsymbol{x}_{k|k} = \mathcal{R}_{22}^{\mathrm{T}} \mathcal{R}_{23}$$

$$\boldsymbol{P}_{k|k} = \mathcal{R}_{22}^{\mathrm{T}} \mathcal{R}_{22}$$

Time update

$$\bar{\mathcal{R}} = qr\left(\begin{bmatrix} \tilde{\boldsymbol{Q}}^{1/2} & \boldsymbol{0}_{N_s \times N_s} & -\tilde{\boldsymbol{Q}}^{-\mathrm{T}/2} \boldsymbol{S}^{(t)} \boldsymbol{R}^{(t)^{-1}} \boldsymbol{y}_k \\ \boldsymbol{P}_{k|k}^{1/2} \tilde{\boldsymbol{A}}^{(t)^{\mathrm{T}}} & \boldsymbol{P}_{k|k}^{1/2} & \boldsymbol{P}_{k|k}^{-\mathrm{T}/2} \boldsymbol{x}_{k|k} \end{bmatrix}\right)$$

$$\boldsymbol{L}_k = \bar{\mathcal{R}}_{12}^{\mathrm{T}} \bar{\mathcal{R}}_{11}^{-\mathrm{T}}, \ \boldsymbol{W}_k = \bar{\mathcal{R}}_{22}^{\mathrm{T}}$$

$$\boldsymbol{P}_{k+1|k} = \bar{\mathcal{R}}_{11}^{\mathrm{T}} \bar{\mathcal{R}}_{11}$$

$$\boldsymbol{x}_{k+1|k} = \bar{\mathcal{R}}_{11}^{\mathrm{T}} \bar{\mathcal{R}}_{13}$$

End For

$\boldsymbol{x}_{N+1}^{(t)} = \bar{\mathcal{R}}_{11}^{\mathrm{T}} (\bar{\mathcal{R}}_{13} + \boldsymbol{n}_0)$ where $\boldsymbol{n}_0 \sim \mathrm{N}(\boldsymbol{0}_{N_s}, \boldsymbol{I}_{N_s})$

For $k = N$ to 1

$\boldsymbol{x}_k^{(t)} = \boldsymbol{x}_{k|k} + \boldsymbol{L}_k (\boldsymbol{x}_{k+1}^{(t)} - \boldsymbol{x}_{k+1|k}) + \boldsymbol{W}_k \boldsymbol{n}_k$ where $\boldsymbol{n}_k \sim \mathrm{N}(\boldsymbol{0}_{N_s}, \boldsymbol{I}_{N_s})$

End For

3) Robust unknown parameters sampling

Compute matrix $\boldsymbol{H}^{(t)}$, $\boldsymbol{J}^{(t)}$, $\boldsymbol{K}^{(t)}$ defined in Eqns. (3.99)-(3.101)(3.48)

$$\boldsymbol{L} = \mathrm{chol}\left(\begin{bmatrix} \boldsymbol{H}^{(t)} & \boldsymbol{J}^{(t)^{\mathrm{T}}} \\ \boldsymbol{J}^{(t)} & \boldsymbol{K}^{(t)} \end{bmatrix}\right)$$

$\boldsymbol{\Sigma}^{(t+1)} \sim \mathrm{IW}(d, \boldsymbol{L}_{22} \boldsymbol{L}_{22}^{\mathrm{T}})$ and $\tilde{\boldsymbol{L}}_{22} = \mathrm{chol}(\boldsymbol{\Sigma}^{(t+1)})$

$\boldsymbol{\Gamma}^{(t+1)} = (\boldsymbol{L}_{21} + \tilde{\boldsymbol{L}}_{22} \tilde{\boldsymbol{n}}) \boldsymbol{L}_{11}^{-1}$ where $\mathrm{vec}(\tilde{\boldsymbol{n}}) \sim \mathrm{N}(\boldsymbol{0}_{N_s(N_o+N_s)}, \boldsymbol{I}_{N_o+N_s} \otimes \boldsymbol{I}_{N_s})$

4) Modal parameters computation following the procedure provided in Section 3.2.1.

End For

Figure 3.4 Robust Gibbs sampler for the OMA.

For the robust implementation of the unknown parameters sampling, we do the Cholesky decomposition

$$\begin{bmatrix} \boldsymbol{H}^{(t)} & \boldsymbol{J}^{(t)^{\mathrm{T}}} \\ \boldsymbol{J}^{(t)} & \boldsymbol{K}^{(t)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{L}_{11} & \boldsymbol{0} \\ \boldsymbol{L}_{21} & \boldsymbol{L}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{L}_{11} & \boldsymbol{0} \\ \boldsymbol{L}_{21} & \boldsymbol{L}_{22} \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} \boldsymbol{L}_{11} \boldsymbol{L}_{11}^{\mathrm{T}} & \boldsymbol{L}_{11} \boldsymbol{L}_{21}^{\mathrm{T}} \\ \boldsymbol{L}_{21} \boldsymbol{L}_{11}^{\mathrm{T}} & \boldsymbol{L}_{21} \boldsymbol{L}_{21}^{\mathrm{T}} + \boldsymbol{L}_{22} \boldsymbol{L}_{22}^{\mathrm{T}} \end{bmatrix} \tag{3.98}$$

where we have defined

$$\boldsymbol{H}^{(t)} = \boldsymbol{\Pi}_0^{-1} + \sum_{k=1}^{N} \boldsymbol{x}_k^{(t)} \boldsymbol{x}_k^{(t)\mathrm{T}} \tag{3.99}$$

$$\boldsymbol{J}^{(t)} = \boldsymbol{\Gamma}_0 \boldsymbol{\Pi}_0^{-1} + \sum_{k=1}^{N} \boldsymbol{z}_k^{(t)} \boldsymbol{x}_k^{(t)\mathrm{T}} \tag{3.100}$$

$$\boldsymbol{K}^{(t)} = \boldsymbol{D}_0 + \sum_{k=1}^{N} \boldsymbol{z}_k^{(t)} \boldsymbol{z}_k^{(t)\mathrm{T}} + \boldsymbol{\Gamma}_0 \boldsymbol{\Pi}_0^{-1} \boldsymbol{\Gamma}_0^{\mathrm{T}} \tag{3.101}$$

By equating the sub-matrices, we have

$$\boldsymbol{\Pi} = \boldsymbol{L}_{11}^{-\mathrm{T}} \boldsymbol{L}_{11}^{-1} \tag{3.102}$$

$$\boldsymbol{\mu}_{\Gamma} = \boldsymbol{L}_{21} \boldsymbol{L}_{11}^{-1} \tag{3.103}$$

$$\boldsymbol{D} = \boldsymbol{L}_{22} \boldsymbol{L}_{22}^{\mathrm{T}} \tag{3.104}$$

For the purpose of efficiently sampling $\boldsymbol{\Gamma}$ after sampling $\boldsymbol{\Sigma}$ based on d and $\boldsymbol{L}_{22}$, we apply the following procedure: do the Cholesky decomposition $\boldsymbol{\Sigma}^{(t+1)} = \tilde{\boldsymbol{L}}_{22} \tilde{\boldsymbol{L}}_{22}^{\mathrm{T}}$, then sample a standard normal matrix $\tilde{\boldsymbol{n}} \in \mathbb{R}^{N_o \times N_s}$, and finally set

$$\boldsymbol{\Gamma}^{(t+1)} = (\boldsymbol{L}_{21} + \tilde{\boldsymbol{L}}_{22} \tilde{\boldsymbol{n}}) \boldsymbol{L}_{11}^{-1} \tag{3.105}$$

As a summary, the procedures for robust implementation of the Gibbs sampler are listed in Figure 3.4.

3.6 Empirical studies

Having derived the EM and VB algorithms and the Gibbs sampler to solve the problem of OMA, this section provides an empirical study of their performance via a mass-spring system, a laboratory model, and a real structure – the One Rincon Tower in San Francisco, CA. The numerical example is used to evaluate the consistency in uncertainty quantification as well as the convergence property of the algorithms. The laboratory model shows their performance in a well-controlled condition, while the third example demonstrates their applicability to large-scale structural identification.

3.6.1 8-DoF mass-spring system

The mass-spring system is an idealization of a real structure. By controlling the true values of the parameters in the model, it is possible to use such a model to compare the performance of different algorithms. We adopt the mass-spring model in Figure 3.5, which has previously been used in by Cara et al. [134].

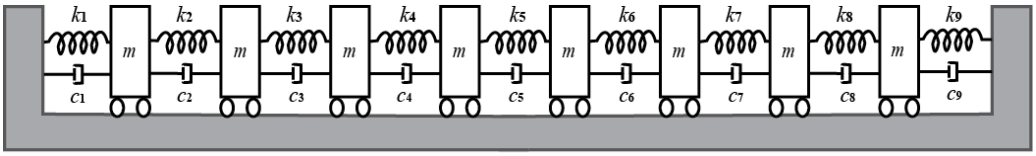


Figure 3.5 Eight DOFs mass-spring system.

The assumed model parameters are: $m = 1$, $k_i = 800i$ for $i = 1, \ldots, 9$, and Rayleigh damping with damping matrix $\boldsymbol{C}_d = 0.68\boldsymbol{M} + 1.743 \times 10^{-4}\boldsymbol{K}$, where $\boldsymbol{M}$ and $\boldsymbol{K}$ are the mass and stiffness matrices, respectively. An identical band-limited, Gaussian white-noise forcing function is applied at each DoF. Acceleration responses at DoFs 2, 4, 6 and 8 are recorded with the sampling frequency 25 Hz for 200 seconds. These "measured" responses are contaminated by an independent Gaussian white-noise process with variance equal to 25% of the largest acceleration variance.

Table 3.1 lists the identified modal frequencies and damping ratios for the three methods. Note that the EM algorithm provides only a point estimate. For VB and Gibbs sampling, the coefficient of variation (COV) or standard deviation (STD) of each estimate is provided. For Gibbs sampling, 1,800 samples are used after discarding burn-in samples. Comparing the point estimates (MLE or posterior means), all three methods are consistent, though slight biases exist. Among the three methods, EM seems to provide the best point estimates, while VB and Gibbs give similar performance. The estimated COVs of VB are always smaller than those of Gibbs sampler, by about 15%. It is noted that the estimated COVs of the modal frequencies by both methods are extremely small.

Table 3.1 Identified Modal Parameters of the mass-spring system.

Mode	Frequencies (Hz)						Damping Ratios (%)					
	True	EM	VB		Gibbs		True	EM	VB		Gibbs	
			Mean	COV(%)	Mean	COV(%)			Mean	STD	Mean	STD
1	2.94	2.92	2.94	0.26	2.93	0.34	2.00	2.00	1.93	0.21	1.87	0.34
2	5.87	5.88	5.87	0.20	5.88	0.21	1.24	1.16	1.30	0.18	1.29	0.21
3	8.60	8.57	8.56	0.14	8.56	0.16	1.10	1.07	1.09	0.14	1.03	0.16
4	11.19	11.21	11.22	0.11	11.20	0.13	1.09	0.79	0.78	0.10	0.84	0.13
5	13.78	13.77	13.74	0.12	13.75	0.15	1.15	1.22	1.24	0.12	1.33	0.16
6	16.52	16.54	16.54	0.11	16.55	0.13	1.23	1.07	1.20	0.11	1.12	0.13
7	19.54	19.54	19.53	0.12	19.56	0.17	1.35	1.34	1.78	0.14	1.33	0.17
8	23.12	23.09	23.08	0.11	23.06	0.14	1.50	1.56	1.77	0.13	1.44	0.15

It is interesting to see that the estimated COVs of the frequencies are almost identical to the estimated standard deviations of the damping ratios for all physical modes. At this time we are unable to explain the reason for this phenomenon. Many

previous studies have shown similar results, see, e.g., Refs [53,57,58], though they did not explicitly state this result.

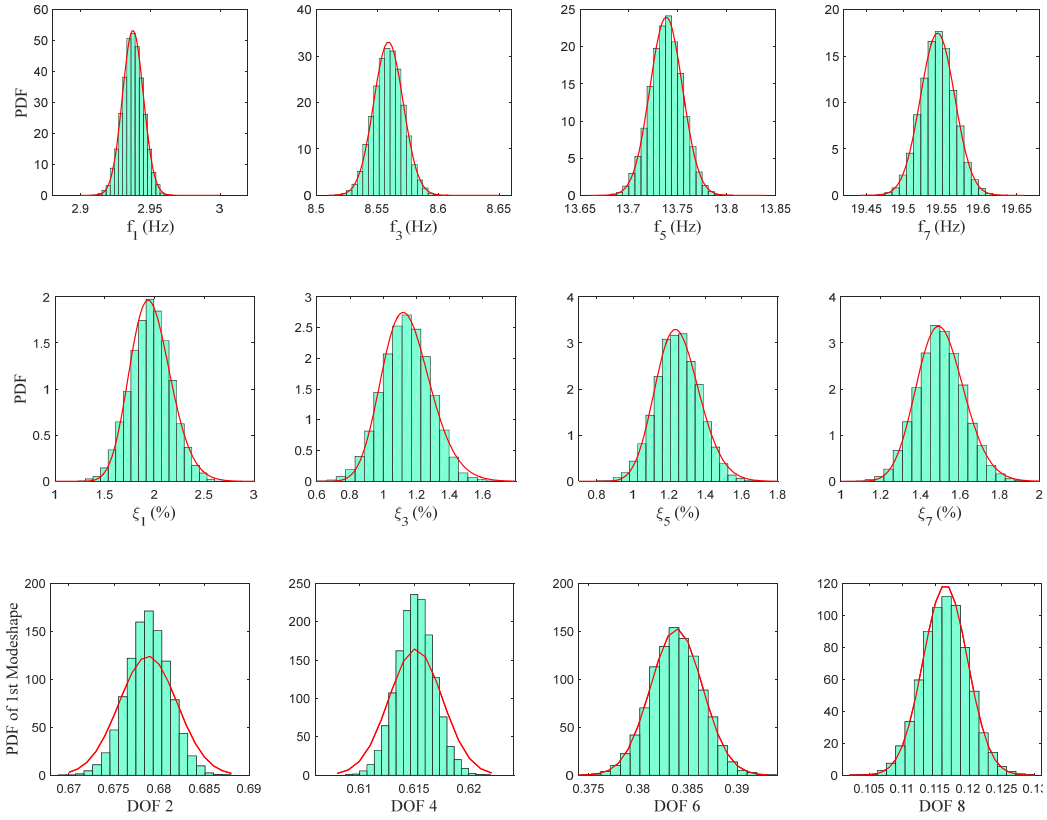


Figure 3.6 First-order approximation of posterior distributions of modal parameters.

In order to examine the accuracy of the posterior distributions of modal parameters obtained by the first-order Taylor series expansion, 10,000 samples are generated from the posterior distribution of parameter $\boldsymbol{\Gamma}$ estimated by VB, as plotted in Figure 3.6. The first-order approximation of the distribution of modal frequencies matches with the normalized frequency diagrams of the generated samples nearly perfectly. The approximation for the damping ratio is also good in spite of the larger COVs. For the mode shapes, discrepancies exist between the distributions obtained by first-order approximation and by sampling, in spite of the small COVs. This has to do with the normalization of the mode shapes in sampling – all mode shape samples are normalized to have unit length. As a result, the normalized frequency diagram does not correctly represent the true distribution of the mode shapes. Nevertheless, the mean values are in close agreement.

Comparison of cumulative distribution functions (CDFs) and empirical CDFs of the modal frequencies and damping ratios as obtained by VB and Gibbs sampler,

respectively, are shown in Figure 3.7 and Figure 3.8, respectively. Again, we can see that the distributions estimated by VB show less dispersion than those obtained from the Gibbs sampler, though the two distributions are relatively close in most cases. Plots of the identified mode shapes, with the data for each DoF shown as a boxplot, are illustrated in Figure 3.9. The same normalization method is used for the two sets of mode shapes. Both the VB and Gibbs sampler provide accurate estimates of the mode shapes, with the variations in identified mode shapes by VB being smaller than those estimated by the Gibbs sampler. In conclusion, the VB estimator is consistent, but tends to under-estimate the uncertainty.

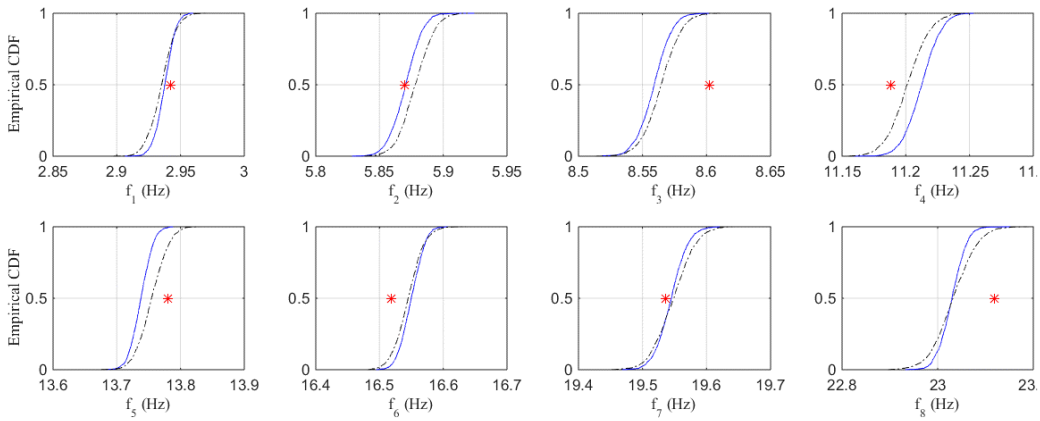


Figure 3.7 Empirical CDF of identified modal frequencies.
Solid line: VB; Dash-dot line: Gibbs sampler; Star: true value

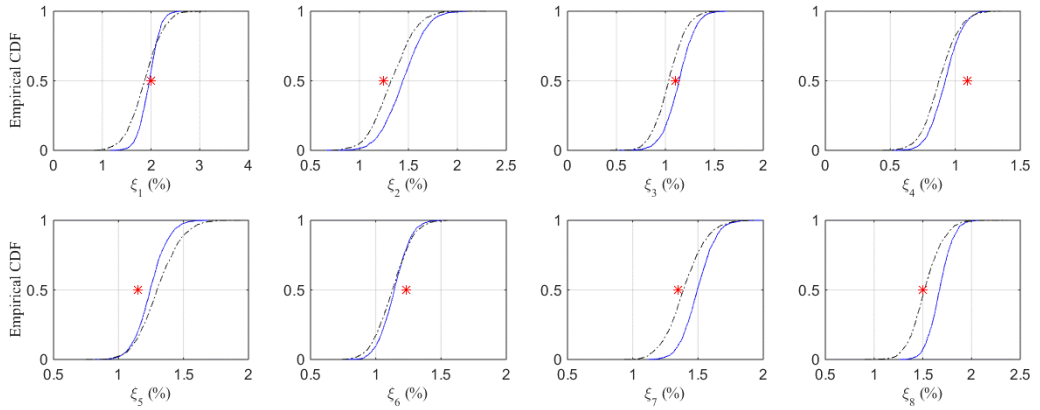


Figure 3.8 Empirical CDF of identified damping ratios.
Solid line: VB; Dash-dot line: Gibbs sampler; Star: true value

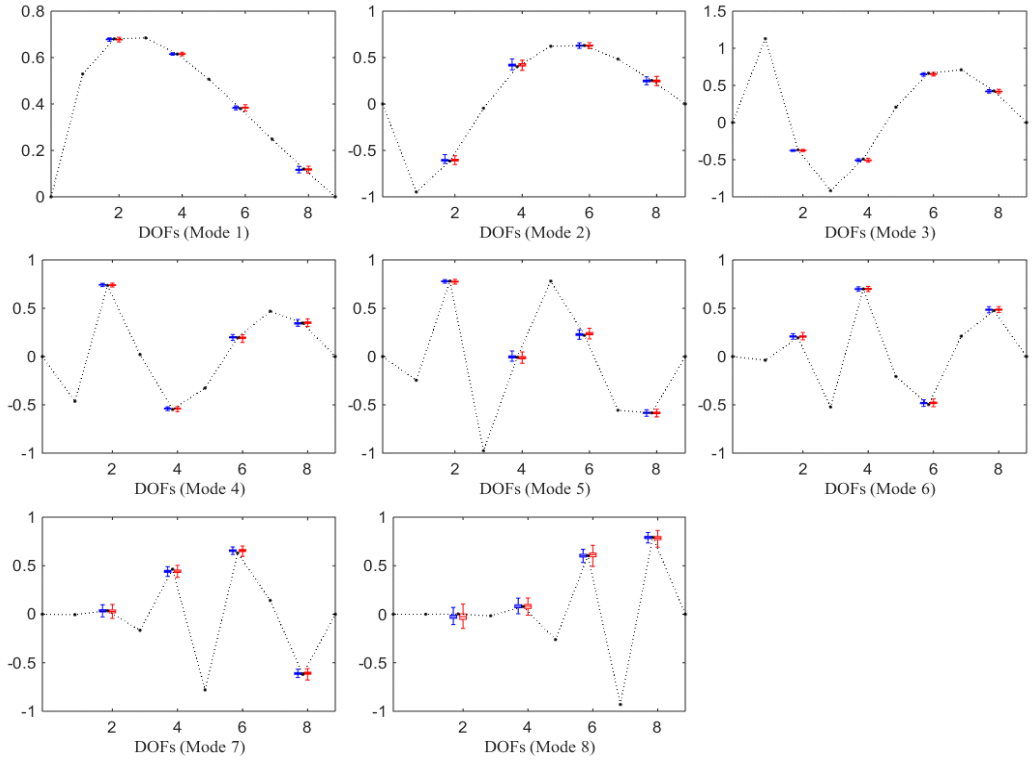


Figure 3.9 Boxplot of identified mode shapes.
Blue: VB; Red: Gibbs sampler; Dot line: true value

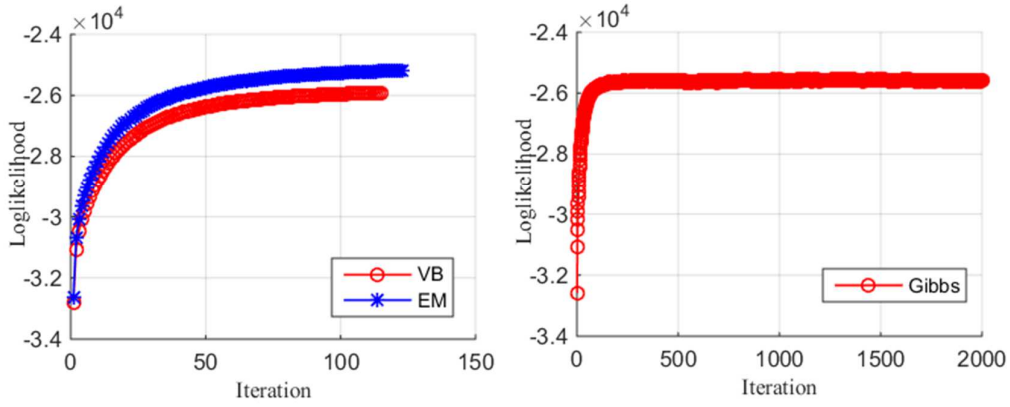


Figure 3.10 Convergence of VB, EM and Gibbs sampler.

We have seen that the VB estimation is consistent with the Gibbs sampler and can provide an approximate posterior distribution; however, as an iterative method, it is computationally demanding, especially when many iterations are needed to converge. Wang & Titterington [98] have proved almost linear convergence of the VB algorithm

for the exponential family of distributions. Here, we empirically investigate its convergence property, as shown in Figure 3.10. The VB converges a little faster than the EM algorithm. Since the EM algorithm has a linear convergence rate [130], this shows that the VB converges at least with a linear rate. For comparison, the convergence curve of the Gibbs sampler is also shown in Figure 3.10. It is observed that the Gibbs sampler converges at about 200 iterations, but many more iterations are needed to provide enough samples to estimate the distribution. In general, the Gibbs sampler is more computationally demanding.

In the SSM-based modal identification, one has to specify the order of the system, i.e. two times of the number of physical modes. Since we do not know the exact order in advance, one favored practice is to specify a high order and then to identify the physical modes from the estimated model. In this dissertation, we do not solve this problem rigorously, but only provide empirical evidence that the physical modes can be identified based on the uncertainty information. For this 8-DoF mass-spring system, an order of 24 is specified in the state-space model, so that there are 4 spurious modes. Figure 3.11 shows the COVs of identified frequencies and standard deviations of damping ratios for both the VB and Gibbs sampler. Spurious modes correspond to the peaks in the graph, i.e. spurious modes have significantly larger uncertainties than the physical modes. The intuitive explanation for this phenomenon, similar to that of a stabilization diagram [45], is that the physical modes should be consistent in different realizations, while the spurious computational and noise modes have no reason to be focused.

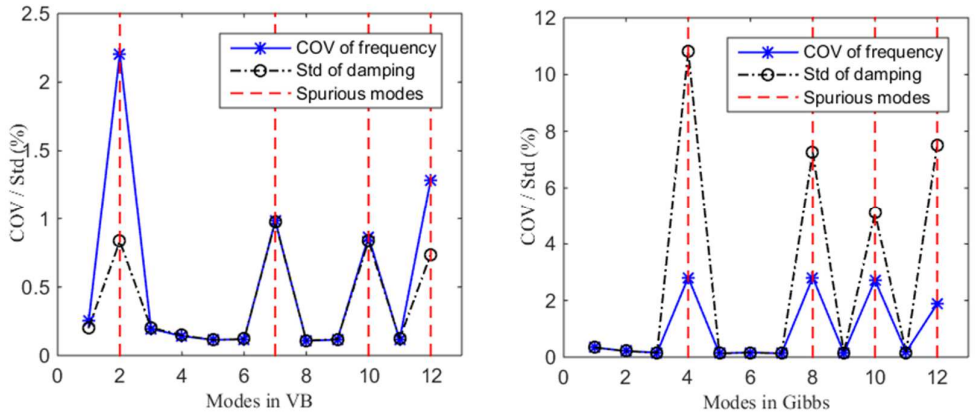
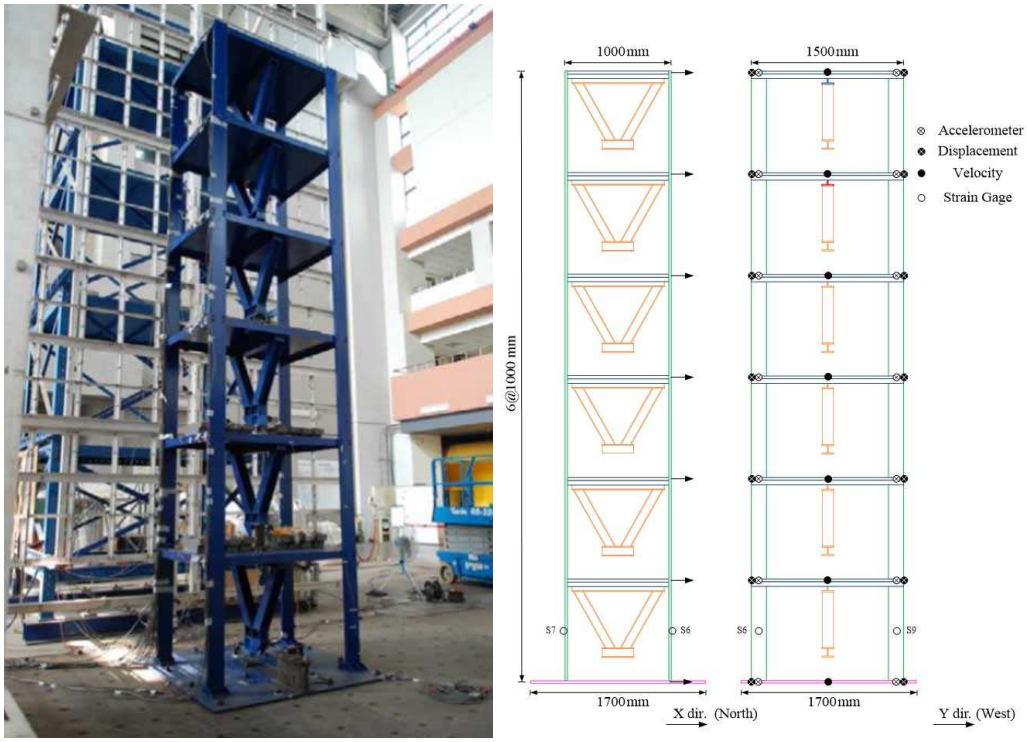


Figure 3.11 Spurious modes identification based on uncertainty.

3.6.2 6-story shear-type building

The second example considered is a six-story 1/4-scale shear-type building [135], which is constructed on a large shaking table (Figure 3.12a) at the National Center for

Research in Earthquake Engineering at National Taiwan University. The structure has a single bay with 1.0 m × 1.5 m floor area and an inter-story height of 1m. The size of the column and beam is 150 mm × 25 mm (rectangular section) and 50 mm × 50 mm (L section), respectively. The dead load is simulated by lead-block units fixed on the steel plate of each floor, yielding a total mass of each floor of 862.85 kg except the roof floor that is 803.98 kg. Because of the rectangular cross section of the columns, the structure has two orthogonal axes: a flexurally weak axis and a flexurally strong axis. In the experiment, only the responses along the weak axis were measured under white-noise base motion using 14 displacement meters, 14 accelerometers and 6 velocimeters shown in Figure 3.12b.



(a) test structure on the NCREE shake table; (b) configuration of sensors

Figure 3.12 Six-story 1/4-scale steel frame building structure

Though many sensors are available, only the measured accelerations are used to identify the modal parameters in this dissertation. The raw data was sampling with 200 Hz for 100 seconds. According to Ref. [136], all lateral modes along the weak axis are below 15 Hz, so that the raw data is resampled down to 40 Hz to remove higher modes.

Initialized with the parameters identified by the SSI algorithm [47], Figure 3.13 shows the convergence of the EM, VB and Gibbs sampler. In this example, the EM algorithm converged after 274 iterations, much slower than the VB that took 180

iterations. It is a well-known drawback of the EM algorithm that it may converge extremely slowly when it approaches to the stationary point [92]. The VB has a higher convergence rate, but there is no theoretical guarantee that it will be always faster than the EM algorithm. In the Gibbs sampler, 2000 samples are generated with the first 200 samples discarded as the burn-in stage.

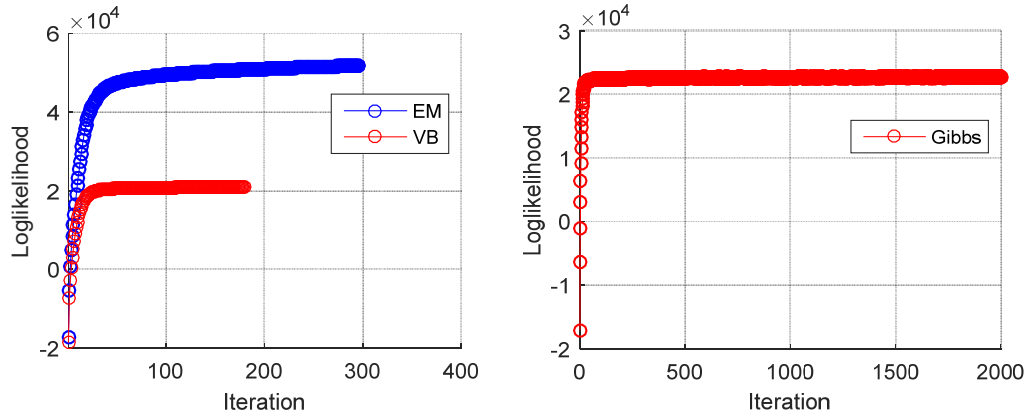


Figure 3.13 Convergence of EM, VB and Gibb sampling.

The system model order was selected as 24, which means at most 12 modes can be extracted. In this model, the identified modes with negative damping ratios are first filtered out because, for a stable structure, the damping ratios must be positive. Then, we plot the COVs of frequencies and standard deviations of damping ratios for all the remaining modes in Figure 3.14. According to the criterion that spurious modes should be much more uncertain than the physical ones, we immediately identify that the first 6 modes in both the VB and Gibbs sampler identification are physical.

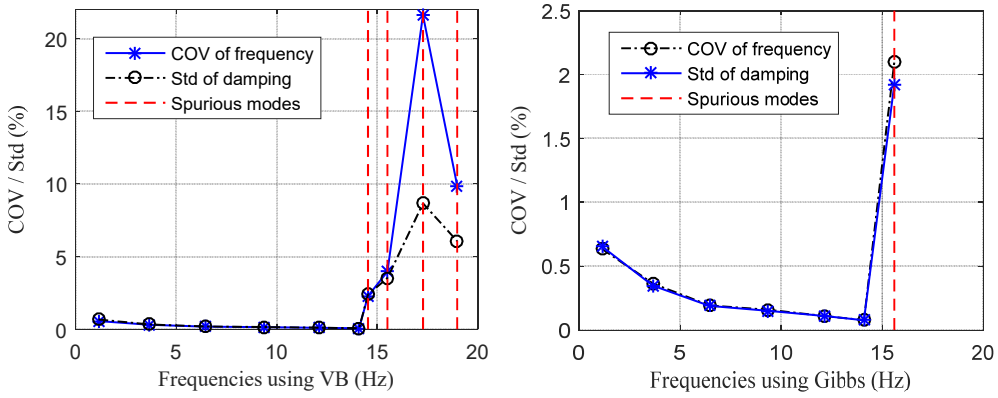


Figure 3.14 Identification of spurious modes.

After removing the spurious modes, the identified frequencies and damping ratios are listed in Table 3.2. The three algorithms give almost identical estimates of modal frequencies, while slight differences exist in the identified damping ratios because there are large uncertainties in these estimates. The COVs or STDs identified by VB correspond well to those identified by the Gibbs sampler. We observed that the VB underestimated the uncertainties in the example of the mass-spring system, but this phenomenon is not apparent in this example. In fact, we can say that the VB gives good estimations of the parameter uncertainties, which is further validated by the CDF plots in Figure 3.15 and Figure 3.16. Again, we see, for each mode, the COV of the frequency is close to the STD of the damping ratio. Furthermore, the information about identified mode shapes are plotted in Figure 3.17, in which the dashed line represents the mode shapes identified by the EM algorithm; the circle and the crossing show the difference between the mean values of mode shapes identified by the VB and Gibbs sampler with those identified by EM, respectively; the associated error bars denote the 3 standard deviation provided in the VB and Gibbs sampler. This graph shows a good correspondence of these algorithms in terms of both the point estimates and the uncertainties.

Table 3.2 Identified Modal Parameters of the shear-type building.

Mode	Frequencies (Hz)					Damping Ratios (%)				
	EM	VB		Gibbs		EM	VB		Gibbs	
		Mean	COV(%)	Mean	COV(%)		Mean	STD	Mean	STD
1	1.15	1.14	0.56	1.13	0.65	1.73	1.30	0.67	1.33	0.63
2	3.67	3.66	0.32	3.66	0.35	1.49	1.29	0.33	1.30	0.36
3	6.47	6.45	0.18	6.46	0.19	0.66	0.74	0.19	0.67	0.19
4	9.37	9.37	0.15	9.37	0.15	0.54	0.67	0.15	0.58	0.16
5	12.12	12.13	0.10	12.12	0.11	0.38	0.44	0.11	0.42	0.11
6	14.10	14.10	0.07	14.10	0.08	0.22	0.21	0.07	0.23	0.08

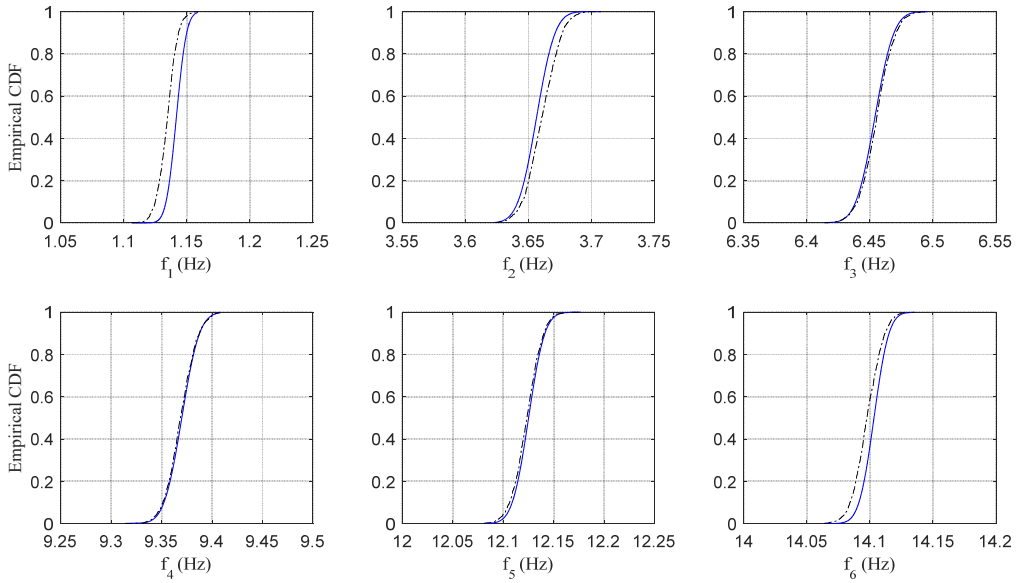


Figure 3.15 CDF plot of frequencies.
Solid line: VB; Dash-dot line: Gibbs sampler

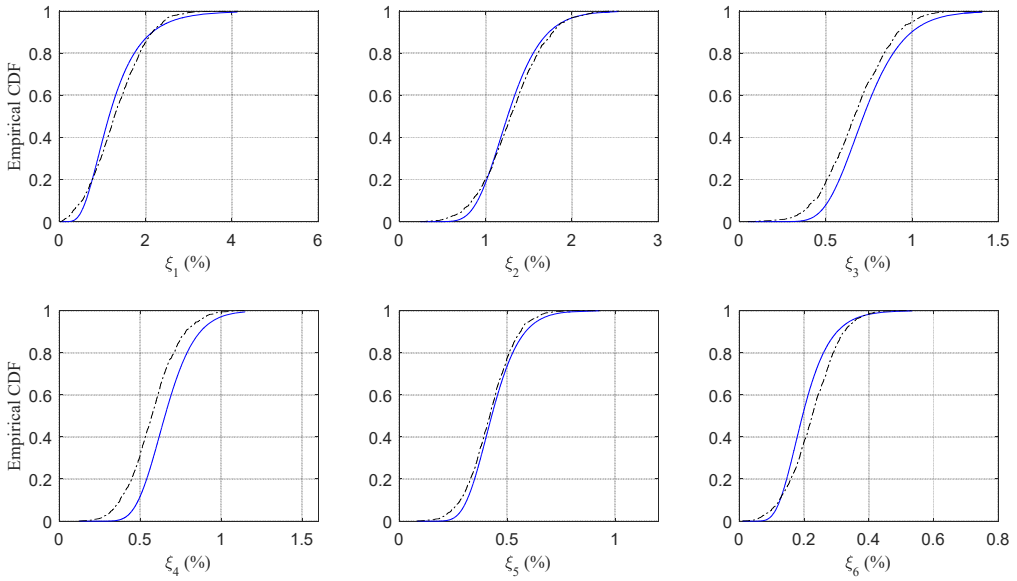


Figure 3.16 CDF plot of damping ratios.
Solid line: VB; Dash-dot line: Gibbs sampler

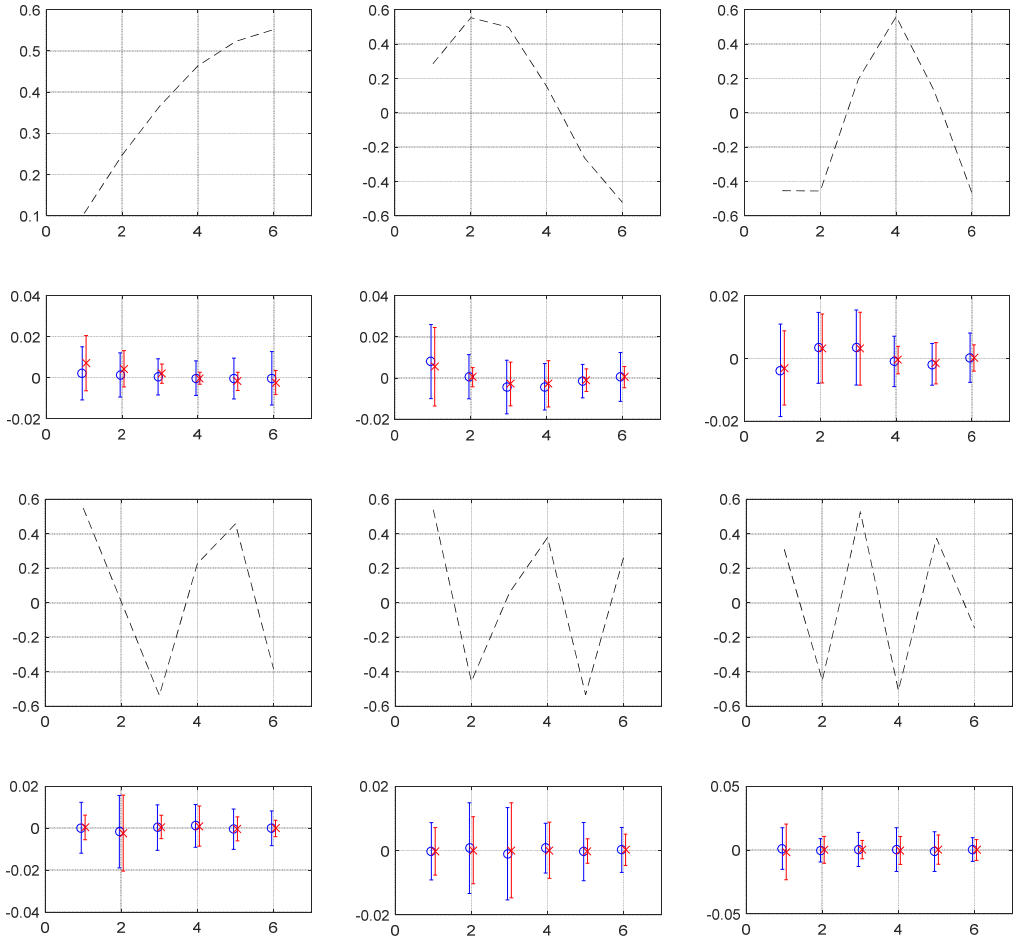


Figure 3.17 Mode shapes and their uncertainties.
Upper six plots: modes 1-3; lower six plots: modes 4-6

3.6.3 One Rincon Hill Tower

The One Rincon Hill Tower (ORHT), a landmark building in San Francisco, California, is a 64-story reinforced-concrete, shear-wall building. It is also the tallest building in the United States designed using performance-based seismic design procedures. In order to evaluate the design and assess the condition of the building after a future earthquake, extensive seismic instrumentation of the building was jointly conducted by the California Strong Motion Instrumentation Program and the National Strong Motion Project and managed by United States Geological Survey [137]. A 72-channel seismic monitoring system (Figure 3.18) was installed to stream real-time acceleration data throughout the building. Two sets of recorded ambient vibration time series are available in the website of the Center for Engineering Strong Motion Data (www.strongmotioncenter.org).

There are a total of 72 accelerometers, of which 4 measure vertical motions and are not considered in this study, since only the lateral modes are of interest. Furthermore, in order to capture the rotational motion of the building, only levels with 3 accelerometers, one in north-south (NS) and two in east-west (EW1 and EW2) directions of each floor, are included in the analysis. The raw data has a sampling frequency of 200 Hz and duration of 230 seconds. According to Ref. [137], modes with frequencies below 7 Hz are of interest. Therefore, for the sake of reducing the computational burden, the raw data is resampled down to 20 Hz so that all higher modes are filtered out.

Initialized with the parameters identified by the SSI algorithm [47] with a model order of 120, the EM, VB and Gibbs sampler are used for the OMA of the ORHW. Figure 2.19 shows the iteration steps of each algorithm, where 178 iterations (about 115 min on a Digital Storm laptop with Intel® Core™ i7 CPU @2.50 GHz and RAM 16.0 GB) are spent for the EM, 84 iterations (about 55 min) are spent for VB, and 800 samples (about 534 min) are used for the Gibbs sampler. One may argue that 800 samples are not enough for the Gibbs sampler, but running more than 10 hours was not practical. One must admit that, while the VB is the most efficient algorithm, all three methods are perhaps too computationally demanding for practical use. One way to accelerate the computations is to loosen the convergence criterion, which requires the relative change of log-likelihood less than 10^{-5} in current setting.

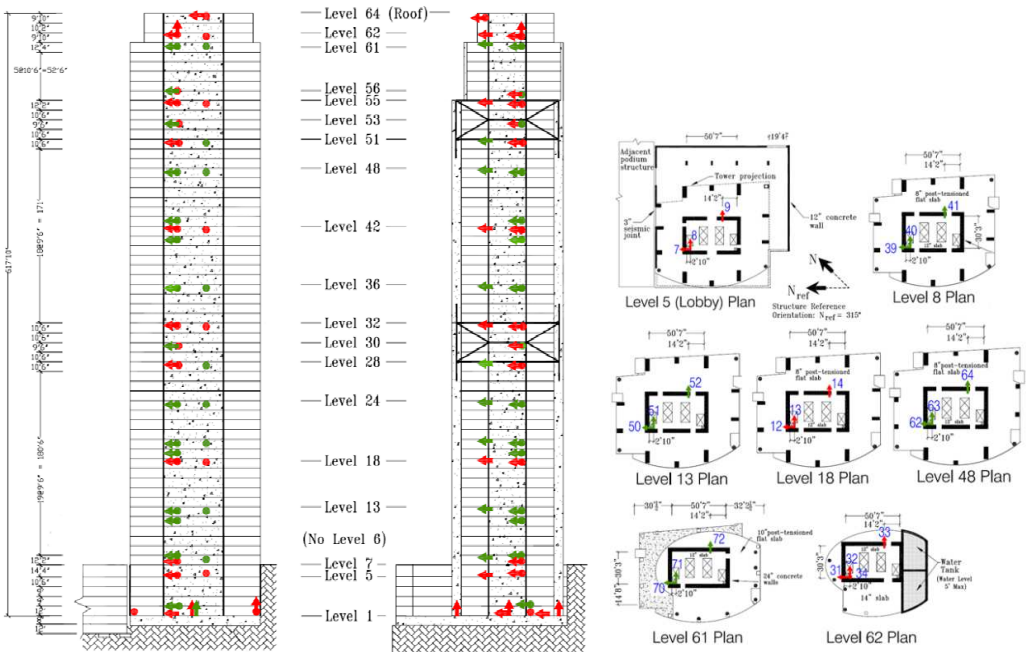


Figure 3.18 Configuration of accelerometers along the vertical and plan of the building (www.strongmotioncenter.org).

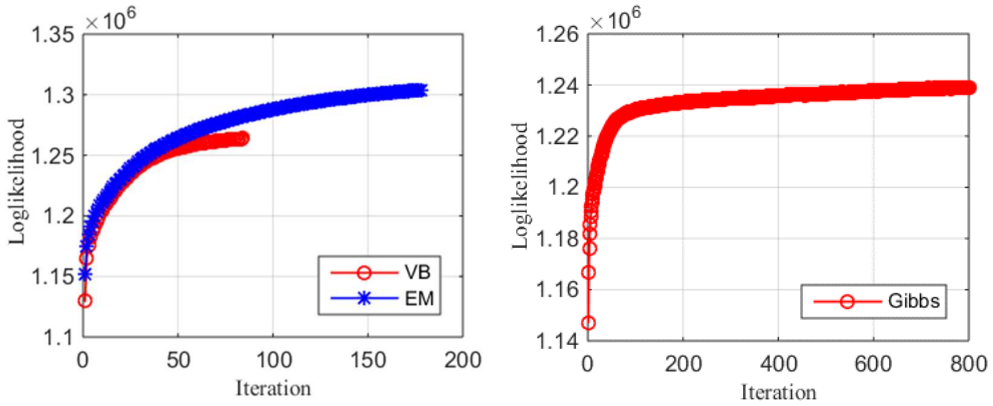


Figure 3.19 Convergence of EM, VB and Gibbs sampler in OMA of ORHT.

Identification of spurious modes is shown in Figure 3.20 for VB and Gibbs sampler, where the averaged power spectral density of measured motions at all locations is also plotted. When the order of the model is high, plots of the COV of frequencies and the standard deviation of damping ratios might be too complex to allow accurate identification of the spurious mode. To address this problem, more requirements should be enforced. Here, we require the parameter the complex mode indicator λ_{min} to be less than 0.01, i.e., the mode shape not to be too complex. The second requirement is that each mode corresponds to a peak in the average power spectral density (PSD) over all measurements.

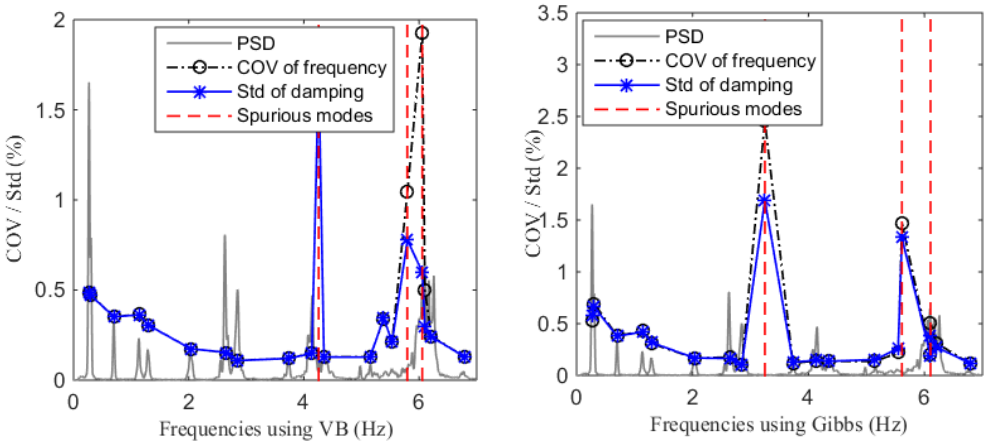


Figure 3.20 Identification of spurious modes in VB and Gibbs sampler.
PSD: scaled power spectrum density

Table 3.3 Identified Modal Parameters using EM, VB and Gibbs Sampler

Mode	Frequencies (Hz)					Damping Ratios (%)					Mode shape
	EM	VB		Gibbs		EM	VB		Gibbs		
		Mean	COV(%)	Mean	COV(%)		Mean	STD	Mean	STD	
1	0.27	0.27	0.49	0.27	0.58	1.01	0.96	0.49	1.09	0.53	1st EW
2	0.30	0.30	0.48	0.30	0.67	1.60	0.94	0.48	1.86	0.67	1st NS
3	0.70	0.70	0.35	0.70	0.38	1.39	1.30	0.36	1.22	0.38	1st Torsion
4	1.14	1.14	0.37	1.14	0.42	2.97	2.10	0.36	2.71	0.43	2nd EW
5	1.30	1.30	0.30	1.30	0.32	2.02	1.75	0.30	1.71	0.31	2nd NS
6	2.03	2.03	0.17	2.03	0.17	1.03	0.90	0.17	0.76	0.17	2nd Torsion
7	2.64	2.65	0.15	2.65	0.17	1.11	0.92	0.15	1.02	0.18	3rd EW
8	2.85	2.85	0.11	2.85	0.11	0.59	0.51	0.11	0.45	0.11	3rd NS
9	3.74	3.73	0.12	3.73	0.13	0.96	0.82	0.12	0.80	0.12	3rd Torsion
10	4.14	4.14	0.15	4.13	0.15	1.13	1.23	0.15	1.14	0.14	4th EW
11	4.35	4.35	0.13	4.34	0.14	0.90	1.00	0.13	1.03	0.14	4th NS
12	5.15	5.15	0.13	5.15	0.15	1.11	1.08	0.13	1.12	0.14	4th Torsion
13	5.31	5.39	0.33	-	-	3.95	4.64	0.33	-	-	5th EW
14	5.54	5.53	0.22	5.54	0.29	2.14	2.17	0.21	2.21	0.23	5th NS
15	6.11	6.10	0.27	6.10	0.20	1.66	1.66	0.27	1.26	0.20	6th EW
16	6.20	6.21	0.25	6.21	0.28	1.76	1.63	0.26	1.70	0.32	6th NS
17	6.81	6.80	0.13	6.78	0.13	1.15	1.08	0.13	0.97	0.12	5th Torsion

After removal of spurious modes, the identified physical modes are listed in Table 3.3. The three approaches yield similar modal parameters with an additional mode identified by the EM and VB algorithms. It is observed that the identified damping ratio is much larger than that of all other modes, because this mode is not well excited and a large estimation error may exist. Çelebi et al. [137] identified only 12 modes using the SSI algorithm; 5 additional modes are identified in this study. As for the uncertainty estimation the COVs of VB are closer to those of the Gibbs sampler, indicating that the VB may perform better for real data. Figure 3.21 and Figure 3.22 show the CDFs and empirical CDFs of modal frequencies and damping ratios, respectively. Generally speaking, the CDFs identified by VB are quite close to those by the Gibbs sampler, and some of them are almost identical. The mode shapes are illustrated in Figure 3.23, though the uncertainty information is not shown for the sake of brevity. Again, we see that the COVs of the estimated frequencies are approximately equal to the standard deviations of the estimated damping ratios.

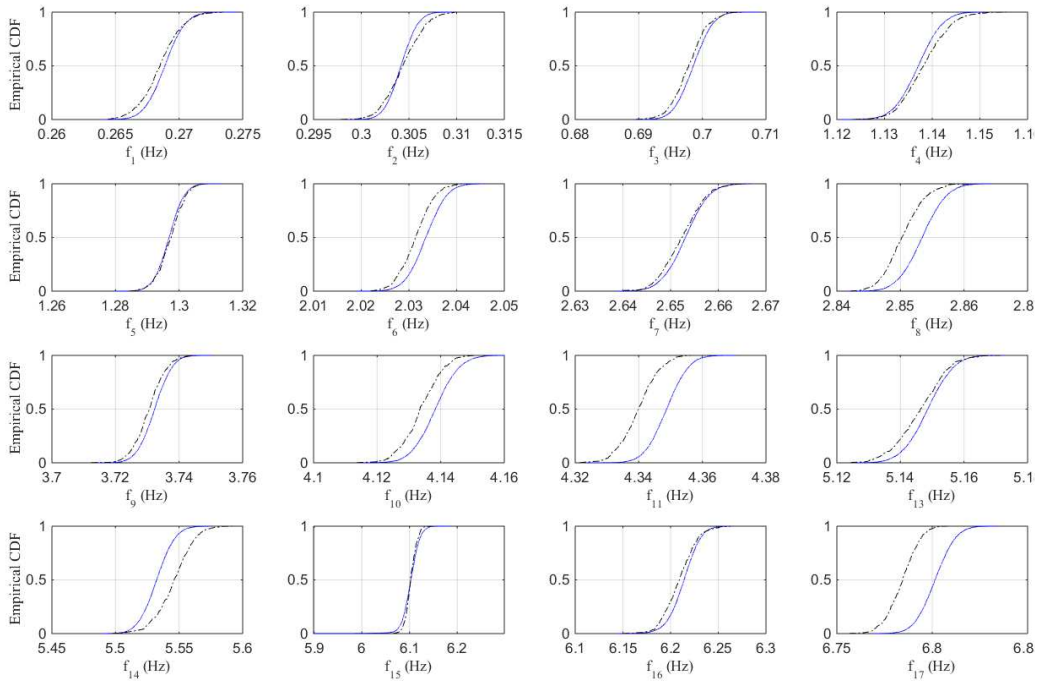


Figure 3.21 Empirical CDF of identified modal frequencies.
Solid line: VB; Dash-dot line: Gibbs sampler

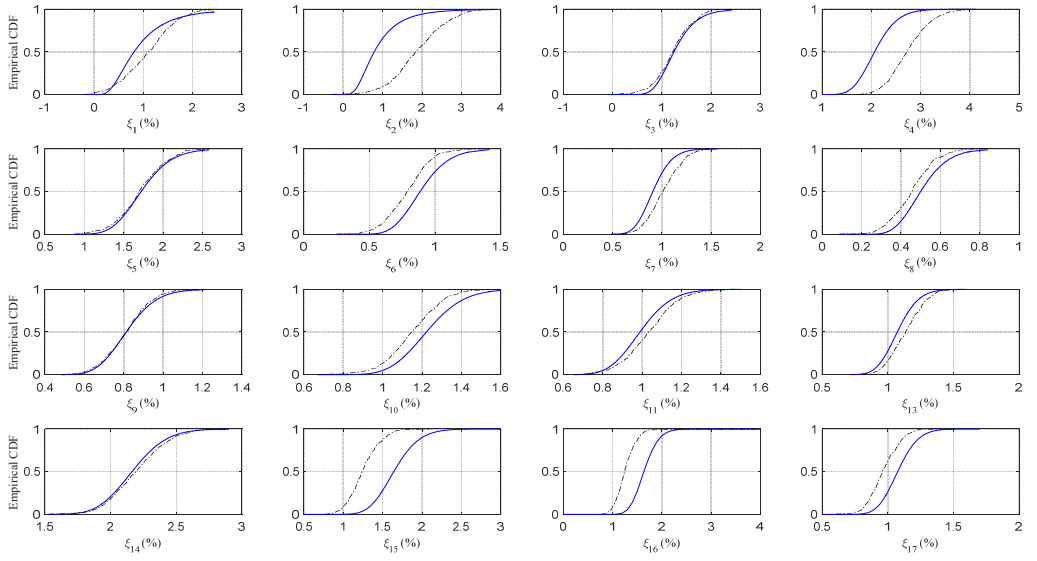


Figure 3.22 Empirical CDF of identified damping ratios.
Solid line: VB; Dash-dot line: Gibbs sampler

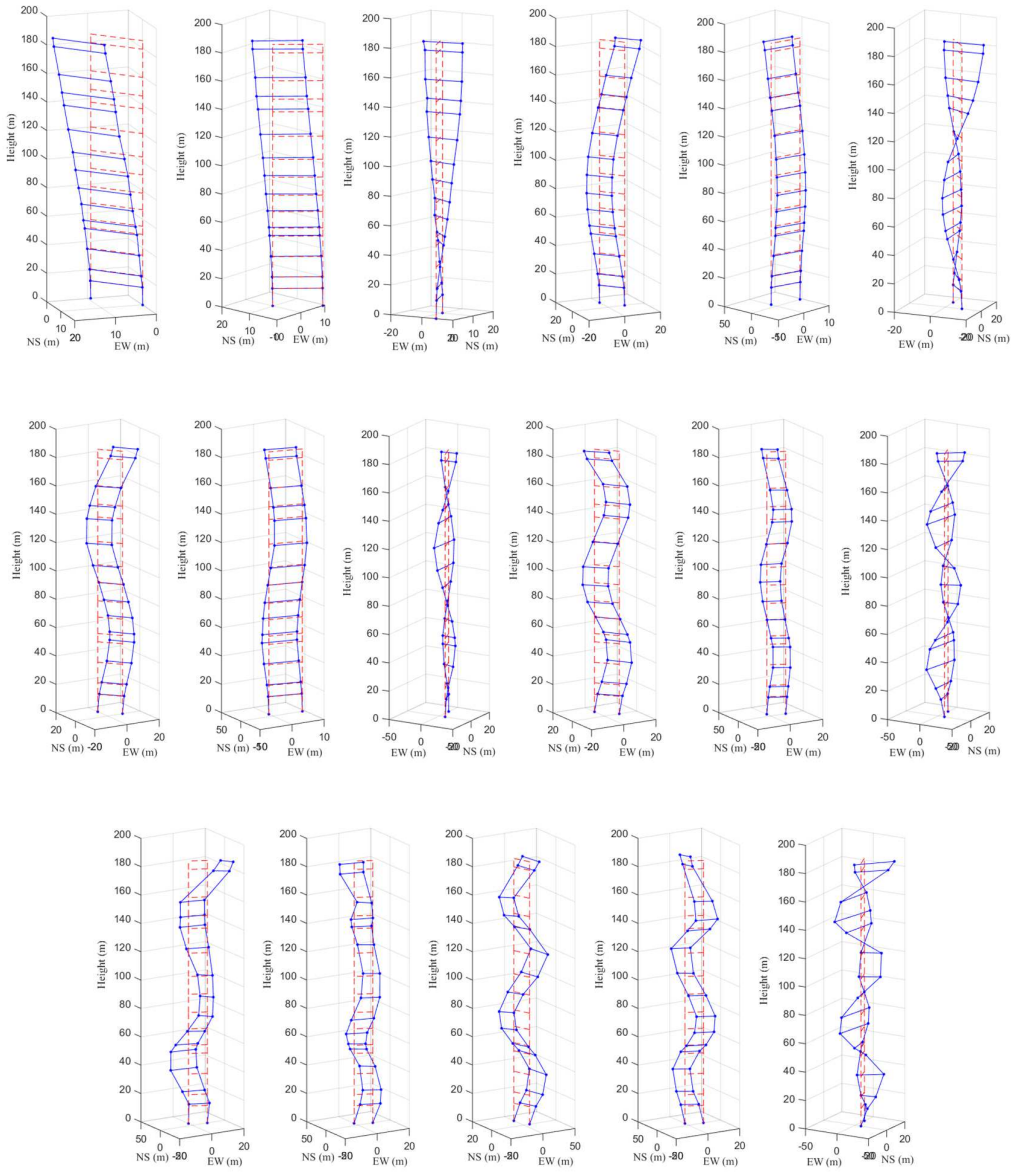


Figure 3.23 Mode shapes identified using VB.

3.7 Conclusions

The problem of operational modal analysis is considered in this chapter. Starting with the state-space representation of the structural system, a probabilistic model is constructed incorporating various sources of uncertainties. After an appropriate selection of error models and priors for the unknown parameters, the constructed

probabilistic model belongs to the conjugate-exponential family, enabling the direct application of the EM algorithm, the variational Bayes and the Gibbs sampler for the Bayesian inference of modal parameters. Three empirical studies, starting with a numerical model and ending with a real super-tall building are conducted to show the applicability of the proposed approaches and to compare their performance. The following conclusions are derived:

(1) The uncertainties in modal parameters are successfully quantified based on the proposed probabilistic model using the VB algorithm and the Gibbs sampler. The modal frequencies and mode shapes can be identified with small uncertainties, while the identified damping ratios generally have a larger variability. It is found that the coefficient of variation of the estimated frequency is approximately equal to the standard deviation of the estimated damping ratio in the same mode.

(2) The EM algorithm, the VB and the Gibbs sampler work in a similar manner, i.e. first inferring the latent variables, then recovering the unknown parameters. The proposed robust implementation strategies based on the square-root filtering and Cholesky decomposition overcomes the accumulated truncation errors, so that they can be reliably used in the OMA.

(3) The three approaches provide consistent point estimates of modal parameters. Though the EM algorithm gives a slightly better estimate, it suffers from the over-fitting problem. The VB tends to underestimate the variability in the estimated modal parameters if the noise level is too high (in the case the mass-spring system), but this shortcoming is not so evident in the practical applications, i.e. the laboratory model and the real building. The Gibbs sampler can work as a baseline to check the performance of other algorithms.

(4) It should be mentioned that the computation burden of all three approaches may be too high for real projects if there are many sensors (e.g. more than 100) and a large number of observable modes (e.g. more than 30). This shortcoming can be alleviated by down-sampling of the raw data and using a short time-period of the data set.

Appendix C

In this appendix we show how the terms in the Taylor series expansion in Eqn. (3.68) can be efficiently computed. The proposed formulation is a modification of the algorithm originally proposed by Vlach and Singhal [138].

For generality, we consider the partial derivatives of the eigenvalue λ and the eigenvector ψ with respect to the parameter θ in the generalized eigenvalue problem

$$\boldsymbol{A}(\theta)\boldsymbol{\psi} = \lambda \boldsymbol{B}(\theta)\boldsymbol{\psi} \tag{3.106}$$

where $\boldsymbol{A}(\theta) \in \mathbb{R}^{n\times n}$ and $\boldsymbol{B}(\theta) \in \mathbb{R}^{n\times n}$. For simplicity of the notation, we do not show the argument θ of $\boldsymbol{A}$ and $\boldsymbol{B}$ in the following derivations. If we set $\boldsymbol{B} = \boldsymbol{I}_n$, Eqn. (3.106) automatically degenerates into the standard eigenvalue problem.

We rewrite Eqn. (3.106) into an equivalent form

$$\boldsymbol{H}\boldsymbol{\psi} = (\boldsymbol{A} - \lambda \boldsymbol{B})\boldsymbol{\psi} = \boldsymbol{0} \tag{3.107}$$

where $\boldsymbol{H} = \boldsymbol{A} - \lambda \boldsymbol{B}$. Consider the LU decomposition of $\boldsymbol{H}$,

$$\boldsymbol{P}\boldsymbol{H} = \boldsymbol{L}\boldsymbol{U} \tag{3.108}$$

where $\boldsymbol{P}$ is a permutation matrix, $\boldsymbol{L}$ is a lower-triangular matrix with unit entries along the diagonal, and $\boldsymbol{U}$ is an upper triangular matrix. Note that $\boldsymbol{L}$ and $\boldsymbol{U}$ are both complex if λ is a complex eigenvalue. Using the property of the determinant of a matrix product,

$$\det(\boldsymbol{H}) = \det(\boldsymbol{L})\det(\boldsymbol{U}) = \det(\boldsymbol{U}) \tag{3.109}$$

where we have used the fact that $\det(\boldsymbol{P}) = 1$ and $\det(\boldsymbol{L}) = 1$. If λ satisfies Eqn. (3.106), $\det(\boldsymbol{H}) = 0$, and thus $\det(\boldsymbol{U}) = 0$, implying that matrix $\boldsymbol{U}$ must have a zero element along its diagonal. Assuming that this zero element is in the last column, i.e. $u_{nn} = 0$, we obtain the following relation

$$\frac{\mathrm{d}u_{nn}}{\mathrm{d}\theta} = \frac{\partial u_{nn}}{\partial \lambda}\frac{\partial \lambda}{\partial \theta} + \frac{\partial u_{nn}}{\partial \theta} = 0 \tag{3.110}$$

To find the terms $\partial u_{nn}/\partial \lambda$ and $\partial u_{nn}/\partial \theta$, we differentiate Eqn. (3.108) with respect to an arbitrary parameter h

$$\boldsymbol{P}\frac{\partial \boldsymbol{H}}{\partial h} = \frac{\partial \boldsymbol{L}}{\partial h}\boldsymbol{U} + \boldsymbol{L}\frac{\partial \boldsymbol{U}}{\partial h} \tag{3.111}$$

We define two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ as follows:

$$\boldsymbol{L}^{*}\boldsymbol{x} = \boldsymbol{e}_n \tag{3.112}$$

$$\boldsymbol{U}\boldsymbol{y} = \boldsymbol{0} \tag{3.113}$$

where $\boldsymbol{e}_n$ is the n-th column of the identity matrix and $*$ means the conjugate transpose. Since $\boldsymbol{y}$ can be arbitrarily normalized, we force its last element to be one, i.e. $y_n = 1$.

Pre- and post-multiplying Eqn. (3.111) by $\boldsymbol{x}^{*}$ and $\boldsymbol{y}$ yields

$$\boldsymbol{x}^*\boldsymbol{P}\frac{\partial \boldsymbol{H}}{\partial h}\boldsymbol{y} = \boldsymbol{x}^*\frac{\partial \boldsymbol{L}}{\partial h}\boldsymbol{U}\boldsymbol{y} + \boldsymbol{x}^*\boldsymbol{L}\frac{\partial \boldsymbol{U}}{\partial h}\,\boldsymbol{y} = \frac{\partial u_{nn}}{\partial h} \tag{3.114}$$

By setting h equal to λ and θ, we can obtain $\partial u_{nn}/\,\partial\lambda$ and $\partial u_{nn}/\,\partial\theta$, and $\partial\lambda/\,\partial\theta$ as a consequence.

Next, consider the partial derivative of the eigenvector. Starting with Eqn. (3.107), we obtain the partial derivatives on the sides of the equation using the chain rule:

$$\left(\frac{\partial \boldsymbol{H}}{\partial\theta} + \frac{\partial \boldsymbol{H}}{\partial\lambda}\frac{\partial\lambda}{\partial\theta}\right)\boldsymbol{\psi} + \boldsymbol{H}\frac{\partial \boldsymbol{\psi}}{\partial\theta} = \boldsymbol{0} \tag{3.115}$$

Rearranging the above equation and inserting Eqn. (3.108) gives

$$\boldsymbol{U}\frac{\partial \boldsymbol{\psi}}{\partial\theta} = -\boldsymbol{L}^{-1}\boldsymbol{P}\left(\frac{\partial \boldsymbol{H}}{\partial\theta} + \frac{\partial \boldsymbol{H}}{\partial\lambda}\frac{\partial\lambda}{\partial\theta}\right)\boldsymbol{\psi} \tag{3.116}$$

Eqn. (3.116) is a linear equation with all terms on the right-hand side known from the previous steps. However, matrix $\boldsymbol{U}$ is singular, so we must enforce one additional constraint. Since eigenvectors are only determined within a constant factor, they are often normalized such that the magnitude of their linear transformation is equal to one, i.e.

$$(\boldsymbol{T}\boldsymbol{\psi})^*(\boldsymbol{T}\boldsymbol{\psi}) = 1 \tag{3.117}$$

Differentiating this constraint gives

$$(\boldsymbol{T}\boldsymbol{\psi})^*\boldsymbol{T}\frac{\partial \boldsymbol{\psi}}{\partial\theta} = 0 \tag{3.118}$$

Combining Eqns. (3.116) and (3.118) yields

$$\begin{bmatrix}\boldsymbol{U}_{1:n-1}\\ (\boldsymbol{T}\boldsymbol{\psi})^*\boldsymbol{T}\end{bmatrix}\frac{\partial \boldsymbol{\psi}}{\partial\theta} = -\boldsymbol{L}^{-1}\boldsymbol{P}\left(\frac{\partial \boldsymbol{H}}{\partial\theta} + \frac{\partial \boldsymbol{H}}{\partial\lambda}\frac{\partial\lambda}{\partial\theta}\right)\boldsymbol{\psi} \tag{3.119}$$

where $\boldsymbol{U}_{1:n-1}$ means the first $n-1$ rows of matrix $\boldsymbol{U}$. Now, the matrix on the left-hand side becomes square and full-rank, so that $\partial\boldsymbol{\psi}/\,\partial\theta$ can be uniquely determined.

In the following we consider the special cases which are used in the calculation of the Jacobian matrices in Eqns. (3.69)-(3.72).

Case 1: $\partial\lambda/\,\partial a_{ij}$ and $\partial\boldsymbol{\psi}/\,\partial a_{ij}$.

We have $\boldsymbol{H} = \boldsymbol{A} - \lambda\boldsymbol{I}_{N_s}$, then $\partial\boldsymbol{H}/\,\partial\lambda = -\boldsymbol{I}_{N_s}$ and $\partial\boldsymbol{H}/\,\partial a_{ij} = \boldsymbol{e}_i\boldsymbol{e}_j^{\mathrm{T}}$.

$$\frac{\partial u_{nn}}{\partial\lambda} = -\boldsymbol{x}^*\boldsymbol{P}\boldsymbol{y} \tag{3.120}$$

$$\frac{\partial u_{nn}}{\partial a_{ij}} = \boldsymbol{x}^*\boldsymbol{P}\boldsymbol{e}_i\boldsymbol{e}_j^{\mathrm{T}}\boldsymbol{y} = (\boldsymbol{x}^*\boldsymbol{P})_i y_j \tag{3.121}$$

According to Eqn. (3.110), we obtain

$$\frac{\partial \lambda}{\partial a_{ij}} = \frac{(\boldsymbol{x}^* \boldsymbol{P})_i y_j}{\boldsymbol{x}^* \boldsymbol{P} \boldsymbol{y}} \tag{3.122}$$

Or we can write it into a matrix form as

$$\frac{\partial \lambda}{\partial \mathrm{vec}(\boldsymbol{A})} = \frac{\boldsymbol{y}^{\mathrm{T}} \otimes (\boldsymbol{x}^* \boldsymbol{P})}{\boldsymbol{x}^* \boldsymbol{P} \boldsymbol{y}} \tag{3.123}$$

Setting $\boldsymbol{T} = \boldsymbol{I}_{N_s}$ and substituting known terms into Eqn. (3.119) gives

$$\begin{bmatrix} \boldsymbol{U}_{1:n-1,:} \\ \boldsymbol{\psi}^* \end{bmatrix} \frac{\partial \boldsymbol{\psi}}{\partial a_{ij}} = \boldsymbol{L}^{-1} \boldsymbol{P} \left(\frac{\partial \lambda}{\partial a_{ij}} \boldsymbol{I}_{N_s} - \boldsymbol{e}_i \boldsymbol{e}_j^{\mathrm{T}} \right) \boldsymbol{\psi} \tag{3.124}$$

Or we can write it into a matrix form as

$$\begin{bmatrix} \boldsymbol{U}_{1:n-1} \\ \boldsymbol{\psi}^* \end{bmatrix} \frac{\partial \boldsymbol{\psi}}{\partial \mathrm{vec}(\mathbf{A})} = \boldsymbol{L}^{-1} \boldsymbol{P} \left(\frac{\partial \lambda}{\partial \mathrm{vec}(\boldsymbol{A})} \otimes \boldsymbol{\psi} - \boldsymbol{\psi}^{\mathrm{T}} \otimes \boldsymbol{I}_{N_s} \right) \tag{3.125}$$

Case 2: $\partial \lambda_{min} / \partial a_{mn}$, $\partial \mathbf{z} / \partial a_{mn}$ and $\partial \lambda_{min} / \partial c_{mn}$, $\partial \mathbf{z} / \partial c_{mn}$.
First, we have

$$\boldsymbol{H} = \boldsymbol{P}_1 \boldsymbol{\varphi}_i^{\mathrm{T}} \boldsymbol{C}^{\mathrm{T}} \boldsymbol{C} \boldsymbol{\varphi}_i \boldsymbol{P}_1 - \lambda_{min} \boldsymbol{P}_2 \boldsymbol{\varphi}_i^{\mathrm{T}} \boldsymbol{C}^{\mathrm{T}} \boldsymbol{C} \boldsymbol{\varphi}_i \boldsymbol{P}_2 \tag{3.126}$$

then

$$\begin{aligned} \frac{\partial \boldsymbol{H}}{\partial a_{mn}} = {} & \boldsymbol{P}_1 \frac{\partial \boldsymbol{\varphi}_i}{\partial a_{mn}}^{\mathrm{T}} \boldsymbol{C}^{\mathrm{T}} \boldsymbol{C} \boldsymbol{\varphi}_i \boldsymbol{P}_1 + \boldsymbol{P}_1 \boldsymbol{\varphi}_i^{\mathrm{T}} \boldsymbol{C}^{\mathrm{T}} \boldsymbol{C} \frac{\partial \boldsymbol{\varphi}_i}{\partial a_{mn}} \boldsymbol{P}_1 \\ & - \lambda_{min} \left(\boldsymbol{P}_2 \frac{\partial \boldsymbol{\varphi}_i}{\partial a_{mn}}^{\mathrm{T}} \boldsymbol{C}^{\mathrm{T}} \boldsymbol{C} \boldsymbol{\varphi}_i \boldsymbol{P}_2 + \boldsymbol{P}_2 \boldsymbol{\varphi}_i^{\mathrm{T}} \boldsymbol{C}^{\mathrm{T}} \boldsymbol{C} \frac{\partial \boldsymbol{\varphi}_i}{\partial a_{mn}} \boldsymbol{P}_2 \right) \end{aligned} \tag{3.127}$$

$$\frac{\partial \boldsymbol{H}}{\partial \lambda_{min}} = - \boldsymbol{P}_2 \boldsymbol{\varphi}_i^{\mathrm{T}} \boldsymbol{C}^{\mathrm{T}} \boldsymbol{C} \boldsymbol{\varphi}_i \boldsymbol{P}_2 \tag{3.128}$$

Thus, we can obtain

$$\frac{\partial u_{22}}{\partial \lambda_{min}} = -\boldsymbol{x}^* \boldsymbol{P} \, \boldsymbol{P}_2 \boldsymbol{\varphi}_i^{\mathrm{T}} \boldsymbol{C}^{\mathrm{T}} \boldsymbol{C} \boldsymbol{\varphi}_i \boldsymbol{P}_2 \boldsymbol{y} \tag{3.129}$$

$$\frac{\partial u_{22}}{\partial a_{mn}} = \boldsymbol{x}^* \boldsymbol{P} \frac{\partial \boldsymbol{H}}{\partial a_{mn}} \boldsymbol{y} \tag{3.130}$$

$$\frac{\partial \lambda_{min}}{\partial a_{mn}} = \boldsymbol{x}^* \boldsymbol{P} \frac{\partial \boldsymbol{H}}{\partial a_{mn}} \boldsymbol{y} / (\boldsymbol{x}^* \boldsymbol{P} \, \boldsymbol{P}_2 \boldsymbol{\varphi}_i^{\mathrm{T}} \boldsymbol{C}^{\mathrm{T}} \boldsymbol{C} \boldsymbol{\varphi}_i \boldsymbol{P}_2 \boldsymbol{y}) \tag{3.131}$$

For the sensitivity of eigenvector,

$$\begin{bmatrix} \boldsymbol{U}_1 \\ \boldsymbol{z}^{\mathrm{T}}\,\boldsymbol{P}_2\boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{\varphi}_i\boldsymbol{P}_2 \end{bmatrix}\frac{\partial \boldsymbol{z}}{\partial a_{mn}} = \boldsymbol{L}^{-1}\boldsymbol{P}\left(\frac{\partial \lambda_{min}}{\partial a_{mn}}\boldsymbol{P}_2\boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{\varphi}_i\boldsymbol{P}_2 - \frac{\partial \boldsymbol{H}}{\partial a_{mn}}\right)\boldsymbol{z} \tag{3.132}$$

Similarly,

$$\begin{aligned}\frac{\partial \boldsymbol{H}}{\partial c_{mn}} = {} & \boldsymbol{P}_1\boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{e}_n\boldsymbol{e}_m^{\mathrm{T}}\boldsymbol{C}\boldsymbol{\varphi}_i\boldsymbol{P}_1 + \boldsymbol{P}_1\boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{e}_m\boldsymbol{e}_n^{\mathrm{T}}\boldsymbol{\varphi}_i\boldsymbol{P}_1 \\ & - \lambda_{min}\left(\boldsymbol{P}_2\boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{e}_n\boldsymbol{e}_m^{\mathrm{T}}\boldsymbol{C}\boldsymbol{\varphi}_i\boldsymbol{P}_2 + \boldsymbol{P}_2\boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{e}_m\boldsymbol{e}_n^{\mathrm{T}}\boldsymbol{\varphi}_i\boldsymbol{P}_2\right)\end{aligned} \tag{3.133}$$

Therefore,

$$\frac{\partial u_{22}}{\partial a_{mn}} = \boldsymbol{x}^*\boldsymbol{P}\frac{\partial \boldsymbol{H}}{\partial c_{mn}}\boldsymbol{y} \tag{3.134}$$

$$\frac{\partial \lambda_{min}}{\partial c_{mn}} = \boldsymbol{x}^*\boldsymbol{P}\frac{\partial \boldsymbol{H}}{\partial c_{mn}}\boldsymbol{y}/(\boldsymbol{x}^*\boldsymbol{P}\,\boldsymbol{P}_2\boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{\varphi}_i\boldsymbol{P}_2\boldsymbol{y} \tag{3.135}$$

For the sensitivity of eigenvector,

$$\begin{bmatrix} \boldsymbol{U}_1 \\ \boldsymbol{z}^{\mathrm{T}}\,\boldsymbol{P}_2\boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{\varphi}_i\boldsymbol{P}_2 \end{bmatrix}\frac{\partial \boldsymbol{z}}{\partial c_{mn}} = \boldsymbol{L}^{-1}\boldsymbol{P}\left(\frac{\partial \lambda_{min}}{\partial c_{mn}}\boldsymbol{P}_2\boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{\varphi}_i\boldsymbol{P}_2 - \frac{\partial \boldsymbol{H}}{\partial c_{mn}}\right)\boldsymbol{z} \tag{3.136}$$

Chapter 4
Finite element model updating

4.1 Introduction

One important application of OMA is to update the structural finite element (FE) model, i.e. to estimate structural parameters (e.g. the stiffness, mass or damping) from the measured modal parameters such that the identified FE model reproduces, as close as possible, the measured modal data. When such an analysis is performed in the aftermath a severe load effect, e.g., an earthquake ground motion, a major loss in the stiffness may be an indication of damage occurred in the structure. Due to the presence of measurement and model errors, uncertainties in identified model parameters are inevitable. For the purpose of reliable damage diagnosis and prognosis, it is important and indispensable to explicitly treat the uncertainties in the finite element model updating (FEMU). In this chapter, a Bayesian methodology is presented for FE model updating using noisy incomplete modal data, assuming the measured structural behavior is linear and classically damped.

The Bayesian method for FE model updating (FEMU) is not new, e.g. see references [74-80], but many previous researches did not formulate the problem correctly. Hence, in Section 4.2, we reformulate the FEMU problem first from the physical model, based on the generalized eigenvalue decomposition, to the probabilistic model by assuming the distribution of errors. The shortcomings in previous formulations of the problem are described in Section 4.2.2. Because of the complexity of the formulated FEMU problem, an analytical posterior distribution of the model parameters is not feasible, so that a Metropolis-within-Gibbs (MwG) sampler is designed to approximate the posterior in Section 4.3, and a robust implementation strategy is also provided. In Section 4.4, the performance of the proposed Bayesian method is illustrated through two examples: a numerical 8-DoF mass-spring system and an experimental 6-story shear-type building.

4.2 Problem formulation

The problem of finite element model updating (FEMU) considered here is to find a good finite element (FE) model to numerically represent the true structure, given the

measured modal frequencies $\omega_{kt} \in \mathbb{R}^+$ and mode shapes $\boldsymbol{\phi}_{kt}^m \in \mathbb{R}^{N_o}$, for $k = 1,2,\dots,N_m$ and $t = 1,2,\dots,T$, in which N_o is the number of measured DoFs over the structure, N_m is the number of measured modes, and T is total number of available data sets. The superscript m on $\boldsymbol{\phi}_{kt}^m$ indicates that only the measured degrees of freedom are represented in this mode shape vector. The basic assumptions on the structural model are that it is linear and classically damped, but it can be time-variant. Since classical damping is presumed, only modal frequencies and mode shapes are necessary in order to identify the mass matrix $\boldsymbol{M} \in \mathbb{R}^{N_d \times N_d}$ and stiffness matrix $\boldsymbol{K} \in \mathbb{R}^{N_d \times N_d}$, where N_d is the number of DoFs of the FE model.

Due to environmental effects and operational conditions, the structural parameters, e.g. stiffness, mass and damping values, as well as boundary and continuity conditions may change over time [139,140]. In that case, it is necessary to explicitly model this time dependence for the sake of accuracy. In fact, since in the Bayesian data analysis a true fixed value is assumed for each parameter [83], the posterior of parameters will vastly underestimate the inherent uncertainty with more and more data collected if this time dependence behavior is ignored. In particular, if we have the samples from the distribution of modal parameters (like the Gibbs sampling used in Chapter 2), the associated uncertainties can be propagated into the identified model parameters by incorporating the time variance.

In this dissertation, we assume that we know the mass matrix $\boldsymbol{M} \in \mathbb{R}^{N_d \times N_d}$, because it usually can be established with sufficient accuracy from the engineering drawings of the structure. The stiffness matrix $\boldsymbol{K} \in \mathbb{R}^{N_d \times N_d}$ is modeled in an affine manner, i.e. as the weighted sum of a nominal stiffness matrix $\boldsymbol{K}_0 \in \mathbb{R}^{N_d \times N_d}$ and a series of substructure stiffness matrices $\boldsymbol{K}_i \in \mathbb{R}^{N_d \times N_d}$ for $i = 1,2,\dots,N_k$

$$\boldsymbol{K}(\boldsymbol{\kappa}_t) = \boldsymbol{K}_0 + \sum_{i=1}^{N_k} \kappa_{it} \boldsymbol{K}_i \tag{4.1}$$

in which $\boldsymbol{\kappa}_t = \left[\kappa_{1t}, \kappa_{2t}, \dots, \kappa_{N_k t}\right]^{\mathrm{T}}$ is a set of dimensionless parameters describing possible modifications of the nominal stiffness in time. Note that under the assumption of linear structural behavior, we can always write the stiffness matrix into the form of Eqn. (4.1) by modeling $\boldsymbol{K}_i$ as a subset of element stiffness matrices. This representation of the stiffness matrix has been used in previous works, see e.g., [74-78].

4.2.1 The physical model

The relation between the unknown parameters $\boldsymbol{\kappa}_t$ and the measured modal parameters ω_{kt} and $\boldsymbol{\phi}_{kt}^m$ is given by the following two equations

$$\det[\boldsymbol{K}(\boldsymbol{\kappa}_t) - \lambda_{kt}\boldsymbol{M}] = 0 \tag{4.2}$$

$$[\boldsymbol{K}(\boldsymbol{\kappa}_t) - \lambda_{kt}\boldsymbol{M}]\boldsymbol{\phi}_{kt} = \boldsymbol{0} \tag{4.3}$$

where $\det[\cdot]$ is the matrix determinant, $\lambda_{kt} = \omega_{kt}^2$ and $\boldsymbol{\phi}_{kt} = \boldsymbol{P}_\phi\left[\boldsymbol{\phi}_{kt}^{m\,\mathrm{T}}\ \boldsymbol{\phi}_{kt}^{u\,\mathrm{T}}\right]^{\mathrm{T}}$, where $\boldsymbol{P}_\phi$ is a known permutation matrix to arrange the elements into appropriate positions, and $\boldsymbol{\phi}_{kt}^u \in \mathbb{R}^{N_d - N_o}$ denotes the unmeasured mode shape components, working as a latent variable in FEMU. Note that the above two equations are the most fundamental ones. Other equations, such as the orthogonality condition of mode shapes, can be derived as a consequence of these equations. Equations (4.2) and (4.3) are not redundant; in particular, note that Eq. (4.2) can be used to estimate the eigenvalue λ_{kt} without knowing the eigenvector $\boldsymbol{\phi}_{kt}$, but Eqn. (4.3) is needed to estimate the eigenvectors.

One drawback of directly using Eqn. (4.2) is that there is usually a large numerical error in computing the matrix determinant, i.e. even if λ_{kt} is an eigenvalue of $[\boldsymbol{K}(\boldsymbol{\kappa}_t) \quad \boldsymbol{M}]$, the computed determinant may still be far away from zero if the eigenvalues of $\boldsymbol{K}(\boldsymbol{\kappa}_t) - \lambda_{kt}\boldsymbol{M}$ are large, which is often the case for civil structures. To overcome this difficulty, we choose to introduce a normalization procedure by use of two scalar factors n_M and λ_0. The first factor, n_M, is the Frobenius norm of the mass matrix $\boldsymbol{M}$ and λ_0 is the median of eigenvalues of $\boldsymbol{K}_0/n_M$ [141]. Thus, we transform Eqns. (4.2) and (4.3) into

$$\det[\boldsymbol{K}'(\boldsymbol{\kappa}_t) - \lambda'_{kt}\boldsymbol{M}'] = 0 \tag{4.4}$$

$$[\boldsymbol{K}'(\boldsymbol{\kappa}_t) - \lambda'_{kt}\boldsymbol{M}']\boldsymbol{\phi}_{kt} = \boldsymbol{0} \tag{4.5}$$

in which $\boldsymbol{K}'(\boldsymbol{\kappa}_t) = \boldsymbol{K}(\boldsymbol{\kappa}_t)/(n_M\lambda_0)$, $\boldsymbol{M}' = \boldsymbol{M}/n_M$ and $\lambda'_{kt} = \lambda_{kt}/\lambda_0$. This normalization does not change $\boldsymbol{\phi}_{kt}$ as an eigenvector and it confines all the eigenvalues of $\boldsymbol{K}'(\boldsymbol{\kappa}_t) - \lambda'_{kt}\boldsymbol{M}'$ to be around 1, except the near-zero eigenvalues. Based on this normalization, Eqn. (4.1) is rewritten as

$$\boldsymbol{K}'(\boldsymbol{\kappa}_t) = \boldsymbol{K}_0/(n_M\lambda_0) + \sum_{i=1}^{N_k} \kappa_{it}\,\boldsymbol{K}_i/(n_M\lambda_0) = \boldsymbol{K}'_0 + \sum_{i=1}^{N_k} \kappa_{it}\,\boldsymbol{K}'_i \tag{4.6}$$

Where $\boldsymbol{K}'_0 = \boldsymbol{K}_0/(n_M\lambda_0)$ and $\boldsymbol{K'}_i = \boldsymbol{K}_i/(n_M\lambda_0)$.

Equations (4.4) and (4.5) provide the deterministic relation between the unknown parameters $\boldsymbol{\kappa}_t$ and the measured modal parameters ω_{kt} and $\boldsymbol{\phi}_{kt}^m$. However, due to the existence of uncertainties, these equations cannot be strictly satisfied. On one hand, $\boldsymbol{K}(\boldsymbol{\kappa}_t)$ is only an idealization of the complex structure with simplifications, so that there are always model errors in $\boldsymbol{K}(\boldsymbol{\kappa}_t)$. On the other hand, ω_{kt} and $\boldsymbol{\phi}_{kt}^m$ cannot be measured exactly as measurement errors are inevitable. Considering these errors, we have the following stochastic equations:

$$\det[\boldsymbol{K}'(\boldsymbol{\kappa}_t) - \lambda'_{kt}\boldsymbol{M}'] = w_{kt} \tag{4.7}$$

$$[\boldsymbol{K}'(\boldsymbol{\kappa}_t) - \lambda'_{kt}\boldsymbol{M}']\boldsymbol{\phi}_{kt} = \boldsymbol{v}_{kt} \tag{4.8}$$

where w_{kt} and $\boldsymbol{v}_{kt}$ represent the combined effects of the model error and the measurement error.

We have explicitly modeled the true structure by a time-variant FEM because of the changing environmental effects and operational conditions. Assuming the existence of a fixed model for a reference environmental condition, we have the following linear constraint equations:

$$\boldsymbol{\kappa}_t = \boldsymbol{\kappa} + \boldsymbol{\varepsilon}_t \tag{4.9}$$

in which $\boldsymbol{\kappa}$ is a deterministic unknown set of parameters of the true reference model and $\boldsymbol{\varepsilon}_t$ models the time-varying environmental and other effects, which are not explicitly measured.

4.2.2 The probabilistic model

The difficulty in constructing the probabilistic model for FEMU lies in the formulation of the conditional probability distributions of λ'_{kt} and $\boldsymbol{\phi}_{kt}$, given $\boldsymbol{\kappa}_t$, i.e. the likelihood function. In this aspect, many previous Bayesian methods [76,78-80] for FEMU are imperfect. In Ref. [76], an equation similar to Eqn. (4.8) is used to construct the conditional distribution of the modal frequencies, but it is not clear how this distribution is obtained by assuming an error distribution. The likelihood used in Ref. [78] fails to connect measurements to unknown parameters, which only appears in the prior. Ref. [79] applies the same equation as Ref. [76] to model the conditional distribution of the modal frequencies, and it is easy to see that the normalization factor in the formulated likelihood is incorrect. Similarly, the likelihood formulated in Ref. [80] is not appropriate because the normalization factor should include the unknown parameters as well. In this section, we propose a correct probabilistic model for FEMU based on the physical model introduced in the previous section by assigning probability distributions to the noise terms w_{kt}, $\boldsymbol{v}_{kt}$ and $\boldsymbol{\varepsilon}_t$ and by specifying prior distributions for the unknown parameters $\boldsymbol{\kappa}_t$.

Following the principle of maximum entropy [108], we assume that w_{kt}, $\boldsymbol{v}_{kt}$ and $\boldsymbol{\varepsilon}_t$ all follow normal distributions and that they are independent of each other. In addition, independence is assumed for w_{kt} and $\boldsymbol{v}_{kt}$ for different modes and different time. Similarly, independence is assumed for values of $\boldsymbol{\varepsilon}_t$ for different t. Specifically, to ensure $\boldsymbol{K}'(\boldsymbol{\kappa}_t)$ is a good representation of the true structure, w_{kt} and $\boldsymbol{v}_{kt}$ are assumed to be zero-mean with the PDFs

$$f(w_{kt}|\alpha_k) = \sqrt{\alpha_k/2\pi}\exp\left\{-\frac{\alpha_k}{2}w_{kt}^2\right\} \tag{4.10}$$

$$f(\boldsymbol{v}_{kt}|\beta_k) = (\beta_k/2\pi)^{N_d/2}\exp\left\{-\frac{\beta_k}{2}\boldsymbol{v}_{kt}^{\mathrm{T}}\boldsymbol{v}_{kt}\right\} \tag{4.11}$$

From Eqn. (4.7), we can see that w_{kt} is a non-injective function of λ'_{kt}, i.e. a many-to-one map from λ'_{kt} to w_{kt}, yielding a multiple-mode distribution of λ'_{kt} given the distribution of w_{kt} following the rules of probability distributions of random variables [14]. In order to derive the conditional distribution of λ'_{kt} for given $\boldsymbol{\kappa}_t$ and α_k, we further assume that the probability masses around the roots of the equation are equal to each other, giving the conditional PDF of λ'_{kt} as

$$\begin{aligned} f(\lambda'_{kt}|\boldsymbol{\kappa}_t,\alpha_k) = {} & N_d^{-1}\sqrt{\alpha_k/2\pi}\,|\mathrm{tr}[\mathrm{adj}(\boldsymbol{K}'(\boldsymbol{\kappa}_t)-\lambda'_{kt}\boldsymbol{M}')\boldsymbol{M}']| \\ & \exp\left\{-\frac{\alpha_k}{2}|\boldsymbol{K}'(\boldsymbol{\kappa}_t)-\lambda'_{kt}\boldsymbol{M}'|^2\right\} \end{aligned} \tag{4.12}$$

where 'adj(·)' denotes the adjugate matrix, which is obtained using the Jacobi's formula [142] to calculate the derivative of the determinant. Although the PDF in Eqn. (4.12) gives a nonzero probability for nonphysical negative frequency values, the amount of probability volume for less than zero values of λ'_{kt} is generally negligible, so truncation of the PDF for negative values followed by a renormalization is not necessary.

Since the above PDF is not standard, it is interesting to see its shape through a simple example. Suppose we have a model with the following parameters:

$$\boldsymbol{K} = \boldsymbol{K}_0 + \kappa\boldsymbol{K}_1 = \begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix} + \kappa\begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix},\ \boldsymbol{M} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \tag{4.13}$$

For $\kappa = 0$ and $\alpha_k = 200$, the PDF in Eqn. (4.12) is plotted in Figure 4.1. It can be seen that the distribution has two modes locating near the eigenvalues of the deterministic model. The two modes share equal probability masses, validating the preceding assumption on the equal probability mass.

Based on the rules of probability distribution of functions of random variables [14], the conditional PDF of $\boldsymbol{\phi}_{kt}$ given $\boldsymbol{\kappa}_t$, λ'_{kt} and β_k is

$$\begin{aligned} f(\boldsymbol{\phi}_{kt}|\boldsymbol{\kappa}_t,\lambda'_{kt},\beta_k) = {} & (\beta_k/2\pi)^{N_d/2}|\boldsymbol{K}'(\boldsymbol{\kappa}_t)-\lambda'_{kt}\boldsymbol{M}'| \\ & \exp\left\{-\frac{\beta_k}{2}\boldsymbol{\phi}_{kt}^{\mathrm{T}}[\boldsymbol{K}'(\boldsymbol{\kappa}_t)-\lambda'_{kt}\boldsymbol{M}']^{\mathrm{T}}[\boldsymbol{K}'(\boldsymbol{\kappa}_t)-\lambda'_{kt}\boldsymbol{M}']\,\boldsymbol{\phi}_{kt}\right\} \end{aligned} \tag{4.14}$$

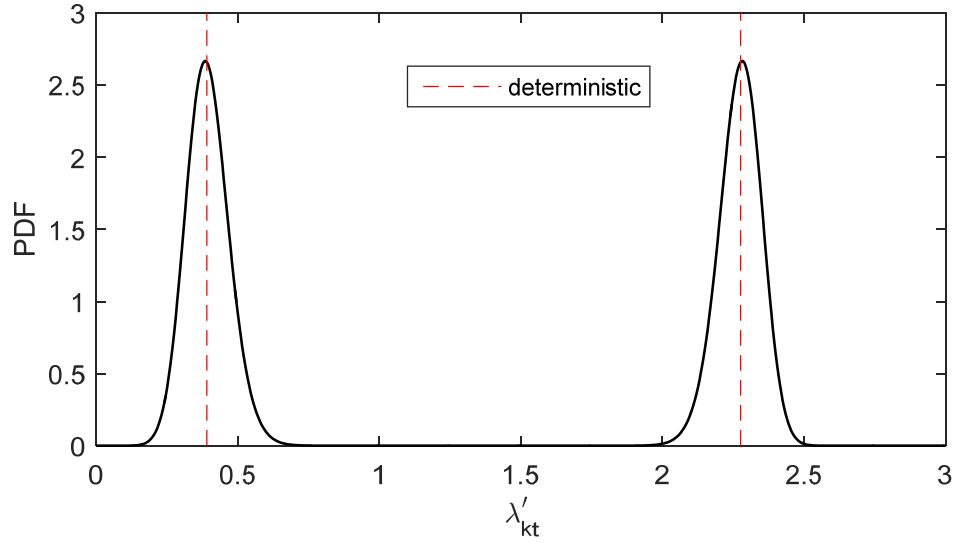


Figure 4.1 PDF of eigenvalues.

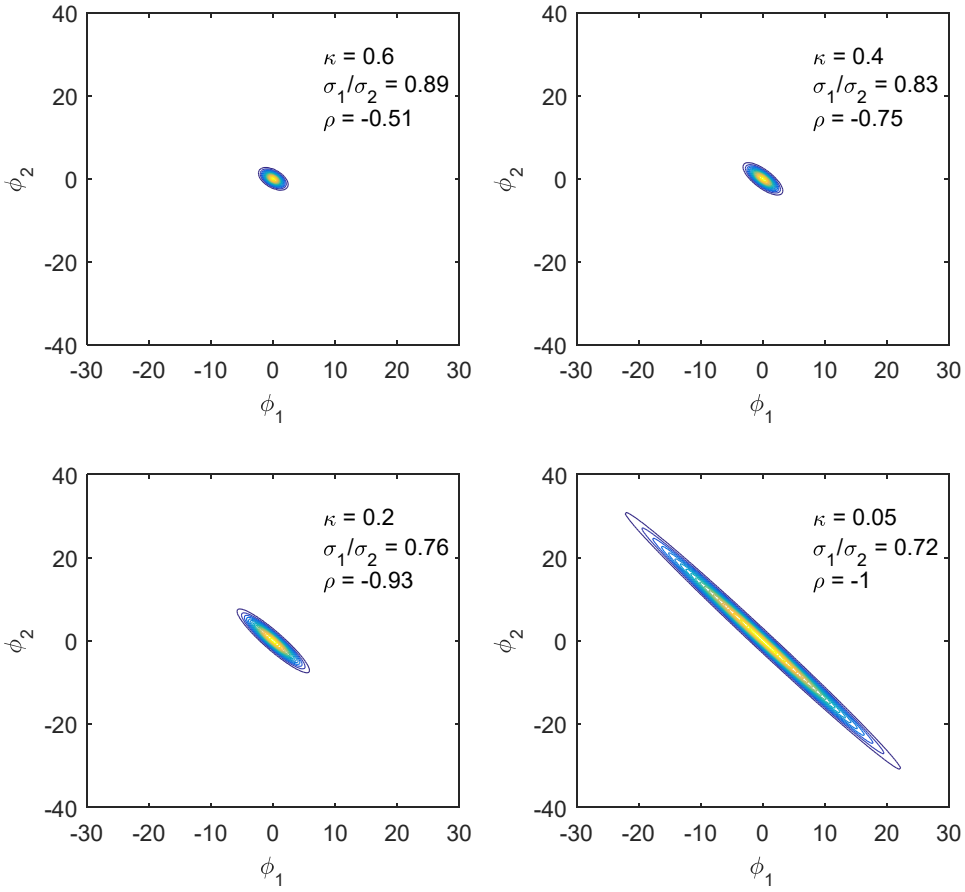


Figure 4.2 Contour plots of the PDF of 2nd mode shape.

This is the multivariate normal distribution with zero means. Since the mode shapes can be arbitrarily normalized, it is easy to verify that the ratio of the standard deviations of two components of a mode is equal to that the ratio of the magnitudes of the components. For the simple model in Eqn. (4.13), the contour plots of the PDF of the 2nd mode shape for selected κ values are illustrated in Figure 4.2, where we have set $\lambda_2' = 2.28$ and $\beta_k = 1$. As the value of κ approaches the true value $\kappa = 0$, the ratio σ_1/σ_2 tends to the nominal value 0.71, and the correlation between mode components ϕ_1 and ϕ_2 becomes higher with the limit of linear correlation $\rho = 1$ when $\kappa = 0$. In this limit condition, the PDF degenerates into a 2-D Dirac delta function.

Similarly, by assuming $\boldsymbol{\varepsilon}_t$ to be a zero-mean multivariate normal random vector with a precision matrix $\boldsymbol{P}$, according to Eqn. (4.9), we have the conditional distribution of $\boldsymbol{\kappa}_t$ given $\boldsymbol{\kappa}$ and $\boldsymbol{P}$ as

$$f(\boldsymbol{\kappa}_t|\boldsymbol{\kappa},\boldsymbol{P}) = \left|\frac{\boldsymbol{P}}{2\pi}\right|^{1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\kappa}_t-\boldsymbol{\kappa})^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{\kappa}_t-\boldsymbol{\kappa})\right\} \tag{4.15}$$

As for the prior distributions for the parameters $\boldsymbol{\theta} = \{\boldsymbol{\kappa},\boldsymbol{P},\boldsymbol{\alpha},\boldsymbol{\beta}\}$, the following conjugate prior distributions are assigned:

$$\pi(\boldsymbol{P}) = \text{Wishart}(\boldsymbol{E}_0, e_0) \propto |\boldsymbol{P}|^{(e_0-N_k-1)/2} \exp[-\text{tr}(\boldsymbol{E}_0^{-1}\boldsymbol{P})/2] \tag{4.16}$$

$$\begin{aligned}\pi(\boldsymbol{\kappa}|\boldsymbol{P}) &= \text{N}(\boldsymbol{\kappa}_0, (p_0\boldsymbol{P})^{-1}) \\ &\propto p_0^{N_k/2}|\boldsymbol{P}|^{1/2} \exp[-p_0(\boldsymbol{\kappa}-\boldsymbol{\kappa}_0)^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{\kappa}-\boldsymbol{\kappa}_0)/2]\end{aligned} \tag{4.17}$$

$$\pi(\boldsymbol{\alpha}) = \prod_{k=1}^{N_m} \text{Gam}(u_{k,0}, h_{k,0}) \propto \prod_{k=1}^{N_m} \alpha_k^{u_{k,0}} \exp(-h_{k,0}\alpha_k) \tag{4.18}$$

$$\pi(\boldsymbol{\beta}) = \prod_{k=1}^{N_m} \text{Gam}(r_{k,0}, s_{k,0}) \propto \prod_{k=1}^{N_m} \beta_k^{r_{k,0}} \exp(-s_{k,0}\beta_k) \tag{4.19}$$

Combing Eqns. (4.12) and (4.14)-(4.19) yields the joint distribution of the observed variables $\boldsymbol{y} = \{\lambda'_{(1:N_m)(1:T)}, \boldsymbol{\phi}^m_{(1:N_m)(1:T)}\}$, latent variables $\boldsymbol{x} = \boldsymbol{\phi}^u_{(1:N_m)(1:T)}$ and parameters $\boldsymbol{\theta}$ as

$$\begin{aligned}f(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\theta}) &= \pi(\boldsymbol{P})\pi(\boldsymbol{\kappa}|\boldsymbol{P})\pi(\boldsymbol{\alpha})\pi(\boldsymbol{\beta}) \\ &\prod_{t=1}^{T}\left\{f(\boldsymbol{\kappa}_t|\boldsymbol{\kappa},\boldsymbol{P})\prod_{k=1}^{N_m}[f(\lambda'_{kt}|\boldsymbol{\kappa}_t,\alpha_k)f(\boldsymbol{\phi}_{kt}|\boldsymbol{\kappa}_t,\lambda'_{kt},\beta_k)]\right\}\end{aligned} \tag{4.20}$$

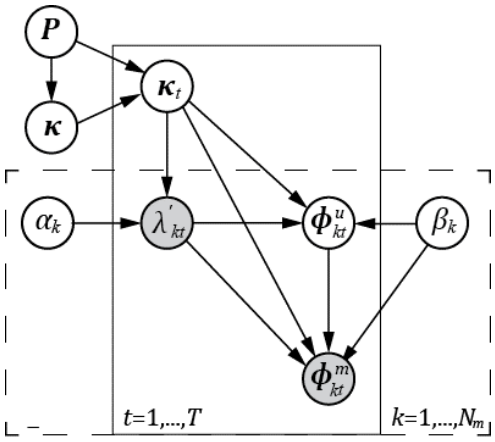


Figure 4.3 Bayesian network for FEMU.

The above joint distribution can be represented by the Bayesian network model illustrated in Figure 4.3, which clearly shows the dependence within the random variables. There are many virtues in constructing this Bayesian model. First, it explicitly models the time variation of the structure, so that the uncertainties caused by environmental effects and operational conditions can be handled. Second, unlike the sensitivity-based methods [65,66], there is no need to match the analytical modes and the measured ones, which is a tedious process in practical applications. Third, the generalized eigenvalue decomposition of the stiffness and mass matrices is not necessary, as it is replaced by directly plugging the eigen-pairs into the eigenvalue equations. However, due to the complexity of the Bayesian model, a closed-form solution does not exist even though many assumptions have been made to simplify the model. As a consequence, we will apply the Metropolis-within-Gibbs sampling approach to approximately solve the problem in the next section.

4.3 Metropolis-within-Gibbs sampler

Because of the complexity of the probabilistic model for the FEMU, it is not possible to obtain the posterior of the unknown parameters in closed form. We can employ the EM algorithm to obtain a MAP solution of the unknown parameters; however, in that case the M step requires a numerical optimization (e.g. the gradient decent method) to find the optimal solution, thus it is nontrivial. Similarly, it is not easy to find a surrogate distribution of the unknown parameters, which restricts the application of the VB algorithm. Therefore, we develop a Metropolis-within-Gibbs (MwG) sampler for the FEMU in this section.

4.3.1 Derivation of the MwG sampler

We start with the logarithm of the joint distribution in Eqn. (4.20)

$$
\begin{aligned}
&\log f(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\theta}) \\
&= \log \pi(\boldsymbol{P}) + \log \pi(\boldsymbol{\kappa}|\boldsymbol{P}) + \log \pi(\boldsymbol{\alpha}) + \log \pi(\boldsymbol{\beta}) + \textstyle\sum_{t=1}^{T} \log f(\boldsymbol{\kappa}_t|\boldsymbol{\kappa}, \boldsymbol{P}) \\
&\quad + \textstyle\sum_{k=1}^{N_m} \sum_{t=1}^{T} \log f(\lambda'_{kt}|\boldsymbol{\kappa}_t, \alpha_k) + \sum_{k=1}^{N_m} \sum_{t=1}^{T} \log f(\boldsymbol{\phi}_{kt}|\boldsymbol{\kappa}_t, \lambda'_{kt}, \beta_k) \\
&\propto \tfrac{e_0 - N_k + T}{2} \log|\boldsymbol{P}| - \tfrac{1}{2}\mathrm{tr}(\boldsymbol{E}_0^{-1}\boldsymbol{P}) + \tfrac{N_k}{2} \log p_0 - \tfrac{p_0}{2}(\boldsymbol{\kappa} - \boldsymbol{\kappa}_0)^{\mathrm{T}} \boldsymbol{P} (\boldsymbol{\kappa} - \boldsymbol{\kappa}_0) \\
&- \tfrac{1}{2} \textstyle\sum_{t=1}^{T} (\boldsymbol{\kappa}_t - \boldsymbol{\kappa})^{\mathrm{T}} \boldsymbol{P} (\boldsymbol{\kappa}_t - \boldsymbol{\kappa}) + \sum_{k=1}^{N_m} \left[\left(u_{k,0} + \tfrac{T}{2} \right) \log \alpha_k - h_{k,0} \alpha_k \right] \\
&+ \textstyle\sum_{k=1}^{N_m} \sum_{t=1}^{T} \left[\log|\mathrm{tr}[\mathrm{adj}(\boldsymbol{K}'(\boldsymbol{\kappa}_t) - \lambda'_{kt}\boldsymbol{M}')\boldsymbol{M}']| - \tfrac{\alpha_k}{2} |\boldsymbol{K}'(\boldsymbol{\kappa}_t) - \lambda'_{kt}\boldsymbol{M}'|^2 \right] \\
&+ \textstyle\sum_{k=1}^{N_m} \left[\left(r_{k,0} + \frac{TN_d}{2} \right) \log \beta_k - s_{k,0} \beta_k \right] + \sum_{k=1}^{N_m} \sum_{t=1}^{T} \log|\boldsymbol{K}'(\boldsymbol{\kappa}_t) - \lambda'_{kt}\boldsymbol{M}'| \\
&- \textstyle\sum_{k=1}^{N_m} \sum_{t=1}^{T} \frac{\beta_k}{2} \boldsymbol{\phi}_{kt}^{\mathrm{T}} [\boldsymbol{K}'(\boldsymbol{\kappa}_t) - \lambda'_{kt}\boldsymbol{M}']^{\mathrm{T}} [\boldsymbol{K}'(\boldsymbol{\kappa}_t) - \lambda'_{kt}\boldsymbol{M}']\, \boldsymbol{\phi}_{kt}
\end{aligned}
\tag{4.21}
$$

The conditional distribution for each of random variable can be obtained by collecting the terms in the joint distribution that include the random variable of interest. Below, we summarize the results.

Conditional distribution of $\boldsymbol{\phi}_{kt}^{u}$

Starting with the latent variable $\boldsymbol{\phi}_{kt}^{u}$, we obtain its conditional posterior by picking up the relevant terms in Eqn. (4.21):

$$
\begin{aligned}
&\log p(\boldsymbol{\phi}_{kt}^{u}|\boldsymbol{\phi}_{kt}^{m}, \lambda'_{kt}, \kappa_t, \beta_k) \\
&\propto -0.5 \boldsymbol{\phi}_{kt}^{u\,\mathrm{T}} \beta_k [\boldsymbol{K}'_u(\boldsymbol{\kappa}_t) - \lambda'_{kt}\boldsymbol{M}'_u]^{\mathrm{T}} [\boldsymbol{K}'_u(\boldsymbol{\kappa}_t) - \lambda'_{kt}\boldsymbol{M}'_u] \boldsymbol{\phi}_{kt}^{u} \\
&\quad + \boldsymbol{\phi}_{kt}^{u\,\mathrm{T}} \beta_k [\boldsymbol{K}'_u(\boldsymbol{\kappa}_t) - \lambda'_{kt}\boldsymbol{M}'_u]^{\mathrm{T}} [\boldsymbol{K}'_m(\boldsymbol{\kappa}_t) - \lambda'_{kt}\boldsymbol{M}'_m] \boldsymbol{\phi}_{kt}^{m} + c_1 \\
&\propto -0.5 \left(\boldsymbol{\phi}_{kt}^{u} - \boldsymbol{\mu}_{\boldsymbol{\phi}_{kt}^{u}} \right)^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{\phi}_{kt}^{u}}^{-1} \left(\boldsymbol{\phi}_{kt}^{u} - \boldsymbol{\mu}_{\boldsymbol{\phi}_{kt}^{u}} \right) \\
&\propto \log \mathrm{N}\left(\boldsymbol{\mu}_{\boldsymbol{\phi}_{kt}^{u}}, \boldsymbol{\Sigma}_{\boldsymbol{\phi}_{kt}^{u}} \right)
\end{aligned}
\tag{4.22}
$$

The above shows that the conditional distribution of $\boldsymbol{\phi}_{kt}^{u}$ follows a multivariate normal distribution with the precision matrix and mean vector as

$$\boldsymbol{\Sigma}_{\boldsymbol{\phi}_{kt}^{u}}^{-1} = \beta_k[\boldsymbol{K}_u'(\boldsymbol{\kappa}_t) - \lambda_{kt}'\boldsymbol{M}_u']^{\mathrm{T}}[\boldsymbol{K}_u'(\boldsymbol{\kappa}_t) - \lambda_{kt}'\boldsymbol{M}_u'] \tag{4.23}$$

$$\boldsymbol{\mu}_{\boldsymbol{\phi}_{kt}^{u}} = \boldsymbol{\Sigma}_{\boldsymbol{\phi}_{kt}^{u}}\beta_k[\boldsymbol{K}_u'(\boldsymbol{\kappa}_t) - \lambda_{kt}'\boldsymbol{M}_u']^{\mathrm{T}}[\lambda_{kt}'\boldsymbol{M}_m' - \boldsymbol{K}_m'(\boldsymbol{\kappa}_t)]\boldsymbol{\phi}_{kt}^{m} \tag{4.24}$$

in which $\boldsymbol{K}_u'(\boldsymbol{\kappa}_t)$ consists of the columns of the stiffness matrix corresponding to unmeasured DoFs, $\boldsymbol{K}_m'(\boldsymbol{\kappa}_t)$ consists of the columns of the stiffness matrix corresponding to the measured DoFs, and $\boldsymbol{M}_u'$ and $\boldsymbol{M}_m'$ are the corresponding columns of the ormalized mass matrix, respectively.

Conditional distribution of $\boldsymbol{\kappa}_t$

Collecting terms in Eqn. (4.21) involving $\boldsymbol{\kappa}_t$ gives

$$\begin{aligned}
&\log p(\boldsymbol{\kappa}_t|\boldsymbol{\phi}_{kt}, \lambda_{kt}', \boldsymbol{\kappa}, \boldsymbol{P}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \\
&\propto -0.5(\boldsymbol{\kappa}_t - \boldsymbol{\kappa})^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{\kappa}_t - \boldsymbol{\kappa}) + \textstyle\sum_{k=1}^{N_m}\log|\boldsymbol{K}'(\boldsymbol{\kappa}_t) - \lambda_{kt}'\boldsymbol{M}'| \\
&+ \textstyle\sum_{k=1}^{N_m}[\log|\mathrm{tr}[\mathrm{adj}(\boldsymbol{K}'(\boldsymbol{\kappa}_t) - \lambda_{kt}'\boldsymbol{M}')\boldsymbol{M}']| - 0.5\alpha_k|\boldsymbol{K}'(\boldsymbol{\kappa}_t) - \lambda_{kt}'\boldsymbol{M}'|^2] \\
&- \textstyle\sum_{k=1}^{N_m} 0.5\beta_k\left(\boldsymbol{K}_{\phi_{kt}}\boldsymbol{\kappa}_t - \boldsymbol{M}_{\phi_{kt}}\right)^{\mathrm{T}}\left(\boldsymbol{K}_{\phi_{kt}}\boldsymbol{\kappa}_t - \boldsymbol{M}_{\phi_{kt}}\right) + c_2
\end{aligned} \tag{4.25}$$

where $\boldsymbol{K}_{\phi_{kt}} = [\boldsymbol{K}_1'\boldsymbol{\phi}_{kt} \quad \dots \quad \boldsymbol{K}_{N_k}'\boldsymbol{\phi}_{kt}]$ and $\boldsymbol{M}_{\phi_{kt}} = \lambda_{kt}'\boldsymbol{M}'\boldsymbol{\phi}_{kt} - \boldsymbol{K}_0'\boldsymbol{\phi}_{kt}$. The above conditional distribution does not belong to any standard family, so it is impossible to directly sample from it. Here, we apply the Metropolis-Hastings (M-H) algorithm [100,101] to asymptotically obtain samples of $\boldsymbol{\kappa}_t$.

In the M-H algorithm, a proposal distribution has to be selected, and it is crucial for the performance of the algorithm [104]. To avoid the trouble of tuning the proposal distribution, we propose an independent M-H algorithm considering the structure of the conditional distribution shown in Eqn. (4.25). More specifically, there are two quadratic terms involving $\boldsymbol{\kappa}_t$, which suggests that we can use the multivariate normal distribution as an independent proposal, and then reject or accept the sample by comparing the remaining terms.

The quadratic terms in Eqn. (4.25) gives a multivariate normal distribution with the precision matrix and mean vector as

$$\boldsymbol{\Sigma}_{\boldsymbol{\kappa}_t}^{-1} = \boldsymbol{P} + \sum_{k=1}^{N_m}\beta_k\boldsymbol{K}_{\phi_{kt}}^{\mathrm{T}}\boldsymbol{K}_{\phi_{kt}} \tag{4.26}$$

$$\boldsymbol{\mu}_{\boldsymbol{\kappa}_t} = \boldsymbol{\Sigma}_{\boldsymbol{\kappa}_t}\left[\boldsymbol{P}\boldsymbol{\kappa} + \sum_{k=1}^{N_m}\beta_k\boldsymbol{K}_{\phi_{kt}}^{\mathrm{T}}\boldsymbol{M}_{\phi_{kt}}\right] \tag{4.27}$$

It is trivial to sample from such a distribution. Since the sampling does not depend on the current sample $\boldsymbol{\kappa}_t^{(s)}$, it is called independent proposal. Once a candidate sample $\boldsymbol{\kappa}_t^{(*)}$ is made, in order to achieve the desired distribution, we accept it with the probability

$$\mathcal{A}\left(\boldsymbol{\kappa}_t^{(s)}, \boldsymbol{\kappa}_t^{(*)}\right) = \min\left\{1, \quad \exp\left[\mathcal{D}\left(\boldsymbol{\kappa}_t^{(*)}\right) - \mathcal{D}\left(\boldsymbol{\kappa}_t^{(s)}\right)\right]\right\} \tag{4.28}$$

where we have defined

$$\begin{aligned}\mathcal{D}(\boldsymbol{\kappa}_t) = \sum_{k=1}^{N_m} \Big[&\log|\boldsymbol{K}'(\boldsymbol{\kappa}_t) - \lambda'_{kt}\boldsymbol{M}'| + \log|\mathrm{tr}[\mathrm{adj}(\boldsymbol{K}'(\boldsymbol{\kappa}_t) - \lambda'_{kt}\boldsymbol{M}')\boldsymbol{M}']| \\ &- \frac{\alpha_k}{2}|\boldsymbol{K}'(\boldsymbol{\kappa}_t) - \lambda'_{kt}\boldsymbol{M}'|^2\Big]\end{aligned} \tag{4.29}$$

It can be shown that the proposed independent M-H algorithm satisfies the detailed balance [104]; therefore, it is a valid sampling scheme, i.e. it asymptotically converges to the desired distribution.

Conditional distribution of $\boldsymbol{\kappa}$ and $\boldsymbol{P}$

A normal-Wishart prior distribution is assigned for $\boldsymbol{\kappa}$ and $\boldsymbol{P}$ with the likelihood (Eqn. (4.15)) being multivariate normal. Thus, their conditional distribution is expected to be the normal-Wishart as well. To see this, collecting the terms involving $\boldsymbol{\kappa}$ and $\boldsymbol{P}$ in Eqn. (4.21) gives

$$\begin{aligned}&\log p(\boldsymbol{\kappa}, \boldsymbol{P}|\boldsymbol{\kappa}_{1:T}) \\ &\propto 0.5(e_0 + T - N_k - 1)\log|\boldsymbol{P}| - 0.5\mathrm{tr}(\boldsymbol{E}_0^{-1}\boldsymbol{P}) + 0.5N_k \log p_0 \\ &-0.5p_0(\boldsymbol{\kappa} - \boldsymbol{\kappa}_0)^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{\kappa} - \boldsymbol{\kappa}_0) - 0.5\sum_{t=1}^{T}(\boldsymbol{\kappa}_t - \boldsymbol{\kappa})^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{\kappa}_t - \boldsymbol{\kappa}) \\ &\propto \log \mathrm{N}(\overline{\boldsymbol{\kappa}}, (p\boldsymbol{P})^{-1}) + \log \mathrm{Wishart}(\boldsymbol{E}, e)\end{aligned} \tag{4.30}$$

in which we have

$$\begin{aligned}p = p_0 + T, \;\; e = e_0 + T \\ \overline{\boldsymbol{\kappa}} = (p_0\boldsymbol{\kappa}_0 + \textstyle\sum_{t=1}^{T}\boldsymbol{\kappa}_t)/p \\ \boldsymbol{E}^{-1} = \boldsymbol{E}_0^{-1} + p_0\boldsymbol{\kappa}_0\boldsymbol{\kappa}_0^{\mathrm{T}} + \textstyle\sum_{t=1}^{\mathrm{T}}\boldsymbol{\kappa}_t\boldsymbol{\kappa}_t^{\mathrm{T}} - p\overline{\boldsymbol{\kappa}}\,\overline{\boldsymbol{\kappa}}^{\mathrm{T}}\end{aligned} \tag{4.31}$$

Conditional distribution of $\boldsymbol{\alpha_k}$ and $\boldsymbol{\beta_k}$

Parameters α_k and β_k relate to the variances and covariances of the distributions of

eigenvalues and mode shapes. Because conjugate priors are assigned to them, their posteriors belong to the same distributions as the priors. Specifically, we have

$$\begin{aligned}&\log p\left(\alpha_k|\boldsymbol{\kappa}_t,\lambda'_{k(1:T)}\right)\\&\propto \left(u_{k,0}+0.5T\right)\log\alpha_k-\alpha_k\left(h_{k,0}+0.5\sum_{t=1}^{T}|\boldsymbol{K}'(\boldsymbol{\kappa}_t)-\lambda'_{kt}\boldsymbol{M}'|^2\right)\\&\propto \log \mathrm{Gam}(u_k,h_k)\end{aligned} \tag{4.32}$$

where

$$\begin{aligned}u_k&=u_{k,0}+0.5T\\h_k&=h_{k,0}+0.5\sum_{t=1}^{T}|\boldsymbol{K}'(\boldsymbol{\kappa}_t)-\lambda'_{kt}\boldsymbol{M}'|^2\end{aligned} \tag{4.33}$$

Similarly, we can compute the conditional distribution of β_k as

$$\begin{aligned}&\log p\left(\beta_k|\boldsymbol{\kappa}_t,\lambda'_{k(1:T)},\boldsymbol{\phi}_{k(1:T)}\right)\\&\propto \left(r_{k,0}+0.5TN_d\right)\log\beta_k\\&\quad-\beta_k\left(s_{k,0}+0.5\sum_{t=1}^{T}\boldsymbol{\phi}_{kt}^{\mathrm{T}}[\boldsymbol{K}'(\boldsymbol{\kappa}_t)-\lambda'_{kt}\boldsymbol{M}']^{\mathrm{T}}[\boldsymbol{K}'(\boldsymbol{\kappa}_t)-\lambda'_{kt}\boldsymbol{M}']\boldsymbol{\phi}_{kt}\right)\\&\propto \log \mathrm{Gam}(r_k,s_k)\end{aligned} \tag{4.34}$$

where

$$\begin{aligned}r_k&=r_{k,0}+0.5TN_d\\s_k&=s_{k,0}+0.5\sum_{t=1}^{T}\boldsymbol{\phi}_{kt}^{\mathrm{T}}[\boldsymbol{K}'(\boldsymbol{\kappa}_t)-\lambda'_{kt}\boldsymbol{M}']^{\mathrm{T}}[\boldsymbol{K}'(\boldsymbol{\kappa}_t)-\lambda'_{kt}\boldsymbol{M}']\boldsymbol{\phi}_{kt}\end{aligned} \tag{4.35}$$

4.3.2 Robust sampling

The covariance matrices $\boldsymbol{\Sigma}_{\boldsymbol{\phi}_{kt}^u}$, $\boldsymbol{\Sigma}_{\boldsymbol{\kappa}_t}$ and $\boldsymbol{E}$ must be symmetric and positive semi-definite. This is essential to develop a robust sampling procedure to reduce the numerical error. For this purpose, the QR decomposition and Cholesky decomposition-based strategies are developed.

Sampling the latent variable $\boldsymbol{\phi}_{kt}^{u}$

A robust sampling of the random variables $\boldsymbol{\phi}_{kt}^{u}$ following $\mathrm{N}\left(\boldsymbol{\mu}_{\boldsymbol{\phi}_{kt}^u},\boldsymbol{\Sigma}_{\boldsymbol{\phi}_{kt}^u}\right)$ can be done as follows:

1) Take the QR decomposition

$$[\boldsymbol{K}'_u(\boldsymbol{\kappa}_t) - \lambda'_{kt}\boldsymbol{M}'_u \quad [\lambda'_{kt}\boldsymbol{M}'_m - \boldsymbol{K}'_m(\boldsymbol{\kappa}_t)]\boldsymbol{\phi}^m_{kt}] = \mathcal{Q}\begin{bmatrix}\mathcal{R}_{11} & \mathcal{R}_{12}\\ \mathbf{0} & \mathcal{R}_{22}\\ \mathbf{0} & \mathbf{0}\end{bmatrix} \tag{4.36}$$

2) Draw a sample $\boldsymbol{n}_\phi$ from a N_d-dimension standard normal distribution; the sample following $\mathrm{N}\left(\boldsymbol{\mu}_{\boldsymbol{\phi}^u_{kt}}, \boldsymbol{\Sigma}_{\boldsymbol{\phi}^u_{kt}}\right)$ is then calculated as

$$\boldsymbol{\phi}^{u\ (s+1)}_{kt} = \mathcal{R}_{11}^{-1}\left(\boldsymbol{n}_\phi/\sqrt{\beta_k} + \mathcal{R}_{12}\right) \tag{4.37}$$

where s is the index of the sampling step.

The validity of the above scheme is shown in Section 3.5.2, thus it is omitted here.

Sampling unknown parameters $\boldsymbol{\kappa_t}$

Following a similar sampling strategy, the robust sampling procedure from $\mathrm{N}\left(\boldsymbol{\mu}_{\boldsymbol{\kappa}_t}, \boldsymbol{\Sigma}_{\boldsymbol{\kappa}_t}\right)$ can be implemented as follows:

1) Do the Cholesky decomposition: $\boldsymbol{P} = \boldsymbol{P}^{\mathrm{T}/2}\boldsymbol{P}^{1/2}$
2) Take the QR decomposition:

$$\begin{bmatrix}\boldsymbol{P}^{1/2} & \boldsymbol{P}^{1/2}\boldsymbol{\kappa}\\ \sqrt{\beta_1}\boldsymbol{K}_{\phi_{1t}} & \sqrt{\beta_1}\boldsymbol{M}_{\phi_{1t}}\\ \vdots & \vdots\\ \sqrt{\beta_{N_m}}\boldsymbol{K}_{\phi_{N_m t}} & \sqrt{\beta_{N_m}}\boldsymbol{M}_{\phi_{N_m t}}\end{bmatrix} = \bar{\mathcal{Q}}\begin{bmatrix}\bar{\mathcal{R}}_{11} & \bar{\mathcal{R}}_{12}\\ \mathbf{0} & \bar{\mathcal{R}}_{22}\\ \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix} \tag{4.38}$$

3) Draw a sample $\boldsymbol{n}_{\kappa_t}$ from a N_k-dimension standard normal distribution; obtain a sample following $\mathrm{N}\left(\boldsymbol{\mu}_{\boldsymbol{\kappa}_t}, \boldsymbol{\Sigma}_{\boldsymbol{\kappa}_t}\right)$ as

$$\boldsymbol{\kappa}_t^{(*)} = \bar{\mathcal{R}}_{11}^{-1}\left(\boldsymbol{n}_{\kappa_t} + \bar{\mathcal{R}}_{12}\right) \tag{4.39}$$

Sampling unknown parameters $\boldsymbol{\kappa}$ and $\boldsymbol{P}$

For the robust sampling of $\boldsymbol{\kappa}$ and $\boldsymbol{P}$, we choose the Cholesky decomposition

$$\begin{bmatrix}p_0 + T & (p_0\boldsymbol{\kappa}_0 + \sum_{t=1}^{T}\boldsymbol{\kappa}_t)^{\mathrm{T}}\\ (p_0\boldsymbol{\kappa}_0 + \sum_{t=1}^{T}\boldsymbol{\kappa}_t) & \boldsymbol{E}_0^{-1} + p_0\boldsymbol{\kappa}_0\boldsymbol{\kappa}_0^{\mathrm{T}} + \sum_{t=1}^{\mathrm{T}}\boldsymbol{\kappa}_t\boldsymbol{\kappa}_t^{\mathrm{T}}\end{bmatrix} = \begin{bmatrix}L_{11} & \mathbf{0}\\ \boldsymbol{L}_{21} & \boldsymbol{L}_{22}\end{bmatrix}\begin{bmatrix}L_{11} & \mathbf{0}\\ \boldsymbol{L}_{21} & \boldsymbol{L}_{22}\end{bmatrix}^{\mathrm{T}} \tag{4.40}$$

By equating the sub-matrices, we have

$$\bar{\boldsymbol{\kappa}} = \boldsymbol{L}_{21}L_{11}^{-1} \tag{4.41}$$

$$\boldsymbol{E}^{-1} = \boldsymbol{L}_{22}\boldsymbol{L}_{22}^{\mathrm{T}} \tag{4.42}$$

For the purpose of efficiently sampling $\boldsymbol{\kappa}$ after sampling $\boldsymbol{P}$ based on the degree of

freedom e and $\boldsymbol{L}_{22}$, we apply the following procedure: Take the Cholesky decomposition $\boldsymbol{P}^{(s+1)} = \widetilde{\boldsymbol{R}}_{22}^{\mathrm{T}}\widetilde{\boldsymbol{R}}_{22}$, then sample a standard multivariate normal vector $\boldsymbol{n}_{\boldsymbol{\kappa}} \in \mathbb{R}^{N_k}$ and set

$$\boldsymbol{\kappa}^{(s+1)} = (\boldsymbol{L}_{21} + \widetilde{\boldsymbol{R}}_{22}^{-1}\boldsymbol{n}_{\boldsymbol{\kappa}})L_{11}^{-1} \tag{4.43}$$

As a summary, the robust implementation of MwG sampler is listed in Figure 4.4.

---

1) Initialization

Choose hyper-parameters $\boldsymbol{\kappa}_0$, p_0, $\boldsymbol{E}_0$, e_0, $u_{k,0}$, $h_{k,0}$, $r_{k,0}$, and $s_{k,0}$

Set $\boldsymbol{\kappa}_t^{(0)} = \boldsymbol{\kappa}_0$, $\boldsymbol{\kappa}^{(0)} = \boldsymbol{\kappa}_0$, $\boldsymbol{P}^{(0)} = e_0\boldsymbol{E}_0$, $\alpha_k^{(0)} = u_{k,0}/h_{k,0}$ and $\beta_k^{(0)} = r_{k,0}/s_{k,0}$

For $s = 0$ to $N_s - 1$

2) Sample $\boldsymbol{\phi}_{kt}^{u}$

$\mathcal{R} = qr\left(\left[\boldsymbol{K}_u'\left(\boldsymbol{\kappa}_t^{(s)}\right) - \lambda_{kt}'\boldsymbol{M}_u' \quad [\lambda_{kt}'\boldsymbol{M}_m' - \boldsymbol{K}_m'(\boldsymbol{\kappa}_t)]\boldsymbol{\phi}_{kt}^{m}\right]\right)$

$\boldsymbol{\phi}_{kt}^{u\ (s+1)} = \mathcal{R}_{11}^{-1}\left(\boldsymbol{n}_{\phi}/\sqrt{\beta_k^{(s)}} + \mathcal{R}_{12}\right)$ where $\boldsymbol{n}_{\phi}\sim\mathrm{N}\left(\boldsymbol{0}_{N_d-N_o}, \boldsymbol{I}_{N_d-N_o}\right)$

3) Sample $\boldsymbol{\kappa}_t$

$$\bar{\mathcal{R}} = qr\left(\begin{bmatrix} \boldsymbol{P}^{(s)^{1/2}} & \boldsymbol{P}^{(s)^{1/2}}\boldsymbol{\kappa}^{(s)} \\ \sqrt{\beta_1^{(s)}}\boldsymbol{K}_{\phi_{1t}}^{(s+1)} & \sqrt{\beta_1^{(s)}}\boldsymbol{M}_{\phi_{1t}}^{(s+1)} \\ \vdots & \vdots \\ \sqrt{\beta_{N_m}^{(s)}}\boldsymbol{K}_{\phi_{N_m t}}^{(s+1)} & \sqrt{\beta_{N_m}^{(s)}}\boldsymbol{M}_{\phi_{N_m t}}^{(s+1)} \end{bmatrix}\right)$$

$\boldsymbol{\kappa}_t^{(*)} = \bar{\mathcal{R}}_{11}^{-1}\left(\boldsymbol{n}_{\kappa_t} + \bar{\mathcal{R}}_{12}\right)$ where $\boldsymbol{n}_{\kappa_t}\sim\mathrm{N}\left(\boldsymbol{0}_{N_k}, \boldsymbol{I}_{N_k}\right)$

Compute $\mathcal{A}\left(\boldsymbol{\kappa}_t^{(s)}, \boldsymbol{\kappa}_t^{(*)}\right)$ defined in Eqn. (4.28)

If $\mathcal{A}\left(\boldsymbol{\kappa}_t^{(s)}, \boldsymbol{\kappa}_t^{(*)}\right) > \mathrm{Unif}(0,1)$ $\boldsymbol{\kappa}_t^{(s+1)} = \boldsymbol{\kappa}_t^{(*)}$

Else $\boldsymbol{\kappa}_t^{(s+1)} = \boldsymbol{\kappa}_t^{(s)}$

4) Sample $\boldsymbol{\kappa}$ and $\boldsymbol{P}$

$$\boldsymbol{L} = \mathrm{chol}\left(\begin{bmatrix} p_0 + T & (p_0\boldsymbol{\kappa}_0 + \sum_{t=1}^{T}\boldsymbol{\kappa}_t^{(s+1)})^{\mathrm{T}} \\ (p_0\boldsymbol{\kappa}_0 + \sum_{t=1}^{T}\boldsymbol{\kappa}_t^{(s+1)}) & \boldsymbol{E}_0^{-1} + p_0\boldsymbol{\kappa}_0\boldsymbol{\kappa}_0^{\mathrm{T}} + \sum_{t=1}^{\mathrm{T}}\boldsymbol{\kappa}_t^{(s+1)}\boldsymbol{\kappa}_t^{(s+1)^{\mathrm{T}}} \end{bmatrix}\right)$$

$\boldsymbol{P}^{(s+1)}\sim\mathrm{Wishart}(e, (\boldsymbol{L}_{22}\boldsymbol{L}_{22}^{\mathrm{T}})^{-1})$ and $\widetilde{\boldsymbol{R}}_{22} = \mathrm{chol}(\boldsymbol{P}^{(s+1)}, \mathrm{'upper'})$

$\boldsymbol{\kappa}^{(s+1)} = (\boldsymbol{L}_{21} + \widetilde{\boldsymbol{R}}_{22}^{-1}\boldsymbol{n}_{\kappa})L_{11}^{-1}$ where $\boldsymbol{n}_{\kappa}\sim\mathrm{N}\left(\boldsymbol{0}_{N_k}, \boldsymbol{I}_{N_k}\right)$

5) Sample α_k and β_k

$u_k = u_{k,0} + \frac{T}{2}$, $h_k = h_{k,0} + \frac{1}{2}\sum_{t=1}^{T}\left|\boldsymbol{K}'\left(\boldsymbol{\kappa}_t^{(s+1)}\right) - \lambda_{kt}'\boldsymbol{M}'\right|^2$

$\alpha_k^{(s+1)}\sim\mathrm{Gam}(u_k, h_k)$

$r_k = r_{k,0} + \frac{TN_d}{2}$, $s_k = s_{k,0} + \frac{1}{2}\sum_{t=1}^{T}\boldsymbol{\phi}_{kt}^{(s+1)^{\mathrm{T}}}\left[\boldsymbol{K}'\left(\boldsymbol{\kappa}_t^{(s+1)}\right) - \lambda_{kt}'\boldsymbol{M}'\right]^{\mathrm{T}}\left[\boldsymbol{K}'\left(\boldsymbol{\kappa}_t^{(s+1)}\right) - \lambda_{kt}'\boldsymbol{M}'\right]\boldsymbol{\phi}_{kt}^{(s+1)}$

$\beta_k^{(s+1)}\sim\mathrm{Gam}(r_k, s_k)$

End For

Figure 4.4 Robust MwG sampler for FEMU.

Deterministic annealing

One drawback of the MwG sampler is that it can be trapped in a local maximum for a long period of time. To accelerate the global convergence, a simple deterministic annealing scheme [143] is employed. As can be seen from Eqns. (4.12) and (4.14), the spread of the likelihood highly depends on the values of precision parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. In order to explore a larger area, these values should be small. Therefore, in the deterministic annealing scheme, we artificially control these values to ensure small values are generated for the first hundreds of samples, then gradually loosen this constraint until it becomes immaterial and correct posterior samples are obtained.

4.4 Empirical studies

In this section, the performance of the MwG sampler for FEMU is empirically studied through two examples: an 8-DoF mass-spring system and a 6-story shear-type experimental building model. The first example is used to illustrate the capability of the proposed method to accurately capture the time variation of the model parameters. The second example demonstrates the uncertainty propagation from the noisy measurements (samples of the Gibbs sampler in Section 3.5) to the identified model parameters and the applicability of the method to a real physical model.

4.4.1 8-DoF mass-spring system

The 8-DoF mass-spring system has the same structure as that used in Section 3.6.1 (Figure 3.5). The nominal model parameters values are $m = 1$ and $k_i = 800i$ for $i = 1, \dots, 9$. The true stiffness factor is assumed to have a multivariate normal distribution $\boldsymbol{\kappa}_t \sim \mathrm{N}(\boldsymbol{\mu}_\kappa, \boldsymbol{P}_\kappa^{-1})$ where $\boldsymbol{\mu}_\kappa = [-0.2, -0.15, -0.1, -0.05, 0, 0.05, 0.1, 0.15, 0.2]^{\mathrm{T}}$ and $\boldsymbol{P}_\kappa = 400\boldsymbol{I}_9$, i.e. the structure is assumed to be time variant with a standard deviation 0.05. In order to synthesize artificial measurements, samples $\boldsymbol{\kappa}_{1:T}$ are first generated from $\mathrm{N}(\boldsymbol{\kappa}_0, \boldsymbol{P}_\kappa^{-1})$; the modal frequencies λ_{kt} and model shapes $\boldsymbol{\phi}_{kt}$ are obtained via the generalized eigenvalue decomposition of the assembled mass and stiffness matrices; finally measurement errors, modeled as zero-mean Gaussian white noises with variances such that the COVs of λ_{kt} and $\boldsymbol{\phi}_{kt}$ are both equal to 0.01, are added. The FEMU problem is to identify $\boldsymbol{\kappa}$ and $\boldsymbol{P}$ given the "measured" data $\{\lambda_{kt}, \boldsymbol{\phi}_{kt}^m, k = 1, \dots, N_m, t = 1, \dots, T\}$.

The mass matrix is assumed to be exactly known, so there is no model error involved in this example. The initial stiffness parameters are selected as $\boldsymbol{\kappa}_0 = \mathbf{0}_{9\times 1}$, i.e. starting from the nominal model. Other hyper-parameters, including p_0, $\boldsymbol{E}_0$, e_0, $u_{k,0}$, $h_{k,0}$, $r_{k,0}$, and $s_{k,0}$, are chosen such that the priors are all sufficiently flat.

Suppose all 8 DoFs and 8 modes are utilized with $T = 500$. Figure 4.5 shows the convergence process of the proposed MwG sampler. The step-like increase of the log likelihood is due to the annealing scheme employed. It can be seen that the parameters converge quickly, and the global optimum is found after less than 300 samples. For posterior analysis, the first 1000 samples are discarded as the burn-in period. The posterior distributions of unknown model parameters are shown in Figure 4.6. The generated samples approximate the true distributions of model parameters $\boldsymbol{\kappa}_t \sim \mathrm{N}(\boldsymbol{\mu}_\kappa, \boldsymbol{P}_\kappa^{-1})$ accurately validated by the marginal distributions and near zero correlation coefficients between each pair of identified model parameters.

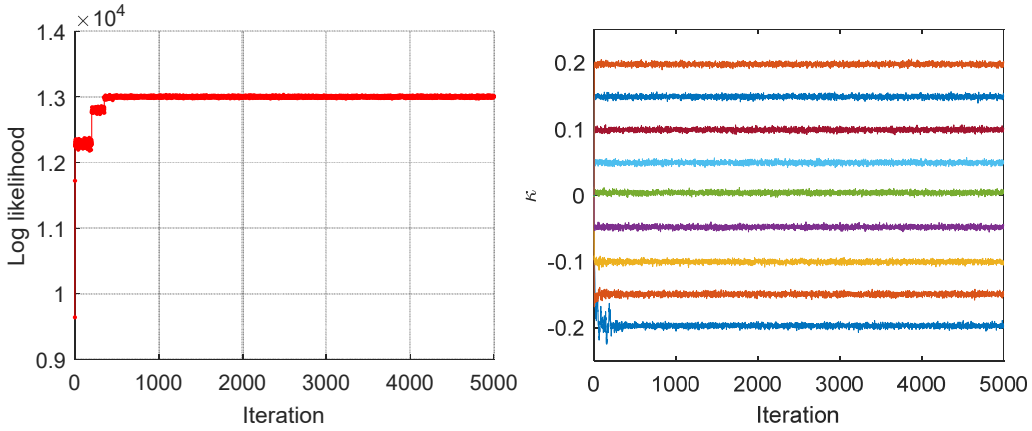
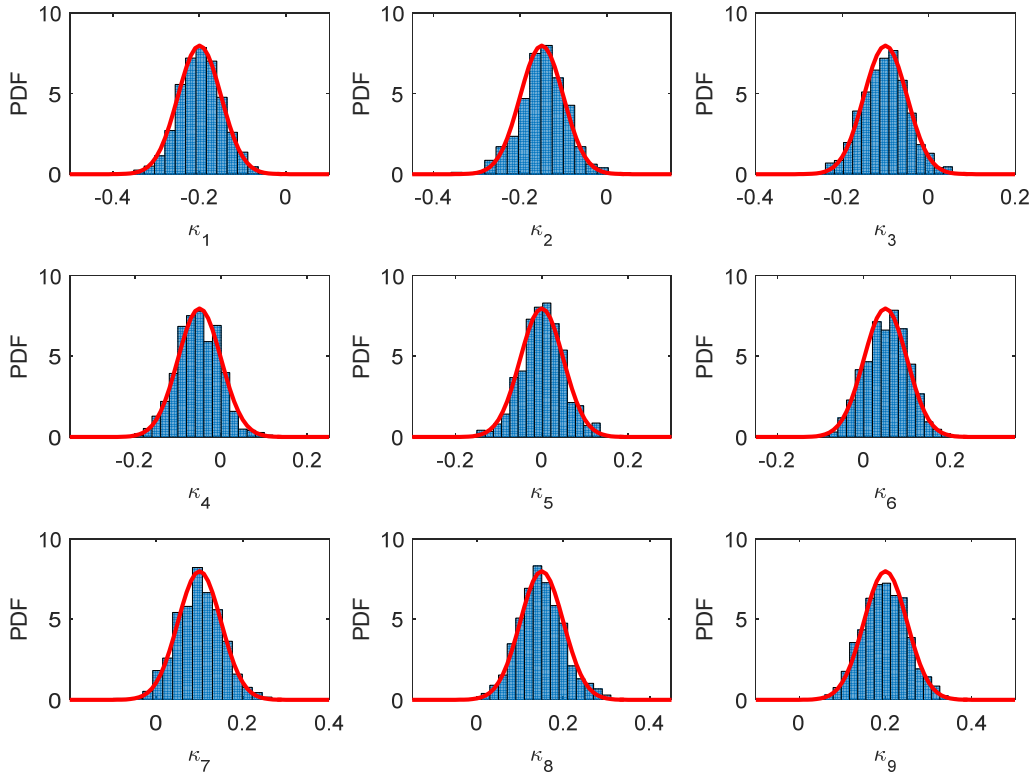
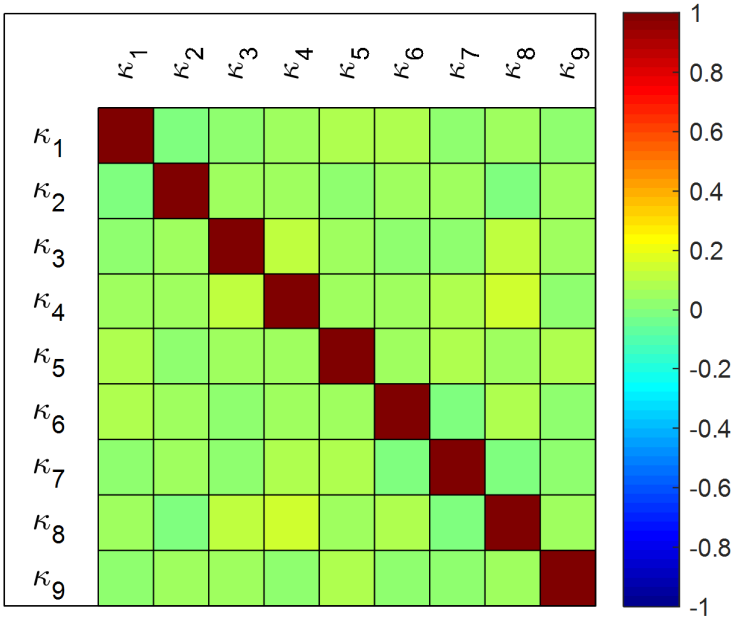


Figure 4.5 Convergence of the MwG sampler.

To investigate the effect of the data length T on the identified model parameters, we artificially generate 600 samples to construct six different subsets of data with $T = \{10,50,100,300,500,600\}$ corresponding to the first 10 samples, first 50 samples, etc. Then, they are used for model updating. The posterior means and standard deviations are reported in Table 4.1 for the considered data subsets. When the number of data is insufficient, a biased posterior distribution can be generated. It is not surprising that the bias and variability reduce when the data length T increases. The required length of data for accurate estimation can be checked from the convergence of the statistics shown in Figure 4.7. The root mean square (RMS) errors of the means and standard deviations converge after $T = 300$. A similar convergence result is also illustrated for the computed modal frequencies. Such large length of data can be collected in monitored structures in a matter of few days although a longer period of data collection is recommended to observe the full range of environmental and ambient variations.



a) Marginal distributions (with true PDF shown as solid line)

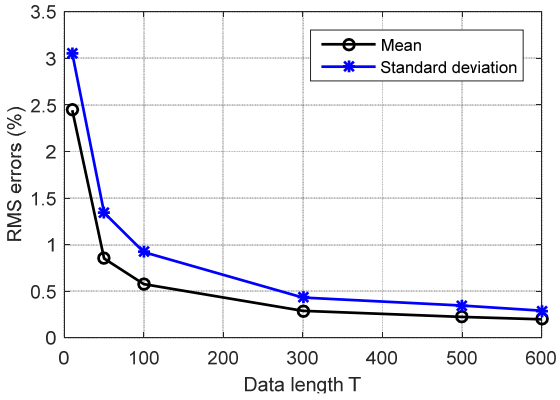


b) Correlation coefficients

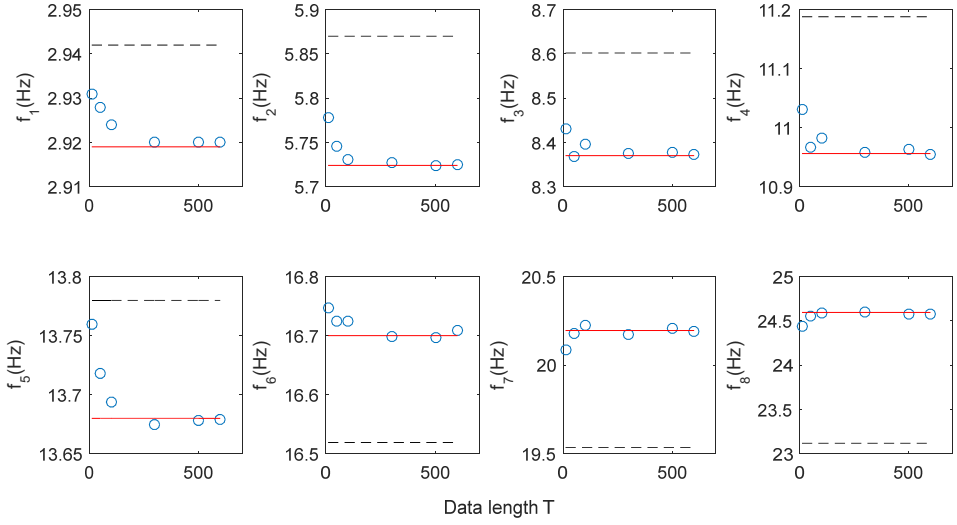
Figure 4.6 Posterior distribution of model parameters.

Table 4.1 Identified model parameters with different data length (standard deviation in parentheses)

True $\boldsymbol{\kappa}$	$T = 10$	$T = 50$	$T = 100$	$T = 300$	$T = 500$	$T = 600$
-0.20(0.05)	-0.164(0.088)	-0.182(0.059)	-0.196(0.058)	-0.197(0.055)	-0.198(0.053)	-0.197(0.053)
-0.15(0.05)	-0.123(0.078)	-0.157(0.065)	-0.140(0.057)	-0.148(0.054)	-0.147(0.055)	-0.150(0.054)
-0.10(0.05)	-0.065(0.073)	-0.088(0.060)	-0.100(0.061)	-0.097(0.055)	-0.100(0.055)	-0.101(0.052)
-0.05(0.05)	-0.035(0.074)	-0.040(0.061)	-0.052(0.057)	-0.053(0.056)	-0.052(0.051)	-0.048(0.051)
0.00(0.05)	0.009(0.076)	0.001(0.061)	0.005(0.057)	0.002(0.054)	0.002(0.051)	0.004(0.051)
0.05(0.05)	0.030(0.081)	0.048(0.067)	0.054(0.062)	0.044(0.051)	0.053(0.052)	0.049(0.054)
0.10(0.05)	0.101(0.081)	0.097(0.066)	0.105(0.059)	0.099(0.052)	0.099(0.054)	0.099(0.053)
0.15(0.05)	0.132(0.086)	0.145(0.060)	0.143(0.062)	0.151(0.054)	0.148(0.054)	0.149(0.053)
0.20(0.05)	0.167(0.084)	0.199(0.068)	0.208(0.060)	0.201(0.056)	0.199(0.054)	0.198(0.053)



a) Root mean square errors of the mean and standard deviation



b) Modal frequencies

Figure 4.7 Effect of data length on identified parameters.

Setting $T = 500$ and $N_m = 8$, consider the case where only partial DoFs are measured for model updating. The results of the identified model parameters are listed in Table 4.2 for 4 different subsets of measured DoFs $N_o = \{2,4,6,8\}$ corresponding to the measured DoFs $\{5,8\}$, $\{2,5,6,8\}$, $\{2-6,8\}$ and $\{1-8\}$. In addition, the RMS errors of the identified model parameters are illustrated in Figure 4.8a. Though the bias and variance slightly reduce with the number of measured DoFs increasing, the identification results are generally insensitive to the number of DoFs used. This is because more measured DoFs directly decreases the variance of $\boldsymbol{\kappa}_t$ as shown in Eqn. (4.26), but the influence on the posterior of $\boldsymbol{\kappa}$ is trivial because its variance mainly explains the time variation of the stiffness.

Table 4.2 Identified model parameters with different measured DoFs (standard deviation in parentheses)

True $\boldsymbol{\kappa}$	$N_o = 2$	$N_o = 4$	$N_o = 6$	$N_o = 8$
-0.20(0.05)	-0.193(0.063)	-0.198(0.053)	-0.198(0.053)	-0.198(0.053)
-0.15(0.05)	-0.145(0.056)	-0.147(0.055)	-0.147(0.055)	-0.147(0.055)
-0.10(0.05)	-0.102(0.057)	-0.100(0.055)	-0.100(0.055)	-0.100(0.055)
-0.05(0.05)	-0.047(0.060)	-0.052(0.051)	-0.052(0.051)	-0.052(0.051)
0.00(0.05)	-0.003(0.056)	0.002(0.051)	0.002(0.051)	0.002(0.051)
0.05(0.05)	0.056(0.055)	0.053(0.052)	0.053(0.052)	0.053(0.052)
0.10(0.05)	0.098(0.055)	0.099(0.054)	0.099(0.054)	0.099(0.054)
0.15(0.05)	0.141(0.061)	0.148(0.054)	0.148(0.054)	0.148(0.054)
0.20(0.05)	0.200(0.055)	0.199(0.054)	0.199(0.054)	0.199(0.054)

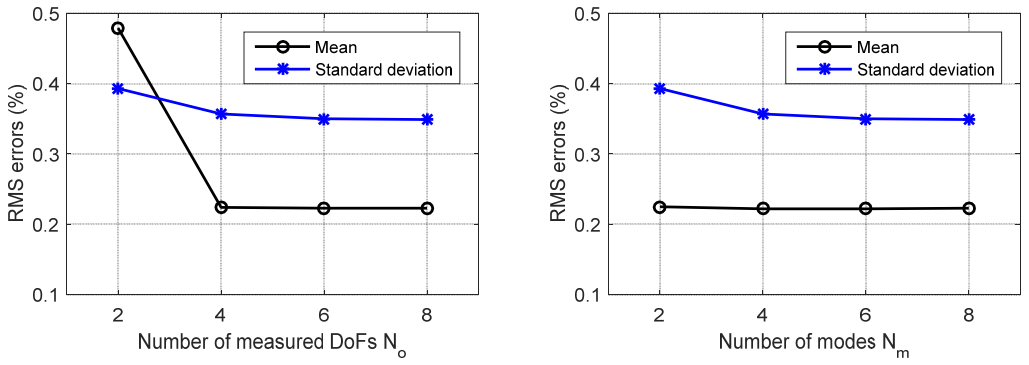


Figure 4.8 Effect of measured modes and DoFs on identified parameters.

The influence of the number of measured modes N_m is also studies with results listed in Table 4.3 and Figure 4.8b. In this case, we choose the data length $T = 500$ and measured DoFs $\{2,5,6,8\}$, and the 4 subsets of measured modes corresponds to the first two, first four, first six and all the eight modes. Again, we see that the generated samples converges to the true values of parameters and the variance is not sensitive to how many modes are utilized. The reason, same as the effect of measured DoFs, is that when more modes included there is a reduction of the variance of $\boldsymbol{\kappa}_t$ as

shown in Eqn. (4.26), but not that of $\boldsymbol{\kappa}$.

Table 4.3 Identified model parameters with different measured modes (standard deviation in parentheses)

True $\boldsymbol{\kappa}$	$N_m = 2$	$N_m = 4$	$N_m = 6$	$N_m = 8$
-0.20(0.05)	-0.198(0.053)	-0.198(0.053)	-0.197(0.053)	-0.198(0.053)
-0.15(0.05)	-0.147(0.056)	-0.147(0.055)	-0.147(0.055)	-0.147(0.055)
-0.10(0.05)	-0.100(0.055)	-0.100(0.055)	-0.100(0.055)	-0.100(0.055)
-0.05(0.05)	-0.053(0.051)	-0.052(0.051)	-0.052(0.051)	-0.052(0.051)
0.00(0.05)	0.002(0.052)	0.002(0.051)	0.002(0.051)	0.002(0.051)
0.05(0.05)	0.053(0.052)	0.053(0.052)	0.053(0.052)	0.053(0.052)
0.10(0.05)	0.099(0.055)	0.099(0.055)	0.099(0.054)	0.099(0.054)
0.15(0.05)	0.147(0.055)	0.148(0.055)	0.148(0.054)	0.148(0.054)
0.20(0.05)	0.199(0.054)	0.199(0.054)	0.199(0.054)	0.199(0.054)

4.4.2 6-story shear-type building

The second example is a six-story 1/4-scale shear-type building, for which the modal parameters have been identified in Section 3.6.2. We assume the building has masses lumped at the floor levels, rigid floors and axially rigid columns, i.e. a shear-type building. According to the design of this experimental structure, all stories have the same stiffness $k_i = 1940.6\, kN/m \;\; (i = 1, \ldots, 6)$, and the total mass of each floor is $m_i = 862.85\, kg \;\; (i = 1, \ldots, 5)$, except the roof, which has $m_6 = 803.98\, kg$. Based on these parameters and the fixed-base shear-type building assumption, a nominal FE model is constructed. As can be seen from Table 4.4, there are large discrepancies between the nominal modal parameters and the measured ones. Hence, it is necessary to update the nominal FE model to match with the measurements.

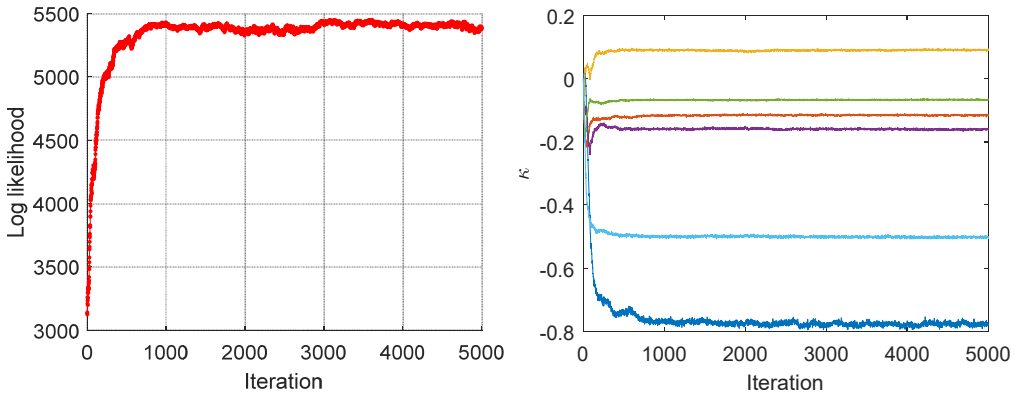
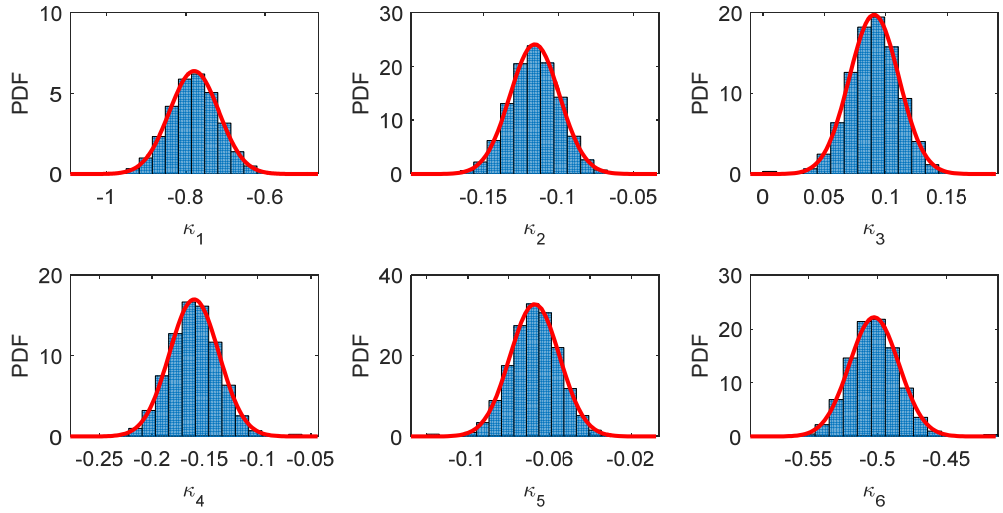


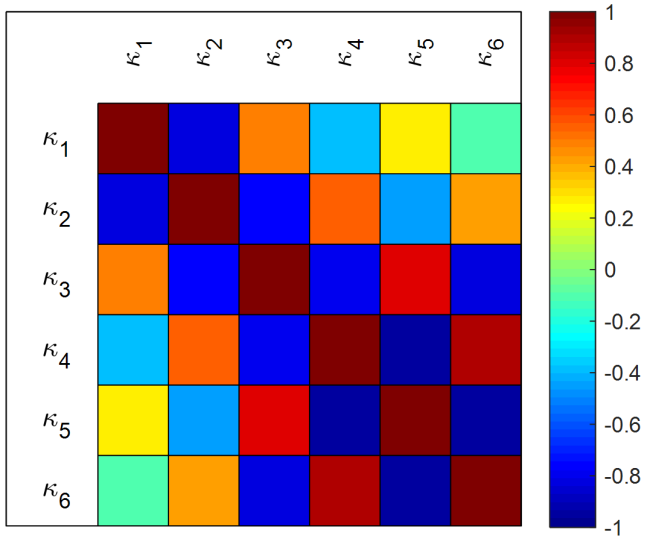
Figure 4.9 Convergence of the MwG sampler.

Unlike the previous example that considered the time variation of the model, this example illustrates the uncertainty propagation from the measured modal parameters to the model parameters. 1800 samples of modal parameters are generated by the

Gibbs samplers introduced in Section 3.5, but these samples are correlated. To reduce the correlation, a 'thinning' step [83] is first implemented, i.e. we discard all but every 5th samples, hence yielding a total of $T = 360$ data sets. Though all DoFs are measured for this structure, we only use the results from DoFs $\{1,3,6\}$ to illustrate the usual practice where only a subset of the DoFs are measured for real structures. All prior settings of the MwG sampler are the same as the previous examples, except the number of unknown parameters, which is now six, i.e. $\kappa_1,\ldots,\kappa_6$.



a) Marginal distributions (with approximated normal PDFs shown as solid line)



b) Correlation coefficients

Figure 4.10 Posterior distribution of model parameters.

Utilizing all measured modes for model updating, the convergence of the algorithm and converged result are shown in Figure 4.9 and Figure 4.10. The MwG sampler converges after 1000 samples, and these samples are discarded as the burn-in period, so that 4000 effective samples are used for posterior analysis. It seems that the posterior of model parameters can be well approximated by normal distributions and interesting patterns for the correlation coefficients exist. For example, κ_1 is negatively correlated with κ_2, κ_4 and κ_6, and positively correlated with κ_3 and κ_5, and the extent of correlation decreases with the distance from κ_1 increasing. This can be understood physically from the composition of the stiffness matrix $\boldsymbol{K}$. For instance, the $(1,1)$ element of the matrix is $\boldsymbol{K}_{11} = \kappa_1 k_1 + \kappa_2 k_2$; Therefore, κ_1 and κ_2 would be negatively correlated. Similarly, κ_2 and κ_3 are negatively correlated, so that κ_1 and κ_3 become positively correlated. Another pattern is that the magnitude of correlation increases along the diagonal direction, e.g. magnitude of correlation between κ_1 and κ_2 is less than that between κ_2 and κ_3. This may be explained by the sensitivity of modal parameters to model parameters. Specifically, modal parameters are more sensitive to the stiffness parameters close to roof in the shear-type building.

Table 4.4 Mean modal frequencies with different target modes

Target modes	f_1(Hz)	f_2(Hz)	f_3(Hz)	f_4(Hz)	f_5(Hz)	f_6(Hz)	RMS
Measured	1.134	3.661	6.455	9.369	12.123	14.098	0
Nominal	1.444	5.317	8.586	11.338	13.388	14.665	1.482
[1,3]	1.131	4.233	6.456	8.642	10.151	11.245	1.466
[5,6]	1.391	4.743	8.141	11.167	12.125	14.097	1.100
[1,3,5]	1.174	4.589	6.467	10.218	12.121	12.484	0.836
[1,4,6]	1.175	4.197	7.518	9.389	13.700	14.097	0.806
[1,3,4,6]	1.182	4.355	6.434	9.391	13.568	14.078	0.655
$[1-6]$	0.952	4.082	6.844	9.361	12.126	14.097	0.246

Table 4.5 Identified model parameters with different target modes (standard deviation in parentheses)

Target modes	κ_1	κ_2	κ_3	κ_4	κ_5	κ_6
[1,3]	-0.221(0.024)	-0.436(0.035)	-0.515(0.037)	-0.350(0.034)	-0.377(0.045)	-0.500(0.050)
[5,6]	0.100(0.073)	-0.155(0.039)	-0.101(0.027)	0.152(0.028)	-0.476(0.054)	-0.046(0.028)
[1,3,5]	-0.233(0.035)	0.022(0.022)	-0.645(0.042)	-0.172(0.043)	-0.332(0.032)	-0.125(0.026)
[1,4,6]	-0.536(0.080)	0.235(0.036)	-0.143(0.035)	-0.539(0.051)	0.345(0.031)	-0.313(0.026)
[1,3,4,6]	-0.528(0.045)	0.349(0.034)	-0.404(0.033)	-0.251(0.037)	0.297(0.058)	-0.556(0.029)
$[1-6]$	-0.777(0.078)	-0.115(0.023)	-0.090(0.025)	-0.160(0.029)	-0.067(0.019)	-0.500(0.037)

One question that often arises when updating an FE model is which measured modes should be used. Based on different target modes, we restart the MwG sampler and all results are reported in Table 4.4 and Table 4.5. It turns out that significantly different FE models are identified. Because of the existence of model error, the constructed shear-type FE model cannot explain the measured modal parameters sufficiently, leading to different posterior distributions under different target modes. If only partial modes are used, the updated model may explain these modes almost exactly, but provide poor performance in reproducing the remaining modes. It is suggested to use as many modes as possible in order to find a more representative model.

Though the MwG sampler finds a FE model to fit the measured modal parameters, we should be careful when interpreting it. For example, when using all modes, the parameter κ_1 has a much smaller value than the nominal ones, and much larger uncertainties than all others, so that it is highly suspicious. This may due to the flexible base that violates the assumption in modeling. In addition, from Table 4.4, the modal parameters computed from the identified model cannot exactly match the measured ones, which may be caused by the misspecification of mass parameters or the over-simplification of the FE model. Therefore, a more complex model to include more unknown parameters is preferred for this example for the purpose to fully explain the measurements. However, a too complex model may yield an unidentifiable situation, so that more thought should be given to balance the identifiability and the complexity of the model.

4.5 Conclusions

A Bayesian method for FEMU is proposed using the measured incomplete and noisy modal data and assuming a linear and classically damped structure. Based on the generalized eigenvalue decomposition of the stiffness and mass matrices and the assumptions on the model errors, a Bayesian model is first formulated, which can incorporate the time-variability, measurement error and model parameter uncertainty. In order to obtain the posterior distributions of model parameters, an MwG sampler is designed with a robust implementation. From the empirical studies for the FEMU of a numerical 8-DoF mass-spring system and an experimental 6-story shear-type building, the following conclusions are derived:

(1) The proposed Bayesian formulation explicitly models the time variance of the structure, so that the uncertainties caused by environmental effects and operational conditions can be handled.

(2) The FEMU procedure does not require matching the analytical and measured

modes; solving the eigenvalue problem of the dynamic model is also unnecessary.

(3) The designed MwG sampler successfully recovers the posterior distribution of model parameters. The introduced robust implementation strategy enhances its capability to reduce the numerical error and to improve the possibility to converge to the global optimum.

(4) It is found that the posterior variance depends on the number of data sets, but t is insensitive to the number of DoFs used and number of modes measured. The correlations between model parameters represent their physical dependence.

(5) In the presence of modeling errors, it is recommended to use as many modes as possible in order to get a more representative model. The FE model should be sufficiently complex so as to fully explain the measured modal data, but not too complex to make the model unidentifiable.

Chapter 5
Concluding Remarks

5.1 Summary of contributions

Three fundamental problems of SHM, i.e. optimal sensor placement, operational modal analysis and finite element model updating, are systematically investigated in this dissertation. A probabilistic model of each problem is constructed from its physical model. Bayesian techniques are then applied to obtain either the best decision or the posterior distributions of unknown parameters. The major contributions of the dissertation are summarized as follows:

(1) A robust optimal sensor placement methodology, considering model and load uncertainty as well as measurement error, is proposed based on the maximum expected utility theory and a linear Bayesian normal model. The major findings are as follows: (a) The optimal placement configuration of displacement meter, velocimeter and accelerometer can be different, and mixed sensor placement becomes possible when accounting for prior information. (b) Since the prior covariance matrix of modal acceleration has a large magnitude, the placement of accelerometer is less influenced by the prior information, justifying the commonly used mode-shape-based accelerometer placement. (c) The magnitude of input-to-noise ratio has a great influence on the optimal configuration of sensors, and it connects the kinetic energy-based and Fisher information-based sensor placement approaches.

(2) Time-domain, state-space model-based approaches for operational modal analysis are investigated, and posterior distributions of modal parameters are obtained. The methods of expectation maximization, variational Bayes and Gibbs sampler for operational modal analysis are theoretically derived and robust implementation strategies are provided. The major findings are as follows: (a) The expectation-maximization yields an accurate point estimation of modal parameters, but the posterior covariance matrix cannot be easily calculated. (b) The Gibbs sampler can be reliably used to approximate the posterior distributions but with a heavy computation burden. (c) The variational Bayes provides a good approximation of posterior distributions in a much more efficient way. (d) It is observed that the modal frequencies and mode shapes can be identified with small uncertainties, while the identified damping ratios generally have large variabilities. In addition, the coefficient of variation of the estimated frequency is approximately equal to the standard deviation of the estimated damping ratio in the same mode.

(3) A Bayesian method for finite element model updating is introduced using the measured incomplete and noisy modal data and assuming linear and classically damped structural behavior. The major findings are as follows: (a) In the proposed method, matching between analytical and experimental modes is not required, and the uncertainties caused by environmental effects and operational conditions as well as the measurement error in modal parameters can be effectively handled. (b) The designed Metropolis-within-Gibbs sampler accurately recovers the posterior distributions of stiffness parameters. (c) The posterior variance depends on the number of data sets, and correlations between stiffness parameters represent their physical dependence. (d) In the presence of modeling errors, it is recommended to use as many modes as possible in order to obtain an accurate model. The model should be sufficiently complex so as to fully explain the measured modal data.

5.2 Future work

This dissertation has laid a foundation for new approaches to damage diagnosis and prognosis and health management in SHM. Various works can be pursued in these areas.

(1) Damage detection is feasible using the identified modal parameters, and the updated finite element model can be used for damage localization and quantification. In order to extract damage information from the overwhelming environmental effect, a long set of monitored data and advanced tools are necessary. The combination of the Bayesian online change point detection [144] and the switching state space [145] is a potential tool to pursue in this direction.

(2) Based on the identified finite element model, the damage prognosis becomes possible if a load evolution model can be constructed. In this respect, a measure of structural reliability can work as a health index, and the surrogate-model accelerated subset simulation [146] can be applied for efficient evaluation of this index. As for the load evolution model, the dynamic Bayesian network [147] will be a good choice.

(3) For the purpose of structural health management, the influence diagram [148] can be an ideal tool by integrating decision and utility nodes into the Bayesian network model. A preliminary research [3] has been done in this area. The ultimate goal is to develop an online and automatic decision-making tool based on the recorded SHM data.

References

[1] American Society of Civil Engineers (ASCE). (2009). Report card for America's infrastructure, American Society of Civil Engineers: Reston, VA.
[2] National Academy of Engineering. (2012) Introduction to the grand challenges for engineering. http://www.engineeringchallenges.org/challenges/16091.aspx.
[3] Goulet, J. A., Der Kiureghian, A., & Li, B. (2015). Pre-posterior optimization of sequence of measurement and intervention actions under structural reliability constraint. Structural Safety, 52, 1-9.
[4] ASCE/SEI-AASHTO Ad-Hoc Group On Bridge Inspection, Rating, Rehabilitation, and Replacement. (2009). White paper on bridge inspection and rating. Journal of Bridge Engineering, 14(1), 1-5.
[5] Farrar, C. R., & Worden, K. (2012). Structural health monitoring: a machine learning perspective. John Wiley & Sons.
[6] Farrar, C. R., & Worden, K. (2007). An introduction to structural health monitoring. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 365(1851), 303-315.
[7] Rytter, A. (1993). Vibration based inspection of civil engineering structures. Aalborg: Aalborg University.
[8] Farrar, C. R., & Lieven, N. A. (2007). Damage prognosis: the future of structural health monitoring. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 365(1851), 623-632.
[9] Kohler, M. D., Davis, P. M., & Safak, E. (2005). Earthquake and ambient vibration monitoring of the steel-frame UCLA Factor building. Earthquake Spectra, 21(3), 715-736.
[10] Ni, Y. Q., Xia, Y., Liao, W. Y., & Ko, J. M. (2009). Technology innovation in developing the structural health monitoring system for Guangzhou New TV Tower. Structural Control and Health Monitoring, 16(1), 73-98.
[11] Ko, J. M., & Ni, Y. Q. (2005). Technology developments in structural health monitoring of large-scale bridges. Engineering structures, 27(12), 1715-1725.
[12] Brownjohn, J. M. W., Magalhaes, F., Caetano, E., & Cunha, A. (2010). Ambient vibration re-testing and operational modal analysis of the Humber Bridge. Engineering Structures, 32(8), 2003-2018.
[13] Koo, K. Y., Brownjohn, J. M. W., List, D. I., & Cole, R. (2013). Structural health monitoring of the Tamar suspension bridge. Structural Control and Health Monitoring, 20(4), 609-625.
[14] Der Kiureghian, A. (2005). Probabilistic methods for engineering risk analysis. Lecture notes, UC Berkeley.

[15] Brownjohn, J. M., De Stefano, A., Xu, Y. L., Wenzel, H., & Aktan, A. E. (2011). Vibration-based monitoring of civil infrastructure: challenges and successes. Journal of Civil Structural Health Monitoring, 1(3-4), 79-95.
[16] Shafer, G. (1976). A mathematical theory of evidence. Princeton: Princeton university press.
[17] Moore, R. E., & Bierbaum, F. (1979). Methods and applications of interval analysis. Philadelphia: Siam.
[18] Klir, G., & Yuan, B. (1995). Fuzzy sets and fuzzy logic. New Jersey: Prentice hall.
[19] Jordan, M. I. (2003). An introduction to probabilistic graphical model. Lecture notes, UC Berkeley.
[20] Roeck, G. D. (2003). The state-of-the-art of damage detection by vibration monitoring: the SIMCES experience. Journal of Structural Control, 10(2), 127-134.
[21] Guan, H. (2006). Vibration-based structural health monitoring of highway bridges. Report, California Department of Transportation, Sacramento, CA.
[22] Chopra, A. K. (2012) Dynamics of Structures: Theory and Applications to Earthquake Engineering (4th Ed), Englewood Cliffs, NJ: Prentice Hall.
[23] Sohn, H., Farrar, C. R., Hemez, F. M., Shunk, D. D., Stinemates, D. W., Nadler, B. R., & Czarnecki, J. J. (2003). A review of structural health monitoring literature: 1996–2001. Los Alamos National Laboratory, USA.
[24] Carden, E. P., & Fanning, P. (2004). Vibration based condition monitoring: a review. Structural health monitoring, 3(4), 355-377.
[25] Kammer, D. C. (1991). Sensor placement for on-orbit modal identification and correlation of large space structures. Journal of Guidance, Control, and Dynamics, 14(2), 251-259.
[26] Udwadia, F. E. (1994). Methodology for optimum sensor locations for parameter identification in dynamic systems. Journal of Engineering Mechanics, 120(2), 368-390.
[27] Borguet, S., & Léonard, O. (2008). The Fisher information matrix as a relevant tool for sensor selection in engine health monitoring. International Journal of Rotating Machinery, 2008.
[28] Papadimitriou, C. (2004). Optimal sensor placement methodology for parametric identification of structural systems. Journal of sound and vibration, 278(4), 923-947.
[29] Papadimitriou, C., & Lombaert, G. (2012). The effect of prediction error correlation on optimal sensor placement in structural dynamics. Mechanical Systems and Signal Processing, 28, 105-127.
[30] Trendafilova, I., Heylen, W., & Van Brussel, H. (2001). Measurement point selection in damage detection using the mutual information concept. Smart materials and structures, 10(3), 528-533.
[31] Li, B., & Ou, J. (2013). Optimal sensor placement for structural health

monitoring based on KL divergence. in Proceeding of Safety, Reliability, Risk and Life-Cycle Performance of Structures and Infrastructures, New York, USA, 2013.
[32] Li, B., (2012). Information theoretic optimal sensor placement in structural health monitoring. Master thesis, Dalian University of Technology, Dalian, China.
[33] Papadopoulos, M., & Garcia, E. (1998). Sensor placement methodologies for dynamic testing. AIAA journal, 36(2), 256-263.
[34] Debnath, N., Dutta, A., & Deb, S. K. (2012). Placement of sensors in operational modal analysis for truss bridges. Mechanical Systems and Signal Processing, 31, 196-216.
[35] Penny, J., Friswell, M., Garvey, S. (1994). Automatic choice of measurement locations for modal survey test, AIAA Journal, 32, 407-414.
[36] Wilson, J. A., & Guhe, S. Y. (2005). Observability matrix condition number in design of measurement strategies. Computer Aided Chemical Engineering, 20, 397-402.
[37] Li, D. S., Zhang, Y., Ren, L., & Li, H. N. (2011). Sensor deployment for structural health monitoring and their evaluation. Advances in Mechanics, 41(1), 39-50.
[38] Li, D. S., Li, H. N., & Fritzen, C. P. (2012). Load dependent sensor placement method: theory and experimental validation. Mechanical Systems and Signal Processing, 31, 217-227.
[39] Brehm, M., Zabel, V., & Bucher, C. (2013). Optimal reference sensor positions using output-only vibration test data. Mechanical systems and signal processing, 41(1), 196-225.
[40] Vinot, P., Cogan, S., & Cipolla, V. (2005). A robust model-based test planning procedure. Journal of Sound and Vibration, 288(3), 571-585.
[41] Castro-Triguero, R., Murugan, S., Gallego, R., & Friswell, M. I. (2013). Robustness of optimal sensor placement under parametric uncertainty. Mechanical Systems and Signal Processing, 41(1), 268-287.
[42] Heredia-Zavoni, E., & Esteva, L. (1998). Optimal instrumentation of uncertain structural systems subject to earthquake ground motions. Earthquake engineering & structural dynamics, 27(4), 343-362.
[43] Yuen, K. V., Katafygiotis, L. S., Papadimitriou, C., & Mickleborough, N. C. (2001). Optimal sensor placement methodology for identification with unmeasured excitation. Journal of dynamic systems, measurement, and control, 123(4), 677-686.
[44] Flynn, E. B., & Todd, M. D. (2010). A Bayesian approach to optimal sensor placement for structural health monitoring with application to active sensing. Mechanical Systems and Signal Processing, 24(4), 891-903.
[45] Reynders, E. (2012). System identification methods for (operational) modal analysis: review and comparison. Archives of Computational Methods in Engineering, 19(1), 51-124.

[46] James, G. H., Carne, T. G., Lauffer, J.P., & Nord, A. R. (1992). Modal testing using natural excitation, in Proceedings of 10th International Modal Analysis Conference, San Diego, USA.
[47] Van Overschee, P., & De Moor, B. (1993). Subspace algorithms for the stochastic identification problem. Automatica, 29(3), 649-660.
[48] Liung, L. (1999). System identification-Theory for the user 2nd ed. PTR Prentice Hall, Upper Saddle River, MJ.
[49] Brincker, R., Zhang, L., & Andersen P. (2000) Modal identification from ambient response using frequency domain decomposition, in Proceedings of IMAC18, San Antonio, USA.
[50] Guillaume, P., Verboven, P., Vanlanduit, S., Van der Auweraer, H., & Peeters, B. (2003). A Poly-reference implementation of the least-squares complex frequency domain estimator, in Proceedings of IMAC21, Kissimmee, USA.
[51] Ciloglu, K., Zhou, Y., Moon, F., & Aktan, A. E. (2012). Impacts of epistemic uncertainty in operational modal analysis. Journal of Engineering Mechanics, 138(9), 1059-1070.
[52] Lam, X. B., Mevel, L. (2011). Uncertainty quantification for eigensystem-realization-algorithm, a class of subspace system identification, in Proceeding of the 18th IFAC World Congress, Milan, Italy.
[53] Reynders, E., Pintelon, R., & De Roeck, G. (2008). Uncertainty bounds on modal parameters obtained from stochastic subspace identification. Mechanical Systems and Signal Processing, 22(4), 948-969.
[54] Dohler, M., Lam, X. B., & Mevel, L. (2011) Uncertainty quantification for stochastic subspace identification on multi-setup measurements, in Proceeding of 50th IEEE Conference on CDC-ECC, Orlando, USA.
[55] Vu, V. H., & Thomas, M. (2014). Uncertainties on modal parameters by operational modal analysis. Mechanics & Industry, 15(2), 153-158.
[56] Cara, F. J., Carpio, J., Juan, J., & Alarcón, E. (2012). An approach to operational modal analysis using the expectation maximization algorithm. Mechanical Systems and Signal Processing, 31, 109-129.
[57] El-Kafafy, M., De Troyer, T., & Guillaume, P. (2014). Fast maximum-likelihood identification of modal parameters with uncertainty intervals: a modal model formulation with enhanced residual term. Mechanical Systems and Signal Processing, 48(1), 49-66.
[58] Yuen, K. V., & Katafygiotis, L. S. (2001). Bayesian time–domain approach for modal updating using ambient data. Probabilistic Engineering Mechanics, 16(3), 219-231.
[59] Katafygiotis, L. S., & Yuen, K. V. (2001). Bayesian spectral density approach for modal updating using ambient data. Earthquake engineering & structural dynamics, 30(8), 1103-1123.

[60] Au, S. K. (2011). Fast Bayesian FFT method for ambient modal identification with separated modes. Journal of Engineering Mechanics, 137(3), 214-226.
[61] Au, S. K. (2012). Fast Bayesian ambient modal identification in the frequency domain, Part II: Posterior uncertainty. Mechanical Systems and Signal Processing, 26, 76-90.
[62] Au, S. K., & Zhang, F. L. (2016). Fundamental two-stage formulation for Bayesian system identification, Part I: General theory. Mechanical Systems and Signal Processing, 66, 31-42.
[63] Berman, A. (1979). Mass matrix correction using an incomplete set of measured modes. AIAA Journal, 17(10), 1147-1148.
[64] Yuen, K. V. (2009). Efficient model correction method with modal measurement. Journal of Engineering Mechanics, 136(1), 91-99.
[65] Friswell, M., & Mottershead, J. E. (1995). Finite element model updating in structural dynamics (Vol. 38). Springer Science & Business Media.
[66] Bakir, P. G., Reynders, E., & De Roeck, G. (2007). Sensitivity-based finite element model updating using constrained optimization with a trust region algorithm. Journal of Sound and Vibration, 305(1), 211-225.
[67] Nocedal, J., & Wright, S. (2006). Numerical optimization. Springer Science & Business Media.
[68] Simoen, E., De Roeck, G., & Lombaert, G. (2015). Dealing with uncertainty in model updating for damage assessment: A review. Mechanical Systems and Signal Processing, 56, 123-149.
[69] Soize, C., Capiez-Lernout, E., & Ohayon, R. (2008). Robust updating of uncertain computational models using experimental modal analysis. AIAA journal, 46(11), 2955-2965.
[70] Govers, Y., & Link, M. (2010). Stochastic model updating—Covariance matrix adjustment from uncertain experimental modal data. Mechanical Systems and Signal Processing, 24(3), 696-706.
[71] Khodaparast, H. H., Mottershead, J. E., & Friswell, M. I. (2008). Perturbation methods for the estimation of parameter variability in stochastic model updating. Mechanical Systems and Signal Processing, 22(8), 1751-1773.
[72] Husain, N. A., Khodaparast, H. H., & Ouyang, H. (2012). Parameter selection and stochastic model updating using perturbation methods with parameter weighting matrix assignment. Mechanical Systems and Signal Processing, 32, 135-152.
[73] Steenackers, G., & Guillaume, P. (2006). Finite element model updating taking into account the uncertainty on the modal parameters estimates. Journal of Sound and Vibration, 296(4), 919-934.
[74] Vanik, M. W., Beck, J. L., & Au, S. (2000). Bayesian probabilistic approach to structural health monitoring. Journal of Engineering Mechanics, 126(7), 738-745.
[75] Ching, J., & Beck, J. L. (2004). New Bayesian model updating algorithm applied

to a structural health monitoring benchmark. Structural Health Monitoring, 3(4), 313-332.
[76] Ching, J., Muto, M., & Beck, J. L. (2006). Structural model updating and health monitoring with incomplete modal data using Gibbs sampler. Computer-Aided Civil and Infrastructure Engineering, 21(4), 242-257.
[77] Ching, J., & Chen, Y. C. (2007). Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging. Journal of engineering mechanics, 133(7), 816-832.
[78] Yuen, K. V., Beck, J. L., & Katafygiotis, L. S. (2006). Efficient model updating and health monitoring methodology using incomplete modal data without mode matching. Structural Control and Health Monitoring, 13(1), 91-107.
[79] Yan, W. J., & Katafygiotis, L. S. (2015). A novel Bayesian approach for structural model updating utilizing statistical modal information from multiple setups. Structural Safety, 52, 260-271.
[80] Sun, H., & Büyüköztürk, O. (2016). Probabilistic updating of building models using incomplete modal data. Mechanical Systems and Signal Processing, 75, 27-40.
[81] Behmanesh, I., Moaveni, B., Lombaert, G., & Papadimitriou, C. (2015). Hierarchical Bayesian model updating for structural identification. Mechanical Systems and Signal Processing, 64, 360-376.
[82] Freedman, D. (1995). Some issues in the foundation of statistics. Foundations of Science, 1(1), 19-39.
[83] Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). Bayesian data analysis. Boca Raton, FL, USA: Chapman & Hall/CRC.
[84] Kass, R. E., & Raftery, A. E. (1995). Bayes factors. Journal of the American Statistical Association, 90(430), 773-795.
[85] Thompson, C. J. (2015). Mathematical statistical mechanics. Princeton University Press.
[86] Tierney, L., & Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. Journal of the American Statistical Association, 81(393), 82-86.
[87] Ghosh, J. K., Delampady, M., & Samanta, T. (2007). An introduction to Bayesian analysis: theory and methods. Springer Science & Business Media.
[88] Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (methodological), 1-38.
[89] Wu, C. J. (1983). On the convergence properties of the EM algorithm. The Annals of Statistics, 95-103.
[90] Sundberg, R. (1974). Maximum likelihood theory for incomplete data from an exponential family. Scandinavian Journal of Statistics, 49-58.
[91] Meng, X. L., & Rubin, D. B. (1991). Using EM to obtain asymptotic

variance-covariance matrices: The SEM algorithm. Journal of the American Statistical Association, 86(416), 899-909.
[92] Petersen, K. B., Winther, O., & Hansen, L. K. (2005). On the slow convergence of EM and VBEM in low-noise linear models. Neural Computation, 17(9), 1921-1926.
[93] McLachlan, G., & Krishnan, T. (2007). The EM algorithm and extensions. John Wiley & Sons.
[94] Beal, M. J. (2003). Variational algorithms for approximate Bayesian inference. Ph.D. dissertation, University of London, London, UK.
[95] Wainwright, M. J., & Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. Foundations and Trends® in Machine Learning, 1(1-2), 1-305.
[96] Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. The Annals of Mathematical Statistics, 22(1), 79-86.
[97] Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2016). Variational inference: a review for statisticians. arXiv preprint arXiv:1601.00670.
[98] Wang, B., & Titterington, D. M. (2004). Convergence and asymptotic normality of variational Bayesian approximations for exponential family models with missing values. In Proceedings of the 20th conference on Uncertainty in artificial intelligence.
[99] Giordano, R. J., Broderick, T., & Jordan, M. I. (2015). Linear response methods for accurate covariance estimates from mean field variational Bayes. In Advances in Neural Information Processing Systems (pp. 1441-1449).
[100] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., & Teller, E. (1953). Equations of state calculations by fast computing machine. Journal of Chemistry and Physics, 21, 1087-1091.
[101] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. Biometrika, 57(1), 97-109.
[102] Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence, (6), 721-741.
[103] Gelfand, A. E., & Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association, 85(410), 398-409.
[104] Robert, C., & Casella, G. (2004). Monte Carlo statistical methods. Springer-Verlag. New York.
[105] Rosenthal, J. S. (2011). Optimal proposal distributions and adaptive MCMC. Handbook of Markov Chain Monte Carlo, 93-112.
[106] Andrieu, C., De Freitas, N., Doucet, A., & Jordan, M. I. (2003). An introduction to MCMC for machine learning. Machine Learning, 50(1-2), 5-43.
[107] Nielsen, T. D., & Jensen, F. V. (2009). Bayesian networks and decision graphs.

Springer Science & Business Media.
[108] Jaynes, E. T. (1957). Information theory and statistical mechanics. Physical review, 106(4), 620.
[109] Press, W. H. (2007). Numerical recipes: The art of scientific computing (3rd Edition). Cambridge university press. New York, USA.
[110] Von Neumann, J., & Morgenstern, O. (2007). Theory of games and economic behavior. Princeton university press. Princeton, USA.
[111] DasGupta, A., & Studden, W. J. (1991). Robust Bayesian experimental designs in normal linear models. The Annals of Statistics, 1244-1256.
[112] Cover, T. M., & Thomas, J. A. (2006). Elements of information theory (2nd Edition). John Wiley & Sons. Hoboken, USA.
[113] Bernardo, J. M. (1979). Expected information as expected utility. The Annals of Statistics, 686-690.
[114] Tiao, G. C., & Afonja, B. (1976). Some Bayesian considerations of the choice of design for ranking, selection and estimation. Annals of the Institute of Statistical Mathematics, 28(1), 167-185.
[115] Eaton, M. L., Giovagnoli, A., & Sebastiani, P. (1996). A predictive approach to the Bayesian design problem with application to normal regression models. Biometrika, 83(1), 111-125.
[116] Malings, C., & Pozzi, M. (2015). Sensor network optimization using Bayesian networks, decision graphs, and value of information. in: Proceeding of 12th International Conference on Applications of Statistics and Probability in Civil Engineering, Vancouver, Canada.
[117] Der Kiureghian, A. (1980). Structural response to stationary excitation. Journal of the Engineering Mechanics Division, 106(6), 1195-1213.
[118] Harichandran, R. S. (1992). Random vibration under propagating excitation: closed-form solutions. Journal of Engineering mechanics, 118(3), 575-586.
[119] Chan, T. F., & Olkin, J. A. (1994). Circulant preconditioners for Toeplitz-block matrices. Numerical Algorithms, 6(1), 89-101.
[120] Van Loan, C. F. (2000). The ubiquitous Kronecker product. Journal of computational and applied mathematics, 123(1), 85-100.
[121] Steeb, W. H., & Shi, T. K. (1997). Matrix calculus and Kronecker product with applications and C++ programs. World Scientific. Singapore.
[122] Genton, M. G. (2007). Separable approximations of space-time covariance matrices. Environmetrics, 18(7), 681-695.
[123] Load Code for the Design of Building Structures (2012), China National Standard, GB 50009-2012, China Architecture and Building Press, Beijing.
[124] Ni, Y. Q., Xia, Y., Liao, W. Y., & Ko, J. M. (2009). Technology innovation in developing the structural health monitoring system for Guangzhou New TV Tower. Structural Control and Health Monitoring, 16(1), 73-98.

[125] Yi, T. H., Li, H. N., & Gu, M. (2011). Optimal sensor placement for structural health monitoring based on multiple optimization strategies. The Structural Design of Tall and Special Buildings, 20(7), 881-900.
[126] Adhikari, S. (2008). Wishart random matrices in probabilistic structural mechanics. Journal of engineering mechanics, 134(12), 1029-1044.
[127] Franklin, G. F., Powell, J.D., & Workman, M. L. (1997). Digital Control of Dynamic Systems, 3rd ed., Ellis-Kagle Press. Half Moon Bay, USA.
[128] Gupta, A. K., & Nagar, D. K. (1999). Matrix variate distributions. CRC Press. Boca Raton, USA.
[129] Fujimoto, K., Satoh, A., & Fukunaga, S. (2011). System identification based on variational Bayes method and the invariance under coordinate transformations, in: Proceeding of 50th IEEE Conference on CDC-ECC, Orlando, USA.
[130] Gibson, S., & Ninness, B. (2005). Robust maximum-likelihood estimation of multivariable dynamic systems. Automatica, 41(10), 1667-1682.
[131] Chiappa, S., & Barber, D. (2007). Unified Inference for Variational Bayesian Linear Gaussian State-Space Models. In Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference.
[132] Adhikari, S., & Friswell, M. I. (2007). Random matrix eigenvalue problems in structural dynamics. International Journal for Numerical Methods in Engineering, 69(3), 562-591.
[133] Shumway, R. H., & Stoffer, D. S. (2010). Time series analysis and its applications: with R examples. Springer Science & Business Media.
[134] Cara, F. J., Juan, J., Alarcón, E., Reynders, E., & De Roeck, G. (2013). Modal contribution and state space order selection in operational modal analysis. Mechanical Systems and Signal Processing, 38(2), 276-298.
[135] Hsu, T. Y., Huang, S. K., Lu, K. C., Loh, C. H., Wang, Y., & Lynch, J. P. (2011). On-line structural damage localization and quantification using wireless sensors. Smart Materials and Structures, 20(10), 105025.
[136] Kim, J., & Lynch, J. P. (2012). Subspace system identification of support-excited structures—part I: theory and black-box system identification. Earthquake engineering & structural dynamics, 41(15), 2235-2251.
[137] Çelebi, M., Huang, M., Shakal, A., Hooper, J., & Klemencic, R. (2013). Ambient response of a unique performance-based design tall building with dynamic response modification features. The Structural Design of Tall and Special Buildings, 22(10), 816-829.
[138] Vlach, J., & Singhal, K. (1983). Computer methods for circuit analysis and design. Springer Science & Business Media. New York, USA.
[139] Sohn, H., Dzwonczyk, M., Straser, E. G., Kiremidjian, A. S., Law, K. H., & Meng, T. (1999). An experimental study of temperature effect on modal parameters of the Alamosa Canyon Bridge. Earthquake engineering & structural dynamics, 28(8),

879-897.
[140] Peeters, B., Maeck, J., & De Roeck, G. (2001). Vibration-based damage detection in civil engineering: excitation sources and temperature effects. Smart materials and Structures, 10(3), 518.
[141] Pintelon, R., & Kollár, I. (2005). On the frequency scaling in continuous-time modeling. IEEE Transactions on Instrumentation and Measurement, 54(1), 318-321.
[142] Bellman R. (1997). Introduction to matrix analysis. Society for Industrial and Applied Mathematics.
[143] Andrieu, C., De Freitas, N., Doucet, A., & Jordan, M. I. (2003). An introduction to MCMC for machine learning. Machine learning, 50(1-2), 5-43.
[144] Adams, R. P., & MacKay, D. J. (2007). Bayesian online changepoint detection. arXiv preprint arXiv:0710.3742.
[145] Ghahramani, Z., & Hinton, G. E. (1996). Switching state-space models. Technical Report CRG-TR-96-3 DRAFT, University of Toronto. Toronto, CA.
[146] Li, B., & Der Kiureghian, A. (2016). Accelerating subset simulation with a surrogate model. IFIP Working Group 7.5 on Reliability and Optimization of Structural Systems, Pittsburgh, USA.
[147] Straub, D. (2009). Stochastic modeling of deterioration processes through dynamic Bayesian networks. Journal of Engineering Mechanics, 135(10), 1089-1099.
[148] Nielsen, T. D., & Jensen, F. V. (2009). Bayesian networks and decision graphs. Springer Science & Business Media.