Sensitivity Metrics for Maximum Likelihood System Identification

Thomas J. Matarazzo, Ph.D., S.M.ASCE[1]; and Shamim N. Pakzad, A.M.ASCE[2]

Abstract: This paper introduces a set of sensitivity metrics to be used along likelihood-based modal identification methods. In maximum likelihood (ML) estimation theory, the precision of ML point estimates can be measured by the curvature of the likelihood function. This paper presents closed-form partial derivatives, observed information, and variance expressions for discrete-time stochastic state-space model parameters as well as state matrix features that influence modal estimates. The results are derived for the observation matrix and the state matrix as well as eigenvalues and eigenvectors of the state matrix; these model entities correspond to natural vibration properties of a structural system. Confidence intervals are constructed for natural frequencies, damping ratios, and mode shapes using the derived asymptotic covariance matrices and the asymptotic normality property of ML estimators. The results are a supplement to the ML-based structural identification using expectation maximization (STRIDE) modal identification algorithm and are applicable to modal identification techniques formulated in the time-domain stochastic state-space model for linear time invariant systems. An application to structural modal identification is included to compare closed-form asymptotic parameter uncertainties to Monte Carlo bootstrap estimates. **DOI: 10.1061/AJRUA6.0000832.**

Author keywords: Estimation precision; Maximum likelihood; BIGDATA; Mobile sensing; Modal identification.

Introduction

The true behavior of a structural system is influenced by the interaction of many epistemic and aleatory random variables in the physical world. These random variables introduce uncertainty into otherwise deterministic systems, consequently reducing confidence in structural assessments. Through techniques such as stochastic modeling or Monte Carlo simulations, engineers can better discern and predict the true performance of a structural system by quantifying the effects of such random variables.

In structural health monitoring (SHM), it is necessary to include field measurements, which contain deterministic and stochastic attributes, within a model that represents the dynamic behavior of a structural system. In general, collected data are analyzed to explain structural behavior (Abdel-Ghaffar and Scanlan 1985; Kwasniewski et al. 2006; Shahidi et al. 2015) or infer properties of existing structures (Juang and Phan 2001; Pakzad and Fenves 2009; Peeters and De Roeck 1999; Pi and Mickleborough 1989). Linear stochastic models can be composed of two subsystems: a deterministic part, which represents an exact mathematical formulation, and a stochastic part, which defines the influence and behavior of random variables (Chang and Pakzad 2013).

This paper is focused on structural modal identification, more specifically, the theoretical precision of maximum likelihood estimators (MLEs), a common class of statistical estimators for modal properties (Andersen; Guillaume et al. 1998; Matarazzo and Pakzad 2015a). The quantification of uncertainty as a result of parameter estimation for a broad class of statistical estimators would prove to be a useful metric for evaluation or selection of identification techniques. The adequacy of an estimator is especially a concern in the rapidly evolving field of SHM; as collected data reach larger sizes and new formats, such as BIGDATA (Matarazzo et al. 2015) or mobile sensing (Matarazzo and Pakzad 2015b), trusted estimation techniques with measureable precision become increasingly valuable.

This work is towards the goal of better understanding how estimated modal parameters may differ from their true values. In a broad sense, it is essential that the variation resulting from the estimation technique is lower than the expected variation of the structural property itself. Without any measure of precision of an estimation approach, this question is left unanswered, leaving practice as the only method to verify the efficiency of such an estimator.

Maximum likelihood (ML) theory provides a popular framework for parameter estimation and model identification of statistical models. With an intuitive goal and measureable performance, ML methods offer an efficient approach. MLEs are a family of statistical estimators with many desirable features provided some regularity conditions. While a long list of these conditions may be solicited for complicated observations [as mentioned by King (1989)], only few are necessary to derive the asymptotic properties of MLE required in this work. More specifically, the proof for consistency, asymptotic normality, and efficiency for the ML estimator assumes the first two derivatives of the likelihood function exist (Stuart et al. 1999). In this paper, such derivatives are computed analytically for model features that influence modal properties.

The following key properties of ML estimates can be derived (Long 1997; Reinert 2009):

- MLEs are consistent;
- MLEs are asymptotically unbiased, but are not necessarily unbiased for finite samples;
- MLEs are asymptotically efficient; they have the smallest asymptotic variance out of all asymptotically unbiased estimators, approximately equal to the Cramér-Rao lower bound; and
- MLEs are asymptotically normally distributed.

[1]Dept. of Civil and Environmental Engineering, Lehigh Univ., 117 ATLSS Dr., Bethlehem, PA 18015 (corresponding author). E-mail: tjm310@lehigh.edu; thomasjmatarazzo@gmail.com

[2]Associate Professor, Dept. of Civil and Environmental Engineering, Lehigh Univ., 117 ATLSS Dr., Bethlehem, PA 18015. E-mail: pakzad@lehigh.edu

Note. This manuscript was submitted on January 1, 2015; approved on May 12, 2015; published online on July 30, 2015. Discussion period open until December 30, 2015; separate discussions must be submitted for individual papers. This paper is part of the ***ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering***, © ASCE, B4015002(15)/$25.00.

Finally, ML estimates that do not satisfy any of the preceding regularity conditions remain useful because they are, by definition, the most likely parameters to have generated the data under the model; such ML estimates may or may not have the asymptotic properties listed previously (King 1989).

In ML estimation theory, the likelihood function contains important information regarding how model parameters represent data. Closed-form likelihood functions and derivatives are valuable because they facilitate enhanced model analyses including precision of point estimates. As King (1989) states, "a measure of the likelihood function's curvature is also a measure of the precision of the ML estimate." As documented in the literature (King 1989; Klein and Neudecker 2000; Rubin et al. 1977; Shumway and Stoffer 2011), the standard errors of the ML point estimators are determined by evaluating the Hessian matrix at the critical likelihood point. It is desirable to determine an analytical solution for information and covariance matrices because they are required by subsequent processes, e.g., likelihood ratio test, Wald test statistic, score test, and interval estimation, especially because the asymptotic variance of the MLE is equal to the Cramér-Rao lower bound.

A majority of studies considering precision of ML estimates have focused on frequency-domain models. The approach in Verboven et al. (2004) used a numerical Jacobian, computed at the final iteration of the Gauss-Newton algorithm, to estimate uncertainty of modal properties. Mahata et al. (2006) analyzed nonparametric noise statistics and derived conditions for consistent estimates. Pintelon et al. (2007) implemented the delta method (Oehlert 1992) to produce asymptotic variance estimates for transfer function model parameters based on a first-order Taylor series likelihood approximation. This method assumed that parameter covariance for system matrices were available along the MLE through the estimation technique, which is rarely the case in state-space ML methods.

In some ML algorithms such as Newton-Raphson or expectation maximization (EM) (Rubin et al. 1977; Shumway and Stoffer 2011), numerical Hessian matrices can be produced at the final iteration. The supplemented EM (SEM) algorithm (Meng and Rubin 1991) provides an approach to computing analytical variance-covariance matrices; however, its applicability to a multivariate state-space model is not covered. It is preferable to have analytical expressions for the sensitivity of the likelihood function with respect to specific state-space model parameters.

Charalambous and Logothetis (2000) developed nonlinear filters to carry out the EM algorithm for continuous-time nonlinear stochastic systems in explicit form; computations for closed-form gradients and Fisher information follow directly from the derivations. There is a volume of work on the uncertainty of identified modal parameters using the Bayesian framework in frequency domain (Au 2014a, b). While frequentist and Bayesian approaches are fundamentally different in theory, they share concepts for measuring the uncertainty of an estimator. For example, the Bayesian fast Fourier transform (FFT) identification method that determines the most probable value (MPV) (Au and Zhang 2012; Au 2011) is analogous to the frequentist, time-domain structural identification using expectation maximization (STRIDE) method that determines maximum likelihood estimates (MLEs) (Matarazzo and Pakzad 2015a) and is considered in this paper. The true relationship between these two perspectives is highly nontrivial and is problem-dependent (Au 2012). This paper provides a deeper insight into the relationship between Bayesian and frequentist parameter uncertainty in system identification, established in Au (2012), by providing additional tools for frequentist approaches.

This paper derives closed-form expected information and covariance matrices for four stochastic state-space model entities: observation matrix, state matrix, eigenvalues of state matrix, and right eigenvectors of state matrix. The importance of the selected model features is that they directly influence modal property estimates of the structural system. Sensitivity metrics on these quantities help engineers better understand the true precision of an estimated modal parameter. Additionally, using the asymptotic covariance matrices and the asymptotic normality property of ML estimators, confidence intervals are constructed for natural frequencies, damping ratios, and mode shapes. The metrics presented in this paper are applicable to modal identification methods within this model class, e.g., auto-regressive (AR) (He and De Roeck 1997), numerical algorithm for subspace state-space system identification (N4SID) (Van Overschee and De Moor 1992), eigenvalue realization algorithm-observer Kalman filter identification of output-only systems (ERA-OKID-OO) (Chang and Pakzad 2013), structural identification using expectation maximization (STRIDE) (Matarazzo and Pakzad 2015a), stochastic subspace identification (SSI) (Peeters and De Roeck 1999), and others.

Conditional Expectation of Log-Likelihood Function

In this section, the stochastic state-space model is briefly reviewed and the corresponding model likelihood is constructed from aleatory model entities. Recall the stochastic state-space model, given in Eqs. (1)–(5), which defines the behavior of a linear discrete time invariant system with time steps $k = 1, 2, \ldots, K$. The size of the observation vector $\mathbf{y}_k$ is O and the size of the state vector $\mathbf{x}_k$ is S; Appendix II gives further model details:

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \boldsymbol{\eta}_k \tag{1}$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \boldsymbol{\upsilon}_k \tag{2}$$

$$\mathbf{x}_1 \sim N(\bar{\boldsymbol{\mu}}, \bar{\mathbf{V}}) \tag{3}$$

$$\boldsymbol{\eta}_k \sim N(\mathbf{0}, \mathbf{Q}) \tag{4}$$

$$\boldsymbol{\upsilon}_k \sim N(\mathbf{0}, \mathbf{R}) \tag{5}$$

The complete-data likelihood function for this model is a product of three independent Gaussian densities [see Hinton and Ghahramani (1996), Matarazzo and Pakzad (2015a), or Shumway and Stoffer (2006) for details]. The log-likelihood function, provided in Eq. (6), represents the same Gaussian mixture as a sum, which is mathematically more convenient than the product. This function is the foundation for determining which set of model parameters best explains the observed data. To ease repeated references, call the collection of all six model parameters the super-parameter $\Psi = (\bar{\boldsymbol{\mu}},\ \bar{\mathbf{V}},\ \mathbf{A},\ \mathbf{Q},\ \mathbf{C},\ \mathbf{R})$ and let the observed data be defined by $Y = \mathbf{y}_1, \ldots, \mathbf{y}_K$ and let the states be $X = \mathbf{x}_1, \ldots, \mathbf{x}_K$:

$$\begin{aligned}\ln[L_{X,Y}(\Psi)] = &-\frac{S}{2}\ln(2\pi) - \frac{1}{2}\ln|\bar{\mathbf{V}}| - \frac{1}{2}(\mathbf{x}_1 - \bar{\boldsymbol{\mu}})^T\bar{\mathbf{V}}^{-1}(\mathbf{x}_1 - \bar{\boldsymbol{\mu}}) \\ &- \frac{KO}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{R}| \\ &- \frac{1}{2}\sum_{k=1}^{K}(\mathbf{y}_k - \mathbf{C}\mathbf{x}_k)^T\mathbf{R}^{-1}(\mathbf{y}_k - \mathbf{C}\mathbf{x}_k) \\ &- \frac{(K-1)S}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{Q}| \\ &- \frac{1}{2}\sum_{k=2}^{K}(\mathbf{x}_k - \mathbf{A}\mathbf{x}_{k-1})^T\mathbf{Q}^{-1}(\mathbf{x}_k - \mathbf{A}\mathbf{x}_{k-1})\end{aligned} \tag{6}$$

Eq. (6) implies complete data, i.e., it assumes state variables are available for this likelihood calculation, which is not the case in state-space. Subsequent sensitivity metrics in this paper assume the state-space model has been identified by an ML method, i.e., $\Psi = \hat{\Psi}$, so that MLEs of the model parameters are available. After identification, the state variable can be computed using the data and ML parameters. For example, with the MLE model parameters, expected state variables and sufficient state statistics can be estimated through Kalman filter (Kalman 1960; Shumway and Stoffer 1981) and Rauch-Tung-Striebel (RTS) smoother recursions (Rauch et al. 1965). As a result, subsequent gradients refer to the conditional expectation of the log-likelihood function at the critical point, given by Eq. (7):

$$G(\Psi)_{\Psi=\hat{\Psi}} = E\{\ln[L_{X,Y}(\hat{\Psi})]\} \tag{7}$$

In summary, the difference between $G(\Psi)$ and $\ln[L_{X,Y}(\Psi)]$ lies in the conditional expectation of the state variable and its variances, given the data and ML parameters. If the estimation of the state variable and its variances is exact, the two log-likelihood functions are identical at the critical point.

Information Metrics

This section discusses the relationship between the information contained within a model parameter and the likelihood function, which is determined by the curvature of the likelihood function. The observed information matrix is the negative expectation of the Hessian [denoted by $H()$] of the likelihood function at the critical point, $G = G(\Psi)_{\Psi=\hat{\Psi}}$, with respect to a parameter. Eq. (8) illustrates the observed information for an MLE element, $\psi \in \Psi = \hat{\Psi}$. The Hessian is the second partial derivative of G:

$$I(\psi) \equiv -E[H(\psi)] = -E\left[\frac{\partial^2 G}{\partial \psi^2}\right] \tag{8}$$

The following subsections focus on the preceding relationship to determine closed-form expected information matrices, then parameter covariances, for four model features: observation matrix, state matrix, eigenvalues of state matrix, and right eigenvectors of state matrix. Derivations begin with appropriate first partial derivatives of the expected likelihood function. Because the likelihood function is a scalar value, the resulting first partial derivatives terms will have the same matrix dimensions as the dependent variable, e.g., $\mathbf{C}$ and $\partial G/\partial \mathbf{C}$ have identical dimensions just as λ_d and $\partial G/\partial \lambda_d$ are both scalars. Second partial derivative terms require the derivative of a matrix with respect to another matrix; these matrices will have dimensions larger than the dependent variable (see Appendix V for details). To present this result in closed-form, numerous linear algebra concepts and indexing techniques are utilized, e.g., Kronecker product, single-entry matrix, matrix vectorization, or the matrix trace. For best comprehension of the mathematics within this paper, see Pedersen et al. (2012). Additional details for selected derivations are provided in Appendix I.

Observed Information of Observation Matrix

Following the definition in Eq. (8), the information for the observation matrix $\mathbf{C}$ is straightforward:

$$I(\mathbf{C}) \equiv -E[H(\mathbf{C})] = -E\left[\frac{\partial^2 G}{\partial \mathbf{C} \partial \mathbf{C}^T}\right] \tag{9}$$

The derivation of the observation matrix information begins with the first partial derivative of the expected likelihood function G. Using the symmetry of $\mathbf{R}$ and $\mathbf{R}^{-1}$, the resulting partial derivative $\partial G/\partial \mathbf{C}$ is a first-order sensitivity matrix with size $O \times S$:

$$\begin{aligned}
\frac{\partial G}{\partial \mathbf{C}} &= -\frac{1}{2}\sum_{k=1}^{K} \frac{\partial}{\partial \mathbf{C}} \{E[(\mathbf{y}_k - \mathbf{C}\mathbf{x}_k)^T \mathbf{R}^{-1}(\mathbf{y}_k - \mathbf{C}\mathbf{x}_k)]\} \\
&= -\frac{1}{2}\sum_{k=1}^{K} -2\mathbf{R}^{-1}(E[\mathbf{y}_k \mathbf{x}_k^T] - \mathbf{C}E[\mathbf{x}_k \mathbf{x}_k^T]) \\
&= \sum_{k=1}^{K} \mathbf{R}^{-1}(E[\mathbf{y}_k \mathbf{x}_k^T] - \mathbf{C}E[\mathbf{x}_k \mathbf{x}_k^T]) \\
&= \mathbf{R}^{-1}\sum_{k=1}^{K} E[\mathbf{y}_k \mathbf{x}_k^T] - \mathbf{R}^{-1}\mathbf{C}\sum_{k=1}^{K} E[\mathbf{x}_k \mathbf{x}_k^T]
\end{aligned} \tag{10}$$

The first line of Eq. (10) comes from the expected value of the sixth term in Eq. (6); the second line follows directly from Eq. (88) in Pedersen et al. (2012). For the Hessian with respect to the observation matrix, the derivative of $\partial G/\partial \mathbf{C}$ must be determined with respect to each element of the observation matrix:

$$\begin{aligned}
\frac{\partial^2 G}{\partial \mathbf{C} \partial \mathbf{C}_{ij}} &= \frac{\partial}{\partial \mathbf{C}_{ij}}\left(\mathbf{R}^{-1}\sum_{k=1}^{K} E[\mathbf{y}_k \mathbf{x}_k^T] - \mathbf{R}^{-1}\mathbf{C}\sum_{k=1}^{K} E[\mathbf{x}_k \mathbf{x}_k^T]\right) \\
&= -\mathbf{R}^{-1}\frac{\partial \mathbf{C}}{\partial \mathbf{C}_{ij}}\sum_{k=1}^{K} E[\mathbf{x}_k \mathbf{x}_k^T] \\
&= -\mathbf{R}^{-1}\boldsymbol{\delta}_{ij}^{O\times S}\sum_{k=1}^{K} E[\mathbf{x}_k \mathbf{x}_k^T]
\end{aligned} \tag{11}$$

where $\boldsymbol{\delta}_{ij}^{O\times S}$ is an $O \times S$ single-entry matrix [a matrix with zeros everywhere except unity at entry ij (Pedersen et al. 2012)] representing the derivative of the observation matrix with respect to one of its entries. The following Hessian is constructed by calculating Eq. (11) for every element of the observation matrix, producing Eq. (12):

$$\frac{\partial^2 G}{\partial \mathbf{C} \partial \mathbf{C}_{ij}} \equiv \begin{bmatrix} \frac{\partial^2 G}{\partial \mathbf{C}_{11} \partial \mathbf{C}_{ij}} & \frac{\partial^2 G}{\partial \mathbf{C}_{12} \partial \mathbf{C}_{ij}} & \cdots & \frac{\partial^2 G}{\partial \mathbf{C}_{1S} \partial \mathbf{C}_{ij}} \\ \frac{\partial^2 G}{\partial \mathbf{C}_{21} \partial \mathbf{C}_{ij}} & \frac{\partial^2 G}{\partial \mathbf{C}_{22} \partial \mathbf{C}_{ij}} & & \vdots \\ \vdots & & \ddots & \\ \frac{\partial^2 G}{\partial \mathbf{C}_{O1} \partial \mathbf{C}_{ij}} & \cdots & & \frac{\partial^2 G}{\partial \mathbf{C}_{OS} \partial \mathbf{C}_{ij}} \end{bmatrix} \tag{12}$$

Next, the preceding element-wise derivative matrices are vectorized and become the columns of the $OS \times OS$ full Hessian (block) matrix (where OS is the product of O and S), $\partial^2 G/(\partial \mathbf{C} \partial \mathbf{C}^T)$, in Eq. (13). When vectorized, each ij row-column coordinate is mapped to a row-wise linear index (Eddins and Shure 2001), i.e., $\text{vec}[\partial^2 G/(\partial \mathbf{C} \partial \mathbf{C}_L)]$ for $L = 1, 2, \ldots, OS$:

$$\frac{\partial^2 G}{\partial \mathbf{C} \partial \mathbf{C}^T} \equiv \left[\text{vec}\left(\frac{\partial^2 G}{\partial \mathbf{C} \partial \mathbf{C}_1}\right) \text{vec}\left(\frac{\partial^2 G}{\partial \mathbf{C} \partial \mathbf{C}_2}\right) \cdots \text{vec}\left(\frac{\partial^2 G}{\partial \mathbf{C} \partial \mathbf{C}_{OS}}\right)\right] \tag{13}$$

Alternatively, because both the state and the observation error covariance are real, this Hessian can be expressed simply using the Kronecker product:

$$\frac{\partial^2 G}{\partial \mathbf{C}\partial \mathbf{C}^T} = \sum_{k=1}^{K} E[\mathbf{x}_k \mathbf{x}_k^T] \otimes (-\mathbf{R}^{-1})^T \tag{14}$$

Finally, the observed information of the observation matrix is obtained:

$$I(\mathbf{C}) = -E\left[\frac{\partial^2 G}{\partial \mathbf{C}\partial \mathbf{C}^T}\right] \tag{15}$$

Observed Information of State Matrix

The state matrix information follows from the definition in Eq. (8):

$$I(A) \equiv -E[H(\mathbf{A})] = -E\left[\frac{\partial^2 G}{\partial \mathbf{A}\partial \mathbf{A}^T}\right] \tag{16}$$

The first partial derivative of the likelihood function with respect to the state matrix is given in the following, utilizing symmetry of input covariance, $\mathbf{Q}$ and $\mathbf{Q}^{-1}$. Then, the second partial derivative is computed, obtaining the likelihood Hessian with respect to the state matrix:

$$\begin{aligned}\frac{\partial G}{\partial \mathbf{A}} &= -\frac{1}{2}\sum_{k=2}^{K}\frac{\partial}{\partial \mathbf{A}}\{E[(\mathbf{x}_k - \mathbf{A}\mathbf{x}_{k-1})^T\mathbf{Q}^{-1}(\mathbf{x}_k - \mathbf{A}\mathbf{x}_{k-1})]\}\\ &= -\frac{1}{2}\sum_{k=2}^{K} -2\mathbf{Q}^{-1}(E[\mathbf{x}_k\mathbf{x}_{k-1}^T] - \mathbf{A}E[\mathbf{x}_{k-1}\mathbf{x}_{k-1}^T])\\ &= \sum_{k=2}^{K}\mathbf{Q}^{-1}(E[\mathbf{x}_k\mathbf{x}_{k-1}^T] - \mathbf{A}E[\mathbf{x}_{k-1}\mathbf{x}_{k-1}^T])\\ &= \mathbf{Q}^{-1}\sum_{k=2}^{K}E[\mathbf{x}_k\mathbf{x}_{k-1}^T] - \mathbf{Q}^{-1}\mathbf{A}\sum_{k=2}^{K}E[\mathbf{x}_{k-1}\mathbf{x}_{k-1}^T]\end{aligned} \tag{17}$$

The previous derivative is similar in form to Eq. (10): the first line of Eq. (17) uses the expected value of the last line in Eq. (6); the second line in Eq. (17) can be computed using Eq. (88) in Pedersen et al. (2012):

$$\begin{aligned}\frac{\partial^2 G}{\partial \mathbf{A}\partial \mathbf{A}_{ij}} &= \frac{\partial}{\partial \mathbf{A}_{ij}}\left(\mathbf{Q}^{-1}\sum_{k=2}^{K}E[\mathbf{x}_k\mathbf{x}_{k-1}^T] - \mathbf{Q}^{-1}\mathbf{A}\sum_{k=2}^{K}E[\mathbf{x}_{k-1}\mathbf{x}_{k-1}^T]\right)\\ &= -\mathbf{Q}^{-1}\frac{\partial \mathbf{A}}{\partial \mathbf{A}_{ij}}\sum_{k=2}^{K}E[\mathbf{x}_{k-1}\mathbf{x}_{k-1}^T]\\ &= -\mathbf{Q}^{-1}\boldsymbol{\delta}_{ij}^{S\times S}\sum_{k=2}^{K}E[\mathbf{x}_{k-1}\mathbf{x}_{k-1}^T]\end{aligned} \tag{18}$$

where $\boldsymbol{\delta}_{ij}^{S\times S}$ is an $S \times S$ single-entry matrix. The matrix provided in Eq. (18) is calculated for each element of the state matrix $\mathbf{A}_{ij}$ and constitutes a block matrix element of the full Hessian:

$$\frac{\partial^2 G}{\partial \mathbf{A}\partial \mathbf{A}_{ij}} \equiv \begin{bmatrix} \frac{\partial^2 G}{\partial \mathbf{A}_{11}\partial \mathbf{A}_{ij}} & \frac{\partial^2 G}{\partial \mathbf{A}_{12}\partial \mathbf{A}_{ij}} & \cdots & \frac{\partial^2 G}{\partial \mathbf{A}_{1S}\partial \mathbf{A}_{ij}} \\ \frac{\partial^2 G}{\partial \mathbf{A}_{21}\partial \mathbf{A}_{ij}} & \frac{\partial^2 G}{\partial \mathbf{A}_{22}\partial \mathbf{A}_{ij}} & & \vdots \\ \vdots & & \ddots & \\ \frac{\partial^2 G}{\partial \mathbf{A}_{S1}\partial \mathbf{A}_{ij}} & \cdots & & \frac{\partial^2 G}{\partial \mathbf{A}_{SS}\partial \mathbf{A}_{ij}} \end{bmatrix} \tag{19}$$

The preceding element-wise derivative matrices are vectorized, reindexed, and arranged as columns to form the $S^2 \times S^2$ full Hessian matrix in Eq. (20). When vectorized, row-wise linear indexes replace row-column coordinates, i.e., $\text{vec}[\partial^2 G/(\partial \mathbf{A}\partial \mathbf{A}_L)] = \text{vec}[\partial^2 G/(\partial \mathbf{A}\partial \mathbf{A}_{ij})]$ for $L = 1, 2, \ldots, S^2$:

$$\frac{\partial^2 \mathbf{G}}{\partial \mathbf{A}\partial \mathbf{A}^T} \equiv \left[\text{vec}\left(\frac{\partial^2 \mathbf{G}}{\partial \mathbf{A}\partial \mathbf{A}_1}\right)\text{vec}\left(\frac{\partial^2 \mathbf{G}}{\partial \mathbf{A}\partial \mathbf{A}_2}\right) \cdots \text{vec}\left(\frac{\partial^2 \mathbf{G}}{\partial \mathbf{A}\partial \mathbf{A}_{S^2}}\right)\right] \tag{20}$$

This Hessian can further be expressed in a simpler form via use of the Kronecker product because the states and state input covariance are both real:

$$\frac{\partial^2 G}{\partial \mathbf{A}\partial \mathbf{A}^T} = \sum_{k=2}^{K} E[\mathbf{x}_{k-1}\mathbf{x}_{k-1}^T] \otimes (-\mathbf{Q}^{-1})^T \tag{21}$$

Finally, the observed information of the state matrix is available:

$$I(\mathbf{A}) = -E\left[\frac{\partial^2 G}{\partial \mathbf{A}\partial \mathbf{A}^T}\right] \tag{22}$$

Observed Information of State Matrix Eigenvalues

The final two subsections focus on eigenfeatures of the state matrix. Consider the eigendecomposition of the state matrix $\mathbf{A} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Theta}$, where $\boldsymbol{\Gamma}$ is the matrix of right eigenvectors, $\boldsymbol{\Lambda}$ is the diagonal eigenvalue matrix, and $\boldsymbol{\Theta}$ is the matrix of left eigenvectors ($\boldsymbol{\Theta} = \boldsymbol{\Gamma}^{-1}$). As previously, the information of the eigenvalues of $\mathbf{A}$ follows from the definition in Eq. (8):

$$I(\boldsymbol{\Lambda}) \equiv -E[H(\boldsymbol{\Lambda})] = -E\left[\frac{\partial^2 G}{\partial \boldsymbol{\Lambda}\partial \boldsymbol{\Lambda}^T}\right] \tag{23}$$

The likelihood score with respect to a state matrix eigenvalue is a diagonal matrix because the eigenvalue matrix is diagonal:

$$\frac{\partial G}{\partial \boldsymbol{\Lambda}} \equiv \begin{bmatrix} \frac{\partial G}{\partial \lambda_1} & 0 & 0 & 0 \\ 0 & \frac{\partial G}{\partial \lambda_2} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \frac{\partial G}{\partial \lambda_S} \end{bmatrix} \tag{24}$$

In the preceding equation, the eigenvalues λ_d, for $d = 1,2, \ldots, S$, are the diagonal elements of the eigenvalue matrix $\boldsymbol{\Lambda}$. In Eq. (25), the chain rule is implemented to formulate the likelihood score with respect to an individual state matrix eigenvalue because $\mathbf{A} = f(\boldsymbol{\Lambda})$. The value of $\partial G/\partial \lambda_d$ is a scalar, while $\partial G/\partial \mathbf{A}$ and $\partial \mathbf{A}/\partial \lambda_d$ are both $N \times N$ matrices. In matrix differentiation, the chain rule can be expressed in many forms (Pedersen et al. 2012). For example, the chain rule for the likelihood score $\partial G/\partial \lambda_d$ can be shown as an element-wise summation of products [Eq. (25)]. Alternatively, the inner product of vectorized partial derivatives or the trace of the product of partial derivatives may be used to establish the likelihood score with respect to a state matrix eigenvalue [Eq. (26)]. Throughout this paper, the vectorized representation is implemented:

$$\frac{\partial G}{\partial \lambda_d} = \sum_{i=1}^{S}\sum_{j=1}^{S}\frac{\partial G}{\partial \mathbf{A}_{ij}}\frac{\partial \mathbf{A}_{ij}}{\partial \lambda_d} \tag{25}$$

$$\frac{\partial G}{\partial \lambda_d} = \text{vec}\left(\frac{\partial G}{\partial \mathbf{A}}\right) \cdot \text{vec}\left(\frac{\partial \mathbf{A}}{\partial \lambda_d}\right) = tr\left[\left(\frac{\partial G}{\partial \mathbf{A}}\right)^T \frac{\partial \mathbf{A}}{\partial \lambda_d}\right] \tag{26}$$

The $\partial G/\partial \mathbf{A}$ term in Eq. (26), i.e., the partial derivative of likelihood with respect to the state matrix, has been determined in the previous section. The partial derivative of the state matrix with respect to one of its eigenvalues, the second term in the preceding equation, is computed using the product rule:

$$\frac{\partial \mathbf{A}}{\partial \lambda_d} = \frac{\partial}{\partial \lambda_d}(\mathbf{\Gamma \Lambda \Theta}) = \mathbf{\Gamma}\frac{\partial \mathbf{\Lambda}}{\partial \lambda_d}\mathbf{\Theta} = \mathbf{\Gamma \delta}_d^{S\times S}\mathbf{\Theta} \tag{27}$$

where $\partial \mathbf{\Lambda}/\partial \lambda_d \equiv \mathbf{\delta}_d^{S\times S}$ is an $S \times S$ single-entry matrix (a matrix with zeros everywhere except unity at the dth diagonal). The diagonal nature of $\mathbf{\Lambda}$ means single-entry matrices for nondiagonal terms are equal to $S \times S$ matrices of zeros, i.e., $\mathbf{\delta}_{L\neq d}^{S\times S} = \mathbf{0}^{S\times S}$, verifying that $\partial G/\partial \mathbf{\Lambda}$ is diagonal as presented in Eq. (24). The term $\partial \mathbf{A}/\partial \mathbf{\Lambda}$ is an $S^2 \times S^2$ block diagonal matrix containing $S \times S$ matrix elements as shown in Eq. (28):

$$\frac{\partial \mathbf{A}}{\partial \mathbf{\Lambda}} \equiv \begin{bmatrix} \left[\frac{\partial \mathbf{A}}{\partial \lambda_1}\right] & \mathbf{0}^{S\times S} & \mathbf{0}^{S\times S} & \mathbf{0}^{S\times S} \\ \mathbf{0}^{S\times S} & \left[\frac{\partial \mathbf{A}}{\partial \lambda_2}\right] & \mathbf{0}^{S\times S} & \mathbf{0}^{S\times S} \\ \mathbf{0}^{S\times S} & \mathbf{0}^{S\times S} & \ddots & \mathbf{0}^{S\times S} \\ \mathbf{0}^{S\times S} & \mathbf{0}^{S\times S} & \mathbf{0}^{S\times S} & \left[\frac{\partial \mathbf{A}}{\partial \lambda_S}\right] \end{bmatrix} \tag{28}$$

For the diagonal elements of $\partial G/\partial \mathbf{\Lambda}$, the terms of Eq. (25) are obtained through Eqs. (17) and (27):

$$\frac{\partial G}{\partial \lambda_d} = \text{vec}\left(\mathbf{Q}^{-1}\sum_{k=2}^{K} E[\mathbf{x}_k \mathbf{x}_{k-1}^T] - \mathbf{Q}^{-1}\mathbf{A}\sum_{k=2}^{K} E[\mathbf{x}_{k-1}\mathbf{x}_{k-1}^T]\right) \cdot \text{vec}(\mathbf{\Gamma \delta}_d^{S\times S}\mathbf{\Theta}) \tag{29}$$

Appendix I has a discussion on the role of Eq. (29) in the M-step of the EM algorithm for state-space models (Matarazzo and Pakzad 2015a; Shumway and Stoffer 1982). The Hessian of G with respect to the diagonals of $\mathbf{\Lambda}$ is an $S \times S$ matrix that describes the likelihood function curvature with respect to a state matrix eigenvalue:

$$\frac{\partial^2 G}{\partial \mathbf{\Lambda}\partial \mathbf{\Lambda}^T} \equiv \begin{bmatrix} \frac{\partial^2 G}{\partial \lambda_1 \partial \lambda_1} & \frac{\partial^2 G}{\partial \lambda_1 \partial \lambda_2} & \cdots & \frac{\partial^2 G}{\partial \lambda_1 \partial \lambda_S} \\ \frac{\partial^2 G}{\partial \lambda_2 \partial \lambda_1} & \frac{\partial^2 G}{\partial \lambda_2 \partial \lambda_2} & & \vdots \\ \vdots & & \ddots & \\ \frac{\partial^2 G}{\partial \lambda_S \partial \lambda_1} & \cdots & & \frac{\partial^2 G}{\partial \lambda_S \partial \lambda_S} \end{bmatrix} \tag{30}$$

The Hessian entries are in terms of two diagonal elements of $\mathbf{\Lambda}$, $\partial^2 G/(\partial \lambda_d \partial \lambda_h)$, with $d, h = 1,2, \ldots, S$. The chain rule is applied to compute $\partial^2 G/(\partial \lambda_d \partial \lambda_h)$, a scalar value:

$$\frac{\partial^2 G}{\partial \lambda_d \partial \lambda_h} = \frac{\partial}{\partial \lambda_h}\left(\frac{\partial G}{\partial \lambda_d}\right) \tag{31}$$

Implementing the product rule and temporarily omitting the vec() operator to simplify notation used within the definition of $\partial G/\partial \lambda_d$ in Eq. (29), appropriate matrices will be vectorized after simplification:

$$\frac{\partial^2 G}{\partial \lambda_d \partial \lambda_h} = \frac{\partial}{\partial \lambda_h}\left(\frac{\partial G}{\partial \mathbf{A}}\right)\frac{\partial \mathbf{A}}{\partial \lambda_d} + \frac{\partial G}{\partial \mathbf{A}}\frac{\partial^2 \mathbf{A}}{\partial \lambda_d \partial \lambda_h} \tag{32}$$

The term $\partial^2 \mathbf{A}/(\partial \lambda_d \partial \lambda_h) = \mathbf{0}^{S\times S}$, so that only the first term of Eq. (32) remains. The likelihood curvature with respect to state matrix eigenvalues can be calculated via Eq. (33), thus constructing the elements of the full Hessian in Eq. (30):

$$\begin{aligned} \frac{\partial^2 G}{\partial \lambda_d \partial \lambda_h} &= \frac{\partial}{\partial \lambda_h}\left(\frac{\partial G}{\partial \mathbf{A}}\right)\frac{\partial \mathbf{A}}{\partial \lambda_d} \\ &= \frac{\partial}{\partial \lambda_h}\left(\mathbf{Q}^{-1}\sum_{k=2}^{K} E[\mathbf{x}_k \mathbf{x}_{k-1}^T] - \mathbf{Q}^{-1}\mathbf{A}\sum_{k=2}^{K} E[\mathbf{x}_{k-1}\mathbf{x}_{k-1}^T]\right)\mathbf{\Gamma \delta}_d^{S\times S}\mathbf{\Theta} \\ &= \left(-\mathbf{Q}^{-1}\mathbf{\Gamma \delta}_h^{S\times S}\mathbf{\Theta}\sum_{k=2}^{K} E[\mathbf{x}_{k-1}\mathbf{x}_{k-1}^T]\right)\mathbf{\Gamma \delta}_d^{S\times S}\mathbf{\Theta}^{-1} \\ &= \text{vec}\left(-\mathbf{Q}^{-1}\mathbf{\Gamma \delta}_h^{S\times S}\mathbf{\Theta}\sum_{k=2}^{K} E[\mathbf{x}_{k-1}\mathbf{x}_{k-1}^T]\right) \cdot \text{vec}(\mathbf{\Gamma \delta}_d^{S\times S}\mathbf{\Theta}) \end{aligned} \tag{33}$$

Finally, the information matrix for the state matrix eigenvalues is obtained:

$$I(\mathbf{\Lambda}) = -E\left[\frac{\partial^2 G}{\partial \mathbf{\Lambda}\partial \mathbf{\Lambda}^T}\right] \tag{34}$$

Observed Information of State Matrix Eigenvectors

Following the definition of parameter information, the observed information of the right eigenvectors of the state matrix is given in the following:

$$I(\mathbf{\Gamma}) \equiv -E[H(\mathbf{\Gamma})] = -E\left[\frac{\partial^2 G}{\partial \mathbf{\Gamma}\partial \mathbf{\Gamma}^T}\right] \tag{35}$$

Repeating the procedures of the previous sections, the first partial derivative of the likelihood function with respect to the eigenvector matrix is computed:

$$\frac{\partial G}{\partial \mathbf{\Gamma}} \equiv \begin{bmatrix} \frac{\partial G}{\partial \mathbf{\Gamma}_{11}} & \frac{\partial G}{\partial \mathbf{\Gamma}_{12}} & \cdots & \frac{\partial G}{\partial \mathbf{\Gamma}_{1S}} \\ \frac{\partial G}{\partial \mathbf{\Gamma}_{21}} & \frac{\partial G}{\partial \mathbf{\Gamma}_{22}} & & \vdots \\ \vdots & & \ddots & \\ \frac{\partial G}{\partial \mathbf{\Gamma}_{S1}} & \cdots & & \frac{\partial G}{\partial \mathbf{\Gamma}_{SS}} \end{bmatrix} \tag{36}$$

Each element of Eq. (36) is the likelihood score with respect to an element of $\mathbf{\Gamma}$, $\partial G/\partial \mathbf{\Gamma}_{ij}$ with $i, j = 1,2, \ldots, S$. Similar to the eigenvalue scores, implementing the chain rule, the inner product of vectorized partial derivative matrices is found as

ASCE-ASME J. Risk Uncertainty Eng. Syst., Part A: Civ. Eng.

$$\frac{\partial G}{\partial \boldsymbol{\Gamma}_{ij}} = \text{vec}\left(\frac{\partial G}{\partial \mathbf{A}}\right) \cdot \text{vec}\left(\frac{\partial \mathbf{A}}{\partial \boldsymbol{\Gamma}_{ij}}\right) \tag{37}$$

The product rule is then used to expand on the second term of Eq. (37):

$$\frac{\partial \mathbf{A}}{\partial \boldsymbol{\Gamma}_{ij}} = \frac{\partial \mathbf{A}}{\partial \boldsymbol{\Gamma}_{ij}}(\boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Theta}) = \frac{\partial \boldsymbol{\Gamma}}{\partial \boldsymbol{\Gamma}_{ij}}\boldsymbol{\Lambda}\boldsymbol{\Theta} + \boldsymbol{\Gamma}\boldsymbol{\Lambda}\frac{\partial \boldsymbol{\Theta}}{\partial \boldsymbol{\Gamma}_{ij}} \tag{38}$$

where $\partial\boldsymbol{\Gamma}/\partial\boldsymbol{\Gamma}_{ij} \equiv \boldsymbol{\delta}_{ij}^{S\times S}$ is an $S \times S$ single-entry matrix with zeros everywhere except unity at entry ij. The derivative of a left eigenvector element $\boldsymbol{\Theta}_{rc}$ with respect to a right eigenvector element $\boldsymbol{\Gamma}_{ij}$ is formulated below using indexes $r, c = 1,2, \ldots, S$:

$$\frac{\partial \boldsymbol{\Theta}_{rc}}{\partial \boldsymbol{\Gamma}_{ij}} \equiv \left[\frac{\partial \boldsymbol{\Theta}}{\partial \boldsymbol{\Gamma}_{ij}}\right]_{rc} \tag{39}$$

The corresponding block matrix $\partial\boldsymbol{\Theta}/\partial\boldsymbol{\Gamma}$ is constructed from the following element matrices:

$$\frac{\partial \boldsymbol{\Theta}}{\partial \boldsymbol{\Gamma}_{ij}} \equiv \begin{bmatrix} \frac{\partial \boldsymbol{\Theta}_{11}}{\partial \boldsymbol{\Gamma}_{ij}} & \frac{\partial \boldsymbol{\Theta}_{12}}{\partial \boldsymbol{\Gamma}_{ij}} & \cdots & \frac{\partial \boldsymbol{\Theta}_{1S}}{\partial \boldsymbol{\Gamma}_{ij}} \\ \frac{\partial \boldsymbol{\Theta}_{21}}{\partial \boldsymbol{\Gamma}_{ij}} & \frac{\partial \boldsymbol{\Theta}_{22}}{\partial \boldsymbol{\Gamma}_{ij}} & & \vdots \\ \vdots & & \ddots & \\ \frac{\partial \boldsymbol{\Theta}_{S1}}{\partial \boldsymbol{\Gamma}_{ij}} & \cdots & & \frac{\partial \boldsymbol{\Theta}_{SS}}{\partial \boldsymbol{\Gamma}_{ij}} \end{bmatrix} = \begin{bmatrix} -\boldsymbol{\Theta}_{1i}\boldsymbol{\Theta}_{j1} & -\boldsymbol{\Theta}_{1i}\boldsymbol{\Theta}_{j2} & \cdots & -\boldsymbol{\Theta}_{1i}\boldsymbol{\Theta}_{jS} \\ -\boldsymbol{\Theta}_{2i}\boldsymbol{\Theta}_{j1} & -\boldsymbol{\Theta}_{2i}\boldsymbol{\Theta}_{j2} & & \vdots \\ \vdots & & \ddots & \\ -\boldsymbol{\Theta}_{Si}\boldsymbol{\Theta}_{j1} & \cdots & & -\boldsymbol{\Theta}_{Si}\boldsymbol{\Theta}_{jS} \end{bmatrix} \tag{40}$$

The column of the first term and the row of the second term in the left eigenvector product are i and j, respectively; the row of the first term and the column of the second term in the left eigenvector product are defined by their location within the score matrix, i.e., $\partial\boldsymbol{\Theta}_{rc}/\partial\boldsymbol{\Gamma}_{ij} \equiv -\boldsymbol{\Theta}_{ri}\boldsymbol{\Theta}_{jc}$. With all terms in Eq. (40) defined, the evaluation of the likelihood score with respect to a right eigenvector element, a scalar value, is available:

$$\frac{\partial \mathbf{A}}{\partial \boldsymbol{\Gamma}_{ij}} = \boldsymbol{\delta}_{ij}^{S\times S}\boldsymbol{\Lambda}\boldsymbol{\Theta} + \boldsymbol{\Gamma}\boldsymbol{\Lambda}\frac{\partial \boldsymbol{\Theta}}{\partial \boldsymbol{\Gamma}_{ij}} \tag{41}$$

$$\begin{aligned} \frac{\partial G}{\partial \boldsymbol{\Gamma}_{ij}} &= \text{vec}\left(\frac{\partial G}{\partial \mathbf{A}}\right) \cdot \text{vec}\left(\frac{\partial \mathbf{A}}{\partial \boldsymbol{\Gamma}_{ij}}\right) \\ &= \text{vec}\left(\mathbf{Q}^{-1}\sum_{k=2}^{K} E[\mathbf{x}_k\mathbf{x}_{k-1}^T] - \mathbf{Q}^{-1}\mathbf{A}\sum_{k=2}^{K} E[\mathbf{x}_{k-1}\mathbf{x}_{k-1}^T]\right) \\ &\quad \cdot \text{vec}\left(\boldsymbol{\delta}_{ij}^{S\times S}\boldsymbol{\Lambda}\boldsymbol{\Theta} + \boldsymbol{\Gamma}\boldsymbol{\Lambda}\frac{\partial \boldsymbol{\Theta}}{\partial \boldsymbol{\Gamma}_{ij}}\right) \end{aligned} \tag{42}$$

Next, the Hessian of G with respect to a right eigenvector element is defined as a matrix:

$$\frac{\partial^2 G}{\partial \boldsymbol{\Gamma}\partial \boldsymbol{\Gamma}_{mn}} \equiv \begin{bmatrix} \frac{\partial^2 G}{\partial \boldsymbol{\Gamma}_{11}\partial \boldsymbol{\Gamma}_{mn}} & \frac{\partial^2 G}{\partial \boldsymbol{\Gamma}_{12}\partial \boldsymbol{\Gamma}_{mn}} & \cdots & \frac{\partial^2 G}{\partial \boldsymbol{\Gamma}_{1S}\partial \boldsymbol{\Gamma}_{mn}} \\ \frac{\partial^2 G}{\partial \boldsymbol{\Gamma}_{21}\partial \boldsymbol{\Gamma}_{mn}} & \frac{\partial^2 G}{\partial \boldsymbol{\Gamma}_{22}\partial \boldsymbol{\Gamma}_{mn}} & & \vdots \\ \vdots & & \ddots & \\ \frac{\partial^2 G}{\partial \boldsymbol{\Gamma}_{S1}\partial \boldsymbol{\Gamma}_{mn}} & \cdots & & \frac{\partial^2 G}{\partial \boldsymbol{\Gamma}_{SS}\partial \boldsymbol{\Gamma}_{mn}} \end{bmatrix} \tag{43}$$

In general, the preceding Hessian entries with respect to two elements of $\boldsymbol{\Gamma}$ are $\partial^2 G/(\partial\boldsymbol{\Gamma}_{ij}\partial\boldsymbol{\Gamma}_{mn}) \equiv \partial(\partial G/\partial\boldsymbol{\Gamma}_{ij})/\partial\boldsymbol{\Gamma}_{mn}$ with $i, j, m, n = 1,2, \ldots, S$. For calculation of the Hessian entries implement the chain rule and temporarily omit the vec() operator in the definition of $\partial G/\partial\boldsymbol{\Gamma}_{ij}$ to simplify notation:

$$\begin{aligned} \frac{\partial^2 G}{\partial \boldsymbol{\Gamma}_{ij}\partial \boldsymbol{\Gamma}_{mn}} &= \frac{\partial}{\partial \boldsymbol{\Gamma}_{mn}}\left(\frac{\partial G}{\partial \mathbf{A}}\frac{\partial \mathbf{A}}{\partial \boldsymbol{\Gamma}_{ij}}\right) \\ &= \frac{\partial}{\partial \boldsymbol{\Gamma}_{mn}}\left(\frac{\partial G}{\partial \mathbf{A}}\right)\frac{\partial \mathbf{A}}{\partial \boldsymbol{\Gamma}_{ij}} + \frac{\partial G}{\partial \mathbf{A}}\frac{\partial}{\partial \boldsymbol{\Gamma}_{mn}}\left(\frac{\partial \mathbf{A}}{\partial \boldsymbol{\Gamma}_{ij}}\right) \end{aligned} \tag{44}$$

The first term of Eq. (44) is expanded:

$$\begin{aligned} &\frac{\partial}{\partial \boldsymbol{\Gamma}_{mn}}\left(\frac{\partial G}{\partial \mathbf{A}}\right)\frac{\partial \mathbf{A}}{\partial \boldsymbol{\Gamma}_{ij}} \\ &= \frac{\partial}{\partial \boldsymbol{\Gamma}_{mn}}\left(\mathbf{Q}^{-1}\sum_{k=2}^{K} E[\mathbf{x}_k\mathbf{x}_{k-1}^T] - \mathbf{Q}^{-1}\mathbf{A}\sum_{k=2}^{K} E[\mathbf{x}_{k-1}\mathbf{x}_{k-1}^T]\right) \\ &\quad \times \left(\boldsymbol{\delta}_{ij}^{S\times S}\boldsymbol{\Lambda}\boldsymbol{\Theta} + \boldsymbol{\Gamma}\boldsymbol{\Lambda}\frac{\partial \boldsymbol{\Theta}}{\partial \boldsymbol{\Gamma}_{ij}}\right) \\ &= -\mathbf{Q}^{-1}\frac{\partial}{\partial \boldsymbol{\Gamma}_{mn}}(\boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Theta})\sum_{k=2}^{K} E[\mathbf{x}_{k-1}\mathbf{x}_{k-1}^T]\left(\boldsymbol{\delta}_{ij}^{S\times S}\boldsymbol{\Lambda}\boldsymbol{\Theta} + \boldsymbol{\Gamma}\boldsymbol{\Lambda}\frac{\partial \boldsymbol{\Theta}}{\partial \boldsymbol{\Gamma}_{ij}}\right) \\ &= -\mathbf{Q}^{-1}\left(\boldsymbol{\delta}_{mn}^{S\times S}\boldsymbol{\Lambda}\boldsymbol{\Theta} + \boldsymbol{\Gamma}\boldsymbol{\Lambda}\frac{\partial \boldsymbol{\Theta}}{\partial \boldsymbol{\Gamma}_{mn}}\right)\sum_{k=2}^{K} E[\mathbf{x}_{k-1}\mathbf{x}_{k-1}^T] \\ &\quad \times \left(\boldsymbol{\delta}_{ij}^{S\times S}\boldsymbol{\Lambda}\boldsymbol{\Theta} + \boldsymbol{\Gamma}\boldsymbol{\Lambda}\frac{\partial \boldsymbol{\Theta}}{\partial \boldsymbol{\Gamma}_{ij}}\right) \end{aligned} \tag{45}$$

where $\boldsymbol{\delta}_{mn}^{S\times S}$ is an $S \times S$ single-entry matrix with zeros everywhere except unity at entry mn.

The second term of Eq. (44) is expanded:

$$\begin{aligned} &\frac{\partial G}{\partial \mathbf{A}}\frac{\partial}{\partial \boldsymbol{\Gamma}_{mn}}\left(\frac{\partial \mathbf{A}}{\partial \boldsymbol{\Gamma}_{ij}}\right) \\ &= \left(\mathbf{Q}^{-1}\sum_{k=2}^{K} E[\mathbf{x}_k\mathbf{x}_{k-1}^T] - \mathbf{Q}^{-1}\mathbf{A}\sum_{k=2}^{K} E[\mathbf{x}_{k-1}\mathbf{x}_{k-1}^T]\right) \\ &\quad \times \frac{\partial}{\partial \boldsymbol{\Gamma}_{mn}}\left(\boldsymbol{\delta}_{ij}^{S\times S}\boldsymbol{\Lambda}\boldsymbol{\Theta} + \boldsymbol{\Gamma}\boldsymbol{\Lambda}\frac{\partial \boldsymbol{\Theta}}{\partial \boldsymbol{\Gamma}_{ij}}\right) \\ &= \left(\mathbf{Q}^{-1}\sum_{k=2}^{K} E[\mathbf{x}_k\mathbf{x}_{k-1}^T] - \mathbf{Q}^{-1}\mathbf{A}\sum_{k=2}^{K} E[\mathbf{x}_{k-1}\mathbf{x}_{k-1}^T]\right) \\ &\quad \times \left(\boldsymbol{\delta}_{ij}^{S\times S}\boldsymbol{\Lambda}\frac{\partial \boldsymbol{\Theta}}{\partial \boldsymbol{\Gamma}_{mn}} + \boldsymbol{\delta}_{mn}^{S\times S}\boldsymbol{\Lambda}\frac{\partial \boldsymbol{\Theta}}{\partial \boldsymbol{\Gamma}_{ij}} + \boldsymbol{\Gamma}\boldsymbol{\Lambda}\frac{\partial^2 \boldsymbol{\Theta}}{\partial \boldsymbol{\Gamma}_{ij}\partial \boldsymbol{\Gamma}_{mn}}\right) \end{aligned} \tag{46}$$

ASCE-ASME J. Risk Uncertainty Eng. Syst., Part A: Civ. Eng.

In this equation:

$$\frac{\partial^2\boldsymbol{\Theta}}{\partial\boldsymbol{\Gamma}_{ij}\partial\boldsymbol{\Gamma}_{mn}} \equiv \begin{bmatrix} \frac{\partial^2\boldsymbol{\Theta}_{11}}{\partial\boldsymbol{\Gamma}_{ij}\partial\boldsymbol{\Gamma}_{mn}} & \frac{\partial^2\boldsymbol{\Theta}_{12}}{\partial\boldsymbol{\Gamma}_{ij}\partial\boldsymbol{\Gamma}_{mn}} & \cdots & \frac{\partial^2\boldsymbol{\Theta}_{1S}}{\partial\boldsymbol{\Gamma}_{11}\partial\boldsymbol{\Gamma}_{mn}} \\ \frac{\partial^2\boldsymbol{\Theta}_{21}}{\partial\boldsymbol{\Gamma}_{ij}\partial\boldsymbol{\Gamma}_{mn}} & \frac{\partial^2\boldsymbol{\Theta}_{22}}{\partial\boldsymbol{\Gamma}_{ij}\partial\boldsymbol{\Gamma}_{mn}} & & \vdots \\ \vdots & & \ddots & \\ \frac{\partial^2\boldsymbol{\Theta}_{S1}}{\partial\boldsymbol{\Gamma}_{ij}\partial\boldsymbol{\Gamma}_{mn}} & \cdots & & \frac{\partial^2\boldsymbol{\Theta}_{SS}}{\partial\boldsymbol{\Gamma}_{ij}\partial\boldsymbol{\Gamma}_{mn}} \end{bmatrix} = \frac{\partial}{\partial\boldsymbol{\Gamma}_{mn}} \begin{bmatrix} -\boldsymbol{\Theta}_{1i}\boldsymbol{\Theta}_{j1} & -\boldsymbol{\Theta}_{1i}\boldsymbol{\Theta}_{j2} & \cdots & -\boldsymbol{\Theta}_{1i}\boldsymbol{\Theta}_{jS} \\ -\boldsymbol{\Theta}_{2i}\boldsymbol{\Theta}_{j1} & -\boldsymbol{\Theta}_{2i}\boldsymbol{\Theta}_{j2} & & \vdots \\ \vdots & & \ddots & \\ -\boldsymbol{\Theta}_{Si}\boldsymbol{\Theta}_{j1} & \cdots & & -\boldsymbol{\Theta}_{Si}\boldsymbol{\Theta}_{jS} \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{\Theta}_{1m}\boldsymbol{\Theta}_{ni}\boldsymbol{\Theta}_{j1} + \boldsymbol{\Theta}_{1i}\boldsymbol{\Theta}_{jn}\boldsymbol{\Theta}_{n1} & \boldsymbol{\Theta}_{1m}\boldsymbol{\Theta}_{ni}\boldsymbol{\Theta}_{j2} + \boldsymbol{\Theta}_{1i}\boldsymbol{\Theta}_{jn}\boldsymbol{\Theta}_{n2} & \cdots & \boldsymbol{\Theta}_{1m}\boldsymbol{\Theta}_{ni}\boldsymbol{\Theta}_{jS} + \boldsymbol{\Theta}_{1i}\boldsymbol{\Theta}_{jn}\boldsymbol{\Theta}_{nS} \\ \boldsymbol{\Theta}_{2m}\boldsymbol{\Theta}_{ni}\boldsymbol{\Theta}_{j1} + \boldsymbol{\Theta}_{2i}\boldsymbol{\Theta}_{jn}\boldsymbol{\Theta}_{n1} & \boldsymbol{\Theta}_{2m}\boldsymbol{\Theta}_{ni}\boldsymbol{\Theta}_{j2} + \boldsymbol{\Theta}_{2i}\boldsymbol{\Theta}_{jn}\boldsymbol{\Theta}_{n2} & & \vdots \\ \vdots & & \ddots & \\ \boldsymbol{\Theta}_{Sm}\boldsymbol{\Theta}_{ni}\boldsymbol{\Theta}_{j1} + \boldsymbol{\Theta}_{Si}\boldsymbol{\Theta}_{jn}\boldsymbol{\Theta}_{n1} & \cdots & & \boldsymbol{\Theta}_{Sm}\boldsymbol{\Theta}_{ni}\boldsymbol{\Theta}_{jS} + \boldsymbol{\Theta}_{Si}\boldsymbol{\Theta}_{jn}\boldsymbol{\Theta}_{nS} \end{bmatrix} \tag{47}$$

the elements of the Hessian of the left eigenvector matrix with respect to the right eigenvector are constructed using the following definition for $i, j, m, n, r, c = 1,2, \ldots, S$:

$$\frac{\partial^2\boldsymbol{\Theta}_{rc}}{\partial\boldsymbol{\Gamma}_{ij}\partial\boldsymbol{\Gamma}_{mn}} \equiv \frac{\partial}{\partial\boldsymbol{\Gamma}_{mn}}\left(\frac{\partial\boldsymbol{\Theta}_{rc}}{\partial\boldsymbol{\Gamma}_{ij}}\right) = \frac{\partial}{\partial\boldsymbol{\Gamma}_{mn}}(-\boldsymbol{\Theta}_{ri}\boldsymbol{\Theta}_{jc}) = \boldsymbol{\Theta}_{rm}\boldsymbol{\Theta}_{ni}\boldsymbol{\Theta}_{jc} + \boldsymbol{\Theta}_{ri}\boldsymbol{\Theta}_{jm}\boldsymbol{\Theta}_{nc} \tag{48}$$

Next, in application of the chain rule to $\partial^2 G/(\partial\boldsymbol{\Gamma}_{ij}\partial\boldsymbol{\Gamma}_{mn})$, the terms are vectorized, so the result is a scalar and Eq. (45), from the first term of Eq. (44), becomes

$$\frac{\partial}{\partial\boldsymbol{\Gamma}_{mn}}\left(\frac{\partial G}{\partial\mathbf{A}}\right)\frac{\partial\mathbf{A}}{\partial\boldsymbol{\Gamma}_{ij}} = \text{vec}\left[-\mathbf{Q}^{-1}\left(\boldsymbol{\delta}_{mn}^{S\times S}\boldsymbol{\Lambda}\boldsymbol{\Theta} + \boldsymbol{\Gamma}\boldsymbol{\Lambda}\frac{\partial\boldsymbol{\Theta}}{\partial\boldsymbol{\Gamma}_{mn}}\right)\sum_{k=2}^{K}E[\mathbf{x}_{k-1}\mathbf{x}_{k-1}^T]\right] \cdot \text{vec}\left(\boldsymbol{\delta}_{ij}^{S\times S}\boldsymbol{\Lambda}\boldsymbol{\Theta} + \boldsymbol{\Gamma}\boldsymbol{\Lambda}\frac{\partial\boldsymbol{\Theta}}{\partial\boldsymbol{\Gamma}_{ij}}\right) \tag{49}$$

Similarly, Eq. (46), from the second term of Eq. (44), becomes

$$\frac{\partial G}{\partial\mathbf{A}}\frac{\partial}{\partial\boldsymbol{\Gamma}_{mn}}\left(\frac{\partial\mathbf{A}}{\partial\boldsymbol{\Gamma}_{ij}}\right) = \text{vec}\left(\mathbf{Q}^{-1}\sum_{k=2}^{K}E[\mathbf{x}_k\mathbf{x}_{k-1}^T] - \mathbf{Q}^{-1}\mathbf{A}\sum_{k=2}^{K}E[\mathbf{x}_{k-1}\mathbf{x}_{k-1}^T]\right) \cdot \text{vec}\left(\boldsymbol{\delta}_{ij}^{S\times S}\boldsymbol{\Lambda}\frac{\partial\boldsymbol{\Theta}}{\partial\boldsymbol{\Gamma}_{mn}} + \boldsymbol{\delta}_{mn}^{S\times S}\boldsymbol{\Lambda}\frac{\partial\boldsymbol{\Theta}}{\partial\boldsymbol{\Gamma}_{ij}} + \boldsymbol{\Gamma}\boldsymbol{\Lambda}\frac{\partial^2\boldsymbol{\Theta}}{\partial\boldsymbol{\Gamma}_{ij}\partial\boldsymbol{\Gamma}_{mn}}\right) \tag{50}$$

Finally, through combination of Eqs. (44), (49), and (50), Eq. (51) is determined:

$$\frac{\partial^2 G}{\partial\boldsymbol{\Gamma}_{ij}\partial\boldsymbol{\Gamma}_{mn}} = \text{vec}\left[-\mathbf{Q}^{-1}\left(\boldsymbol{\delta}_{mn}^{S\times S}\boldsymbol{\Lambda}\boldsymbol{\Theta} + \boldsymbol{\Gamma}\boldsymbol{\Lambda}\frac{\partial\boldsymbol{\Theta}}{\partial\boldsymbol{\Gamma}_{mn}}\right)\sum_{k=2}^{K}E[\mathbf{x}_{k-1}\mathbf{x}_{k-1}^T]\right] \cdot \text{vec}\left(\boldsymbol{\delta}_{ij}^{S\times S}\boldsymbol{\Lambda}\boldsymbol{\Theta} + \boldsymbol{\Gamma}\boldsymbol{\Lambda}\frac{\partial\boldsymbol{\Theta}}{\partial\boldsymbol{\Gamma}_{ij}}\right) + \text{vec}\left(\mathbf{Q}^{-1}\sum_{k=2}^{K}E[\mathbf{x}_k\mathbf{x}_{k-1}^T] - \mathbf{Q}^{-1}\mathbf{A}\sum_{k=2}^{K}E[\mathbf{x}_{k-1}\mathbf{x}_{k-1}^T]\right) \cdot \text{vec}\left(\boldsymbol{\delta}_{ij}^{S\times S}\boldsymbol{\Lambda}\frac{\partial\boldsymbol{\Theta}}{\partial\boldsymbol{\Gamma}_{mn}} + \boldsymbol{\delta}_{mn}^{S\times S}\boldsymbol{\Lambda}\frac{\partial\boldsymbol{\Theta}}{\partial\boldsymbol{\Gamma}_{ij}} + \boldsymbol{\Gamma}\boldsymbol{\Lambda}\frac{\partial^2\boldsymbol{\Theta}}{\partial\boldsymbol{\Gamma}_{ij}\partial\boldsymbol{\Gamma}_{mn}}\right) \tag{51}$$

Covariance Matrices

This section relates the information matrices derived in the previous sections to the corresponding asymptotic covariance. From ML estimation theory, the asymptotic parameter covariance matrix is the inverse of the parameter information matrix (King 1989):

$$\text{cov}(\psi) \equiv [I(\psi)]^{-1} \equiv \{-E[H(\psi)]\}^{-1} = \left\{-E\left[\frac{\partial^2 G}{\partial\psi^2}\right]\right\}^{-1} \tag{52}$$

Furthermore, the asymptotic parameter variances and standard errors can be extracted from the diagonals of the covariance matrix:

$$\text{var}(\psi) \equiv \text{diag}[\text{cov}(\psi)] \equiv \text{diag}\{[I(\psi)]^{-1}\} \tag{53}$$

$$\sigma(\psi) \equiv \sqrt{\text{diag}[\text{cov}(\psi)]} \equiv \sqrt{\text{diag}\{[I(\psi)]^{-1}\}} \tag{54}$$

In most practical purposes, ML estimates will differ from the true model parameters. Thus, it is important to make the distinction between true asymptotic parameter covariance, which would result from a very large number of samples, and the estimated asymptotic parameter covariance for finite sample sizes. The computations are identical in form to the equations in this paper; true model parameters are substituted with MLE, namely $\psi = \hat{\psi}$, for each parameter.

Parameter Variance Results

The asymptotic parameter covariance Eqs. (55)–(58) directly follow from Eq. (52) when implementing the Hessian expressions from Eqs. (14), (21), (33), and (51):

$$\text{cov}(\mathbf{C}) \equiv \left\{-E\left[\frac{\partial^2 G}{\partial \mathbf{C}\partial \mathbf{C}^T}\right]\right\}^{-1} \tag{55}$$

$$\text{cov}(\mathbf{A}) \equiv \left\{-E\left[\frac{\partial^2 G}{\partial \mathbf{A}\partial \mathbf{A}^T}\right]\right\}^{-1} \tag{56}$$

$$\text{cov}(\mathbf{\Lambda}) \equiv \left\{-E\left[\frac{\partial^2 G}{\partial \mathbf{\Lambda}\partial \mathbf{\Lambda}^T}\right]\right\}^{-1} \tag{57}$$

$$\text{cov}(\mathbf{\Gamma}) \equiv \left\{-E\left[\frac{\partial^2 G}{\partial \mathbf{\Gamma}\partial \mathbf{\Gamma}^T}\right]\right\}^{-1} \tag{58}$$

The mode shapes are computed using the observation matrix and right state eigenvector matrix. Similarly, these estimated parameters and their respective variances [Eqs. (55) and (58)] are used to compute the mode shape variance. Because the mode shape matrix is defined as the product of two random variables $\mathbf{\Phi} = \mathbf{C\Gamma}$, an entry of the mode shape matrix is $\mathbf{\Phi}_{ij} = \mathbf{C}_i\mathbf{\Gamma}_j$, computed using the ith row of the observation matrix and the jth column of the right state eigenvector matrix. The variance of the mode shapes is constructed element-wise:

$$\text{var}(\mathbf{\Phi}_{ij}) \equiv \text{var}(\mathbf{C}_i)\text{var}(\mathbf{\Gamma}_j) + \text{var}(\mathbf{C}_i)(\mathbf{\Gamma}_j^2) + \mathbf{C}_i^2\text{var}(\mathbf{\Gamma}_j) \tag{59}$$

$$\text{var}(\mathbf{\Phi}) \equiv \begin{bmatrix} \text{var}(\mathbf{\Phi}_{11}) & \text{var}(\mathbf{\Phi}_{12}) & \cdots & \text{var}(\mathbf{\Phi}_{1S}) \\ \text{var}(\mathbf{\Phi}_{21}) & \text{var}(\mathbf{\Phi}_{22}) & & \vdots \\ \vdots & & \ddots & \\ \text{var}(\mathbf{\Phi}_{O1}) & \cdots & & \text{var}(\mathbf{\Phi}_{OS}) \end{bmatrix} \tag{60}$$

The frequency and damping ratio variances are not as simple because they comprise nonlinear functions:

$$f_n = \frac{2\pi}{\Delta t}\ln(\lambda_d) \tag{61}$$

$$\zeta_n = -\cos\{\measuredangle[\ln(\lambda_d)]\} \tag{62}$$

Because an analytical form of these variances is desirable, the authors initially implemented a first-order Taylor series expansion to approximate the moments of the modal properties. However, this approximation did not prove to be consistent with Monte Carlo simulations; this is a result of the highly nonlinear behavior of the logarithm function in the bounded domain of the eigenvalues [the norm of each eigenvalue is less than 1 (Juang and Phan 2001)]. Despite this challenge, the variance of the frequencies and damping ratios can be accurately computed after constructing their 95% confidence intervals through a perturbation technique detailed in the following section.

Asymptotic Distribution and Interval Estimation

Depending on the asymptotic distribution of the parameter, the computed standard errors are used to construct an estimation interval. Under the previously defined regularity conditions of the ML point estimate, say a scalar estimator $\hat{\psi}$, the distribution of the MLE is Gaussian:

$$\hat{\psi} - \psi \sim N\{0, \quad [I(\psi)]^{-1}\} \tag{63}$$

$$\hat{\psi} - \psi \sim N[0, \text{var}(\psi)] \tag{64}$$

The distribution of the MLE is also asymptotically centered on the true parameter because the MLE bias diminishes to zero at rate $O(K^{-1})$ (Quenouille 1956):

$$\hat{\psi} \sim N[\psi, \text{var}(\psi)] \tag{65}$$

The application of Eq. (63) to the elements of the observation and state matrices is a logical extension of ML theory. It is clear that all six model parameters are governed by this theory, but does this apply to the eigenvalues and eigenvectors of the state matrix? This paper demonstrates that the eigenvalues are also MLEs and are therefore asymptotically normal (see Appendix I for details). At this point, the eigenvectors were assumed to also be asymptotically normal and this assumption was verified through Monte Carlo simulations. Finally, with $\sigma(\psi)$, the $1-\alpha$ confidence interval is constructed as

$$\hat{\psi} \pm z_{1-\alpha/2}\sigma(\psi) \tag{66}$$

The general expression Eq. (66) is extended to the multivariate parameter case using the vec operator, the parameter covariance matrices of the previous section, and the definition of standard deviation [Eq. (54)]:

$$\text{vec}(\hat{\mathbf{\Lambda}}) \pm z_{1-\alpha/2}\text{vec}[\sigma(\hat{\mathbf{\Lambda}})] \tag{67}$$

$$\text{vec}(\hat{\mathbf{\Phi}}) \pm z_{1-\alpha/2}\text{vec}[\sigma(\hat{\mathbf{\Phi}})] \tag{68}$$

Confidence intervals for each frequency and damping ratio can be accurately computed through perturbation of the eigenvalue from the MLE as shown in Eqs. (69) and (70); the following intervals have been verified to be within 1% of Monte Carlo simulations:

$$[\hat{f}_{LL}\hat{f}_{UL}] = \frac{2\pi}{\Delta t}\ln[\lambda_d \pm z_{1-\alpha/2}\sigma(\lambda_d)] \tag{69}$$

$$[\hat{\zeta}_{LL}\hat{\zeta}_{UL}] = -\cos(\measuredangle\{\ln[\lambda_d \pm z_{1-\alpha/2}\sigma(\lambda_d)]\}) \tag{70}$$

If variances are desired for frequencies or damping, the limits of the interval can be rearranged to solve for the standard errors directly:

$$f_n \pm z_{1-\alpha/2}\sigma(f_n) = [\hat{f}_{LL}\hat{f}_{UL}] \tag{71}$$

$$\zeta_n \pm z_{1-\alpha/2}\sigma(\zeta_n) = [\hat{\zeta}_{LL}\hat{\zeta}_{UL}] \tag{72}$$

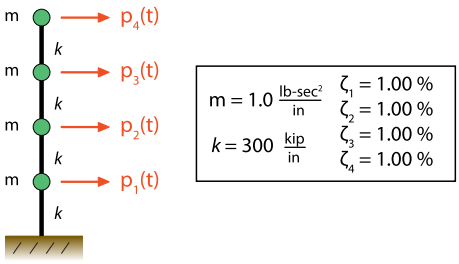

Fig. 1. Four-story shear structure, structural properties, damping, and loading

Table 1. True Frequencies, Gaussian ML Results, Bootstrap Statistics, and Relative Deviation

True value (Hz)	Gaussian ML: $\hat{f}_n, \sigma(\hat{f}_n)$ Estimate and 95% confidence interval	Gaussian ML: Standard deviation	Bootstrap stats: $f_n^*, \sigma(f_n^*)$ Mean and 95% confidence interval	Bootstrap stats: Standard deviation	Relative deviation $\sigma(f_n^*)/\sigma(\hat{f}_n)$
$f_1 = 0.9574$	0.9594 ± 0.0021	0.0011	0.9592 ± 0.0055	0.0028	2.6
$f_2 = 2.7566$	2.7566 ± 0.0136	0.0069	2.7568 ± 0.0103	0.0052	0.8
$f_3 = 4.2234$	4.2239 ± 0.0087	0.0045	4.2240 ± 0.0133	0.0068	1.5
$f_4 = 5.1808$	5.1778 ± 0.0031	0.0016	5.1783 ± 0.0135	0.0069	4.4

Application

This section implements tools developed throughout the first three sections in a practical system identification application. The identification of a four degree-of-freedom (DOF) shear structure, documented in Fig. 1, is considered for illustration. Closed-form Gaussian ML (GML) parametric variances and confidence bounds are compared with those approximated through nonparametric Monte Carlo bootstrapping.

Monte Carlo Bootstrap

Bootstrap methods have become a popular strategy for estimating the bias and variance of desired statistics, in this case model parameters, due to their broad applicability and ease of implementation. Efron (1979) introduced the bootstrap as a generalized version of the Quenouille-Tukey jackknife (Quenouille 1956), of which an informative review is available from Miller (1974). Bootstrapping methods are an effective alternative to closed-form variance estimation and have vast applications in time series problems (Efron and Tibshirani 1986; Masset 2008; Pakzad et al. 2009).

Furthermore, bootstrap results contain desirable asymptotic properties, e.g., when the number of Monte Carlo simulations and the number of samples become very large, the bootstrap estimator is equivalent to the ML estimator (Bickel and Freedman 1981). In ML estimation, bootstrapping is especially favorable for models involving complex likelihood functions, for which closed-form derivatives are arduous or impossible. Bickel and Freedman (1981) provided additional properties of the bootstrap estimator.

The application in this section follows the approach for bootstrapping the Gaussian ML estimator for the state-space model that is presented by Stoffer and Wall (1991). A bootstrapped dataset is generated by resampling the original data with replacement and corresponding perturbations in the model parameters are recorded. The procedure is repeated a large number of times, computing equivalently many bootstrap data realizations from which the true distribution of the MLE is estimated by the distribution of the bootstrapped model parameters.

Table 2. Closed-Form Gaussian ML and Bootstrap Asymptotic 95% Frequency Confidence Intervals

Mode	Gaussian ML $\hat{f}_{LL}$ (Hz)	Gaussian ML $\hat{f}_{UL}$ (Hz)	Gaussian ML $\frac{(\hat{f}_{UL}-\hat{f}_{LL})}{f_n}$ (%)	Bootstrap stats f^*_{LL} (Hz)	Bootstrap stats f^*_{UL} (Hz)	Bootstrap stats $\frac{(f^*_{UL}-f^*_{LL})}{f_n}$ (%)
1	0.9573	0.9615	0.44	0.9538	0.9647	1.1
2	2.7431	2.7702	0.98	2.7466	2.7670	0.74
3	4.2151	4.2326	0.42	4.2106	4.2373	0.63
4	5.1747	5.1809	0.12	5.1648	5.1918	0.52

ML Identification of a Four-Degree-of-Freedom Shear Structure

The four DOF shear structure shown in Fig. 1 is identified using the output-only STRIDE algorithm. The shear structure has four natural frequencies under 6 Hz and 1% damping ratios for all modes. A successful identification of this structure includes accurate frequency, damping, and mode shape estimates. The goal of this application is to compute closed-form sensitivity metrics of ML estimates and compare to bootstrap estimates. The structural response of the shear structure was simulated through the theoretical state-space model for a structural system (Juang and Phan 2001). More specifically, the model observations were noisy story accelerations, driven by an independent and identically distributed random input, i.e., $\boldsymbol{\eta}_k \sim N(\mathbf{0}, \mathbf{Q})$ and $\boldsymbol{\upsilon}_k \sim N(\mathbf{0}, \mathbf{R})$, with $\mathbf{Q} \propto I^{8\times 8}$ and $\mathbf{R} \propto I^{4\times 4}$.

STRIDE (Matarazzo and Pakzad 2015a) is an iterative output-only method for modal identification which embeds the EM algorithm (Rubin et al. 1977). STRIDE determines ML estimates of a model through maximizing the conditional expectation of the likelihood function. The algorithm begins with initial parameter estimates for model parameters that are updated at each iteration in a manner that guarantees an increase in the conditional likelihood function (Rubin et al. 1977; Wu 1983). A slope threshold θ is utilized to determine when the conditional expectation of the likelihood function has practically attained its maximum value. For a

Table 3. True Damping, Gaussian ML Results, Bootstrap Statistics, and Relative Deviation

True value (%)	Gaussian ML: $\hat{\zeta}_n, \sigma(\hat{\zeta}_n)$ Estimate and 95% confidence interval	Gaussian ML: Standard deviation	Bootstrap statistics: $\zeta_n^*, \sigma(\zeta_n^*)$ Mean and 95% confidence interval	Bootstrap statistics: Standard deviation	Relative deviation $\sigma(\zeta_n^*)/\sigma(\hat{\zeta}_n)$
$\zeta_1 = 1.0000$	0.6958 ± 0.3949	0.2015	0.8067 ± 0.6099	0.3112	1.5
$\zeta_2 = 1.0000$	0.9256 ± 0.0604	0.0308	0.9673 ± 0.3732	0.1904	6.2
$\zeta_3 = 1.0000$	1.1154 ± 0.1586	0.0809	1.1389 ± 0.3277	0.1672	2.1
$\zeta_4 = 1.0000$	0.9675 ± 0.1337	0.0682	1.9834 ± 0.2804	0.1431	2.1

given data set and model order, smaller slope thresholds are conducive to higher likelihood values at convergence and yield smaller parameter variances because the corresponding critical point is technically a superior MLE.

Theoretically, the absolute maximum in the likelihood function ensures an unbiased, consistent, and efficient estimate (Gupta and Mehra 1974), while the EM algorithm may not necessarily converge to this point. The conditional likelihood in EM converges monotonically, from below, to some maximum value of the sequence, so if this likelihood has several maximums and stationary points, the choice of starting point determines the type of point to which the EM sequence will converge. Thus, for a successful EM implementation, it is important to choose a good starting value as incorporated in the development of STRIDE (Matarazzo and Pakzad 2015a).

The observed data consisted of 2,000 samples at 12 Hz. STRIDE was implemented at minimum model order two ($p = 2$) with default slope threshold $\theta = 5 \times 10^{-4}$ to obtain the MLE of the superparameter. Various equations presented in this paper were computed at these MLEs to determine information and covariance metrics. The results that incorporate the MLE with these equations will be henceforth denoted as GML.

The bootstrap consisted of 1,000 Monte Carlo simulations of the procedure presented in Stoffer and Wall (1991). In this approach, the standardized innovation sequence is sampled K times without replacement to produce bootstrapped observations. With these observations and the superparameter fixed at the MLE from STRIDE, the state-space model is complete. Finally, the likelihood is constructed and a bootstrapped superparameter can be computed for each Monte Carlo run; these bootstrapped superparameters

Table 4. Closed-Form Gaussian ML and Bootstrap Asymptotic 95% Damping Confidence Intervals

	Gaussian ML			Bootstrap stats		
Mode	ζ_{LL} (%)	ζ_{UL} (%)	$\frac{(\zeta_{UL}-\zeta_{LL})}{\zeta_n}$ (%)	$\hat{\zeta}_{LL}$	$\hat{\zeta}_{UL}$	$\frac{(\hat{\zeta}_{UL}-\hat{\zeta}_{LL})}{\zeta_n}$ (%)
1	0.3009	1.0895	79	0.1968	1.4166	122
2	0.8652	0.9891	12	0.5940	1.3405	75
3	0.9570	1.2740	32	0.8112	1.4666	66
4	0.8343	1.1012	27	0.7030	1.2638	56

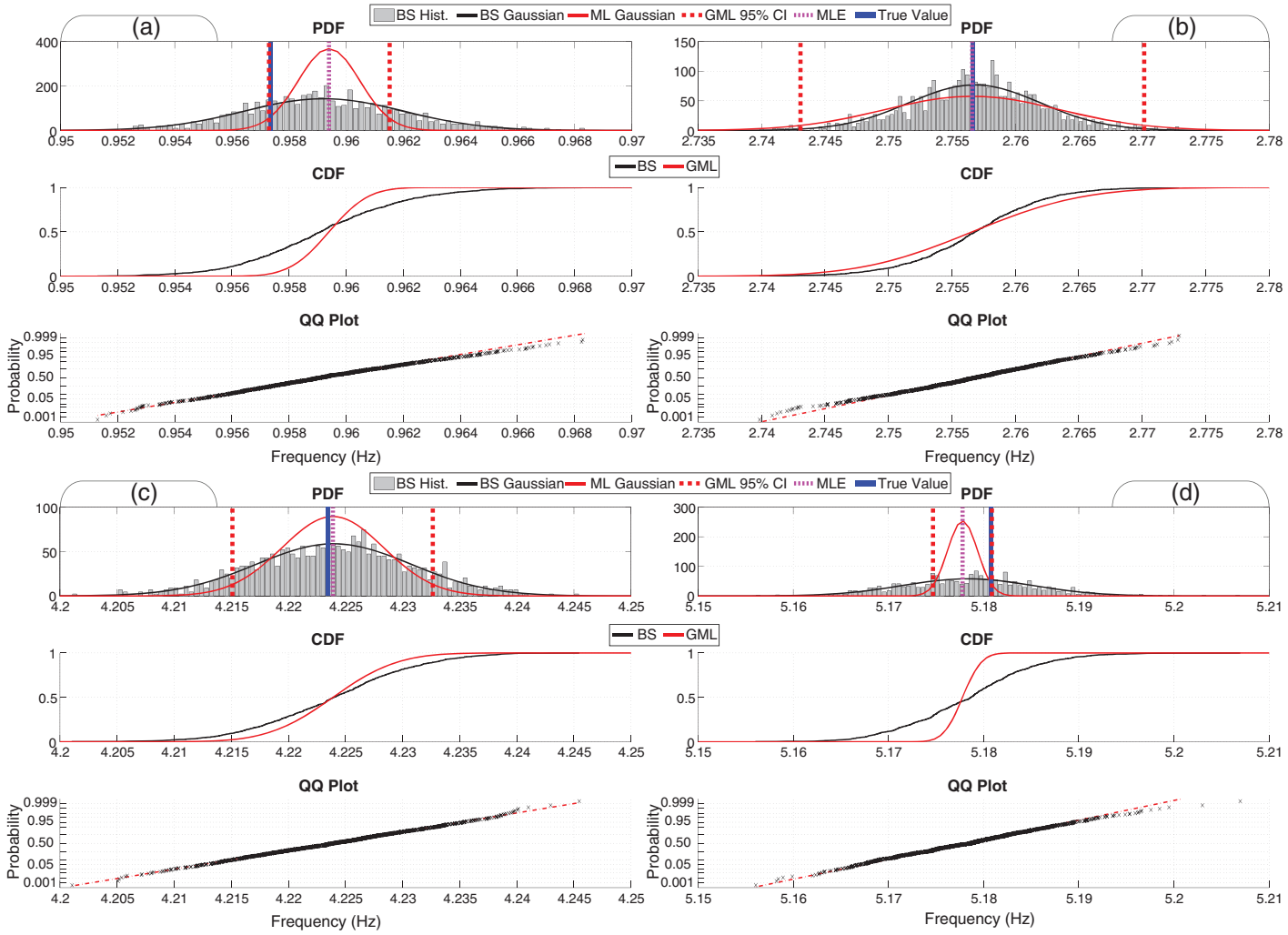

Fig. 2. Precision of (a) first frequency estimate; (b) second frequency estimate; (c) third frequency estimate; (d) fourth frequency estimate: bootstrap histogram, bootstrap Gaussian fit probability density function (PDF), Gaussian ML PDF, Gaussian 95% confidence intervals, and true values; bootstrap cumulative distribution function (CDF) and Gaussian ML CDF; quantile-quantile (QQ) plot for bootstrap

ASCE-ASME J. Risk Uncertainty Eng. Syst., Part A: Civ. Eng.

represent a sample of the true parameter population in ML estimation. It is expected that the variances and confidence intervals obtained through closed-form formulas will be smaller than the bootstrap results.

Discussion

Tables 1 and 2 provide 95% confidence intervals and compare natural frequency results. True values, MLEs, GML standard errors, bootstrap means, and bootstrap standard errors are presented. First, all four true frequencies are enclosed within both GML and bootstrap confidence bounds. Second, in most cases, GML standard errors are considerably smaller than bootstrap standard errors; this result is most evident from the last column of Table 1. As a result, GML confidence intervals are tighter than bootstrap confidence intervals. The second frequency is the exception, where the GML variance is slightly larger than that of the bootstrap.

Tables 3 and 4 are analogous to Tables 1 and 2 for damping ratios. In general, the GML damping ratios are two orders of magnitude less precise than GML frequency estimates. Nevertheless, true damping values are included within the GML (the second mode is on the boundary) and bootstrap confidence bounds. However, bootstrap standard errors are overall noticeably larger than those from GML. For example, consider the second damping ratio in row two of Table 3, in which the GML has nearly six times the precision of the bootstrap. In short, the formulas in this paper are superior to bootstrapping for estimating MLE damping precision.

Figs. 2(a–d) provide detailed comparisons between GML and bootstrap results for each frequency estimate. The top panel of each figure superimposes the asymptotic GML probability density function (PDF) over bootstrap histograms and a bootstrap Gaussian fit. GML confidence bounds and the true value are also indicated in this subplot. In the four top panels, the GML PDFs generally have lighter tails than the bootstrap PDF. The middle panel of these figures shows the GML cumulative distribution function (CDF) and bootstrap CDF. GML and booststrap medians are nearly coincident for all frequencies. The bottom panel displays a QQ plot for the bootstrapped frequency estimates and demonstrates that the normality assumption is valid. Figs. 3(a–d) show QQ plots for the bootstrapped damping estimates. The drifting tails in these plots indicate that the normality assumption is not valid for damping, at least for the sample size considered.

Fig. 4 shows the true mode shapes for the shear structure with superimposed GML and bootstrap estimates and confidence bounds. After identification, the mode shapes were scaled so that the absolute value of the maximum ordinate was equal to 1, purely for convenient presentation. MLE mode shapes were identical to the true mode shapes. Bootstrap mode shape estimates were biased and had larger confidence intervals, and therefore less precision than MLE. To reiterate, the equations in this paper are applicable to complex mode shapes. Although the mode shape covariance matrix is complex, in this application, the computed complex values were practically zero.

Overall, the GML results are consistent with ML theory, with confidence bounds that enclosed the true values and lower parameter variances than those estimated by bootstrapping. As the sample size becomes very large, both GML and bootstrap variance estimates approach the Cramér-Rao lower bound; for finite samples, the tools presented in this paper provide valid estimates of precision.

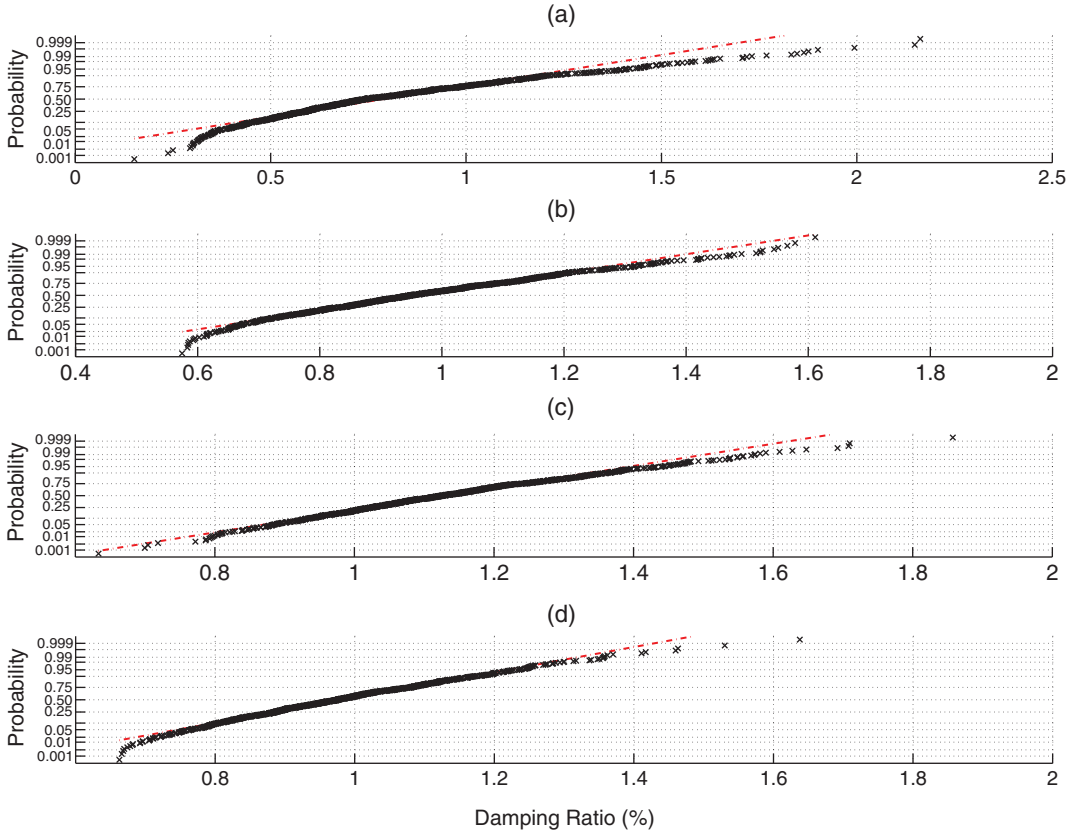

Fig. 3. QQ plots for bootstrap results for (a) first mode damping; (b) second mode damping; (c) third mode damping; (d) fourth mode damping

ASCE-ASME J. Risk Uncertainty Eng. Syst., Part A: Civ. Eng.

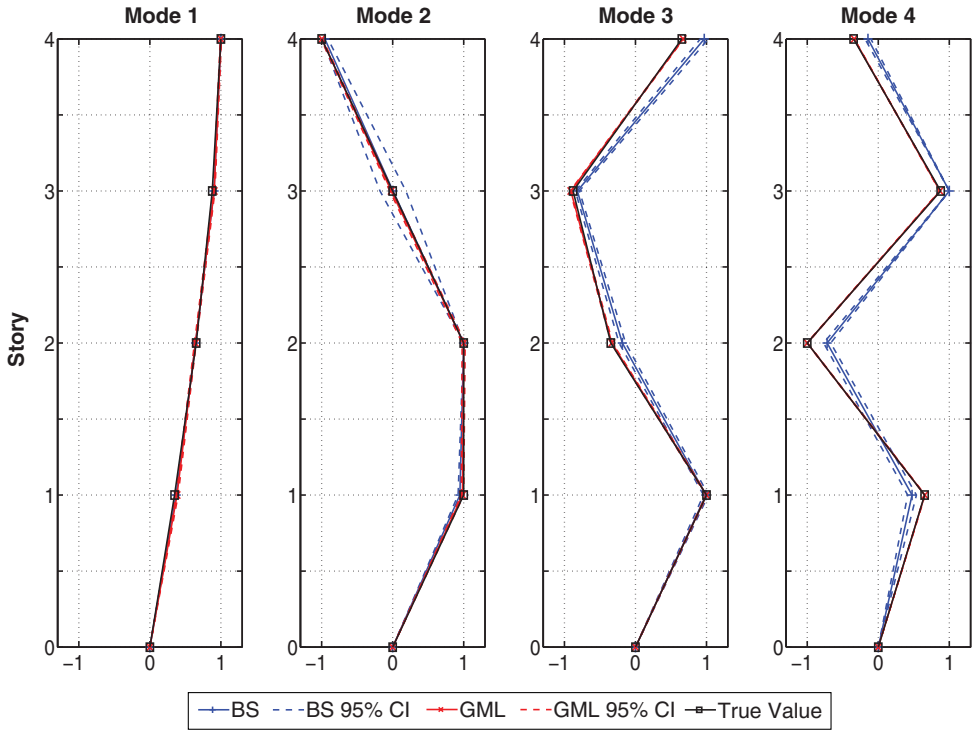

Fig. 4. Bootstrap mode shapes with 95% confidence intervals, Gaussian MLE with 95% confidence intervals, and true values

Conclusion

This paper presented a set of sensitivity metrics to be used along likelihood-based modal identification methods. The closed-form partial derivatives of the likelihood function are directly related to observed information and variance expressions for discrete-time stochastic state-space model parameters. Derivatives were provided for the observation matrix and the state matrix as well as eigenvalues and eigenvectors of the state matrix. Standard error formulas and confidence intervals were constructed for natural frequencies, damping ratios, and mode shapes by implementing asymptotic characteristics of ML estimators. While the equations supplement the STRIDE modal identification algorithm, they are applicable to modal identification techniques formulated within the state-space model.

An application to structural modal identification compared closed-form asymptotic parameter metrics to Monte Carlo bootstrap estimates. For frequency estimates, Gaussian ML PDFs showed lighter tails than bootstrap histograms and Gaussian fits indicating a lower estimation uncertainty. Closed-form 95% confidence bounds for the frequency and damping ratio included modal properties. The normality assumption for asymptotic frequency distribution was validated; however, this assumption did not hold for damping estimates. It was demonstrated that damping ratio standard errors were better represented by the MLE formulas than by bootstrapping. MLE mode shapes were coincident with the true values while bootstrap means were biased.

Appendix I contains a brief proof that eigenvalues for an MLE state matrix are also MLEs. This follows directly from the state matrix M-step of the EM algorithm because it simultaneously equates the first derivative of the likelihood function with respect to an eigenvalue to zero. This paper applied ML theory to establish a better understanding of how modal parameters estimated through ML may differ from their true values. The precision of MLE was demonstrated to be higher than bootstrapping while still enclosing the true values. In short, the asymptotic advantages of MLEs are distinct and valuable, further supporting ML methods for structural modal identification.

Appendix I. Note on the M-Step Update Formula for EM in the State-Space Model

Recall the first derivative of the conditional likelihood with respect to an eigenvalue of the state matrix:

$$\frac{\partial G}{\partial \lambda_d} = \text{vec}\left(\mathbf{Q}^{-1}\sum_{k=2}^{K} E[\mathbf{x}_k \mathbf{x}_{k-1}^T] - \mathbf{Q}^{-1}\mathbf{A}\sum_{k=2}^{K} E[\mathbf{x}_{k-1}\mathbf{x}_{k-1}^T]\right) \cdot \text{vec}(\boldsymbol{\Gamma}\boldsymbol{\delta}_d^{S\times S}\boldsymbol{\Theta}) \tag{73}$$

In the EM algorithm, M-step parameters are chosen to optimize the likelihood function by equating the first partial derivative to zero:

$$\frac{\partial G}{\partial \psi} = 0 \tag{74}$$

The M-step update formula for the state matrix satisfies Eq. (75) (Matarazzo and Pakzad 2015a):

$$\frac{\partial G}{\partial \mathbf{A}} = \mathbf{Q}^{-1}\sum_{k=2}^{K} E[\mathbf{x}_k \mathbf{x}_{k-1}^T] - \mathbf{Q}^{-1}\mathbf{A}\sum_{k=2}^{K} E[\mathbf{x}_{k-1}\mathbf{x}_{k-1}^T] = 0 \tag{75}$$

In consideration of the eigendecomposition of a matrix to be solely a function of the matrix, the state matrix eigenvalues are purely a function of the state matrix, i.e., $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}(\mathbf{A})$, $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}(\mathbf{A})$, and $\boldsymbol{\Theta} = \boldsymbol{\Theta}(\mathbf{A})$. Eq. (75) also satisfies $\partial G/\partial \lambda_d = 0$. Through maximization of the likelihood in terms of the state matrix, the likelihood is also optimized with respect to the state matrix eigenvalues. In other words, the eigenvalues of an ML state matrix estimate are also MLE, and thus these eigenvalues share the asymptotic advantages of ML estimators.

Appendix II. State-Space Model Parameters, Features, and Other Terms

Matrix	Size	Description	Element
O	Scalar	Observation size	—
S	Scalar	State-space size	—
G	Scalar	Conditional expectation of log-likelihood function	—
$\mathbf{C}$	$O \times S$	Observation matrix	$\mathbf{C}_{ij}$
$\mathbf{A}$	$S \times S$	State matrix	$\mathbf{A}_{ij}$
$\boldsymbol{\Gamma}$	$S \times S$	Eigenvector matrix of state matrix	$\boldsymbol{\Gamma}_{ij}$
$\boldsymbol{\Theta}$	$S \times S$	Inverse eigenvector matrix (left eigenvectors)	$\boldsymbol{\Theta}_{ij}$
$\boldsymbol{\Lambda}$	$S \times S$	Diagonal eigenvalue matrix; lowercase for diagonal element	λ_d
$\mathbf{Q}$	$S \times S$	Structural loading covariance	$\mathbf{Q}_{ij}$
$\sum_{k=1}^{K} E[\mathbf{x}_k \mathbf{x}_{k-1}^T]$	$S \times S$	Sum of mean square statistics for states at times k and $k-1$	$[\sum_{k=1}^{K} E[\mathbf{x}_k \mathbf{x}_{k-1}^T]]_{ij}$
$\sum_{k=1}^{K} E[\mathbf{x}_{k-1} \mathbf{x}_{k-1}^T]$	$S \times S$	Sum of mean square statistics for states at times $k-1$ and $k-1$	$[\sum_{k=1}^{K} E[\mathbf{x}_{k-1} \mathbf{x}_{k-1}^T]]_{ij}$
$\boldsymbol{\delta}_{mn}^{R \times C}$	$R \times C$	Single-entry matrix: zeros everywhere, except entry mn has unity	$[\boldsymbol{\delta}_{mn}^{R \times C}]_{ij}$

Appendix III. Index, Subscript, and Dimension Descriptions

Index	Description
i	Row subscript
j	Column subscript
m	Alternate row subscript
n	Alternate column subscript
L	Linear index subscript
P	Alternate linear index subscript
d	Diagonal subscript
h	Alternate diagonal subscript
r	Second alternate row subscript
c	Second alternate column subscript
R	Alternate row size
C	Alternate column size
$[\,]$	Matrix
$[\,]_{ij}$	Alternate notation for matrix element at row i, column j
k	Time-step subscript
K	Total number of time samples

Appendix IV. First Partial Derivative Matrices, Vectors, and Elements

Matrix	Size	Description	Element
$\frac{\partial G}{\partial \mathbf{C}}$	$O \times S$	First partial (score) of G with regard to $\mathbf{C}$	$\frac{\partial G}{\partial \mathbf{C}_{ij}}$
$\frac{\partial G}{\partial \mathbf{A}}$	$S \times S$	First partial (score) of G with regard to $\mathbf{A}$	$\frac{\partial G}{\partial \mathbf{A}_{ij}}$
$\mathrm{vec}\left(\frac{\partial G}{\partial \mathbf{A}}\right)$	$S^2 \times 1$	Vectorized $\frac{\partial G}{\partial \mathbf{A}}$	$\frac{\partial G}{\partial \mathbf{A}_L}$
$\frac{\partial G}{\partial \boldsymbol{\Gamma}}$	$S \times S$	First partial (score) of G with regard to $\boldsymbol{\Gamma}$	$\frac{\partial G}{\partial \boldsymbol{\Gamma}_{ij}}$
$\frac{\partial G}{\partial \boldsymbol{\Theta}}$	$S \times S$	First partial (score) of G with regard to $\boldsymbol{\Theta}$	$\frac{\partial G}{\partial \boldsymbol{\Theta}_{ij}}$
$\frac{\partial G}{\partial \boldsymbol{\Lambda}}$	$S \times S$	First partial (score) of G with regard to $\boldsymbol{\Lambda}$; diagonal matrix	$\frac{\partial G}{\partial \lambda_d}$
$\frac{\partial \mathbf{A}}{\partial \lambda_d}$	$S \times S$	First partial of $\mathbf{A}$ with respect to eigenvalue λ_d	$\frac{\partial \mathbf{A}_{ij}}{\partial \lambda_d}$
$\mathrm{vec}\left(\frac{\partial \mathbf{A}}{\partial \lambda_d}\right)$	$S^2 \times 1$	Vectorized $\frac{\partial \mathbf{A}_{ij}}{\partial \lambda_d}$	$\frac{\partial \mathbf{A}_L}{\partial \lambda_d}$
$\frac{\partial \mathbf{A}}{\partial \boldsymbol{\Gamma}_{mn}}$	$S \times S$	First partial of $\mathbf{A}$ with regard to eigenvector element $\boldsymbol{\Gamma}_{mn}$	$\frac{\partial \mathbf{A}_{ij}}{\partial \boldsymbol{\Gamma}_{mn}} = \left[\frac{\partial \mathbf{A}}{\partial \boldsymbol{\Gamma}_{mn}}\right]_{ij}$
$\mathrm{vec}\left(\frac{\partial \mathbf{A}}{\partial \boldsymbol{\Gamma}_P}\right)$	$S^2 \times 1$	Vectorized $\frac{\partial \mathbf{A}}{\partial \boldsymbol{\Gamma}_{mn}}$	$\frac{\partial \mathbf{A}_L}{\partial \boldsymbol{\Gamma}_P}$
$\frac{\partial \boldsymbol{\Gamma}}{\partial \boldsymbol{\Gamma}_{mn}}$	$S \times S$	First partial of right eigenvector matrix with regard to $\boldsymbol{\Gamma}_{mn}$	$\frac{\partial \boldsymbol{\Gamma}_{ij}}{\partial \boldsymbol{\Gamma}_{mn}} = \left[\frac{\partial \boldsymbol{\Gamma}}{\partial \boldsymbol{\Gamma}_{mn}}\right]_{ij}$
$\frac{\partial \boldsymbol{\Theta}}{\partial \boldsymbol{\Gamma}_{mn}}$	$S \times S$	First partial of left eigenvector matrix with regard to $\boldsymbol{\Gamma}_{mn}$	$\frac{\partial \boldsymbol{\Theta}_{ij}}{\partial \boldsymbol{\Gamma}_{mn}} = \left[\frac{\partial \boldsymbol{\Theta}}{\partial \boldsymbol{\Gamma}_{mn}}\right]_{ij}$

Appendix V. Second Partial Derivative Terms: Block Matrices, Vectors, and Elements

Matrix	Size	Description	Element
$\frac{\partial^2 G}{\partial \mathbf{C} \partial \mathbf{C}^T}$	$OS \times OS$	Second partial (Hessian) of G with regard to $\mathbf{C}$	$\left[\frac{\partial^2 G}{\partial \mathbf{C} \partial \mathbf{C}_{ij}}\right]$
$\mathrm{vec}\left(\frac{\partial^2 G}{\partial \mathbf{C} \partial \mathbf{C}_P}\right)$	$OS \times 1$	Vectorized $\frac{\partial^2 G}{\partial \mathbf{C} \partial \mathbf{C}_{ij}}$	$\frac{\partial^2 G}{\partial \mathbf{C}_L \partial \mathbf{C}_P}$
$\frac{\partial^2 G}{\partial \mathbf{A} \partial \mathbf{A}^T}$	$S^2 \times S^2$	Second partial (Hessian) of G with regard to $\mathbf{A}$	$\left[\frac{\partial^2 G}{\partial \mathbf{A} \partial \mathbf{A}_{ij}}\right]$
$\mathrm{vec}\left(\frac{\partial^2 G}{\partial \mathbf{A} \partial \mathbf{A}_P}\right)$	$S^2 \times 1$	Vectorized $\frac{\partial^2 G}{\partial \mathbf{A} \partial \mathbf{A}_{ij}}$	$\frac{\partial^2 G}{\partial \mathbf{A}_L \partial \mathbf{A}_P}$
$\frac{\partial^2 G}{\partial \boldsymbol{\Gamma} \partial \boldsymbol{\Gamma}_{mn}}$	$S \times S$	Second partial (Hessian) of G with regard to $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}_{mn}$	$\frac{\partial^2 G}{\partial \boldsymbol{\Gamma}_{ij} \partial \boldsymbol{\Gamma}_{mn}}$
$\mathrm{vec}\left(\frac{\partial^2 G}{\partial \boldsymbol{\Gamma} \partial \boldsymbol{\Gamma}_P}\right)$	$S^2 \times 1$	Vectorized $\frac{\partial^2 G}{\partial \boldsymbol{\Gamma} \partial \boldsymbol{\Gamma}_{mn}}$	$\frac{\partial^2 G}{\partial \boldsymbol{\Gamma}_L \partial \boldsymbol{\Gamma}_P}$
$\frac{\partial^2 G}{\partial \boldsymbol{\Gamma} \partial \boldsymbol{\Gamma}^T}$	$S^2 \times S^2$	Second partial (block Hessian) of G with regard to $\boldsymbol{\Gamma}$; columns are $\mathrm{vec}\left(\frac{\partial^2 G}{\partial \boldsymbol{\Gamma} \partial \boldsymbol{\Gamma}_P}\right)$	$\left[\frac{\partial^2 G}{\partial \boldsymbol{\Gamma} \partial \boldsymbol{\Gamma}_P}\right]$
$\frac{\partial^2 G}{\partial \boldsymbol{\Lambda} \partial \lambda_h}$	$S \times S$	Second partial (Hessian) of G with regard to $\boldsymbol{\Lambda}$ and λ_h; zeros for entries with $h \neq d$	$\frac{\partial^2 G}{\partial \lambda_d \partial \lambda_h}$
$\mathrm{vec}\left(\frac{\partial^2 G}{\partial \boldsymbol{\Lambda} \partial \lambda_h}\right)$	$S \times 1$	Vectorized diagonals of $\frac{\partial^2 G}{\partial \boldsymbol{\Lambda} \partial \lambda_h}$	$\frac{\partial^2 G}{\partial \lambda_d \partial \lambda_h}$
$\frac{\partial^2 G}{\partial \boldsymbol{\Lambda} \partial \boldsymbol{\Lambda}}$	$S \times S$	Second partial (block Hessian) of G with regard to the diagonals of $\boldsymbol{\Lambda}$; columns are $\mathrm{vec}\left(\frac{\partial^2 G}{\partial \boldsymbol{\Lambda} \partial \lambda_h}\right)$	$\left[\frac{\partial^2 G}{\partial \boldsymbol{\Lambda} \partial \lambda_h}\right]$

Note: Block matrix elements are encapsulated with brackets, e.g., [].

Acknowledgments

Research funding is partially provided by the National Science Foundation through Grant No. CMMI-1351537 by Hazard Mitigation and Structural Engineering program, and by a grant from the Commonwealth of Pennsylvania, Department of Community and Economic Development, through the Pennsylvania Infrastructure Technology Alliance (PITA).

References

Abdel-Ghaffar, A., and Scanlan, R. H. (1985). "Ambient vibration studies of Golden Gate Bridge: 1. Suspended structure." *J. Eng. Mech.*, 10.1061/(ASCE)0733-9399(1985)111:4(463), 463–482.

Andersen, P. (1997). "Identification of civil engineering structures using vector ARMA models." Ph.D. dissertation, Aalborg Univ., Aalborg, Denmark.

Au, S.-K. (2011). "Fast Bayesian FFT method for ambient modal identification with separated modes." *J. Eng. Mech.*, 10.1061/(ASCE)EM.1943-7889.0000213, 214–226.

Au, S.-K. (2012). "Connecting Bayesian and frequentist quantification of parameter uncertainty in system identification." *Mech. Syst. Signal Process.*, 29, 328–342.

Au, S.-K. (2014a). "Uncertainty law in ambient modal identification—Part I: Theory." *Mech. Syst. Signal Process.*, 48(1–2), 15–33.

Au, S.-K. (2014b). "Uncertainty law in ambient modal identification—Part II: Implication and field verification." *Mech. Syst. Signal Process.*, 48–34 ,(2–1)48.

Au, S.-K., and Zhang, F. L. (2012). "Fast Bayesian ambient modal identification incorporating multiple setups." *J. Eng. Mech.*, 10.1061/(ASCE)EM.1943-7889.0000385, 800–815.

Bickel, P. J., and Freedman, D. A. (1981). "Some asymptotic theory for the bootstrap." *Ann. Stat.*, 9(6), 1196–1217.

Chang, M., and Pakzad, S. N. (2013). "Observer Kalman filter identification for output-only systems using interactive structural modal identification toolsuite (SMIT)." *J. Bridge Eng.*, 10.1061/(ASCE)BE.1943-5592.0000530, 1–11.

Charalambous, C. D., and Logothetis, A. (2000). "Maximum likelihood parameter estimation from incomplete data via the sensitivity equations: The continuous-time case." *IEEE Trans. Autom. Control*, 45(5), 928–934.

Eddins, S., and Shure, L. (2001). "Matrix indexing in MATLAB." Mathworks Newsletter, ⟨http://www.mathworks.com/company/newsletters/articles/matrix-indexing-in-matlab.html⟩ (Sep. 13, 2014).

Efron, B. (1979). "Bootstrap methods: Another look at the jacknife." *Ann. Stat.*, 7(1), 1–26.

Efron, B., and Tibshirani, R. (1986). "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy." *Stat. Sci.*, 1(1), 54–75.

Guillaume, P., Verboven, P., and Vanlanduit, S. (1998). "Frequency-domain maximum likelihood identification of modal parameters with confidence intervals." *Proc., Int. Conf. on Noise and Vibration Engineering*, Vol. 1, Katholieke Universiteit Leuven, Leuven, Belgium, 359–366.

Gupta, N., and Mehra, R. (1974). "Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculations." *IEEE Trans. Autom. Control*, 19(6), 774–783.

He, X., and De Roeck, G. (1997). "System identification of mechanical structures by a high-order multivariate autoregressive model." *Comput. Struct.*, 64(1–4), 341–351.

Hinton, G. E., and Ghahramani, Z. (1996). "Parameter estimation for linear dynamical systems." *Technical Rep. CRG-TR-96-2*, Univ. of Toronto, Toronto.

Juang, J.-N., and Phan, M. Q. (2001). *Identification and control of mechanical systems*, Cambridge University Press, Cambridge, U.K.

Kalman, R. E. (1960). "A new approach to linear filtering and prediction problems." *Trans. ASME J. Basic Eng.*, 82(1), 35–45.

King, G. (1989). *Unifying political methodology: The likelihood theory of statistical inference*, Cambridge University Press, Cambridge, U.K.

Klein, A., and Neudecker, H. (2000). "A direct derivation of the exact fisher information matrix of gaussian vector state space models." *Linear Algebra Appl.*, 321(1–3), 233–238.

Kwasniewski, L., Wekezer, J., Roufa, G., Li, H., Ducher, J., and Malachowski, J. (2006). "Experimental evaluation of dynamic effects for a selected highway bridge." *J. Perform. Constr. Facil.*, 20(3), 253–260.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*, Sage, Thousand Oaks, CA.

Mahata, K., Pintelon, R., and Schoukens, J. (2006). "On parameter estimation using nonparametric noise models." *IEEE Trans. Autom. Control*, 51(10), 1602–1612.

Masset, P. (2008). "Analysis of financial time-series using Fourier and wavelet methods." 1–36.

Matarazzo, T. J., and Pakzad, S. N. (2015a). "Structural identification using expectation maximization (STRIDE): An iterative output-only method for modal identification." *J. Eng. Mech.*, 10.1061/(ASCE)EM.1943-7889.0000951, in press.

Matarazzo, T. J., and Pakzad, S. N. (2015b). "Structural modal identification for mobile sensing with missing data." *J. Eng. Mech.*, in press.

Matarazzo, T. J., Shahidi, S. G., Chang, M., and Pakzad, S. N. (2015). "Are today's SHM procedures suitable for tomorrow's BIGDATA?" *Proc., Society of Experimental Mechanics IMAC XXXIII, Structural Health Monitoring and Damage Detection*, Springer, Heidelberg, 59–65.

Meng, X.-L., and Rubin, D. B. (1991). "Variance-covariance using EM to obtain asymptotic matrices: The SEM algorithm." *J. Am. Stat. Assoc.*, 86(416), 899–909.

Miller, R. G. (1974). "The jackknife—A review." *Biometrika*, 61(1), 1–15.

Oehlert, G. W. (1992). "A note on the delta method." *Am. Stat.*, 46(1), 27–29.

Pakzad, S. N., Dryden, M., and Fenves, G. L. (2009). "Parametric bootstrap for system identification of a scaled reinforced concrete bridge." *Structures Congress*, L. Griffis, T. Helwig, M. Waggoner, and M. Hoit, eds., ASCE, Reston, VA, 397–405.

Pakzad, S. N., and Fenves, G. L. (2009). "Statistical analysis of vibration modes of a suspension bridge using spatially dense wireless sensor network." *J. Struct. Eng.*, 10.1061/(ASCE)ST.1943-541X.0000033, 863–872.

Pedersen, M. S., et al. (2012). "The matrix cookbook." ⟨http://matrixcookbook.com⟩ (May 12, 2014).

Peeters, B., and De Roeck, G. (1999). "Reference-based stochastic subspace identification for output-only modal analysis." *Mech. Syst. Signal Process.*, 13(6), 855–878.

Pi, Y. L., and Mickleborough, N. C. (1989). "Modal identification of vibrating structures using ARMA model." *J. Eng. Mech.*, 10.1061/(ASCE)0733-9399(1989)115:10(2232), 2232–2250.

Pintelon, R., Guillaume, P., and Schoukens, J. (2007). "Uncertainty calculation in (operational) modal analysis." *Mech. Syst. Signal Process.*, 21(6), 2359–2373.

Quenouille, M. H. (1956). "Notes on bias in estimation." *Biometrika*, 43(3/4), 353–360.

Rauch, H. E., Striebel, C. T., and Tung, F. (1965). "Maximum likelihood estimates of linear dynamic systems." *AIAA J.*, 3(8), 1445–1450.

Reinert, G. (2009). *Statistical theory*, Univ. of Oxford, Oxford, U.K.

Rubin, D. B., Dempster, A. P., and Laird, N. M. (1977). "Maximum likelihood from incomplete data via the EM algorithm." *J. R. Stat. Soc.*, 39(1), 1–38.

Shahidi, S. G., Pakzad, S. N., Ricles, J. M., Martin, J. R., Olgun, C. G., and Godfrey, E. A. (2015). "Behavior and damage of the Washington monument during the 2011 Mineral, Virginia, earthquake." *Geol. Soc. Am.*, 509, 235–252.

Shumway, R. H., and Stoffer, D. S. (1981). *An approach to time series smoothing and forecasting using the EM algorithm*, Univ. of California, Davis, Davis, CA.

Shumway, R. H., and Stoffer, D. S. (1982). "An approach to time series smoothing and forecasting using the EM algorithm." *J. Time Ser. Anal.*, 3(4), 253–264.

Shumway, R. H., and Stoffer, D. S. (2006). *Time series analysis and its applications—With R examples*, Springer, New York.
Shumway, R. H., and Stoffer, D. S. (2011). *Time series analysis and its applications with R examples*, Springer, New York.
Stoffer, D. S., and Wall, K. D. (1991). "Bootstrapping state-space models: Gaussian and the Kalman filter maximum likelihood estimation." *J. Am. Stat. Assoc.*, 86(416), 1024–1033.
Stuart, A., Ord, J. K., and Arnold, S. (1999). *Kendall's advanced theory of statistics, classical inference and the linear model*, Vol. 2A, Arnold, London.
Van Overschee, P., and De Moor, B. (1992). "N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems." *Automatica*, 30(1), 75–93.
Verboven, P., Guillaume, P., Cauberghe, B., Vanlanduit, S., and Parloo, E. (2004). "Modal parameter estimation from input-output Fourier data using frequency-domain maximum likelihood identification." *J. Sound Vib.*, 276(3–5), 957–979.
Wu, C. F. J. (1983). "On the convergence properties of the EM algorithm." *Ann. Stat.*, 11(1), 95–103.