Posterior Uncertainty, Asymptotic Law and Cramér-Rao Bound

Siu-Kui Au [a], Binbin Li [b*]

[a]Institute for Risk & Uncertainty and Center for Engineering Dynamics, University of Liverpool, Liverpool, L69 3GH, UK. Email: siukuiau@liverpool.ac.uk
[b]Institute for Risk & Uncertainty and Center for Engineering Dynamics, University of Liverpool, Liverpool, L69 3GH, UK. Email: bbl@liverpool.ac.uk

Abstract: In a globally identifiable Bayesian system identification problem, the uncertainty of model parameters can be quantified by their 'posterior covariance matrix' calculated for a particular data set. When the data is modeled to be distributed as the likelihood function (i.e., no modeling error), a statistical law analogous to the Law of Large Numbers results, where the posterior covariance matrix is asymptotic to a deterministic quantity that depends on the 'information content' of data rather than its particular (stochastic) details. This was referred as the 'uncertainty law' in a recent study of the achievable precision of modal parameters in operational modal analysis. Deriving the uncertainty law involves asymptotics techniques and leveraging on the mathematical structure of the likelihood function, which was found to be tedious. As a sequel to the development, this work shows that for large data size and up to a Gaussian approximation of the posterior distribution, the uncertainty law is asymptotic to the inverse of the Fisher information matrix, which coincides with the tightest Cramér-Rao Bound in classical statistics. A parametric study is presented to illustrate the theoretical results in the context of operational modal analysis. As a direct application with practical relevance, the relationship provides a systematic means for deriving the uncertainty laws in operational modal analysis. It can also be applied in general Bayesian system identification problems.

Key words: Uncertainty law; Bayesian system identification; Fisher information; Operational modal analysis

1 Introduction

Bayesian system identification is concerned with making inference about some set of model parameters $\boldsymbol{\theta}$ based on measured data $\boldsymbol{z}$ in the context of a mathematical model that relates $\boldsymbol{\theta}$ and $\boldsymbol{z}$. Without much loss of generality, assume that $\boldsymbol{\theta}$ and $\boldsymbol{z}$ are continuous-valued. The identification result about $\boldsymbol{\theta}$ is encapsulated by its 'posterior' (i.e., given data) probability density function (PDF) $p(\boldsymbol{\theta}|\boldsymbol{z})$. Using Bayes theorem, it is given by

$$p(\boldsymbol{\theta}|\boldsymbol{z}) = \frac{1}{p(\boldsymbol{z})} p(\boldsymbol{z}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) \tag{1}$$

* Corresponding author

where $p(\boldsymbol{z}|\boldsymbol{\theta})$ is called the 'likelihood function', giving the modeled PDF of $\boldsymbol{z}$ for a given $\boldsymbol{\theta}$; $p(\boldsymbol{\theta})$ is called the 'prior distribution', reflecting one's knowledge about $\boldsymbol{\theta}$ in the absence of data; $p(\boldsymbol{z})$ is a normalizing constant because it does not depend on $\boldsymbol{\theta}$. All PDFs here are conditional on the assumed model but this has been omitted in notation for simplicity.

The identification uncertainty of $\boldsymbol{\theta}$ can be quantified by its covariance matrix associated with the posterior PDF. In globally identifiable problems [1–3], the posterior PDF has a unique regular maximum at the 'most probable value' (MPV) of parameters, or equivalently, a unique regular minimum of the function $-\ln[p(\boldsymbol{z}|\boldsymbol{\theta})p(\boldsymbol{\theta})]$, referred as 'NF' here. A second order Taylor approximation of the NL leads to a Gaussian approximation of the posterior PDF. This approximation can be justified by the centralized shape of the posterior PDF. Under some regularity conditions [4], it becomes asymptotically correct for large data size, which has been investigated in the mathematics of 'Laplace type integrals' [5–7]. Under the Gaussian approximation, the posterior covariance matrix is equal to the inverse of the Hessian of NF at the MPV. When the system is not globally identifiable, the approximation is not applicable and one may resort to more general tools such as Monte Carlo methods [8–10].

Beyond quantification, managing identification uncertainty requires one to understand how it depends on the test configuration. This has both scientific significance in discovering identification precision limits and practical significance in experimental design and budgeting. Typically, the MPV and covariance matrix corresponding to the posterior PDF are calculated in a Bayesian identification algorithm for a given data set. The value of the covariance matrix reflects the identification uncertainty of $\boldsymbol{\theta}$ for the given data set, but it does not offer any insight on how the identification uncertainty depends on the test configuration that leads to the data. A single set of result does not serve the purpose of uncertainty management. Investigating the relationship through a parametric study with data sets of different configurations could be one option. This is only empirical, however, not to mention that the effect of test configuration has yet to be quantified.

It turns out that when the distribution of data is assumed to follow the likelihood function, i.e., no modeling error, a statistical law analogous to the Law of Large Numbers (LLN) in classical probability results. The posterior covariance matrix is asymptotic to a deterministic quantity that depends on the 'information content' of data rather than its particular details. This was referred as the 'uncertainty law' in a recent study of the achievable precision of modal parameters in operational modal analysis [11,12]. Deriving the uncertainty law was found to be tedious, involving asymptotics techniques and leveraging on the mathematical structure of the likelihood function. As a sequel to the development, this work shows that for large data size the uncertainty law is asymptotic to the inverse of the Fisher information matrix (FIM), which coincides with the Cramér-Rao Bound (CRB) in classical statistics [13,14]. This connection in the mathematics between Bayesian and frequentist statistics can be regarded as a continuation of the discussion in [15], which attempted to bridge the two definitions of identification uncertainty in a limited heuristic context. From a utility perspective, the connection established in this work allows the uncertainty laws for other unexplored cases (e.g., close modes, multiple setups) to be derived more systematically and effectively.

This paper is outlined as follow. Bayesian system identification is first reviewed based on Gaussian approximation for globally identifiable problems. The asymptotic equivalence of the posterior covariance matrix and the inverse of FIM is then established, assuming no modeling error and sufficiently long data. As an illustration, the long-data asymptotic expressions of the Hessian of NL obtained in [6] are re-derived by taking advantage of the CRB, showing the utility of the asymptotic equivalence. Finally, empirical studies based synthetic and experimental data are presented to further validate the asymptotic equivalence and the quality of approximation under non-asymptotic situations.

2 Bayesian System Identification

Consider a Bayesian system identification problem with a sufficiently large amount of data $\boldsymbol{z} = \{\boldsymbol{z}_1, \boldsymbol{z}_2, \dots, \boldsymbol{z}_N\}$, where theoretically $N \to \infty$. For instructional purpose, assume that $\{\boldsymbol{z}_1, \boldsymbol{z}_2, \dots, \boldsymbol{z}_N\}$ are independent and identically distributed (i.i.d.) for given $\boldsymbol{\theta}$; comments for the general case shall be made later. The likelihood function is given by

$$p(\boldsymbol{z}|\boldsymbol{\theta}) = \prod_{k=1}^{N} p(\boldsymbol{z}_k|\boldsymbol{\theta}) \tag{2}$$

For large data size, the prior PDF $p(\boldsymbol{\theta})$ is a slowly varying function of $\boldsymbol{\theta}$ compared to the likelihood function $p(\boldsymbol{z}|\boldsymbol{\theta})$. The posterior PDF is then insensitive to the choice of the prior PDF. The latter can be practically taken as a constant over the parameter domain, i.e., a uniform distribution, so that the posterior PDF is directly proportional to the likelihood function. Consequently, one has

$$p(\boldsymbol{\theta}|\boldsymbol{z}) \propto e^{-L(\boldsymbol{\theta},\boldsymbol{z})} \tag{3}$$

where

$$L(\boldsymbol{\theta}, \boldsymbol{z}) = -\ln p(\boldsymbol{z}|\boldsymbol{\theta}) = -\sum_{k=1}^{N} \ln p(\boldsymbol{z}_k|\boldsymbol{\theta}) \tag{4}$$

is the NL. Assuming that the problem is globally identifiable, the MPV $\widehat{\boldsymbol{\theta}}(\boldsymbol{z})$ (say) uniquely minimizes $L(\boldsymbol{\theta}, \boldsymbol{z})$ with respect to (w.r.t.) $\boldsymbol{\theta}$, where its dependence on $\boldsymbol{z}$ has been emphasized. At the MPV $\widehat{\boldsymbol{\theta}}(\boldsymbol{z})$, the gradient of $L(\boldsymbol{\theta}, \boldsymbol{z})$ w.r.t. $\boldsymbol{\theta}$ is a zero vector and the Hessian matrix w.r.t. $\boldsymbol{\theta}$ is a positive definite matrix, i.e.,

$$\begin{aligned} \nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \boldsymbol{z})|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\boldsymbol{z})} &= \boldsymbol{0} \\ \nabla^2_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \boldsymbol{z})\big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\boldsymbol{z})} &> \boldsymbol{0} \end{aligned} \tag{5}$$

where the '$>$' for a matrix denotes that it is a positive definite matrix and the subscript $\boldsymbol{\theta}$ in ∇ denotes that the gradient is taken w.r.t. $\boldsymbol{\theta}$. Since $\{\boldsymbol{z}_1, \boldsymbol{z}_2, \dots, \boldsymbol{z}_N\}$ are independent,

$$\begin{aligned} \nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \boldsymbol{z}) &= -\textstyle\sum_{k=1}^{N} \nabla_{\boldsymbol{\theta}} \ln p(\boldsymbol{z}_k|\boldsymbol{\theta}) \\ \nabla^2_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \boldsymbol{z}) &= -\textstyle\sum_{k=1}^{N} \nabla^2_{\boldsymbol{\theta}} \ln p(\boldsymbol{z}_k|\boldsymbol{\theta}) \end{aligned} \tag{6}$$

Under a second order approximation of the NL about the MPV, the posterior PDF is approximated by a Gaussian PDF with mean $\widehat{\boldsymbol{\theta}}(\boldsymbol{z})$ and covariance matrix

$$\widehat{\boldsymbol{C}}(\boldsymbol{z}) = \left[\nabla_{\boldsymbol{\theta}}^2 L\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{z}\right)\right]^{-1} \tag{7}$$

where the dependence on $\boldsymbol{z}$ has been emphasized; and $\nabla_{\boldsymbol{\theta}}^2 L\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{z}\right)$ denotes the Hessian of NL w.r.t. $\boldsymbol{\theta}$ and evaluated at $\widehat{\boldsymbol{\theta}}(\boldsymbol{z})$. This approximation is simple and elegant. All one needs is the MPV and the Hessian at that location. One does not need to explore the whole posterior PDF since a Gaussian PDF is completely defined by the mean vector and the covariance matrix. The quality of Gaussian approximation to the posterior PDF $p(\boldsymbol{\theta}|\boldsymbol{z})$ depends on how close $L(\boldsymbol{\theta}, \boldsymbol{z})$ is to a quadratic function of $\boldsymbol{\theta}$. Under some regularity conditions [4], the posterior PDF is asymptotically Gaussian, in which case the Gaussian approximation is equivalent to the Laplace asymptotic approximation.

3 Long-data asymptotic behavior of posterior covariance matrix

Assume no modeling error, so that the data $\boldsymbol{z}$ is indeed distributed as the modeled likelihood function $p(\boldsymbol{z}|\boldsymbol{\theta}_0)$ with some 'actual parameter' value $\boldsymbol{\theta}_0$. The existence of $\boldsymbol{\theta}_0$ is a frequentist concept introduced here to bridge the mathematics between Bayesian and frequentist statistics. We shall argue that, for large N,

$$\widehat{\boldsymbol{C}}(\boldsymbol{z}) = \boldsymbol{J}(\boldsymbol{\theta}_0)^{-1}\left[\boldsymbol{I} + O_{\boldsymbol{z}}(N^{-1/2})\right] \tag{8}$$

where $\boldsymbol{I}$ denotes the identity matrix; the term $O_{\boldsymbol{z}}(N^{-1/2})$ denotes that the remainder in the bracket depends on $\boldsymbol{z}$ and is of the order $N^{-1/2}$;

$$\boldsymbol{J}(\boldsymbol{\theta}_0) = \mathrm{E}[\nabla_{\boldsymbol{\theta}}^2 L(\boldsymbol{\theta}_0, \boldsymbol{z})|\boldsymbol{\theta}_0] = \int \nabla_{\boldsymbol{\theta}}^2 L(\boldsymbol{\theta}_0, \boldsymbol{z}) p(\boldsymbol{z}|\boldsymbol{\theta}_0) d\boldsymbol{z} \tag{9}$$

is the 'Fisher information matrix' (FIM) for $\boldsymbol{\theta}$; $\mathrm{E}[\cdot\,|\boldsymbol{\theta}_0]$ denotes the expectation when $\boldsymbol{z}$ is distributed as $p(\boldsymbol{z}|\boldsymbol{\theta}_0)$; and

$$\nabla_{\boldsymbol{\theta}}^2 L(\boldsymbol{\theta}_0, \boldsymbol{z}) = \nabla_{\boldsymbol{\theta}}^2 L(\boldsymbol{\theta}, \boldsymbol{z})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \tag{10}$$

The FIM is originally defined in a classical statistical, i.e., 'non-Bayesian' context. It indicates how difficult it is to estimate $\boldsymbol{\theta}_0$: parameters with greater information can be estimated more easily, requiring less data to achieve a required precision. The FIM has been used for experimental design; see, e.g., [16] in stochastic structural dynamics.

In terms of asymptotic relationship, Eqn. (8) reads

$$\widehat{\boldsymbol{C}} \sim \boldsymbol{J}(\boldsymbol{\theta}_0)^{-1}, \quad N \to \infty \tag{11}$$

As uncertainty law is defined as the leading order of the posterior covariance under some asymptotic conditions, the above shows that it is simply the inverse of the FIM for asymptotically large data size. In a frequentist context where $\boldsymbol{z}$ is assumed to be distributed as

$p(\boldsymbol{z}|\boldsymbol{\theta}_0)$, the inverse of the FIM is the Cramér-Rao bound (CRB), which says that any unbiased estimator $\boldsymbol{G}(\boldsymbol{z})$ for $\boldsymbol{\theta}_0$ will have a covariance matrix bounded by the following inequality:

$$\text{cov}[\boldsymbol{G}(\boldsymbol{z})|\boldsymbol{\theta}_0] \geq \boldsymbol{J}(\boldsymbol{\theta}_0)^{-1} \tag{12}$$

Here, $\text{cov}[\boldsymbol{G}(\boldsymbol{z})|\boldsymbol{\theta}_0]$ denotes the covariance matrix of $\boldsymbol{G}(\boldsymbol{z})$ when $\boldsymbol{z}$ is distributed as $p(\boldsymbol{z}|\boldsymbol{\theta}_0)$. Equation (12) is a matrix inequality that should be interpreted as the LHS minus the RHS being a positive semi-definite matrix. Note that the CRB here is the 'tightest' one involving all parameters in the identification problem. In statistical applications, looser (weaker) bounds with a lower-dimensional FIM involving only some of the parameters are often derived for mathematical tractability. See, e.g., [17] that derives the bounds for modal identification with free vibration data. The asymptotic result of the posterior covariance matrix bridges the mathematics between the Bayesian and non-Bayesian quantification of identification uncertainty, although their philosophies are different.

Equation (8) can be argued as follow. For a given $\boldsymbol{z}$, the MPV $\widehat{\boldsymbol{\theta}}(\boldsymbol{z})$ maximizes the likelihood function $p(\boldsymbol{z}|\boldsymbol{\theta})$ w.r.t. $\boldsymbol{\theta}$. It is then mathematically equivalent to the maximum likelihood estimator, which is known to be an asymptotically unbiased estimator of $\boldsymbol{\theta}_0$ [18]. In addition, reasoning from Eqn. (6), the posterior covariance matrix is $O_{\boldsymbol{z}}(N^{-1})$. One can then write

$$\widehat{\boldsymbol{\theta}}(\boldsymbol{z}) = \left[\boldsymbol{I} + O_{\boldsymbol{z}}(N^{-1/2})\right]\boldsymbol{\theta}_0 \tag{13}$$

Assuming that the NF has third order derivatives w.r.t. $\boldsymbol{\theta}$, we have

$$\nabla^2_{\boldsymbol{\theta}} L\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{z}\right) = \nabla^2_{\boldsymbol{\theta}} L(\boldsymbol{\theta}_0, \boldsymbol{z})\left[\boldsymbol{I} + O_{\boldsymbol{z}}(N^{-1/2})\right] \tag{14}$$

Since $\{\boldsymbol{z}_1, \boldsymbol{z}_2, \dots, \boldsymbol{z}_N\}$ are i.i.d, the second derivatives in the sum on the RHS of Eqn. (6) are i.i.d. as well. By the LLN, their average converges to the expectation of a single term:

$$\tfrac{1}{N}\nabla^2_{\boldsymbol{\theta}} L(\boldsymbol{\theta}_0, \boldsymbol{z}) \to \text{E}[\nabla^2_{\boldsymbol{\theta}} \ln p(\boldsymbol{z}_k|\boldsymbol{\theta}_0)\,|\boldsymbol{\theta}_0] = \boldsymbol{J}_k(\boldsymbol{\theta}_0) \tag{15}$$

where $\boldsymbol{J}_k$ is the FIM based on the sample $\boldsymbol{z}_k$. Again, $\{\boldsymbol{z}_1, \boldsymbol{z}_2, \dots, \boldsymbol{z}_N\}$ are i.i.d. and so $\boldsymbol{J}(\boldsymbol{\theta}_0) = N\boldsymbol{J}_k(\boldsymbol{\theta}_0)$. Combining these gives

$$\nabla^2_{\boldsymbol{\theta}} L\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{z}\right) = \boldsymbol{J}(\boldsymbol{\theta}_0)\left[\boldsymbol{I} + O_{\boldsymbol{z}}(N^{-1/2})\right] \tag{16}$$

Equation (8) then follows because matrix inverse is a continuous mapping.

An intuitive observation of the above result is that the uncertainty law is asymptotically the expectation of the posterior covariance, under the 'frequentist' context of 'identically repeated experiments'. When interpreting the uncertainty law, the data is unknown. For conceptual understanding or planning purpose, it is intuitive to integrate out the effect of stochastic data $\boldsymbol{z}$ to obtain an averaged value, i.e. to take expectation of $\widehat{\boldsymbol{C}}(\boldsymbol{z})$ with $\boldsymbol{z}$ distributed as $p(\boldsymbol{z}|\boldsymbol{\theta}_0)$ for some actual parameter value $\boldsymbol{\theta}_0$. Understanding the uncertainty law as the expectation of the posterior covariance provides an intuitive perspective to see their relationship. The fusion of Bayesian and frequentist concept here is adopted to bridge the two schools of thoughts. It should not be

confused with the Bayesian nature of the posterior covariance matrix, which depends on a particular data set and there is no notion of actual parameter or 'inherent uncertainty'.
The asymptotic behavior of posterior covariance developed above is based on models with i.i.d. observations. For the general case when the observations may be correlated or not identically distributed, the argument can be extended to hold by virtue of the general version of the LLN [19], but we do not pursue it here. Rather, we explore the utility of this asymptotic equivalence to derive the uncertainty laws in operational modal analysis.

4 Operational Modal Analysis

The mathematical equivalence between the uncertainty law and the inverse of FIM holds for general globally identifiable problems under the assumptions mentioned in the last section. Here, we focus on the uncertainty law of operational modal analysis (OMA), which aims at identifying the modal parameters (e.g. modal frequencies, damping ratios and mode shapes) of a structure using ambient vibration data [20–22]. The uncertainty law of OMA for well-separated modes was derived in [11] from first principle. The asymptotic conditions involve long data and small damping. Substituting the stochastic representation of the scaled fast Fourier transform (FFT) of data into the second derivatives of NF leads to an expression comprising sums of random terms, whose leading order is its expectation for long data. The derivation was found to be tedious, although the final expressions were remarkably simple and intuitive. Since OMA is a globally identifiable problem, the theory in the last section can be applied to allow for a more systematic derivation of the uncertainty laws for long data via the FIM. For this purpose, we re-derive the uncertainty law for long data. The objective is to illustrate how the asymptotic equivalence can lead to a systematic and effective derivation for future exploration of uncertainty laws for other cases, e.g., multiple modes and multiple setups.

4.1 Problem statement

Let the acceleration time history data measured at n degrees of freedom (DoFs) of a structure under ambient vibration be $\{\hat{\ddot{\boldsymbol{y}}}_j \in \mathbb{R}^n : j = 0,1,\dots,N-1\}$ and abbreviated as $\{\hat{\ddot{\boldsymbol{y}}}_j\}$, where N is the number of samples per data channel. Modeled as a stationary stochastic process, its scaled FFT is defined as

$$\hat{\boldsymbol{\mathcal{F}}}_k = \sqrt{\Delta t/N} \sum_{j=0}^{N-1} \hat{\ddot{\boldsymbol{y}}}_j e^{-2\pi \mathbf{i} jk/N} \tag{17}$$

where Δt (sec) is the sampling period and $\mathbf{i}$ is the imaginary unit. For $k \le N_q$, $\hat{\boldsymbol{\mathcal{F}}}_k$ corresponds to the frequency $\mathrm{f}_k = k/N\Delta t$ (Hz), where $N_q = \mathrm{int}[N/2] + 1$ (int[·] denotes the integer part) is the index at the Nyquist frequency. The FFT in Eqn. (17) is scaled by the factor $\sqrt{\Delta t/N}$ so that the expectation of $\hat{\boldsymbol{\mathcal{F}}}_k \hat{\boldsymbol{\mathcal{F}}}_k^*$ ('*' denotes conjugate transpose) is equal to the power spectral density (PSD) matrix of the data process.
In practice, only the FFTs on a selected frequency band, denoted by $\{\hat{\boldsymbol{\mathcal{F}}}_k\}$, containing the mode(s) of interest are used for identification. This trades off between the information used for identification (the wider the better) and modeling error risk (the narrower the better) [23]. Within the selected band, it is assumed that

$$\widehat{\boldsymbol{\mathcal{F}}}_k = \boldsymbol{\mathcal{F}}_k + \boldsymbol{\varepsilon}_k \tag{18}$$

where $\boldsymbol{\mathcal{F}}_k$ and $\boldsymbol{\varepsilon}_k$ denote respectively the scaled FFT of the theoretical structural dynamic response and the prediction error (e.g., data noise). For well-separated modes, one can select a frequency band dominated by only one vibration mode such that

$$\boldsymbol{\mathcal{F}}_k = \boldsymbol{\varphi} h_k p_k \tag{19}$$

where $\boldsymbol{\varphi} \in \mathbb{R}^n$ is the partial mode shape confined to measured DoFs; $p_k \in \mathbb{C}$ is the scaled FFT of the modal excitation; $h_k \in \mathbb{C}$ is the transfer function given by

$$h_k = \frac{1}{(1-\beta_k^2) - \mathbf{i}(2\zeta\beta_k)} \tag{20}$$

ζ is the damping ratio; $\beta_k = f/\mathrm{f}_k$ is the ratio of the natural frequency f to the FFT frequency f_k. In OMA, the modal excitation p_k is not measured and it is modeled statistically. Assuming zero-mean stationary modal excitation, p_k has a complex Gaussian distribution. Its variance (i.e., PSD of the process) is assumed to be a constant S in the selected band. The prediction errors at different channels are also assumed to be complex Gaussian distributed with zero mean and variance S_e within the selected band, which can be justified based on the principle of maximum entropy [1]. Further assuming statistical independence between the modal excitation and prediction errors yields a jointly independent complex Gaussian distribution for $\{\widehat{\boldsymbol{\mathcal{F}}}_k\}$ with zero mean and each with a covariance matrix (or theoretical PSD of data) given by

$$\boldsymbol{E}_k = SD_k \overline{\boldsymbol{\varphi}}\overline{\boldsymbol{\varphi}}^{\mathrm{T}} + S_e \boldsymbol{I}_n \tag{21}$$

where $\boldsymbol{I}_n \in \mathbb{R}^{n\times n}$ denotes the identity matrix;

$$D_k = \frac{1}{(1-\beta_k^2)^2 + (2\zeta\beta_k)^2} \tag{22}$$

is a dynamic amplification factor; and

$$\overline{\boldsymbol{\varphi}} = \|\boldsymbol{\varphi}\|^{-1}\boldsymbol{\varphi} \tag{23}$$

is the normalized mode shape, where $\|\cdot\|$ denotes the Euclidean norm, i.e., square root of the sum of squares.

The complex Gaussian distribution for $\{\widehat{\boldsymbol{\mathcal{F}}}_k\}$ is supported by the theoretical result that the scaled FFTs of a stationary process are asymptotically independent complex Gaussian distributed for long data [24]. As the data in OMA is stationary and its size is usually large enough, this assumption holds in general. The above frequency-domain model only makes use of the FFT information in the selected band, which significantly simplifies the identification model. The PSD of the modal excitation and prediction error need only be flat within the selected band, relaxing the conventional white noise assumption and making the method more robust than time-domain methods. Other bands with irrelevant information or which are difficult to model are legitimately ignored, therefore avoiding modeling error. This does not require any signal pre-processing such as filtering or averaging.

In summary, the problem of OMA is to identify parameters $\boldsymbol{\theta} = \{f, \zeta, \boldsymbol{\varphi}, S, S_e\}$ given the FFTs $\{\widehat{\boldsymbol{\mathcal{F}}}_k\}$. The likelihood function is

$$p\big(\{\widehat{\boldsymbol{\mathcal{F}}}_k\}\big|\boldsymbol{\theta}\big) = \prod_k p\big(\widehat{\boldsymbol{\mathcal{F}}}_k\big|\boldsymbol{\theta}\big) \tag{24}$$

where the product is taken over all frequency indices k in the selected band;

$$p\big(\widehat{\boldsymbol{\mathcal{F}}}_k\big|\boldsymbol{\theta}\big) = \frac{\pi^{-n}}{|\boldsymbol{E}_k|}\exp[-\widehat{\boldsymbol{\mathcal{F}}}_k^* \boldsymbol{E}_k^{-1}\widehat{\boldsymbol{\mathcal{F}}}_k] \tag{25}$$

and $|\cdot|$ denotes the matrix determinant; $\boldsymbol{E}_k$ is given in Eqn. (21). Accordingly, the NF is given by

$$L\big(\boldsymbol{\theta}, \{\widehat{\boldsymbol{\mathcal{F}}}_k\}\big) = nN_f \ln \pi + \sum_k \ln|\boldsymbol{E}_k| + \sum_k \widehat{\boldsymbol{\mathcal{F}}}_k^* \boldsymbol{E}_k^{-1}\widehat{\boldsymbol{\mathcal{F}}}_k \tag{26}$$

where the sum is over all k in the selected band. Without loss of generality, a joint uniform prior PDF is assumed for $\boldsymbol{\theta}$. The posterior distribution is then given by

$$p\big(\boldsymbol{\theta}\big|\{\widehat{\boldsymbol{\mathcal{F}}}_k\}\big) \propto e^{-L\left(\boldsymbol{\theta}, \{\widehat{\boldsymbol{\mathcal{F}}}_k\}\right)} \tag{27}$$

The data size in OMA is typically long and a Gaussian approximation is applied to approximate the posterior PDF as in Section 2. The Bayesian FFT approach was first formulated in [25]. Efficient algorithms have been developed for determining the MPV and the Hessian of the NF. A recent review with applications in civil engineering can be found in [26].

4.2 Deriving uncertainty laws via FIM

The connection between the uncertainty law and the FIM allows one to derive it in a systematic manner through an expectation rather than asymptotic analysis of the NL, which turns out to involve much insights on the mathematical structure of the NL [11] . Here we re-derive the uncertainty law for well-separated modes via the FIM. Let $\hat{L}^{(xy)}$ denote the derivative of the NF w.r.t. variables x and y, and evaluated at the MPV for given $\{\widehat{\boldsymbol{\mathcal{F}}}_k\}$ distributed as the likelihood function $p\big(\{\widehat{\boldsymbol{\mathcal{F}}}_k\}\big|\boldsymbol{\theta}\big)$. In the present context, Eqn. (16) reads

$$\hat{L}^{(xy)} = J_{xy}\big[1 + O(N_f^{-1/2})\big] \tag{28}$$

where J_{xy} is the entry corresponding to parameter pair (x, y) in the FIM; and N_f is the number of FFT points in the selected band. This implies that $\hat{L}^{(xy)} \sim J_{xy}$ if $J_{xy} \neq 0$. Otherwise, the leading order of $\hat{L}^{(xy)}$ needs to be determined by other means. As the FFT data is zero-mean complex Gaussian, it is a standard result in statistics that the FIM is given by [27]

$$J_{xy} = \textstyle\sum_k \mathrm{tr}[\boldsymbol{E}_k^{-1}\boldsymbol{E}_k^{(x)}\boldsymbol{E}_k^{-1}\boldsymbol{E}_k^{(y)}] \tag{29}$$

where a superscripted variable denotes a partial derivative w.r.t. it, and $\mathrm{tr}[\cdot]$ denotes the 'trace' of the subject matrix, i.e., sum of diagonal entries. Although the FIM is the expectation of second-

order partial derivative of NF, it turns out that only the first-order partial derivative of $\boldsymbol{E}_k$ is needed in the case of zero-mean complex Gaussian distribution.

The trace in Eqn. (29) can be evaluated effectively using the eigenvector decomposition of $\boldsymbol{E}_k$. Let $\{\boldsymbol{b}_i \in \mathbb{R}^n : i = 1,2,\dots,n\}$ be an orthonormal basis in $\mathbb{R}^n$ with $\boldsymbol{b}_1 = \bar{\boldsymbol{\varphi}}$. Substituting $\boldsymbol{I}_n = \sum_{i=1}^n \boldsymbol{b}_i \boldsymbol{b}_i^{\mathrm{T}}$ into Eqn. (21) yields

$$\boldsymbol{E}_k = (SD_k + S_e)\boldsymbol{b}_1\boldsymbol{b}_1^{\mathrm{T}} + S_e \sum_{i=2}^n \boldsymbol{b}_i \boldsymbol{b}_i^{\mathrm{T}} \tag{30}$$

This shows that the eigenvectors of $\boldsymbol{E}_k$ are $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, …, $\boldsymbol{b}_n$ with corresponding eigenvalues $SD_k + S_e$ and $S_e, \dots, S_e$. Its inverse can then be obtained by operating on the eigenvalues:

$$\boldsymbol{E}_k^{-1} = (SD_k + S_e)^{-1}\boldsymbol{b}_1\boldsymbol{b}_1^{\mathrm{T}} + S_e^{-1} \sum_{i=2}^n \boldsymbol{b}_i \boldsymbol{b}_i^{\mathrm{T}} \tag{31}$$

The partial derivatives can be obtained conveniently using Eqn. (30) by noting that the eigenvectors only depend on the mode shape $\bar{\boldsymbol{\varphi}}$ and the eigenvalues only depend on the remaining parameters $\{f, \zeta, S, S_e\}$. Details can be found in Appendix A and the results are summarized in Table 1. In the table, derivatives involving the dynamic amplification factor are written in terms of the derivatives of its reciprocal, which are mathematically easier:

$$\begin{aligned} (D_k^{-1})^{(f)} &= 4f^{-1}\beta_k^2(\beta_k^2 - 1 + 2\zeta^2) \\ (D_k^{-1})^{(\zeta)} &= 8\zeta\beta_k^2 \end{aligned} \tag{32}$$

Table 1: The entry J_{xy} in the FIM for $N_f \to \infty$; $e_k = S_e/SD_k$ and $\|\boldsymbol{\varphi}\| = 1$

x \ y	f, ζ	S	S_e	$\boldsymbol{\varphi}$
f, ζ	$\sum_k \frac{D_k^2 (D_k^{-1})^{(x)} (D_k^{-1})^{(y)}}{(1+e_k)^2}$	sym.		
S	$-S^{-1} \sum_k \frac{D_k (D_k^{-1})^{(y)}}{(1+e_k)^2}$	$S^{-2} \sum_k (1+e_k)^{-2}$		
S_e	$-S^{-1} \sum_k \frac{(D_k^{-1})^{(y)}}{(1+e_k)^2}$	$S^{-2} \sum_k \frac{D_k^{-1}}{(1+e_k)^2}$	$S_e^{-2}\left[(n-1)N_f + \sum_k \frac{e_k^2}{(1+e_k)^2}\right]$	
$\boldsymbol{\varphi}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$	$\left[2\sum_k (1+e_k)^{-1} e_k^{-1}\right](\boldsymbol{I}_n - \bar{\boldsymbol{\varphi}}\bar{\boldsymbol{\varphi}}^{\mathrm{T}})$

Note that the symbols $\{f, \zeta, \boldsymbol{\varphi}, S, S_e\}$ in Table 1 denote the actual parameter value that lead to the data, following the context in Section 3. In addition to long data asymptotics as is done here, small damping asymptotics and asymptotic decoupling are needed to obtain the final asymptotic expressions that relate the posterior uncertainty explicitly to test configurations (instead of sums as in Table 1). These are beyond the scope of this paper. Interested readers may refer to [11] for more details.

5 Empirical Studies

In order to validate the long-data asymptotic behavior of the posterior covariance matrix, empirical studies are presented here for Bayesian OMA with synthetic and experimental data of increasing lengths. Data length effect is studied through the 'normalized data length' N_c = data duration / natural period. Note that N_c is proportional to the number of FFT points in the selected band (N_f). It is directly related to the maximum amount of information available in data for inferring the mode of interest and is therefore an intuitive measure in the context of modal identification [13]. In the examples, three quantities of uncertainty should be distinguished. The 'posterior variance' (denoted by a cross in the figures) refers to the conventional posterior variance of a parameter calculated in the Bayesian identification algorithm for a given data set. It is equal to the diagonal entry (corresponding to the parameter) in the posterior covariance matrix, which is the inverse of the Hessian of NL at the MPV (for a given data set). The CRB (denoted by a triangle) refers to the (tightest) Cramér-Rao Bound, equal to the inverse of the FIM (involving all parameters). The FIM is equal to the expectation of the Hessian of NL at the *actual parameter value*, which is a legitimate quantity in the examples where the data is generated synthetically. The 'uncertainty law' (denoted by a circle) refers to the leading order of the posterior variance for long data (but not necessarily small damping), calculated as the inverse of the FIM using the expressions in Table 1 but *evaluated at the MPV*. Note that the FIM involved in the calculation of uncertainty law is not equal to the actual value involved in the CRB, which is evaluated at the actual parameter value. Calculating the uncertainty law does not involve the actual parameter value; the latter is not a legitimate quantity in a Bayesian context and need not exist for real data. As the examples will illustrate, the above three quantities converge to the same value as the data size increases. In real applications with given data, the posterior covariance and uncertainty law can be calculated, but not the CRB or FIM because the actual parameter value is not known or does not even exist. Despite its name, the CRB (the inverse of the FIM) does not lower-bound either the posterior variance or the uncertainty law. The name is used merely to be consistent with the conventional terminology in classical statistics.

5.1 Single-DoF system

The first example is a single-DoF system under Gaussian white noise excitation $W(t)$ with a modal force (per unit mass) PSD of $S = 1(\mu g)^2/Hz$ ($1\ \mu g = 9.81 \times 10^{-6}\ m/s^2$), satisfying the governing equation

$$\ddot{x} + 2\zeta\omega\dot{x} + \omega^2 x = W(t) \tag{33}$$

where $\omega = 2\pi\ rad/sec$ (i.e., $1\ Hz$) and $\zeta = 2\%$. Synthetic acceleration data is generated with a sampling frequency of $10\ Hz$ and contaminated by i.i.d. channel noise with a PSD $S_e = 25\ (\mu g)^2/Hz$. The resulting modal signal-to-noise (s/n) ratio, defined as the PSD ratio of response to noise at the natural frequency, is $\gamma = S/4S_e\zeta^2 = 25$. This quality of data can be readily achieved in typical ambient vibration tests. The PSD calculated using 1000 sec of data is shown in Figure 1.

Selecting a frequency band of $[0.7,1.3]\ Hz$, the Bayesian FFT method is applied to identify the modal parameters. The results are shown in Figure 2 for different data lengths. The dashed

line denotes the actual value of the modal parameter that generated the data. The identified result for each data length is shown with a dot at the posterior MPV and an error bar covering two posterior standard deviations. It is seen that as the data length increases the error bar systematically narrows. There is no evidence of bias, as reflected by the fact that the error bar covers the actual value regardless of data length.

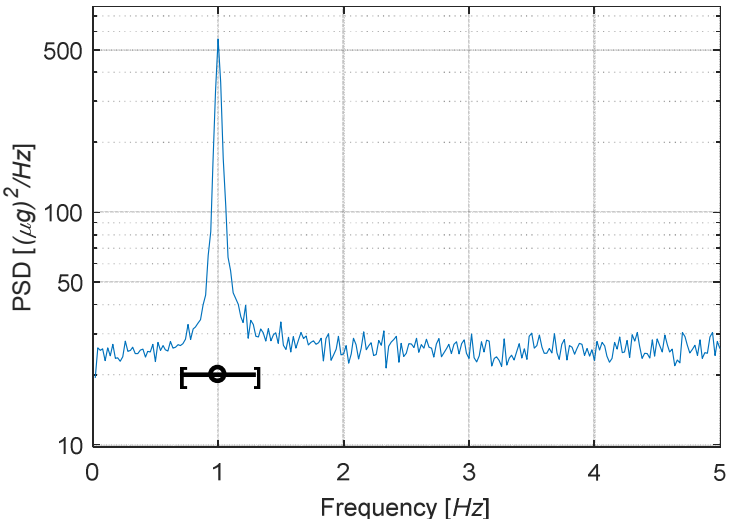


Figure 1: PSD spectrum, single-DoF example.
Bracket: frequency band used in Bayesian FFT algorithm.

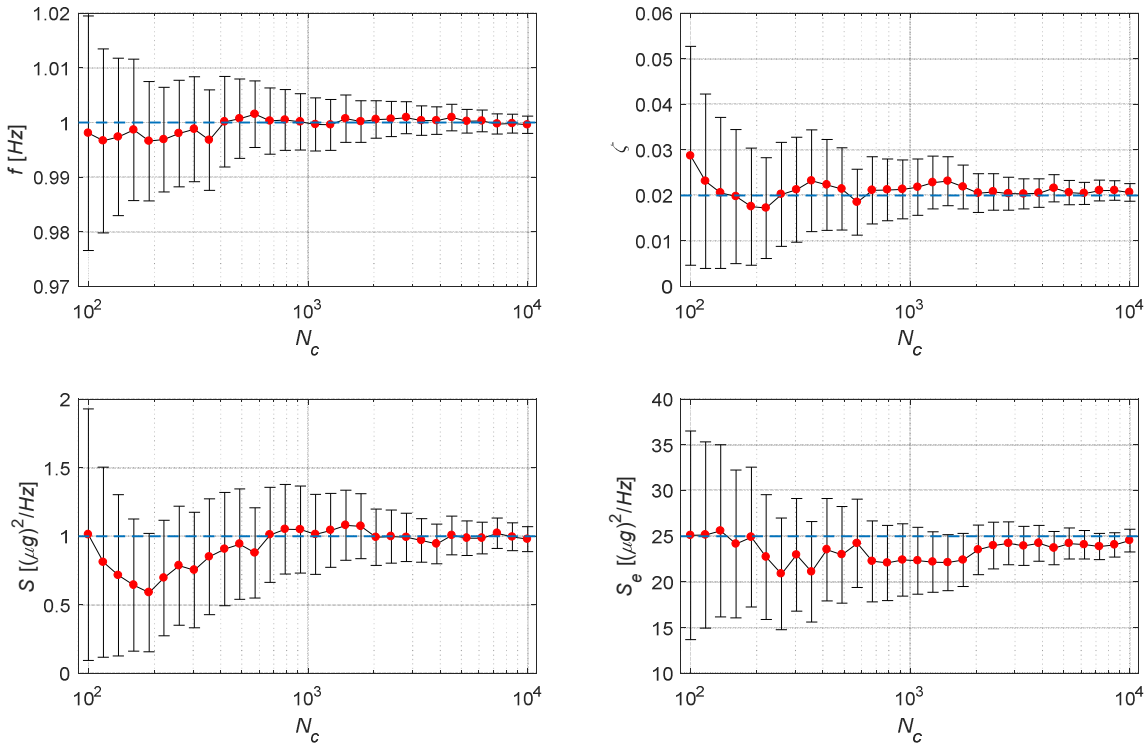


Figure 2: Identified modal parameters, single-DoF example.
Dot: MPV; error bar: ±two standard deviations. Dashed line: actual value. N_c = data length/natural period

The posterior standard deviation and covariance of modal parameters, both denoted by a circle, are shown in Figure 3 and Figure 4, respectively. The former is equal to the square root of the posterior covariance and the latter is equal to the corresponding off-diagonal entry in the posterior covariance matrix. As expected, the posterior standard deviation generally decreases with data length, by virtual of increasing amount of information for identification. The decrease need not be monotonic, however, as the identification uncertainty for a given data set still depends on the particular details of data (which is stochastic). The uncertainty law (circle) also has a general decreasing trend. Recalling that the uncertainty law is calculated via the analytical expressions in Table 1, its apparently random fluctuation is due to the fluctuation in the MPV calculated for different data lengths, which is close but not equal to the actual value. Evaluating the expressions at the actual value (and taking inverse) gives the CRB (triangle), which displays a smooth monotonic decreasing trend. The posterior covariance, uncertainty law and CRB all converges to the same value as the data length increases, validating the long-data asymptotic equivalence established in Eqn. (8). The uncertainty law appears to be closer to the posterior covariance than the CRB especially for short durations. This is an observation that is generally true (see also the next example), although it is not accounted by the theory. Note that the CRB cannot be calculated in real applications because the actual parameter value is not known or it does not even exist. As mentioned before this subsection, the CRB does not lower-bound either the posterior variance or the uncertainty law.

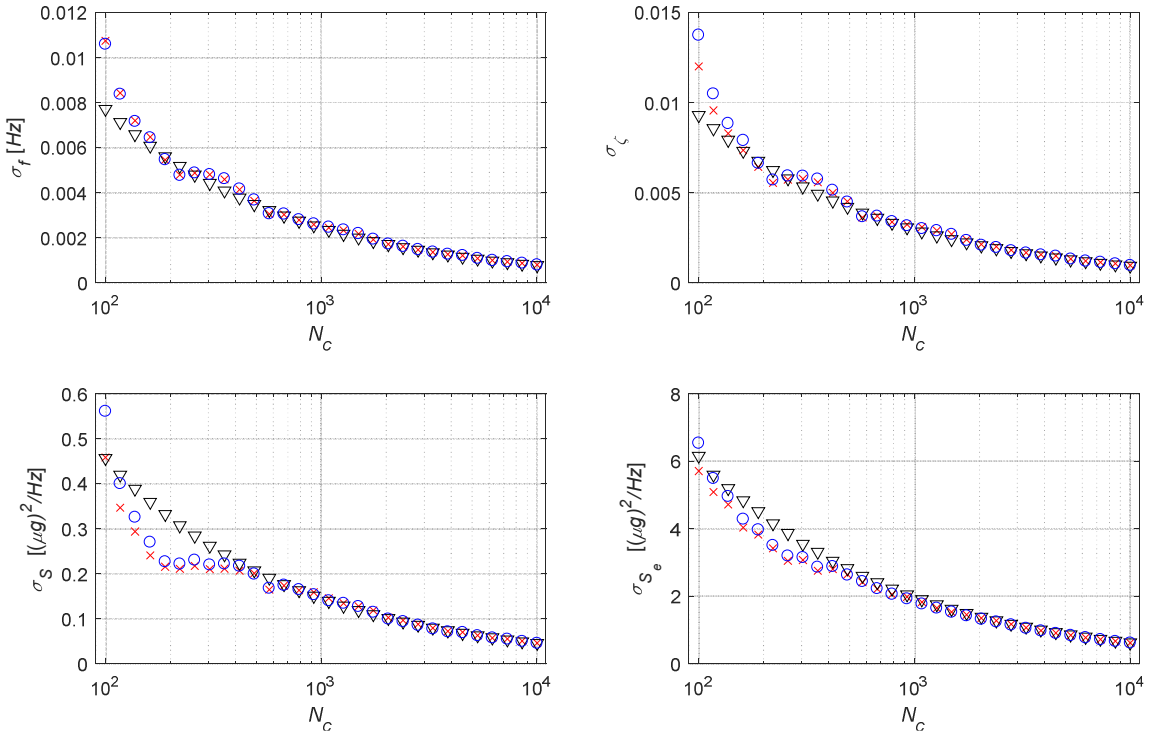


Figure 3: Posterior standard deviation of modal parameters (cross), single-DoF example. Circle: uncertainty law (MPV); triangular: CRB. N_c = data length/natural period

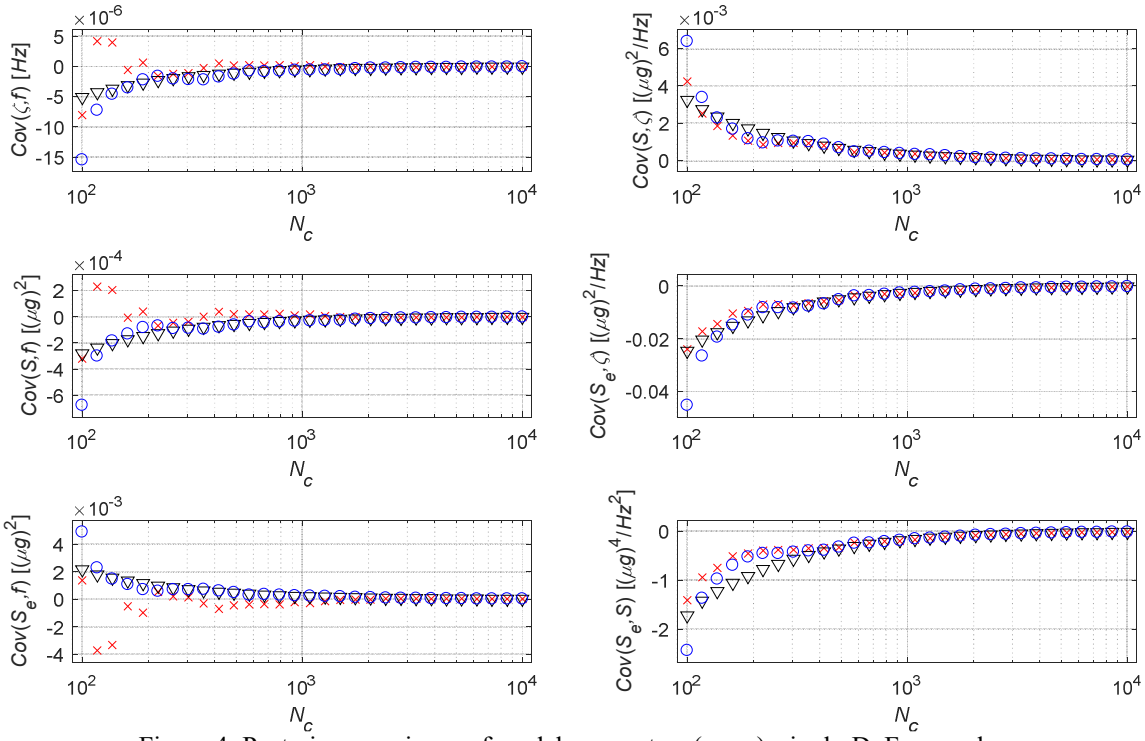


Figure 4: Posterior covariance of modal parameters (cross), single-DoF example.
Circle: uncertainty law (MPV); triangular: CRB. N_c = data length/natural period

5.2 Eight-DoF shear-type building

The second example is an eight-storied shearing-type building with a uniform floor mass of 100 ton and an interstory stiffness of 115.94 kN/mm. The first three frequencies are calculated to be 1, 2.97 and 4.83 Hz. A damping ratio of 1% is assumed for all modes. The structure is subjected to i.i.d. Gaussian white noise excitations at all floors, each with a PSD of 0.8 N^2/Hz. Synthetic acceleration data is generated at a sampling frequency of 50 Hz and contaminated by i.i.d. channel noise with a PSD of 50 $(\mu g)^2/Hz$. The first three mode shapes (normalized with unit norm) are shown in Figure 5. Uniaxial acceleration data is measured at the 3rd, 4th and 7th floors. Correspondingly, the modal s/n ratios are $\gamma = 16.2$, 20.4 and 4.3 for the first to third modes. Note that the modal s/n ratio is proportional to the sum of squares of mode shape values at the measured DoFs. Figure 6 show the PSD and singular value (SV) spectrum (i.e., eigenvalues of PSD matrix) of data (1000 sec duration), respectively.

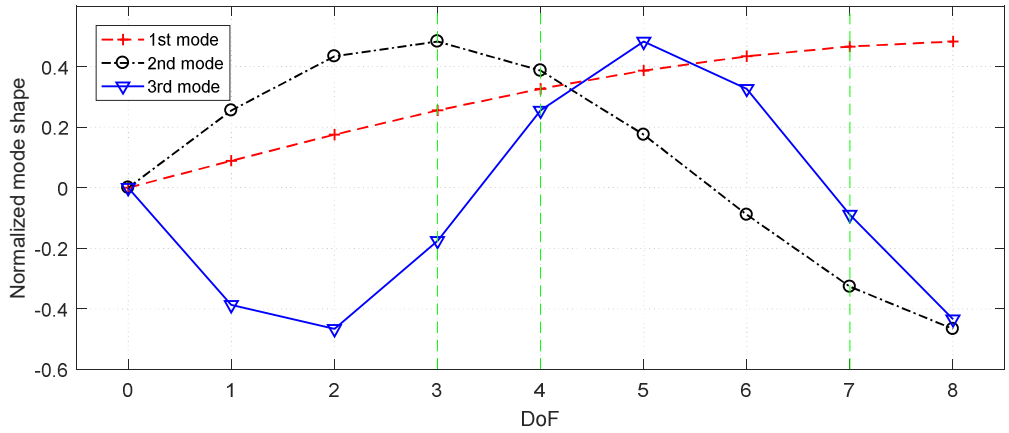


Figure 5: Mode shapes of the first three modes, eight-DoF example.
Vertical dashed line: measured DoFs.

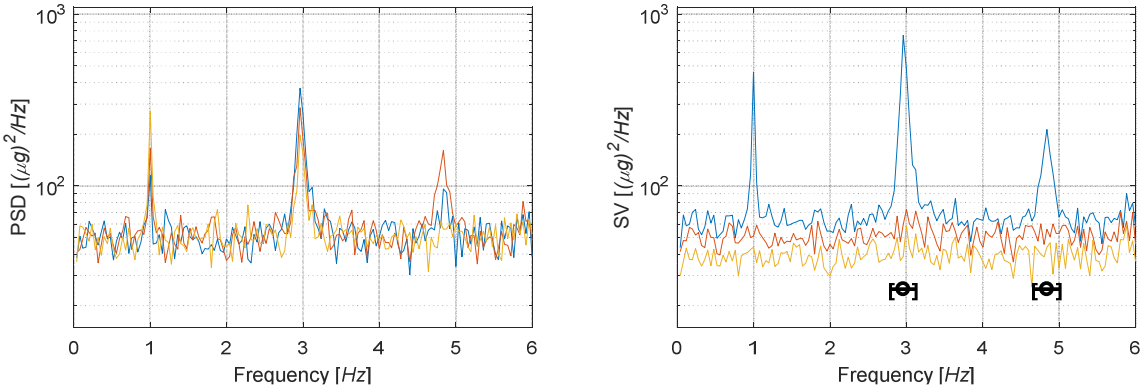


Figure 6: PSD and SV spectrum, eight-DoF example.
Bracket: frequency bands used in Bayesian FFT algorithm.

Second mode

Consider identifying the second mode based on FFT data in the frequency band $[2.8, 3.12]\ Hz$. Figure 7 shows the results for normalized data length from 100 to 10000. Similar to Figure 2, as the data length increases, the MPV (dot) converges (in a random manner) to the actual value (dashed line) and the error bar shortens. The corresponding posterior standard deviations of modal parameters (circles) are shown in Figure 8 and Figure 9. For mode shape which is a vector-valued quantity, the coefficient of variation (c.o.v.) is defined as the square root of the trace of posterior covariance matrix [11]. The differences between the posterior covariance, uncertainty law and CRB are apparent for short data durations but they diminish for long durations. Again, this validates the asymptotic equivalence established in this work.

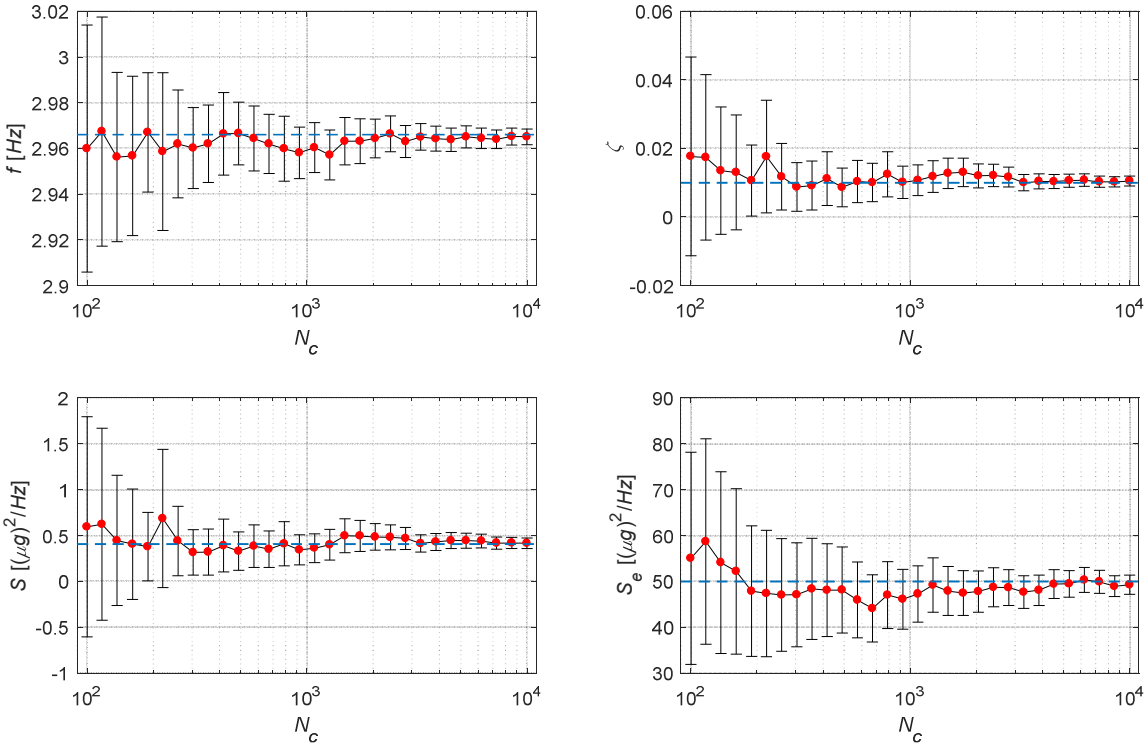


Figure 7: Identified modal parameters, second mode, eight-DoF example.
Dot: MPV; error bar: ±two standard deviations. Dashed line: actual value. N_c = data length/natural period

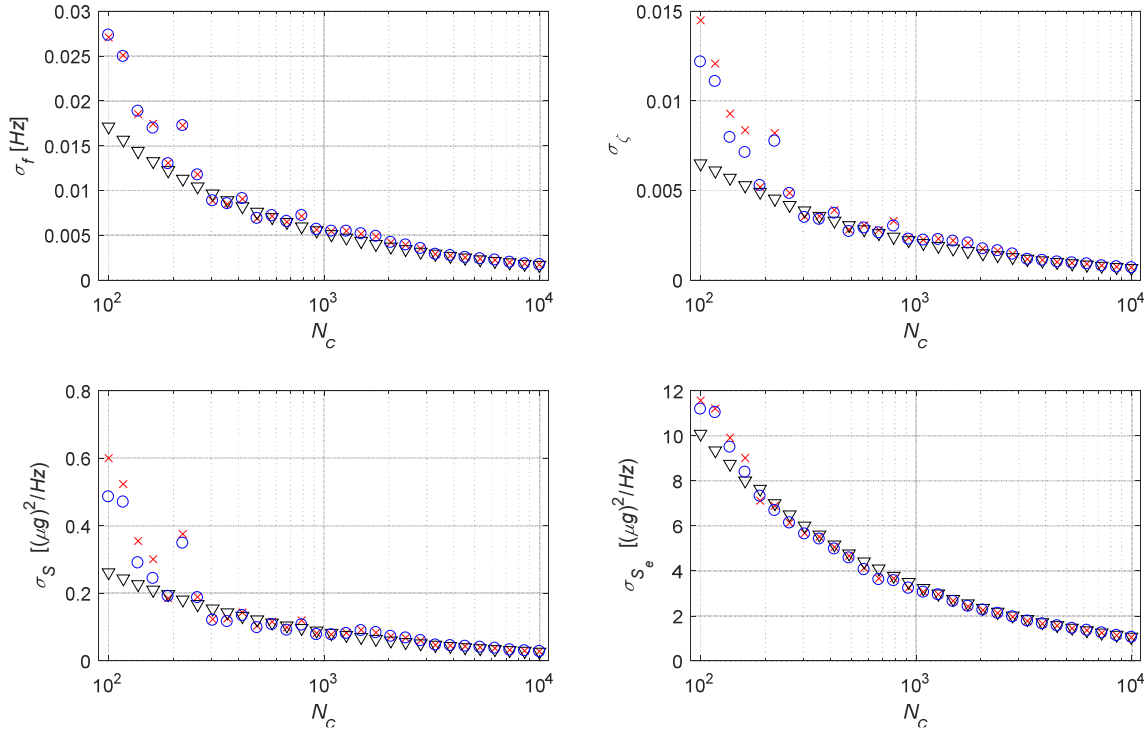


Figure 8: Posterior standard deviation of modal parameters (cross), second mode, eight-DoF example.
Circle: uncertainty law (MPV); triangular: CRB. N_c = data length/natural period

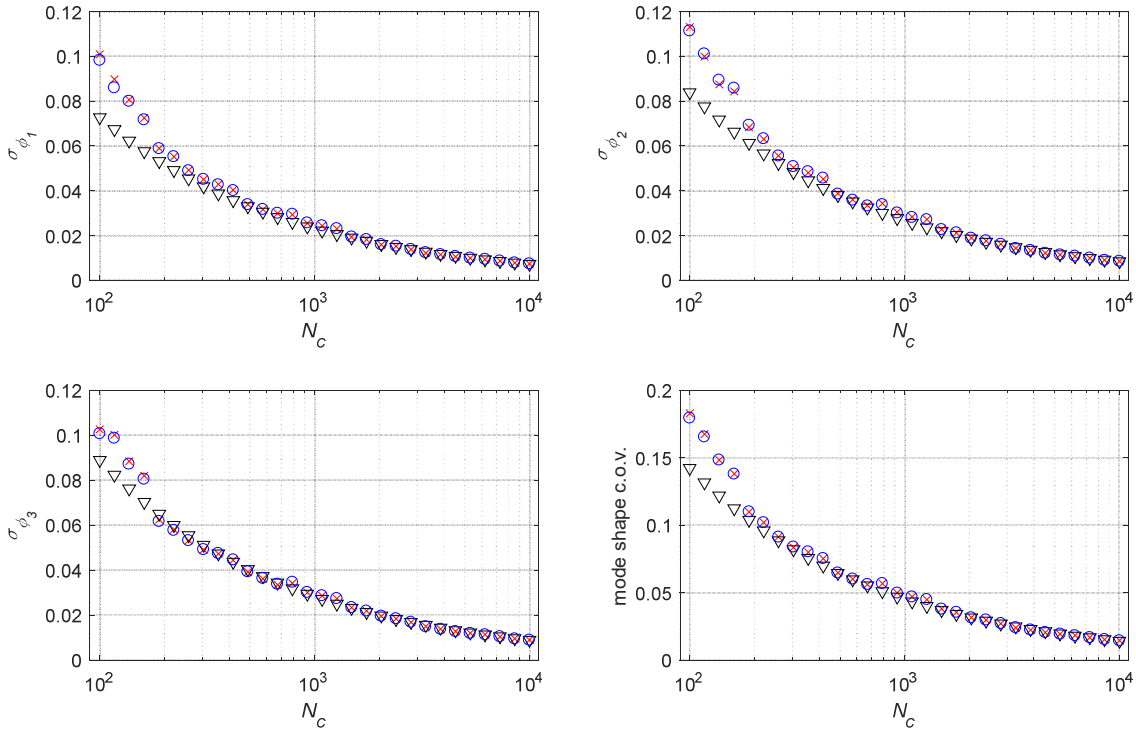


Figure 9: Posterior standard deviation of mode shape (cross), second mode, eight-DoF example. Circle: uncertainty law; triangular: CRB. N_c = data length/natural period

Third mode

Next, consider identifying the third mode using FFT data within the frequency band $[4.68, 5.0]\ Hz$. This mode is considered to illustrate the effect of low s/n ratio. As mentioned in the beginning of this section, the modal s/n of this mode is only $\gamma = 4.3$, which is quite low. Figure 10 shows the identification results, analogous to Figure 7. The identification uncertainty is significantly larger than that of the second mode. In Figure 11 and Figure 12, the posterior standard deviation and uncertainty law have a similar (or slightly lower) proximity as those for the second mode, but now they differ significantly from the CRB especially for short data. Such difference is due to the difference between the MPV and the actual value of modal parameters, the latter being used by the CRB. The difference between the MPV and the actual value can be large for short data but it diminishes as the data length increases. The posterior standard deviation, uncertainty law and CRB still converge to visually the same value, as predicted by the asymptotic equivalence.

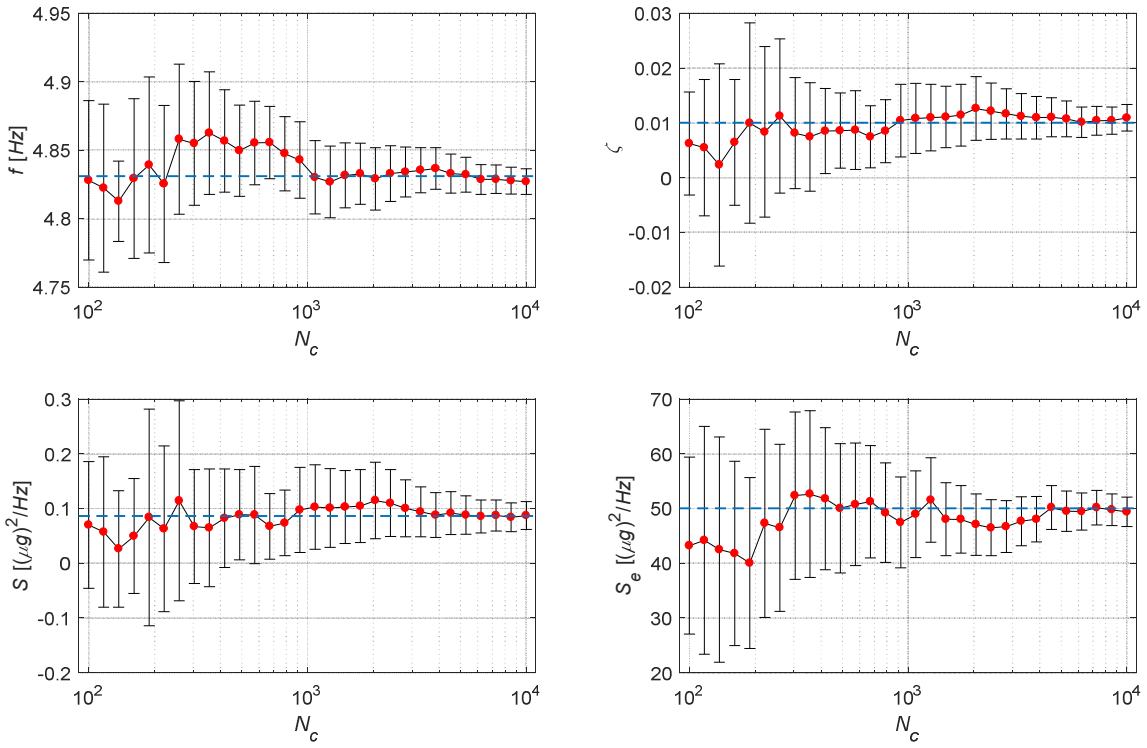


Figure 10: Identified modal parameters, third mode, eight-DoF example.
Dot: MPV; error bar: ±two standard deviations. Dashed line: actual value. N_c = data length/natural period

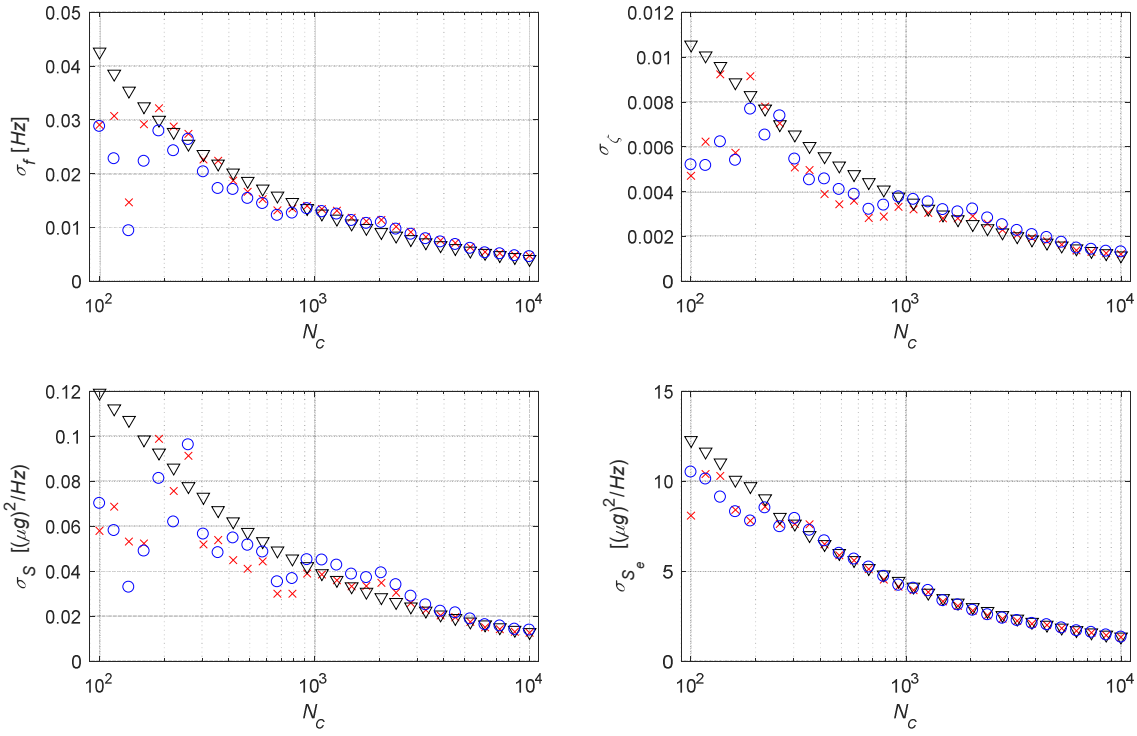


Figure 11: Posterior standard deviation of parameters, third mode, eight-DoF example.
Circle: uncertainty law; triangular: CRB. N_c = data length/natural period

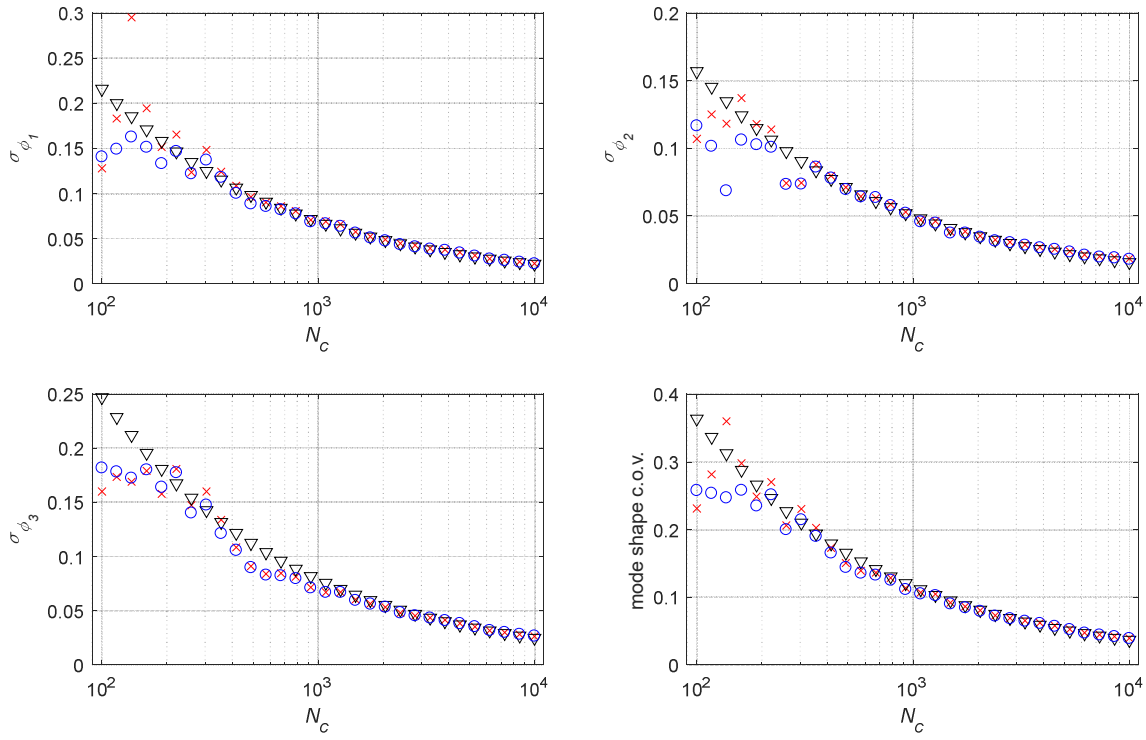
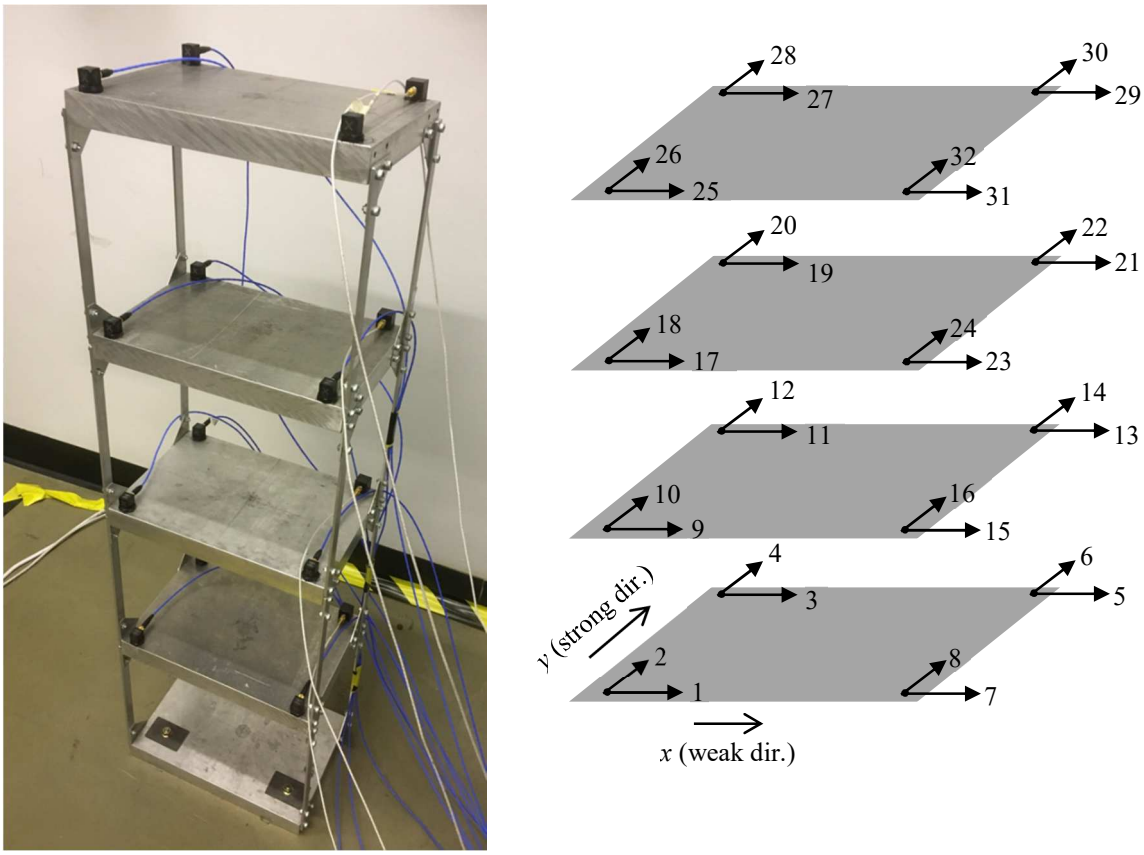


Figure 12: Posterior standard deviation of mode shape, third mode, eight-DoF example.
Circle: uncertainty law; triangular: CRB. N_c = data length/natural period

5.3 Four-story laboratory frame

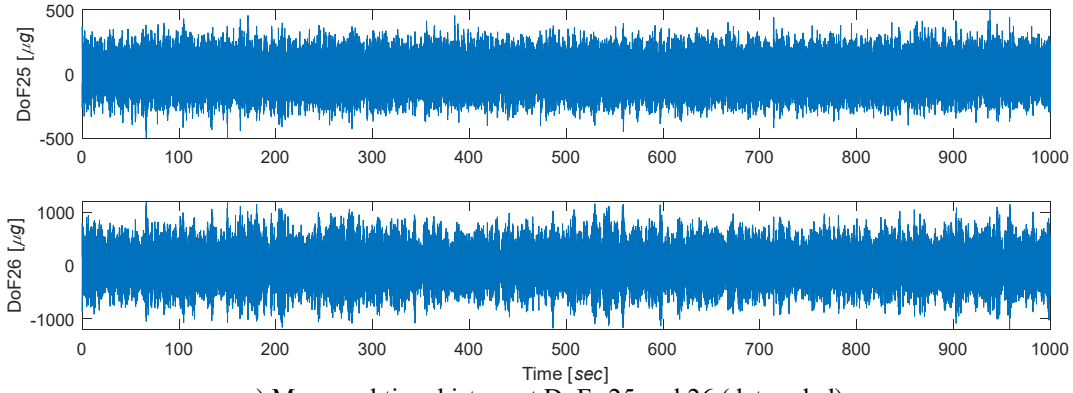
An example with laboratory data is next presented to illustrate the connection of uncertainty law and the FIM with real data. A laboratory OMA test was performed on a four-story aluminium frame, measuring 30 cm × 20 cm in plan and with a uniform story height of 25 cm, as shown in Figure 13a. The four corners of each floor were instrumented with two piezoelectric accelerometers along the horizontal x- and y- directions, giving a total of 32 measured DoFs (Figure 13b). Ambient data was recorded for 1000 sec at 2048 Hz. It was later decimated to 256 Hz for analysis. The measured time history at DoFs 25 and 26 are shown in Figure 14a). The PSD and SV spectrum of data (1000 sec duration) are shown in Figure 14b).



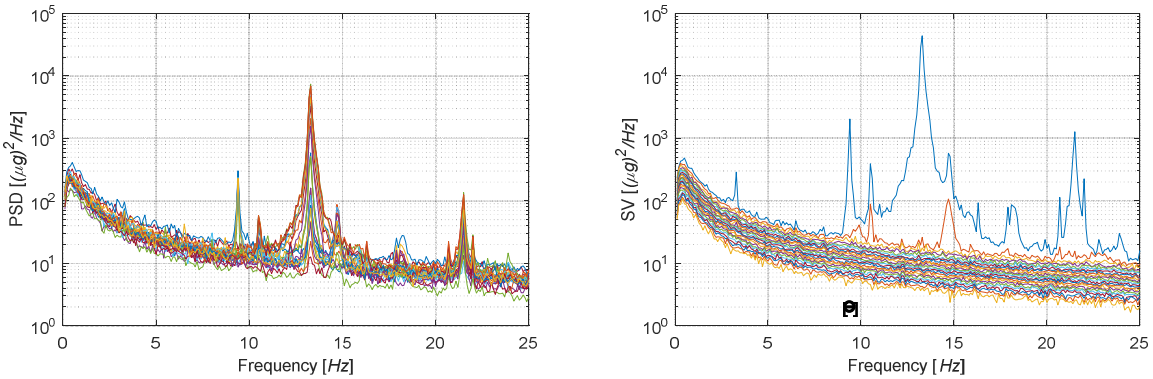
a) Photo b) Measured DoFs

Figure 13: Laboratory shear frame.

Consider identifying the mode based on FFT data in the frequency band $[9.1, 9.7]$ Hz. Figure 15 shows the results for normalized data length from 100 to 10000 and the mode shape (x-translation) is the MPV at $N_c = 10000$. Unlike previous examples based on synthetic data, there is no 'true model' in this example and the notion of the actual parameter values of modal properties is irrelevant. For the MPV, although intuition may expect it to converge as the data length increases, it need not do so in real experiments because of model errors due to, e.g., changing environment, damping mechanisms. Consistent with the intuition about increasing information, the error bar generally shortens as the data duration increases, though not necessarily in a monotonic manner. The corresponding posterior standard deviations of modal parameters (circles) are shown in Figure 16 and Figure 17. The CRB is not relevant in this case. Similar to the previous examples, the uncertainty laws are at good proximity to the posterior variances even for short data.

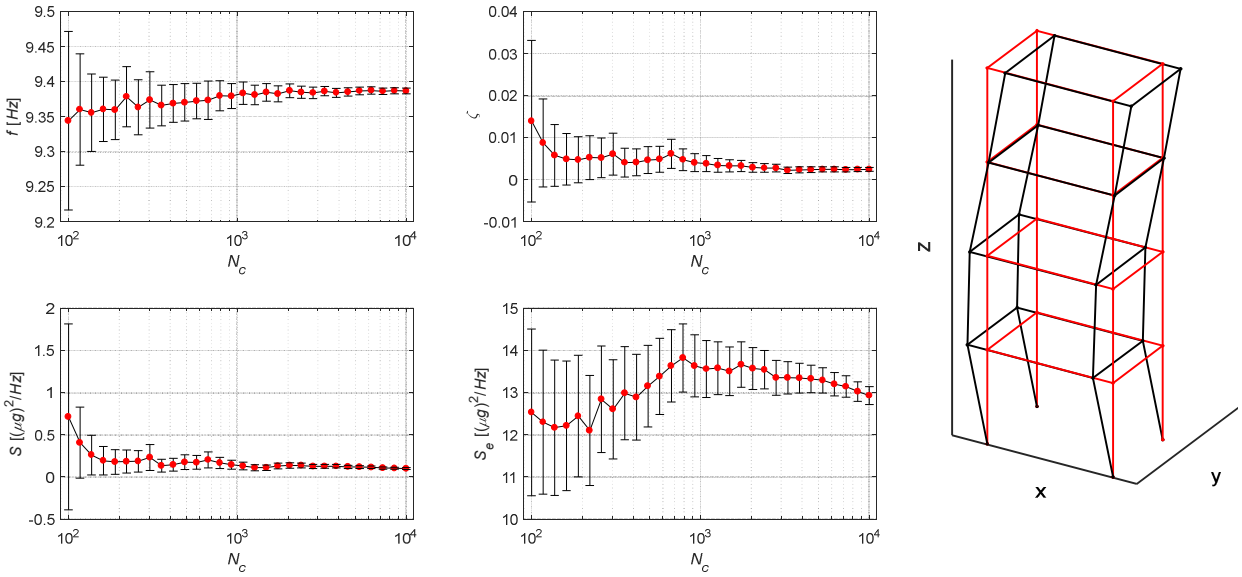


a) Measured time history at DoFs 25 and 26 (detrended)



b) PSD and SV spectrum (Bracket: frequency bands used in Bayesian FFT algorithm)

Figure 14: Measured data, laboratory shear frame.



a) Modal parameters b) Mode shape

Figure 15: Identified modal parameters, laboratory shear frame.
Dot: MPV; error bar: ±two standard deviations.

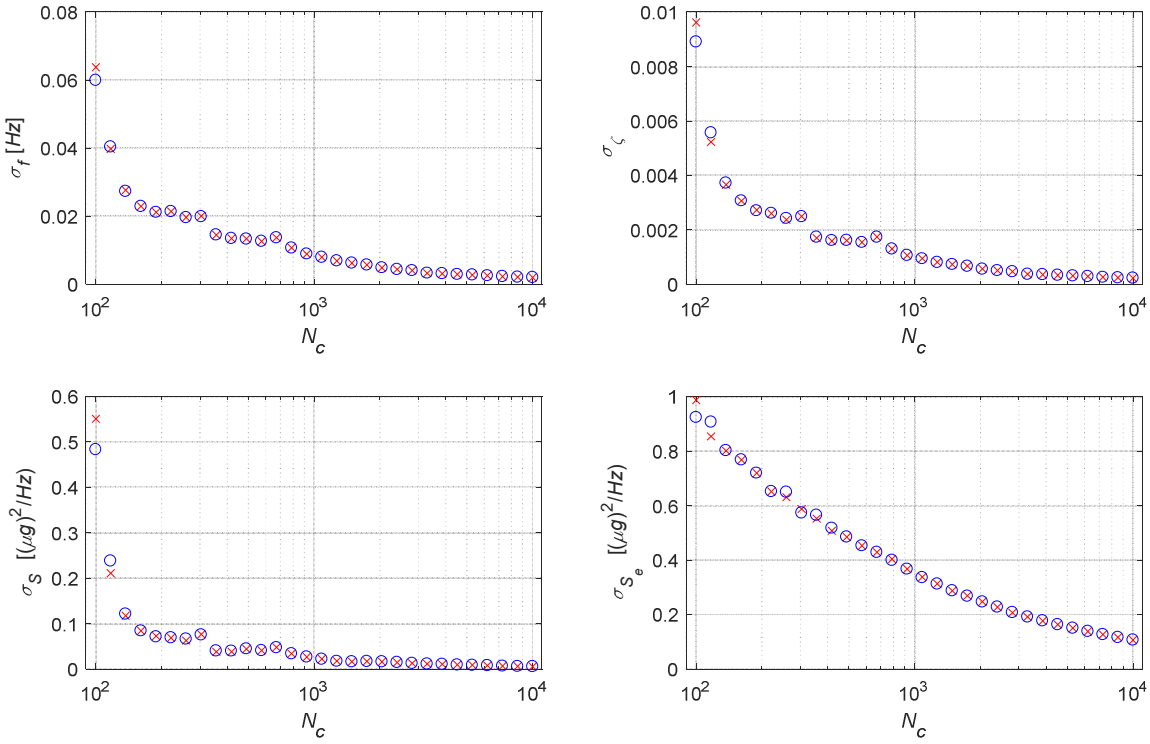


Figure 16: Posterior standard deviation of modal parameters (cross), laboratory shear frame. Circle: uncertainty law (MPV). N_c = data length/natural period

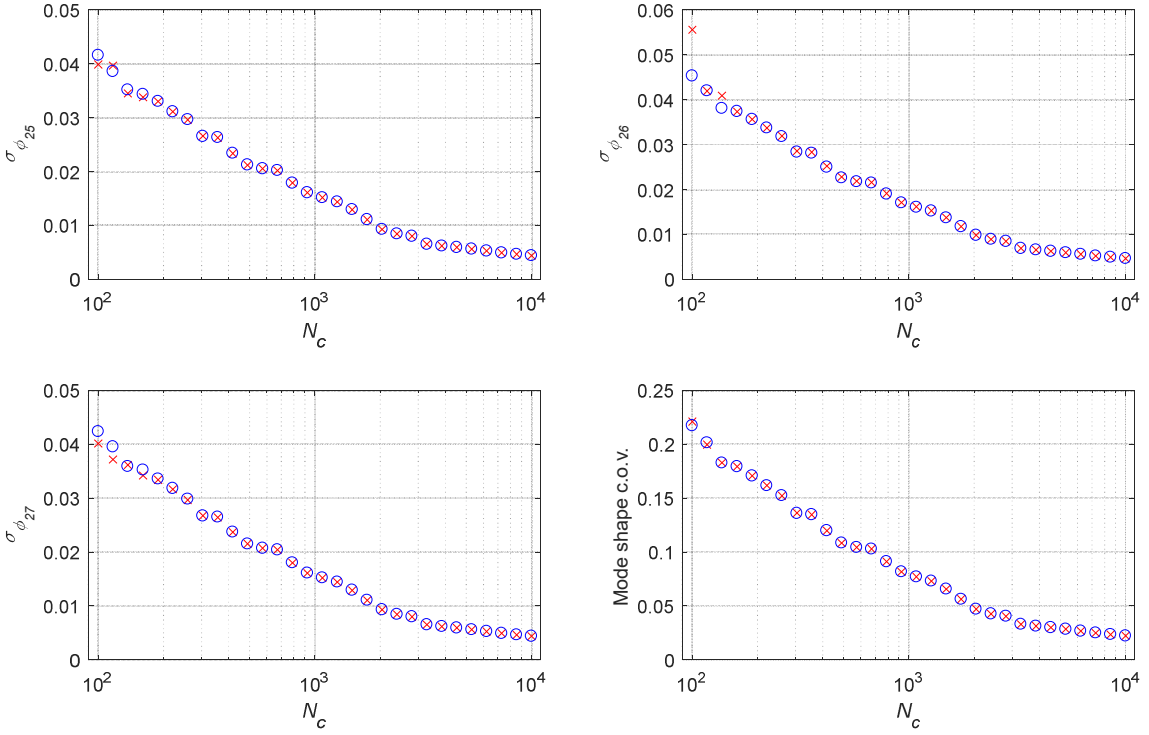


Figure 17: Posterior standard deviation of mode shape (cross), laboratory shear frame. Circle: uncertainty law. N_c = data length/natural period

6 Conclusions

This paper has investigated the asymptotic behavior of the posterior covariance matrix in Bayesian system identification. Based on Gaussian approximation of the posterior distribution in a globally identifiable problem, it is found that the posterior covariance matrix is asymptotically equal to the inverse of the Fisher information matrix (FIM), which coincides with the tightest Cramér-Rao Bound (CRB, involving all parameters). Despite its name, the CRB does not lower-bound either the posterior variance (for given data) or its leading order (uncertainty law).

The mathematical connection with the CRB offers a systematic means for deriving the uncertainty laws in operational modal analysis, which aims at understanding the relationship between identification uncertainty and test configuration. To illustrate the utility of this connection, the uncertainty law for well-separated modes is re-derived through the FIM and examined by synthetic and experimental data. As the FIM has an analytical form, the derivation is more systematic and elegant than the previous one that required insights about the asymptotic mathematical structure of the log-likelihood function. The synthetic data examples demonstrate that when there is no modeling error the posterior variance, uncertainty law (long data) and CRB all converge to the same value as the data duration increases. The synthetic and experimental data examples suggest that the uncertainty law is at good proximity with the posterior variance even under non-asymptotic situations.

7 Acknowledgments

This work is supported by the UK Engineering and Physical Sciences Research Council (Responsive Mode Grant EP/N017897/1).

8 Appendix A

In this appendix, we derive the expressions for the leading order of $\hat{L}^{(xy)}$ shown in Tab. 1 based on its relation with the FIM. Starting with the eigenvector decomposition of $\boldsymbol{E}_k$ in Eqn. (30), for $x = f, \zeta, S, S_e$, differentiation simply operates on the eigenvalues:

$$\boldsymbol{E}_k^{(x)} = (SD_k + S_e)^{(x)} \boldsymbol{b}_1 \boldsymbol{b}_1^{\mathrm{T}} + S_e^{(x)} \sum_{i=2}^{n} \boldsymbol{b}_i \boldsymbol{b}_i^{\mathrm{T}} \tag{A.1}$$

The same also applies to the product. For $x = f, \zeta, S, S_e$,

$$\boldsymbol{E}_k^{-1} \boldsymbol{E}_k^{(x)} \boldsymbol{E}_k^{-1} \boldsymbol{E}_k^{(y)} = \frac{(SD_k+S_e)^{(x)}(SD_k+S_e)^{(y)}}{(SD_k+S_e)^2} \boldsymbol{b}_1 \boldsymbol{b}_1^{\mathrm{T}} + \frac{S_e^{(x)} S_e^{(y)}}{S_e^2} \sum_{i=2}^{n} \boldsymbol{b}_i \boldsymbol{b}_i^{\mathrm{T}} \tag{A.2}$$

Taking trace, noting $\mathrm{tr}\left[\boldsymbol{b}_i \boldsymbol{b}_i^{\mathrm{T}}\right] = 1$ and summing over k yields

$$J_{xy} = \sum_k \frac{(SD_k+S_e)^{(x)}(SD_k+S_e)^{(y)}}{(SD_k+S_e)^2} + (n-1) N_f \frac{S_e^{(x)} S_e^{(y)}}{S_e^2} \tag{A.3}$$

The terms $J_{x\boldsymbol{\varphi}}$ ($x = f, \zeta, S, S_e$) and $J_{\boldsymbol{\varphi}\boldsymbol{\varphi}}$ involve differentiation w.r.t. $\boldsymbol{\varphi}$. Analysis can be simplified by making use of the orthogonality between $\boldsymbol{\varphi}$ and the gradient of $\bar{\boldsymbol{\varphi}} = \|\boldsymbol{\varphi}\|^{-1} \boldsymbol{\varphi}$. Let $\boldsymbol{\varphi} = [\phi_1, \dots, \phi_n]^{\mathrm{T}}$ and $\boldsymbol{e}_i$ be a $n \times 1$ vector whose ith entry is the only non-zero entry equal to 1. Then,

$$\overline{\boldsymbol{\varphi}}^{(\phi_i)} = \nabla\overline{\boldsymbol{\varphi}}\boldsymbol{e}_i \quad \nabla\overline{\boldsymbol{\varphi}} = \left[\frac{\partial\overline{\boldsymbol{\varphi}}}{\partial\phi_1} \quad \cdots \quad \frac{\partial\overline{\boldsymbol{\varphi}}}{\partial\phi_n}\right] = \boldsymbol{I}_n - \overline{\boldsymbol{\varphi}}\overline{\boldsymbol{\varphi}}^{\mathrm{T}} \quad (\|\boldsymbol{\varphi}\| = 1) \tag{A.4}$$

It follows that

$$(\nabla\overline{\boldsymbol{\varphi}})^{\mathrm{T}} = \nabla\overline{\boldsymbol{\varphi}} \quad (\nabla\overline{\boldsymbol{\varphi}})(\nabla\overline{\boldsymbol{\varphi}}) = \nabla\overline{\boldsymbol{\varphi}} \quad (\nabla\overline{\boldsymbol{\varphi}})\boldsymbol{\varphi} = \boldsymbol{0} \quad \boldsymbol{\varphi}^{\mathrm{T}}(\nabla\overline{\boldsymbol{\varphi}}) = \boldsymbol{0} \tag{A.5}$$

$$\boldsymbol{E}_k^{(\phi_i)} = SD_k\left[\overline{\boldsymbol{\varphi}}^{(\phi_i)}\overline{\boldsymbol{\varphi}}^{\mathrm{T}} + \overline{\boldsymbol{\varphi}}\overline{\boldsymbol{\varphi}}^{(\phi_i)\mathrm{T}}\right] = SD_k\left[\nabla\overline{\boldsymbol{\varphi}}\boldsymbol{e}_i\overline{\boldsymbol{\varphi}}^{\mathrm{T}} + \overline{\boldsymbol{\varphi}}\boldsymbol{e}_i^{\mathrm{T}}\nabla\overline{\boldsymbol{\varphi}}\right] \tag{A.6}$$

Using these results, it can be shown that, for $\|\boldsymbol{\varphi}\| = 1$ and $x = f, \zeta, S, S_e$,

$$\boldsymbol{E}_k^{-1}\boldsymbol{E}_k^{(x)} = \frac{(SD_k+S_e)^{(x)}}{SD_k(1+e_k)}\overline{\boldsymbol{\varphi}}\overline{\boldsymbol{\varphi}}^{\mathrm{T}} + \frac{S_e^{(x)}}{S_e}\nabla\overline{\boldsymbol{\varphi}} \tag{A.7}$$

$$\boldsymbol{E}_k^{-1}\boldsymbol{E}_k^{(\phi_i)} = (1+e_k)^{-1}\overline{\boldsymbol{\varphi}}\boldsymbol{e}_i^{\mathrm{T}}(\nabla\overline{\boldsymbol{\varphi}}) + e_k^{-1}(\nabla\overline{\boldsymbol{\varphi}})\boldsymbol{e}_i\overline{\boldsymbol{\varphi}}^{\mathrm{T}} \tag{A.8}$$

$$\boldsymbol{E}_k^{-1}\boldsymbol{E}_k^{(x)}\boldsymbol{E}_k^{-1}\boldsymbol{E}_k^{(\phi_i)} = \frac{(SD_k+S_e)^{(x)}}{SD_k(1+e_k)^2}\overline{\boldsymbol{\varphi}}\boldsymbol{e}_i^{\mathrm{T}}(\nabla\overline{\boldsymbol{\varphi}}) + \frac{S_e^{(x)}}{S_e e_k}\nabla\overline{\boldsymbol{\varphi}}\boldsymbol{e}_i\overline{\boldsymbol{\varphi}}^{\mathrm{T}} \tag{A.9}$$

$$\boldsymbol{E}_k^{-1}\boldsymbol{E}_k^{(\phi_i)}\boldsymbol{E}_k^{-1}\boldsymbol{E}_k^{(\phi_j)} = (1+e_k)^{-1}e_k^{-1}\left[\overline{\boldsymbol{\varphi}}\boldsymbol{e}_i^{\mathrm{T}}(\nabla\overline{\boldsymbol{\varphi}})\boldsymbol{e}_j\overline{\boldsymbol{\varphi}}^{\mathrm{T}} + (\nabla\overline{\boldsymbol{\varphi}})\boldsymbol{e}_i\boldsymbol{e}_j^{\mathrm{T}}(\nabla\overline{\boldsymbol{\varphi}})\right] \tag{A.10}$$

where $e_k = S_e/SD_k$.
Taking trace on Eqn. (A.9) and summing over k gives

$$J_{x\phi_i} = 0 \tag{A.11}$$

because $\mathrm{tr}\left[\overline{\boldsymbol{\varphi}}\boldsymbol{e}_i^{\mathrm{T}}(\nabla\overline{\boldsymbol{\varphi}})\right] = \mathrm{tr}\left[(\nabla\overline{\boldsymbol{\varphi}})\overline{\boldsymbol{\varphi}}\boldsymbol{e}_i^{\mathrm{T}}\right] = 0$ and $\mathrm{tr}[\nabla\overline{\boldsymbol{\varphi}}\boldsymbol{e}_i\overline{\boldsymbol{\varphi}}^{\mathrm{T}}] = \mathrm{tr}[\overline{\boldsymbol{\varphi}}^{\mathrm{T}}\nabla\overline{\boldsymbol{\varphi}}\boldsymbol{e}_i] = 0$ (cyclic property of trace). This indicates that the leading order of $\hat{L}^{(x\phi_i)}$ is not given by FIM. Direct investigation of $\hat{L}^{(x\phi_i)}$, which is a sum of N_f random terms, shows that it is $O(N_f^{1/2})$.
On the other hand, taking trace on Eqn. (A.10) and summing over k gives

$$J_{\phi_i\phi_j} = [2\textstyle\sum_k(1+e_k)^{-1}e_k^{-1}]\boldsymbol{e}_i^{\mathrm{T}}(\nabla\overline{\boldsymbol{\varphi}})\boldsymbol{e}_j \tag{A.12}$$

where we have simplified using the cyclic property of trace:

$$\begin{aligned}\mathrm{tr}\left[\overline{\boldsymbol{\varphi}}\boldsymbol{e}_i^{\mathrm{T}}(\nabla\overline{\boldsymbol{\varphi}})\boldsymbol{e}_j\overline{\boldsymbol{\varphi}}^{\mathrm{T}}\right] &= \mathrm{tr}\left[\overline{\boldsymbol{\varphi}}^{\mathrm{T}}\overline{\boldsymbol{\varphi}}\boldsymbol{e}_i^{\mathrm{T}}(\nabla\overline{\boldsymbol{\varphi}})\boldsymbol{e}_j\right] = \boldsymbol{e}_i^{\mathrm{T}}(\nabla\overline{\boldsymbol{\varphi}})\boldsymbol{e}_j \\ \mathrm{tr}\left[(\nabla\overline{\boldsymbol{\varphi}})\boldsymbol{e}_i\boldsymbol{e}_j^{\mathrm{T}}(\nabla\overline{\boldsymbol{\varphi}})\right] &= \mathrm{tr}\left[\boldsymbol{e}_j^{\mathrm{T}}(\nabla\overline{\boldsymbol{\varphi}})(\nabla\overline{\boldsymbol{\varphi}})\boldsymbol{e}_i\right] = \boldsymbol{e}_j^{\mathrm{T}}(\nabla\overline{\boldsymbol{\varphi}})\boldsymbol{e}_i = \boldsymbol{e}_i^{\mathrm{T}}(\nabla\overline{\boldsymbol{\varphi}})\boldsymbol{e}_j\end{aligned} \tag{A.13}$$

Since $\boldsymbol{e}_i^{\mathrm{T}}(\nabla\overline{\boldsymbol{\varphi}})\boldsymbol{e}_j$ is simply the (i,j)-entry of $\nabla\overline{\boldsymbol{\varphi}}$, Eqn. (A.12) implies

$$J_{\boldsymbol{\varphi}\boldsymbol{\varphi}} = [2\textstyle\sum_k(1+e_k)^{-1}e_k^{-1}](\nabla\overline{\boldsymbol{\varphi}}) = [2\textstyle\sum_k(1+e_k)^{-1}e_k^{-1}](\boldsymbol{I}_n - \overline{\boldsymbol{\varphi}}\overline{\boldsymbol{\varphi}}^{\mathrm{T}}) \tag{A.14}$$

Applying Eqns. (A.3), (A.11) and (A.14) gives the expressions in Table 1.

9 References

1. Beck JL. Bayesian system identification based on probability logic. *Structural Control and Health Monitoring* 2010; **17**(7): 825–847. DOI: 10.1002/stc.424.

2. Beck JL, Katafygiotis LS. Updating models and their uncertainties. I: bayesian statistical framework. *Journal of Engineering Mechanics* 1998; **124**(4): 463–467. DOI: 10.1061/(ASCE)0733-9399(1998)124:4(463).

3. Papadimitriou C, Beck JL, Katafygiotis LS. Updating robust reliability using structural test data. *Probabilistic Engineering Mechanics* 2001; **16**(2): 103–113. DOI: 10.1016/S0266-8920(00)00012-6.

4. Ghosh JK, Delampady M, Samanta T. *An introduction to Bayesian analysis: theory and methods*. New York: Springer Science & Business Media; 2007.

5. Erdélyi A. *Asymptotic expansions*. New York: Dover Publications; 1956.

6. Bleistein N, Handelsman RA. *Asymptotic expansions of integrals*. New York: Dover Publications; 1986.

7. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian data analysis*. Third Edit. Boca Raton: Chapman & Hall/CRC; 2014.

8. Liu JS. *Monte Carlo strategies in scientific computing*. New York: Springer; 2001.

9. Robert CP, Casella G. *Monte Carlo Statistical Methods*. New York: Springer; 2004.

10. Straub D, Papaioannou I. Bayesian Updating with Structural Reliability Methods. *Journal of Engineering Mechanics* 2015; **141**(3): 4014134. DOI: 10.1061/(ASCE)EM.1943-7889.0000839.

11. Au SK. Uncertainty law in ambient modal identification - Part I: Theory. *Mechanical Systems and Signal Processing* 2014; **48**(1–2): 15–33. DOI: 10.1016/j.ymssp.2013.07.016.

12. Au SK. Uncertainty law in ambient modal identification---Part II: Implication and field verification. *Mechanical Systems and Signal Processing* 2014; **48**(1–2): 34–48. DOI: 10.1016/j.ymssp.2013.07.017.

13. Cramér H. *Mathematical Methods of Statistics*. Princeton: Princeton University Press; 1946.

14. Rao CR. Information and the accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society* 1945; **37**: 81–89.

15. Au SK. Connecting Bayesian and frequentist quantification of parameter uncertainty in system identification. *Mechanical Systems and Signal Processing* 2012; **29**: 328–342. DOI: 10.1016/j.ymssp.2012.01.010.

16. Castro-Triguero R, Saavedra Flores EI, DiazDelaO FA, Friswell MI, Gallego R. Optimal sensor placement in timber structures by means of a multi-scale approach with material uncertainty. *Structural Control and Health Monitoring* 2014; **21**(12): 1437–1452. DOI: 10.1002/stc.1654.

17. Hernandez EM, Polanco NR. A lower bound for the variance of frequency and damping ratio identified from noisy vibration measurements. *Structural Control and Health Monitoring* 2016; **23**(1): 5–19. DOI: 10.1002/stc.1757.

18. Keener RW. *Theoretical statistics : topics for a core course*. Springer; 2010.

19. DasGupta A. *Asymptotic Theory of Statistics and Probability*. 1st ed. New York, NY: Springer New York; 2008. DOI: 10.1007/978-0-387-75971-5.

20. Çatbaş FN, Kijewski-Correa T, Aktan AE, editors. *Structural Identification of Constructed Systems: Approaches, Methods, and Technologies for Effective Practice of St-Id*. Reston, VA: American Society of Civil Engineers; 2013. DOI: 10.1061/9780784411971.

21. Reynders E. System Identification Methods for (Operational) Modal Analysis: Review and Comparison. *Archives of Computational Methods in Engineering* 2012; **19**(1): 51–124. DOI: 10.1007/s11831-012-9069-x.

22. Brincker R, Ventura C. *Introduction to operational modal analysis*. Chichester: Wiley; 2015.

23. Au SK. Model validity and frequency band selection in operational modal analysis. *Mechanical Systems and Signal Processing* 2016; **81**: 339–359. DOI: 10.1016/j.ymssp.2016.03.025.

24. Brillinger DR. *Time series : data analysis and theory*. Philadelphia: SIAM; 2001. DOI: http://dx.doi.org/10.1137/1.9780898719246.

25. Yuen K, Katafygiotis L. Bayesian fast Fourier transform approach for modal updating using ambient data. *Advances in Structural Engineering* 2003; **6**(2): 81–95. DOI: 10.1260/136943303769013183.

26. Au SK, Zhang FL, Ni YC. Bayesian operational modal analysis: Theory, computation, practice. *Computers & Structures* 2013; **126**: 3–14. DOI: 10.1016/j.compstruc.2012.12.015.

27. Tong YL. *The Multivariate Normal Distribution*. New York, NY: Springer New York; 1990. DOI: 10.1007/978-1-4613-9655-0.