A probabilistic approach to vibration-based modal identification using mobile sensors

by

Rajdip Nayek

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Civil Engineering

Waterloo, Ontario, Canada, 2019

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Dr. Juan M. Caicedo Professor, Chair, Dept. of Civil & Environmental Engineering University of South Carolina
Supervisor:	Dr. Sriram Narasimhan Professor, Department of Civil & Environmental Engineering, University of Waterloo
Internal Members:	Dr. Adil Al-Mayah Associate Prof., Dept. of Civil & Environmental Engineering, University of Waterloo
	Dr. Giovanni Cascante Professor, Department of Civil & Environmental Engineering, University of Waterloo
Internal-External Member:	Dr. Amir Khajepour Professor, Dept. of Mechanical & Mechatronics Engineering, University of Waterloo

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

The contents in Chapter 3 of this thesis have been published in the following journal article co-authored by myself, Dr. Suparno Mukhopadhyay (Assistant Professor at Indian Institute of Technology, Kanpur), and my supervisor.

R. Nayek, S. Mukhopadhyay and S. Narasimhan. "Mass normalized mode shape identification of bridge structures using a single actuator-sensor pair". *Structural Control and Health Monitoring* 25.11 (2018): e2244

The methodology was jointly developed by myself and Dr. Mukhopadhyay. All numerical and experimental studies were conducted by me. Dr. Mukhopadhyay also provided editorial assistance in writing the manuscript.

Section 6.3 of Chapter 6 has been included in the following journal article co-authored by myself, Dr. Souvik Chakraborty (Postdoctoral fellow, University of Notre Dame), and my supervisor.

R. Nayek, S. Chakraborty and S. Narasimhan. A Gaussian process latent force model for joint input-state estimation in linear structural systems. *Mechanical Systems and Signal Processing* 128: 497-530, (2019)

The development of the methodology and simulation of numerical examples was conducted by me. Dr. Chakraborty provided technical guidance on Gaussian process latent force modelling and provided editorial assistance in writing the manuscript.

Abstract

Vibration-based modal identification is the process of analyzing vibration measurements in order to identify the modal properties of the structure, namely the modal frequencies, damping ratios, and mode shapes. The quality of the estimated modal parameters, particularly the mode shapes, depends to a great extent on the spatial resolution of the measurements on the structure. Traditionally, measurements are obtained using networks of static sensors (i.e., sensors that remain fixed at certain locations on the structure) which are prone to poor spatial resolution when using a few sensors. To circumvent this issue, a large, dense network is required which has negative cost and implementation implications. An appealing alternative to large static sensor networks is the use of a much smaller network of mobile sensors. Mobile sensors are sensors mounted on carrier vehicles (such as robots or cars) that can be sequentially conveyed to various locations on a structure, thereby achieving dense spatial resolution with relatively few sensors.

Two measurement strategies can be used with mobile sensors: re-configurable mobile sensing, which involves repositioning the mobile sensors in a series of static configurations and collecting measurement data while not in motion, and in-motion mobile sensing which involves continuously collecting measurement data while the sensors traverse the spatial domain of the structure. Conventional modal identification methods – developed for use with measurement data from static sensor networks – are not directly compatible with data obtained from mobile sensor networks. This motivates the need to develop new modal identification algorithms specifically to process mobile sensor data. Furthermore, due to various factors such as motion over rough surfaces, data transmission errors, vehicle-structure interaction, mobile sensors are typically subjected to more sources of error compared to static sensors. Therefore, quantifying the uncertainties in the modal parameter estimates is important and should be included in the algorithms using mobile sensor data.

The current state of research on system identification methods using mobile sensors is still developing. This thesis addresses some of the theoretical and algorithmic challenges encountered in system identification using both re-configurable and in-motion mobile sensing strategies. In the context of re-configurable sensing, the feasibility of using a single mobile actuator-sensor pair for input-output modal identification of a structure is studied.

An input-output balance method is developed and it is shown that high-resolution mass-normalized mode shapes can be obtained with a single mobile actuator-sensor pair and the approximate knowledge of the signs of the mode shapes.

The problem of output-only modal identification using in-motion mobile sensors is addressed from two perspectives: an incomplete-data perspective and a complete-data perspective. From an incomplete data perspective, the data matrix generated by mobile sensors is treated as a static sensor data matrix with missing entries. A Bayesian inference framework based on a stochastic linear time-invariant state-space model is introduced to obtain the posterior distribution over the modal parameters. Three computational algorithms, namely the Expectation-Maximization (EM), the Variational Bayes (VB) and the Gibbs Sampler (GS) are employed for modal parameter estimation from the data matrix with missing entries. The EM provides point estimates whereas VB and GS provide posterior distributions of the modal parameters. From a complete-data perspective, the mobile sensor data obtained from a network of in-motion mobile sensors are represented by a single stacked data matrix characterized by spatial discontinuities along each sensor channel. To facilitate modal parameter estimation with the stacked data matrix, a novel stochastic linear time-varying modal state model (MSM) is proposed. The EM, VB and GS algorithms are suitably modified to facilitate inference of the modal parameters via the MSM. The modal parameter estimation framework employing the MSM is shown to be flexible and capable of providing high-resolution mode shape estimates along with posterior uncertainties.

Finally, a methodology to suppress the effect of carrier-vehicle dynamics on the recorded mobile sensor responses is proposed. The vehicle responses (i.e., responses recorded by a sensor mounted on top of the carrier-vehicle) are modified versions of the true contact-point structural responses due to the low-pass filtering effect of the vehicle dynamics. Contact-point responses – the responses at the contact point of the vehicle with the structure – are deemed to be better suited for modal identification than measured vehicle responses. Using the knowledge of the vehicle system, the problem of estimating the contact-point response from the measured vehicle response is cast as an input reconstruction problem, and a Gaussian process latent force model (GPLFM) with Kalman filtering is proposed to recover the contact-point responses. Through various numerical studies, it is shown

that the recovered contact-point responses are superior – the effect of vehicle dynamics is reduced and the prominence of higher structural modes is significantly improved.

In summary, this thesis presents a suite of numerical algorithms for modal parameter estimation using mobile sensor data. The performances of the proposed numerical algorithms are evaluated through a series of numerical simulations and bench-scale experimental tests. The results (i) confirm the advantages of mobile sensing over large static sensor networks, (ii) verify the different perspectives of analyzing mobile sensor data, and (iii) demonstrate the capability of the algorithms to include uncertainty quantification.

Acknowledgements

All along my PhD journey I have been fortunate to receive help and support from various people. Attempting to thank all of them will be an insurmountable task. This page provides me the opportunity to express my sincere gratitude to at least a few of them.

First of all, I would like to thank my supervisor Prof. Sriram Narasimhan for providing me the opportunity to pursue doctoral studies at the University of Waterloo and for allowing me independence in doing my research. His constant guidance and unrestrained encouragement during our countless meetings have always instilled in me the confidence to pursue and develop my ideas.

I would like to thank Prof. Suparno Mukhopadhyay with whom I collaborated during the years 2016-17. My collaboration with him has since then grown into a friendship which I will cherish forever. His guidance and critical questioning skills have played a significant role in the development of the work presented in Chapter 4. A fruitful collaboration also developed with Dr. Souvik Chakraborty in the spring of 2018 with whom I had regular discussions on various machine learning tools.

I am grateful to my PhD defense committee members, Prof. Juan Caicedo, Prof. Adil Al-Mayah, Prof. Giovanni Cascante, and Prof. Amir Khajepour for reviewing my thesis and offering me valuable suggestions.

Prof. CS Manohar, my Master's advisor, was the first to formally introduce me to research. I would like to take this opportunity to thank him for his inspiration which has motivated me to not only pursue doctoral studies but also to strive to be a better person in life.

I owe my deepest gratitude to my colleague Dr. Kevin Goorts, who has not only been a wonderful friend but also a great mentor since the beginning of my PhD journey. My heartfelt thanks goes to various past and present members of the Structural Dynamics, Identification and Control (SDIC) lab; Dylan Dowling, for taking time out of his busy schedule and helping me with the DIC tests; Stan Fong, for being a good friend and also showing me how to operate the linear shaker; Dirk Friesen, for helping me out with the lab test setups; Prof. Pampa Dey, for her encouragement and inspiration, and Prof. Guru

Prakash, for inspiring me with his spiritual outlook on life. Thanks are also due to my friends (in no particular order) Piyius, Rakesh, Marshal, Endrina, Roya, Nick, Stephen, Wanis, Thiru, Indranil, Shankha, and Jobanmeet; it has always been fun being around them. Also, thanks to Prof. Budhaditya Hazra, Dr. VS Sundar, Dr. Subhayan De, and Dr. Jinane Harmouche for providing me encouragement in times of need. Special thanks are reserved for Subhajit Majhi, my housemate, with whom I shared many unforgettable memories in the last five years. His brotherly support through all these years is much appreciated.

Last but not the least, I would like to thank my parents for their unconditional love from across the world. They have always trusted and encouraged me in my endeavours. My mother, in particular, has played an instrumental role in keeping me sane throughout my PhD journey. Without her unwavering love and support, this journey would not have been possible.

Dedication

This is dedicated to my mother.

Table of Contents

List of Tables

List of Figures

List of Symbols

Notational conventions

$\mathbf{A}^{-1}$	inverse of matrix
$a_{ij}, a_{i,j}$	element at ith row and jth column of matrix $\mathbf{A}$
$\boldsymbol{a}_j$	jth column of matrix $\mathbf{A}$
$\left\|\mathbf{A}\right\|$	determinant of matrix $\mathbf{A}$
$\mathrm{tr}\left\{\mathbf{A}\right\}$	trace of matrix $\mathbf{A}$
$\mathrm{vec}\left(\mathbf{A}\right)$	vectorized form of matrix $\mathbf{A}$
$\mathrm{COV}(\mathbf{x})$	covariance of random variable $\boldsymbol{x}$
$\frac{d\boldsymbol{x}}{dt}$	time derivative of $\boldsymbol{x}(t)$
$\frac{\partial g_i(\boldsymbol{x})}{\partial x_j}$	partial derivative of g_i with respect to x_j
$p\left(\boldsymbol{x}\right)$	probability density function of random variable $\boldsymbol{x}$
$\mathrm{Re}\left(a\right)$, $\mathrm{Im}\left(a\right)$	real and imaginary parts of a complex number a
$\boldsymbol{x} \sim p\left(\boldsymbol{x}\right)$	random variable $\boldsymbol{x}$ has probability density function $p\left(\boldsymbol{x}\right)$
$\left\langle \boldsymbol{x} \right\rangle$	expectation of $\boldsymbol{x}$; $\left\langle \boldsymbol{x} \right\rangle = \int \boldsymbol{x} p\left(\boldsymbol{x}\right) d\boldsymbol{x}$
$\left\langle \boldsymbol{x} \right\rangle_{q(\boldsymbol{x})}$	expectation of $\boldsymbol{x}$ w.r.t. $q\left(\boldsymbol{x}\right)$; $\left\langle \boldsymbol{x} \right\rangle_{q(\boldsymbol{x})} = \int \boldsymbol{x} q\left(\boldsymbol{x}\right) d\boldsymbol{x}$
$\left\langle \boldsymbol{x} \middle\| \boldsymbol{y} \right\rangle$	conditional expectation of $\boldsymbol{x}$ given $\boldsymbol{y}$; $\left\langle \boldsymbol{x} \middle\| \boldsymbol{y} \right\rangle = \int \boldsymbol{x} p\left(\boldsymbol{x} \middle\| \boldsymbol{y}\right) d\boldsymbol{x}$
$\otimes$	Kronecker product

Common symbols

$\mathbf{A}_c$, $\mathbf{B}_c$	continuous-time transition and input state space matrices
$\mathbf{A}$, $\mathbf{B}$	discrete-time transition and input state space matrices
$\boldsymbol{d}(t), \dot{\boldsymbol{d}}(t), \ddot{\boldsymbol{d}}(t)$	vector of structural displacements, velocities and acceleration

$\boldsymbol{d}_0, \dot{\boldsymbol{d}}_0$	vector of initial displacements and velocities
$\mathcal{F}\left(\boldsymbol{\theta}\middle\vert\boldsymbol{\theta}^{(j)}\right)$	expected value of log-likelihood function $\mathcal{L}(\boldsymbol{\theta})$ given $\boldsymbol{Y}$ and $\boldsymbol{\theta}^{(j)}$
$\mathcal{F}_{ve}$	variational free energy
$\mathbf{G}$, $\mathbf{J}$	observation and feed-through matrices
$\mathbf{I}$	identity matrix
$\mathbf{K}_k$, $\boldsymbol{e}_k$, $\mathbf{S}_k$	Kalman gain matrix, innovation vector and its covariance matrix at the kth time step of Kalman filtering
$\mathrm{KL}(q(\boldsymbol{x})\|\|p(\boldsymbol{x}))$	Kullback-Liebler divergence between distributions $q(\boldsymbol{x})$ and $p(\boldsymbol{x})$
$\mathcal{L}(\boldsymbol{\theta})$	complete-data log likelihood $(= \ln p(\boldsymbol{Y}, \boldsymbol{X}\|\boldsymbol{\theta}))$
$\mathbf{L}$	input location matrix
L	length of a bridge
$\mathbf{M}, \mathbf{C}, \mathbf{K}$	mass, damping and stiffness matrices
$\mathcal{N}(\cdot)$	Gaussian probability density function
$\mathbf{N}_k$	smoother gain matrix at the kth time step of RTS smoothing
N	total discrete time samples
n	degrees of freedom of the structural model
n_m	number of participating modes
n_{ms}	number of mobile sensors
n_s	dimension of state vector $\boldsymbol{x}$
n_u	dimension of input vector $\boldsymbol{u}$
n_o	size of observation vector $\boldsymbol{y}_k$
$\boldsymbol{p}_r$	rth force participation factor
$p(\boldsymbol{x}_{k+1}\|\boldsymbol{x}_k, \boldsymbol{\theta})$	probability density function of $\boldsymbol{x}_{k+1}$ given $\boldsymbol{x}_k$ and $\boldsymbol{\theta}$ are known
$p(\boldsymbol{y}_k\|\boldsymbol{x}_k, \boldsymbol{\theta})$	probability density function of $\boldsymbol{y}_k$ given $\boldsymbol{x}_k$ and $\boldsymbol{\theta}$ are known
$p(\boldsymbol{Y})$	joint probability density function over $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N\}$
$p(\boldsymbol{Y}, \boldsymbol{X})$	joint probability density function over
$\mathbf{Q}$	process noise covariance
$q(\cdot)$	variational distribution
$q^{(j)}(\cdot)$	variational distribution at the jth iteration
$\mathbb{R}^{a\times b}$	space of real $a \times b$ matrices
$\mathbf{R}$	measurement noise covariance

$\mathbf{S}_{dis}, \mathbf{S}_{vel}, \mathbf{S}_{acc}$	selection matrices for displacements, velocities and accelerations
Δt	sampling period
$\boldsymbol{u}(t)$	vector of input forces
$\hat{\mathbf{V}}_{k\|k}$	conditional covariance of $\boldsymbol{x}_k$ given measurements $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k\}$
$\hat{\mathbf{V}}_{k\|N}$	conditional covariance of $\boldsymbol{x}_k$ given measurements $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N\}$
$\hat{\mathbf{V}}_{k+1,k\|N}$	lag-one smoother covariance matrix of $\boldsymbol{x}_{k+1}$ and $\boldsymbol{x}_k$ given measurements $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N\}$
$\boldsymbol{v}_k$	measurement noise vector at kth time step
$\boldsymbol{w}_k$	process noise vector at kth time step
$\hat{\boldsymbol{x}}_{k\|N}$	conditional mean of $\boldsymbol{x}_k$ given measurements $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N\}$
$\boldsymbol{X}$	sequence of states $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{N+1}\}$
$\boldsymbol{X}^{(j)}$	sequence of states $\boldsymbol{X}$ at the jth iteration
$\boldsymbol{x}(t)$	continuous-time state vector
$\boldsymbol{x}_k$	discrete-time state vector at kth time step
$\hat{\boldsymbol{x}}_{k\|k}$	conditional mean of $\boldsymbol{x}_k$ given measurements $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k\}$
$\boldsymbol{Y}$	sequence of measurements $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N\}$
$\boldsymbol{y}(t)$	continuous-time measurement vector
$\boldsymbol{y}_k$	discrete-time measurement vector at kth time step
$\boldsymbol{\eta}(t)$	vector of modal displacements
η_r	rth modal displacement
$\boldsymbol{\Phi}$	mass-normalized mode shape matrix
$\boldsymbol{\phi}_r$	rth mass normalized mode shape vector
$\boldsymbol{\phi}_r^{un}$	rth un-normalized mode shape vector
ω_r	rth modal frequency
$\pi_{pr}(\cdot)$	prior probability density function
ρ, E, I	mass-density, Young's modulus and moment of inertia
ξ_r	rth modal damping ratio
$\hat{\boldsymbol{\theta}}_{\text{MAP}}$	maximum *a posteriori* estimate of $\boldsymbol{\theta}$
$\boldsymbol{\theta}$	set of state space system parameters
$\boldsymbol{\theta}^{(j)}$	set of state space system parameters at the jth iteration
$\boldsymbol{\theta}^{(0)}$	initial set of state space system parameters

$\mathbf{0}$	zero matrix

A list of chapter-specific notations are listed as follows:

Chapter 3

$\mathbf{C}_m$	diagonalized damping matrix
$\mathcal{D}^{\text{NCAS}}$	set of input-output pairs for NCAS case
$\mathcal{D}^{\text{CAS}}$	set of input-output pairs for CAS case
$\underline{\boldsymbol{d}}^{(s_j,s_i)}$	vector of sampled measurements at s_j due to input at s_i
$\mathbf{F}^{(s_i)}$	matrix of discrete-time modal responses due to input at station s_i
$f_{\text{exp}}, \xi_{\text{exp}}$	experimentally determined frequencies and damping ratios
$\mathbf{I}(\cdot)$	indicator function
$\mathcal{P}$	set of all stations
P	total number of stations
$\boldsymbol{\Lambda}$	diagonal matrix of eigenvalues
s_i	ith station on beam
$u^{(s_i)}$	input excitation provided at station s_i
$y^{(s_j,s_i)}$	output measured at station s_j due to input at input at s_i
$\boldsymbol{\alpha}^{(s_j,s_i)}$	estimate of product of mode shape ordinates
$\tilde{\eta}_r, \dot{\tilde{\eta}}_r, \ddot{\tilde{\eta}}_r$	pseudo-modal displacement, velocity and acceleration of the rth mode
$\phi_{s_i,r}\phi_{s_j,r}$	product of s_ith and s_jth mode shape components of the rth mode
$\phi_{s_i,r}$	true value of rth mode shape component at station s_i
$\hat{\phi}_{s_i,r}$	estimate of rth mode shape component at station s_i
$\texttt{sgn}(\cdot)$	signum function
$\vartheta, \vartheta_1, \vartheta_2$	magnitudes of perturbation

Chapter 4

$\mathbf{A}^{\text{SSI}}$, $\mathbf{G}^{\text{SSI}}$	estimate of state space matrices $\mathbf{A}$ and $\mathbf{G}$ from subspace system identification (SSI) method
$\mathbf{A}^{(j)}, \mathbf{G}^{(j)}, \mathbf{Q}^{(j)}, \mathbf{R}^{(j)}$	estimated state space parameters at the jth iteration of EM

$\widetilde{\mathbf{A}}, \widetilde{\mathbf{G}}, \widetilde{\mathbf{Q}}, \widetilde{\mathbf{R}}$	parameters of the modified state space model for Kalman filtering within VB algorithm
b	subsampling period in GS
$\mathbf{D}_{\mathbf{Q}}$, $d_{\mathbf{Q}}$	positive definite scale matrix and degrees of freedom associated with $\mathcal{IW}$ distribution over $\mathbf{Q}$
$\mathbf{D}_{\mathbf{R}}$, $d_{\mathbf{R}}$	positive definite scale matrix and degrees of freedom associated with $\mathcal{IW}$ distribution over $\mathbf{R}$
$\mathbf{D}_{\mathbf{Q}}^{(j)}$, $d_{\mathbf{Q}}^{(j)}$	positive definite scale matrix and degrees of freedom associated with variational $\mathcal{IW}$ distribution over $\mathbf{Q}$ at the jth iteration of VB
$\mathbf{D}_{\mathbf{R}}^{(j)}$, $d_{\mathbf{R}}^{(j)}$	positive definite scale matrix and degrees of freedom associated with variational $\mathcal{IW}$ distribution over $\mathbf{R}$ at the jth iteration of VB
D	number of independent Markov chains simulated for multiple imputation procedure
$\mathcal{E}_{av}^{(j)}$	average energy at the jth VB iteration
$\hat{\mathcal{F}}^{(j)}$	estimate of log-likelihood at jth iteration of EM
$\mathcal{F}_{ve}^{(j)}$	variational free energy at the jth VB iteration
$\mathcal{H}^{(j)}$	differential entropy at the jth VB iteration
$\mathcal{IW}(\mathbf{\Psi}, \nu)$	inverse Wishart distribution with positive definite scale matrix $\mathbf{\Psi}$ and degrees of freedom ν
J	total number of samples simulated in a Markov chain using GS
$\mathbf{K}_k^*$, $\boldsymbol{e}_k^*$, $\mathbf{S}_k^*$	Kalman gain matrix, innovation vector and its covariance matrix at the kth time step of Kalman filtering (in EM for missing data)
$\widetilde{\mathbf{K}}_k^*$, $\widetilde{\boldsymbol{e}}_k^*$, $\widetilde{\mathbf{S}}_k^*$	Kalman gain matrix, innovation vector and its covariance matrix at the kth time step of Kalman filtering (in VB for missing data)
$\mathcal{MN}(\mathbf{M}, \mathbf{U}, \mathbf{V})$	matrix normal distribution with mean matrix $\mathbf{M}$, positive definite scale matrices $\mathbf{U}$ and $\mathbf{V}$
$\mathbf{N}_k^*$	smoother gain matrix at the kth time step of RTS smoothing (in EM for missing data)
$\widetilde{\mathbf{N}}_k^*$	smoother gain matrix at the kth time step of RTS smoothing (in VB for missing data)
n_{burnin}	number of burn-in samples discarded in the Markov chain

$\mathbf{P}_{xx}^{(d,j)}, \mathbf{P}_{fx}^{(d,j)}, \mathbf{P}_{ff}^{(d,j)}$	sufficient statistics using states at the jth sampling step for dth imputation a in GS algorithm
$\mathbf{P}_{yx}^{(d,j)}, \mathbf{P}_{yy}^{(d,j)}$	sufficient statistics using states and measurements at the jth sampling step for dth imputation a in GS algorithm
$\mathbf{S}_{ff}$, $\mathbf{S}_{fx}$, $\mathbf{S}_{xx}$	sufficient statistics using states in EM algorithm
$\mathbf{S}_{yx}$, $\mathbf{S}_{yy}$	sufficient statistics using states and measurements for EM algorithm
$\mathbf{S}_{ff}^{(j)}$, $\mathbf{S}_{fx}^{(j)}$, $\mathbf{S}_{xx}^{(j)}$	sufficient statistics using states with respect to $q^{(j)}(\boldsymbol{X})$ for VB algorithm
$\mathbf{S}_{yx}^{(j)}$, $\mathbf{S}_{yy}^{(j)}$	sufficient statistics using states and measurements with respect to $q^{(j)}(\boldsymbol{X})$ for VB algorithm
$\boldsymbol{Y}_{miss}, \boldsymbol{Y}_{obs}$	sequence of missing data and observed data
$\boldsymbol{Y}_{miss}^{(d)}$	sequence of missing data using d
$\boldsymbol{y}_k^{obs}, \boldsymbol{y}_k^{miss}$	vector of non-missing and missing observations extracted from $\boldsymbol{y}_k$
ϵ_{tol}	stopping tolerance threshold for iterations (in EM and VB)
$\Gamma_p(\cdot)$	multivariate gamma function
$\gamma_p(\cdot)$	multivariate digamma function
$\boldsymbol{\upsilon}_k^{(1)}, \mathbf{I}_k^{(1)}$	permutation matrices for observed data
$\boldsymbol{\upsilon}_k^{(2)}, \mathbf{I}_k^{(2)}$	permutation matrices for missing data
$\boldsymbol{\mu}_{\mathbf{A}}$, $\boldsymbol{\mu}_{\mathbf{G}}$	mean matrix hyperparameters of $\mathcal{MN}$ distributions over random matrices $\mathbf{A}$ and $\mathbf{G}$
$\boldsymbol{\mu}_{\mathbf{A}}^{(j)}$, $\boldsymbol{\mu}_{\mathbf{G}}^{(j)}$	mean matrix hyperparameters of variational $\mathcal{MN}$ distributions over random matrices $\mathbf{A}$ and $\mathbf{G}$ at the jth iteration of VB
$\boldsymbol{\mu}_{\mathbf{A}}^*$, $\boldsymbol{\mu}_{\mathbf{G}}^*$, $\mathbf{D}_{\mathbf{Q}}^*$, $\mathbf{D}_{\mathbf{R}}^*$, $\boldsymbol{\Pi}^*$, $d_{\mathbf{Q}}^*$, $d_{\mathbf{R}}^*$	converged hyperparameters of the variational distributions over $\mathbf{A}$, $\mathbf{G}$, $\mathbf{Q}$, $\mathbf{R}$
$\overline{\boldsymbol{\mu}}_{\mathbf{A}}^{(d,j)}, \overline{\boldsymbol{\mu}}_{\mathbf{G}}^{(d,j)}, \overline{\mathbf{D}}_{\mathbf{Q}}^{(d,j)}, \overline{\mathbf{D}}_{\mathbf{R}}^{(d,j)}, \overline{\boldsymbol{\Pi}}^{(d,j)}, \bar{d}_{\mathbf{Q}}^{(d,j)}, \bar{d}_{\mathbf{R}}^{(d,j)}$	hyperparameters of the sampling distribution of $\mathbf{A}$, $\mathbf{G}$, $\mathbf{Q}$ and $\mathbf{R}$ at the jth sampling step for dth imputation using GS
$\boldsymbol{\mu}_1, \mathbf{V}_1$	mean vector and covariance matrix of initial state $\boldsymbol{x}_1$
$\boldsymbol{\mu}_1^{(j)}$, $\mathbf{V}_1^{(j)}$	mean vector and covariance matrix of initial state $\boldsymbol{x}_1$ at jth EM iteration

μ_{f_i}, μ_{ξ_i}, $\boldsymbol{\mu}_{\boldsymbol{\phi}_i^{un}}$, $\sigma^2_{f_i}$, $\sigma^2_{\xi_i}$, $\boldsymbol{\Sigma}_{\boldsymbol{\phi}_i^{un}}$	hyperparameters of the posterior distributions over ith modal frequency, damping ratio and un-normalized mode shape vector
$\hat{\boldsymbol{\Phi}}_{\text{ML}}$	maximum likelihood estimate of matrix of mode shapes
$\boldsymbol{\Pi}$	scale matrix hyperparameter shared by $\mathcal{MN}$ distributions of random matrices $\mathbf{A}$ and $\mathbf{G}$
$\boldsymbol{\Pi}^{(j)}$	scale matrix hyperparameter shared by $\mathcal{MN}$ variational distributions of random matrices $\mathbf{A}$ and $\mathbf{G}$ at the jth iteration of VB
$\boldsymbol{\theta}_h$	hyperparameters of probability distribution over state space parameters
$\boldsymbol{\theta}_h^{(j)}$	hyperparameters of variational probability distribution over state space parameters at jth VB iteration
$\hat{\boldsymbol{\theta}}_{\text{ML}}$	maximum likelihood estimate of $\boldsymbol{\theta}$

Chapter 5

$\overline{\mathbf{A}}, \overline{\mathbf{G}}$	transition and observation matrices of MSM
$\overline{\mathbf{A}}^{\text{SSI}}$, $\overline{\mathbf{G}}^{\text{SSI}}$	estimate of state space matrices $\overline{\mathbf{A}}$ and $\overline{\mathbf{G}}$ from SSI
$\overline{\mathbf{A}}^{(j)}, \overline{\mathbf{G}}^{(j)}, \overline{\mathbf{Q}}^{(j)}, \mathbf{R}^{(j)}$	estimated MSM parameters at the jth iteration of EM
$\widetilde{\overline{\mathbf{A}}}, \widetilde{\overline{\mathbf{G}}}, \widetilde{\overline{\mathbf{Q}}}, \widetilde{\overline{\mathbf{R}}}$	parameters of the modified state space model for Kalman filtering within VB algorithm
$\bar{\boldsymbol{a}}$, $\bar{\boldsymbol{g}}$	vectorized form of matrices $\overline{\mathbf{A}}$ and $\overline{\mathbf{G}}$
$\mathbf{D}_{\overline{\mathbf{Q}}}$, $d_{\overline{\mathbf{Q}}}$	positive definite scale matrix and degrees of freedom associated with $\mathcal{IW}$ distribution over $\overline{\mathbf{Q}}$
$\mathbf{D}_{\mathbf{R}}$, $d_{\mathbf{R}}$	positive definite scale matrix and degrees of freedom associated with $\mathcal{IW}$ distribution over $\mathbf{R}$
$\mathbf{D}_{\overline{\mathbf{Q}}}^{(j)}$, $d_{\overline{\mathbf{Q}}}^{(j)}$	positive definite scale matrix and degrees of freedom associated with variational $\mathcal{IW}$ distribution over $\overline{\mathbf{Q}}$ at the jth iteration of VB
$\mathbf{D}_{\mathbf{R}}^{(j)}$, $d_{\mathbf{R}}^{(j)}$	positive definite scale matrix and degrees of freedom associated with variational $\mathcal{IW}$ distribution over $\mathbf{R}$ at the jth iteration of VB
$\hat{\overline{\mathcal{F}}}^{(j)}$	estimate of log-likelihood at jth iteration of EM
n_α	number of VSLs
n_o	number of mobile sensors ($= n_{ms}$)

$\mathbf{P}^{(j)}_{\bar{x}\bar{x}}, \mathbf{P}^{(j)}_{\bar{f}\bar{x}}, \mathbf{P}^{(j)}_{\bar{f}\bar{f}}$	sufficient statistics using MSM states at jth sampling step of GS algorithm
$\mathbf{P}^{(j)}_{y\bar{x},k}, \mathbf{P}^{(j)}_{yy,k}$	sufficient statistics using MSM states and measurements at jth sampling step of GS algorithm
$\overline{\mathbf{Q}}$	covariance matrix of process noise vector $\overline{\boldsymbol{w}}$
$\mathbf{S}_{\bar{f}\bar{f}}$, $\mathbf{S}_{\bar{f}\bar{x}}$, $\mathbf{S}_{\bar{x}\bar{x}}$	sufficient statistics using MSM states in EM algorithm
$\mathbf{S}_{y\bar{x}}$, $\mathbf{S}_{yy}$	sufficient statistics using MSM states and measurements for EM algorithm
$\mathbf{S}^{(j)}_{ff}$, $\mathbf{S}^{(j)}_{fx}$, $\mathbf{S}^{(j)}_{xx}$	sufficient statistics using MSM states with respect to $q^{(j)}\left(\overline{\boldsymbol{X}}\right)$ for VB algorithm
$\mathbf{S}^{(j)}_{yx}$, $\mathbf{S}^{(j)}_{yy}$	sufficient statistics using MSM states and measurements with respect to $q^{(j)}\left(\overline{\boldsymbol{X}}\right)$ for VB algorithm
$\boldsymbol{s}^{\alpha}$	vector of locations of VSLs
$\boldsymbol{s}^{o}_{k}$	vector of mobile sensing locations at kth time step
Δs^{α}	distance between VSLs
$\hat{\overline{\mathbf{V}}}_{k\|k}$	conditional covariance of $\overline{\boldsymbol{x}}_k$ given measurements $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k\}$
$\hat{\overline{\mathbf{V}}}_{k\|N}$	conditional covariance of $\overline{\boldsymbol{x}}_k$ given measurements $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N\}$
$\hat{\overline{\mathbf{V}}}_{k+1,k\|N}$	lag-one smoother covariance matrix of $\overline{\boldsymbol{x}}_{k+1}$ and $\overline{\boldsymbol{x}}_k$ given measurements $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N\}$
$\overline{\boldsymbol{w}}_k$	process noise vector of the stochastic MSM at kth time step
$\overline{\boldsymbol{X}}$	sequence of states $\{\overline{\boldsymbol{x}}_1, \ldots, \overline{\boldsymbol{x}}_{N+1}\}$
$\overline{\boldsymbol{x}}_k$	state vector of the stochastic MSM at kth time step
$\hat{\overline{\boldsymbol{x}}}_{k\|k}$	conditional mean of $\overline{\boldsymbol{x}}_k$ given measurements $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k\}$
$\hat{\overline{\boldsymbol{x}}}_{k\|N}$	conditional mean of $\overline{\boldsymbol{x}}_k$ given measurements $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N\}$
$\overline{\boldsymbol{\mu}}_1, \overline{\mathbf{V}}_1$	mean vector and covariance matrix of initial state $\overline{\boldsymbol{x}}_1$
$\boldsymbol{\mu}_{\bar{\boldsymbol{a}}}$, $\boldsymbol{\mu}_{\bar{\boldsymbol{g}}}$	vectorized representations of matrices $\boldsymbol{\mu}_{\overline{\mathbf{A}}}$ and $\boldsymbol{\mu}_{\overline{\mathbf{G}}}$
$\boldsymbol{\mu}_{\bar{\boldsymbol{a}}}$, $\boldsymbol{\mu}_{\bar{\boldsymbol{g}}}$	mean vector hyperparameters of normal distributions over vectors $\bar{\boldsymbol{a}}$ and $\bar{\boldsymbol{g}}$
$\boldsymbol{\mu}^{(j)}_{\bar{\boldsymbol{a}}}$, $\boldsymbol{\mu}^{(j)}_{\bar{\boldsymbol{g}}}$	mean vector hyperparameters of variational $\mathcal{N}$ distributions over random vectors $\bar{\boldsymbol{a}}$ and $\bar{\boldsymbol{g}}$ at the jth iteration of VB

$\overline{\boldsymbol{\mu}}^*_{\bar{\boldsymbol{a}}}$, $\overline{\boldsymbol{\mu}}^*_{\bar{\boldsymbol{g}}}$, $\overline{\boldsymbol{\Sigma}}^*_{\bar{\boldsymbol{a}}}$, $\overline{\boldsymbol{\Sigma}}^*_{\bar{\boldsymbol{g}}}$, $\mathbf{D}^*_{\overline{\mathbf{Q}}}$, $\mathbf{D}^*_{\mathbf{R}}$, $d^*_{\overline{\mathbf{Q}}}$, $d^*_{\mathbf{R}}$	converged hyperparameters of the variational distributions over $\bar{\boldsymbol{a}}$, $\bar{\boldsymbol{g}}$, $\overline{\mathbf{Q}}$, $\mathbf{R}$
$\widetilde{\boldsymbol{\mu}}^{(j)}_{\bar{\boldsymbol{a}}}$, $\widetilde{\boldsymbol{\Sigma}}^{(j)}_{\bar{\boldsymbol{a}}}$, $\widetilde{\boldsymbol{\mu}}^{(j)}_{\bar{\boldsymbol{g}}}$, $\widetilde{\boldsymbol{\Sigma}}^{(j)}_{\bar{\boldsymbol{g}}}$, $\widetilde{\mathbf{D}}^{(j)}_{\overline{\mathbf{Q}}}$, $\widetilde{d}^{(j)}_{\overline{\mathbf{Q}}}$, $\widetilde{\mathbf{D}}^{(j)}_{\mathbf{R}}$, $\widetilde{d}^{(j)}_{\mathbf{R}}$	hyperparameters of the posterior distribution over $\bar{\boldsymbol{a}}$, $\bar{\boldsymbol{g}}$, $\overline{\mathbf{Q}}$, $\mathbf{R}$ updated using GS
$\boldsymbol{\Omega}_k$	mode shape regression (MSR) matrix at kth time step
$\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}$, $\boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}$	covariance matrix hyperparameters of normal distributions over vectors $\bar{\boldsymbol{a}}$ and $\bar{\boldsymbol{g}}$

Chapter 6

$\mathbf{A}_{vc}, \mathbf{B}_{vc}, \mathbf{G}_v$	parameters of continuous-time state space model of the vehicle
$\mathbf{A}_{vd}, \mathbf{B}_{vd}$	parameters of discrete-time state space model of the vehicle
$d_b(x,t)$	displacement of bridge at time t and space x from its left support
$d_{v,rel}$	vehicle displacement response relative to contact-point displacement response
$d_v(t), \dot{d}_v(t), \ddot{d}_v(t)$	displacement, velocity and acceleration of vehicle
$\ddot{d}_l(t), \ddot{d}_r(t)$	support excitations at left and right supports
EI, ρ_b, μ_b	flexural rigidity, mass and damping per unit length of bridge
$\boldsymbol{e}_k, \mathbf{S}_k$	innovation vector and its covariance at kth time step in Kalman filtering algorithm
$\mathbf{F}_{ac}, \mathbf{L}_{ac}, \mathbf{G}_{ac}$	parameters of continuous-time GPLFM
$\mathbf{F}_{ad}, \mathbf{L}_{ad}, \mathbf{G}_{ad}$	parameters of discrete-time GPLFM
$\mathbf{F}_{cf}, \mathbf{L}_{cf}, \mathbf{H}_{cf}$	parameters of continuous-time GP SSM
$\boldsymbol{f}_a(t), \boldsymbol{d}_a(t)$	augmented force and response vectors in vehicle-bridge interaction equation
$f(t)$	force acting on the bridge as a function of time
$\mathcal{GP}(0, \kappa)$	Gaussian process with zero mean function and covariance function κ
g	acceleration due to gravity
L	length (or span) of bridge

$\mathbf{M}_a(t), \mathbf{C}_a(t), \mathbf{K}_a(t)$	time-varying parameter matrices for vehicle-bridge interaction equation
m_v, c_v, k_v	mass, damping and stiffness of vehicle
$\mathbf{Q}_d$	covariance matrix of $\boldsymbol{w}_k$
t	continuous-time coordinate
$u_c(t)$	contact-point vehicle displacement response
v	speed of vehicle
$\widetilde{v}_k$	discrete-time measurement noise at kth time step
$\boldsymbol{w}(t)$	vector of continuous-time white noise process
$\boldsymbol{w}_k$	vector of discrete-time white noise at kth time step
$w(t)$	scalar white noise process with power spectral density σ_w
$\boldsymbol{x}_v(t)$	continuous-time vehicle state
$\boldsymbol{x}_c^a(t)$	continuous-time GPLFM state vector
$\boldsymbol{x}_k^a$	discrete-time GPLFM state vector at kth time step
x	continuous-space coordinate
$y_c(t,k), y_{v,k}$	discrete-time vehicle acceleration response
$\boldsymbol{z}(t)$	continuous-time GP SSM state vector
$\delta(\cdot)$	Dirac's delta function
$\kappa_{5/2}(\tau;\boldsymbol{\theta}_{cf})$	Matérn covariance function with smoothness parameter $\frac{5}{2}$ and hyperparameters $\boldsymbol{\theta}_{cf}$
$\hat{\boldsymbol{\theta}}_cf$	maximum likelihood estimate of $\boldsymbol{\theta}_{cf}$
ς, l	signal variance and lengthscale hyperparameters of $\boldsymbol{\theta}_{cf}$

Abbreviations

CAS	collocated actuator sensor
CDF	cumulative distribution function
CoV	coefficient of variation
DACS	deployable autonomous control system
DoF	degrees of freedom
EM	expectation maximization
ERA-OKID	eigensystem realization algorithm with observer Kalman filter identification
FE	finite element
FRF	frequency response function
GP	Gaussian process
GPLFM	Gaussian process latent force model
GS	Gibbs sampler
KLD	Kullback-Leibler divergence
LGSSM	linear Gaussian state space model
LSFD	least squares frequency domain
LTI	linear time-invariant
MAC	modal assurance criterion
MAP	maximum a posteriori
MAR	missing at random
MCMC	Markov chain Monte Carlo
MH	Metropolis-Hastings
ML	maximum likelihood

MNIW	Matrix-Normal-Inverse-Wishart
MSM	modal state model
MSR	mode shape regression
N4SID	numerical algorithm for subspace state space system identification
NCAS	non-collocated actuator sensor
PDF	probability density function
PoSD	probability of sign discrepancy
PSD	power spectral density
RMS	root-mean-square
RTS	Rauch-Tung-Striebel
SD	standard deviation
SNR	signal-to-noise ratio
SSI	stochastic subspace identification
SSID	structural system identification
SSM	state space model
TPM	truncated physical model
VB	variational Bayes
VBI	vehicle-bridge interaction
VSL	virtual static location

Chapter 1

Introduction

In recent years, there has been a notable increase in public awareness of aging infrastructure around the world due to its direct impact on human well-being and the global economy. For example, in 2017, the American Society of Civil Engineers (ASCE) reported that nearly 40% of the 614,387 bridges in the United States are at least 50 years old and that 9.1% of all bridges are considered structurally deficient [1]. Meanwhile, the Canadian Infrastructure Report Card – a condition-based survey of existing infrastructure published in 2016 – reported that although most bridges are currently in acceptable condition, the present levels of investment for maintenance and rehabilitation are insufficient and will likely lead to a rapid decline in the condition of existing bridges in the near future [2]. These findings have emphasized the importance of structural health monitoring (SHM) methods which play a key role in decision-making and can be used to prioritize the needs of major infrastructural systems. A numerical model of the structure is at the core of nearly all existing SHM strategies [3]. These numerical models are used to evaluate the structural performance under specific conditions such as heavy loading [4], earthquake motion [5], wind loading [6], or human activity [7]. Therefore, developing accurate models of existing structures is critical in order to evaluate vulnerability [8], detect damage [9], study retrofit alternatives [10], and to predict the remaining useful life [11].

Traditional structural system identification (SSID) techniques can be used to characterize numerical models of existing structures based on measurements of the dynamic

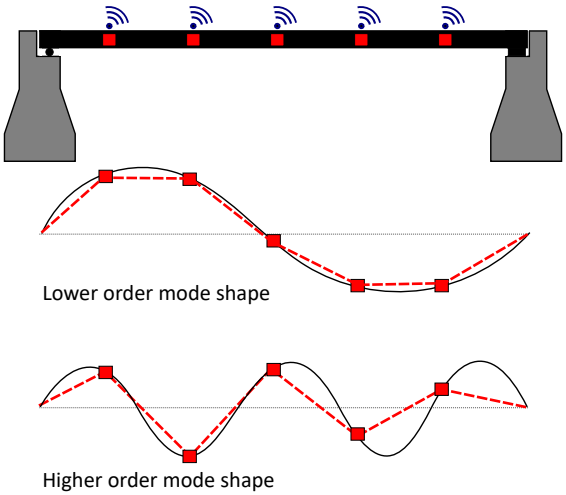
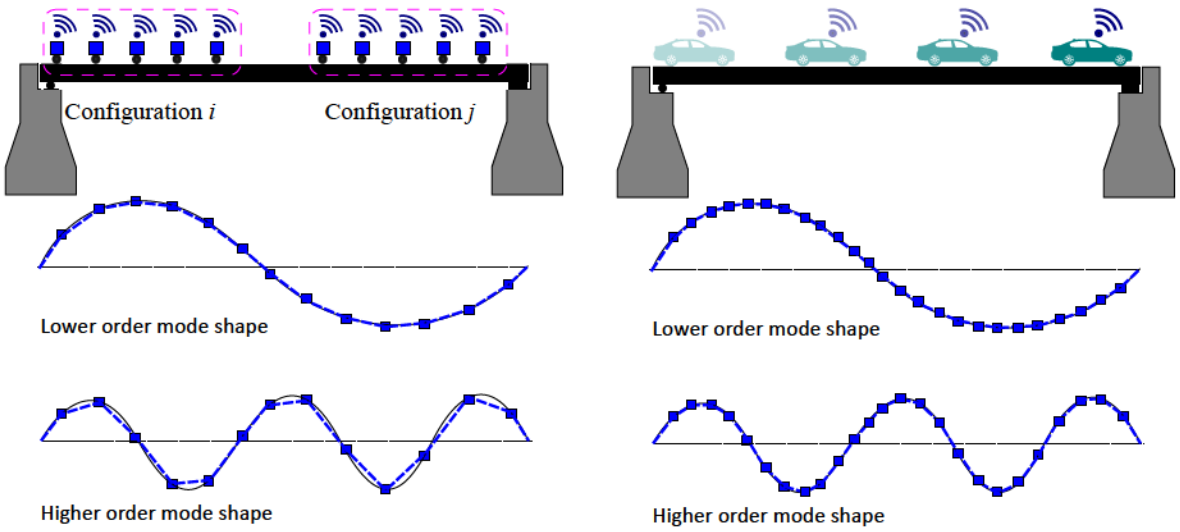


Figure 1.1: Impact of spatial resolution on estimating mode shapes using static sensors.

responses (i.e., accelerations). Most techniques assume the structural system behaves linearly and estimate modal features such as natural frequencies, damping ratios, and mode shapes. The estimated modal features can then be used to tune the numerical model via model updating procedures [12]. Conventional SSID techniques rely on data collected from a network of strategically placed static sensors that remain fixed at specific locations on the structure over an extended period of time. Although estimating the natural frequencies and damping ratios requires measurement data from only a few sensors, mode shapes, being spatial features, require a large number of sensors to achieve sufficient spatial resolution. Figure 1.1 illustrates the impact of low spatial resolution on estimating mode shapes. As shown, the use of a limited number of static sensors may lead to low quality estimates of higher-order mode shapes which, in turn, could negatively impact the results of the model updating procedure.

The most direct method to circumvent this problem is to use spatially dense networks of static sensors. However, the increased costs and setup time of dense networks coupled with large cabling requirements limit the practicality of this method to small or relatively simple structures. Although the use of wireless sensor networks [13–15] eliminates many of the cabling requirements, dense deployments of wireless sensors still remain a relatively expensive proposition.

Recently, the use of mobile sensors for modal identification has demonstrated a number of advantages over static sensor networks [16–21]. In addition to reducing the number of sensors needed, the advantages also include ease of implementation and increased resolution of mode shape estimates. Mobile sensors are defined as sensors that are able to reposition themselves at different locations on a structure. As such, there are two strategies of acquiring vibration data using mobile sensors: *re-configurable sensing* which involves repositioning the mobile sensor(s) in a sequence of static configurations and collecting data while the sensors are not in motion (see Figure 1.2a), and *in-motion sensing* which involves continuously collecting data while the sensors traverse the structure (see Figure 1.2b). In both methods, high spatial resolution can be achieved using a relatively small number of sensors.



(a) *Re-configurable* mobile sensing (b) *In-motion* mobile sensing

Figure 1.2: Use of mobile sensors result in higher resolution of mode shapes.

A key advantage in employing the re-configurable mobile sensing strategy is the ability to apply conventional SSID algorithms (there has been an extensive array of algorithms developed in this area) to the individual datasets obtained from each static configuration. Combining the local results from each setup yields high resolution mode shapes for the overall structure. Furthermore, the use of robotic systems for repositioning the sensors has

automated the data collection process and enabled data collection from several, possibly inaccessible, locations. One such example is the flexure-based mobile sensing nodes (FMSNs) developed by Zhu et al. [16]. This approach utilized magnetic wall-climbing robots capable of repositioning to several locations on the structure and pausing to collect data at those locations. Despite the use of robotic systems, the time needed to acquire data from multiple configurations is a drawback of re-configurable mobile sensing. Furthermore, reference sensors are needed in order to combine the local mode shape estimates from each configuration into high resolution global mode shapes, and thus the configurations must overlap at selected reference locations. Mobile robotic platforms that contain both sensors and an actuator such as the deployable autonomous control system (DACS) developed by Goorts et al. [22], have great potential for re-configurable mobile sensing applications. These devices offer the ability to (locally) excite the structure using the control actuator while simultaneously measuring the structural response thereby yielding efficient data collection from several locations on the structure. Given the knowledge of the applied excitation, mass-normalized mode shapes can be extracted which is useful for controller design. Since the actuator and sensor pair are collocated at the device position, reference sensors are needed to apply conventional SSID algorithms. However, the fact that a single device is sufficient for vibration control applications (as demonstrated in Goorts et al. [22]) motivates the need for a modal identification methodology using a single mobile actuator-sensor pair.

As regards in-motion mobile sensing, its motivation stems from the fact that a sensor mounted on a moving device (e.g., mobile robot, car) can be used as a receiver of rich structural information. A potential future application of in-motion mobile sensing could be to harness crowd-sensed data (i.e. data collected from smartphones or sensors in passenger cars while they cross a bridge) which can yield large amounts of information in addition to steep reductions in cost and time requirements. However, the drawback of in-motion mobile sensing is that the resulting datasets are incompatible with traditional SSID algorithms. In this context, the research on modal identification using multiple (or a network of) in-motion mobile sensors has only recently garnered attention. The newly developed schemes for system identification using in-motion mobile sensors [18–21] mostly yield point estimates of the identified modal parameters and do not incorporate uncer-

tainty quantification. Datasets from in-motion mobile sensors typically have more sources of noise compared to static sensors; the noise derives from factors such as measurement noise induced by sensor-motion dynamics, road roughness, and instrument imperfections, which leads to increased uncertainty in the identified parameters. Quantification of uncertainty in the identified parameters provides a degree of confidence in the resulting estimates which is crucial for model updating and subsequent decision-making. This calls for the development of algorithms for in-motion mobile sensing that allow simultaneous parameter identification and uncertainty quantification.

1.1 Research objectives and scope

The overarching goal of this thesis is to develop a principled framework for modal parameter identification and uncertainty quantification using mobile sensors. Along these lines, the proposed research aims are two-fold:

1. To develop numerical algorithms to estimate structural modal parameters using reconfigurable and in-motion mobile sensor measurements.

2. To incorporate uncertainty quantification into the framework of the modal parameter identification.

The scope of this thesis is limited to developments of comprehensive modal identification algorithms – that provide estimates of modal frequencies, damping ratios and mode shapes – validated with bench-scale tests. More specific objectives in the pursuit of the overarching goals are provided in Section 2.4 of the next chapter following a background and literature survey.

1.2 Organization of thesis

The thesis contains 7 chapters and is organized as follows:

- **Chapter 1** provides a brief introduction and motivation for the framework of modal identification using mobile sensors and presents the overarching research goal.
- **Chapter 2** provides background on existing modal identification methodologies using mobile sensors and reviews Bayesian computational methodologies for parameter estimation and uncertainty quantification. The research gap areas are identified and specific research objectives are outlined.
- **Chapter 3** presents a feasibility study for extracting high resolution mode shapes using input-output data from a single *re-configurable* actuator-sensor pair.
- **Chapter 4** presents Bayesian inference algorithms for output-only modal identification and uncertainty quantification using a network of *in-motion* mobile sensors following an incomplete (or missing) data approach.
- **Chapter 5** presents Bayesian inference algorithms for output-only modal identification and uncertainty quantification using a network of *in-motion* mobile sensors following a complete (or stacked) data approach.
- **Chapter 6** presents a probabilistic framework to obtain contact-point responses from carrier-vehicle responses using an input reconstruction approach.
- Finally, a number of conclusions resulting from the thesis work are discussed in **Chapter 7**, followed by several recommendations for future study.

Chapter 2

Background

In line with the research objectives presented in the previous chapter, this chapter provides background on two key aspects: (a) existing system identification methods for modal parameter estimation using mobile sensor data and (b) Bayesian data analysis methods for incorporating uncertainty quantification into the system identification framework. Accordingly, this chapter starts with the description of system models that are common in time-domain identification of structural systems. The problem of modal parameter identification is then reviewed in the light of emerging mobile sensing strategies, where the research developments on modal identification using *re-configurable* mobile sensors and *in-motion* mobile sensors are presented. Next, a section is devoted to introduce a class of Bayesian parameter inference methods using state space models, including the variational Bayesian methods and the Markov Chain Monte Carlo sampling methods. Finally, key research gaps in the literature are identified and specific research goals are outlined.

2.1 Vibration-based modal identification

Vibration-based modal identification involves the process of estimating the modal features (i.e., the natural frequencies, damping ratios, mode shapes, etc.) using measured structural vibration responses (most commonly accelerations). Over the past decades a number of

deterministic and stochastic techniques for modal parameter estimation have been formulated in both the time and frequency domains [23–25]. Such extracted modal parameters have been used to detect, localize or quantify potential damage in the structure, or to update finite element models. A non-exhaustive set of examples of such applications in the civil and mechanical engineering fields includes finite element model calibration and updating [3, 12], structural health monitoring [26, 27], non-destructive damage assessment [28, 29], and vibration mitigation and control [30, 31].

2.1.1 System modelling

This subsection provides an overview of the time-domain numerical models commonly used to represent dynamic structural systems.

Physical model

The physical model of a structure is derived from the physical laws governing its vibration. These models are commonly developed using finite element (FE) software yielding a physics-based FE model. In general, the structure is assumed to be linear time-invariant (LTI), and the resulting equation of motion for a discretized n-degrees of freedom (DoF) structure is represented by the following second order differential equation

$$\mathbf{M}\ddot{\boldsymbol{d}}(t) + \mathbf{C}\dot{\boldsymbol{d}}(t) + \mathbf{K}\boldsymbol{d}(t) = \mathbf{L}\boldsymbol{u}(t) \tag{2.1}$$

where $\boldsymbol{d}(t) \in \mathbb{R}^n$ is the vector of displacements corresponding to the DoFs and $\mathbf{M}$, $\mathbf{C}$ and $\mathbf{K} \in \mathbb{R}^{n\times n}$ represent the mass, damping, and stiffness matrices of the structural system, respectively. The initial displacement and velocity vectors are given by $\boldsymbol{d}(0) = \boldsymbol{d}_0$ and $\dot{\boldsymbol{d}}(0) = \dot{\boldsymbol{d}}_0$, $\boldsymbol{u}(t) \in \mathbb{R}^{n_u}$ denotes the vector of external forces, and $\mathbf{L} \in \mathbb{R}^{n\times n_u}$ is the input location matrix representing the spatial influence of external input force(s).

Modal model

By assuming a classically damped system, Eq.(2.1) can be decoupled by introducing the following modal transformation

$$\boldsymbol{d}(t) = \boldsymbol{\Phi}\boldsymbol{\eta}(t) \tag{2.2}$$

where

$$\begin{aligned} \boldsymbol{\eta}(t) &= \begin{bmatrix} \eta_1(t) & \eta_2(t) & \dots & \eta_n(t) \end{bmatrix}^T \in \mathbb{R}^n \\ \boldsymbol{\Phi} &= \begin{bmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \dots & \boldsymbol{\phi}_n \end{bmatrix}^T \in \mathbb{R}^{n\times n} \end{aligned} \tag{2.3}$$

are the modal displacement vector and the mass-normalized mode shape matrix respectively. The modal transformation yields r decoupled modal equations

$$\ddot{\eta}_r + 2\xi_r\omega_r\dot{\eta}_r + \omega_r^2\eta_r = \boldsymbol{p}_r^T\,\boldsymbol{u} \tag{2.4}$$

for $r = 1, 2, \dots, n$, where

$$\begin{aligned} \omega_r &= \sqrt{\boldsymbol{\phi}_r^T\left(\mathbf{M}^{-1}\mathbf{K}\right)\boldsymbol{\phi}_r} \\ \xi_r &= \sqrt{\boldsymbol{\phi}_r^T\left(\mathbf{M}^{-1}\mathbf{C}\right)\boldsymbol{\phi}_r}/\,2\omega_r \\ \boldsymbol{p}_r &= \boldsymbol{\phi}_r^T\left(\mathbf{M}^{-1}\mathbf{L}\right) \end{aligned} \tag{2.5}$$

are the rth circular natural frequency, damping ratio, and force participation factor, respectively. The circular natural frequency ω_r and mode shape vectors $\boldsymbol{\phi}_r$ satisfy the following eigenvalue problem:

$$\mathbf{K}\boldsymbol{\phi}_r = \omega_r^2\mathbf{M}\boldsymbol{\phi}_r. \tag{2.6}$$

State space model

The state space model (SSM) is a widely used representation of dynamic systems. The SSM, originating in modern control theory, is a set of first order differential equations representing the dynamics of a system. A key advantage of the SSM is its ability to include the characteristics of the underlying dynamic system while offering fast computation and

suitability in optimization-related problems. This feature has motivated several time domain SSID methods to adopt the SSM for parameter identification including eigensystem realization algorithm (ERA) [25] and numerical algorithms for subspace state space system identification (N4SID) [32] and stochastic subspace identification (SSI) [32].

In developing a SSM, a state vector is first defined as

$$\boldsymbol{x}(t) = \begin{bmatrix} \boldsymbol{d}(t) \\ \dot{\boldsymbol{d}}(t) \end{bmatrix} \tag{2.7}$$

where $\boldsymbol{x}(t) \in \mathbb{R}^{n_s}$, n_s is the state dimension and $n_s = 2n$. The set of second-order differential equations in Eq.(2.1) (or equivalently Eq.(2.4)) are then converted into the following set of first-order differential equations known as the continuous-time *process equation*

$$\dot{\boldsymbol{x}}(t) = \mathbf{A}_c \boldsymbol{x}(t) + \mathbf{B}_c \boldsymbol{u}(t) \tag{2.8}$$

where the continuous-time system matrices $\mathbf{A}_c \in \mathbb{R}^{n_s \times n_s}$ and $\mathbf{B}_c \in \mathbb{R}^{n_s \times n_u}$ are defined as

$$\mathbf{A}_c = \begin{bmatrix} \mathbf{0}_{n\times n} & \mathbf{I}_{n\times n} \\ -\mathbf{M}^{-1}\mathbf{K} & -\mathbf{M}^{-1}\mathbf{C} \end{bmatrix}, \quad \mathbf{B}_c = \begin{bmatrix} \mathbf{0}_{n\times n_u} \\ \mathbf{M}^{-1}\mathbf{L} \end{bmatrix} \tag{2.9}$$

and $\mathbf{I}$ denotes the identity matrix.

If a combination of displacements, velocities, and accelerations are measured, the output vector $\boldsymbol{y}(t)$, containing n_o measured quantities, assumes the following form

$$\boldsymbol{y}(t) = \begin{bmatrix} \mathbf{S}_{dis} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{vel} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_{acc} \end{bmatrix} \begin{bmatrix} \boldsymbol{d}(t) \\ \dot{\boldsymbol{d}}(t) \\ \ddot{\boldsymbol{d}}(t) \end{bmatrix}, \tag{2.10}$$

where, $\mathbf{S}_{dis}$, $\mathbf{S}_{vel}$, and $\mathbf{S}_{acc}$ are the selection matrices for displacements, velocities and accelerations, respectively. The *observation equation* of the SSM is given by

$$\boldsymbol{y}(t) = \mathbf{G}\boldsymbol{x}(t) + \mathbf{J}\boldsymbol{u}(t), \tag{2.11}$$

where the output influence matrix $\mathbf{G} \in \mathbb{R}^{n_o \times n_s}$ and direct transmission matrix $\mathbf{J} \in \mathbb{R}^{n_o \times n_u}$ are defined as

$$\mathbf{G} = \begin{bmatrix} \mathbf{S}_{dis} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{vel} \\ -\mathbf{S}_{acc}\mathbf{M}^{-1}\mathbf{K} & -\mathbf{S}_{acc}\mathbf{M}^{-1}\mathbf{C} \end{bmatrix}, \quad \mathbf{J} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{S}_{acc}\mathbf{M}^{-1}\mathbf{L} & \mathbf{0} \end{bmatrix}. \tag{2.12}$$

When combined, equations (2.8) and (2.11) represent a continuous-time SSM:

$$\begin{aligned} \dot{\boldsymbol{x}}(t) &= \mathbf{A}_c \boldsymbol{x}(t) + \mathbf{B}_c \boldsymbol{u}(t) \\ \boldsymbol{y}(t) &= \mathbf{G}\boldsymbol{x}(t) + \mathbf{J}\boldsymbol{u}(t). \end{aligned} \tag{2.13}$$

In practice, continuous time outputs $\boldsymbol{y}(t)$ are not observed but rather obtained via sampling of the system response at discrete time instances. As such, for numerical implementation the continuous-time SSM can be converted to the discrete-time form following the zero-order hold assumption

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \mathbf{A}\boldsymbol{x}_k + \mathbf{B}\boldsymbol{u}_k, \\ \boldsymbol{y}_k &= \mathbf{G}\boldsymbol{x}_k + \mathbf{J}\boldsymbol{u}_k \end{aligned} \tag{2.14}$$

for $k = 1, \ldots, N$, where $\mathbf{A} = \exp(\mathbf{A}_c \Delta t)$ is the transition matrix, $\mathbf{B} = [\mathbf{A} - \mathbf{I}]\mathbf{A}_c^{-1}\mathbf{B}_c$ is the input-influence matrix and Δt is the sampling time interval. Eq.(2.14) depicts a *deterministic input-output* SSM.

A *stochastic input-output* SSM, on the other hand, is represented by

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \mathbf{A}\boldsymbol{x}_k + \mathbf{B}\boldsymbol{u}_k + \boldsymbol{w}_k, \\ \boldsymbol{y}_k &= \mathbf{G}\boldsymbol{x}_k + \mathbf{J}\boldsymbol{u}_k + \boldsymbol{v}_k \end{aligned} \tag{2.15}$$

where $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ are process noise and measurement noise terms that account for uncertainties in the SSM (such as unmodelled dynamics or measurement errors) and are modelled as temporally independent zero-mean Gaussian white-noise processes, with the following covariance structure

$$\left\langle \begin{bmatrix} \boldsymbol{w}_k \\ \boldsymbol{v}_k \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_k^T & \boldsymbol{v}_k^T \end{bmatrix} \right\rangle = \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix}. \tag{2.16}$$

where $\langle \cdot \rangle$ denotes the expectation operation. For identification of civil infrastructure, applying a sufficiently large controlled excitation is often impractical. Thus, a common practice is to rely on ambient excitation which despite not being measurable, can be represented by realizations of a stochastic process. In this case, the following *output-only stochastic* SSM is used for identification purposes:

$$\boldsymbol{x}_{k+1} = \mathbf{A}\boldsymbol{x}_k + \boldsymbol{w}_k \tag{2.17a}$$

$$\boldsymbol{y}_k = \mathbf{G}\boldsymbol{x}_k + \boldsymbol{v}_k. \tag{2.17b}$$

The modal parameters of the underlying structure (i.e., natural frequencies, damping ratios, and mode shapes) are preserved in the SSM and can be extracted from the $\mathbf{A}$ and $\mathbf{G}$ matrices via an eigenvalue transformation. An algorithm for extracting real-valued modal parameters from discrete-time SSM matrices is provided in Appendix B.1.

2.1.2 Mobile sensing paradigm

Data collection for structural system identification (SSID) has traditionally relied on static sensor networks [14, 33] which are comprised of sensors that remain fixed to strategic locations on the structure during the data collection period. The spatial coverage, which is a function of the size and the arrangement of sensors in the network, determines the quality of information that can be extracted by the SSID techniques. Since mode shapes are spatial features, their spatial resolution is directly related to the spatial coverage of the sensor networks as the mode shape ordinates are only identified at sensor locations. Thus, the use of limited static sensors may result in low spatial resolution of the estimated mode shapes. Although this problem may be alleviated through the use of dense static (wired or wireless) sensor networks, the increased cost associated with instrumentation and maintenance is a serious impediment in practical, large-scale implementations. Optimal sensor placement techniques [34, 35] have been explored in this regard in an effort to minimize the number of sensors needed to achieve adequate mode shape information from the data; however, the underlying challenges associated with static sensor networks, including low mode shape resolution, have not yet been solved.

An appealing alternative to static sensor networks is the concept of mobile sensing. Mounting sensors on robots or other vehicles and having them traverse to different positions on the structure enables efficient data collection from a large number of locations. As such, the use of a few mobile sensors has the potential to imitate large networks of static sensors with the implication being that high resolution mode shapes may be obtained from data collected by a few mobile sensors. The various attempts in the literature to implement mobile sensing can be broadly categorized as either *re-configurable* or *in-motion* mobile

sensing. These two approaches are distinguished by the method in which data is collected; *re-configurable* mobile sensing collects data only when the sensors are stationary while *in-motion* mobile sensing acquires a continuous stream of measurements from the moving sensors. Both approaches lead to incomplete datasets which pose challenges for estimating structural modal parameters. The two methods, along with the challenges posed by the datasets are discussed next.

Re-configurable **mobile sensing**

In re-configurable mobile sensing, vibration data is collected at different locations using multiple non-simultaneous configurations; the data collection happens when the sensors are not in motion. An example of re-configurable mobile sensing is the flexure based mobile sensing nodes developed by Zhu et al. [16]. These nodes, each consisting of a pair of two-wheeled cars and an accelerometer, function by temporarily mounting the accelerometer directly on the structural surface for data collection and feature automated repositioning to achieve data collection at several locations on the structure in a sequential manner. In a subsequent study [17], the authors deployed four such nodes on a steel truss pedestrian bridge for modal parameter identification. A total of five sequential sensing configurations were used where the nodes paused to collect ambient acceleration data. With the aid of two static reference sensors, three mode shapes of the bridge were identified using ERA-NExT [36] with a higher resolution compared to a fixed sensor network of same size (i.e., 4 sensors).

The nature of the data collected during each configuration is similar to that of static sensor networks; however, unlike static sensor networks which provide coverage of the entire structure, each configuration only covers a portion of the structure resulting in a sequential collection of datasets from different portions of the structure. This property is illustrated in Figure 2.1. Each dataset (coming from each configuration) can be individually processed by conventional SSID techniques to obtain partial (i.e., spatially incomplete) mode shapes. Subsequently, with the aid of some reference sensors, the individual partial mode shapes can be merged into global mode shapes. Since the partial mode shapes are dense with respect to the portion of the structure they measured, fusing several partial mode shapes

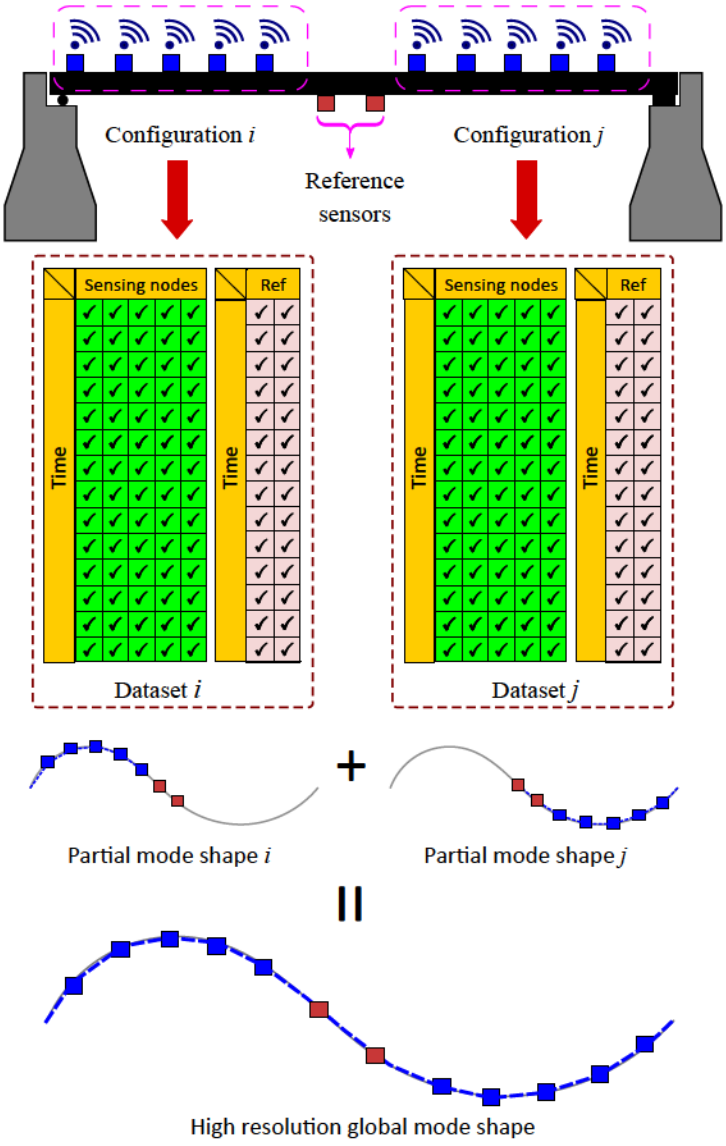


Figure 2.1: Identification with *re-configurable* mobile sensors using two configurations.

together yields high resolution global mode shapes. This is the key concept underlying the identification of high resolution mode shapes with re-configurable mobile sensors.

The potential for mobile platforms to carry both sensors and actuators presents a novel application of *re-configurable* mobile sensing. The DACS, recently developed by Goorts et al. [22] is capable of repositioning at different locations on a structure and applying desired control actions. The platform consists of an unmanned ground vehicle equipped with an active mass damper and accelerometers. Although the platform was initially intended for vibration control applications, this type of system can also be used as a collocated actuator-sensor pair for modal identification purposes. A key advantage of such a system is that mass-normalized mode shapes could be estimated using knowledge of the input excitation. However, the feasibility of using a single collocated actuator-sensor pair for modal identification has yet to be successfully demonstrated in the literature.

In-motion mobile sensing

In-motion mobile sensing differs from re-configurable mobile sensing in that the vibration data is collected while the sensors are in motion. An advantage of in-motion mobile sensing is that it improves spatial coverage in a reduced amount of time as the sensors need not pause at prescribed locations to collect data. However, the datasets obtained from in-motion mobile sensing are always characterized by spatial discontinuities, meaning that each entry in a sensor channel can be associated with a different spatial location compared to its neighbouring entries. This feature prevents a direct application of conventional modal identification algorithms on mobile sensor data.

Attempts to extract modal parameters from measurements collected using in-motion mobile sensors began almost a decade ago and the ensuing research since then has been quite diverse. A growing body of research exists that is focused on using the dynamic response of an instrumented vehicle – such as a truck fitted with a sensor on its axle – crossing a bridge for extracting the bridge modal parameters (see reviews [37, 38]). The concept underlying this methodology has its roots in the theory of vehicle-bridge interaction (VBI); as the vehicle crosses the bridge, both the vehicle and bridge vibrate and the vehicle response is influenced by the bridge vibration. By analyzing the vehicle response,

one aims to identify the modal characteristics of the bridge structure. The idea of extracting the fundamental bridge frequency from the dynamic response of a passing vehicle was first proposed by Yang et al. [39] and verified experimentally by Lin and Yang [40] using an accelerometer-instrumented cart towed by a light truck over a bridge. The fundamental frequency of a bridge was extracted from a fast Fourier transform of the response measurements recorded by the instrumented accelerometer. Later, Yang and Chang [41] adopted an empirical mode decomposition (EMD) technique for preprocessing the vehicle response measurements in order to enhance the visibility of higher bridge frequencies which typically get suppressed in the vehicle responses. More recently, Yang and Chen [42] proposed a modified stochastic subspace identification to extract the bridge modal frequencies from the response of a moving instrumented vehicle.

The identification of bridge damping ratios and mode shapes using VBI has been attempted by few researchers. Identification of bridge damping ratios was proposed by González et al. [43] where the authors developed a six-step iterative algorithm using acceleration responses at two axles of a half-car model. The effectiveness of the algorithm under the influence of measurement noise, road roughness and modelling errors was studied and it was shown that bridge damping could be identified with reasonable accuracy. Compared to bridge frequencies and damping ratios, VBI-based estimation of mode shapes has been looked into very recently. For a simply-supported bridge excited by the crossing of a single vehicle, Yang et al. [44] proposed extracting proxy mode shapes of the bridge using Hilbert transform. The measured vehicle response was separated into multiple monotone component responses using band-pass filtering. Then using a Hilbert transform on the component responses the instantaneous amplitude histories of the component responses were extracted; the envelope of these instantaneous amplitude histories resembled the mode shapes of the simply supported bridge. Malekjafarian and O'Brien [45] proposed to split up the vehicle responses into multiple short segments corresponding to artificial segmented spans of a bridge; short-time frequency domain decomposition applied to these segmented vehicle responses yielded mode shape ordinates at the mid-point of the segmented spans. It is to be mentioned that although the aforementioned techniques based on VBI theory are developing, a comprehensive modal parameter identification technique where all three modal parameters (i.e., modal frequencies, damping ratios and mode shapes) are identified

simultaneously, is still lacking. There is no unifying mathematical framework underpinning the aforementioned VBI-based approaches for identifying the modal parameters, which is a deterrent to the development of a unified modal identification framework in the context of in-motion mobile sensors.

Marulanda et al. [19] developed a high resolution mode shape identification procedure using just two sensors (one static and one mobile). The procedure first identified the natural frequencies and damping ratios, and then using them spatially dense mode shapes were extracted from a space-frequency representation of the responses constructed with short-time Fourier transforms. Using a numerical example, assuming noise-free data and known natural frequencies, the authors identified first three mode shapes at 479 points with great accuracy. The authors conducted a laboratory scale experiment on a simply supported beam where a moving sensor car instrumented with a wireless iMote2 sensor was used for in-motion mobile sensing. The beam was excited by white noise using a dynamic shaker and manually with a rubber hammer. Three mode shapes, with 21 ordinates each, were successfully identified.

Matarazzo and Pakzad [20] posed the problem of modal identification using mobile sensors as a missing data problem. By assuming the mobile sensing nodes (i.e., the locations on the structure where the responses were recorded) coincided with the DoFs of the underlying numerical model, the authors showed that the generation of data from in-motion mobile sensing can be framed as an equivalent static sensor data matrix with unobserved (or missing) entries. The missing entries include all locations that do not coincide with the mobile sensor paths, as shown in Figure 2.2, and the mobile sensor data was asserted to be a specific subset of the full, dense static sensor dataset, with the subset corresponding to the movement of the mobile sensors. The problem of modal parameter identification using the missing data matrix was tackled [20] in a SSM-based maximum likelihood (ML) framework and a modified expectation maximization (EM) algorithm was proposed to obtain the SSM parameters. The modal parameters of the structure were subsequently extracted from the estimated SSM parameter matrices. In an application of modal identification of the Golden Gate bridge using ambient vibration data, the authors [20] simulated a mobile sensor network with 10 moving sensors which resulted in 82% missing entries in the data matrix. Nevertheless, 19 vibration modes were identified with an accuracy comparable to

that from a dense static sensor network. However, the uncertainty in identification results, particularly due to the large amount of missing entries in data matrix, was not addressed.

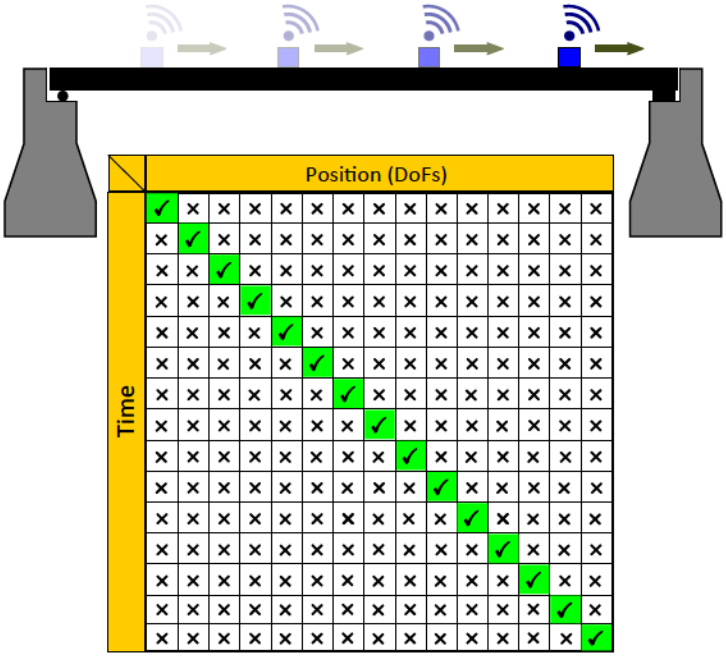


Figure 2.2: Mobile sensor data matrix with missing observations obtained using a single in-motion mobile sensor.

In another study, Matarazzo and Pakzad [46] proposed the idea of converting mobile sensor data collected at different locations to an equivalent static sensor data at some chosen locations on the structure. To facilitate this, they introduced the truncated physical model (TPM) – a linear time-varying state space model – that incorporates a *sinc* function-based spatial interpolator to enable mapping information recorded at a large number of spatial locations into a relatively small number of user-chosen static locations on the structure. Unlike the missing data matrix, the mobile sensor observations could then be stored in a smaller data matrix – termed as the stacked data matrix[1] where the matrix entries are stamped with time and location of record, as shown in Figure 2.3. The user-chosen

[1]Originally termed as the dynamic sensor network (DSN) data matrix [46]

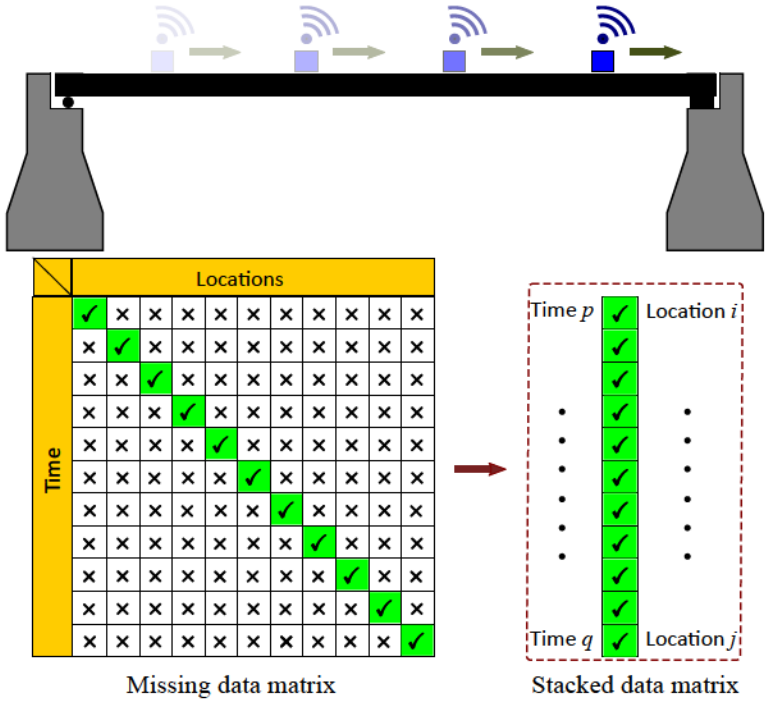


Figure 2.3: Stacked mobile sensor data matrix with compact storage of data from a single in-motion mobile sensor (compared to missing data matrix); each entry in the stacked data matrix is tagged with location and time of record.

static locations are termed as virtual static locations (VSLs)[2]. The mode shape ordinates are identified only at the VSLs and the number of the VSLs determine the size of the TPM. Modal parameter estimation using the TPM was conducted using STRIDEX [21] – a modified EM algorithm compatible with the TPM. The STRIDEX formulation however is not robust[3] and the authors [21] banked on a minimum model size criterion to prevent a certain numerical invertibility issue. This criterion requires the number of VSLs to be equal to the number of mobile sensors, and as such the flexibility of the formulation becomes limited. Nevertheless, to identify mode shape ordinates at a large number of VSLs using a few mobile sensors, the authors proposed to construct multiple TPMs with

[2]Originally termed as virtual probing locations (VPLs) [46]

[3]During this thesis work, an error was uncovered in the derivation of STRIDEX update equations [21]

smaller subsets of VSLs and then run STRIDEX algorithm on each TPM to yield partial mode shapes. These partial mode shapes were then assembled into high resolution global mode shapes. The authors demonstrated this procedure experimentally [47] on a 3.66 m long simply supported steel beam scanned by two toy cars equipped with Imote2 wireless accelerometers. A total of 144 subsets of VSLs were analyzed (144 STRIDEX runs) which corresponded to 288 modal ordinates. The first mode shape of the beam was extracted at 248 points on the beam, illustrating the dense spatial information that can be obtained using in-motion mobile sensors.

In the quest of a unified framework for modal identification using in-motion mobile sensors, an assumption that the moving sensors are able to directly measure the bridge responses of a structure is commonly employed. However, in practice the sensors mounted on carrier-vehicles (e.g., robots, cars) measure the vehicular responses, which are the result of the convolution of the 'true' contact-point bridge responses – the response at the contact point of the vehicle with the bridge – with the carrier-vehicle system. Yang and Chen [42] reported that the the bridge modal features can be masked in the vehicle response due to dominant vehicle dynamics and this could prevent successful identification of bridge modal parameters. Yang et al. [48] proposed to use contact-point response to get better modal identification results. Considering a simply supported beam and ignoring vehicle and bridge damping, an approximate closed-from solution was derived to obtain contact-point response from the measured vehicle response. The proposed method was assessed using numerical simulations and it was shown that the contact-point responses outperformed the vehicle response in extracting the frequencies and mode shapes of a bridge. However, the approach to obtain contact-point responses is approximate in nature and a more generalized procedure to obtain contact-point responses is lacking.

2.2 Bayesian inference using state space models

A certain level of uncertainty is inherent in all methods of modal identification however the use of mobile sensor data creates additional sources of uncertainty that may impact the accuracy of the modal parameter estimates. In general, modal identification using mobile sensor data is subjected to uncertainty from any combination of the following sources:

- Uncertainty due to unmeasured external loads (e.g., wind, traffic, environmental effects) in case of output-only mobile sensor measurements,
- Measurement error in the sensory system (i.e., instrument noise, transmission error, missing data),
- Large measurement error induced by the mobile platform moving over a rough road surface,
- Modelling error due to unmodelled dynamics of sensor-structure interaction and assumption of linear dynamics for the underlying structural system,
- Statistical uncertainty arises due to limited data and reduces with more data.

Uncertainty is typically classified as aleatory (or Type I) and epistemic (or Type II). Epistemic uncertainty (or reducible uncertainty) stems from gaps in knowledge whereas aleatory uncertainty (or irreducible uncertainty) is attributed to unknown random variations when an experiment or physical phenomenon is repeated. In a modelling context, the assumptions made during the development of a model, such as the linear or nonlinear behavior, type of boundary condition, material characteristics, etc. result in epistemic uncertainty. Aleatory uncertainty is often modeled as a random variable added to the output of a model, and it is used to express the randomness observed when the same experiment is performed more than once. A detailed exposition on classification of uncertainties that arise in structural modelling for risk analysis is provided in [49].

In a Bayesian perspective, uncertainty is expressed via probability distributions which characterize the state of knowledge. Random variables express our knowledge about the

model parameters and hence in this perspective, uncertainty is only epistemic and can always be reduced by improving one's knowledge. Quantifying the uncertainty associated with estimated modal parameters not only provides a sense of fidelity but is also useful in further applications such as Bayesian model updating, risk assessment and decision making [3, 50, 51]. Most of the existing Bayesian formulations of modal parameter identification using static sensors are confined to the frequency-domain [52–56], where a Fourier transform is first implemented and then modal analysis is performed. Time-domain Bayesian modal parameter identification with static sensor data on the other hand has been attempted only very recently [57, 58].

Bayesian inference [59, 60] is a widely used estimation framework that directly provides the uncertainties in parameter estimation. The underlying principles of the Bayesian framework are based on probability distributions and the rules of probability mathematics. It is therefore essential that all mathematical models used within the Bayesian framework are formulated in terms of probability distributions; this allows the computation rules of probability theory to be valid for Bayesian inference of such models. In Bayesian statistics, the outcome to any inference over a set of random variables is always the posterior probability distribution, defined as the joint probability distribution of the random variables given the measurements. Thus, instead of resulting in a single value, the outcome is the probability density function of the quantities of interest given the data.

The class of mathematical models that are of concern in this thesis are the stochastic SSMs (e.g. Eq.(2.17)). The two key characteristics of stochastic SSMs are: the decoupled nature of the system dynamics and measurements (i.e., the process and measurement equations) and the inclusion of process noise and measurement noise – which induces a probability distribution over the states and the measurements. The sequence of system states $\boldsymbol{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{N+1}\}$ contains sufficient information about the system, however they are typically *hidden* (or *latent*) and the inference on the states has to be made entirely based on the noisy measurements $\boldsymbol{Y} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\}$. For example, in structural dynamical systems, the states contain the displacements and velocities at the DoFs of the structural model, however in most cases only noisy acceleration measurements are available.

In SSMs representing dynamic systems there are often unknown or uncertain parameters $\boldsymbol{\theta}$ which should be estimated along with the states. The Bayesian approach to addressing

unknown parameters is to model them as random variables with a certain prior distribution $\pi_{pr}(\boldsymbol{\theta})$. A stochastic SSM (as in Eq.(2.17)) with unknown parameters can be written in the following probabilistic notation:

$$\begin{aligned} \boldsymbol{\theta} &\sim \pi_{pr}(\boldsymbol{\theta}) \\ \boldsymbol{x}_1 &\sim p(\boldsymbol{x}_1 \mid \boldsymbol{\theta}) \\ \boldsymbol{x}_{k+1} &\sim p(\boldsymbol{x}_{k+1} \mid \boldsymbol{x}_k, \boldsymbol{\theta}) \\ \boldsymbol{y}_k &\sim p(\boldsymbol{y}_k \mid \boldsymbol{x}_k, \boldsymbol{\theta}). \end{aligned} \tag{2.18}$$

Regarding the notation, $p(\cdot \mid \cdot)$ is used to denote a generic conditional probability density function (PDF) specified by its arguments and the difference between the notations for random variables and their realizations is suppressed. Following Bayes' rule, the joint posterior distribution of the states and the parameters can be written as:

$$p(\boldsymbol{X}, \boldsymbol{\theta} \mid \boldsymbol{Y}) = \frac{p(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{\theta})\, p(\boldsymbol{X} \mid \boldsymbol{\theta})\, \pi_{pr}(\boldsymbol{\theta})}{p(\boldsymbol{Y})}. \tag{2.19}$$

To obtain the parameters $\boldsymbol{\theta}$, the states are integrated out, which yields the marginal posterior distribution of parameters:

$$p(\boldsymbol{\theta} \mid \boldsymbol{Y}) = \int p(\boldsymbol{X}, \boldsymbol{\theta} \mid \boldsymbol{Y})\, d\boldsymbol{X}. \tag{2.20}$$

The above equation involves a high-dimensional integral (i.e., integration over all state variables $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{N+1}\}$) which is analytically intractable. Hence four computational methods for parameter estimation, namely Laplace approximation, Expectation maximization, variational Bayes and Markov chain Monte Carlo methods are discussed next, all of which are based on approximating the marginal posterior distribution (without explicitly forming the joint posterior distribution of the states and parameters).

2.2.1 Maximum *a posteriori* and Laplace approximation

The simplest approximation to the posterior distribution is a point estimate based on the location of the maximum of the posterior distribution. This is known as the maximum a

posteriori (MAP) estimate and defined as:

$$\hat{\boldsymbol{\theta}}_{\text{MAP}} = \arg\max_{\boldsymbol{\theta}} \ \ln\left[p\left(\boldsymbol{Y} \mid \boldsymbol{\theta}\right)\pi_{pr}\left(\boldsymbol{\theta}\right)\right]. \tag{2.21}$$

The disadvantage of the MAP estimate is that it ignores the spread of the posterior distribution and as such lacks uncertainty information. The Laplace approximation, [61] which performs a Taylor series expansion at the MAP estimate, is one way of improving the MAP estimate. The idea here is to approximate the posterior with a normal distribution centered at the MAP estimate,

$$p\left(\boldsymbol{\theta} \mid \boldsymbol{Y}\right) \approxeq \mathcal{N}\left(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}_{\text{MAP}}, \left[\mathcal{H}\left(\hat{\boldsymbol{\theta}}_{\text{MAP}}\right)\right]^{-1}\right) \tag{2.22}$$

where $\mathcal{H}$ is the second order derivative of the negative log joint PDF at the MAP estimate:

$$\mathcal{H}\left(\hat{\boldsymbol{\theta}}_{\text{MAP}}\right) = \left.\frac{\partial^2 \ln\left[p\left(\boldsymbol{Y} \mid \boldsymbol{\theta}\right)\pi_{pr}\left(\boldsymbol{\theta}\right)\right]}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\text{MAP}}}. \tag{2.23}$$

The approximation is based on the asymptotic normality of the posterior under a set of regularity conditions, as the number of observed data points tends to infinity [62]. However, the assumption of a normal distribution may misrepresent the posterior uncertainty for small datasets. Furthermore, the Laplace approximation is not suitable for bounded, constrained, or strictly positive parameters or for multi-modal probability distributions. Finally, the second order derivative may be difficult to compute, particularly when the number of unknown parameters is large.

2.2.2 Expectation Maximization

The expectation-maximization (EM) algorithm [63–65] provides an iterative approach to find a ML or MAP estimate of the parameters in stochastic SSMs. In order to obtain the ML estimate of $\boldsymbol{\theta}$, the EM algorithm maximizes the following lower bound of the optimization problem:

$$\max_{\boldsymbol{\theta}}\left[\ln p\left(\boldsymbol{Y} \mid \boldsymbol{\theta}\right)\right] = \max_{\boldsymbol{\theta}}\left[\ln \int p\left(\boldsymbol{Y}, \boldsymbol{X} \mid \boldsymbol{\theta}\right) d\boldsymbol{X}\right]$$

$$\geq \max_{\boldsymbol{\theta}} \left[\int \ln p\left(\boldsymbol{Y}, \boldsymbol{X} \mid \boldsymbol{\theta}\right) d\boldsymbol{X} \right] \tag{2.24}$$

where $\mathcal{L}\left(\boldsymbol{\theta}\right) \coloneqq \ln p\left(\boldsymbol{Y}, \boldsymbol{X} \mid \boldsymbol{\theta}\right)$ is referred as the *complete-data* log likelihood function and is much easier to optimize than $\log p\left(\boldsymbol{Y} \mid \boldsymbol{\theta}\right)$. Starting from an initial guess $\boldsymbol{\theta}^{(0)}$, the EM algorithm seeks to maximize Eq.(2.24) by iteratively alternating between two distinct subroutines called the E-step and the M-step:

1. Expectation step (E-step): Compute the expectation of the log likelihood function with respect to the current conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{Y}$ and the current estimate of the parameter $\boldsymbol{\theta}^{(j)}$

$$\mathcal{F}\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(j)}\right) = \left\langle \mathcal{L}(\boldsymbol{\theta}) \mid \boldsymbol{Y}, \boldsymbol{\theta}^{(j)} \right\rangle. \tag{2.25}$$

2. Maximization step (M-step): Find the parameter that maximizes the $\mathcal{F}\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(j)}\right)$,

$$\boldsymbol{\theta}^{(j+1)} = \arg\max_{\boldsymbol{\theta}} \ \mathcal{F}\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(j)}\right). \tag{2.26}$$

The iterations are continued until convergence in likelihood is achieved. The EM algorithm is a local maximizer and has the property that the likelihood increases monotonically [66] with each EM iteration. The EM algorithm is particularly useful when the likelihood function belongs to an exponential family of probability distributions; in such a case, the E-step becomes a sum of *expectations of sufficient statistics* [65] and the M-step involves a closed-form update for each step [67]. Application of the EM algorithm for maximum likelihood estimation of modal parameters using static sensor data has been proposed in [68–70].

A drawback of the EM algorithm is that it does not provide an estimate of the covariance matrix corresponding to the parameter estimates. The Supplemented EM algorithm [71] was proposed to address this limitation by providing the covariances of the estimated parameters calculated from observed Fischer information matrix. The other drawback associated with the EM algorithm is the potential for slow convergence, even in seemingly innocuous problems or in problems where there is too much incomplete information [72]. A detailed presentation of the EM algorithm and its variants is provided in [73].

2.2.3 Variational Bayes

The variational Bayes (VB) approach [74, 75] casts the Bayesian inference of unobserved variables (i.e., unknown states and unknown parameters) as a deterministic optimization problem where an approximate posterior distribution is fit to the true posterior distribution. A problem often encountered in Bayesian inference is the need to evaluate the log-marginal likelihood [74, 76] of the observed data $\boldsymbol{Y}$ given by

$$\ln p\left(\boldsymbol{Y}\right) = \ln \int p\left(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\theta}\right) d\boldsymbol{X} d\boldsymbol{\theta}. \tag{2.27}$$

The right hand side of Eq.(2.27) involves a high-dimensional integration and is often analytically burdensome or even intractable. In the VB framework, a lower bound approximation to the log marginal likelihood is sought and iteratively maximized such that the lower bound eventually converges to a value close to the true log marginal likelihood.

The joint posterior distribution $p\left(\boldsymbol{X}, \boldsymbol{\theta} \mid \boldsymbol{Y}\right)$, in Eq.(2.19), is approximated by an arbitrary probability distribution $q\left(\boldsymbol{X}, \boldsymbol{\theta}\right)$, known as the "variational distribution". The distribution $q\left(\boldsymbol{X}, \boldsymbol{\theta}\right)$ is usually selected from a family of distributions more simple than $p\left(\boldsymbol{X}, \boldsymbol{\theta} \mid \boldsymbol{Y}\right)$, with the intention of making $q\left(\boldsymbol{X}, \boldsymbol{\theta}\right)$ similar to the true posterior distribution. The following composition of the log marginal likelihood forms the core of the VB approach:

$$\begin{aligned}
\ln p\left(\boldsymbol{Y}\right) &= \int \int q\left(\boldsymbol{X}, \boldsymbol{\theta}\right) \ln p\left(\boldsymbol{Y}\right) d\boldsymbol{X} d\boldsymbol{\theta} \\
&= \int \int q\left(\boldsymbol{X}, \boldsymbol{\theta}\right) \ln \frac{p\left(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\theta}\right)}{p\left(\boldsymbol{X}, \boldsymbol{\theta} \mid \boldsymbol{Y}\right)} d\boldsymbol{X} d\boldsymbol{\theta} \\
&= \int \int q\left(\boldsymbol{X}, \boldsymbol{\theta}\right) \ln \frac{p\left(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\theta}\right)}{q\left(\boldsymbol{X}, \boldsymbol{\theta}\right)} d\boldsymbol{X} d\boldsymbol{\theta} + \int \int q\left(\boldsymbol{X}, \boldsymbol{\theta}\right) \ln \frac{q\left(\boldsymbol{X}, \boldsymbol{\theta}\right)}{p\left(\boldsymbol{X}, \boldsymbol{\theta} \mid \boldsymbol{Y}\right)} d\boldsymbol{X} d\boldsymbol{\theta} \\
&\triangleq \mathcal{F}_{ve}\left(q\left(\boldsymbol{X}, \boldsymbol{\theta}\right)\right) + \mathrm{KL}\left(q\left(\boldsymbol{X}, \boldsymbol{\theta}\right) || p\left(\boldsymbol{X}, \boldsymbol{\theta} \mid \boldsymbol{Y}\right)\right).
\end{aligned} \tag{2.28}$$

The log marginal likelihood comprises the sum of two information-theoretic quantities: $\mathcal{F}_{ve}\left(q\left(\boldsymbol{X}, \boldsymbol{\theta}\right)\right)$, referred to as the "variational free energy", and $\mathrm{KL}\left(q\left(\boldsymbol{X}, \boldsymbol{\theta}\right) || p\left(\boldsymbol{X}, \boldsymbol{\theta} \mid \boldsymbol{Y}\right)\right)$, referred to as the Kullback-Leibler divergence (KLD) [77] between the true posterior distribution $p\left(\boldsymbol{X}, \boldsymbol{\theta} \mid \boldsymbol{Y}\right)$ and the variational distribution $q\left(\boldsymbol{X}, \boldsymbol{\theta}\right)$. The KLD is always greater

than or equal to zero and can be intuitively understood as the distance between two probability distributions: if the two probability distributions $q(\boldsymbol{X}, \boldsymbol{\theta})$ and $p(\boldsymbol{X}, \boldsymbol{\theta} \mid \boldsymbol{Y})$ are very dissimilar, the $\mathrm{KL}(q(\boldsymbol{X}, \boldsymbol{\theta}) || p(\boldsymbol{X}, \boldsymbol{\theta} \mid \boldsymbol{Y}))$ returns a high positive value. If and only if the two probability distributions $q(\boldsymbol{X}, \boldsymbol{\theta})$ and $p(\boldsymbol{X}, \boldsymbol{\theta} \mid \boldsymbol{Y})$ are identical, it returns zero. Due to the non-negativity of KLD, the variational free energy $\mathcal{F}_{ve}(q(\boldsymbol{X}, \boldsymbol{\theta}))$ is always smaller than the log marginal likelihood and thus forms the lower bound of the log marginal likelihood (i.e. $\mathcal{F}_{ve}(q(\boldsymbol{X}, \boldsymbol{\theta})) \leq \ln p(\boldsymbol{Y})$). This fact is exploited in the numerical application of the VB approach: since the log marginal likelihood is a constant quantity, maximizing the variational free energy is equivalent to minimizing the KLD.

A common choice for the variational distribution $q(\boldsymbol{X}, \boldsymbol{\theta})$ over the unobserved variables $\boldsymbol{X}$ and $\boldsymbol{\theta}$ is a factorization over sets $\boldsymbol{X}$ and $\boldsymbol{\theta}$:

$$q(\boldsymbol{X}, \boldsymbol{\theta}) = q(\boldsymbol{X}) q(\boldsymbol{\theta}). \tag{2.29}$$

However, due to the complexity of the integrals involved, a simultaneous analytical maximization of the variational free energy with respect to both its arguments is often tedious and a "coordinate-wise" approach that first maximizes with respect to $q(\boldsymbol{X})$ and second with respect to $q(\boldsymbol{\theta})$ is preferred [74]. The VB algorithm maximizes the variational free energy $\mathcal{F}_{ve}(q(\boldsymbol{X}, \boldsymbol{\theta}))$ in Eq.(2.28) with respect to the variational distributions $q(\boldsymbol{X})$ and $q(\boldsymbol{\theta})$ by alternately applying the following two steps [78, 79]:

1. VB Expectation (VBE-step): Compute the variational distribution of the latent states

$$\begin{aligned} q^{(j+1)}(\boldsymbol{X}) &\propto \exp\left(\int \ln p(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\theta}) \, q^{(j)}(\boldsymbol{\theta}) \, d\boldsymbol{\theta}\right) \\ &\propto \exp\left(\langle \ln p(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\theta}) \rangle_{q^{(j)}(\boldsymbol{\theta})}\right). \end{aligned} \tag{2.30}$$

2. VB Maximization (VBM-step): Compute the variational distribution of the unknown parameters

$$\begin{aligned} q^{(j+1)}(\boldsymbol{\theta}) &\propto \exp\left(\int \ln p(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\theta}) \, q^{(j+1)}(\boldsymbol{X}) \, d\boldsymbol{X}\right) \\ &\propto \exp\left(\langle \ln p(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\theta}) \rangle_{q^{(j+1)}(\boldsymbol{X})}\right). \end{aligned} \tag{2.31}$$

The above VBE-step and VBM-step are computed iteratively until the variational free energy converges to a stationary value. Application of the VB algorithm for time-domain Bayesian inference of modal parameters using static sensor data has been proposed in [57].

The limitations of this method are: (a) there is only one approximating distribution limiting its effectiveness in representing multi-modal posteriors; however, this can be alleviated by considering a mixture of approximate distributions [80, 81]; (b) similar to EM, the VB is also sensitive to the initial point estimate since VB acts as a local maximizer and can get stuck in a local maximum; and (c) the spread (covariance) of the posterior distribution may be underestimated [82, 83] in the variational family.

2.2.4 Markov chain Monte Carlo

A more commonly used class of algorithms is the Markov chain Monte Carlo (MCMC) [84] algorithms, which are based on sampling from probability distributions via Markov chains. Like variational inference, MCMC algorithms starts with a random draw from an initial distribution of the unobserved variables (i.e., unknown states and unknown parameters). However, rather than optimizing the hyperparameters of this distribution, the MCMC algorithms construct a Markov chain (i.e. a sequence of random variables in which the distribution of each element depends on the value of the previous one) by repeatedly simulating samples in the chain. After simulating for a sufficiently long time, the desired posterior distribution is eventually obtained as the stationary distribution of the Markov chain.

The MCMC algorithms [85] are particularly well-suited for simulating samples from Bayesian posterior distributions $p(\boldsymbol{\theta} \mid \boldsymbol{Y})$ as their implementation only requires knowledge of the unnormalized posterior distribution $p(\boldsymbol{Y} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})$. As such, there is no need to evaluate the intractable marginal likelihood $p(\boldsymbol{Y})$. Landmark developments in MCMC tools for Bayesian computation include the Metropolis-Hastings algorithm [86, 87] and the Gibbs sampler [88]. The use of MCMC algorithms in the context of SSMs has been discussed in [89, 90].

The Metropolis-Hastings (MH) algorithm is the most common type of MCMC algo-

rithm. It uses a proposal density $q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(j)}\right)$ for suggesting the next sample, given the current sample $\boldsymbol{\theta}^{(j)}$. Starting from an initial point $\boldsymbol{\theta}^{(0)}$, the Markov chain in MH algorithm is constructed by iterating the following steps:

1. Sample a candidate point $\boldsymbol{\theta}^*$ from the proposal distribution $\boldsymbol{\theta}^* \sim q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(j)}\right)$,
2. Evaluate acceptance probability

$$\mathcal{A}_j = \min\left\{1\ ,\ \frac{p\left(\boldsymbol{\theta}^* \mid \boldsymbol{Y}\right) q\left(\boldsymbol{\theta}^{(j)} \mid \boldsymbol{\theta}^*\right)}{p\left(\boldsymbol{\theta}^{(j)} \mid \boldsymbol{Y}\right) q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(j)}\right)}\right\}$$

3. Generate a uniform random variable $\mathtt{u} \sim \mathcal{U}(0,1)$ and set

$$\boldsymbol{\theta}^{(j+1)} = \begin{cases} \boldsymbol{\theta}^* & \text{if } \mathtt{u} \leq \mathcal{A}_j \\ \boldsymbol{\theta}^{(j)} & \text{otherwise} \end{cases}$$

The choice of the proposal distribution is crucial for performance of the MH algorithm and it has become good practice to adaptively tune the proposal to achieve an optimal acceptance rate [91].

The Gibbs sampler (GS) is another popular MCMC algorithm, which samples components (or blocks) of random variables one (block) at a time from their conditional distributions given the other random variables [88]. It can be shown that the sequence of samples constitutes a Markov chain and the stationary distribution of that Markov chain is the desired joint distribution [59]. In its basic form, GS is a special case of the Metropolis-Hastings algorithm and is particularly applicable when the joint distribution is not known explicitly or is difficult to sample from directly while the conditional distribution of each variable (or block of variables) is known and is easy to sample from.

Gibbs sampling is well suited for SSMs [92] as they are typically specified as a collection of conditional distributions as depicted in Eq.(2.18). Starting with an initial parameter set $\boldsymbol{\theta}^{(0)}$, the GS performs the following steps to sample the SSM with unknown parameters $\boldsymbol{\theta}$, latent states $\boldsymbol{X}$ and observed variables $\boldsymbol{Y}$:

1. Sample $\boldsymbol{X}^{(j+1)} \sim p\left(\boldsymbol{X} \mid \boldsymbol{\theta}^{(j)}, \boldsymbol{Y}\right)$,

2. Sample $\boldsymbol{\theta}^{(j+1)} \sim p\left(\boldsymbol{\theta} \mid \boldsymbol{X}^{(j+1)}, \boldsymbol{Y}\right)$

Although MCMC algorithms can provide *asymptotically* exact results of posterior distributions in Bayesian inference, there are a number of disadvantages to using them in practice. First, MCMC algorithms are typically much more computationally expensive compared to variational inference algorithms due to the time required for convergence of the Markov chain to the desired posterior distribution. Furthermore, verifying convergence of the Markov chain may require multiple runs of a Markov chain from varying starting points further adding to the computational demands. Second, the initial samples of the Markov chains typically come from some non-stationary distribution; these initial samples are referred to as "burn-in" samples and need to be discarded. Third, the samples generated from MCMC algorithms are correlated, which can increase the number of samples needed to achieve good precision. To obtain independent samples, a subsampling procedure is usually implemented that involves discarding all but every kth sample.

2.3 Research gaps in mobile sensing based modal parameter identification

A summary of the main gap areas in the existing research on modal parameter identification using mobile sensing is described below:

- Although modal parameter identification using a network of re-configurable mobile sensors has been studied in the literature, identification of high resolution mode shapes using minimal instrumentation, specifically using a single reconfigurable mobile actuator-sensor pair, has not been investigated.

- From an incomplete-data perspective, the mobile sensor data matrix generated using in-motion mobile sensors was shown to be equivalent to a dense static sensor data matrix with missing observations [20]. Inference in the presence of large amounts

of missing data can add significant uncertainty to the identified modal parameters. The fidelity of the identified parameters, particularly the mode shapes, is directly related to the degree of missing information, and therefore a framework that allows quantifying the uncertainty induced by missing observations in the identified modal parameters is important and has thus far not been explored in the literature.

- The stacked data matrix [46] presents a complete-data perspective to data from in-motion mobile sensors. However, a direct identification with the stacked data matrix is not straight-forward due to the spatial discontinuities of the data entries. The ML-based STRIDEX [46] algorithm, introduced to enable identification with stacked data matrix, is inflexible due to the constraint of minimum model size and needs to be improved. Furthermore, uncertainty information of the estimated modal parameters was not quantified and needs to be accounted for.

- The direct use of in-motion mobile sensor responses for modal identification may lead to poor estimation results for higher bridge modes. As the sensors are instrumented on mobile carrier-vehicles, the responses recorded are in fact vehicle responses and not true contact-point structural responses. Thus the structural dynamics may get masked in the vehicle responses and this may prevent successful identification of the structural modal parameters. Contact-point responses are shown to be a better choice for modal parameter identification over vehicle responses [48], however a generalized technique to estimate contact-point response from measured vehicle response has not yet been developed.

2.4 Specific Objectives

Based on the identified gap areas, the specific research objectives of this thesis are as follows:

1. To determine the feasibility of re-configurable mobile sensing with a single actuator-sensor pair (i.e., minimal instrumentation) for identification of high resolution mode shapes. This entails

(a) developing a methodology for modal identification using input-output data from a single actuator-sensor pair,

(b) conducting a sensitivity study for performance assessment, and

(c) experimentally verifying the proposed approach.

2. To propose a Bayesian framework for SSM-based output-only modal parameter identification following the missing data approach to in-motion mobile sensing. This entails

   (a) developing a Bayesian modal parameter identification methodology that can handle missing observations,

   (b) quantifying the uncertainty in the identified modal parameter estimates due to factors including extent of missing entries in data matrix, and

   (c) experimentally verifying the proposed approach.

3. To propose a Bayesian framework for SSM-based output-only modal parameter identification following the stacked data approach to in-motion mobile sensing. This entails

   (a) developing an improved modelling and estimation framework that is flexible,

   (b) establishing a Bayesian framework for modal parameter identification and uncertainty quantification, and

   (c) experimentally verifying the proposed approach.

4. To propose a generalized framework for deriving the contact-point response from measured vehicle response. This entails

   (a) establishing a mathematical relation between the vehicle response and the contact-point response, and

   (b) establishing an estimation framework that recovers contact-point response from measured vehicle response.

The dynamics associated with mobile sensor platforms will be neglected in pursuit of the first three objectives and addressed in detail in the fourth objective.

Chapter 3

Modal identification using a single re-configurable actuator-sensor pair

This chapter considers the problem of input-output modal identification using a single *re-configurable* mobile actuator-sensor pair.[1] Although, output-only modal identification using the ambient response of a structure is more convenient, without knowledge of the input excitation, the resulting mode shape estimates are arbitrarily normalized [93]. This poses a challenge for applications such as structural control, frequency response function (FRF) synthesis, or force estimation where mass-normalized mode shapes are required.

A novel input-output mode shape identification method is developed in this study whereby a single mobile actuator and a single mobile sensor are used within the *re-configurable* mobile sensing approach. Two different implementations of the actuator and sensor, namely non-collocated actuator sensor (NCAS) and collocated actuator sensor (CAS) are considered in this study. In the NCAS case, the actuator and sensor move independently and therefore may or may not be collocated whereas in the CAS case the actuator and sensor are considered to be a single system and thus always collocated.

The statement of the problem is formally presented next followed by a detailed description of the solution methodology. The proposed approach and effects of different types of

[1] A portion of the work presented in this chapter was developed in collaboration with Dr. Mukhopadhyay. For details, refer to the Statement of Contributions

identification errors are studied through numerical examples and later verified experimentally using a laboratory-scale beam model.

3.1 Problem Statement

Consider an n-DoF FE model of a classically damped Euler-Bernoulli beam (representing a bridge), with the equation of motion

$$\mathbf{M}\ddot{\boldsymbol{d}} + \mathbf{C}\dot{\boldsymbol{d}} + \mathbf{K}\boldsymbol{d} = \mathbf{L}u \tag{3.1}$$

where the $n \times n$ matrices $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$, are respectively the mass, damping and stiffness matrices of the beam, and $\mathbf{L}$ is the $n \times 1$ input location vector. The beam is excited using a single actuator; when the actuator applies an input u to the jth DoF, $L_j = 1$ and $L_i = 0$, $\forall\, i \neq j$. The mass of the actuator is assumed to be negligibly small compared to the beam, and actuator-structure interaction is ignored. The undamped eigenvalue problem for the system is given by

$$\mathbf{K}\boldsymbol{\Phi} = \mathbf{M}\boldsymbol{\Phi}\boldsymbol{\Lambda} \tag{3.2}$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{n\times n}$ is the mode shape matrix and $\boldsymbol{\Lambda} \in \mathbb{R}^{n\times n}$ is the diagonal matrix of eigenvalues. The objective is to estimate the modal parameters of the above system using a single sensor and single actuator. This will be demonstrated using both the NCAS and CAS approaches, as discussed next.

3.1.1 Non-collocated actuator-sensor (NCAS) case

The NCAS case can be implemented in three different setups derived from the positioning of the actuator and sensor. The three setups are defined as: (a) mobile actuator - static sensor, and (b) static actuator - mobile sensor. For each case, a static actuator or sensor will remain fixed at a specific location on the beam while a mobile actuator or sensor sequentially repositions itself at different locations along the length of the beam. Once the mobile sensor/actuator arrives at a new location, a test is performed whereby the actuator excites the structure and the response is measured by the sensor. This process results in

a collection of data sets each corresponding to a different actuator/sensor position. As will soon become evident in the solution methodology, a test in which the actuator and sensor are collocated must be performed, regardless of the setup, in order achieve unique identification of mass-normalized mode shapes. The solution methodology presented in Section 3.2 will focus specifically on setup (a); however, the methodology and associated equations can be extended to the other setup. Within setup (a), consider an arbitrary test in which the mobile actuator is positioned on the beam at station s_i, and the sensor is fixed at station s_j as shown in Figure 3.1.

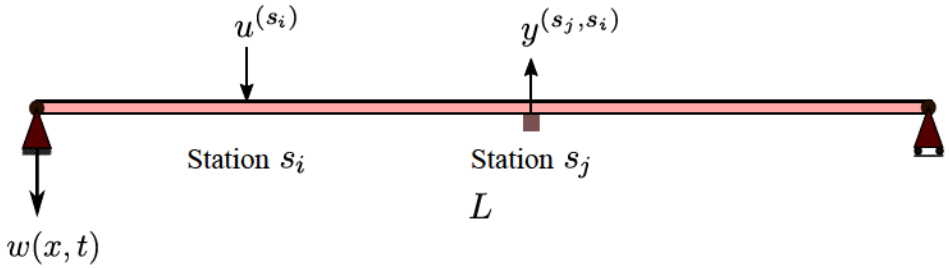


Figure 3.1: Beam with a mobile actuator and a static sensor.

This input-output pair is indexed as $\left(u^{(s_i)}, y^{(s_j,s_i)}\right)$, where $u^{(s_i)}$ is the input force from the actuator at station s_i, and $y^{(s_j,s_i)}$ is the response (displacement/velocity/acceleration) measured by the sensor at station s_j due to the force exerted at station s_i. In the next test the actuator repositions itself to another station (i.e., any admissible location) on the structure, and this procedure is repeated. Let $\mathcal{P}$ be the set of all such stations s_i, $i = \{1, 2, \ldots, P\}$, where the actuator sequentially stations itself. Also, assume that one of these stations is s_j in order to achieve collocation with sensor. Traversing the actuator to the complete set of stations while measuring the response for each test at s_j results in a set $\mathcal{D}^{\text{NCAS}}$ of input-output pairs, each indexed by their corresponding input location, as shown below:

$$\mathcal{D}^{\text{NCAS}} = \left\{ \left(u^{(s_i)}, y^{(s_j,s_i)}\right) : s_i, s_j \in \mathcal{P},\ i = 1, 2, \ldots, P \right\}. \tag{3.3}$$

3.1.2 Collocated actuator-sensor (CAS) case

In the CAS case, the actuator-sensor pair move as a single unit, exciting and sensing different points on the beam, sequentially. The set of input-output pairs in this case is indexed as

$$\mathcal{D}^{\text{CAS}} = \left\{ \left(u^{(s_i)}, y^{(s_i)} \right) : s_i \in \mathcal{P},\ i = 1, 2, \ldots, P \right\}, \tag{3.4}$$

where $y^{(s_i)}$ is the output measured by the collocated sensor at station s_i, due to the input force $u^{(s_i)}$ applied at s_i. Given either dataset (i.e., $\mathcal{D}^{\text{NCAS}}$ or $\mathcal{D}^{\text{CAS}}$) of sequential point input-output measurements, the objective is to estimate the modal parameters (i.e., natural frequencies, damping ratios, and mode shapes), of the beam.

3.2 Mass Normalized Mode Shape Identification

In this section, a formulation to obtain mode shape ordinates is derived for both NCAS and CAS cases based on an input-output balance [94]. It will be shown that although the NCAS case can estimate mode shapes directly, the CAS case requires an additional step which involves approximating the signs of mode shape components. The signs may be obtained from *apriori* knowledge of the structure or mode shapes approximated by a FE model.

NCAS with Acceleration Output: First consider the case of NCAS, with acceleration output measurements (i.e. $y^{(s_j,s_i)} = \ddot{d}^{(s_j,s_i)}$). Assume that only the first n_m modes ($n_m < n$) are dominant in the measured input-output data in $\mathcal{D}^{\text{NCAS}}$. Furthermore, assume that any suitable system identification technique can identify all n_m dominant modal frequencies and damping ratios from $\mathcal{D}^{\text{NCAS}}$. These identified parameters are denoted as: (ω_r, ξ_r) for $r = 1, \ldots, n_m$ where ω_r and ξ_r are the circular natural frequency and damping ratio corresponding to the rth mode, respectively. The objective is to identify the corresponding n_m mode shapes at all the locations excited by the mobile actuator, starting from the identified (ω_r, ξ_r)'s and measured $\left(u^{(s_i)}, \ddot{d}^{(s_j,s_i)} \right)$'s.

By modal superposition, the measured acceleration responses with the n_m dominant modes, can be expressed as:

$$\ddot{d}^{(s_j,s_i)} \approx \sum_{r=1}^{n_m} \phi_{s_j,r}\ \ddot{\eta}_r \tag{3.5}$$

where $\phi_{s_j,r}$ is the component of the rth mode shape at the sensor location s_j, and $\ddot{\eta}_r$ is the rth modal response. For mass normalized mode shapes, substituting Eq.(3.5) in Eq.(3.1), and noting that in a generic test only a single input $u^{(s_i)}$ acts at location s_i, we get the n_m modal equations of motion for $r = 1, \ldots, n_m$:

$$\ddot{\eta}_r + 2\xi_r\omega_r\dot{\eta}_r + \omega_r^2\eta_r = \phi_{s_i,r}\ u^{(s_i)}. \tag{3.6}$$

Since the mass normalized mode shapes have yet to be identified, define a set of n_m *pseudo-modal* equations of motion [94]:

$$\ddot{\tilde{\eta}}_r + 2\xi_r\omega_r\dot{\tilde{\eta}}_r + \omega_r^2\tilde{\eta}_r = u^{(s_i)} \tag{3.7}$$

where $\eta_r = \phi_{s_i,r}\ \tilde{\eta}_r$, and $\tilde{\eta}_r$'s are the *pseudo-modal* responses. Knowing $u^{(s_i)}$ and the identified (ω_r, ξ_r)'s, the pseudo-modal acceleration responses, $\ddot{\tilde{\eta}}_r$ for $r = 1, \ldots, n_m$, can be obtained by solving Eq.(3.7). The measured acceleration response, $\ddot{d}^{(s_j,s_i)}$, can then be expressed as a linear combination of these computed pseudo-modal responses as:

$$\ddot{d}^{(s_j,s_i)} \approx \sum_{r=1}^{n_m} \phi_{s_j,r}\ \ddot{\eta}_r = \sum_{r=1}^{n_m} \phi_{s_j,r}\phi_{s_i,r}\ \ddot{\tilde{\eta}}_r \tag{3.8}$$

such that the coefficients in the linear combination are the products of mass normalized mode shape components at input and output locations. Assuming that the input/output sequences consist of measured values at N time instants ($N > n_m$), estimates of these coefficients can be obtained as a least squares solution to Eq.(3.8) as:

$$\boldsymbol{\alpha}^{(s_j,s_i)} = \mathbf{F}^{(s_i)^\dagger}\underline{\boldsymbol{d}}^{(s_j,s_i)} \tag{3.9}$$

where † denotes the Moore-Penrose pseudo-inverse;

$$\boldsymbol{\alpha}^{(s_j,s_i)} = \left[\hat{\phi}_{s_j,1}\hat{\phi}_{s_i,1}\ ,\ \hat{\phi}_{s_j,2}\hat{\phi}_{s_i,2}\ , \ldots,\ \hat{\phi}_{s_j,n_m}\hat{\phi}_{s_i,n_m}\right]^T \tag{3.10a}$$

$$\underline{\boldsymbol{d}}^{(s_j,s_i)} = \left[\ddot{d}_1^{(s_j,s_i)},\ \ddot{d}_2^{(s_j,s_i)},\ \ldots,\ \ddot{d}_N^{(s_j,s_i)}\right]^T \tag{3.10b}$$

with $\hat{\phi}_{s_i,r}$ denoting an estimate of the corresponding $\phi_{s_i,r}$ and $\ddot{d}_k^{(s_j,s_i)}$ denoting the measured value of $\ddot{d}^{(s_j,s_i)}$ at time instant $k\Delta t$, Δt being the sampling time; and

$$\mathbf{F}^{(s_i)} = \begin{bmatrix} \ddot{\tilde{\eta}}_{1,1} & \ddot{\tilde{\eta}}_{2,1} & \cdots & \ddot{\tilde{\eta}}_{n_m,1} \\ \ddot{\tilde{\eta}}_{1,2} & \ddot{\tilde{\eta}}_{2,2} & \cdots & \ddot{\tilde{\eta}}_{n_m,2} \\ \vdots & \vdots & \ddots & \vdots \\ \ddot{\tilde{\eta}}_{1,N} & \ddot{\tilde{\eta}}_{2,N} & \cdots & \ddot{\tilde{\eta}}_{n_m,N} \end{bmatrix} \tag{3.11}$$

with any $\ddot{\tilde{\eta}}_{r,k}$ denoting the computed value of $\ddot{\tilde{\eta}}_r$ at time instant $k\Delta t$. Repeating Eqs. (3.5)-(3.11) for all the P input-output pairs, one can thus obtain estimates of the products of mode shape components at input-output locations for all the dominant modes, i.e.:

$$\hat{\phi}_{s_j,r}\hat{\phi}_{s_i,r} \qquad \text{for} \;\; r = 1, \ldots, n_m, \;\; s_i \in \mathcal{P}. \tag{3.12}$$

It is evident from Eq.(3.12) that in order to estimate the individual mode shape components from the identified products, there should be at least one test with *collocated input-output*. For the test when the mobile actuator is stationed at the sensor location s_j, the estimates of the mass normalized mode shape coefficients at s_j can be first obtained by taking the square root of the identified products, $\hat{\phi}^2_{s_j,r}$ for $r = 1, \ldots, n_m$. The mode shape components at the remaining actuator locations, i.e. $\hat{\phi}_{s_i,r} \;\forall\; s_i \in \mathcal{P}, \;\; s_i \neq s_j$, can be then obtained by dividing the corresponding products, $\hat{\phi}_{s_j,r}\hat{\phi}_{s_i,r}$, with the estimated $\hat{\phi}_{s_j,r}$, for $r = 1, \ldots, n_m$.

The above method for estimating the mass normalized mode shape components can be similarly applied for the case of a static actuator-mobile sensor (setup (b)). The only condition for unique identification is that at least one test should have the actuator and sensor collocated. In practice, it may sometimes be necessary to have more than one test with collocated actuator-sensor, with the locations of collocation changing. This necessity will arise in situations where, for example, the first location of collocation (say s_j) is a node for one of the modes (say mode r); then, the estimated $\hat{\phi}_{s_j,r}$ will be zero, making it impossible to estimate the remaining components of the rth mode shape. Having test data from additional tests with different locations of the collocated actuator-sensor will avoid such issues. Hence, the case where both the actuator and the sensor are mobile is, in general, preferable.

CAS with Acceleration Output: When the actuator and sensor are always collocated ($s_j = s_i \ \forall \ s_i \in \mathcal{P}$), applying Eqs. (3.5)-(3.11) for all P input-output pairs gives the estimates of the squares of the mass normalized mode shape components, at all input-output locations and for all the dominant modes:

$$\hat{\phi}^2_{s_i,r} \qquad \text{for} \ \ r = 1, \ldots, n_m, \ \ s_i \in \mathcal{P}. \tag{3.13}$$

However, in this case, the individual mode shape components with appropriate signs cannot be obtained from the measured data alone. While the magnitude of the mode shape components can be obtained as the square roots of the estimated mode shape squares, the signs need to be assigned based on some additional information. For this purpose, it is assumed that the sign of any estimated mode shape component is the same as the sign of the corresponding component obtained from an approximate FE model of the structure. It is important to underscore that a FE model is not needed in this approach *per se*, however the relative mode shape signs are necessary to address the sign ambiguity. In this way, the estimated mass normalized mode shape components are obtained as:

$$\hat{\phi}_{s_i,r} = \texttt{sgn}\big(\phi^0_{s_i,r}\big)\sqrt{\hat{\phi}^2_{s_i,r}} \qquad \text{for} \ \ r = 1, \ldots, n_m, \ \ s_i \in \mathcal{P} \tag{3.14}$$

where '`sgn`' is the signum function, and $\phi^0_{s_i,r}$ is the component at location s_i of the rth FE mode shape.

Alternative Formulation for Acceleration Output: Here, the least squares estimation in Eq.(3.9) is re-formulated in terms of pseudo-modal velocity and pseudo-modal displacement responses, while still using measured acceleration outputs. As will be demonstrated in the numerical study, this alternative formulation leads to improved estimates of the fundamental mode shape, possibly due to the lower contribution of the fundamental mode to acceleration as compared to velocity/displacement. Recalling that the estimated mode shapes are mass normalized, and thus $\mathbf{M}^{-1} = \mathbf{\Phi}\mathbf{\Phi}^T$, it is possible to obtain the following relations:

$$\mathbf{M}^{-1}\mathbf{K}\mathbf{\Phi} = \mathbf{\Phi}\mathbf{\Phi}^T\mathbf{K}\mathbf{\Phi} = \mathbf{\Phi}\mathbf{\Lambda} \tag{3.15a}$$

$$\mathbf{M}^{-1}\mathbf{C}\mathbf{\Phi} = \mathbf{\Phi}\mathbf{\Phi}^T\mathbf{C}\mathbf{\Phi} = \mathbf{\Phi}\mathbf{C}_m \tag{3.15b}$$

where $\mathbf{C}_m$ is a diagonal matrix with elements $2\xi_r\omega_r$, $r = 1, 2, \ldots, n$, along the diagonals. Considering the modal superposition $\boldsymbol{d} = \boldsymbol{\Phi\eta}$ in Eq.(3.1) along with the above relations for the r dominant modes, one can write:

$$\begin{aligned}\ddot{\boldsymbol{d}} &= \mathbf{M}^{-1}\mathbf{L}u - \mathbf{M}^{-1}\mathbf{C}\boldsymbol{\Phi}\dot{\boldsymbol{\eta}} - \mathbf{M}^{-1}\mathbf{K}\boldsymbol{\Phi\eta} \\ &= \mathbf{M}^{-1}\mathbf{L}u - \boldsymbol{\Phi}\mathbf{C}_m\dot{\boldsymbol{\eta}} - \boldsymbol{\Phi\Lambda\eta} \\ &\approx \mathbf{M}^{-1}\mathbf{L}u - \sum_{r=1}^{n_m} 2\xi_r\omega_r\boldsymbol{\phi}_r\dot{\eta}_r - \sum_{r=1}^{n_m}\omega_r^2\boldsymbol{\phi}_r\eta_r \end{aligned} \tag{3.16}$$

where $\boldsymbol{\phi}_r$ is the rth mode shape. The acceleration output at s_j, due to the single input applied at s_i, can then be expressed in terms of the pseudo-modal velocities and displacements as:

$$\ddot{d}^{(s_j,s_i)} = \mathbf{M}^{-1}_{s_j,s_i}u^{(s_i)} - \sum_{r=1}^{n_m}\phi_{s_i,r}\phi_{s_j,r}\left(2\xi_r\omega_r\dot{\tilde{\eta}}_r - \omega_r^2\tilde{\eta}_r\right) \tag{3.17}$$

where $\mathbf{M}^{-1}_{s_j,s_i}$ is the (s_j, s_i)th element of $\mathbf{M}^{-1}$. This leads to an alternative least squares solution to the products of mode shape components at input and output locations:

$$\boldsymbol{\alpha}^{(s_j,s_i)} = \mathbf{H}^{(s_i)^\dagger}\mathbf{Z}^{(s_j,s_i)} \tag{3.18}$$

where $\mathbf{Z}^{(s_j,s_i)} = \left[z_1^{(s_j,s_i)},\ z_2^{(s_j,s_i)},\ \ldots,\ z_N^{(s_j,s_i)}\right]^T$, with any $z_k^{(s_j,s_i)} = \mathbf{M}^{-1}_{s_j,s_i}u_k^{(s_i)} - \ddot{d}_k^{(s_j,s_i)}$; and

$$\mathbf{H}^{(s_j,s_i)} = \begin{bmatrix} h_{1,1} & h_{2,1} & \cdots & h_{n_m,1} \\ h_{1,2} & h_{2,2} & \cdots & h_{n_m,2} \\ \vdots & \vdots & \ddots & \vdots \\ h_{1,N} & h_{2,N} & \cdots & h_{n_m,N} \end{bmatrix} \tag{3.19}$$

with any $h_{r,k} = 2\xi_r\omega_r\dot{\tilde{\eta}}_r(k\Delta t) - \omega_r^2\tilde{\eta}_r(k\Delta t)$. The scalar $\mathbf{M}^{-1}_{s_j,s_i}$ is the direct feedthrough term in a state space model of this system; any appropriate system identification technique can be used to identify such a model from the measured single input-single output data $\left(u^{(s_i)}, \ddot{d}^{(s_j,s_i)}\right)$, thus giving this scalar.

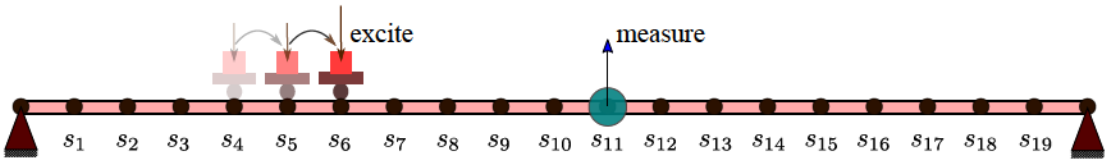


Figure 3.2: Experimental setup in case of NCAS: static sensor at s_{11}, mobile actuator sequentially exciting at s_1 to s_{19}.

3.3 Numerical Study

In this section, the performance of the proposed approach is assessed using numerically simulated data from a simply-supported beam of length $L = 20$m, flexural rigidity 3.2×10^8Nm^2, and mass density 2000kg/m. A FE model consisting of 60 Euler-Bernoulli beam elements is used for simulating the input-output measurements. With this model, the natural frequencies for the first five modes are 1.57, 6.28, 14.14, 25.14, and 39.28Hz. The damping matrix $\mathbf{C}$ is constructed by specifying the modal damping ratios: 2.5% of critical for the first five modes, and 5% for the remaining modes.

A total of 19 measurement stations ($s_1 - s_{19}$), equally spaced at 1m intervals, are considered where input-output measurements are collected. For the NCAS case (Figure 3.2), the response is assumed to be measured at station s_{11} in the vertical direction, while all 19 vertical DoFs at stations $s_1 - s_{19}$ are sequentially excited by a mobile actuator. For the CAS case (Figure 3.3), both the actuator and the sensor are roved together as an integrated mobile unit, sequentially exciting and measuring all vertical DoFs at stations $s_1 - s_{19}$. Zero mean and unit variance Gaussian random excitations bandlimited such that the energy drops gradually after 25 Hz (see Figure 3.4), are used as input forces. Different types of outputs (displacement, velocity, acceleration) are considered. For acceleration outputs, the estimation using pseudo-modal accelerations is termed Approach I, while the estimation using pseudo-modal velocities and displacements is termed Approach II. The simulations are performed with a sampling frequency of 500 Hz.

Based on the actuation bandwidth, only the first four modes can be identified from the input-output data. The eigensystem realization algorithm with observer Kalman filter

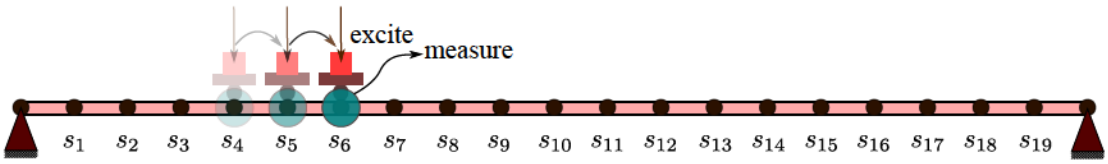


Figure 3.3: Experimental setup in case of CAS: mobile actuator-sensor pair sequentially exciting and measuring at s_1 to s_{19}.

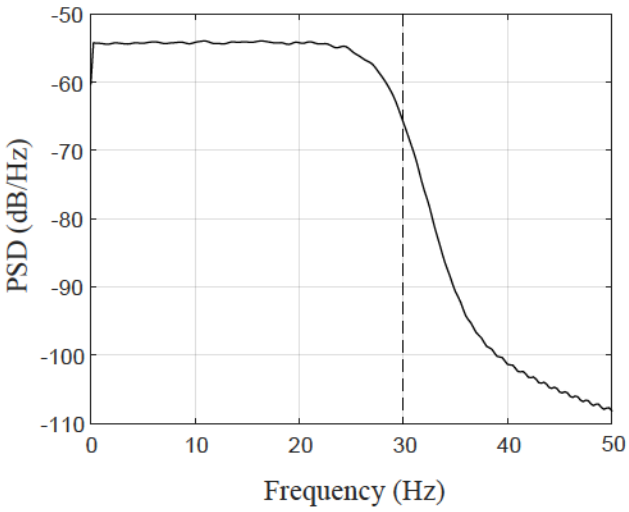


Figure 3.4: Power spectral density of typical random excitation used in numerical simulations.

Table 3.1: Comparison of MAC values between identified NCAS and FE mode shapes.

Output	NCAS			
	Mode 1	Mode 2	Mode 3	Mode 4
Acceleration (Approach I)	0.999	0.999	0.999	0.998
Acceleration (Approach II)	1.000	0.999	0.999	0.997
Velocity	1.000	1.000	1.000	1.000
Displacement	1.000	1.000	1.000	1.000

identification (ERA-OKID) [25] is used to identify these four sets of natural frequencies and modal damping ratios from the simulated input-output data. The proposed approach is then used to identify the 19 vertical components (at stations $s_1 - s_{19}$) of each of the first four mode shapes. In the case of CAS, the signs of the corresponding FE mode shape components are used with the magnitudes of the mode shape components identified from data.

The identified mode shapes are compared with the FE mode shapes using the modal assurance criterion (MAC) [95]. The MAC values obtained using different types of outputs are listed in Tables 3.1 and 3.2 for the NCAS and CAS cases respectively. It is observed that the estimates obtained using velocity and displacement outputs are generally more accurate than the estimates from acceleration outputs. This may be due to the increased effect of modal truncation in case of acceleration, which has higher contribution from high frequency modes as compared to velocity and displacement. For possibly the same reason, when using acceleration outputs, Approach II is found to provide slightly better estimates of the fundamental mode shape than Approach I. The least squares fit obtained using lower frequency pseudo-modal velocities and displacements may be less affected by modal truncation than the fit obtained using lower frequency pseudo-modal accelerations. Nonetheless, the estimated mode shapes in all cases can be considered to be sufficiently accurate from a practical perspective. This is evident from Figures 3.5 and 3.6, where the identified mode shapes are graphically compared with the corresponding FE mode shapes for the NCAS and CAS cases.

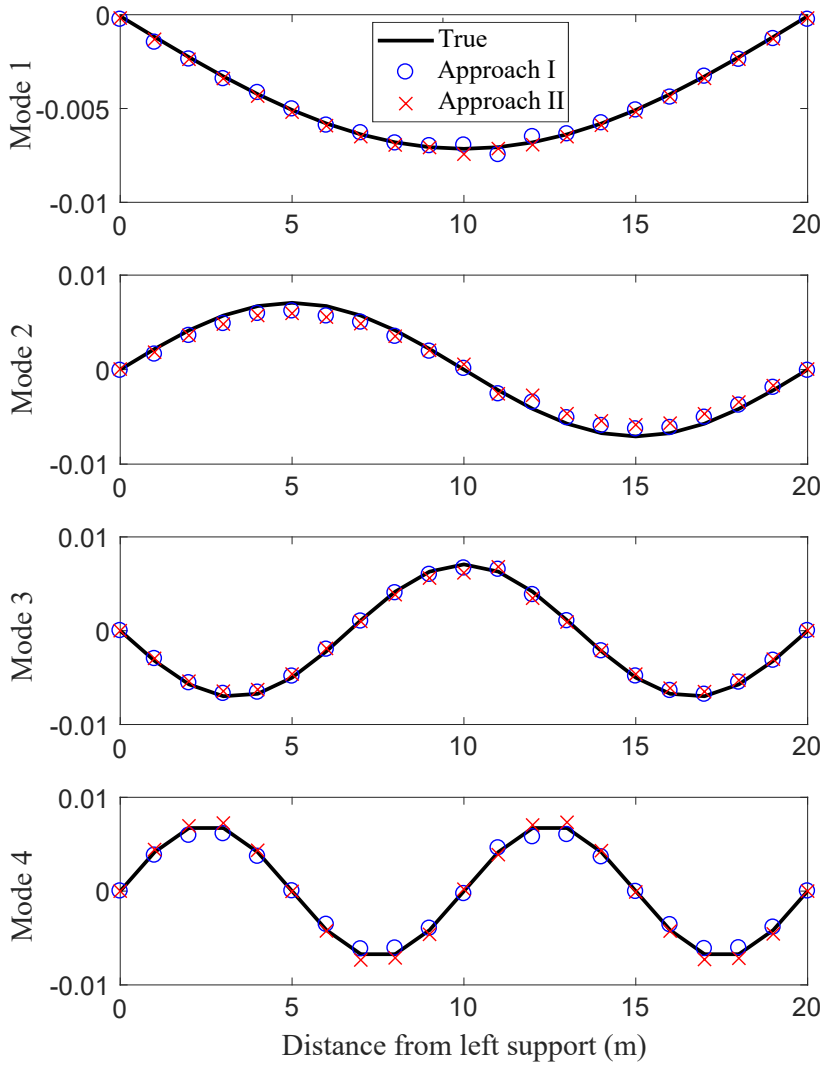


Figure 3.5: Comparison of identified mode shapes from acceleration outputs with FE mode shapes, for the case of NCAS.

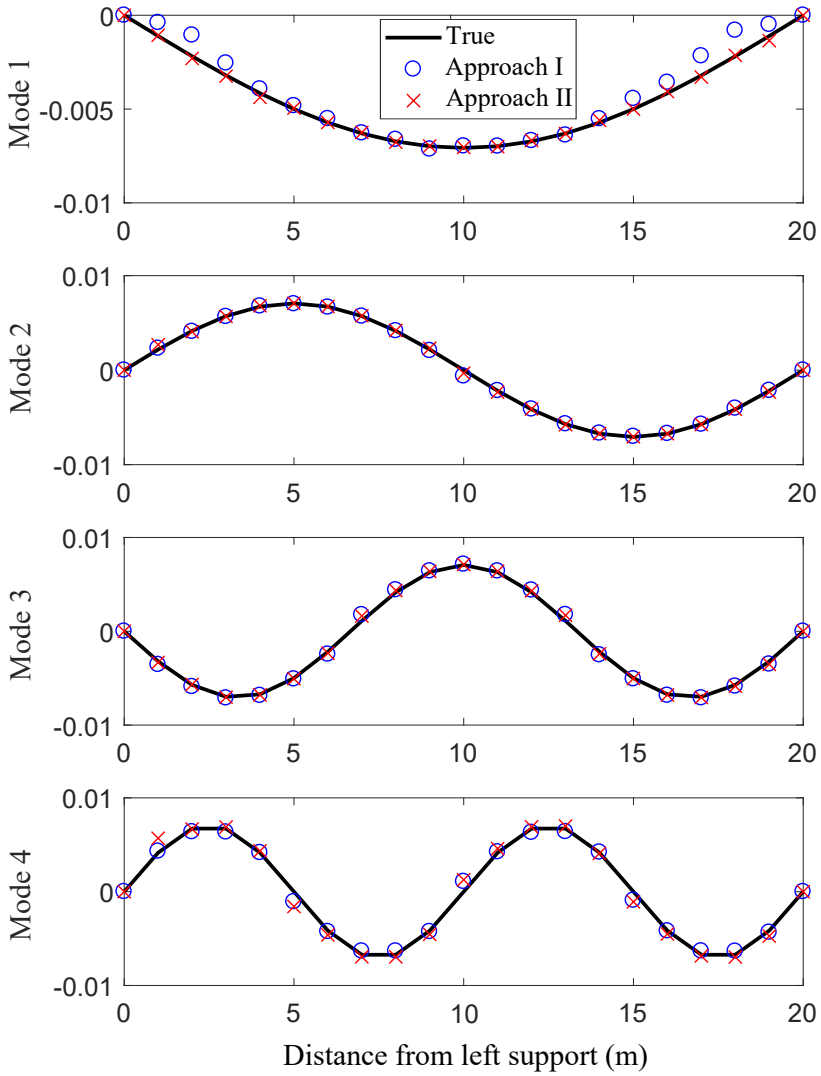


Figure 3.6: Comparison of identified mode shapes from acceleration outputs with FE mode shapes, for the case of CAS.

Table 3.2: Comparison of MAC values between identified CAS and FE mode shapes.

Output	CAS			
	Mode 1	Mode 2	Mode 3	Mode 4
Acceleration (Approach I)	0.988	0.999	0.998	0.992
Acceleration (Approach II)	0.999	0.999	0.998	0.992
Velocity	0.999	0.999	1.000	0.994
Displacement	0.999	0.999	0.999	0.985

3.3.1 Sensitivity studies

As demonstrated in the methodology, the accuracy of the identified mode shape components is not only affected by the accuracy of the identified modal frequencies and damping ratios, but also the presence of noise in the measured input-output data. Furthermore, when using Approach II for acceleration outputs, any error in the estimated direct feedthrough term of an identified state space model will affect the estimates of the mode shapes. For the case of CAS, the estimated mode shape components will also be affected by any erroneous assignment of sign from the FE mode shapes. In this section, the effect of these various sources of errors on the accuracy of the identified mode shapes is studied. Since the underlying methodology for both the CAS and NCAS cases are same, the sensitivity studies reported here are based on simulations of the NCAS case only, unless the CAS case is explicitly mentioned. The same numerical example considered in Section 3.3 is used here for the sensitivity studies.

Sensitivity to errors in modal frequencies

It is found that errors in the estimates of modal frequencies can produce significant errors in the estimates of $\boldsymbol{\alpha}^{(s_j,s_i)}$, and consequently, the identified mode shape components. This is illustrated in Figure 3.7, which compares the estimates of $\hat{\phi}^2_{11,r}$ obtained from acceleration outputs (both Approach I and II), due to an under-estimation and an over-estimation, by 5%, of the 1st natural frequency (3.7(a) and 3.7(b)), or of the 4th natural frequency (3.7(c)

and 3.7(d)).

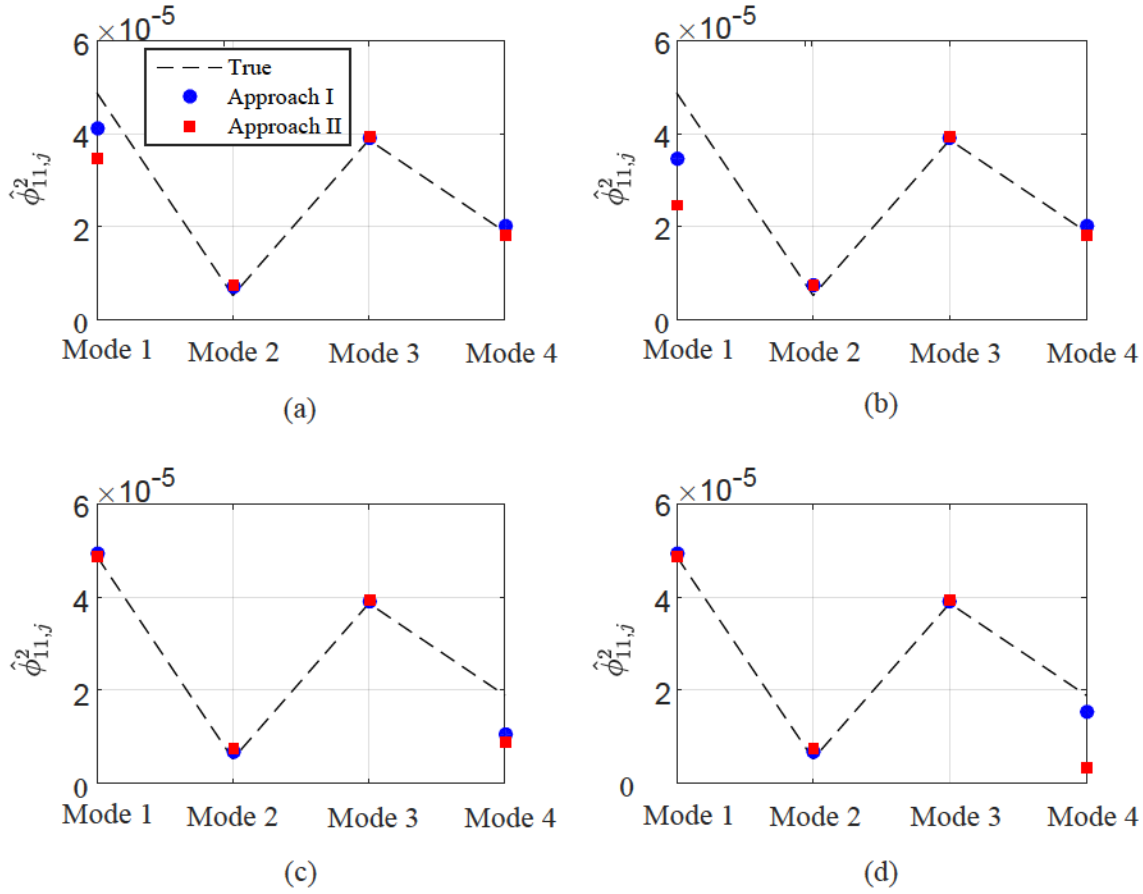


Figure 3.7: Estimates of $\hat{\phi}^2_{11,r}$ from acceleration outputs for: (a) 5% underestimation and (b) 5% overestimation of the 1st natural frequency; and (c) 5% underestimation and (d) 5% overestimation of the 4th natural frequency.

It can be inferred from these plots that an error in any identified frequency largely affects the estimated mode shape for the same mode. This is further illustrated in Figure 3.8, where MAC values comparing the estimated (from acceleration output, Approach I) and FE mode shapes are plotted, for different levels of perturbations in the 1st to 4th natural frequencies. The perturbed frequencies are obtained by multiplying the corresponding true

frequencies with a scalar factor β varying between 0.9 and 1.1 (i.e. between $\pm 10\%$ error).

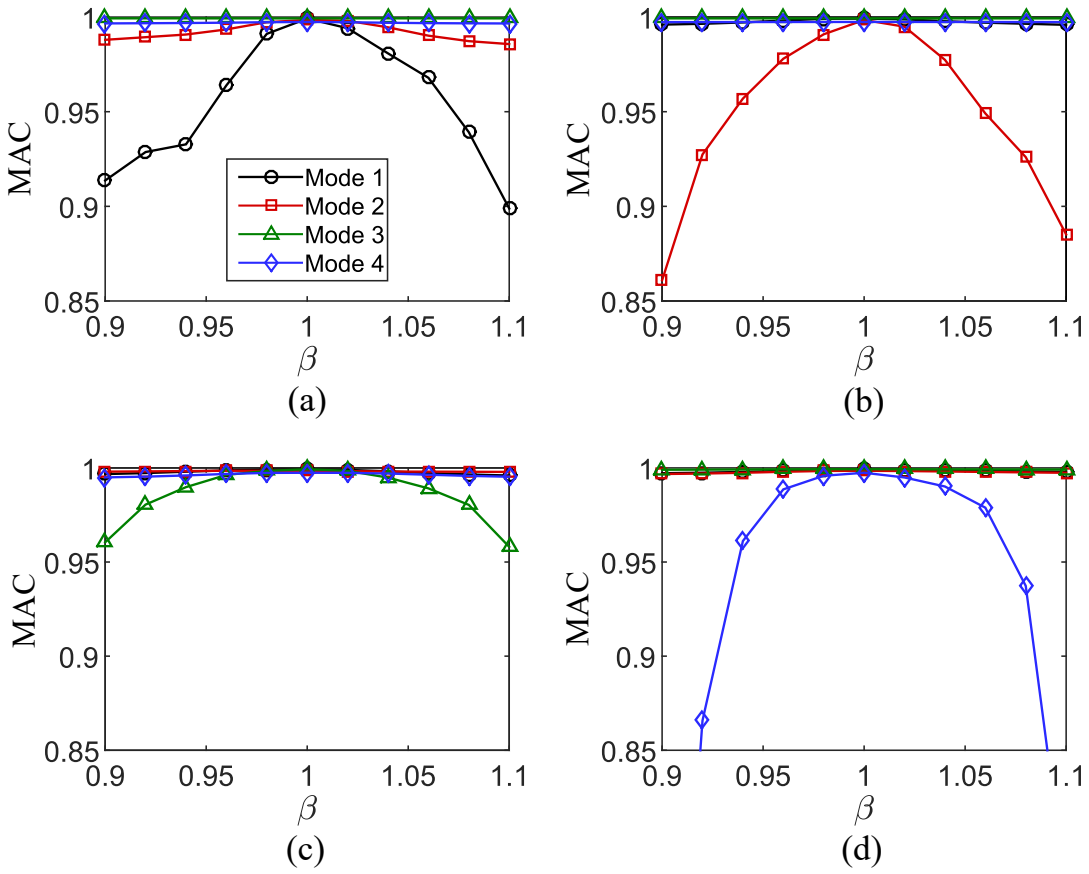


Figure 3.8: Sensitivity of MAC values comparing estimated and FE mode shapes to errors in frequency of: (a) 1st mode, (b) 2nd mode, (c) 3rd mode, and (d) 4th mode, for acceleration output (Approach I).

However, this behavior of erroneous frequency estimates mostly affecting corresponding mode shape estimates is not true for all output types. This can be seen from Figure 3.9, where MAC values comparing estimated and FE mode shapes, for different types of outputs, are plotted against the different levels of perturbations in the 1st and 4th natural frequencies.

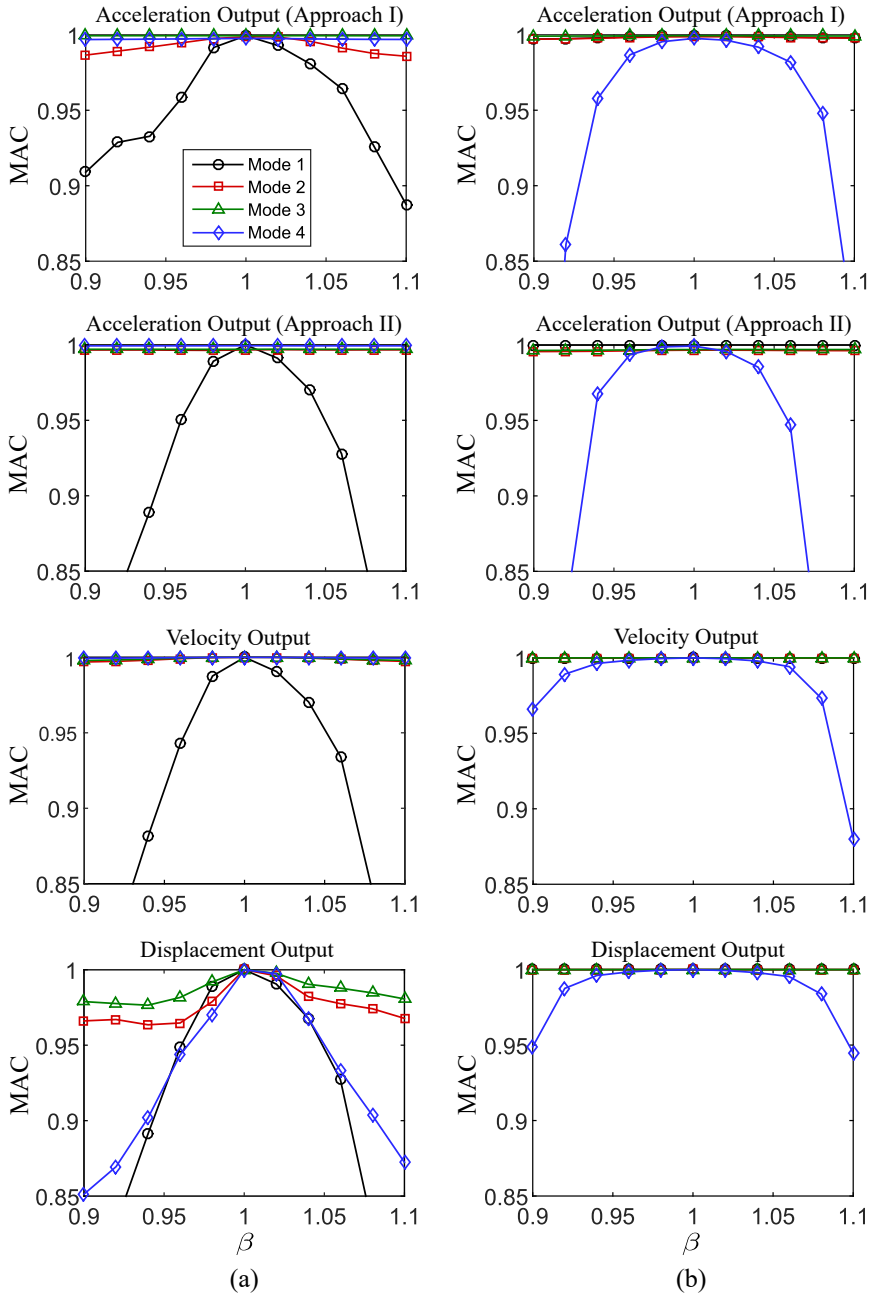


Figure 3.9: Sensitivity of MAC values comparing estimated and FE mode shapes to errors in frequency of: (a) 1st mode, and (b) 4th mode, for different output types.

It is evident that erroneous frequency induced errors in estimated mode shapes vary with the output type. Further, unlike acceleration/velocity outputs, for displacement outputs, errors in the 1st frequency affects the mode shape estimates for all modes, albeit by different extents. This is possibly owing to the fact that displacement responses are more sensitive to lower frequencies, as compared to acceleration/velocity responses.

Sensitivity to errors in modal damping ratios

Compared to modal frequencies, errors in estimated damping ratios are found to affect the mode shape estimates to a considerably lesser extent. This is illustrated in Figure 3.10,

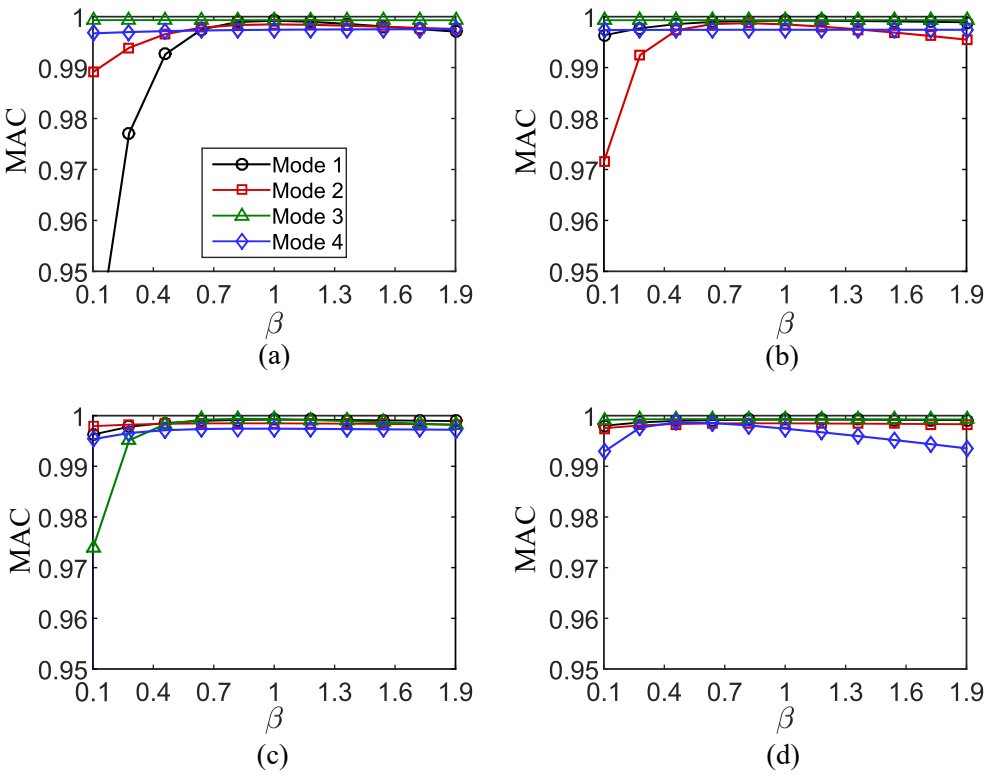


Figure 3.10: Sensitivity of MAC values comparing estimated and FE mode shapes to errors in damping ratio of: (a) 1st mode, (b) 2nd mode, (c) 3rd mode, and (d) 4th mode, for acceleration output (Approach I).

where the MAC values comparing estimated (from acceleration output, Approach I) and FE mode shapes are plotted, for different levels of perturbations in the 1st to 4th modal damping ratios.

Similar to the sensitivity testing for frequencies, the perturbed damping ratios are obtained by multiplying the corresponding true ratios with a scalar factor β; in this case, β is varied between 0.1 and 1.9 (i.e. between $\pm 90\%$ error), to account for the large errors usually observed in identified damping ratios. Similar to frequencies, the largest effect of an erroneous rth modal damping ratio is found to be on the estimated rth mode shape. Nonetheless, the effect of erroneous damping ratios on mode shape estimates can be observed to be much lower as compared to the effect of erroneous frequencies.

3.3.2 Sensitivity to noise in input-output data

The unavoidable presence of measurement noise in the collected input-output data will not only affect the estimates of the mode shapes directly, in Eqs. (3.9) or (3.18), but also indirectly, through the estimation errors in frequencies and damping ratios. Since the effect of erroneous frequencies and damping ratios on mode shape estimates have already been studied, the focus here is on the impact of measurement noise directly on the identified mode shapes. To this end, both the input and output measurements are corrupted with additive noise, while the true values of the modal frequencies and damping ratios are used, when obtaining the mode shape estimates. The noise added to the input and output data are zero mean Gaussian white noise sequences, having root-mean-square (RMS) values equal to $x\%$ of the RMS of corresponding "true" signals; x is varied between 5 to 30 in this study. Figure 3.11 shows the MAC values comparing the estimated mode shapes with the FE mode shapes, for varying levels of noise, when using different types of outputs. It is evident that the direct effect of measurement noise on the estimated mode shapes is considerably less for all output types, with velocity output being the most robust to noise.

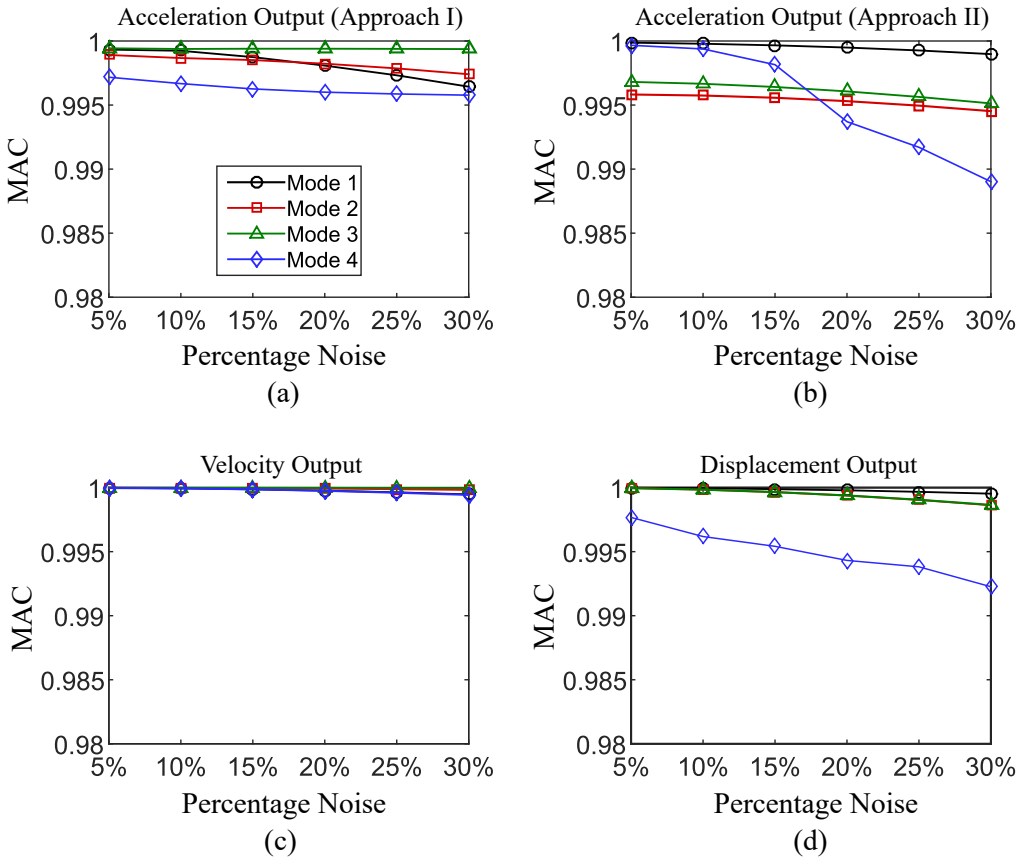


Figure 3.11: Sensitivity of MAC values comparing estimated and FE mode shapes to varying percentage of noise in input-output data, when using different outputs: (a) acceleration (Approach I), (b) acceleration (Approach II), (c) velocity, and (d) displacement.

Although the direct effect of measurement noise is found to be insignificant, the indirect effect of noisy data can often be significant. To study the effect of measurement noise on the frequencies and damping ratios identified using ERA-OKID, input-output data corrupted with different levels of noise, i.e., 10, 20 and 30% of RMS, is used. Acceleration response is used here as the output. To differentiate between the structural modes and spurious (noise) modes, stabilization diagram [96] using an averaged response frequency response function (FRF) is used with typical stabilization criteria of 1% variation for frequency and 5% variation for damping ratio. Figure 3.12 shows a typical stabilization diagram used to

select the structural modes when the input-output measurements are corrupted by 20% noise.

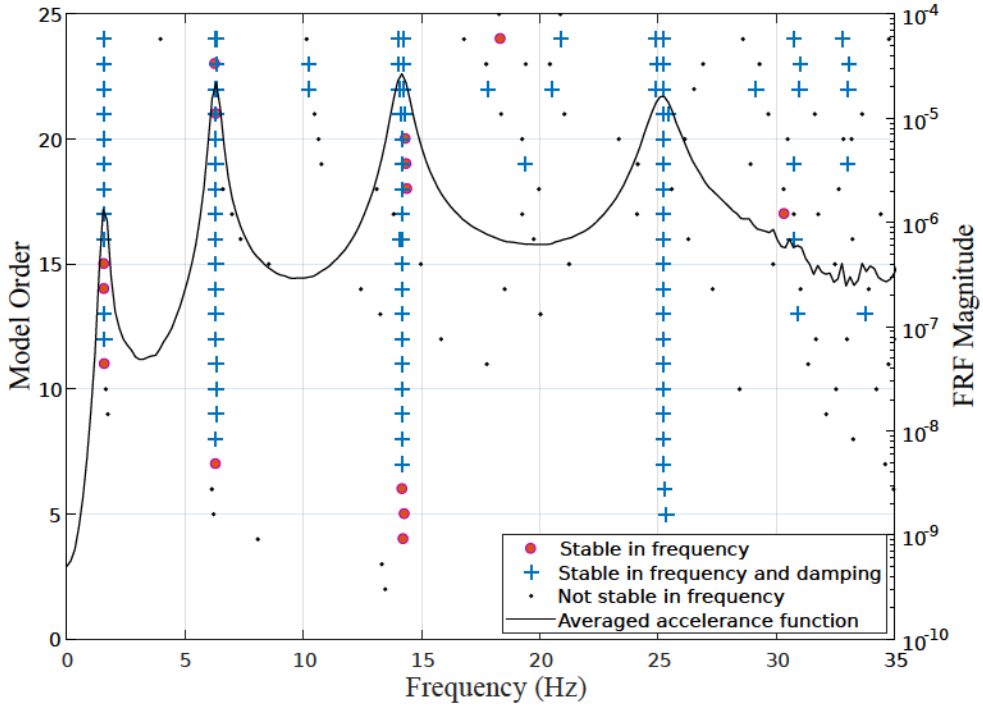


Figure 3.12: Stabilization diagram in case of input-output (acceleration) measurements corrupted with 20% noise.

Once the stable modes are identified from the stabilization diagram, the ERA-OKID estimates of the corresponding natural frequencies and damping ratios are selected for all input-output pairs, and the estimates are averaged for each mode (over all input-output pairs). Table 3.3 shows the absolute percentage errors in these averaged estimates of natural frequencies and damping ratio, obtained using the 19 input-output pairs. The absolute percentage errors are computed as the absolute values of the percentage errors in the estimated quantities with respect to the corresponding true quantities. While the errors in the estimated natural frequencies are found to mostly lie within $0-5\%$, the errors in the estimates of the damping ratios can be significantly high (nearly 500%). Higher errors

Table 3.3: Absolute percentage error in natural frequencies and damping ratios identified using ERA-OKID, for different levels of noise (10, 20 and 30%) in input-output (acceleration) data.

Mode	f_{true} (Hz)	$\lvert\text{error}_f\rvert$ (in %)			ξ_{true} (%)	$\lvert\text{error}_\xi\rvert$ (in %)		
		10%	**20%**	**30%**		**10%**	**20%**	**30%**
1	1.57	1.28	2.86	4.82	2.50	118.02	187.83	480.84
2	6.28	0.42	0.83	1.92	2.50	29.84	105.72	170.61
3	14.14	0.00	0.13	0.20	2.50	0.57	1.30	12.24
4	25.14	0.02	0.22	0.80	2.50	0.12	0.14	0.32

in damping estimates is also supported by other studies [97, 98] where the uncertainty in estimated damping ratios has been shown to be higher as compared to estimated natural frequencies. It can also be seen that the errors are higher for Mode 1, and they decrease considerably for Modes 3 and 4. This is possibly due to higher energies in Modes 2 to 4 compared to Mode 1, as can be seen from the accelerance FRF in Figure 3.12.

Although the errors in frequency estimates are considerably lower than in damping, even such low errors would adversely affect the mode shape estimation, as shown earlier. Furthermore, the significantly high errors in damping would also affect the identified mode shapes, even though the mode shape estimation is less sensitive to errors in damping. Hence, to improve the estimates of the modal frequencies and damping ratios, prior to mode shape identification, the following nonlinear least squares optimization step may be incorporated to obtain optimal estimates of the modal frequencies and damping ratios, starting from the initial estimates obtained using ERA-OKID:

$$\hat{\boldsymbol{\gamma}}_p^* = \arg\min_{\hat{\boldsymbol{\gamma}}_p} \sum_{k=1}^{N} \sum_{i=1}^{P} \left(\ddot{d}_k^{(s_i)} - \ddot{\hat{d}}_k^{(s_i)} \left(\hat{\boldsymbol{\gamma}}_p \right) \right)^2 \tag{3.20}$$

In Eq.(3.20), $\hat{\boldsymbol{\gamma}}_p = \left\{ \hat{\omega}_1, \ldots, \hat{\omega}_r, \hat{\xi}_1, \ldots, \hat{\xi}_r \right\}$ denotes the parameter vector of frequencies and damping ratios; $\ddot{\hat{d}}_k^{(s_i)} \left(\hat{\boldsymbol{\gamma}}_p \right)$ is the reconstructed response at location s_i at time instant $k\Delta t$ computed using the proposed approach for any $\hat{\boldsymbol{\gamma}}_p$; and $\hat{\boldsymbol{\gamma}}_p^*$ is the optimal $\hat{\boldsymbol{\gamma}}_p$ which

Table 3.4: Absolute error (AE) percentage in optimized estimates of natural frequencies and damping ratios, for different levels of noise (10, 20 and 30%) in input-output (acceleration) data.

Mode	f_{true} (Hz)	AE_f (in %)			ξ_{true} (%)	AE_ξ (in %)		
		10%	**20%**	**30%**		**10%**	**20%**	**30%**
1	1.57	0.29	0.56	1.23	2.50	41.02	57.83	78.84
2	6.28	0.05	0.17	0.41	2.50	3.54	5.26	12.51
3	14.14	0.00	0.00	0.01	2.50	0.19	0.34	1.00
4	25.14	0.01	0.02	0.04	2.50	0.08	0.10	0.13

minimizes the sum of squared errors (over all stations and all time instants) between the measured and reconstructed responses. Table 3.4 shows the errors in the optimal frequency and damping ratio estimates obtained using Eq.(3.20). Comparing with Table 3.3, a significant improvement, especially in the damping ratios, is evident.

Finally, to examine the performance of the proposed mode shape estimation method with noisy input-output data, including both the direct and indirect effects of noise, the identified optimal values of the modal frequencies and damping ratios are used to identify the mode shapes from the noisy data. Table 3.5 shows the MAC values comparing these identified modes with FE modes, for cases with different levels of noise (i.e., 10%, 20% and 30%); both approaches using acceleration outputs are considered. It is evident that the mode shapes identified using the proposed approach are reasonably accurate, even in the presence of significant noise in the input-output data. For comparative purposes, the MAC values for modes identified using an alternative technique, the least squares frequency domain (LSFD) [99], are also included in Table 3.5. It is observed that, while the performance of the proposed approach and LSFD in identifying the higher modes are comparable (with MAC values ≈ 0.99 in all cases), the proposed approach performs distinctly better than LSFD in estimating the fundamental mode shape in presence of higher noise.

Note that, in the identification using Approach II, noisy data may have another indirect

Table 3.5: MAC values comparing estimated and FE mode shapes, for identification using noisy input-output (acceleration) data, with different noise levels (10%, 20% and 30%), in the case of NCAS.

Method	**Mode 1**	**Mode 2**	**Mode 3**	**Mode 4**	**Noise**
Approach I	0.998	0.998	0.998	0.999	
Approach II	0.999	0.999	0.998	0.995	10%
LSFD	0.988	0.999	0.999	0.998	
Approach I	0.996	0.994	0.998	0.998	
Approach II	0.999	0.996	0.997	0.990	20%
LSFD	0.978	0.997	0.998	0.997	
Approach I	0.994	0.993	0.998	0.997	
Approach II	0.999	0.996	0.997	0.989	30%
LSFD	0.956	0.996	0.997	0.997	

effect on the estimated mode shapes. In this approach, an estimate of the direct feedthrough term in the state space model, is necessary prior to identifying the mode shapes. The direct feedthrough term is estimated from the state space model identified using ERA-OKID. In this investigation, since each test consists of a single input - single output measurement, the identified direct feedthrough is a scalar. Presence of noise in the input-output data will affect the estimate of direct feedthrough, which in turn may affect the estimated mode shapes. However, comparing the MAC values for Approaches I and II in Table 3.5, it can be seen that noise-induced errors in the direct feedthrough term do not produce any significant additional errors in the estimated mode shapes.

3.3.3 On signs of mode shape components for the CAS case

In the case of CAS, the signs of the mode shape components are assigned from corresponding approximate FE mode shapes. This could lead to erroneous estimates of the mode shapes if the assigned signs are incorrect. Such mismatch between the sign of an FE mode shape component and the "true" sign of that component may be expected at

locations close to the nodes for mode shapes, especially in situations when there are significant changes in the stiffness distribution (due to damage). Hence, this issue is particularly relevant in structural health monitoring applications. It is assumed that a FE model of the original structure is available, while the test is performed on the structure in its current deteriorated condition (referred to as the "test structure"). To study the frequency of such deterioration-induced sign disparities, three possible types of degradation in the test structure are considered:

- *Case 1 – Distributed change in stiffness* : In this case, the test structure is affected by a reduction of stiffness over the entire span of the structure leading to an approximately uniform reduction in stiffness of the structure. This case is numerically simulated by introducing random stiffness reductions of $\vartheta\%$ in all 60 elements of the beam used in Section 3.3. For any element, the reduction α is sampled from the uniform distribution on interval (0.13, 0.17), resulting in a mean stiffness reduction of 15% across all elements.

- *Case 2 – Localized damage*: In this case, the test structure is affected by localized damage, e.g. due to fatigue cracks, represented here as a localized reduction in stiffness. Two different damage locations are considered, with the stiffness of the corresponding affected element reduced by $\vartheta\%$: (a) damage at mid-span, affecting element 30, and (b) damage at quarter-span, affecting element 15. In both situations, the stiffness reduction ϑ is sampled from uniform distribution on interval (0.18, 0.22).

- *Case 3 – Partial fixity at support*: In this case, the supports of the test structure develop some degree of fixity, e.g. due to accumulation of corrosion products at joints. This is modeled here by introducing rotational springs at the two supports of the beam, with stiffness values: $k_r^{\text{left}} = \vartheta_1 \times 10^{10}$ Nm/rad and $k_r^{\text{right}} = \vartheta_2 \times 10^{10}$ Nm/rad, with the values of ϑ_1 and ϑ_2 sampled from uniform distribution on interval (50,200).

For each of the above cases, 10000 different test structures are simulated by sampling the ϑ values from the respective uniform distributions. The event of sign discrepancy at any location s_i, for any mode r, occurs when $\text{sgn}(\phi_{s_i,r}^{\text{FE}}) \neq \text{sgn}(\phi_{s_i,r}^{\text{test structure}})$, where the FE model is the model of the beam from Section 3.3 in all cases. To ensure that both the FE and

test structure mode shapes have the same global directions (e.g. the first mode shape is in the downward direction, as in Figure 3.5), the signs of corresponding mode shapes at the first translational DoF are kept same. Based on the 10000 simulations in each case, the probability of sign discrepancy (PoSD), at location s_i for mode r, is defined as:

$$\text{PoSD}(s_i, r) = \frac{1}{10000} \sum_{i=1}^{10000} \mathbf{I}\left(\text{sgn}(\phi^{\text{FE}}_{s_i,r}) \neq \text{sgn}(\phi^{\text{test structure}}_{s_i,r})\right) \tag{3.21}$$

where $\mathbf{I}(\cdot)$ is the indicator function. For each case, the computed PoSD values for modes 2 to 4 are shown in Figure 3.13; mode 1 is not included as it has a zero PoSD at all locations in all cases. Comparing these results with the mode shapes plotted in Figure 3.5, it is evident that the mismatch in sign between the FE and test structure mode shape components occur only near the nodes of the FE mode shapes. While the discrepancies in sign occur in around 50% of the simulated test structures when there is distributed deterioration, they occur in almost all test structures when there is localized damage or partial fixity at the supports. It is also shown that, local damage at mid-span induces sign changes near nodes of modes all along the span (Figure 3.13(b)), but with a lesser frequency for symmetric modes (PoSD is zero for mode 2, 0.5 for mode 4, but 1.0 for mode 3). On the other hand, the sign changes induced by a local damage away from the center does not occur along the entire span, but only around the nodes nearer to the damage location (Figure 3.13(c)); the frequency of sign change in this case is however high for all modes. This is possibly due to local damage near the center creating a more global effect on the structure, as compared to damage away from the center. However, note that since the components near nodes of any mode shape are themselves very small, it is expected that the estimated mode shapes, even with incorrect signs of such near-node components, may still be used for further applications (e.g. condition monitoring), without significantly affecting the results of such applications.

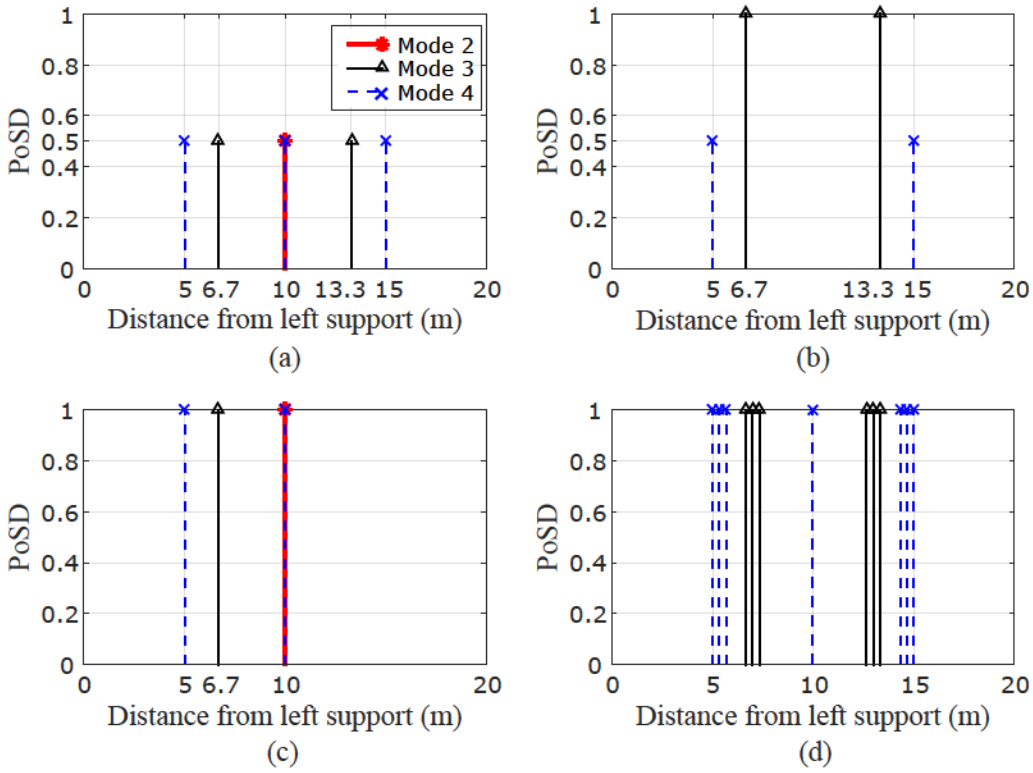


Figure 3.13: Probability of sign discrepancy between FE and test structure modes, for test structure with: (a) distributed change in stiffness, (b) localized damage at mid-span, (c) localized damage at quarter span, and (d) partial fixity of supports.

3.4 Experimental Application

In this section, the proposed approach is applied on a laboratory scale simply supported steel beam. The experimental set-up is shown in Figure 3.14, with a schematic representation of the beam and the tests shown in Figure 3.15. The tested beam is 1.4m in length, and has a uniform 63 mm wide $\times$ 3 mm deep rectangular cross-section (see Figure 3.15). Along the length of the beam, 19 locations, uniformly spaced at 0.07m intervals, are selected as stations where the beam is excited and/or the response of the beam is measured.

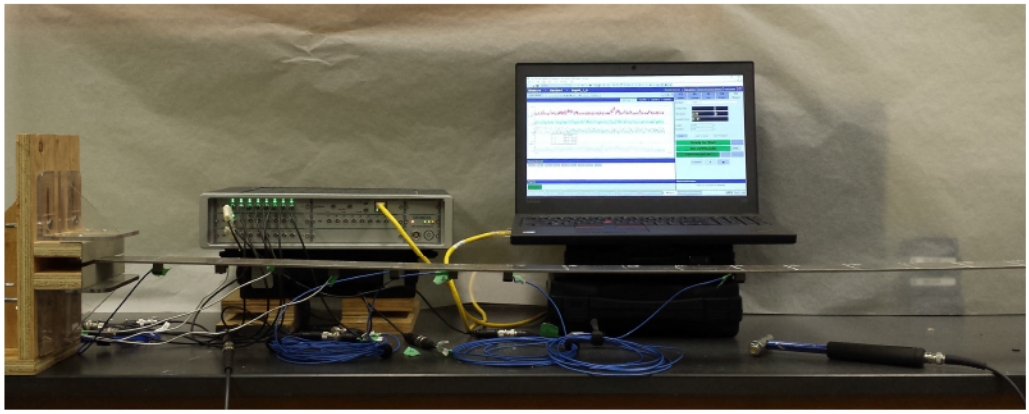


Figure 3.14: Experimental set-up in the laboratory.

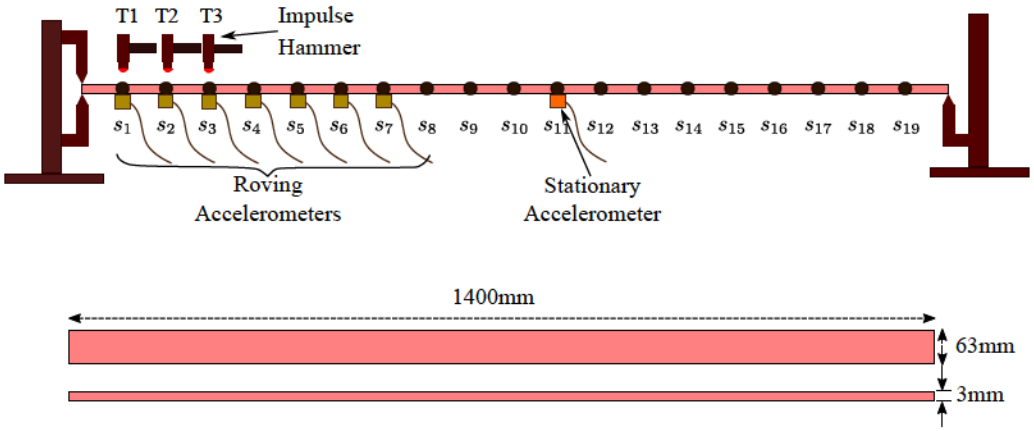


Figure 3.15: Schematic of laboratory tests.

An impulse hammer (Dytran Dynapulse model 5800B4) with soft plastic impact tip (model 6250PS) is used to provide pulse excitations as inputs. To compare the performance of the proposed (single sensor) approach against traditional identification with multiple sensors, the beam is instrumented using eight uniaxial piezoelectric accelerometers (PCB Piezotronics Model 333B40) which measure the vertical accelerations of the beam. One accelerometer is kept stationary at station s_{11} for all tests, while the others are roved to cover all the 19 measurement stations. A total of 19 tests are performed by sequentially

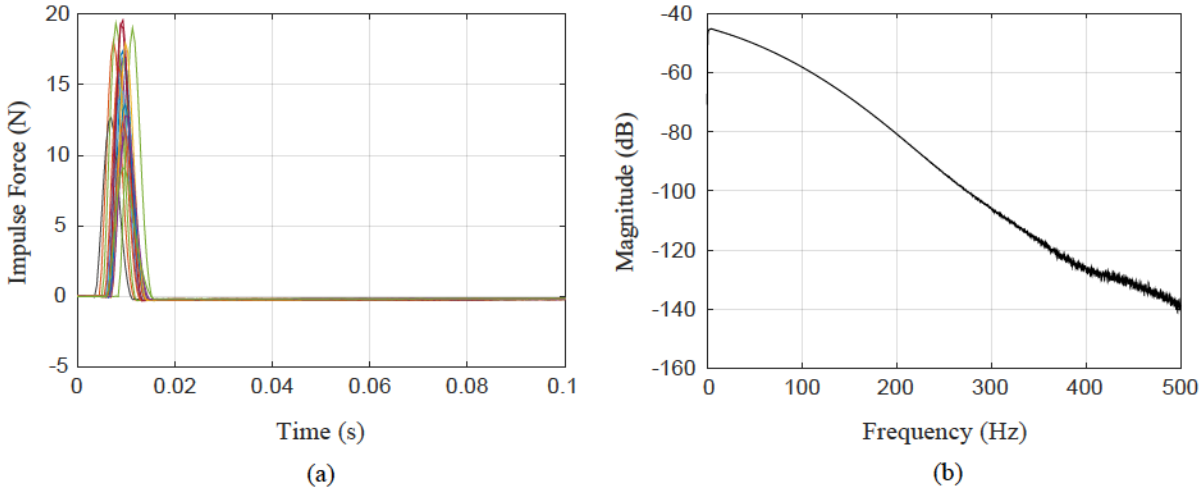


Figure 3.16: Inputs from impulse hammer in experiments: (a) Time histories, and (b) Averaged Fourier spectra.

exciting the beam at stations $s_1$to s_{19}; these tests are indicated as T1, T2, etc. in Figure 3.15. The mobile accelerometers are mounted at stations $s_1 - s_7$ during tests T1–T7, at $s_8 - s_{15}$ during tests T8–T15, and finally at $s_{16} - s_{19}$ during tests T16–T19.

The NCAS data consists of 19 sets of single input – single output measurements: the 19 inputs in tests T1–T19, and the corresponding 19 outputs at s_{11}. The CAS data also consists of 19 sets of single input – single output measurements: the 19 inputs in tests T1–T19, and corresponding outputs measured at the same stations as the inputs. For the reference modal identification, the data consists of three sets of single input – multi-output measurements: the 3 inputs in tests {T4, T9, T16}, and all corresponding measured outputs. In each test, the data is collected for a duration of 10s, with a sampling frequency of 1000 Hz. Figure 3.16 shows the initial 0.1s time histories of the impulsive inputs applied in all 19 tests, as well as the averaged (over all 19 tests) Fourier amplitude spectra of these inputs. Sample acceleration outputs used in the NCAS and CAS cases are shown in Figure 3.17. These accelerations are recorded in test T5, and suggest a very low damping for the system. The averaged Fourier amplitude spectra of all output accelerations used in the NCAS and CAS cases are shown in Figure 3.18.

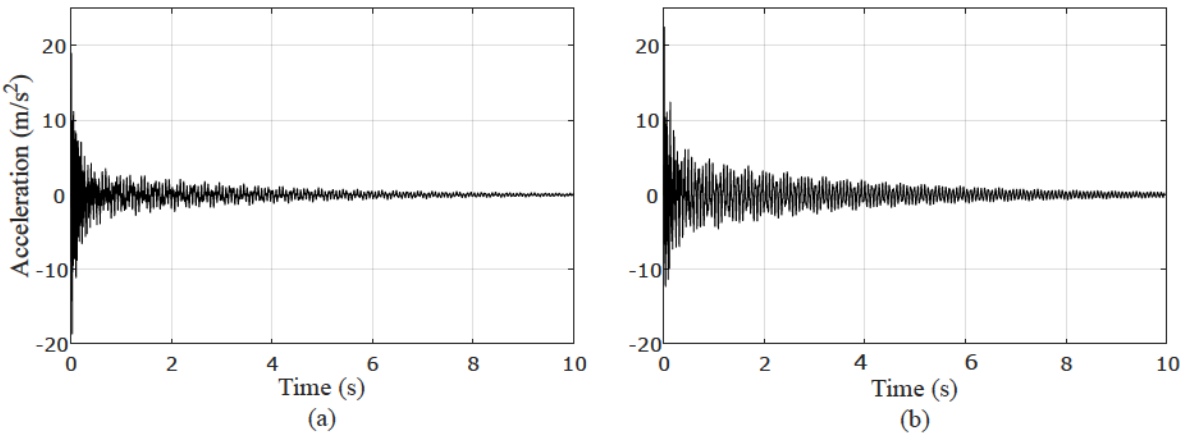


Figure 3.17: Sample acceleration outputs obtained in test T5: (a) at station s_{11} – used in the NCAS case, and (b) at station s_5 – used in the CAS case.

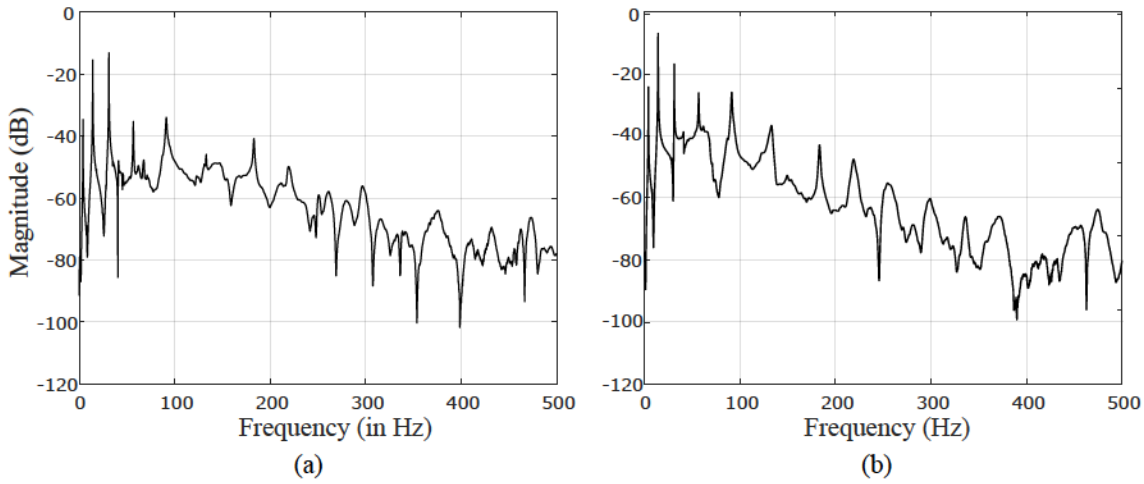


Figure 3.18: Averaged Fourier spectra of measured acceleration outputs: (a) at station s_{11} for NCAS case, and (b) at stations s_1 to s_{19} for CAS case.

The first eight modes of the beam are clearly visible from the spectra in Figure 3.18. Based on these spectra, the measured input and output data are filtered using a low pass 8th order Butterworth filter with a cutoff frequency of 250Hz, so as to reduce the effect of high frequency noise. The filtered data are then used in ERA-OKID to identify the modal frequencies and damping ratios, in both NCAS and CAS cases. The stable modes are distinguished using stabilization diagrams with averaged accelerance FRFs, and the criteria of 1% variation for frequencies and 5% variation for damping ratios. The stabilization diagram in case of NCAS measurements is shown as an example in Figure 3.19. For the reference modal identification with multiple sensors, ERA-OKID assisted with stabilization diagram is used to identify the modal frequencies, damping ratios, as well as the mode shapes, from the filtered data. In all three cases, the first eight modal frequencies and damping ratios could be identified. The means and standard deviations of these identified modal parameters for the reference case (using multiple sensors) are listed in Table 3.6 and that for NCAS and CAS cases are listed in Table 3.7. These mean and standard deviation values are computed from the individual values identified in all the tests for each case: 19 tests in case of NCAS and CAS, and 3 tests in case of the reference identification. It is evident from Table 3.7 that the identified modal parameters in both the NCAS and CAS cases agree reasonably well with the corresponding reference identified parameters in Table 3.6. As expected from the example output time histories in Figure 3.17, the identified damping values are mostly very low. It can also be seen that the uncertainty in the frequency estimates are quite low in all cases for all modes, while the uncertainty in the damping ratio estimates are higher. This is often encountered in modal identification. However, none of the damping ratio estimates appear to be unrealistic.

After identifying the modal frequencies and damping ratios, the mode shapes are identified for the NCAS and CAS cases, using both Approaches I and II. For assigning the signs of the mode shape components in case of CAS, the FE model of a simply supported uniform beam, with length $L = 1.4$m, Young's modulus $E = 220 \times 10^9 \text{N/m}^2$, moment of inertia $I = 1.42 \times 10^{-10}\text{m}^4$, and mass density $\rho = 8095\text{kg/m}^3$, is used; the FE model is obtained by discretizing the beam into 20 Euler-Bernoulli beam elements. The mode shapes identified in both NCAS and CAS cases are compared with the corresponding reference identified mode shapes graphically in Figure 3.21 and through the MAC values in Table

Table 3.6: Natural frequencies and damping ratios identified from experimental data using multiple sensors (Reference case).

Mode	f_{exp} (Hz)		ξ_{exp} (%)	
	Mean	SD	Mean	SD
1	3.99	0.02	1.01	0.25
2	14.27	0.10	0.58	0.13
3	31.03	0.08	0.35	0.07
4	57.10	0.85	0.80	0.33
5	91.30	0.66	0.80	0.33
6	132.89	1.06	1.52	0.86
7	183.11	1.45	0.68	0.35
8	219.52	1.89	1.31	0.73

Table 3.7: Natural frequencies and damping ratios identified from experimental data using NCAS and CAS cases.

Mode	NCAS				CAS			
	f_{exp} (in Hz)		ξ_{exp} (in %)		f_{exp} (Hz)		ξ_{exp} (%)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	3.99	0.02	0.96	0.18	3.99	0.02	0.84	0.20
2	14.23	0.12	0.43	0.21	14.22	0.16	0.48	0.16
3	31.02	0.13	0.41	0.11	31.08	0.14	0.39	0.04
4	56.61	1.08	0.70	0.68	56.49	1.13	1.43	1.17
5	91.31	0.90	0.75	0.27	90.90	1.11	0.67	0.32
6	132.80	1.23	1.36	1.08	132.63	1.17	1.18	0.96
7	182.83	2.07	0.57	0.42	182.99	1.76	0.73	0.43
8	219.34	1.94	2.21	1.33	219.10	2.13	3.07	1.59

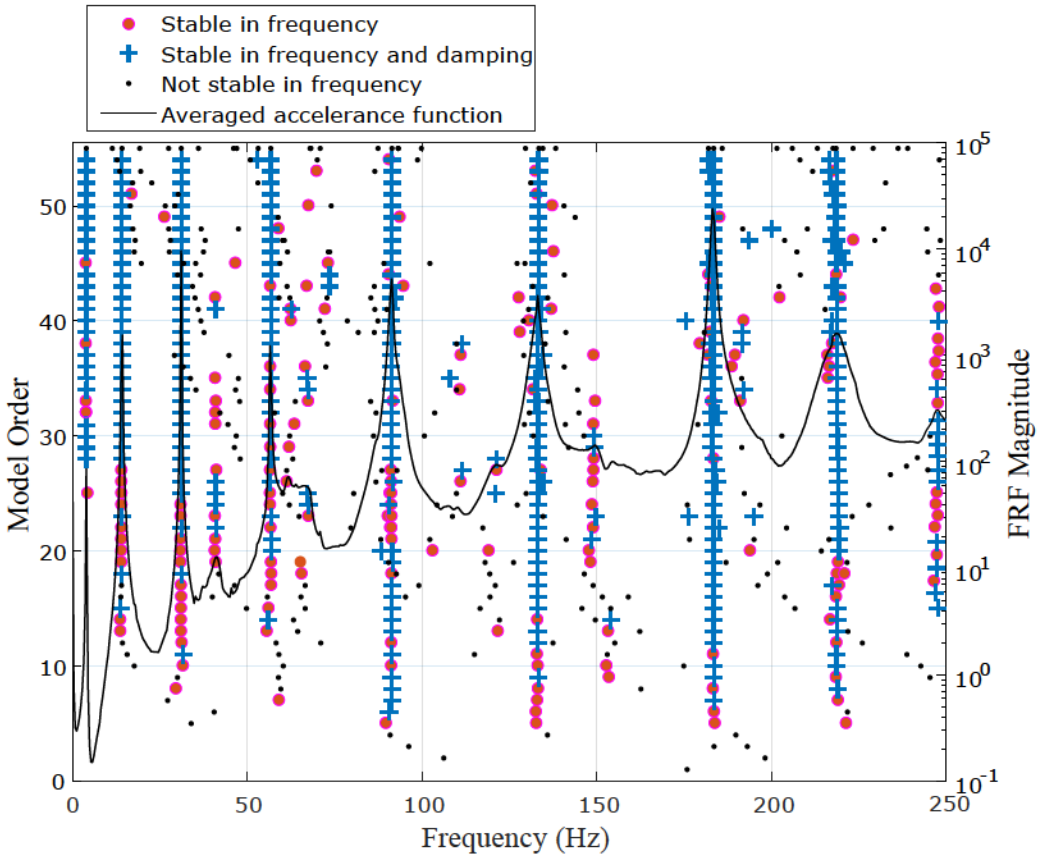


Figure 3.19: Stabilization diagram obtained in case of NCAS input-output data.

3.8.

In NCAS, the eighth mode shape could not be identified as, for this mode, the square of the mode shape component at s_{11}, obtained in the test with collocated input at s_{11}, is estimated as negative. This negative estimate is possibly owing to the location s_{11} being close to a node for Mode 8, as marked in Figure 3.21, resulting in an output response with very low signal-to-noise ratio (SNR). Although s_{11} is similarly next to a node for Modes 2, 4 and 6, this problem is not encountered in case of these modes. This may be due to Mode 8 having higher damping relative to the other modes, resulting in a faster decay of the modal response. Further, it is expected that Mode 8 is more affected by high frequency noise compared to the other modes, especially with the frequency of Mode 8 (≈ 219 Hz) being close to the cutoff frequency of 250 Hz. In fact, as can be seen from the the accelerance FRF plot in Figure 3.20a, the energy in Mode 8 is much lesser than in the other modes for the collocated input-output data at s_{11}.

The issue of negative estimates of the square of mode shape components is also encountered in the case of CAS, for all Modes 4 to 8. The components affected by this problem could not be estimated, and are marked with a green $\times$ in Figure 3.21. However, unlike the NCAS case, the other mode shape components can be estimated in the case of CAS, as the estimation of different components are not linked in this case. When computing the MAC values in Table 3.8, the components affected by this problem have been excluded. From Figure 3.21, it can again be seen that locations close to nodes are the most prone to this problem. However, in case of Mode 8, many non-nodal locations are also affected; for this mode, the estimation is very poor even in locations not affected by this problem. This is again possibly owing to the cumulative effect of a higher damping as well as higher effect of noise in case of Mode 8. As can be observed from Figure 3.20(b), for the collocated input-output data at s_4, the energy associated with Modes 5 and 8 are lesser than the other modes. These are the two modes for which the problem of negative estimates of squared mode shape components is encountered at location s_4 (Figure 3.21), illustrating that this problem may mostly arise in situations when a mode is not sufficiently excited at a collocated input-output location.

Although the mode shape for Mode 8 could not be estimated in both NCAS and CAS cases, as evident from both Figure 3.21 as well as the MAC values in Table 3.8, the

estimates of all the remaining mode shapes agree reasonably well with the corresponding reference identified mode shapes. It is also seen that, while for the lower modes both Approaches I and II perform equally well, for the higher modes Approach II sometimes tend to overestimate the mode shape components.

Table 3.8: MAC values comparing mode shapes identified using proposed approach with identified reference mode shapes.

Mode	NCAS		CAS	
	Approach I	Approach II	Approach I	Approach II
1	0.999	0.999	0.996	0.999
2	0.998	0.998	0.996	0.996
3	0.992	0.992	0.985	0.989
4	0.978	0.979	0.983	0.981
5	0.977	0.981	0.983	0.986
6	0.966	0.964	0.967	0.947
7	0.967	0.949	0.969	0.950
8	–	–	0.662	0.638

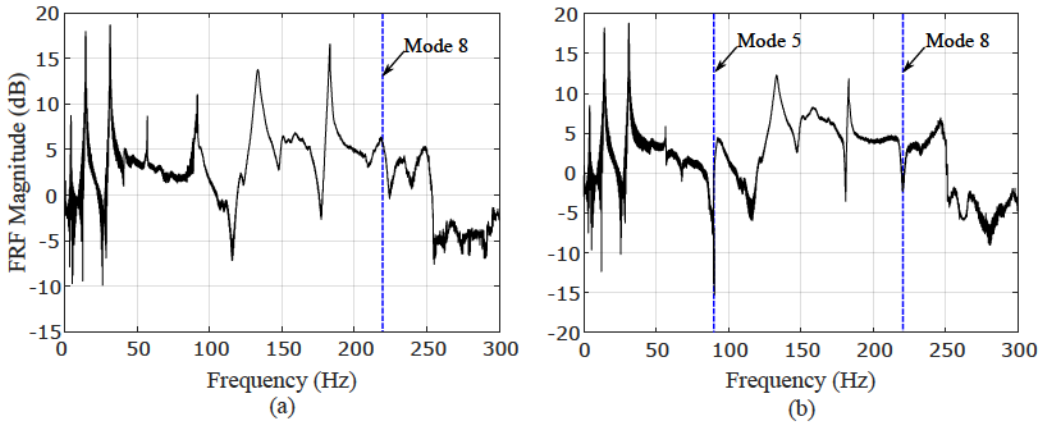


Figure 3.20: Accelerance FRFs for collocated input-output measurements: (a) at station s_{11}, and (b) at station s_4.

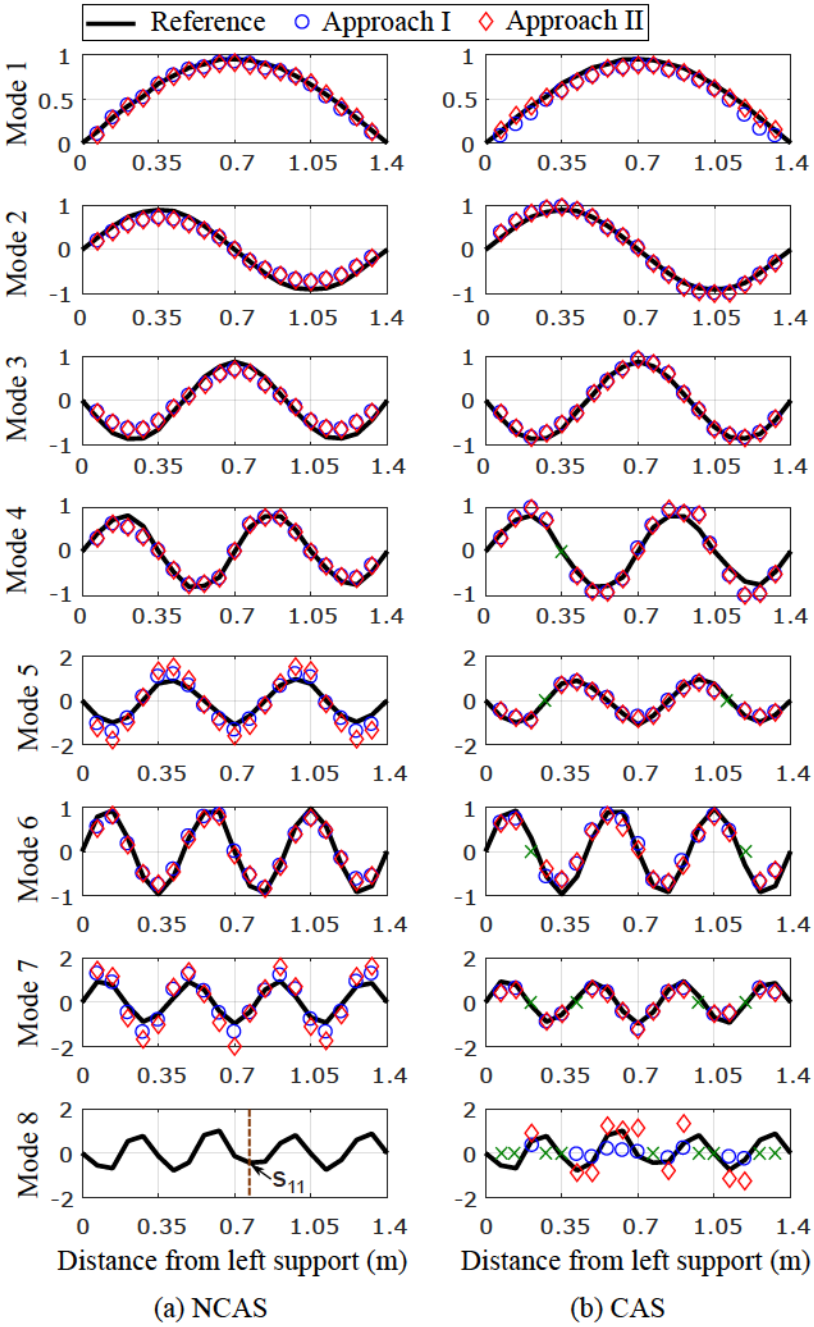


Figure 3.21: Mode shapes identified from experimental data in case of: (a) NCAS data and (b) CAS data.

3.5 Summary

In this chapter, the feasibility of using a single re-configurable pair of actuator and sensor for modal identification of bridge decks is studied. The proposed identification strategy involves performing a suite of tests, with either: (a) a static actuator - mobile sensor or a mobile actuator - static sensor (NCAS case), or (b) a combined mobile actuator and sensor system (CAS case). Of particular interest is the case of a combined mobile actuator and sensor which resembles a mobile device fitted with an actuator and a sensor.

The suite of tests involve sequentially repositioning the mobile actuator/sensor system to various locations on the structure. At each location the actuator excites the structure while the sensor measures the response. The output data from the tests is first used to identify the modal frequencies and damping ratios, using any conventional system identification method such as ERA-OKID. The identified frequencies and damping ratios, along with the measured input-output data, are then used to identify the the mass normalized mode shape components. A time-domain technique, based on input-output balance using pseudo-modal responses, was developed for estimating the mode shapes. It was shown that when the actuator and sensor are always collocated (CAS case), which corresponds to case of using a single mobile actuator-sensor pair, the mode shapes components can not be directly obtained. Instead only their absolute values can be extracted and an approximate knowledge of the signs of mode shapes is required for mode shape estimation. However, no such prior information is needed for the NCAS case. Based on the numerical and experimental results, the proposed strategy may be considered as a potentially feasible alternative to traditional static-sensing schemes for modal identification.

The proposed strategy however could be prohibitive in practice due to two reasons: (a) it requires the use of an actuator which may not be available, and (b) the testing strategy is time-consuming and may require closing down of the structure. These drawbacks can be avoided by an output-only testing strategy with a network of in-motion mobile sensors. The next two chapters will focus on the development of modal identification algorithms for use with output-only vibration data from in-motion mobile sensors.

Chapter 4

Bayesian output-only modal identification using in-motion mobile sensors: Missing data approach

4.1 Introduction

In the previous chapter, a reconfigurable mobile actuation-sensing strategy was employed for input-output based modal identification which required pausing at several locations on the structure to apply excitation and collect data. In contrast to this, output-only modal identification with in-motion mobile sensors has the advantages that: (1) it does not need the explicit input information (hence the name output-only) and (2) the mobile sensors do not need to pause to collect data allowing greater spatial coverage in a reduced amount of time. Along these lines, this chapter introduces the use of *in-motion* mobile sensors for *output-only* modal identification. However, as mentioned previously in Section 2.1.2, the datasets collected using in-motion mobile sensors are characterized by spatial discontinuities which prevents the direct application of conventional SSID algorithms for modal identification.

In this chapter, the problem of modal identification is posed from an incomplete data

perspective, that is, the data matrix of structural responses collected using in-motion sensors is treated equivalent to a static sensor data matrix with missing entries – referred to as a missing data matrix. A time-domain state space modelling approach is adopted for modal parameter estimation from the missing data matrix. This involves fitting a parametric SSM to the missing data matrix and then extracting the modal parameters from the fitted SSM parameters. The state space parameter identification is handled in a Bayesian framework which provides a principled approach towards parameter estimation and uncertainty quantification; the uncertainty is caused not only due to variability in the observed data but also due to missing information inherent in the approach presented in this study.

This chapter is organized as follows: first, the mechanism of constructing the missing data matrix is described; next, a stochastic linear SSM is introduced with unknown parameters which are to be identified using the missing data matrix; and finally, three computational algorithms – EM, VB and GS – are presented for SSM parameter identification with from missing data matrices, among which VB and GS enable uncertainty quantification following Bayesian principles.

4.2 Construction of the missing data matrix

Static sensor networks capture structural responses at a fixed set of spatial points for the entire test duration. This sensing method yields a data matrix that is *complete* in the sense that an entire time-history of the vibration response is available for all predetermined spatial locations. By contrast, in-motion mobile sensors record structural responses while simultaneously moving through the space of the structure. Considering the same set of spatial locations, the continuous movement of sensors from one location to another removes the ability to capture the entire time history of vibration responses at any given location. As such, treating the mobile sensor data in a similar fashion to data collected from static sensor networks (where each column represents a time history for a given location) results in a data matrix with many unrecorded or *missing* entries. This matrix, referred to as a *missing data matrix*, is constructed by inputting the measured data at the corresponding

time and spatial indices.

Figure 4.1 illustrates the differences between a complete data matrix and missing data matrix obtained from a static sensor network and single mobile sensor respectively. In both cases, three sensing nodes are considered and the path of the mobile sensor is indicated by an arrow.

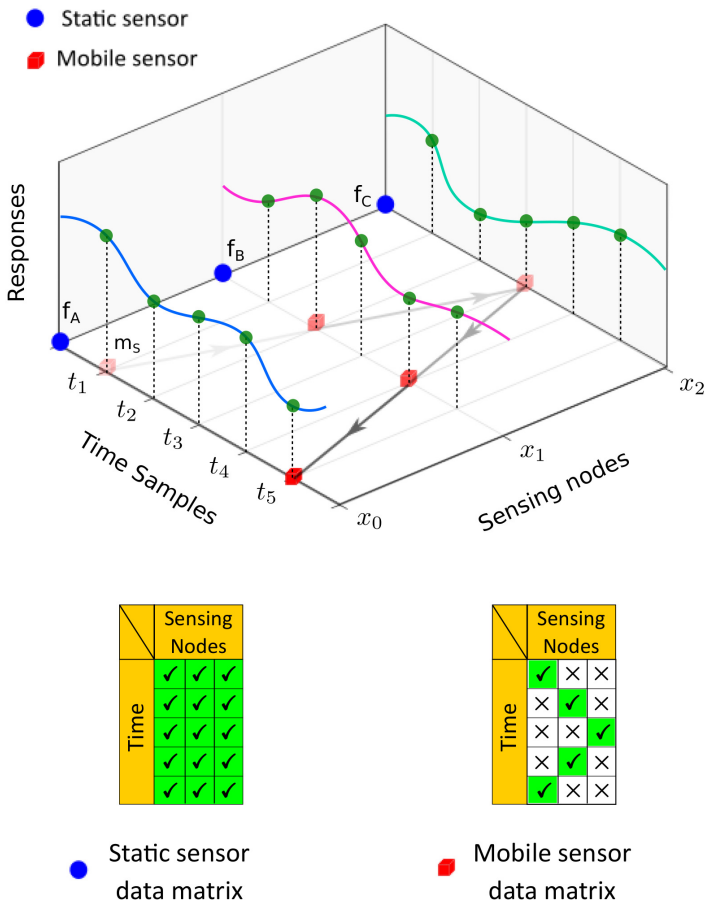


Figure 4.1: An example of data collection with both static sensors and in-motion mobile sensors. Three static sensors generate a complete data matrix whereas one in-motion mobile sensor generates a missing data matrix, as at any time instant, the mobile sensor is able to sample only one out of three spatial locations.

Despite having the advantage of improved spatial coverage and reduced testing time, the main limitation associated with in-motion mobile sensing is the collected data, when treated as a missing data matrix, is incompatible for processing with conventional system identification algorithms. That is, the widely-used output-only (or operational or stochastic) structural system identification algorithms such as ERA-OKID-OO [100], ERA-NEXT [36], AR [101], SSI [24], N4SID [102], cannot be readily used to with data containing missing observations. In this study, banking on the state space modelling framework, two novel Bayesian techniques (i.e., VB and GS) are introduced and modified appropriately to handle Bayesian inference with missing data.

4.3 Linear Gaussian state space model

The discrete-time linear SSM in Eq.(2.17) is considered once again:

$$\boldsymbol{x}_{k+1} = \mathbf{A}\boldsymbol{x}_k + \boldsymbol{w}_k \tag{4.1a}$$

$$\boldsymbol{y}_k = \mathbf{G}\boldsymbol{x}_k + \boldsymbol{v}_k. \tag{4.1b}$$

The dimensions of the variables are, $\boldsymbol{x}_k \in \mathbb{R}^{n_s}$, $\boldsymbol{y}_k \in \mathbb{R}^{n_o}$, $\mathbf{A} \in \mathbb{R}^{n_s \times n_s}$, $\mathbf{G} \in \mathbb{R}^{n_o \times n_s}$, $\boldsymbol{w}_k \in \mathbb{R}^{n_s}$ and $\boldsymbol{v}_k \in \mathbb{R}^{n_o}$. The modal parameters of the structure are related to the parameters of the SSM through an eigen-transformation as outlined in Appendix B.1. Eq.(4.1) specifies a SSM with a multivariate joint Gaussian distribution defined over the latent state variables $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{N+1}\}$ and the observed variables $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N\}$, and is henceforth referred to as the linear Gaussian state space model (LGSSM). Specifically, due to the Gaussian properties of $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ (i.e. $\boldsymbol{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ and $\boldsymbol{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$), Eq.(4.1) leads to the following conditional probability distributions over state and observed variables $\boldsymbol{x}_k$ and $\boldsymbol{y}_k$:

$$p(\boldsymbol{x}_{k+1} \mid \boldsymbol{x}_k, \mathbf{A}, \mathbf{Q}) = \mathcal{N}(\mathbf{A}\boldsymbol{x}_k, \mathbf{Q}), \tag{4.2a}$$

$$p(\boldsymbol{y}_k \mid \boldsymbol{x}_k, \mathbf{G}, \mathbf{R}) = \mathcal{N}(\mathbf{G}\boldsymbol{x}_k, \mathbf{R}), \tag{4.2b}$$

$$p(\boldsymbol{x}_1 \mid \boldsymbol{\mu}_1, \mathbf{V}_1) = \mathcal{N}(\boldsymbol{\mu}_1, \mathbf{V}_1), \tag{4.2c}$$

and the joint distribution over the sequence of states and observations is expressed as

$$p\left(\boldsymbol{X}, \boldsymbol{Y} \mid \mathbf{A}, \mathbf{G}, \mathbf{Q}, \mathbf{R}, \boldsymbol{\mu}_1, \mathbf{V}_1\right) = p\left(\boldsymbol{x}_1 \mid \boldsymbol{\mu}_1, \mathbf{V}_1\right) \prod_{k=1}^{N} p\left(\boldsymbol{x}_{k+1} \mid \boldsymbol{x}_k, \mathbf{A}, \mathbf{Q}\right) p\left(\boldsymbol{y}_k \mid \boldsymbol{x}_k, \mathbf{G}, \mathbf{R}\right) \tag{4.3}$$

where $\boldsymbol{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{N+1}\}$ and $\boldsymbol{Y} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N\}$ and $\mathcal{N}(\cdot\,,\,\cdot)$ denotes multivariate joint Gaussian distribution specified by its arguments (i.e. mean and covariance). Eq.(4.3) states that the joint distribution over all variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ given the parameters $\boldsymbol{\theta} = \{\mathbf{A}, \mathbf{G}, \mathbf{Q}, \mathbf{R}\}$ (with auxiliary parameters $\boldsymbol{\mu}_1, \mathbf{V}_1$) of the LGSSM is given by the product of Gaussian marginal and conditional distributions over $\boldsymbol{X}$ and $\boldsymbol{Y}$ respectively.

4.3.1 Notations for observed and missing data

Before commencing, a few notations are introduced for the observed and missing elements of the data. The $n_o \times 1$ vector $\boldsymbol{y}_k$ denotes the observations at time k. If some elements of the $\boldsymbol{y}_k$ are missing, the missing elements are set to zero:

$$\boldsymbol{y}_k = \begin{bmatrix} y_1 \\ 0 \\ y_3 \\ y_4 \\ 0 \end{bmatrix}. \tag{4.4}$$

Denote the non-missing observations as $\boldsymbol{y}_k^{obs}$ and the missing observations as $\boldsymbol{y}_k^{miss}$. Also, define matrix $\boldsymbol{\upsilon}_k^{(1)}$ which extracts only $\boldsymbol{y}_k^{obs}$ from $\boldsymbol{y}_k$, and matrix $\boldsymbol{\upsilon}_k^{(2)}$ which extracts only $\boldsymbol{y}_k^{miss}$ from $\boldsymbol{y}_k$:

$$\begin{aligned} \boldsymbol{y}_k^{obs} &= \boldsymbol{\upsilon}_k^{(1)} \boldsymbol{y}_k, \quad \boldsymbol{\upsilon}_k^{(1)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \\ \boldsymbol{y}_k^{miss} &= \boldsymbol{\upsilon}_k^{(2)} \boldsymbol{y}_k, \quad \boldsymbol{\upsilon}_k^{(2)} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \end{aligned} \tag{4.5}$$

Define another set of matrices that zeros out the missing and non-missing observations. Let $\mathbf{I}_k^{(1)}$ denote the diagonal matrix that zeros out the missing entries in $\boldsymbol{y}_k$ and $\mathbf{I}_k^{(2)}$ denote a matrix that zeros out the observed entries in $\boldsymbol{y}_k$. Considering the above example, the matrices $\mathbf{I}_k^{(1)}$ and $\mathbf{I}_k^{(2)}$ would be

$$\mathbf{I}_k^{(1)} = \mathfrak{V}_k^{(1)^T}\mathfrak{V}_k^{(1)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{I}_k^{(2)} = \mathfrak{V}_k^{(2)^T}\mathfrak{V}_k^{(2)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{4.6}$$

The matrices defined above will be used for succinctly representing relations involving missing data.

4.3.2 Convention for parameters

It is also helpful to establish some linguistic conventions with respect to variables involved in a Bayesian treatment of Eq.(4.3). In a maximum likelihood (ML) setting, i.e. the case where $\boldsymbol{\theta}$ is not governed by probability distributions, $\boldsymbol{Y}$ is classically referred to as "observed variables" and $\boldsymbol{X}$ are referred to as "latent" (or "unobserved" or 'hidden") variables, and $\boldsymbol{\theta}$ is referred to as "parameter" [64]. In a Bayesian setting, the quantity $\boldsymbol{\theta}$ itself becomes governed by probability distributions, and hence the parameters that one is interested in are the parameters governing the distribution of $\boldsymbol{\theta}$ which will be referred to as the "hyperparameters" $\boldsymbol{\theta}_h$ in this chapter.

Given the definition of LGSSM and the conventions for missing observations in measurements, three computational algorithms for inference with missing data, namely the EM, VB and GS, will be presented next.

4.4 ML estimation of LGSSM via EM

This section studies the use of the EM algorithm for computing ML estimates of modal parameters from missing data matrices. In this setting, the model parameters are treated as deterministic (non-random). The objective is to solve for the parameters $\boldsymbol{\theta} = \{\mathbf{A}, \mathbf{G}, \mathbf{Q}, \mathbf{R}\}$ that maximizes the likelihood of the LGSSM given the incomplete data.

In the case of no missing information, the complete-data log-likelihood of the LGSSM would be given by

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) =& \ln\left(p\left(\boldsymbol{X}, \boldsymbol{Y} \mid \boldsymbol{\theta}\right)\right) \\
=& \ln\left(p\left(\boldsymbol{x}_1 \mid \boldsymbol{\mu}_1, \mathbf{V}_1\right) \prod_{k=1}^{N} p\left(\boldsymbol{x}_{k+1} \mid \boldsymbol{x}_k, \mathbf{A}, \mathbf{Q}\right) \prod_{k=1}^{N} p\left(\boldsymbol{y}_k \mid \boldsymbol{x}_k, \mathbf{G}, \mathbf{R}\right)\right) \\
=& -\frac{n_s}{2}\ln 2\pi - \frac{1}{2}\ln|\mathbf{V}_1| - \frac{1}{2}\left(\boldsymbol{x}_1 - \boldsymbol{\mu}_1\right)^T \mathbf{V}_1^{-1}\left(\boldsymbol{x}_1 - \boldsymbol{\mu}_1\right) \\
& - \frac{Nn_s}{2}\ln 2\pi - \frac{N}{2}\ln|\mathbf{Q}| - \frac{1}{2}\sum_{k=1}^{N}\left(\boldsymbol{x}_{k+1} - \mathbf{A}\boldsymbol{x}_k\right)^T \mathbf{Q}^{-1}\left(\boldsymbol{x}_{k+1} - \mathbf{A}\boldsymbol{x}_k\right) \\
& - \frac{Nn_o}{2}\ln 2\pi - \frac{N}{2}\ln|\mathbf{R}| - \frac{1}{2}\sum_{k=1}^{N}\left(\boldsymbol{y}_k - \mathbf{G}\boldsymbol{x}_k\right)^T \mathbf{R}^{-1}\left(\boldsymbol{y}_k - \mathbf{G}\boldsymbol{x}_k\right)
\end{aligned} \tag{4.7}$$

and the goal to obtain the ML estimate, $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$, is achieved by maximizing $\mathcal{L}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. However, due to the missing entries in $\boldsymbol{Y}$ and unavailability of state sequence $\boldsymbol{X}$, a direct maximization of $\mathcal{L}(\boldsymbol{\theta})$ given incomplete data becomes problematic. The EM algorithm solves this problem by iteratively alternating between making guesses about the latent states $\boldsymbol{X}$ and missing entries in $\boldsymbol{Y}$, and finding the $\boldsymbol{\theta}$ that maximizes $p\left(\boldsymbol{Y}, \boldsymbol{X} \mid \boldsymbol{\theta}\right)$. Algorithmically, EM starts from an initial point $\boldsymbol{\theta}^{(0)} = \left\{\mathbf{A}^{(0)}, \mathbf{G}^{(0)}, \mathbf{Q}^{(0)}, \mathbf{R}^{(0)}\right\}$ and repeats the following two steps iteratively until the likelihood converges to a stationary value:

1. **E-step**: Compute the conditional expectation $\mathcal{F}\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(j)}\right) := \left\langle \mathcal{L}(\boldsymbol{\theta}) \mid \boldsymbol{Y}, \boldsymbol{\theta}^{(j)} \right\rangle$,
2. **M-step**: Obtain $\boldsymbol{\theta}^{(j+1)}$ by maximizing $\mathcal{F}\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(j)}\right)$.

Here, $\langle\cdot\rangle$ denotes the expectation operator, and the superscript (j) denotes the jth iteration of EM. Since direct maximization of $\mathcal{L}(\boldsymbol{\theta})$ is difficult, the EM guesses the expected value

$\mathcal{F}\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(j)}\right)$ and maximizes it instead in each iteration. It has been shown in literature that the convergence of EM is monotonic [63]. For the LGSSM in Eq.(4.1), the required expectations in the E-step are easily evaluated using the Kalman filter and the Rauch-Tung-Striebel (RTS) smoother, and the maximization in the M-step can be obtained in closed form using the results from the smoother.

4.4.1 Derivation of the E-step

The EM algorithm evaluates the conditional expectation $\mathcal{F}\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(j)}\right)$ which can be expressed as

$$\begin{aligned}\mathcal{F}\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(j)}\right) = &-\frac{1}{2}\ln|\mathbf{V}_1| - \frac{1}{2}\mathrm{tr}\left\{\mathbf{V}_1{}^{-1}\left\langle(\boldsymbol{x}_1-\boldsymbol{\mu}_1)(\boldsymbol{x}_1-\boldsymbol{\mu}_1)^T\right\rangle\right\}\\ &-\frac{N}{2}\ln|\mathbf{Q}| - \frac{1}{2}\sum_{k=1}^{N}\mathrm{tr}\left\{\mathbf{Q}^{-1}\left\langle(\boldsymbol{x}_{k+1}-\mathbf{A}\boldsymbol{x}_k)(\boldsymbol{x}_{k+1}-\mathbf{A}\boldsymbol{x}_k)^T\right\rangle\right\}\\ &-\frac{N}{2}\ln|\mathbf{R}| - \frac{1}{2}\sum_{k=1}^{N}\mathrm{tr}\left\{\mathbf{R}^{-1}\left\langle(\boldsymbol{y}_k-\mathbf{G}\boldsymbol{x}_k)(\boldsymbol{y}_k-\mathbf{G}\boldsymbol{x}_k)^T\right\rangle\right\} + \text{constant}.\end{aligned}$$

On further simplification, $\mathcal{F}\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(j)}\right)$ can be written as

$$\begin{aligned}\mathcal{F}\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(j)}\right) = &-\frac{1}{2}\ln|\mathbf{V}_1| - \frac{1}{2}\mathrm{tr}\left\{\mathbf{V}_1{}^{-1}\left[\hat{\mathbf{V}}_{1|N} + \left(\hat{\boldsymbol{x}}_{1|N}-\boldsymbol{\mu}_1\right)\left(\hat{\boldsymbol{x}}_{1|N}-\boldsymbol{\mu}_1\right)^T\right]\right\}\\ &-\frac{N}{2}\ln|\mathbf{Q}| - \frac{1}{2}\mathrm{tr}\left\{\mathbf{Q}^{-1}\left[\mathbf{S}_{ff} - \mathbf{A}\mathbf{S}_{fx}^T - \mathbf{S}_{fx}\mathbf{A}^T + \mathbf{A}\mathbf{S}_{xx}\mathbf{A}^T\right]\right\}\\ &-\frac{N}{2}\ln|\mathbf{R}| - \frac{1}{2}\mathrm{tr}\left\{\mathbf{R}^{-1}\left[\mathbf{S}_{yy} - \mathbf{G}\mathbf{S}_{yx}^T - \mathbf{S}_{yx}\mathbf{G}^T + \mathbf{G}\mathbf{S}_{xx}\mathbf{G}^T\right]\right\},\end{aligned} \tag{4.8}$$

where the variables

$$\hat{\boldsymbol{x}}_{k|N} = \left\langle \boldsymbol{x}_k \mid \boldsymbol{Y}, \boldsymbol{\theta}^{(j)}\right\rangle, \tag{4.9a}$$

$$\hat{\mathbf{V}}_{k|N} = \mathrm{COV}(\boldsymbol{x}_k) = \left\langle \left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|N}\right)\left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|N}\right)^T \mid \boldsymbol{Y}, \boldsymbol{\theta}^{(j)}\right\rangle, \tag{4.9b}$$

$$\hat{\mathbf{V}}_{k+1,k|N} = \mathrm{COV}(\boldsymbol{x}_{k+1}, \boldsymbol{x}_k) = \left\langle \left(\boldsymbol{x}_{k+1} - \hat{\boldsymbol{x}}_{k+1|N}\right)\left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|N}\right)^T \mid \boldsymbol{Y}, \boldsymbol{\theta}^{(j)}\right\rangle \tag{4.9c}$$

are outputs obtained from a slightly modified Kalman filter and RTS smoother.

The terms $\mathbf{S}_{ff}$, $\mathbf{S}_{fx}$, $\mathbf{S}_{xx}$, $\mathbf{S}_{yx}$ and $\mathbf{S}_{yy}$ in Eq.(4.8) are referred to as the *expectations of sufficient statistics* and represent the following expectations:

$$\begin{aligned} \mathbf{S}_{ff} = \sum_{k=1}^{N} \left\langle \boldsymbol{x}_{k+1}\boldsymbol{x}_{k+1}^T \right\rangle, \quad \mathbf{S}_{fx} = \sum_{k=1}^{N} \left\langle \boldsymbol{x}_{k+1}\boldsymbol{x}_{k}^T \right\rangle \quad \mathbf{S}_{xx} = \sum_{k=1}^{N} \left\langle \boldsymbol{x}_{k}\boldsymbol{x}_{k}^T \right\rangle \\ \mathbf{S}_{yy} = \sum_{k=1}^{N} \left\langle \boldsymbol{y}_{k}\boldsymbol{y}_{k}^T \right\rangle, \quad \mathbf{S}_{yx} = \sum_{k=1}^{N} \left\langle \boldsymbol{y}_{k}\boldsymbol{x}_{k}^T \right\rangle. \end{aligned} \tag{4.10}$$

$\mathbf{S}_{ff}$, $\mathbf{S}_{fx}$, $\mathbf{S}_{xx}$, involve only the states and are evaluated using the estimated states from RTS smoother:

$$\mathbf{S}_{ff} = \sum_{k=1}^{N} \left\langle \boldsymbol{x}_{k+1}\boldsymbol{x}_{k+1}^T \right\rangle = \sum_{k=1}^{N} \left(\hat{\boldsymbol{x}}_{k+1|N}\hat{\boldsymbol{x}}_{k+1|N}^T + \hat{\mathbf{V}}_{k+1|N} \right), \tag{4.11a}$$

$$\mathbf{S}_{fx} = \sum_{k=1}^{N} \left\langle \boldsymbol{x}_{k+1}\boldsymbol{x}_{k}^T \right\rangle = \sum_{k=1}^{N} \left(\hat{\boldsymbol{x}}_{k+1|N}\hat{\boldsymbol{x}}_{k|N}^T + \hat{\mathbf{V}}_{k+1,k|N} \right), \tag{4.11b}$$

$$\mathbf{S}_{xx} = \sum_{k=1}^{N} \left\langle \boldsymbol{x}_{k}\boldsymbol{x}_{k}^T \right\rangle = \sum_{k=1}^{N} \left(\hat{\boldsymbol{x}}_{k|N}\hat{\boldsymbol{x}}_{k|N}^T + \hat{\mathbf{V}}_{k|N} \right), \tag{4.11c}$$

whereas $\mathbf{S}_{yx}$ and $\mathbf{S}_{yy}$ involve both states and missing observations and are computed as follows:

$$\mathbf{Z}_k = \mathbf{I}_{n_o \times n_o} - \mathbf{R}\mathcal{O}_k^{(1)^T} \left(\mathcal{O}_k^{(1)} \mathbf{R} \mathcal{O}_k^{(1)^T} \right)^{-1} \mathcal{O}_k^{(1)} \tag{4.12a}$$

$$\hat{\boldsymbol{y}}_k = \langle \boldsymbol{y}_k \rangle = \boldsymbol{y}_k - \mathbf{Z}_k \left(\boldsymbol{y}_k - \mathbf{G}\hat{\boldsymbol{x}}_{k|N} \right) \tag{4.12b}$$

$$\mathbf{S}_{yy} = \sum_{k=1}^{N} \left\langle \boldsymbol{y}_{k}\boldsymbol{y}_{k}^T \right\rangle = \sum_{k=1}^{N} \left(\mathbf{I}_k^{(2)} \mathbf{Z}_k \left(\mathbf{R} + \mathbf{G}\hat{\mathbf{V}}_{k|N}\mathbf{G}^T \right) \mathbf{Z}_k^T \mathbf{I}_k^{(2)} + \hat{\boldsymbol{y}}_k \hat{\boldsymbol{y}}_k^T \right) \tag{4.12c}$$

$$\mathbf{S}_{yx} = \sum_{k=1}^{N} \left\langle \boldsymbol{y}_{k}\boldsymbol{x}_{k}^T \right\rangle = \sum_{k=1}^{N} \left(\mathbf{Z}_k \mathbf{G} \hat{\mathbf{V}}_{k|N} + \hat{\boldsymbol{y}}_k \hat{\boldsymbol{x}}_{k|N}^T \right). \tag{4.12d}$$

These expectations of sufficient statistics are used in the M-step for deriving closed form expressions for maximizing the parameters. The detailed derivations in Eq.(4.12) can be found in Appendix C.

Given the parameters, the state estimation problem in the presence of complete data with no missing values can be handled using the classical Kalman filter (see Appendix F).

However, state estimation in the presence of missing observations calls for a modification to these state estimation tools. Shumway and Stoffer [64, 103] recommended modifications to the Kalman filter to enable unbiased filtering operation in the presence of missing observations. For running the modified Kalman filter with missing data, it is convenient to first compute a sequence of matrices $\left\{\mathbf{I}_k^{(1)}, \mathbf{I}_k^{(2)}\right\}_{1:N}$ following the notations introduced in Section 4.3.1. Then, the Kalman filter equations for obtaining predicted and filtered states and error covariance matrices from incomplete data can be written as:

Kalman filter with missing data : Do for $k = 1, \ldots, N$

$$\boldsymbol{y}_k^* = \mathbf{I}_k^{(1)} \boldsymbol{y}_k \tag{4.13a}$$
$$\mathbf{G}_k^* = \mathbf{I}_k^{(1)} \mathbf{G} \tag{4.13b}$$
$$\mathbf{R}_k^* = \mathbf{I}_k^{(1)} \mathbf{R} \mathbf{I}_k^{(1)} + \mathbf{I}_k^{(2)} \mathbf{I}_k^{(2)} \tag{4.13c}$$
$$\boldsymbol{e}_k^* = \boldsymbol{y}_k^* - \mathbf{G}_k^* \hat{\boldsymbol{x}}_{k|k-1} \tag{4.13d}$$
$$\mathbf{S}_k^* = \mathbf{G}_k^* \hat{\mathbf{V}}_{k|k-1} \mathbf{G}_k^{*T} + \mathbf{R}_k^* \tag{4.13e}$$
$$\mathbf{K}_k^* = \hat{\mathbf{V}}_{k|k-1} \mathbf{G}_k^{*T} \mathbf{S}_k^{*-1} \tag{4.13f}$$
$$\hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k-1} + \mathbf{K}_k^* \boldsymbol{e}_k^* \tag{4.13g}$$
$$\hat{\mathbf{V}}_{k|k} = \hat{\mathbf{V}}_{k|k-1} - \mathbf{K}_k^* \mathbf{S}_k^* \mathbf{K}_k^{*T} \tag{4.13h}$$
$$\hat{\boldsymbol{x}}_{k+1|k} = \mathbf{A} \hat{\boldsymbol{x}}_{k|k} \tag{4.13i}$$
$$\hat{\mathbf{V}}_{k+1|k} = \mathbf{A} \hat{\mathbf{V}}_{k|k} \mathbf{A} + \mathbf{Q}. \tag{4.13j}$$

Here, $\hat{\boldsymbol{x}}_{k|k-1}$ and $\hat{\boldsymbol{x}}_{k|k}$ represent the kth predicted and filtered state estimate respectively, and, $\hat{\mathbf{V}}_{k|k-1}$ and $\hat{\mathbf{V}}_{k|k}$ denote the kth predicted and filtered state error covariance matrices respectively. The Kalman filter recursion is started from an initial state estimate $\hat{\boldsymbol{x}}_{1|0}$ and an initial covariance estimate $\hat{\mathbf{V}}_{1|0}$. For computational purposes, it is more convenient to calculate the log-likelihood at the jth EM iteration using the innovations from the Kalman filter

$$\hat{\mathcal{F}}^{(j)} = -\frac{1}{2} \sum_{k=1}^{N} \left(\ln |\mathbf{S}_k^*| + \boldsymbol{e}_k^{*T} \mathbf{S}_k^{*-1} \boldsymbol{e}_k^* \right) \tag{4.14}$$

compared to calculating the actual log-likelihood by Eq.(4.8). Following the filtering step, the (fixed interval) smoothing recursions given by the RTS smoother are computed as

follows:

Kalman smoother : Do for $k = N, \ldots, 1$

$$\mathbf{N}_k = \hat{\mathbf{V}}_{k|k}\mathbf{A}^T \left(\hat{\mathbf{V}}_{k+1|k}\right)^{-1} \tag{4.15a}$$

$$\hat{\boldsymbol{x}}_{k|N} = \hat{\boldsymbol{x}}_{k|k} + \mathbf{N}_k \left(\hat{\boldsymbol{x}}_{k+1|N} - \hat{\boldsymbol{x}}_{k+1|k}\right) \tag{4.15b}$$

$$\hat{\mathbf{V}}_{k|N} = \hat{\mathbf{V}}_{k|k} + \mathbf{N}_k \left(\hat{\mathbf{V}}_{k+1|N} - \hat{\mathbf{V}}_{k+1|k}\right) \mathbf{N}_k^T \tag{4.15c}$$

$$\hat{\mathbf{V}}_{k+1,k|N} = \hat{\mathbf{V}}_{k+1|N}\mathbf{N}_k^T, \tag{4.15d}$$

where $\hat{\mathbf{V}}_{k+1,k|N}$ is called the lag-one smoother covariance matrix [64].

4.4.2 Derivation of the M-step

Maximizing $\mathcal{F}\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(j)}\right)$ with respect to the parameters $\boldsymbol{\theta}$ at iteration j constitutes the M-step of EM. The usefulness of EM in this particular case — where the full model parameters (i.e. $\mathbf{A}, \mathbf{G}, \mathbf{Q}, \mathbf{R}$) are estimated — stems from the fact that the M-step maximizations are obtained in closed form by setting the gradient $\frac{\partial \mathcal{F}\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)}\right)}{\partial \boldsymbol{\theta}}$ to zero for each $\boldsymbol{\theta} = \{\mathbf{A}, \mathbf{G}, \mathbf{Q}, \mathbf{R}\}$. The closed form expressions of $\boldsymbol{\theta}^{(j+1)}$ are given as follows:

$$\mathbf{A}^{(j+1)} = \mathbf{S}_{fx}\mathbf{S}_{xx}^{-1} \tag{4.16a}$$

$$\mathbf{Q}^{(j+1)} = \frac{1}{N}\left[\mathbf{S}_{ff} - \mathbf{S}_{fx}\mathbf{A}^{(j+1)^T} - \mathbf{A}^{(j+1)}\mathbf{S}_{fx}^T + \mathbf{A}^{(j+1)}\mathbf{S}_{xx}\mathbf{A}^{(j+1)^T}\right] \tag{4.16b}$$

$$\mathbf{G}^{(j+1)} = \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1} \tag{4.17a}$$

$$\mathbf{R}^{(j+1)} = \frac{1}{N}\left[\mathbf{S}_{yy} - \mathbf{S}_{yx}\mathbf{G}^{(j+1)^T} - \mathbf{G}^{(j+1)}\mathbf{S}_{yx}^T + \mathbf{G}^{(j+1)}\mathbf{S}_{xx}\mathbf{G}^{(j+1)^T}\right]. \tag{4.17b}$$

The auxiliary parameters $\boldsymbol{\mu}_1$ and $\mathbf{V}_1$ are set as

$$\boldsymbol{\mu}_1^{(j+1)} = \hat{\boldsymbol{x}}_{1|N} \tag{4.18a}$$

$$\mathbf{V}_1^{(j+1)} = \hat{\mathbf{V}}_{1|N}. \tag{4.18b}$$

The derivation of the M-step update rules can be found in [68, 70].

4.4.3 Initialization of EM

The EM algorithm, being an iterative optimizer, requires an initial estimate of the system, $\boldsymbol{\theta}^{(0)}$, to commence. The choice of the initial estimate determines how quickly the algorithm converges to a solution. Typically, EM is initialized using estimates from other identification algorithms such as numerical algorithm for subspace state space system identification (N4SID) [104] and SSI [24]. Another preferred strategy is to use random initializations of the parameters, however some difficulties (i.e. system instability, inaccurate solutions) have been reported [68] when dealing with higher order SSM. In this study, the initial point $\boldsymbol{\theta}^{(0)} = \left\{\mathbf{A}^{(0)}, \mathbf{G}^{(0)}, \mathbf{Q}^{(0)}, \mathbf{R}^{(0)}\right\}$ are set as follows:

$$\begin{aligned} \mathbf{A}^{(0)} &= \mathbf{A}^{\text{SSI}}, \qquad & \mathbf{G}^{(0)} &= \mathbf{G}^{\text{SSI}} \\ \mathbf{Q}^{(0)} &= \mathbf{I}_{n_s \times n_s}, \qquad & \mathbf{R}^{(0)} &= \mathbf{I}_{n_o \times n_o}, \end{aligned} \tag{4.19}$$

where $\mathbf{A}^{\text{SSI}}$ and $\mathbf{G}^{\text{SSI}}$ are estimates from SSI.

4.4.4 Convergence Criterion of EM

A commonly used convergence criterion for the EM algorithm is based on the change in the log-likelihood values between iterations as given by:

$$\frac{\hat{\mathcal{F}}^{(j+1)} - \hat{\mathcal{F}}^{(j)}}{\left(\hat{\mathcal{F}}^{(j+1)} + \hat{\mathcal{F}}^{(j)}\right)/2} \leq \epsilon_{tol}, \tag{4.20}$$

where ϵ_{tol} is typically in the range of $10^{-4} \sim 10^{-6}$. The EM iterations are stopped once the convergence criterion is satisfied, and the parameters from the final iteration are treated as the ML estimates of the parameters, $\hat{\boldsymbol{\theta}}_{\text{ML}} = \left\{\hat{\mathbf{A}}_{\text{ML}}, \hat{\mathbf{G}}_{\text{ML}}, \hat{\mathbf{Q}}_{\text{ML}}, \hat{\mathbf{R}}_{\text{ML}}\right\}$. The ML estimate of the modal parameters, $\left\{\hat{f}_i, \hat{\xi}_i, \hat{\boldsymbol{\phi}}_i^{un}\right\}$, can then be obtained by feeding $\hat{\mathbf{A}}_{\text{ML}}, \hat{\mathbf{G}}_{\text{ML}}$ to Appendix B.1. The complete list of steps for applying the EM algorithm to the missing data case is provided in Algorithm 1.

Algorithm 1 Modal parameter estimation using EM with missing data

Input: $\boldsymbol{Y}$, $\mathbf{A}^{(0)}$, $\mathbf{G}^{(0)}$, $\mathbf{Q}^{(0)}$, $\mathbf{R}^{(0)}$, $\boldsymbol{\mu}_1^{(0)}$, $\mathbf{V}_1^{(0)}$, ϵ_{tol}, $\left\{\mathfrak{V}_k^{(1)}, \mathfrak{V}_k^{(2)}, \mathbf{I}_k^{(1)}, \mathbf{I}_k^{(2)}\right\}_{1:N}$

E-step:

1. Use Eqs.(4.13) and (4.15) for Kalman filter and RTS smoother
2. Use Eq.(4.14) to compute $\hat{\mathcal{F}}^{(0)}$
3. Calculate the expectations of sufficient statistics using Eqs.(4.11) and (4.12)

while not converged **do**

1. **M**-step: Use Eqs.(4.16), (4.17), and (4.18) to update parameters
   $\rightarrow \mathbf{A}^{(j+1)}, \mathbf{G}^{(j+1)}, \mathbf{Q}^{(j+1)}, \mathbf{R}^{(j+1)}, \boldsymbol{\mu}_1^{(j+1)}, \mathbf{V}_1^{(j+1)}$
2. **E**-step:
   (a) Use Eq.(4.13) and (4.15) for Kalman filter and RTS smoother
   (b) Use Eq.(4.14) to compute $\mathcal{F}^{(j+1)}$
   (c) Calculate the expectations of sufficient statistics using Eqs.(4.11) and (4.12)
3. Use Eq.(4.20) to check if convergence criterion is satisfied

end while

Return $\hat{\mathbf{A}}_{\text{ML}} \leftarrow \mathbf{A}^{(final)}$, $\mathbf{G}_{\text{ML}} \leftarrow \mathbf{G}^{(final)}$

Extract modal frequencies, damping ratios and un-normalized mode shapes from $\hat{\mathbf{A}}_{\text{ML}}, \hat{\mathbf{G}}_{\text{ML}}$ using Appendix B.1

Output: $\hat{f}_i, \hat{\xi}_i, \hat{\boldsymbol{\phi}}_i^{un}$ for all modes $i = 1, 2, \ldots$

4.5 Bayesian estimation of LGSSM via VB

While the EM algorithm computes ML point estimates of each parameter, VB computes (an approximation to) the entire posterior distribution of the parameters and latent variables. The variational Bayesian approximation assumes a factored distribution over the parameters and the latent variables (as mentioned in Section 2.2.3) and the VB algorithm updates their approximate posterior distributions in an alternating fashion (similar to the

alternating structure of EM).

First, consider applying the VB framework for the LGSSM assuming a complete data matrix is available (we will later show how the framework is adopted to the missing data case). To start, the joint distribution of all variables in the LGSSM must be set up as follows:

$$p\left(\boldsymbol{Y}, \boldsymbol{X}, \mathbf{A}, \mathbf{Q}, \mathbf{G}, \mathbf{R}\right) = p\left(\boldsymbol{Y}, \boldsymbol{X} \mid \mathbf{A}, \mathbf{Q}, \mathbf{G}, \mathbf{R}\right) p\left(\mathbf{A}, \mathbf{Q}, \mathbf{G}, \mathbf{R}\right) \tag{4.21}$$

where

$$p\left(\boldsymbol{x}_1 \mid \boldsymbol{\mu}_1, \mathbf{V}_1\right) = \mathcal{N}\left(\boldsymbol{\mu}_1, \mathbf{V}_1\right) = \frac{\exp\left(-\frac{1}{2}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1)^T \mathbf{V}_1^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1)\right)}{(2\pi)^{n_s/2}\left|\mathbf{V}_1\right|^{1/2}} \tag{4.22a}$$

$$p\left(\boldsymbol{x}_{k+1} \mid \boldsymbol{x}_k, \mathbf{A}, \mathbf{Q}\right) = \mathcal{N}\left(\mathbf{A}\boldsymbol{x}_k, \mathbf{Q}\right) = \frac{\exp\left(-\frac{1}{2}(\boldsymbol{x}_{k+1} - \mathbf{A}\boldsymbol{x}_k)^T \mathbf{Q}^{-1}(\boldsymbol{x}_{k+1} - \mathbf{A}\boldsymbol{x}_k)\right)}{(2\pi)^{n_s/2}\left|\mathbf{Q}\right|^{1/2}} \tag{4.22b}$$

$$p\left(\boldsymbol{y}_k \mid \boldsymbol{x}_k, \mathbf{G}, \mathbf{R}\right) = \mathcal{N}\left(\mathbf{G}\boldsymbol{x}_k, \mathbf{R}\right) = \frac{\exp\left(-\frac{1}{2}(\boldsymbol{y}_k - \mathbf{G}\boldsymbol{x}_k)^T \mathbf{R}^{-1}(\boldsymbol{y}_k - \mathbf{G}\boldsymbol{x}_k)\right)}{(2\pi)^{n_o/2}\left|\mathbf{R}\right|^{1/2}}. \tag{4.22c}$$

For the LGSSM parameters, a factorized prior distribution is assumed:

$$p\left(\mathbf{A}, \mathbf{Q}, \mathbf{G}, \mathbf{R}\right) = p\left(\mathbf{A} \mid \mathbf{Q}\right) p\left(\mathbf{Q}\right) p\left(\mathbf{G} \mid \mathbf{R}\right) p\left(\mathbf{R}\right) \tag{4.23}$$

where random covariance matrices $\mathbf{Q}$ and $\mathbf{R}$ are assumed independent of each other. The forms of the prior distributions are assumed as follows:

$$p\left(\mathbf{A} \mid \mathbf{Q}\right) = \mathcal{MN}\left(\boldsymbol{\mu}_{\mathbf{A}}, \mathbf{Q}, \boldsymbol{\Pi}\right) = \frac{\exp\left(-\frac{1}{2}\operatorname{tr}\left\{\boldsymbol{\Pi}^{-1}(\mathbf{A} - \boldsymbol{\mu}_{\mathbf{A}})^T \mathbf{Q}^{-1}(\mathbf{A} - \boldsymbol{\mu}_{\mathbf{A}})\right\}\right)}{(2\pi)^{n_s^2/2}\left|\boldsymbol{\Pi}\right|^{n_s/2}\left|\mathbf{Q}\right|^{n_s/2}}, \tag{4.24a}$$

$$p\left(\mathbf{G} \mid \mathbf{R}\right) = \mathcal{MN}\left(\boldsymbol{\mu}_{\mathbf{G}}, \mathbf{R}, \boldsymbol{\Pi}\right) = \frac{\exp\left(-\frac{1}{2}\operatorname{tr}\left\{\boldsymbol{\Pi}^{-1}(\mathbf{G} - \boldsymbol{\mu}_{\mathbf{G}})^T \mathbf{R}^{-1}(\mathbf{G} - \boldsymbol{\mu}_{\mathbf{G}})\right\}\right)}{(2\pi)^{n_s n_o/2}\left|\boldsymbol{\Pi}\right|^{n_o/2}\left|\mathbf{R}\right|^{n_s/2}}, \tag{4.24b}$$

$$p(\mathbf{Q}) = \mathcal{IW}\left(d_{\mathbf{Q}}, \mathbf{D}_{\mathbf{Q}}\right) = \frac{\left|\mathbf{D}_{\mathbf{Q}}\right|^{d_{\mathbf{Q}}/2}\left|\mathbf{Q}\right|^{-(d_{\mathbf{Q}}+n_s+1)/2}\exp\left(-\frac{1}{2}\operatorname{tr}\left\{\mathbf{D}_{\mathbf{Q}}\mathbf{Q}^{-1}\right\}\right)}{2^{n_s d_{\mathbf{Q}}/2}\ \ \Gamma_{n_s}\left(\frac{d_{\mathbf{Q}}}{2}\right)}, \tag{4.24c}$$

$$p(\mathbf{R}) = \mathcal{IW}\left(d_{\mathbf{R}}, \mathbf{D}_{\mathbf{R}}\right) = \frac{\left|\mathbf{D}_{\mathbf{R}}\right|^{d_{\mathbf{R}}/2}\left|\mathbf{R}\right|^{-(d_{\mathbf{R}}+n_o+1)/2}\exp\left(-\frac{1}{2}\operatorname{tr}\left\{\mathbf{D}_{\mathbf{R}}\mathbf{R}^{-1}\right\}\right)}{2^{n_o d_{\mathbf{R}}/2}\ \ \Gamma_{n_o}\left(\frac{d_{\mathbf{R}}}{2}\right)}, \tag{4.24d}$$

where $\mathcal{MN}(\cdot)$ is the *Matrix Normal* distribution and $\mathcal{IW}(\cdot)$ is the *Inverse Wishart* distribution. The joint prior distributions of the parameters, $p\left(\mathbf{A}, \mathbf{G}, \mathbf{Q}, \mathbf{R}\right)$, is described

by a Matrix-Normal-Inverse-Wishart (MNIW) distribution [105, 106]. The set of prior hyperparameters is denoted by $\boldsymbol{\pi}_h = (\boldsymbol{\mu}_{\mathbf{A}}, \boldsymbol{\mu}_{\mathbf{G}}, \boldsymbol{\Pi}, \mathbf{D}_{\mathbf{Q}}, \mathbf{D}_{\mathbf{R}}, d_{\mathbf{Q}}, d_{\mathbf{R}})$. The choice of prior distributions over the LGSSM parameters is motivated by conjugacy – the posterior distribution of the parameters will also have a MNIW distribution and therefore becomes convenient for obtaining closed form update expressions for the hyperparameters of the MNIW distribution. It must be mentioned that the assumption of statistical independence between the matrices $\mathbf{A}$ and $\mathbf{G}$ is employed to make the VB inference amenable to closed-form updates; although, in practice they would actually be correlated. For the same reason, statistical independence is also assumed between the process and measurement noise which would be correlated when acceleration measurements are used as outputs.

Following the VB framework, a factorized form of variational approximation to the posterior distribution (see Eq.(2.29)) over the parameters $\boldsymbol{\theta}$ and the latent states $\boldsymbol{X}$ is assumed:

$$\begin{aligned} p\left(\mathbf{A}, \mathbf{Q}, \mathbf{G}, \mathbf{R}, \boldsymbol{X} \mid \boldsymbol{Y}\right) &\approx q\left(\mathbf{A}, \mathbf{Q}\right) q\left(\mathbf{G}, \mathbf{R}\right) q\left(\boldsymbol{X}\right) \\ &\approx q\left(\mathbf{A} \mid \mathbf{Q}\right) q\left(\mathbf{Q}\right) q\left(\mathbf{G} \mid \mathbf{R}\right) q\left(\mathbf{R}\right) q\left(\boldsymbol{X}\right). \end{aligned} \tag{4.25}$$

The variational free energy is maximized by setting the variational distributions as follows [107]:

$$q\left(\mathbf{A}, \mathbf{Q}\right) \propto \exp\left(\left\langle \ln p\left(\boldsymbol{Y}, \boldsymbol{X}, \mathbf{A}, \mathbf{Q}, \mathbf{G}, \mathbf{R}\right)\right\rangle_{q(\mathbf{G}|\mathbf{R})q(\mathbf{R})q(\boldsymbol{X})}\right) \tag{4.26a}$$

$$q\left(\mathbf{G}, \mathbf{R}\right) \propto \exp\left(\left\langle \ln p\left(\boldsymbol{Y}, \boldsymbol{X}, \mathbf{A}, \mathbf{Q}, \mathbf{G}, \mathbf{R}\right)\right\rangle_{q(\mathbf{A}|\mathbf{Q})q(\mathbf{Q})q(\boldsymbol{X})}\right) \tag{4.26b}$$

$$q\left(\boldsymbol{X}\right) \propto \exp\left(\left\langle \ln p\left(\boldsymbol{Y}, \boldsymbol{X}, \mathbf{A}, \mathbf{Q}, \mathbf{G}, \mathbf{R}\right)\right\rangle_{q(\mathbf{A}|\mathbf{Q})q(\mathbf{Q})q(\mathbf{G}|\mathbf{R})q(\mathbf{R})}\right) \tag{4.26c}$$

where the $\left\langle f(x)\right\rangle_{q(x)} := \int f(x) q\left(x\right) dx$. In the following, an iterative scheme based on the above equations is derived by indexing the variational distributions $q\left(\cdot\right)$ as $q^{(j)}\left(\cdot\right)$ at the current VB iteration and as $q^{(j+1)}\left(\cdot\right)$ as the variational distributions in the next VB iteration:

$$q^{(j)}\left(\boldsymbol{X}\right) \propto \exp\left(\left\langle \ln p\left(\boldsymbol{Y}, \boldsymbol{X}, \mathbf{A}, \mathbf{Q}, \mathbf{G}, \mathbf{R}\right)\right\rangle_{q^{(j)}(\mathbf{A}|\mathbf{Q})q^{(j)}(\mathbf{Q})q^{(j)}(\mathbf{G}|\mathbf{R})q^{(j)}(\mathbf{R})}\right) \tag{4.27a}$$

$$q^{(j+1)}\left(\mathbf{A}, \mathbf{Q}\right) \propto \exp\left(\left\langle \ln p\left(\boldsymbol{Y}, \boldsymbol{X}, \mathbf{A}, \mathbf{Q}, \mathbf{G}, \mathbf{R}\right)\right\rangle_{q^{(j)}(\mathbf{G}|\mathbf{R})q^{(j)}(\mathbf{R})q^{(j)}(\boldsymbol{X})}\right) \tag{4.27b}$$

$$q^{(j+1)}\left(\mathbf{G}, \mathbf{R}\right) \propto \exp\left(\left\langle \ln p\left(\boldsymbol{Y}, \boldsymbol{X}, \mathbf{A}, \mathbf{Q}, \mathbf{G}, \mathbf{R}\right)\right\rangle_{q^{(j)}(\mathbf{A}|\mathbf{Q})q^{(j)}(\mathbf{Q})q^{(j)}(\boldsymbol{X})}\right). \tag{4.27c}$$

Here the jth variational distributions are defined as

$$\begin{aligned} q^{(j)}\left(\mathbf{A} \mid \mathbf{Q}\right) &= \mathcal{MN}\left(\boldsymbol{\mu}_{\mathbf{A}}^{(j)}, \mathbf{Q}^{(j)}, \boldsymbol{\Pi}^{(j)}\right) \\ &= \frac{\exp\left(-\frac{1}{2}\operatorname{tr}\left\{\boldsymbol{\Pi}^{(j)^{-1}}\left(\mathbf{A}-\boldsymbol{\mu}_{\mathbf{A}}^{(j)}\right)^{T}\mathbf{Q}^{(j)^{-1}}\left(\mathbf{A}-\boldsymbol{\mu}_{\mathbf{A}}^{(j)}\right)\right\}\right)}{(2\pi)^{n_s^2/2}\left|\boldsymbol{\Pi}^{(j)}\right|^{n_s/2}\left|\mathbf{Q}^{(j)}\right|^{n_s/2}} \end{aligned} \tag{4.28a}$$

$$\begin{aligned} q^{(j)}\left(\mathbf{G} \mid \mathbf{R}\right) &= \mathcal{MN}\left(\boldsymbol{\mu}_{\mathbf{G}}^{(j)}, \mathbf{R}^{(j)}, \boldsymbol{\Pi}^{(j)}\right) \\ &= \frac{\exp\left(-\frac{1}{2}\operatorname{tr}\left\{\boldsymbol{\Pi}^{(j)^{-1}}\left(\mathbf{G}-\boldsymbol{\mu}_{\mathbf{G}}^{(j)}\right)^{T}\mathbf{R}^{(j)^{-1}}\left(\mathbf{G}-\boldsymbol{\mu}_{\mathbf{G}}^{(j)}\right)\right\}\right)}{(2\pi)^{n_s n_o/2}\left|\boldsymbol{\Pi}^{(j)}\right|^{n_o/2}\left|\mathbf{R}^{(j)}\right|^{n_s/2}} \end{aligned} \tag{4.28b}$$

$$\begin{aligned} q^{(j)}\left(\mathbf{Q}\right) &= \mathcal{IW}\left(d_{\mathbf{Q}}^{(j)}, \mathbf{D}_{\mathbf{Q}}^{(j)}\right) \\ &= \frac{\left|\mathbf{D}_{\mathbf{Q}}^{(j)}\right|^{d_{\mathbf{Q}}^{(j)}/2}\left|\mathbf{Q}\right|^{-(d_{\mathbf{Q}}^{(j)}+n_s+1)/2}\exp\left(-\frac{1}{2}\operatorname{tr}\left\{\mathbf{D}_{\mathbf{Q}}^{(j)}\mathbf{Q}^{-1}\right\}\right)}{2^{n_s d_{\mathbf{Q}}^{(j)}/2}\ \Gamma_{n_s}\left(\frac{d_{\mathbf{Q}}^{(j)}}{2}\right)} \end{aligned} \tag{4.28c}$$

$$\begin{aligned} q^{(j)}\left(\mathbf{R}\right) &= \mathcal{IW}\left(d_{\mathbf{R}}^{(j)}, \mathbf{D}_{\mathbf{R}}^{(j)}\right) \\ &= \frac{\left|\mathbf{D}_{\mathbf{R}}^{(j)}\right|^{d_{\mathbf{R}}^{(j)}/2}\left|\mathbf{R}\right|^{-(d_{\mathbf{R}}^{(j)}+n_o+1)/2}\exp\left(-\frac{1}{2}\operatorname{tr}\left\{\mathbf{D}_{\mathbf{R}}^{(j)}\mathbf{R}^{-1}\right\}\right)}{2^{n_o d_{\mathbf{R}}^{(j)}/2}\ \Gamma_{n_o}\left(\frac{d_{\mathbf{R}}^{(j)}}{2}\right)} \end{aligned} \tag{4.28d}$$

Algorithmically, the variational Bayesian algorithm starts from an initial point of the hyperparameters

$$\boldsymbol{\theta}_h^{(0)} = \left\{\boldsymbol{\mu}_{\mathbf{A}}^{(0)}, \boldsymbol{\mu}_{\mathbf{G}}^{(0)}, \boldsymbol{\Pi}^{(0)}, \mathbf{D}_{\mathbf{Q}}^{(0)}, \mathbf{D}_{\mathbf{R}}^{(0)}, d_{\mathbf{Q}}^{(0)}, d_{\mathbf{R}}^{(0)}\right\} \tag{4.29}$$

and repeats the following two steps iteratively until the free energy $\mathcal{F}_{ve}\left(q^{(j)}\left(\boldsymbol{\theta}, \boldsymbol{X}\right)\right)$ (refer Eq.(2.28)) converges to a stationary value:

1. **VBE-step**: Given $\boldsymbol{\theta}_h^{(j)}$, compute $q^{(j)}\left(\boldsymbol{X}\right)$,
2. **VBM-step**: Using $q^{(j)}\left(\boldsymbol{X}\right)$, obtain the hyperparameters $\boldsymbol{\theta}_h^{(j+1)}$ of $q^{(j+1)}\left(\boldsymbol{\theta}\right)$.

4.5.1 Evaluation of variational expectation

To derive the update equations for the VBE and VBM steps, the expectation of the log joint PDF of the hidden LGSSM variables (i.e., the states and parameters) with respect to the variational PDFs has to be evaluated. One can use Eqs. (4.21) and (4.24) to arrive at the following expression for the expected value of the log joint PDF:

$$
\begin{aligned}
&\left\langle \ln\left(p\left(\boldsymbol{Y},\boldsymbol{X},\mathbf{A},\mathbf{Q},\mathbf{G},\mathbf{R}\right)\right)\right\rangle_{q^{(j)}(\mathbf{A}|\mathbf{Q})q^{(j)}(\mathbf{Q})q^{(j)}(\mathbf{G}|\mathbf{R})q^{(j)}(\mathbf{R})q^{(j)}(\boldsymbol{X})} \\
\propto &-\frac{1}{2}\left\langle (\boldsymbol{x}_1-\boldsymbol{\mu}_1)^T\mathbf{V}_1^{-1}(\boldsymbol{x}_1-\boldsymbol{\mu}_1)\right\rangle_{q^{(j)}(\boldsymbol{x}_1)} - \frac{n_s}{2}\left\langle \ln|\mathbf{Q}|\right\rangle_{q^{(j)}(\mathbf{Q})} - \frac{n_o}{2}\left\langle \ln|\mathbf{R}|\right\rangle_{q^{(j)}(\mathbf{R})} \\
&-\frac{1}{2}\left\langle \mathrm{tr}\left\{\mathbf{Q}^{-1}\left(\mathbf{S}_{ff}^{(j)}-\mathbf{A}\mathbf{S}_{fx}^{(j)T}-\mathbf{S}_{fx}^{(j)}\mathbf{A}^T+\mathbf{A}\mathbf{S}_{xx}^{(j)}\mathbf{A}^T\right)\right\}\right\rangle_{q^{(j)}(\mathbf{A},\mathbf{Q})} \\
&-\frac{1}{2}\left\langle \mathrm{tr}\left\{\mathbf{R}^{-1}\left(\mathbf{S}_{yy}^{(j)}-\mathbf{G}\mathbf{S}_{yx}^{(j)T}-\mathbf{S}_{yx}^{(j)}\mathbf{G}^T+\mathbf{G}\mathbf{S}_{xx}^{(j)}\mathbf{G}^T\right)\right\}\right\rangle_{q^{(j)}(\mathbf{G},\mathbf{R})} \\
&-\frac{1}{2}\left\langle \mathrm{tr}\left\{\mathbf{Q}^{-1}\left(\boldsymbol{\mu}_{\mathbf{A}}\boldsymbol{\Pi}^{-1}\boldsymbol{\mu}_{\mathbf{A}}^T-\mathbf{A}\boldsymbol{\Pi}^{-1}\boldsymbol{\mu}_{\mathbf{A}}^T-\boldsymbol{\mu}_{\mathbf{A}}\boldsymbol{\Pi}^{-1}\mathbf{A}^T+\mathbf{A}\boldsymbol{\Pi}^{-1}\mathbf{A}^T\right)\right\}\right\rangle_{q^{(j)}(\mathbf{A},\mathbf{Q})} \\
&-\frac{1}{2}\left\langle \mathrm{tr}\left\{\mathbf{R}^{-1}\left(\boldsymbol{\mu}_{\mathbf{G}}\boldsymbol{\Pi}^{-1}\boldsymbol{\mu}_{\mathbf{G}}^T-\mathbf{G}\boldsymbol{\Pi}^{-1}\boldsymbol{\mu}_{\mathbf{G}}^T-\boldsymbol{\mu}_{\mathbf{G}}\boldsymbol{\Pi}^{-1}\mathbf{G}^T+\mathbf{G}\boldsymbol{\Pi}^{-1}\mathbf{G}^T\right)\right\}\right\rangle_{q^{(j)}(\mathbf{G},\mathbf{R})} \\
&-\frac{d_{\mathbf{Q}}+N+n_s+1}{2}\left\langle \ln|\mathbf{Q}|\right\rangle_{q^{(j)}(\mathbf{Q})} - \frac{1}{2}\left\langle \mathrm{tr}\left\{\mathbf{D}_{\mathbf{Q}}\mathbf{Q}^{-1}\right\}\right\rangle_{q^{(j)}(\mathbf{Q})} \\
&-\frac{d_{\mathbf{R}}+N+n_o+1}{2}\left\langle \ln|\mathbf{R}|\right\rangle_{q^{(j)}(\mathbf{R})} - \frac{1}{2}\left\langle \mathrm{tr}\left\{\mathbf{D}_{\mathbf{R}}\mathbf{R}^{-1}\right\}\right\rangle_{q^{(j)}(\mathbf{R})}.
\end{aligned}
\tag{4.30}
$$

The matrices $\mathbf{S}_{xx}^{(j)}$, $\mathbf{S}_{fx}^{(j)}$, $\mathbf{S}_{ff}^{(j)}$, $\mathbf{S}_{yx}^{(j)}$ and $\mathbf{S}_{yy}^{(j)}$ are the expectations of sufficient statistics needed in the VBM step and are defined as:

$$
\begin{aligned}
\mathbf{S}_{xx}^{(j)} &= \sum_{k=1}^{N}\left\langle \boldsymbol{x}_k\boldsymbol{x}_k^T\right\rangle_{q^{(j)}(\boldsymbol{X})}, \\
\mathbf{S}_{fx}^{(j)} &= \sum_{k=1}^{N}\left\langle \boldsymbol{x}_{k+1}\boldsymbol{x}_k^T\right\rangle_{q^{(j)}(\boldsymbol{X})}, \\
\mathbf{S}_{ff}^{(j)} &= \sum_{k=1}^{N}\left\langle \boldsymbol{x}_{k+1}\boldsymbol{x}_{k+1}^T\right\rangle_{q^{(j)}(\boldsymbol{X})},
\end{aligned}
\tag{4.31a}
$$

$$
\mathbf{S}_{yx}^{(j)} = \sum_{k=1}^{N}\left\langle \boldsymbol{y}_k\boldsymbol{x}_k^T\right\rangle_{q^{(j)}(\boldsymbol{X})}, \quad \mathbf{S}_{yy}^{(j)} = \sum_{k=1}^{N}\left\langle \boldsymbol{y}_k\boldsymbol{y}_k^T\right\rangle_{q^{(j)}(\boldsymbol{X})}. \tag{4.31b}
$$

The procedures to calculate the expectations of sufficient statistics in Eq.(4.31a) and (4.31b) are described in the next section.

4.5.2 Derivation of VBE-step: Evaluating $q^{(j)}(\boldsymbol{X})$

The evaluation of the expectations needed to compute the expectations of sufficient statistics, as given in Eqs.(4.31a) and (4.31b), constitutes the VBE-step. Under the assumption of non-random, known $\boldsymbol{\theta}$, the evaluation of the variational distribution $q^{(j)}(\boldsymbol{X})$ is equivalent to the evaluation of the conditional distribution over the variables $\boldsymbol{X}$ given an observed data sequence $\boldsymbol{Y}$. This corresponds to classical state estimation which can accomplished using a Kalman filter (and smoother). However, the values of $\boldsymbol{\theta}$ in the ensuing variational approximation are not fixed, but are governed by probability distributions. This fact renders the standard algorithms inappropriate for evaluating $q^{(j)}(\boldsymbol{X})$.

A potential solution is the modified state space model, proposed by Barber and Chiappa [108], which permits the use of a Kalman filter and RTS smoother for complete data matrices. In this thesis, the modified state space model is adopted for the missing-data case by applying permutation matrices to separate the missing data from the observed date (see Section 4.3.1). The sequence of permutation matrices $\left\{\mathfrak{V}_k^{(1)}, \mathfrak{V}_k^{(2)}, \mathbf{I}_k^{(1)}, \mathbf{I}_k^{(2)}\right\}_{1:N}$ are computed based on the missingness pattern of the entries in the incomplete data matrix. The following terms associated with the modified state space model can then be defined as:

$$\widetilde{\mathbf{A}} = \langle \mathbf{A} \rangle_{q^{(j)}(\mathbf{A}|\mathbf{Q})} = \boldsymbol{\mu}_{\mathbf{A}}^{(j)} \tag{4.32a}$$

$$\widetilde{\mathbf{Q}} = \langle \mathbf{Q} \rangle_{q^{(j)}(\mathbf{Q})} = \frac{\mathbf{D}_Q^{(j)}}{d_{\mathbf{Q}}^{(j)} - n_s - 1} \tag{4.32b}$$

$$\underline{\mathbf{G}}_k = \mathbf{I}_k^{(1)} \langle \mathbf{G} \rangle_{q^{(j)}(\mathbf{G}|\mathbf{R})} = \mathbf{I}_k^{(1)} \boldsymbol{\mu}_{\mathbf{G}}^{(j)} \tag{4.32c}$$

$$\underline{\mathbf{R}}_k = \mathbf{I}_k^{(1)} \langle \mathbf{R} \rangle_{q^{(j)}(\mathbf{R})} \mathbf{I}_k^{(1)} + \mathbf{I}_k^{(2)} \mathbf{I}_k^{(2)} = \frac{\mathbf{I}_k^{(1)} \mathbf{D}_R^{(j)} \mathbf{I}_k^{(1)}}{d_{\mathbf{R}}^{(j)} - n_o - 1} + \mathbf{I}_k^{(2)} \mathbf{I}_k^{(2)} \tag{4.32d}$$

$$\mathbf{U}_k^T \mathbf{U}_k = \left\langle \mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A} \right\rangle_{q^{(j)}(\mathbf{A},\mathbf{Q})} - \left\langle \mathbf{A}^T \right\rangle_{q^{(j)}(\mathbf{A}|\mathbf{Q})} \left\langle \mathbf{Q}^{-1} \right\rangle_{q^{(j)}(\mathbf{Q})} \langle \mathbf{A} \rangle_{q^{(j)}(\mathbf{A}|\mathbf{Q})}$$

$$
\begin{aligned}
&- \left\langle \underline{\mathbf{G}}_k^T \underline{\mathbf{R}}_k^{-1} \underline{\mathbf{G}}_k \right\rangle_{q^{(j)}(\mathbf{G},\mathbf{R})} - \left\langle \underline{\mathbf{G}}_k^T \right\rangle_{q^{(j)}(\mathbf{G}|\mathbf{R})} \left\langle \underline{\mathbf{R}}_k^{-1} \right\rangle_{q^{(j)}(\mathbf{R})} \left\langle \underline{\mathbf{G}}_k \right\rangle_{q^{(j)}(\mathbf{G}|\mathbf{R})} \\
= \;& (n_s + n_{obs,k}) \boldsymbol{\Pi}^{(j)} \qquad (4.32e)
\end{aligned}
$$

where $n_{obs,k}$ is the number of observed entries at time step k, and

$$
\widetilde{\boldsymbol{y}}_k = \begin{bmatrix} \mathbf{I}_k^{(1)} \boldsymbol{y}_k \\ \mathbf{0}_{n_s} \end{bmatrix}, \qquad \widetilde{\mathbf{G}}_k = \begin{bmatrix} \underline{\mathbf{G}}_k \\ \mathbf{U}_k \end{bmatrix}, \qquad \widetilde{\mathbf{R}}_k = \begin{bmatrix} \underline{\mathbf{R}}_k & \mathbf{0}_{n_o \times n_s} \\ \mathbf{0}_{n_o \times n_s}^T & \mathbf{I}_{n_s} \end{bmatrix}. \tag{4.33a}
$$

Furthermore, the Kalman filter and RTS smoother for the modified model are as follows:

Kalman filter for modified LGSSM : Do for $k = 1, \ldots, N$

$$
\begin{aligned}
\widetilde{\boldsymbol{e}}_k^* &= \widetilde{\boldsymbol{y}}_k - \widetilde{\mathbf{G}}_k \hat{\boldsymbol{x}}_{k|k-1} && (4.34a)\\
\widetilde{\mathbf{H}}_k^* &= \widetilde{\mathbf{G}}_k \hat{\mathbf{V}}_{k|k-1} \widetilde{\mathbf{G}}_k^T + \widetilde{\mathbf{R}}_k && (4.34b)\\
\widetilde{\mathbf{K}}_k^* &= \hat{\mathbf{V}}_{k|k-1} \widetilde{\mathbf{G}}_k^T \left(\widetilde{\mathbf{H}}_k^* \right)^{-1} && (4.34c)\\
\hat{\boldsymbol{x}}_{k|k} &= \hat{\boldsymbol{x}}_{k|k-1} + \widetilde{\mathbf{K}}_k^* \widetilde{\boldsymbol{e}}_k^* && (4.34d)\\
\hat{\mathbf{V}}_{k|k} &= \hat{\mathbf{V}}_{k|k-1} - \widetilde{\mathbf{K}}_k^* \widetilde{\mathbf{H}}_k^* \left(\widetilde{\mathbf{K}}_k^* \right)^T && (4.34e)\\
\hat{\boldsymbol{x}}_{k+1|k} &= \widetilde{\mathbf{A}} \hat{\boldsymbol{x}}_{k|k} && (4.34f)\\
\hat{\mathbf{V}}_{k+1|k} &= \widetilde{\mathbf{A}} \hat{\mathbf{V}}_{k|k} \widetilde{\mathbf{A}} + \widetilde{\mathbf{Q}}, && (4.34g)
\end{aligned}
$$

Kalman smoother : Do for $k = N, \ldots, 1$

$$
\begin{aligned}
\widetilde{\mathbf{N}}_k^* &= \hat{\mathbf{V}}_{k|k} \widetilde{\mathbf{A}}^T \left(\hat{\mathbf{V}}_{k+1|k} \right)^{-1} && (4.35a)\\
\hat{\boldsymbol{x}}_{k|N} &= \hat{\boldsymbol{x}}_{k|k} + \widetilde{\mathbf{N}}_k^* \left(\hat{\boldsymbol{x}}_{k+1|N} - \hat{\boldsymbol{x}}_{k+1|k} \right) && (4.35b)\\
\hat{\mathbf{V}}_{k|N} &= \hat{\mathbf{V}}_{k|k} + \widetilde{\mathbf{N}}_k^* \left(\hat{\mathbf{V}}_{k+1|N} - \hat{\mathbf{V}}_{k+1|k} \right) \left(\widetilde{\mathbf{N}}_k^* \right)^T && (4.35c)\\
\hat{\mathbf{V}}_{k+1,k|N} &= \hat{\mathbf{V}}_{k+1|N} \left(\widetilde{\mathbf{N}}_k^* \right)^T. && (4.35d)
\end{aligned}
$$

Given the estimated states, the three of the expectations of sufficient statistics can be computed as follows:

$$
\mathbf{S}_{xx}^{(j)} = \sum_{k=1}^{N} \left\langle \boldsymbol{x}_k \boldsymbol{x}_k^T \right\rangle_{q^{(j)}(\boldsymbol{X})} = \sum_{k=1}^{N} \left(\hat{\boldsymbol{x}}_{k|N} \hat{\boldsymbol{x}}_{k|N}^T + \hat{\mathbf{V}}_{k|N} \right) \tag{4.36a}
$$

$$\mathbf{S}_{fx}^{(j)} = \sum_{k=1}^{N} \left\langle \boldsymbol{x}_{k+1}\boldsymbol{x}_k^T \right\rangle_{q^{(j)}(\boldsymbol{X})} = \sum_{k=1}^{N} \left(\hat{\boldsymbol{x}}_{k+1|N}\hat{\boldsymbol{x}}_{k|N}^T + \hat{\mathbf{V}}_{k+1,k|N} \right) \tag{4.36b}$$

$$\mathbf{S}_{ff}^{(j)} = \sum_{k=1}^{N} \left\langle \boldsymbol{x}_{k+1}\boldsymbol{x}_{k+1}^T \right\rangle_{q^{(j)}(\boldsymbol{X})} = \sum_{k=1}^{N} \left(\hat{\boldsymbol{x}}_{k+1|N}\hat{\boldsymbol{x}}_{k+1|N}^T + \hat{\mathbf{V}}_{k+1|N} \right). \tag{4.36c}$$

Unlike the expectations of sufficient statistics $\mathbf{S}_{xx}^{(j)}$, $\mathbf{S}_{fx}^{(j)}$, $\mathbf{S}_{ff}^{(j)}$ which depend only on the estimated states, $\mathbf{S}_{yy}^{(j)}$ and $\mathbf{S}_{yx}^{(j)}$ also depend on the complete observations. Since the observations have missing components, the expected values of the missing components, conditional on the estimated states, are obtained prior to the calculation of $\mathbf{S}_{yy}^{(j)}$ and $\mathbf{S}_{yx}^{(j)}$:

$$\mathbf{R}^{(j)} = \langle \mathbf{R} \rangle_{q^{(j)}(\mathbf{R})} = \frac{\mathbf{D}_R^{(j)}}{d_{\mathbf{R}}^{(j)} - n_o - 1} \tag{4.37a}$$

$$\mathbf{Z}_k^{(j)} = \mathbf{I}_{n_o \times n_o} - \mathbf{R}^{(j)} \boldsymbol{\upsilon}_k^{(1)^T} \left(\boldsymbol{\upsilon}_k^{(1)} \mathbf{R}^{(j)} \boldsymbol{\upsilon}_k^{(1)^T} \right)^{-1} \boldsymbol{\upsilon}_k^{(1)} \tag{4.37b}$$

$$\hat{\boldsymbol{y}}_k = \langle \boldsymbol{y}_k \rangle_{q^{(j)}(\boldsymbol{X})} = \boldsymbol{y}_k - \mathbf{Z}_k^{(j)} \left(\boldsymbol{y}_k - \boldsymbol{\mu}_{\mathbf{G}}^{(j)} \hat{\boldsymbol{x}}_{k|N} \right) \tag{4.37c}$$

$$\mathbf{S}_{yy}^{(j)} = \sum_{k=1}^{N} \left\langle \boldsymbol{y}_k \boldsymbol{y}_k^T \right\rangle_{q^{(j)}(\boldsymbol{X})} = \sum_{k=1}^{N} \left(\mathbf{I}_k^{(2)} \mathbf{Z}_k^{(j)} \left(\mathbf{R}^{(j)} + \boldsymbol{\mu}_{\mathbf{G}}^{(j)} \hat{\mathbf{V}}_{k|N} \left(\boldsymbol{\mu}_{\mathbf{G}}^{(j)} \right)^T \right) \left(\mathbf{Z}_k^{(j)} \right)^T \mathbf{I}_k^{(2)} + \hat{\boldsymbol{y}}_k \hat{\boldsymbol{y}}_k^T \right) \tag{4.37d}$$

$$\mathbf{S}_{yx}^{(j)} = \sum_{k=1}^{N} \left\langle \boldsymbol{y}_k \boldsymbol{x}_k^T \right\rangle_{q^{(j)}(\boldsymbol{X})} = \sum_{k=1}^{N} \left(\mathbf{Z}_k^{(j)} \boldsymbol{\mu}_{\mathbf{G}}^{(j)} \hat{\mathbf{V}}_{k|N} + \hat{\boldsymbol{y}}_k \hat{\boldsymbol{x}}_{k|N}^T \right) \tag{4.37e}$$

Eq.(4.37c) imputes the missing observations by conditional mean values of the observations given the expected values of the states.

4.5.3 Derivation of VBM-step: Computing $q^{(j+1)}(\boldsymbol{\theta})$

This section presents the update rules for the hyperparameters of the variational distributions $q^{(j+1)}(\mathbf{A}, \mathbf{Q})$ and $q^{(j+1)}(\mathbf{G}, \mathbf{R})$. The update expressions are derived following the maximization rule in Eq.(4.27). The expectations of sufficient statistics in the jth iteration (refer Eqs.(4.36) and (4.37)) and the set of prior hyperparameters $\boldsymbol{\pi}_h = \{\boldsymbol{\mu}_{\mathbf{A}}, \boldsymbol{\mu}_{\mathbf{G}}, \boldsymbol{\Pi}, \mathbf{D}_{\mathbf{Q}}, \mathbf{D}_{\mathbf{R}}, d_{\mathbf{Q}}, d_{\mathbf{R}}\}$ (refer Eq.(4.24)) are used to derive the expressions. The details of the derivations of the update rules can be found in Appendix D.1.

Update for $q^{(j+1)}(\mathbf{A}, \mathbf{Q})$

The hyperparameters of $q^{(j+1)}(\mathbf{A}, \mathbf{Q})$ are updated using the following expressions:

$$\boldsymbol{\Pi}^{(j+1)} = \left(\mathbf{S}_{xx}^{(j)} + \boldsymbol{\Pi}^{-1}\right)^{-1} \tag{4.38a}$$

$$\boldsymbol{\mu}_{\mathbf{A}}^{(j+1)} = \left(\mathbf{S}_{fx}^{(j)} + \boldsymbol{\mu}_{\mathbf{A}}\boldsymbol{\Pi}^{-1}\right)\boldsymbol{\Pi}^{(j+1)} \tag{4.38b}$$

$$d_{\mathbf{Q}}^{(j+1)} = d_{\mathbf{Q}} + N \tag{4.38c}$$

$$\mathbf{D}_{\mathbf{Q}}^{(j+1)} = \mathbf{D}_{\mathbf{Q}} + \mathbf{S}_{ff}^{(j)} + \boldsymbol{\mu}_{\mathbf{A}}\boldsymbol{\Pi}^{-1}\boldsymbol{\mu}_{\mathbf{A}}^{T} - \boldsymbol{\mu}_{\mathbf{A}}^{(j+1)}{\boldsymbol{\Pi}^{(j+1)}}^{-1}{\boldsymbol{\mu}_{\mathbf{A}}^{(j+1)}}^{T}. \tag{4.38d}$$

Update for $q^{(j+1)}(\mathbf{G}, \mathbf{R})$

The hyperparameters of $q^{(j+1)}(\mathbf{G}, \mathbf{R})$ are updated using the following expressions

$$\boldsymbol{\mu}_{\mathbf{G}}^{(j+1)} = \left(\mathbf{S}_{yx}^{(j)} + \boldsymbol{\mu}_{\mathbf{G}}\boldsymbol{\Pi}^{-1}\right)\boldsymbol{\Pi}^{(j+1)} \tag{4.39a}$$

$$d_{\mathbf{R}}^{(j+1)} = d_{\mathbf{R}} + N \tag{4.39b}$$

$$\mathbf{D}_{\mathbf{R}}^{(j+1)} = \mathbf{D}_{\mathbf{R}} + \mathbf{S}_{ff}^{(j)} + \boldsymbol{\mu}_{\mathbf{G}}\boldsymbol{\Pi}^{-1}\boldsymbol{\mu}_{\mathbf{G}}^{T} - \boldsymbol{\mu}_{\mathbf{G}}^{(j+1)}{\boldsymbol{\Pi}^{(j+1)}}^{-1}{\boldsymbol{\mu}_{\mathbf{G}}^{(j+1)}}^{T}. \tag{4.39c}$$

The auxiliary parameters $\boldsymbol{\mu}_1$ and $\mathbf{V}_1$ are set as

$$\boldsymbol{\mu}_1^{(j+1)} = \hat{\boldsymbol{x}}_{1|N} \tag{4.40a}$$

$$\mathbf{V}_1^{(j+1)} = \hat{\mathbf{V}}_{1|N}. \tag{4.40b}$$

4.5.4 Evaluation of variational free energy

The computation of the variational free energy $\mathcal{F}_{ve}^{(j)} \coloneqq \mathcal{F}_{ve}\left(q^{(j)}(\boldsymbol{X}, \boldsymbol{\theta})\right)$ (refer Eq.(2.28)) at each iteration helps to ensure that the lower bound of the marginal likelihood is maximized monotonically. To conveniently evaluate the variational free energy, the "average energy" and an "entropy" decomposition [107] is used

$$\begin{aligned}\mathcal{F}_{ve}^{(j)} &= \int\int q^{(j)}(\boldsymbol{X}, \boldsymbol{\theta}) \ln \frac{p(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\theta})}{q^{(j)}(\boldsymbol{X}, \boldsymbol{\theta})} d\boldsymbol{X}\, d\boldsymbol{\theta} \\ &= \int\int q^{(j)}(\boldsymbol{X}, \boldsymbol{\theta}) \ln p(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\theta})\, d\boldsymbol{X}\, d\boldsymbol{\theta} + \int\int -q^{(j)}(\boldsymbol{X}, \boldsymbol{\theta}) \ln q^{(j)}(\boldsymbol{X}, \boldsymbol{\theta})\, d\boldsymbol{X}\, d\boldsymbol{\theta}\end{aligned}$$

$$= \mathcal{E}_{av}^{(j)} + \mathcal{H}^{(j)} \tag{4.41}$$

The detailed calculations of $\mathcal{E}_{av}^{(j)}$ and $\mathcal{H}^{(j)}$ are presented in Appendix D.2. The final expressions needed for their computations are presented as follows:

$$\begin{aligned}
\mathcal{E}_{av}^{(j)} = \frac{1}{2}\Bigg[& - c_3 \ln(2\pi) + (c_4 n_s + c_5 n_o - c_6)\ln 2 - 2\ln\Gamma_{n_s}\left(\frac{d_{\mathbf{Q}}}{2}\right) - 2\ln\Gamma_{n_o}\left(\frac{d_{\mathbf{R}}}{2}\right) \\
& - \ln|\mathbf{V}_1| - (n_s + n_o)\ln\left|\mathbf{\Pi}^{-1}\right| + d_{\mathbf{Q}}\ln|\mathbf{D}_{\mathbf{Q}}| + d_{\mathbf{R}}\ln|\mathbf{D}_{\mathbf{R}}| \\
& + c_4 \sum_{i=1}^{n_s}\gamma\left(\frac{d_{\mathbf{Q}}^{(j)} - i + 1}{2}\right) + c_5 \sum_{i=1}^{n_o}\gamma\left(\frac{d_{\mathbf{R}}^{(j)} - i + 1}{2}\right) - 2\mathrm{tr}\left\{\left(\mathbf{S}_{xx} + \mathbf{\Pi}^{-1}\right)\mathbf{\Pi}^{(j)}\right\} \\
& - c_4 \ln\left|\mathbf{D}_{\mathbf{Q}}^{(j)}\right| - c_5 \ln\left|\mathbf{D}_{\mathbf{R}}^{(j)}\right| - \mathrm{tr}\left\{\mathbf{V}_1^{-1}\mathbf{E}_1 + d_{\mathbf{Q}}^{(j)}{\mathbf{D}_{\mathbf{Q}}^{(j)}}^{-1}\mathbf{E}_2 + d_{\mathbf{R}}^{(j)}{\mathbf{D}_{\mathbf{R}}^{(j)}}^{-1}\mathbf{E}_3\right\} \Bigg],
\end{aligned} \tag{4.42}$$

where $\Gamma_p(\cdot)$ and $\gamma_p(\cdot)$ are gamma and digamma functions [109] and

$$\begin{aligned}
c_3 &= n_s + N n_s + N n_o + n_s^2 + n_s n_o \\
c_4 &= d_{\mathbf{Q}} + N + 2n_s + 1 \\
c_5 &= d_{\mathbf{R}} + N + n_s + n_o + 1 \\
c_6 &= d_{\mathbf{Q}} n_s + d_{\mathbf{R}} n_o,
\end{aligned} \tag{4.43}$$

and

$$\begin{aligned}
\mathbf{E}_1 &= \left\langle \boldsymbol{x}_1 \boldsymbol{x}_1^T \right\rangle_{q(\boldsymbol{x}_1)} - \left\langle \boldsymbol{x}_1 \right\rangle_{q(\boldsymbol{x}_1)} \boldsymbol{\mu}_1^T - \boldsymbol{\mu}_1 \left\langle \boldsymbol{x}_1^T \right\rangle_{q(\boldsymbol{x}_1)} + \boldsymbol{\mu}_1 \boldsymbol{\mu}_1^T \\
&= \left(\hat{\mathbf{V}}_{1|N} + \hat{\boldsymbol{x}}_{1|N}\hat{\boldsymbol{x}}_{1|N}^T\right) - \hat{\boldsymbol{x}}_{1|N}\boldsymbol{\mu}_1^T - \boldsymbol{\mu}_1 \hat{\boldsymbol{x}}_{1|N}^T + \boldsymbol{\mu}_1 \boldsymbol{\mu}_1^T \\
\mathbf{E}_2 &= \mathbf{D}_{\mathbf{Q}} + \boldsymbol{\mu}_{\mathbf{A}} \mathbf{\Pi}^{-1} \boldsymbol{\mu}_{\mathbf{A}}^T + \mathbf{S}_{ff} - \boldsymbol{\mu}_{\mathbf{A}}^{(j)} \left(\mathbf{S}_{fx} + \boldsymbol{\mu}_{\mathbf{A}} \mathbf{\Pi}^{-1}\right)^T - \left(\mathbf{S}_{fx} + \boldsymbol{\mu}_{\mathbf{A}} \mathbf{\Pi}^{-1}\right) {\boldsymbol{\mu}_{\mathbf{A}}^{(j)}}^T \\
&\quad + \boldsymbol{\mu}_{\mathbf{A}}^{(j)} \left(\mathbf{S}_{xx} + \mathbf{\Pi}^{-1}\right) {\boldsymbol{\mu}_{\mathbf{A}}^{(j)}}^T \\
\mathbf{E}_3 &= \mathbf{D}_{\mathbf{R}} + \boldsymbol{\mu}_{\mathbf{G}} \mathbf{\Pi}^{-1} \boldsymbol{\mu}_{\mathbf{G}}^T + \mathbf{S}_{yy} - \boldsymbol{\mu}_{\mathbf{G}}^{(j)} \left(\mathbf{S}_{yx} + \boldsymbol{\mu}_{\mathbf{G}} \mathbf{\Pi}^{-1}\right)^T - \left(\mathbf{S}_{yx} + \boldsymbol{\mu}_{\mathbf{G}} \mathbf{\Pi}^{-1}\right) {\boldsymbol{\mu}_{\mathbf{G}}^{(j)}}^T \\
&\quad + \boldsymbol{\mu}_{\mathbf{G}}^{(j)} \left(\mathbf{S}_{xx} + \mathbf{\Pi}^{-1}\right) {\boldsymbol{\mu}_{\mathbf{G}}^{(j)}}^T.
\end{aligned} \tag{4.44}$$

The entropy term $\mathcal{H}\left(q^{(j)}(\boldsymbol{X}, \boldsymbol{\theta})\right)$ can be written as entropy over a product of independent random variables:

$$\mathcal{H}^{(j)} = \mathcal{H}\left(q^{(j)}(\mathbf{A} \mid \mathbf{Q})\, q^{(j)}(\mathbf{Q})\, q^{(j)}(\mathbf{G} \mid \mathbf{R})\, q^{(j)}(\mathbf{R})\, q^{(j)}(\boldsymbol{X})\right). \tag{4.45}$$

Then, $\mathcal{H}\left(q^{(j)}\left(\boldsymbol{X},\boldsymbol{\theta}\right)\right)$ can be computed using the additivity property of differential entropy for independent variables [110]:

$$\mathcal{H}\left(q^{(j)}\left(\boldsymbol{X},\boldsymbol{\theta}\right)\right)=\mathcal{H}\left(q^{(j)}\left(\mathbf{A}\mid\mathbf{Q}\right)\right)+\mathcal{H}\left(q^{(j)}\left(\mathbf{Q}\right)\right)+\mathcal{H}\left(q^{(j)}\left(\mathbf{G}\mid\mathbf{R}\right)\right)+\mathcal{H}\left(q^{(j)}\left(\mathbf{R}\right)\right)+\mathcal{H}\left(q^{(j)}\left(\boldsymbol{X}\right)\right). \tag{4.46}$$

The expression for the component-wise entropies are given as:

$$\mathcal{H}\left(q^{(j)}\left(\mathbf{A}\mid\mathbf{Q}\right)\right)=\frac{1}{2}\ln\left|\boldsymbol{\Pi}^{(j)}\otimes\mathbf{Q}^{(j)}\right|+\frac{n_s^2}{2}\left(1+\ln(2\pi)\right) \tag{4.47a}$$

$$\mathcal{H}\left(q^{(j)}\left(\mathbf{Q}\right)\right)=\ln\Gamma_{n_s}\left(\frac{d_{\mathbf{Q}}^{(j)}}{2}\right)+\frac{d_{\mathbf{Q}}^{(j)}n_s}{2}+\frac{n_s+1}{2}\ln\left|\frac{\mathbf{D}_{\mathbf{Q}}^{(j)}}{2}\right|-\frac{d_{\mathbf{Q}}^{(j)}+n_s+1}{2}\sum_{i=1}^{n_s}\gamma\left(\frac{d_{\mathbf{Q}}^{(j)}-i+1}{2}\right) \tag{4.47b}$$

$$\mathcal{H}\left(q^{(j)}\left(\mathbf{G}\mid\mathbf{R}\right)\right)=\frac{1}{2}\ln\left|\boldsymbol{\Pi}^{(j)}\otimes\mathbf{R}^{(j)}\right|+\frac{n_sn_o}{2}\left(1+\ln(2\pi)\right) \tag{4.47c}$$

$$\mathcal{H}\left(q^{(j)}\left(\mathbf{R}\right)\right)=\ln\Gamma_{n_o}\left(\frac{d_{\mathbf{R}}^{(j)}}{2}\right)+\frac{d_{\mathbf{R}}^{(j)}n_o}{2}+\frac{n_o+1}{2}\ln\left|\frac{\mathbf{D}_{\mathbf{R}}^{(j)}}{2}\right|-\frac{d_{\mathbf{R}}^{(j)}+n_o+1}{2}\sum_{i=1}^{n_o}\gamma\left(\frac{d_{\mathbf{R}}^{(j)}-i+1}{2}\right) \tag{4.47d}$$

$$\mathcal{H}\left(q^{(j)}\left(\boldsymbol{X}\right)\right)=\frac{1}{2}\left[\ln\left|\hat{\mathbf{V}}_{1|N}\right|+\sum_{k=1}^{N}\left(\ln\left|\hat{\mathbf{V}}_{k+1|N}-\hat{\mathbf{V}}_{k+1,k|N}\hat{\mathbf{V}}_{k|N}^{-1}\hat{\mathbf{V}}_{k+1,k|N}^{T}\right|\right)+c_7\right] \tag{4.47e}$$

where the constant $c_7=\left(Nn_s^2-(N-1)n_s\right)\left(1+\ln(2\pi)\right)$ and '$\otimes$' denotes the Kronecker product.

The total differential entropy $\mathcal{H}\left(q^{(j)}\left(\boldsymbol{X},\boldsymbol{\theta}\right)\right)$ is obtained by substituting Eq.(4.47) in Eq.(4.46). Finally, the variational free energy $\mathcal{F}_{ve}^{(j)}$ as a sum of Eq.(4.42) and (4.46) can be calculated as in Eq.(4.41). This completes the calculation of the variational free energy at each VB iteration.

4.5.5 Initialization and convergence criterion for VB

The VB algorithm is iterative in nature, and the the hyperparameters $\boldsymbol{\theta}_h$ of the parameter distributions are updated iteratively using Eqs.(4.38)-(4.40). For commencing the VB algorithm, the prior as well as the initial variational distributions are needed. First, the

hyperparameters of the prior distributions $\boldsymbol{\pi}_h = \{\boldsymbol{\mu}_{\mathbf{A}}, \boldsymbol{\mu}_{\mathbf{G}}, \boldsymbol{\Pi}, d_{\mathbf{Q}}, d_{\mathbf{R}}, \mathbf{D}_{\mathbf{Q}}, \mathbf{D}_{\mathbf{R}}\}$ are set as follows:

$$\boldsymbol{\mu}_{\mathbf{A}} = \mathbf{A}^{(\mathrm{SSI})}, \qquad \boldsymbol{\mu}_{\mathbf{G}} = \mathbf{G}^{(\mathrm{SSI})} \tag{4.48a}$$

$$\boldsymbol{\Pi} = 10^{15} \times \mathbf{I}_{n_s \times n_s}, \qquad \mathbf{D}_{\mathbf{Q}} = \mathbf{I}_{n_s \times n_s}, \qquad \mathbf{D}_{\mathbf{R}} = \mathbf{I}_{n_o \times n_o} \tag{4.48b}$$

$$d_{\mathbf{Q}} = n_s + 2, \qquad d_{\mathbf{R}} = n_o + 2. \tag{4.48c}$$

Note the means of prior distributions of $\mathbf{A}$ and $\mathbf{G}$ are set equal to their SSI estimates and a large value ($\sim 10^{15}$) of covariance is assigned to make the prior sufficiently flat (and non-informative). Next, the hyperparameters of initial variational distribution $\boldsymbol{\theta}_h^{(0)}$ in Eq.(4.29) are set equal to the hyperparameters of the prior distributions.

$$\boldsymbol{\mu}_{\mathbf{A}}^{(0)} = \mathbf{A}^{(\mathrm{SSI})}, \qquad \boldsymbol{\mu}_{\mathbf{G}}^{(0)} = \mathbf{G}^{(\mathrm{SSI})} \tag{4.49a}$$

$$\mathbf{D}_{\mathbf{Q}}^{(0)} = \mathbf{I}_{n_s \times n_s}, \qquad \mathbf{D}_{\mathbf{R}}^{(0)} = \mathbf{I}_{n_o \times n_o} \tag{4.49b}$$

$$d_{\mathbf{Q}}^{(0)} = n_s + 2, \qquad d_{\mathbf{R}}^{(0)} = n_o + 2 \tag{4.49c}$$

$$\boldsymbol{\mu}_1^{(0)} = \mathbf{0}_{n_s}, \qquad \mathbf{V}_1^{(0)} = \mathbf{I}_{n_s \times n_s}. \tag{4.49d}$$

The convergence criterion for VB algorithm is based on the change in the variational free energy $\mathcal{F}_{ve}^{(j)}$ between iterations and is given by:

$$\frac{\mathcal{F}_{ve}^{(j+1)} - \mathcal{F}_{ve}^{(j)}}{\left(\mathcal{F}_{ve}^{(j+1)} + \mathcal{F}_{ve}^{(j)}\right)/2} \leq \epsilon_{tol}. \tag{4.50}$$

Note that $\mathcal{F}_{ve}^{(j)}$ is calculated using Eq.(4.41). Typically, ϵ_{tol} is taken in the range of $10^{-4} \sim 10^{-6}$ [57]. The VB iterations are stopped once the convergence criterion is satisfied, and the hyperparameters obtained at the final iteration $\boldsymbol{\theta}_h^* = \left\{\boldsymbol{\mu}_{\mathbf{A}}^*, \boldsymbol{\mu}_{\mathbf{G}}^*, \boldsymbol{\Pi}^*, \mathbf{D}_{\mathbf{Q}}^*, \mathbf{D}_{\mathbf{R}}^*, d_{\mathbf{Q}}^*, d_{\mathbf{R}}^*\right\}$ are treated as the converged hyperparameters governing the posterior distribution of the parameters $\boldsymbol{\theta}$.

4.5.6 Posterior distributions of modal parameters from VB

The VB algorithm outputs the hyperparameters governing the posterior distributions of the variables $\mathbf{A}, \mathbf{G}, \mathbf{Q}, \mathbf{R}$. The marginal distributions of $\mathbf{A}$ and $\mathbf{G}$, in theory, follows matrix

variate-t distributions [111] and as the degrees of freedom $d^*_{\mathbf{Q}}$ and $d^*_{\mathbf{R}}$ increase, the matrix variate-t distributions of $\mathbf{A}$ and $\mathbf{G}$ converge to matrix normal distributions. In this study, the degrees of freedom $d^*_{\mathbf{Q}}$ and $d^*_{\mathbf{R}}$ are fairly large due to large number of time samples and hence the posterior distributions of $\mathbf{A}$ and $\mathbf{G}$ can be considered matrix normal. However, the posterior distribution of the modal parameters do not have a closed-form solution due to required eigenvalue transformation. Following the work of Li and Kiureghian [57], a first-order Taylor series expansion about the posterior means $\boldsymbol{\mu}^*_{\mathbf{A}}$ and $\boldsymbol{\mu}^*_{\mathbf{G}}$ is used to approximate the posterior distributions of the modal parameters. This approximation is expected to provide a good estimation, because the joint distribution is generally unimodal and the coefficient of variation of the modal parameters are usually small. Since the modal frequencies and damping ratios are always positive for stable structures, the first-order Taylor expansion is operated on their logarithms:

$$\ln f_i(\mathbf{A}) \approx \ln f_i(\boldsymbol{\mu}^*_{\mathbf{A}}) + \left.\frac{\partial \ln f_i(\mathbf{A})}{\partial \mathrm{vec}\,(\mathbf{A})}\right|_{\mathbf{A}=\boldsymbol{\mu}^*_{\mathbf{A}}} (\mathrm{vec}\,(\mathbf{A}) - \mathrm{vec}\,(\boldsymbol{\mu}^*_{\mathbf{A}})) \tag{4.51a}$$

$$\ln \xi_i(\mathbf{A}) \approx \ln \xi_i(\boldsymbol{\mu}^*_{\mathbf{A}}) + \left.\frac{\partial \ln \xi_i(\mathbf{A})}{\partial \mathrm{vec}\,(\mathbf{A})}\right|_{\mathbf{A}=\boldsymbol{\mu}^*_{\mathbf{A}}} (\mathrm{vec}\,(\mathbf{A}) - \mathrm{vec}\,(\boldsymbol{\mu}^*_{\mathbf{A}})) \tag{4.51b}$$

$$\begin{aligned}\boldsymbol{\phi}^{un}_i(\mathbf{A},\mathbf{G}) \approx \boldsymbol{\phi}^{un}_i(\boldsymbol{\mu}^*_{\mathbf{A}},\boldsymbol{\mu}^*_{\mathbf{G}}) &+ \left.\frac{\partial \boldsymbol{\phi}^{un}_i(\mathbf{A},\mathbf{G})}{\partial \mathrm{vec}\,(\mathbf{A})}\right|_{\mathbf{A}=\boldsymbol{\mu}^*_{\mathbf{A}}} (\mathrm{vec}\,(\mathbf{A}) - \mathrm{vec}\,(\boldsymbol{\mu}^*_{\mathbf{A}})) \\ &+ \left.\frac{\partial \boldsymbol{\phi}^{un}_i(\mathbf{A},\mathbf{G})}{\partial \mathrm{vec}\,(\mathbf{G})}\right|_{\mathbf{G}=\boldsymbol{\mu}^*_{\mathbf{G}}} (\mathrm{vec}\,(\mathbf{G}) - \mathrm{vec}\,(\boldsymbol{\mu}^*_{\mathbf{G}}))\,.\end{aligned} \tag{4.51c}$$

Here f_i, ξ_i and $\boldsymbol{\phi}^{un}_i$ are the ith natural frequency, damping ratio and un-normalized mode shape respectively, and the first-order partial derivatives form the Jacobian matrices which define the sensitivity of the modal parameters with respect to the elements of $\mathbf{A}$ and $\mathbf{G}$. The calculation of the Jacobian matrices are detailed in Appendix B.2.

The vectors $\mathrm{vec}\,(\mathbf{A})$ and $\mathrm{vec}\,(\mathbf{G})$ are approximately normally distributed with mean $\mathrm{vec}\,(\boldsymbol{\mu}^*_{\mathbf{A}})$ and $\mathrm{vec}\,(\boldsymbol{\mu}^*_{\mathbf{G}})$, and covariances $\frac{\boldsymbol{\Pi}^*\otimes\mathbf{D}^*_{\mathbf{Q}}}{d^*_{\mathbf{Q}}-n_s-1}$ and $\frac{\boldsymbol{\Pi}^*\otimes\mathbf{D}^*_{\mathbf{R}}}{d^*_{\mathbf{R}}-n_o-1}$ respectively, where '$\otimes$' stands for the Kronecker product. The modal parameters are expressed approximately as linear equations in Eq.(4.51), which implies that the ith identified modal parameters will follow multivariate normal distributions:

$$\ln f_i(\mathbf{A}) \sim \mathcal{N}\left(\mu_{\ln f_i},\ \sigma^2_{\ln f_i}\right) \tag{4.52a}$$

$$\ln \xi_i(\mathbf{A}) \sim \mathcal{N}\left(\mu_{\ln \xi_i},\ \sigma^2_{\ln \xi_i}\right) \tag{4.52b}$$

$$\boldsymbol{\phi}_i^{un}(\mathbf{A}, \mathbf{G}) \sim \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{\phi}_i^{un}},\ \boldsymbol{\Sigma}_{\boldsymbol{\phi}_i^{un}}\right) \tag{4.52c}$$

with

$$\mu_{\ln f_i} = \ln f_i(\boldsymbol{\mu}^*_{\mathbf{A}}) \tag{4.53a}$$

$$\mu_{\ln \xi_i} = \ln \xi_i(\boldsymbol{\mu}^*_{\mathbf{A}}) \tag{4.53b}$$

$$\boldsymbol{\mu}_{\boldsymbol{\phi}_i^{un}} = \boldsymbol{\phi}_i^{un}(\boldsymbol{\mu}^*_{\mathbf{A}}, \boldsymbol{\mu}^*_{\mathbf{G}}) \tag{4.53c}$$

$$\sigma^2_{\ln f_i} = \left[\frac{\partial \ln f_i(\mathbf{A})}{\partial \mathrm{vec}\,(\mathbf{A})}\bigg|_{\mathbf{A}=\boldsymbol{\mu}^*_{\mathbf{A}}}\right]\left[\frac{\boldsymbol{\Pi}^* \otimes \mathbf{D}^*_{\mathbf{Q}}}{d^*_{\mathbf{Q}} - n_s - 1}\right]\left[\frac{\partial \ln f_i(\mathbf{A})}{\partial \mathrm{vec}\,(\mathbf{A})}\bigg|_{\mathbf{A}=\boldsymbol{\mu}^*_{\mathbf{A}}}\right]^T \tag{4.53d}$$

$$\sigma^2_{\ln \xi_i} = \left[\frac{\partial \ln \xi_i(\mathbf{A})}{\partial \mathrm{vec}\,(\mathbf{A})}\bigg|_{\mathbf{A}=\boldsymbol{\mu}^*_{\mathbf{A}}}\right]\left[\frac{\boldsymbol{\Pi}^* \otimes \mathbf{D}^*_{\mathbf{Q}}}{d^*_{\mathbf{Q}} - n_s - 1}\right]\left[\frac{\partial \ln \xi_i(\mathbf{A})}{\partial \mathrm{vec}\,(\mathbf{A})}\bigg|_{\mathbf{A}=\boldsymbol{\mu}^*_{\mathbf{A}}}\right]^T \tag{4.53e}$$

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{\phi}_i^{un}} = &\left[\frac{\partial \boldsymbol{\phi}_i^{un}(\mathbf{A}, \mathbf{G})}{\partial \mathrm{vec}\,(\mathbf{A})}\bigg|_{\mathbf{A}=\boldsymbol{\mu}^*_{\mathbf{A}}}\right]\left[\frac{\boldsymbol{\Pi}^* \otimes \mathbf{D}^*_{\mathbf{Q}}}{d^*_{\mathbf{Q}} - n_s - 1}\right]\left[\frac{\partial \boldsymbol{\phi}_i^{un}(\mathbf{A}, \mathbf{G})}{\partial \mathrm{vec}\,(\mathbf{A})}\bigg|_{\mathbf{A}=\boldsymbol{\mu}^*_{\mathbf{A}}}\right]^T + \\ &\left[\frac{\partial \boldsymbol{\phi}_i^{un}(\mathbf{A}, \mathbf{G})}{\partial \mathrm{vec}\,(\mathbf{G})}\bigg|_{\mathbf{G}=\boldsymbol{\mu}^*_{\mathbf{G}}}\right]\left[\frac{\boldsymbol{\Pi}^* \otimes \mathbf{D}^*_{\mathbf{R}}}{d^*_{\mathbf{R}} - n_o - 1}\right]\left[\frac{\partial \boldsymbol{\phi}_i^{un}(\mathbf{A}, \mathbf{G})}{\partial \mathrm{vec}\,(\mathbf{G})}\bigg|_{\mathbf{G}=\boldsymbol{\mu}^*_{\mathbf{G}}}\right]^T.\end{aligned} \tag{4.53f}$$

Any arbitrary normalized mode shape $\boldsymbol{\phi}_i^n$ can be obtained from un-normalized mode shape $\boldsymbol{\phi}_i^{un}$ by scaling with $\nu_i \in \mathbb{R}$ (i.e., $\boldsymbol{\phi}_i^n = \nu_i \boldsymbol{\phi}_i^{un}$), and the distribution of new arbitrarily normalized mode shape would be given by

$$\boldsymbol{\phi}_i^n(\mathbf{A}, \mathbf{G}) \sim \mathcal{N}\left(\nu_i \boldsymbol{\mu}_{\boldsymbol{\phi}_i^{un}}(\boldsymbol{\mu}^*_{\mathbf{A}}, \boldsymbol{\mu}^*_{\mathbf{G}}),\ \nu_i^2 \boldsymbol{\Sigma}_{\boldsymbol{\phi}_i^{un}}\right). \tag{4.54}$$

Since the logarithms of the modal frequency and damping ratio follow normal distributions, the modal frequency and damping ratio have lognormal distributions and can be converted to equivalent normal distributions using transformation:

$$\begin{aligned}\textbf{Mean}: \mu_{f_i} &= \exp\left(\mu_{\ln f_i} + \frac{\sigma^2_{\ln f_i}}{2}\right), \quad \mu_{\xi_i} = \exp\left(\mu_{\ln \xi_i} + \frac{\sigma^2_{\ln \xi_i}}{2}\right) \\ \textbf{Variance}: \sigma^2_{f_i} &= \exp\left(2\mu_{\ln f_i} + \sigma^2_{\ln f_i}\right)\exp\left(\sigma^2_{\ln f_i} - 1\right), \\ \sigma^2_{\xi_i} &= \exp\left(2\mu_{\ln \xi_i} + \sigma^2_{\ln \xi_i}\right)\exp\left(\sigma^2_{\ln \xi_i} - 1\right).\end{aligned} \tag{4.55}$$

The mode shapes, however, follow multivariate normal distributions. The complete list of steps for applying the VB algorithm to the missing data case is provided in Algorithm 2.

Algorithm 2 Modal parameter estimation with VB using missing data

Input: $\boldsymbol{Y}$, ϵ_{tol}, $\left\{\mho_k^{(1)}, \mho_k^{(2)}, \mathbf{I}_k^{(1)}, \mathbf{I}_k^{(2)}\right\}_{1:N}$

Set prior hyperparameter $\boldsymbol{\pi}_h$ according to Eq.(4.48)

Initialize hyperparameters $\boldsymbol{\theta}_h^{(0)} = \left\{\boldsymbol{\mu}_{\mathbf{A}}^{(0)}, \boldsymbol{\mu}_{\mathbf{G}}^{(0)}, \mathbf{D}_{\mathbf{Q}}^{(0)}, \mathbf{D}_{\mathbf{R}}^{(0)}, \boldsymbol{\mu}_1^{(0)}, \mathbf{V}_1^{(0)}\right\}$ using Eq.(4.49)

VB E-step:

1. Set $\mathbf{A}^{(0)}$, $\mathbf{G}^{(0)}$, $\mathbf{Q}^{(0)}$ and $\mathbf{R}^{(0)}$ to the mean values of their respective distributions
2. Use Eqs.(4.13) and (4.15) for Kalman filter and RTS smoother
3. Calculate the expectations of sufficient statistics using Eqs.(4.36) and (4.37)

while not converged **do**

1. **VB M**-step: Use Eqs.(4.38), (4.39) and (4.40) to update parameters
   $\rightarrow \boldsymbol{\mu}_{\mathbf{A}}^{(j+1)}, \boldsymbol{\mu}_{\mathbf{G}}^{(j+1)}, \boldsymbol{\Pi}^{(j+1)}, \mathbf{D}_{\mathbf{Q}}^{(j+1)}, \mathbf{D}_{\mathbf{R}}^{(j+1)}, d_{\mathbf{Q}}^{(j+1)}, d_{\mathbf{R}}^{(j+1)}, \boldsymbol{\mu}_1^{(j+1)}, \mathbf{V}_1^{(j+1)}$
2. **VB E**-step:
   (a) Run Kalman filter and RTS smoother using Eqs.(4.32) – (4.35)
   (b) Calculate the expectations of sufficient statistics usingEqs.(4.36) and (4.37)
3. Use Eqs.(4.41), (4.42) and (4.46) to compute $\mathcal{F}_{ve}^{(j+1)}$
4. Use Eq.(4.50) to check if convergence criterion is satisfied

end while

Return $\boldsymbol{\mu}_{\mathbf{A}}^* \leftarrow \boldsymbol{\mu}_{\mathbf{A}}^{(final)}$, $\boldsymbol{\mu}_{\mathbf{G}}^* \leftarrow \boldsymbol{\mu}_{\mathbf{G}}^{(final)}$, $\mathbf{D}_{\mathbf{Q}}^* \leftarrow \mathbf{D}_{\mathbf{Q}}^{(final)}$, $\mathbf{D}_{\mathbf{R}}^* \leftarrow \mathbf{D}_{\mathbf{R}}^{(final)}$, $\boldsymbol{\Pi}^* \leftarrow \boldsymbol{\Pi}^{(final)}$, $d_{\mathbf{Q}}^* \leftarrow d_{\mathbf{Q}}^{(final)}$, $d_{\mathbf{R}}^* \leftarrow d_{\mathbf{R}}^{(final)}$

Set $\mathbf{A} = \boldsymbol{\mu}_{\mathbf{A}}^*$ and $\mathbf{G} = \boldsymbol{\mu}_{\mathbf{G}}^*$. Then extract the modal parameters i.e. frequency f_i, damping ratio ξ_i, un-normalized mode shape $\boldsymbol{\phi}_i^{un}$, using Appendix B.1

for $f_i, \xi_i, \boldsymbol{\phi}_i^{un}$ **do**

Compute the sensitivities using Eq.(B.6) and Appendix B.3

Compute the means and covariances of the posterior distribution of the identified modal parameters using Eqs.(4.52) – (4.55)

end for

Output: $\mu_{f_i}, \mu_{\xi_i}, \boldsymbol{\mu}_{\boldsymbol{\phi}_i^{un}}, \sigma_{f_i}^2, \sigma_{\xi_i}^2, \boldsymbol{\Sigma}_{\boldsymbol{\phi}_i^{un}}$ for all modes $i = 1, 2, \ldots$

4.6 Bayesian estimation of LGSSM via GS

Another inference method for the LGSSM is formulated using a Gibbs Sampler (GS). The main advantage of using a GS is the ability to achieve arbitrarily accurate approximation of the posterior distribution (i.e., converges to the true posterior distribution with an arbitrarily defined error); however, this comes with a cost of heavier computational burden.

In this section, the standard GS [59, 88] is used within a *multiple imputation* framework for Bayesian inference with missing data. This method will serve as a baseline for comparison of the different approaches in this study. Multiple imputation [60, 112, 113] in combination with GS is based on the premise of "filling in" the missing data values and subsequently applying Bayesian analysis to the complete data matrix.

4.6.1 Multiple imputation procedure

Multiple imputation using GS shares the same underlying philosophy as EM and VB: solving an incomplete-data problem by repeatedly solving the complete-data version. Multiple imputation assumes that the missing data are missing at random (MAR), that is, the probability that an observation is missing may depend on $\boldsymbol{Y}_{obs}$ but not on $\boldsymbol{Y}_{miss}$. This MAR assumption holds true for the missing data matrix generated from in-motion mobile sensing because the pattern of missingness depends only on the path of the mobile sensors and does not depend on the missing responses $\boldsymbol{Y}_{miss}$. In multiple imputation, the unknown missing dataset, $\boldsymbol{Y}_{miss}$, is replaced by imputed datasets $\boldsymbol{Y}_{miss}^{(1)}, \boldsymbol{Y}_{miss}^{(2)}, \ldots, \boldsymbol{Y}_{miss}^{(D)}$. Each of the D completed datasets is analyzed by standard complete-data methods. The variability among the results of the D analyses provides a measure of the uncertainty due to missing data.

The multiple imputation procedure follows three steps. First, for every missing entry in the data, imputation is conducted by drawing samples from the posterior predictive distribution D times, creating D sets of complete data, denoted $\boldsymbol{Y}^{(d)} = \left\{\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{miss}^{(d)}\right\}$, $d = 1, \ldots, D$. Let $\mathcal{Q}$ be scalar quantity of interest, with $\hat{\mathcal{Q}}$ as its estimator and $\mathcal{U}$ be the associated variance. With D imputations, one can calculate D different versions of $\hat{\mathcal{Q}}$ and $\mathcal{U}$. Let $\hat{\mathcal{Q}}^{(d)}$ and $\mathcal{U}^{(d)}$ be the point and variance estimates using the dth imputed dataset

$\boldsymbol{Y}^{(d)} = \left\{\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{miss}^{(d)}\right\}$, $d = 1, \ldots, D$, then valid inferences can be drawn using Rubin's combination rule [60]. The multiple imputation point estimate $\hat{\mathcal{Q}}$ is the average of the D-complete data estimates

$$\hat{\mathcal{Q}} = \frac{1}{D}\sum_{d=1}^{D}\hat{\mathcal{Q}}^{(d)}. \tag{4.56a}$$

The total variance estimate associated with $\hat{\mathcal{Q}}$ is

$$\mathcal{T} = \bar{\mathcal{U}} + \left(1 + \frac{1}{D}\right)\mathcal{B} \tag{4.56b}$$

where

$$\textit{Within-chain-imputation variance}: \quad \bar{\mathcal{U}} = \frac{1}{D}\sum_{d=1}^{D}\hat{\mathcal{U}}^{(d)} \tag{4.56c}$$

$$\textit{Between-chain-imputation variance}: \quad \mathcal{B} = \frac{1}{D-1}\sum_{d=1}^{D}\left(\hat{\mathcal{Q}}^{(d)} - \hat{\mathcal{Q}}\right)^2. \tag{4.56d}$$

The joint distribution of the unknown parameters $\boldsymbol{\theta} = \{\mathbf{A}, \mathbf{G}, \mathbf{Q}, \mathbf{R}\}$, the hidden states $\boldsymbol{X}$ and the measurements $\boldsymbol{Y} = \{\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{miss}\}$ can be written as

$$p\left(\boldsymbol{Y}, \boldsymbol{X}, \mathbf{A}, \mathbf{G}, \mathbf{Q}, \mathbf{R}\right) = p\left(\mathbf{A} \mid \mathbf{Q}\right) p\left(\mathbf{Q}\right) p\left(\mathbf{G} \mid \mathbf{R}\right) p\left(\mathbf{R}\right) p\left(\boldsymbol{x}_1\right) \tag{4.57}$$

$$\times \prod_{k=1}^{N} p\left(\boldsymbol{x}_{k+1} \mid \boldsymbol{x}_k, \mathbf{A}, \mathbf{Q}\right) p\left(\boldsymbol{y}_k \mid \boldsymbol{x}_k, \mathbf{G}, \mathbf{R}\right). \tag{4.58}$$

To apply the multiple imputation technique, one needs to obtain D completed datasets which requires running D Markov chains. The Markov chains are constructed using GS, whereby the elements of the Markov chains are drawn from fully specified conditional distributions. Given the observed measurements $\boldsymbol{Y}_{obs}$, the required conditional distributions can be derived and samples can be drawn during each iteration j of the dth Markov chain in the following cyclic fashion:

1. Given $\boldsymbol{\theta}^{(d,j)} = \left\{\mathbf{A}^{(d,j)}, \mathbf{G}^{(d,j)}, \mathbf{Q}^{(d,j)}, \mathbf{R}^{(d,j)}\right\}$ and observed measurements $\boldsymbol{Y}_{obs}$, a sequence of hidden state variables are sampled according to

$$\boldsymbol{X}^{(d,j)} \sim p\left(\boldsymbol{X} \mid \boldsymbol{Y}_{obs}, \boldsymbol{\theta}^{(d,j)}\right). \tag{4.59}$$

2. Given $\boldsymbol{X}^{(d,j)}$, a sequence of missing measurements is sampled according to

$$\boldsymbol{Y}_{miss}^{(d,j)} \sim p\left(\boldsymbol{Y}_{miss} \mid \boldsymbol{Y}_{obs}, \boldsymbol{X}^{(d,j)}, \boldsymbol{\theta}^{(d,j)}\right). \tag{4.60}$$

3. Given $\boldsymbol{X}^{(d,j)}$ and $\boldsymbol{Y}_{miss}^{(d,j)}$, the unknown parameters $\boldsymbol{\theta}^{(d,j+1)}$ are sampled according to

$$\boldsymbol{\theta}^{(d,j+1)} \sim p\left(\boldsymbol{\theta} \mid \boldsymbol{Y}_{obs}, \boldsymbol{Y}_{miss}^{(d,j)}, \boldsymbol{X}^{(d,j)}\right). \tag{4.61a}$$

4.6.2 Forward filtering backward sampling for states

The state sequence $\boldsymbol{X}$ is sampled following the *forward filtering backward sampling* strategy elaborated in Wills et al. [92] (also in [114]). The strategy consists of two parts: forward filtering and backward sampling. The forward filtering part sequentially predicts and updates the states and state-error covariances using the Kalman filter, as in Eq.(4.13), to obtain $p\left(\boldsymbol{x}_k \mid \boldsymbol{Y}_{obs}, \boldsymbol{\theta}\right)$ for all $k = 1, \ldots, N+1$. Post forward filtering, the backward sampling simulates state vectors from time $k = N+1$ to $k = 1$. The joint distribution of $\boldsymbol{x}_{k+1}$ and $\boldsymbol{x}_k$ conditioned on observed data $\boldsymbol{Y}_{obs}$ [65] can be written as

$$p\left(\boldsymbol{x}_{k+1}, \boldsymbol{x}_k \mid \boldsymbol{Y}_{obs}\right) = \mathcal{N}\left(\begin{bmatrix}\boldsymbol{x}_{k+1} \\ \boldsymbol{x}_k\end{bmatrix} \middle| \tilde{\boldsymbol{m}}_1, \tilde{\mathbf{P}}_1\right) \tag{4.62}$$

where

$$\begin{aligned}
\tilde{\boldsymbol{m}}_1 &= \begin{bmatrix} \hat{\boldsymbol{x}}_{k+1|N} \\ \hat{\boldsymbol{x}}_{k|k} + \mathbf{N}_k\left(\hat{\boldsymbol{x}}_{k+1|N} - \mathbf{A}\hat{\boldsymbol{x}}_{k|k}\right)\end{bmatrix} \\
\tilde{\mathbf{P}}_1 &= \begin{bmatrix} \hat{\mathbf{V}}_{k+1|N} & \hat{\mathbf{V}}_{k+1|N}\mathbf{N}_k^T \\ \mathbf{N}_k\hat{\mathbf{V}}_{k+1|N}^T & \hat{\mathbf{V}}_{k|k} + \mathbf{N}_k\left(\hat{\mathbf{V}}_{k+1|N} - \hat{\mathbf{V}}_{k+1|k}\right)\mathbf{N}_k^T\end{bmatrix}.
\end{aligned} \tag{4.63}$$

The matrix $\mathbf{N}_k = \hat{\mathbf{V}}_{k|k}\mathbf{A}^T\left(\hat{\mathbf{V}}_{k+1|k}\right)^{-1}$ is the smoother gain matrix at the k time instant, as calculated in RTS smoother Eq.(4.15a). Using the property of the multivariate normal distribution, the conditional distribution of $\boldsymbol{x}_k$ given $\boldsymbol{x}_{k+1}$ follows a multivariate normal distribution (refer Eq. (8.10) in [65])

$$p\left(\boldsymbol{x}_k \mid \boldsymbol{x}_{k+1}, \boldsymbol{Y}_{obs}\right) = \mathcal{N}\left(\boldsymbol{x}_k \mid \tilde{\boldsymbol{m}}_2, \tilde{\mathbf{P}}_2\right) \tag{4.64}$$

where

$$\begin{aligned} \tilde{\boldsymbol{m}}_2 &= \hat{\boldsymbol{x}}_{k|k} + \mathbf{N}_k\left(\boldsymbol{x}_{k+1} - \mathbf{A}\hat{\boldsymbol{x}}_{k|k}\right) \\ \tilde{\mathbf{P}}_2 &= \hat{\mathbf{V}}_{k+1|k} - \mathbf{N}_k\hat{\mathbf{V}}_{k+1|k}\mathbf{N}_k^T. \end{aligned} \tag{4.65}$$

For the jth iteration of the dth Markov chain, given $\boldsymbol{x}_{k+1}^{(d,j)}$, one successfully samples $\boldsymbol{x}_k^{(d,j)}$ from the multivariate normal distribution in Eq.(4.64). Thus, given $\mathbf{A}^{(d,j)}, \mathbf{G}^{(d,j)}, \mathbf{Q}^{(d,j)}, \mathbf{R}^{(d,j)}$ and $\boldsymbol{Y}_{obs}$, one can generate the sequence $\boldsymbol{X}^{(d,j)}$.

4.6.3 Sampling the missing observations

Given $\mathbf{A}$, $\mathbf{G}$, $\mathbf{Q}$, $\mathbf{R}$ and the state sequence $\boldsymbol{X}$, the conditional distribution of $\boldsymbol{y}_k = \left(\boldsymbol{y}_k^{obs}, \boldsymbol{y}_k^{miss}\right)$ can be written as

$$\begin{bmatrix} \boldsymbol{y}_k^{obs} \mid \boldsymbol{x}_k \\ \boldsymbol{y}_k^{miss} \mid \boldsymbol{x}_k \end{bmatrix} = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\upsilon}_k^{(1)}\mathbf{G}\boldsymbol{x}_k \\ \boldsymbol{\upsilon}_k^{(2)}\mathbf{G}\boldsymbol{x}_k \end{bmatrix}, \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{21}^T \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{bmatrix}\right)$$

where $\boldsymbol{\upsilon}_k^{(1)}$ and $\boldsymbol{\upsilon}_k^{(2)}$ are matrices defined based on the missingness pattern in $\boldsymbol{Y}$, and $\mathbf{R}_{11} = \boldsymbol{\upsilon}_k^{(1)}\mathbf{R}\boldsymbol{\upsilon}_k^{(1)^T}$, $\mathbf{R}^{12} = \boldsymbol{\upsilon}_k^{(1)}\mathbf{R}\boldsymbol{\upsilon}_k^{(2)^T}$, $\mathbf{R}^{21} = \boldsymbol{\upsilon}_k^{(2)}\mathbf{R}\boldsymbol{\upsilon}_k^{(1)^T}$ and $\mathbf{R}^{22} = \boldsymbol{\upsilon}_k^{(2)}\mathbf{R}\boldsymbol{\upsilon}_k^{(2)^T}$. Using the property of conditional distributions for multivariate normal distribution, one can obtain

$$\left(\boldsymbol{y}_k^{miss} \mid \boldsymbol{y}_k^{obs}, \boldsymbol{x}_k\right) \sim \mathcal{N}\left(\ddot{\boldsymbol{\mu}}, \ddot{\boldsymbol{\Sigma}}\right) \tag{4.66}$$

where

$$\ddot{\boldsymbol{\mu}} = \boldsymbol{\upsilon}_k^{(2)}\mathbf{G}\boldsymbol{x}_k + \mathbf{R}_{21}\mathbf{R}_{11}^{-1}\boldsymbol{\upsilon}_k^{(1)}\left(\boldsymbol{y}_k - \mathbf{G}\boldsymbol{x}_k\right)$$

$$= \boldsymbol{\upsilon}_k^{(2)} \left(\boldsymbol{y}_k - \mathbf{Z}_k \left(\boldsymbol{y}_k - \mathbf{G} \hat{\boldsymbol{x}}_{k|N} \right) \right) \tag{4.67a}$$

$$\ddot{\boldsymbol{\Sigma}} = \mathbf{R}_{22} - \mathbf{R}_{21} \mathbf{R}_{11}^{-1} \mathbf{R}_{21}^T$$

$$= \boldsymbol{\upsilon}_k^{(2)} \mathbf{Z}_k \mathbf{R} {\boldsymbol{\upsilon}_k^{(2)}}^T \tag{4.67b}$$

and $\mathbf{Z}_k = \mathbf{I} - \mathbf{R} \boldsymbol{\upsilon}_k^{(1)^T} \left(\boldsymbol{\upsilon}_k^{(1)} \mathbf{R} \boldsymbol{\upsilon}_k^{(1)^T} \right)^{-1} \boldsymbol{\upsilon}_k^{(1)}$. Therefore, given $\boldsymbol{X}^{(d,j)}$ and $\boldsymbol{\theta}^{(d,j)}$, one can sample $\boldsymbol{Y}_{miss}^{(d,j)}$ using Eqs.(4.66) and (4.67).

Sampling unknown parameters

The posterior distribution of the parameters $\boldsymbol{\theta} = \{\mathbf{A}, \mathbf{G}, \mathbf{Q}, \mathbf{R}\}$ given a sample of the state sequence $\boldsymbol{X}^{(d,j)}$ and the complete measurements $\boldsymbol{Y}^{(d,j)} = \left\{ \boldsymbol{Y}_{obs}, \boldsymbol{Y}_{miss}^{(d,j)} \right\}$ is given by

$$p\left(\boldsymbol{\theta} \mid \boldsymbol{X}^{(d,j)}, \boldsymbol{Y}^{(d,j)} \right) \propto p\left(\boldsymbol{X}^{(d,j)}, \boldsymbol{Y}^{(d,j)} \mid \boldsymbol{\theta} \right) p\left(\boldsymbol{\theta} \right). \tag{4.68}$$

Since the prior $p\left(\boldsymbol{\theta} \right)$ is assumed to be described by a MNIW distribution, the posterior distribution $p\left(\boldsymbol{\theta} \mid \boldsymbol{X}^{(d,j)}, \boldsymbol{Y}^{(d,j)} \right)$ is also a MNIW distribution for the LGSSM, and therefore one can readily obtain samples of parameters from it. More specifically,

$$\begin{aligned} p\left(\mathbf{A}, \mathbf{Q} \right) &= p\left(\mathbf{A} \mid \mathbf{Q} \right) p\left(\mathbf{Q} \right) \\ p\left(\mathbf{G}, \mathbf{R} \right) &= p\left(\mathbf{G} \mid \mathbf{R} \right) p\left(\mathbf{R} \right). \end{aligned} \tag{4.69}$$

With the prior distribution in place, the logarithm of joint likelihood term can be written as

$$\begin{aligned} \ln p\left(\boldsymbol{X}^{(d,j)}, \boldsymbol{Y}^{(d,j)} \mid \boldsymbol{\theta} \right) \propto\ & N \ln |\mathbf{Q}| + N \ln |\mathbf{R}| \\ & + \operatorname{tr}\left\{ \mathbf{Q}^{-1} \left(\mathbf{P}_{ff}^{(d,j)} - \mathbf{A} \left(\mathbf{P}_{fx}^{(d,j)} \right)^T - \mathbf{P}_{fx}^{(d,j)} \mathbf{A}^T + \mathbf{A} \mathbf{P}_{xx}^{(d,j)} \mathbf{A}^T \right) \right\} \\ & + \operatorname{tr}\left\{ \mathbf{R}^{-1} \left(\mathbf{P}_{yy}^{(d,j)} - \mathbf{G} \left(\mathbf{P}_{yx}^{(d,j)} \right)^T - \mathbf{P}_{yx}^{(d,j)} \mathbf{G}^T + \mathbf{G} \mathbf{P}_{xx}^{(d,j)} \mathbf{G}^T \right) \right\} \end{aligned} \tag{4.70}$$

where $\mathbf{P}_{xx}^{(d,j)}, \mathbf{P}_{fx}^{(d,j)}, \mathbf{P}_{ff}^{(d,j)}, \mathbf{P}_{yx}^{(d,j)}, \mathbf{P}_{yy}^{(d,j)}$ are the expectations of sufficient statistics described as

$$\mathbf{P}_{xx}^{(d,j)} = \sum_{k=1}^{N} \boldsymbol{x}_k^{(d,j)} \left(\boldsymbol{x}_k^{(d,j)} \right)^T, \quad \mathbf{P}_{fx}^{(d,j)} = \sum_{k=1}^{N} \boldsymbol{x}_{k+1}^{(d,j)} \left(\boldsymbol{x}_k^{(d,j)} \right)^T, \quad \mathbf{P}_{ff}^{(d,j)} = \sum_{k=1}^{N} \boldsymbol{x}_{k+1}^{(d,j)} \left(\boldsymbol{x}_{k+1}^{(d,j)} \right)^T \tag{4.71a}$$

$$\mathbf{P}_{yx}^{(d,j)} = \sum_{k=1}^{N} \boldsymbol{y}_k^{(d,j)} \left(\boldsymbol{x}_k^{(d,j)}\right)^T, \qquad \mathbf{P}_{yy}^{(d,j)} = \sum_{k=1}^{N} \boldsymbol{y}_k^{(d,j)} \left(\boldsymbol{y}_k^{(d,j)}\right)^T. \tag{4.71b}$$

It follows (see West and Harrison [115]) that the posterior distribution is MNIW and is given by

$$p\left(\mathbf{A}, \mathbf{Q} \mid \boldsymbol{X}^{(d,j)}, \boldsymbol{Y}^{(d,j)}\right) = p\left(\mathbf{A} \mid \boldsymbol{X}^{(d,j)}, \boldsymbol{Y}^{(d,j)}, \mathbf{Q}\right) p\left(\mathbf{Q} \mid \boldsymbol{X}^{(d,j)}, \boldsymbol{Y}^{(d,j)}\right) \tag{4.72a}$$

$$p\left(\mathbf{G}, \mathbf{R} \mid \boldsymbol{X}^{(d,j)}, \boldsymbol{Y}^{(d,j)}\right) = p\left(\mathbf{G} \mid \boldsymbol{X}^{(d,j)}, \boldsymbol{Y}^{(d,j)}, \mathbf{R}\right) p\left(\mathbf{R} \mid \boldsymbol{X}^{(d,j)}, \boldsymbol{Y}^{(d,j)}\right) \tag{4.72b}$$

with

$$p\left(\mathbf{Q} \mid \boldsymbol{X}^{(d,j)}, \boldsymbol{Y}^{(d,j)}\right) = \mathcal{IW}\left(\bar{d}_{\mathbf{Q}}^{(d,j+1)}, \overline{\mathbf{D}}_{\mathbf{Q}}^{(d,j+1)}\right) \tag{4.73a}$$

$$p\left(\mathbf{A} \mid \boldsymbol{X}^{(d,j)}, \boldsymbol{Y}^{(d,j)}, \mathbf{Q}^{(d,j+1)}\right) = \mathcal{MN}\left(\bar{\boldsymbol{\mu}}_{\mathbf{A}}^{(d,j+1)}, \mathbf{Q}^{(d,j+1)}, \overline{\boldsymbol{\Pi}}^{(d,j+1)}\right) \tag{4.73b}$$

$$p\left(\mathbf{R} \mid \boldsymbol{X}^{(d,j)}, \boldsymbol{Y}^{(d,j)}\right) = \mathcal{IW}\left(\bar{d}_{\mathbf{R}}^{(d,j+1)}, \overline{\mathbf{D}}_{\mathbf{R}}^{(d,j+1)}\right) \tag{4.73c}$$

$$p\left(\mathbf{G} \mid \boldsymbol{X}^{(d,j)}, \boldsymbol{Y}^{(d,j)}, \mathbf{R}^{(d,j+1)}\right) = \mathcal{MN}\left(\bar{\boldsymbol{\mu}}_{\mathbf{G}}^{(d,j+1)}, \mathbf{R}^{(d,j+1)}, \overline{\boldsymbol{\Pi}}^{(d,j+1)}\right). \tag{4.73d}$$

Here,

$$\overline{\boldsymbol{\Pi}}^{(d,j+1)} = \left(\mathbf{P}_{xx}^{(d,j)} + \boldsymbol{\Pi}^{-1}\right)^{-1} \tag{4.74a}$$

$$\bar{\boldsymbol{\mu}}_{\mathbf{A}}^{(d,j+1)} = \left(\mathbf{P}_{fx}^{(d,j)} + \boldsymbol{\mu}_{\mathbf{A}}\boldsymbol{\Pi}^{-1}\right) \overline{\boldsymbol{\Pi}}^{(d,j+1)} \tag{4.74b}$$

$$\bar{d}_{\mathbf{Q}}^{(d,j+1)} = d_{\mathbf{Q}} + N \tag{4.74c}$$

$$\overline{\mathbf{D}}_{\mathbf{Q}}^{(d,j+1)} = \mathbf{D}_{\mathbf{Q}} + \mathbf{P}_{ff}^{(d,j)} + \boldsymbol{\mu}_{\mathbf{A}}\boldsymbol{\Pi}^{-1}\boldsymbol{\mu}_{\mathbf{A}}^T - \bar{\boldsymbol{\mu}}_{\mathbf{A}}^{(d,j+1)} \left(\overline{\boldsymbol{\Pi}}^{(d,j+1)}\right)^{-1} \left(\bar{\boldsymbol{\mu}}_{\mathbf{A}}^{(d,j+1)}\right)^T \tag{4.74d}$$

$$\bar{\boldsymbol{\mu}}_{\mathbf{G}}^{(d,j+1)} = \left(\mathbf{P}_{yx}^{(d,j)} + \boldsymbol{\mu}_{\mathbf{G}}\boldsymbol{\Pi}^{-1}\right) \overline{\boldsymbol{\Pi}}^{(d,j+1)} \tag{4.74e}$$

$$\bar{d}_{\mathbf{R}}^{(d,j+1)} = d_{\mathbf{R}} + N \tag{4.74f}$$

$$\overline{\mathbf{D}}_{\mathbf{R}}^{(d,j+1)} = \mathbf{D}_{\mathbf{R}} + \mathbf{P}_{yy}^{(d,j+1)} + \boldsymbol{\mu}_{\mathbf{G}}\boldsymbol{\Pi}^{-1}\boldsymbol{\mu}_{\mathbf{G}}^T - \bar{\boldsymbol{\mu}}_{\mathbf{G}}^{(d,j+1)} \left(\overline{\boldsymbol{\Pi}}^{(d,j+1)}\right)^{-1} \left(\bar{\boldsymbol{\mu}}_{\mathbf{G}}^{(d,j+1)}\right)^T. \tag{4.74g}$$

Thus, given the state sequence $\boldsymbol{X}^{(d,j)}$ and the complete observations $\boldsymbol{Y}^{(d,j)} = \left\{\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{miss}^{(d,j)}\right\}$ at the jth GS iteration of the dth Markov chain, the updated parameters $\boldsymbol{\theta}^{(d,j+1)}$ can be sampled using Eqs.(4.71), (4.73) and (4.74). For the purposes of implementation, first $\mathbf{Q}^{(d,j+1)}$ and $\mathbf{R}^{(d,j+1)}$ are sampled from the respective inverse Wishart distributions, and

then conditioned on the sampled $\mathbf{Q}^{(d,j+1)}$ and $\mathbf{R}^{(d,j+1)}$, $\mathbf{A}^{(d,j+1)}$ and $\mathbf{G}^{(d,j+1)}$ are sampled from their respective matrix normal distributions. Once $\mathbf{A}^{(d,j+1)}$ and $\mathbf{G}^{(d,j+1)}$ are sampled, one obtains samples of modal parameters using the procedure in Appendix B.1.

4.6.4 Initialization of GS

In GS, the desired posterior distribution is obtained after the Markov chain reaches its stationary distribution. The number of samples needed for a Markov chain to reach its stationary distribution depends, to a great extent, on the percentage of missing data. The higher the percentage of missing data, the more samples required to reach stationarity and the longer the Markov chain needs to be simulated. Since D Markov chains are required for multiple imputation, the GS is run in parallel for D Markov chains, each of which is initialized with different parameter estimates from SSI. One way to generate different parameter estimates $\left\{\mathbf{A}^{\text{SSI}}, \mathbf{G}^{\text{SSI}}\right\}$ from SSI is by randomly varying the number of block rows inputted to the SSI algorithm. Another way is to obtain an estimate and then generate more estimates by randomly perturbing the parameter estimate from SSI. The initial parameter estimates can be selected in the same way as mentioned in Eq.(4.49) with the hyperparameters of the prior distributions selected using Eq.(4.48).

4.6.5 Convergence monitoring of Markov chains

The first few samples of each Markov chain are discarded to remove the transient samples; this operation is termed as *burn-in* and the number of samples used for burn-in is denoted by n_b. A major consideration in MCMC simulations is that of convergence to the stationary distribution of the Markov chain. The Gelman-Rubin diagnostic evaluates MCMC convergence by analyzing the difference between multiple Markov chains. The convergence is assessed by comparing the estimated between-chains and within-chain variances for each model parameter. Large differences between these variances indicate non-convergence. A Gelman-Rubin diagnostic value of 1.2 is typically recommended for this difference [116, 117].

4.6.6 Computation of posterior statistics

Typically, the samples of a Markov chain are correlated; to minimize the correlation between samples one may choose to subsample the Markov chain. If $\mathcal{Q}^{(d,1)}, \mathcal{Q}^{(d,2)}, \ldots, \mathcal{Q}^{(d,J)}$ denote the output from the dth Markov chain of length J after discarding n_b samples from an initial burn-in period, the averaged estimate of $\hat{\mathcal{Q}}^{(d)}$ and unbiased variance $\mathcal{U}^{(d)}$ over every bth sample of the dth Markov chain is given by

$$\hat{\mathcal{Q}}^{(d)} = \frac{1}{L}\sum_{l=1}^{L} \mathcal{Q}^{(d,bl)} \tag{4.75a}$$

$$\mathcal{U}^{(d)} = \frac{1}{L-1}\sum_{l=1}^{L} \left(\mathcal{Q}^{(d,bl)} - \hat{\mathcal{Q}}^{(d)}\right)^2 \tag{4.75b}$$

where $L = J/b$ and $d = 1, \ldots, D$. Here, $\hat{\mathcal{Q}}^{(d)}$ would represent the averaged estimates of frequencies, damping ratios and mode shapes, and $\mathcal{U}^{(d)}$ would represent their respective variances. The averaged estimates for the modal parameters using outputs for D chains are obtained using Eq.(4.56) following Rubin's combination rule.

The complete list of steps for applying the GS algorithm to the missing data case is provided in Algorithm 3.

Algorithm 3 Modal parameter estimation with GS using missing data

Input: $\boldsymbol{Y}$, ϵ_{tol}, $\left\{\mathfrak{V}_k^{(1)}, \mathfrak{V}_k^{(2)}, \mathbf{I}_k^{(1)}, \mathbf{I}_k^{(2)}\right\}_{1:N}$, number of chains D, chain length J, burn-in samples n_b, subsampling period b

Set prior hyperparameter $\boldsymbol{\pi}_h$ using Eq.(4.48) and initialize $\boldsymbol{\theta}_h^{(0)}$ using Eq.(4.49)

for Markov chain d **do**

for $j = 1 : J$ **do**

1. **Forward filtering backward sampling for states**
   (a) Use Eq.(4.13) for Kalman filter recursions
   (b) Calculate smoother gains using Eq.(4.15a)
   (c) Generate sequence $\boldsymbol{X}^{(j)}$ using Eqs.(4.64) and (4.65)
2. **Sampling missing observations**
   Use Eqs. (4.66) and (4.67) to sample missing observations $\boldsymbol{Y}_{miss}^{(j)}$
3. **Sampling parameters**
   (a) Compute expectations of sufficient statistics using Eq.(4.71)
   (b) Update the distributional hyperparameters using Eq.(4.74)
   (c) Sample the covariance parameters $\mathbf{Q}^{(d,j+1)}$ and $\mathbf{R}^{(d,j+1)}$ using probability distributions given by Eqs.(4.73a) and (4.73c) respectively
   (d) Using samples $\mathbf{Q}^{(d,j+1)}$ and $\mathbf{R}^{(d,j+1)}$ from previous step, sample the process parameters $\mathbf{A}^{(d,j+1)}$ and $\mathbf{G}^{(d,j+1)}$ using probability distributions given by Eqs.(4.73b) and (4.73d) respectively
   (e) Use Appendix B.1 to extract the modal parameters

end for

Using Eq.(4.75), compute the averaged point estimate $\hat{\mathcal{Q}}^{(d)}$ and variance $\mathcal{U}^{(d)}$ for each modal parameter for the dth Markov chain (after discarding n_b initial samples)

end for

for $f_i, \xi_i, \boldsymbol{\phi}_i^{un}$ **do**

Compute the averaged point estimate and total variance using Eq.(4.56)

end for

Output: $\mu_{f_i}, \mu_{\xi_i}, \boldsymbol{\mu}_{\boldsymbol{\phi}_i^{un}}, \sigma_{f_i}^2, \sigma_{\xi_i}^2, \boldsymbol{\Sigma}_{\boldsymbol{\phi}_i^{un}}$ for all modes $i = 1, 2, \ldots$

4.7 Numerical Demonstration

This section presents a numerical example based on an eight-DoF lumped mass beam model (adopted from Cara et al. [69]). Figure 4.2 depicts a schematic of the mass-spring-damper model. The performance of the EM, VB and GS for modal parameter estimation is demonstrated on data from in-motion mobile sensors following the missing data approach. The structural parameters are: $m = 1$ kg, $k_i = 800i$ N/m for $i = 1, \ldots, 9$ and damping matrix $\mathbf{C} = 0.68\mathbf{M} + 1.743 \times 10^{-4}\mathbf{K}$ Ns/m (Rayleigh damping), where $\mathbf{M}$ and $\mathbf{K}$ are the mass and stiffness matrices, respectively.

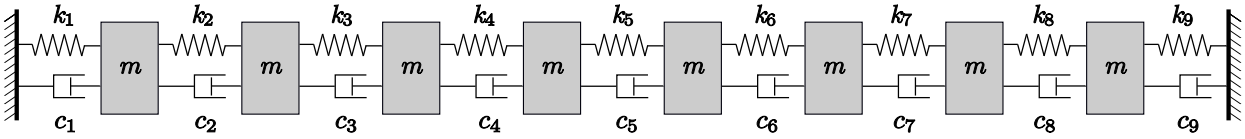


Figure 4.2: Eight-DoF mass spring damper model.

The modal frequencies and damping ratios of the above system are tabulated in Table 4.1.

Table 4.1: Modal frequencies (f) and damping ratios (ξ) of the eight-DoF model.

Mode	1	2	3	4	5	6	7	8
f (Hz)	2.942	5.870	8.602	11.188	13.780	16.519	19.536	23.188
ξ (%)	2.000	1.243	1.100	1.096	1.147	1.232	1.347	1.500

The numerical simulations of the structural responses are computed using the following:

- Identical bandlimited Gaussian white noise inputs are applied to all dofs of the lumped-mass beam model,
- Sampling period $\Delta t = 0.02$s. Total duration of 200s (10000 time steps),
- Acceleration outputs are observed at all eight vertical DoFs of the beam,

- Zero mean Gaussian white noise sequences with RMS values equal to 20% of the RMS of the corresponding "true" acceleration signals were added to obtain noisy acceleration outputs,

- A group of n_{ms} mobile sensors is assumed to scan the eight DoFs, shifting back and forth in a cyclical fashion. At every time step, the locations of the mobile sensors change and only n_{ms} observations are sampled. This leads to a missing data matrix. Figure 4.3 illustrates a missing data matrix for two mobile sensors (i.e., $n_{ms} = 2$). The velocity of the mobile sensors are assumed to be such that each mobile sensor proceeds from one DoF to the next in each sampling instant.

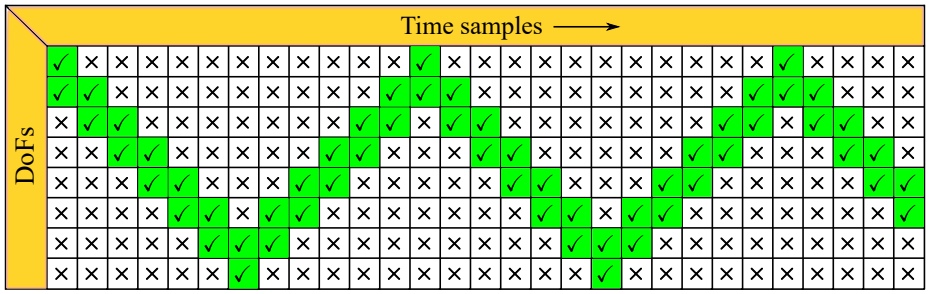


Figure 4.3: An example of missing data matrix for the eight-DoF lumped mass model when observations are sampled using two adjacent mobile sensors ($n_{ms} = 2$) moving in a cyclical fashion. The green boxes with ticks represent the mobile sensors' trajectory; the white boxes with crosses indicate positions where the data are absent (or missing).

In SSM based modal identification, the model order is first specified, which is typically twice the number of physical modes. In this case, the theoretical model order is 16 (i.e. twice the number of modes included in the acceleration responses). However, in practice, the theoretical model order would not be known in advance, and hence a larger model order is used to identify the physical modes. In this example, a model order of 20 is chosen. Furthermore, the physical modes of the structure are identified based on two conditions: (a) they should correspond to eigenvalues (of $\mathbf{A}$) that appear in complex conjugate pairs, and (b) they should have positive damping ratios below 5%. If a mode passes both these

conditions it is considered a physical mode, else it is labelled as a spurious mode and discarded from the identification process.

The averaged power spectral density (PSD) estimate from eight acceleration outputs is shown in Figure 4.3 where it can be seen that the modal energy contribution of the first mode is highest while the energies from the higher modes are diminishing.

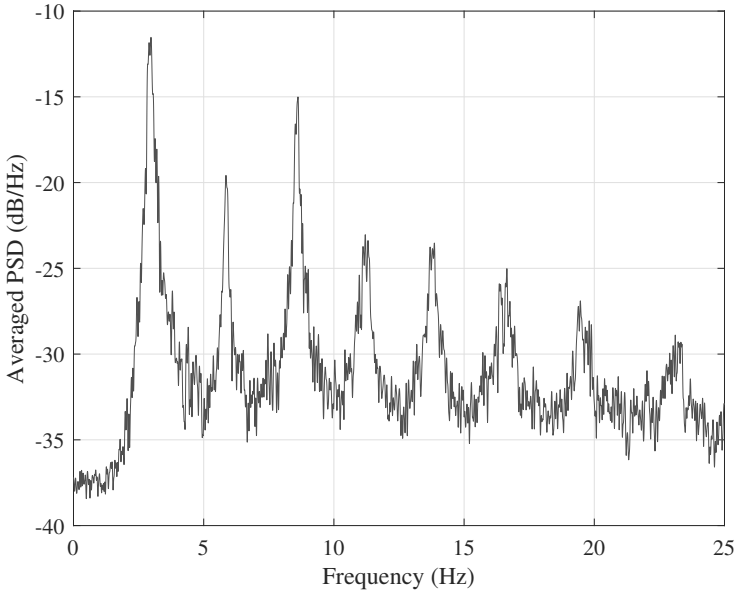


Figure 4.4: Averaged power spectral density from eight acceleration outputs.

For a comprehensive investigation of the modal parameter estimation performance, three cases of mobile sensor groups (or networks) are considered, as listed in Table 4.2. The three cases employ different number of mobile sensors for data collection, thus giving rise to incomplete data matrices with varying degrees (or percentages) of missing data.

Table 4.2: Cases of mobile sensor groups considered in this study.

Case	n_{ms}	% data missing ($^{n_{ms}}/_{8} \times 100$)
`MS2`	2	75%
`MS4`	4	50%
`MS6`	6	25%

Estimation algorithms, namely EM, VB and GS, are employed to assess the performance of the identified modal parameters using the missing data modelling approach. The following are used in the inference procedure using EM, VB and GS:

- Initial estimates $\mathbf{A}^{(0)}$ and $\mathbf{G}^{(0)}$ are obtained from SSI (with model order 20) using output responses from a slightly perturbed variant of the true system model. The perturbed system model was set up with perturbed values of frequencies, damping ratios and mode shapes having random perturbations in the range of 10%, 40% and 30%, respectively about their corresponding true values. Thus, the initial estimates $\mathbf{A}^{(0)}$ and $\mathbf{G}^{(0)}$ from SSI produced frequencies $f^{(0)}$, damping ratios $\xi^{(0)}$ and mode shapes $\boldsymbol{\phi}^{(0)}$ that had errors in the range of 10%, 40% and 30% respectively. These initial errors are introduced to check the robustness of the employed algorithms when subjected to initialization errors. $\mathbf{A}^{(0)}$ and $\mathbf{G}^{(0)}$ were set as initial estimates for EM. For VB and GS, $\boldsymbol{\mu}_{\mathbf{A}}^{(0)}$ and $\boldsymbol{\mu}_{\mathbf{G}}^{(0)}$ were set equal to $\mathbf{A}^{(0)}$ and $\mathbf{G}^{(0)}$ respectively.
- $\mathbf{Q}^{(0)} = \mathbf{I}_{20\times 20}$ and $\mathbf{R}^{(0)} = \mathbf{I}_{8\times 8}$
- A tolerance value of $\epsilon_{tol} = 5 \times 10^{-5}$ is set for both EM and VB for the stopping criteria
- For GS, 20 Markov chains – each having 2500 samples – are simulated. A typical plot of Gelman-Rubin statistic is shown in Figure 4.5. The initial 500 samples of each chain are discarded as burn-in and the rest 2000 samples are further subsampled with a period of 2 to obtain 1000 samples for each chain, totalling 20000 (20×1000) samples for estimation of posterior distribution.

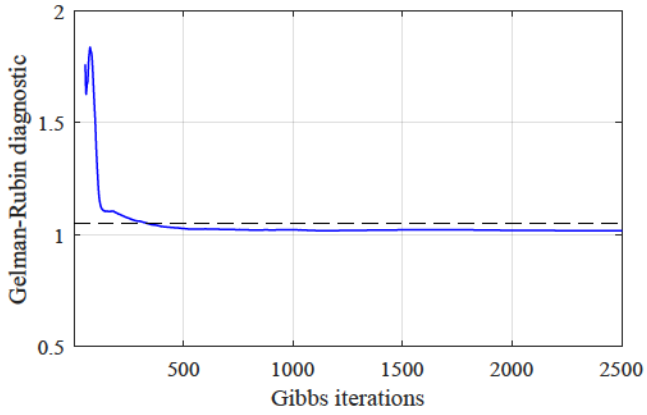


Figure 4.5: Convergence monitoring for GS using Gelman-Rubin diagnostic; the *dotted line* corresponds a Gelman-Rubin diagnostic value of 1.05.

4.7.1 Estimation with missing data from two mobile sensors

The missing data matrix generated using two mobile sensors amounts to 75% of missing data, which is the highest among the three cases considered in Table 4.2. The modal frequencies and damping ratios identified using EM, VB and GS are listed in Table 4.3 and 4.4. Note that EM provides only point estimates whereas VB and GS provides the posterior distributions from which the mean and coefficient of variation (CoV) (= standard deviation (SD)/mean $\times$ 100) for each modal parameter can be calculated.

The estimates of frequencies from all three methods are quite consistent and accurate, while for damping, the estimates from EM are found to be more accurate than those from VB and GS. The CoV of the identified frequencies from VB and GS are substantially small (max around 0.25%), as natural frequencies are much easier to estimate. The damping ratios are identified with a much higher CoV ranging between 5-15%. This is not unnatural as damping ratios are in general much harder to estimate compared to natural frequencies. It is to be noted that all three algorithms show good performance in converging close to the true frequencies and damping ratios, as they were commenced with erroneous initial values of $\mathbf{A}^{(0)}$ and $\mathbf{G}^{(0)}$; the corresponding erroneous initial frequencies $f^{(0)}$ and damping ratios $\xi^{(0)}$ are shown in the third columns of Tables 4.3 and 4.4).

Table 4.3: Identified modal frequencies (in Hz) using EM, VB and GS.

Mode	f_{true}	$f^{(0)}$	EM	VB			GS		
				Mean	SD ($\times 10^{-2}$)	CoV (%)	Mean	SD ($\times 10^{-2}$)	CoV (%)
1	2.942	3.128	2.939	2.948	0.725	0.246	2.938	0.809	0.275
2	5.870	6.347	5.870	5.873	0.822	0.140	5.870	0.921	0.157
3	8.602	7.964	8.579	8.578	0.910	0.106	8.579	0.976	0.114
4	11.188	12.102	11.188	11.191	1.179	0.105	11.190	1.423	0.127
5	13.780	14.141	13.778	13.782	1.333	0.097	13.781	1.568	0.114
6	16.519	15.197	16.537	16.538	1.516	0.092	16.535	2.015	0.122
7	19.536	18.672	19.536	19.537	1.676	0.086	19.536	2.229	0.114
8	23.118	23.328	23.168	23.157	1.864	0.081	23.168	3.177	0.137

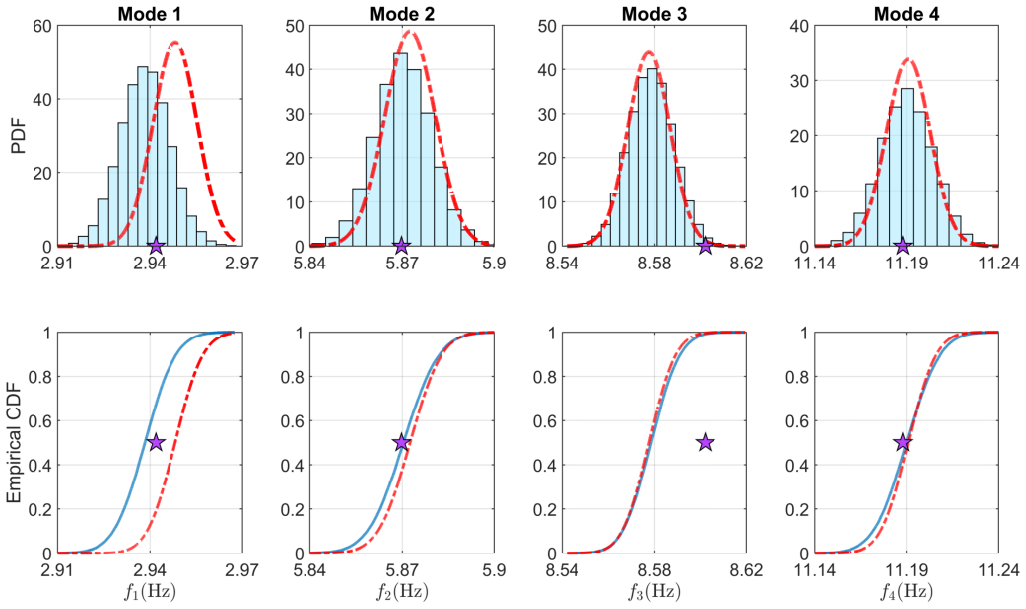
Table 4.4: Identified modal damping ratios (in %) using EM, VB and GS.

Mode	ξ_{true}	$\xi^{(0)}$	EM	VB			GS		
				Mean	SD	CoV (%)	Mean	SD	CoV (%)
1	2.000	1.554	1.984	2.200	0.248	11.293	2.112	0.266	12.255
2	1.243	1.302	1.198	1.371	0.141	10.282	1.219	0.156	11.765
3	1.100	0.661	1.035	1.124	0.107	9.519	1.033	0.115	10.784
4	1.096	1.071	1.230	1.406	0.106	7.550	1.238	0.124	8.264
5	1.147	0.763	1.290	1.410	0.097	6.860	1.271	0.119	7.752
6	1.232	0.936	1.366	1.471	0.092	6.268	1.376	0.118	7.143
7	1.347	1.038	1.338	1.448	0.085	5.867	1.351	0.118	6.667
8	1.500	1.232	1.332	1.556	0.087	5.573	1.393	0.144	6.944

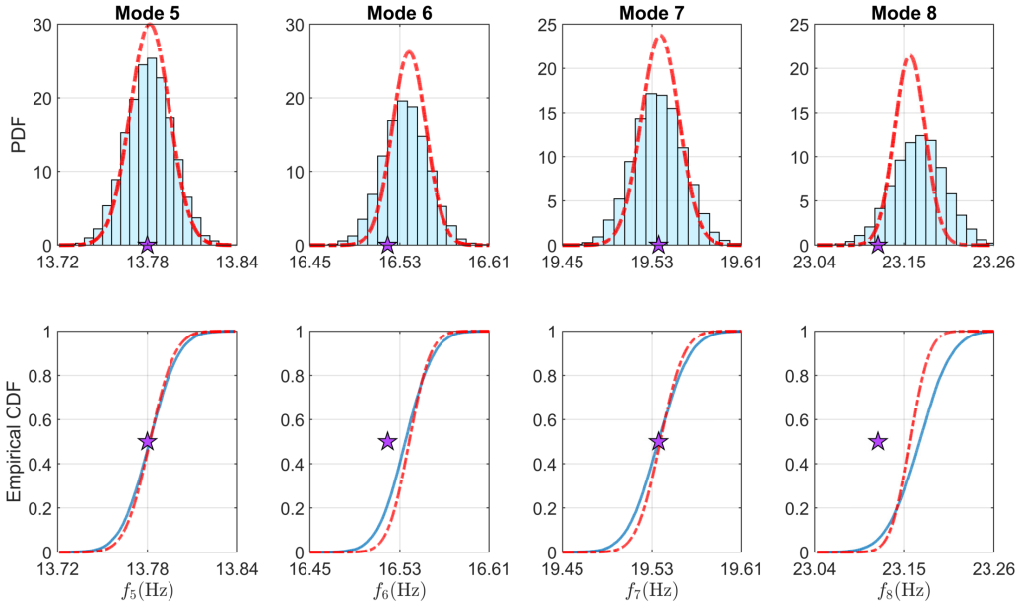
It is also noted that the estimated CoVs of the identified modal frequencies and damping

ratios from VB are typically smaller than those of GS, with a typical underestimation by around 15-25%. Furthermore, it is noticed that the CoVs decrease gradually with higher order modes, particularly for the frequencies.

A comparison of the probability density functions (PDFs) and cumulative distribution functions (CDFs) of the identified modal frequencies and damping ratios are provided in Figure 4.6 and 4.7. A distinct feature noticed in the PDFs estimated by VB is that they consistently show higher peaks than those estimated by the GS. This happens because VB underestimates the spread (variance) of distribution compared to GS, and thus in order to keep the area under the PDF equal to 1, the posterior PDFs from VB show higher peaks than PDFs from GS. Also notice that the mean values of the estimated frequencies and damping ratios from VB vary slightly from that of GS, thus the PDFs from VB appear slightly shifted when plotted on top of the histograms from GS.

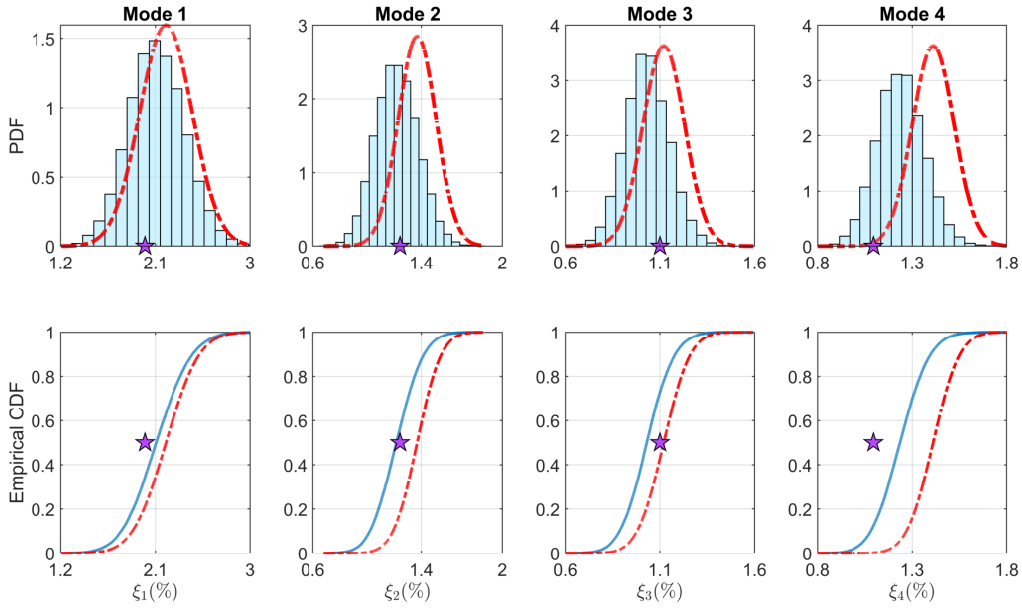


(a) Modes 1 - 4

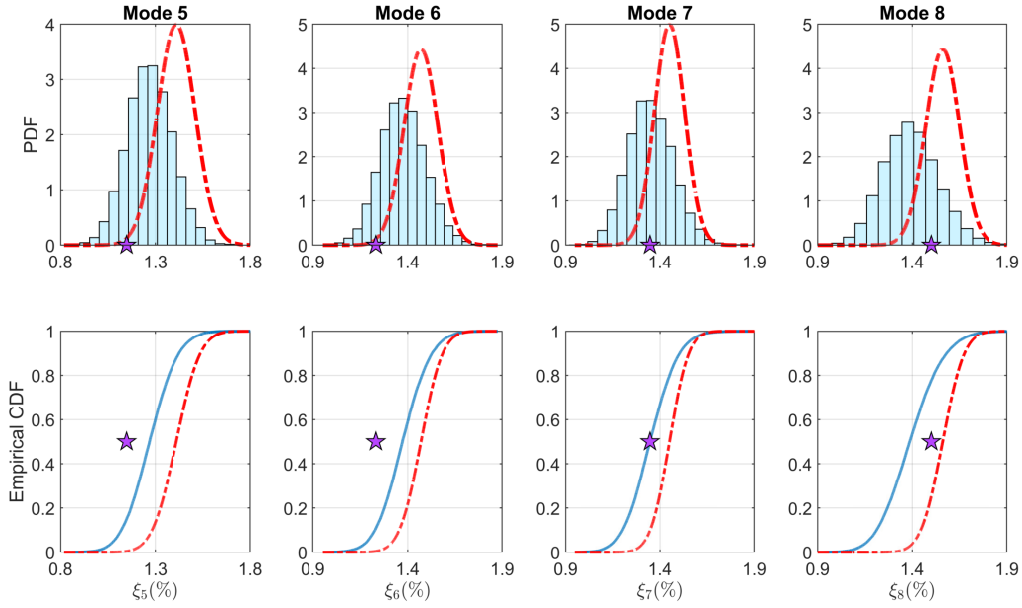


(b) Modes 5 - 8

Figure 4.6: Probability distributions of identified modal frequencies. Solid line: GS; dotted line: variational Bayes; star: true value.



(a) Modes 1 - 4



(b) Modes 5 - 8

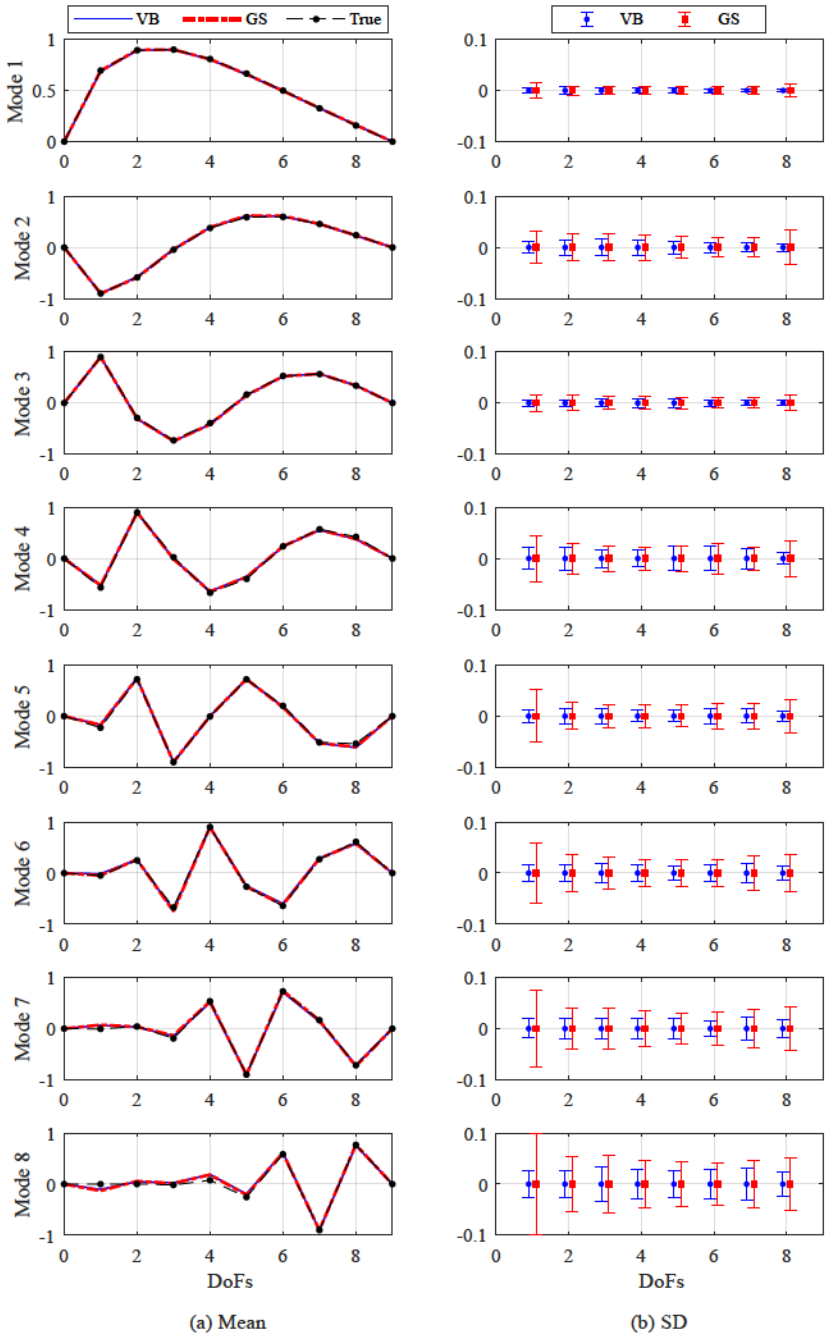
Figure 4.7: Probability distributions of identified modal damping ratios. Solid line: GS; dotted line: variational Bayes; star: true value.

For comparing the estimates of mode shapes from the three algorithms, the modal assurance criterion (MAC) is used as it provides a correlation between the estimated mode shapes and the true mode shapes. The MAC values of the identified mode shapes obtained from EM, VB and GS are tabulated in Table 4.5 alongside the MAC values of initial mode shapes; the initial mode shapes correspond to the mode shapes from $\mathbf{A}^{(0)}$ and $\mathbf{G}^{(0)}$. By comparing the MAC values, it can be seen that the three algorithms, after starting from poor initial estimates of mode shapes, are able to yield good final estimates of mode shapes (with MAC values close to 0.99). However, an exception is noticed for the eighth mode, whose final estimates from the three algorithms are slightly poorer than its initial MAC value.

Table 4.5: Comparison of Modal Assurance Criterion (MAC) values of mode shapes identified using EM, VB and GS.

Mode	True	Initial	EM	VB	GS
1	1.000	0.964	1.000	1.000	1.000
2	1.000	0.990	1.000	1.000	1.000
3	1.000	0.974	1.000	0.999	0.999
4	1.000	0.984	0.998	0.998	0.998
5	1.000	0.955	0.997	0.996	0.997
6	1.000	0.993	0.997	0.996	0.997
7	1.000	0.981	0.996	0.996	0.996
8	1.000	0.992	0.981	0.983	0.981

For a visual comparison, the means and the SDs of the identified mode shapes from VB and GS are plotted in Figures 4.8. In the figures, the right panel plots the SDs of the mode shape ordinates corresponding to their mean values in the left panel. The SDs are plotted as box plots; the SD bars stretch from -SD to +SD at all mode shape ordinates. For example, the mean and SD of the eight mode shape estimated from GS at the 8th DoF is 0.82 ± 0.05 where 0.82 is the mean from left panel and 0.05 is the SD from the right panel corresponding to the 8th DoF of eighth mode shape.



(a) Mean (b) SD

Figure 4.8: The means (left panel) and the SDs (right panel) of the mode shapes ordinates identified using VB and GS.

It is seen that the SDs of the mode shape ordinates are quite low compared to their mean values. Further, it is noted that the SDs of the estimated mode shapes obtained from VB exhibit a 40-60% underestimation compared to that from GS especially near the regions of smaller vibration (or zero crossings). Thus, in the face of large amounts of missing data, the VB uncertainty estimates of mode shape ordinates may not be reliable.

The log-likelihood convergence plots of the three algorithms i.e. EM, VB and GS, are shown in Figure 4.9. It can be seen that all three algorithms converge to similar likelihood values. For the same tolerance $\epsilon_{tol} = 5 \times 10^5$, VB is typically found to converge faster (requiring fewer iterations) than EM.

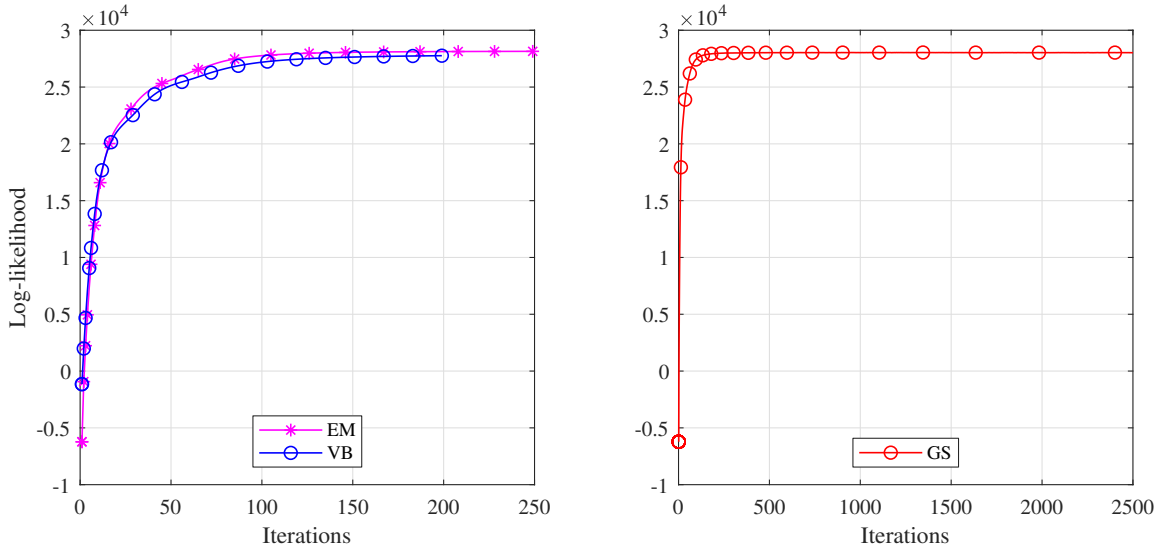


Figure 4.9: Convergence of EM, VB and GS.

Regarding the computational cost of running the algorithms for modal parameter estimation, the EM and VB algorithms are computationally much cheaper than the GS. A comparison of computational runtimes taken by the three algorithms (using SHARCNET [118] servers) is listed in Table 4.6.

Table 4.6: Comparison of computational runtimes for EM, VB and GS; computations were done on SHARCNET server nodes with Intel E5-2683V4 CPU @2.1GHz processor and 4GB RAM.

Algorithm	Nodes used	Time (hrs)
EM	1	0.26
VB	1	0.28
GS	20 (using parallel computing)	6.78

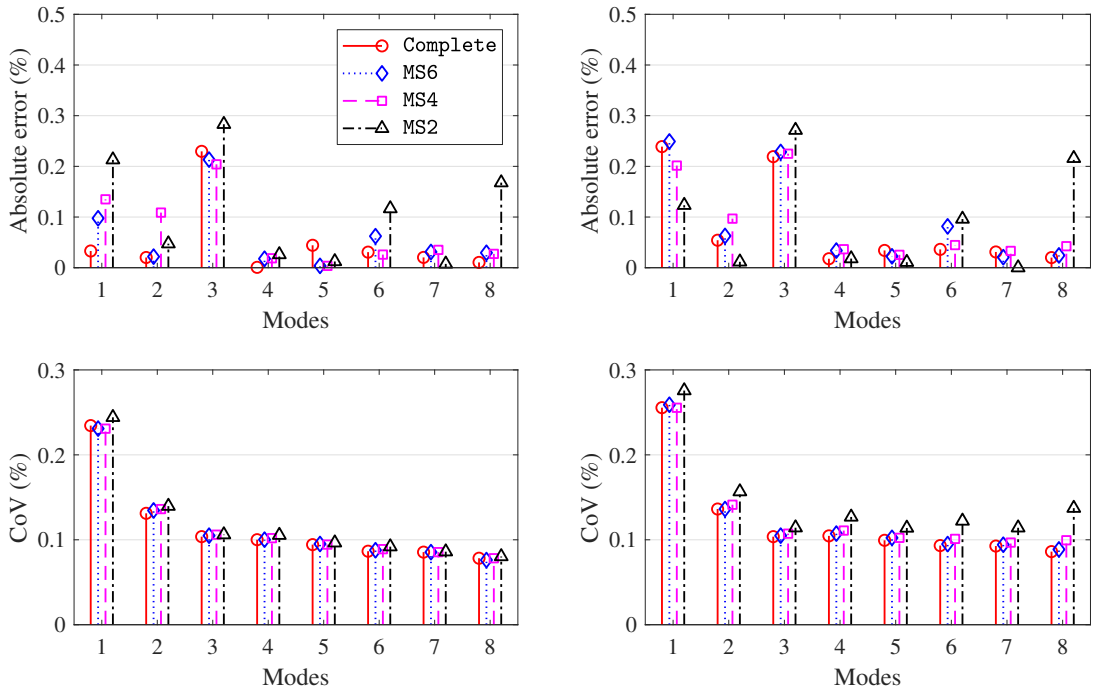
4.7.2 Impact of missing data on modal parameter estimation

The impact of missing data on the accuracy and uncertainty of modal parameter estimates is studied by comparing the performance achieved by the three difference mobile sensor groups: `MS2` (2 sensors), `MS4` (4 sensors) and `MS6` (6 sensors) against a baseline scenario consisting of a full static sensor network (i.e., `Complete` data set). The number of mobile sensors is directly related to the extent of missing information as shown in Table 4.2. The VB and GS approaches will be compared within each sensor group to assess how missing data impacts the parameter estimation and uncertainty quantification. The point estimates from the EM algorithm were generally consistent with the mean estimates from VB and thus are not presented here.

Comparison of modal frequencies and damping ratios

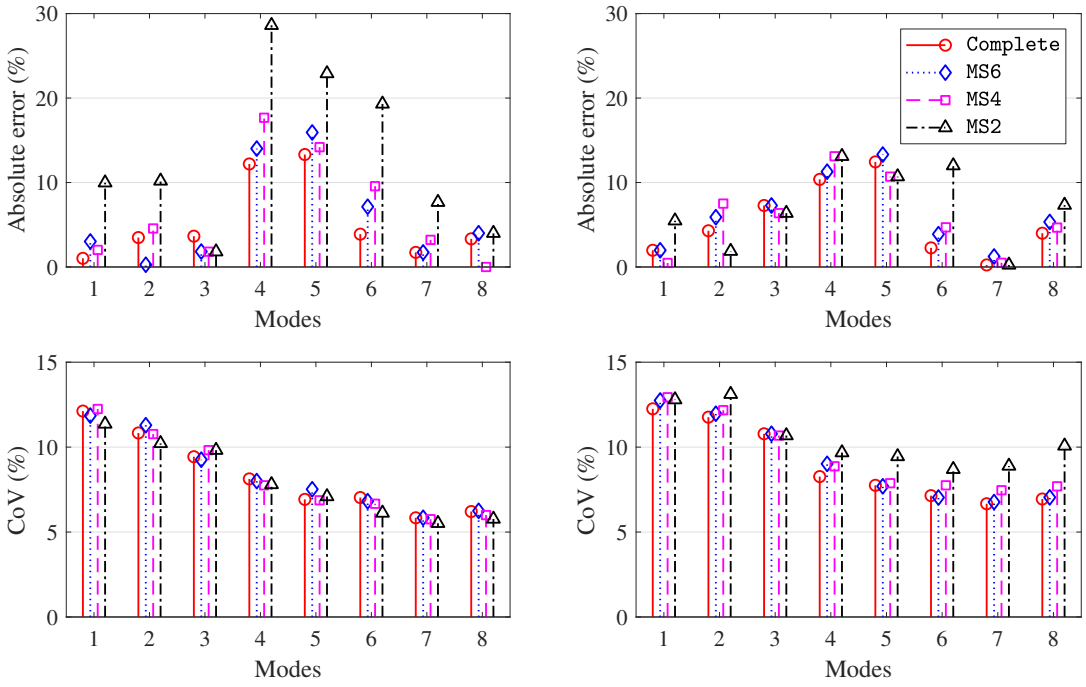
Consider the impact of missing data on estimates of modal frequencies and damping ratios. Figures 4.10 and 4.11 compare the absolute error and CoVs across each sensor group using VB and GS for frequency and damping ratio estimates respectively. The absolute error percentage is obtained by dividing the absolute difference (between the true value and the estimated value) by the true value, multiplied by 100.

From Figures 4.10 and 4.11, the following observations are made:



(a) VB (b) GS

Figure 4.10: Absolute errors and CoVs of the estimated modal frequencies obtained using different number of mobile sensors; panel (a): estimates from VB, and panel (b): estimates from GS.



(a) VB (b) GS

Figure 4.11: Absolute errors and CoVs of the estimated modal damping ratios obtained using different number of mobile sensors; panel (a): estimates from VB, and panel (b): estimates from GS.

1. The estimates from complete dataset (i.e., baseline scenario) contained the lowest errors. This result is expected given the complete dataset contains the most information.

2. The VB and GS approaches yield similar mean estimates of modal frequencies and damping ratios. One exception/anomaly is the case of `MS2` where the errors from VB are much larger than those from GS.

3. The CoVs tends to decrease with higher modes. This trend is observed for both frequency and damping estimates obtained from VB and GS.

4. The VB and GS approaches yield similar CoVs of estimated modal frequencies and damping ratios for all sensor setups. One notable exception is the `MS2` case which yielded slightly higher CoVs for GS estimates.

Comparison of mode shapes

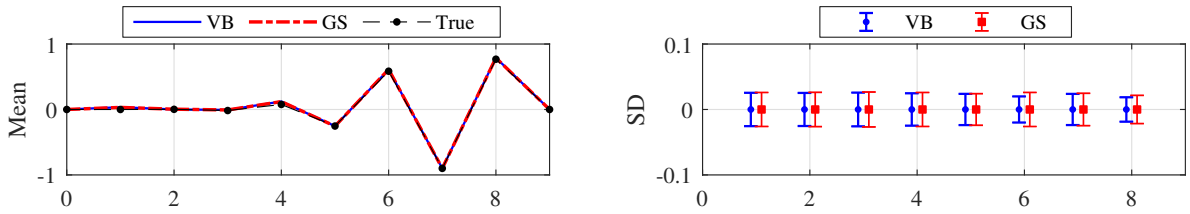
Next, consider the impact of missing data on the mode shape estimates. Table 4.7 summarizes the MAC values of the mode shape estimates obtained from VB and GS for the different mobile sensor networks and baseline scenario. As shown in the MAC values, the mode shape estimates improve as the percentage of missing data decreases. For example, the MAC values of the eighth mode shape estimates improve considerably from 0.981 for `MS2` to 0.996 for `MS4`, demonstrating considerable improvement when the percentage of missing data is reduced from 75% to 50%. However, not a large improvement is obtained when the percentage of missing data is reduced from 50% (4 mobile sensors) to 25% (6 mobile sensors) as the MAC values only marginally improve from 0.996 to 0.998. For a more comprehensive assessment, the mean and the SDs of the mode shapes from VB and GS are plotted for mode 8 in Figure 4.12. In the left panel, as one goes from top (`Complete` case) to bottom (`MS2` case), a decline in the accuracy of the mean estimate of the eighth mode shape is observed. The right panel of Figure 4.12 plots the corresponding SDs and a noticeable increase in the SDs with more missing data is seen while this is not so apparent in the case of VB. It is however noticed that as the number of mobile sensors increases, the amount of data increases and the SDs of the from GS converge closer to that from

Table 4.7: Summary of MAC values of the mode shape estimates identified using dataset from different sensor groups (`Complete`, `MS6`, `MS4`, `MS2`) via VB and GS.

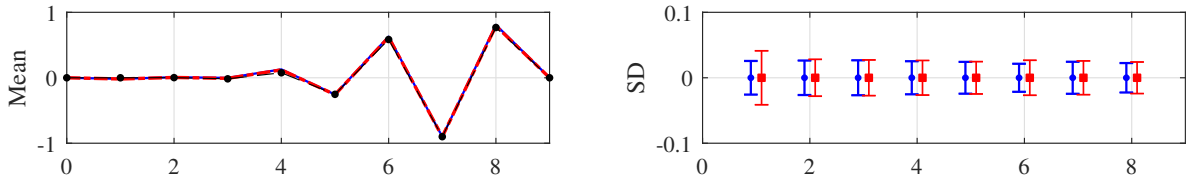
	`Complete`		`MS6`		`MS4`		`MS2`	
Mode	VB	GS	VB	GS	VB	GS	VB	GS
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	1.000	1.000	1.000	1.000	0.999	1.000	0.999	0.999
4	1.000	1.000	1.000	1.000	0.999	0.999	0.998	0.998
5	1.000	1.000	0.999	1.000	1.000	1.000	0.996	0.997
6	0.999	0.999	0.999	0.999	0.999	0.999	0.996	0.997
7	0.999	0.999	0.998	0.998	0.997	0.997	0.996	0.996
8	0.998	0.998	0.998	0.998	0.996	0.996	0.983	0.981

VB. Hence, it can be deduced that VB may fall short in capturing the variability of the parameters when the missing information is large, however, the mean values are estimated quite accurately.

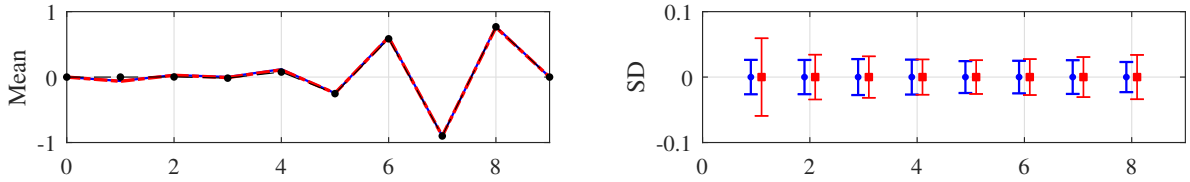
The log-likelihood convergence plots for the cases `Complete`, `MS6`, `MS4` and `MS2` are shown in Figure 4.13. Log-likelihood values provide an idea of the goodness-of-fit of the estimated parameters i.e. higher the stationary log-likelihood, better the fit. From the inspection of the stationary log-likelihood values for the four cases, it can be seen that the stationary log-likelihood value decreases with an increase in missing data. In other words, an incomplete data matrix from a sensor network with larger number of mobile sensors will lead to higher log-likelihoods (due to a lower percentage of missing data). Thus, an increase in the log-likelihood is proportional to the amount of measurements, i.e. more the measurements, higher the likelihood. Furthermore, it is noted that the number of iterations needed to attain log-likelihood convergence for VB (and EM) decreases with a decrease in the missingness of data.



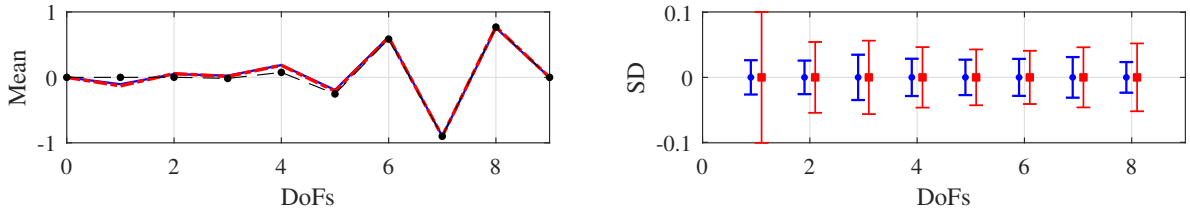
(a) `Complete`



(b) `MS6` ($n_{ms} = 6$)

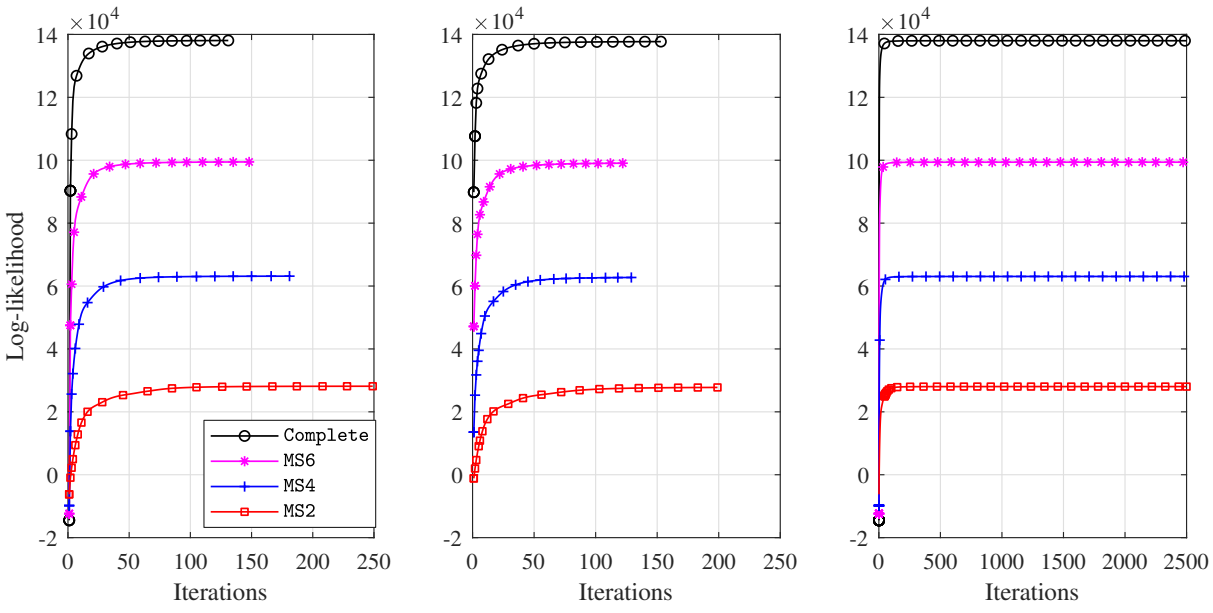


(c) `MS4` ($n_{ms} = 4$)



(d) `MS2` ($n_{ms} = 2$)

Figure 4.12: Comparison of the means and the SDs of the eighth mode shape estimated using VB and GS for the four cases `Complete`, `MS6`, `MS4` and `MS2`. Left panel shows means and right panel shows SDs.



(a) EM (b) VB (c) GS

Figure 4.13: Likelihood plots showing convergence of EM, VB and GS for the four cases `Fixed`, `MS6`, `MS4` and `MS2`.

4.8 Experimental Validation

Modal parameter identification using in-mobile sensing following the missing data approach is experimentally validated using a 3-storey bench-scale test frame. Figure 4.14 illustrates the 3-storey test frame used for testing and a picture of the actual setup is provided in Figure 4.15.

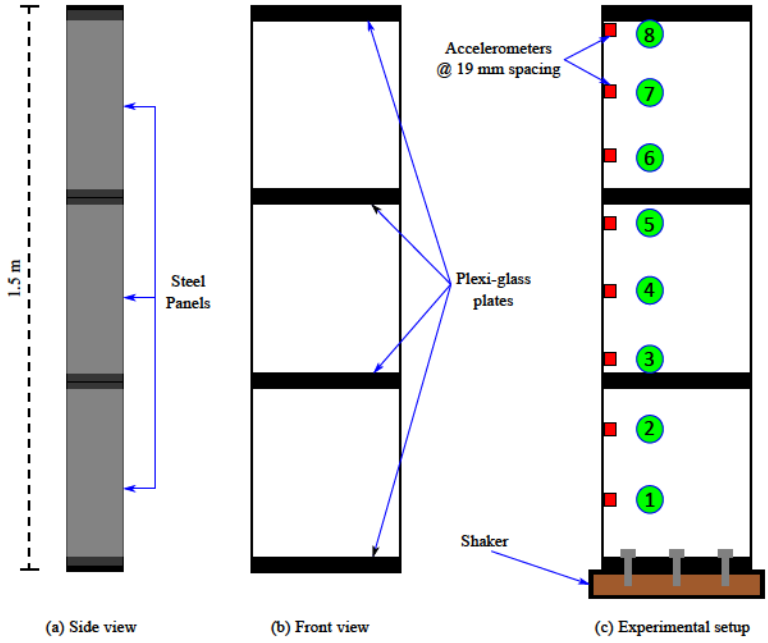


Figure 4.14: Schematic of laboratory scale 3-storey test frame: (a) side view, (b) front view and (c) experimental setup.

The 3-storey test frame consists of six steel panels columns and four plexi-glass plates. Each column was built up with three steel panels (of size 500mm $\times$ 110mm $\times$ 1.75mm). The mass of the plexi glass plates were 3kg each. The steel panels make up two supporting columns and the plexi-glass blocks act as lumped floor masses. The frame is fixed at its base to a one-directional linear dynamic shaker. The test frame is base-excited by the dynamic shaker with a 40 Hz band-limited Gaussian white noise excitation and is instrumented with eight static accelerometers as shown in 4.14c. The eight static accelerometers sampled

structural responses at eight sensing nodes (i.e., locations on the test frame where the sensors are mounted) at 100 Hz for a time duration of 200s.

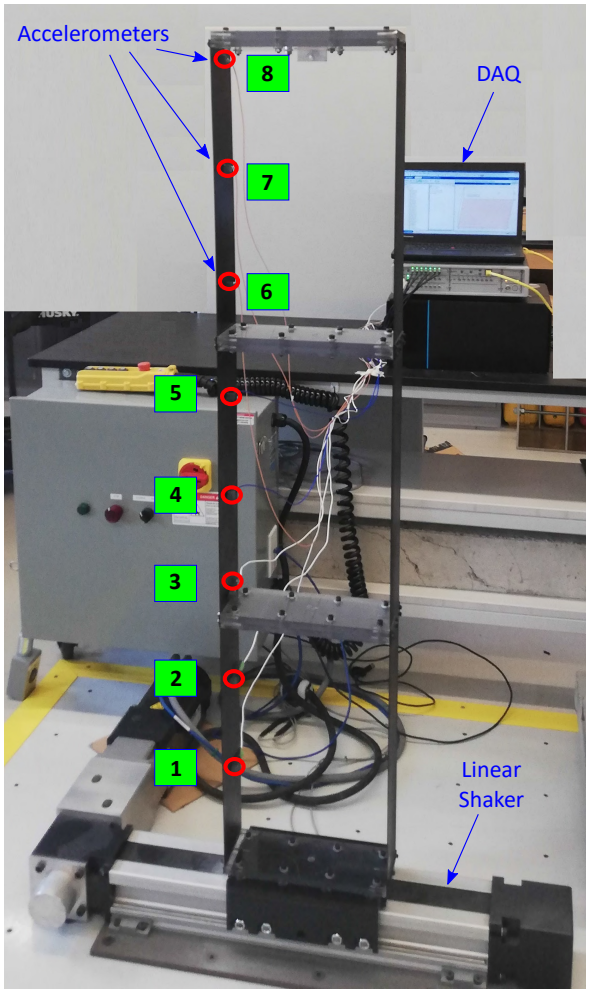


Figure 4.15: Picture of the experimental setup.

The eight measured acceleration responses are shown in Figure 4.16, and their averaged power spectral density is shown in Figure 4.17 which shows the three clear peaks of lateral modes of the test frame.

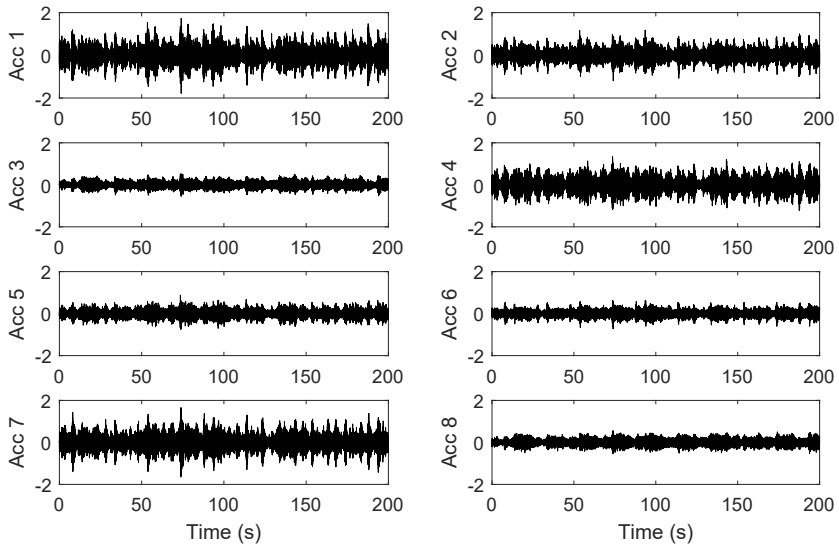


Figure 4.16: Accelerations (in m/s^2) measured at eight sensing nodes (numbered Acc 1-8).

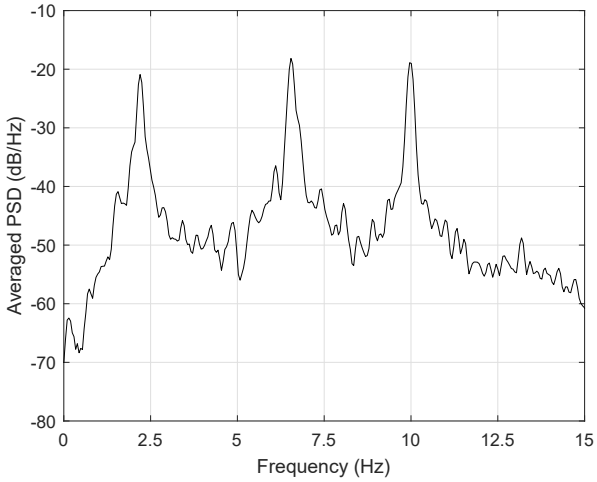


Figure 4.17: Averaged power spectral density of the acceleration data obtained from eight static sensors.

A mobile-sensor strategy employing two mobile sensors is simulated using the static sensor dataset. The two mobile sensors are assumed to scan the eight sensing nodes back

and forth, resulting in a dataset with 75% missing data. The EM, VB and GS algorithms are implemented on the mobile sensor data using a model order of 10. The estimates of $\mathbf{A}$ and $\mathbf{G}$ obtained from SSI are randomly perturbed to produce initial estimates for EM, VB and GS; the perturbations are created by adding 5% Gaussian noise to the values of $\mathbf{A}$ and $\mathbf{G}$ obtained from SSI which led to perturbed initial modal parameters $f^{(0)}$ and $\zeta^{(0)}$. The modal parameters estimated from SSI – employed on the static sensor dataset – are used as baseline estimates for comparing estimates from EM, VB and GS using the mobile sensor dataset.

The estimates of the modal frequencies and damping ratios from the three computational algorithms are summarized in Tables 4.8 and 4.9 respectively.

Table 4.8: Identified modal frequencies (in Hz) using EM, VB and GS.

Mode	f^{SSI}	$f^{(0)}$	EM	VB			GS		
				Mean	SD ($\times 10^{-2}$)	CoV (%)	Mean	SD ($\times 10^{-2}$)	CoV (%)
1	2.186	2.311	2.193	2.191	0.468	0.214	2.192	0.478	0.218
2	6.561	6.226	6.561	6.563	0.422	0.064	6.564	0.456	0.070
3	9.983	10.832	9.985	9.984	0.400	0.034	9.982	0.441	0.044

Table 4.9: Identified modal damping ratios (in %) using EM, VB and GS.

Mode	ξ^{SSI}	$\xi^{(0)}$	EM	VB			GS		
				Mean	SD	CoV (%)	Mean	SD	CoV (%)
1	0.599	0.754	0.707	0.598	0.218	36.455	0.553	0.274	49.548
2	0.288	0.402	0.332	0.340	0.066	19.412	0.358	0.070	19.553
3	0.130	0.361	0.137	0.144	0.036	25.000	0.136	0.040	29.412

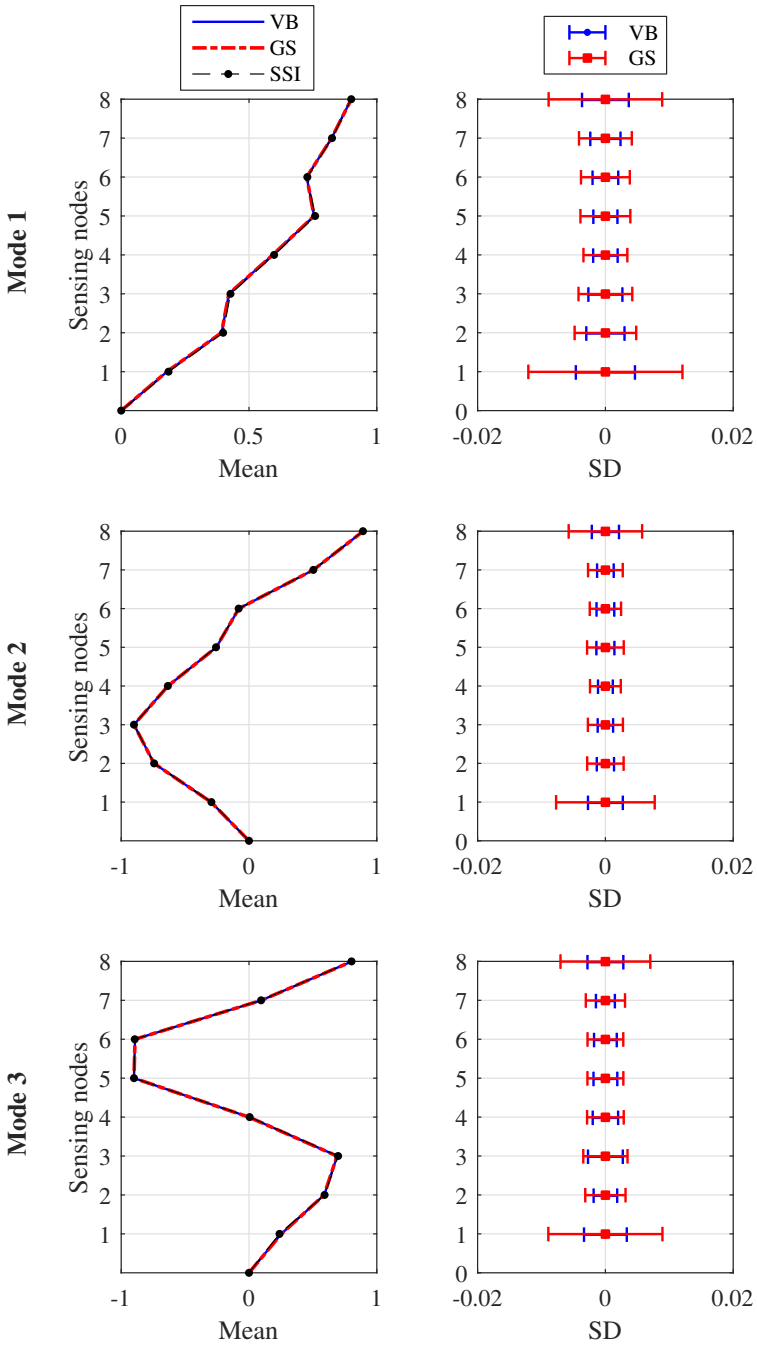


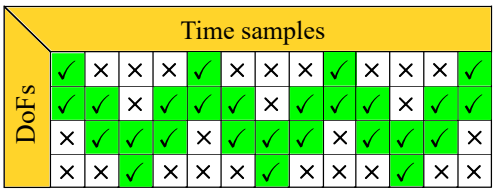
Figure 4.18: Means (left panel) and SDs (right panel) of mode shapes obtained using VB and GS.

It is observed that modal damping ratios are estimated with greater variability than the modal frequencies and that VB underestimates the variability of the modal parameters compared to GS. The mean values and SDs of the estimated mode shapes are illustrated in Figure 4.18 and they are found to be in very good agreement with the mode shapes estimated from SSI. As for the variability of the mode shape ordinates, the VB is found to underestimate the uncertainty by around 40-60% of the uncertainty estimated by GS.

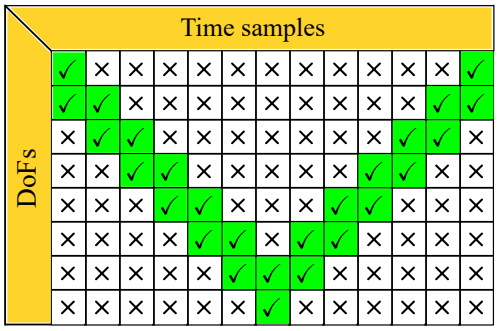
4.9 Summary

In this chapter, the problem of output-only modal parameter estimation using in-motion mobile sensors is posed from an incomplete data perspective. The mobile sensor data is treated equivalent to a static sensor dataset with missing entries at DoFs not coinciding with the mobile sensor paths. Parameters of a linear time-invariant stochastic SSM are fitted using the missing data matrix and the modal parameters are subsequently extracted from these estimated SSM parameters. The estimation of the SSM parameters is handled in a Bayesian framework suitable for obtaining uncertainty information, arising not only due to measurement and modelling errors but also missingness in data matrix. Three computational algorithms, EM, VB and GS, are presented and modified for use with datasets featuring missing data.

The performance of the modal parameter estimation and uncertainty quantification is studied using numerical simulations on a eight-DoF lumped mass beam model as well as on a laboratory scale 3-storey test frame. All three algorithms (i.e. EM, VB and GS) are found to provide good point estimates of the modal parameters even when subjected to large amounts of missing entries. Regarding uncertainty estimation, it was noted that in comparison to GS, the VB underestimates the uncertainty in modal parameter estimates, particularly for the mode shapes, when a large amount of entries (e.g. 75%) are missing. It is observed that the uncertainty estimates in the mode shape ordinates from VB converge close to that from GS as the percentage of missing data reduces. Finally, among the three algorithms, the EM and VB are found to be much faster computationally compared to GS. It must also be mentioned that all the three algorithms can become computationally



(a) 50% missing data



(b) 75% missing data

Figure 4.19: Increase in missing entries due to increase in the number of sensing nodes; cases (a) and (b) correspond to scenarios of measuring 4 DoFs and 8 DoFs using two mobile sensors resulting in 50% and 75% missing entries, respectively.

costly when the model order of the state space parameters is very high.

Finally, it should be mentioned that a shortcoming of the missing data approach is that the estimation does not scale well when there are a large number of sensing nodes. This is due to the fact that the sensing nodes are linked to the modelled DoFs of the structure, and as the number of sensing nodes increases keeping the number of mobile sensors fixed, the degree of missing data also increases. This issue is also illustrated in Figure 4.19. As such, in events when the missing data matrix has 90-99% missing entries, this approach could be computationally inefficient and may not yield statistically meaningful results. The next chapter presents a stacked data approach to represent mobile sensor data which does not feature any missing entries.

Chapter 5

Bayesian output-only modal identification using in-motion mobile sensors: Stacked data approach

5.1 Introduction

The previous chapter dealt with the problem of modal identification using *in-motion* mobile sensors from an incomplete-data perspective. In this approach, the mobile sensor observations were used to construct an equivalent static sensor matrix containing missing entries at the sensing nodes where a mobile sensor was not present. The efficiency of this approach reduces considerably as the percentage of missing entries increases. This occurs when the ratio of sensing nodes to mobile sensors becomes large. This chapter presents a stacked-data approach as an alternative to the missing data approach by reshaping the problem using a complete-data perspective. Here, each mobile sensor is treated as a sensor channel that records time-series data from various points in space and concatenation of these sensor channels yields a stacked data matrix. Unlike the missing data matrix, the stacked data matrix is complete in the sense that it has no missing entries. The transformation from the missing data representation to stacked data representation has been previously illustrated in Figure 2.3.

A defining feature of the stacked data matrix is the presence of *spatial discontinuities* along its channels, that is, the entries in any particular channel come from different spatial locations. By contrast, all entries in a single channel of a static sensor data matrix correspond to a single fixed location in space. Figure 5.1 illustrates the differences between the static sensor data matrix and the stacked mobile sensor data matrix. Moreover, due to spatial discontinuities, it is necessary to tag each entry in the stacked data matrix to its corresponding spatial location and time record. This is achieved by constructing a sensor-position matrix which stores the spatial positions of the mobile sensors at all sampling-time instances and would be needed during the process of modal identification.

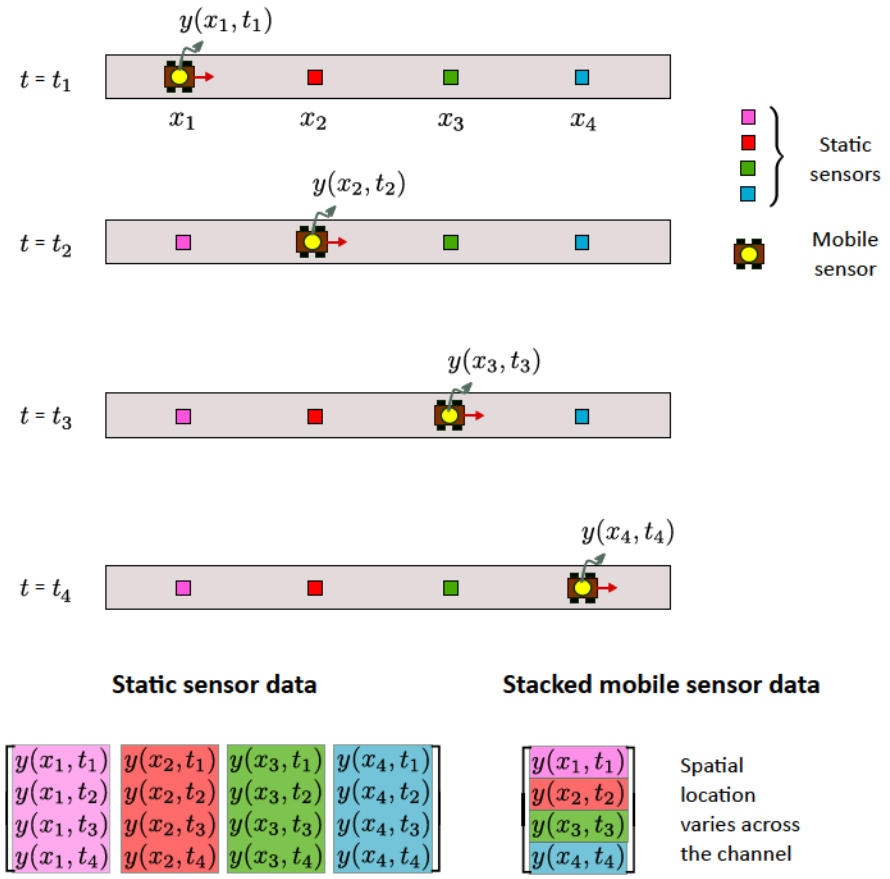


Figure 5.1: Illustration of a single channel stacked data matrix (from one mobile sensor) and four-channel static sensor matrix (from four static sensors).

Conventional SSID techniques (such as SSI, ERA-OKID, etc.) are based on LTI SSMs which are only compatible with static measurements – where each sensor channel measures responses from from a fixed location. As such, LTI SSMs are incompatible with the stacked data matrix due to its inherent spatial discontinuities, and thus the conventional SSID algorithms are unable to handle mobile sensor data when stacked in such a manner. In this study, a new linear time-varying SSM is proposed that can accommodate measurements with spatial discontinuities and is therefore compatible with the stacked data matrix. Furthermore, new update equations are derived for EM, VB and GS in order to apply them for modal parameter estimation with the proposed linear time-varying SSM. The proposed methodology is demonstrated through numerical examples and validated experimentally.

5.2 State space models for identification with stacked data matrix

The typical LTI SSM, used with a static sensor data matrix, is given by:

$$\boldsymbol{x}_{k+1} = \mathbf{A}\boldsymbol{x}_k + \boldsymbol{w}_k \tag{5.1}$$

$$\boldsymbol{y}_k = \mathbf{G}\boldsymbol{x}_k + \boldsymbol{v}_k. \tag{5.2}$$

The fact that $\mathbf{G}$ is a constant matrix implies Eq.(5.1) cannot address spatially varying measurements. This is the main limitation rendering the LTI SSM above incompatible with a stacked data matrix. A way to modify the above SSM to suit the stacked data matrix is to incorporate a spatial interpolator in the measurement equation, such that the mobile sensor measurements from various locations can be mapped to *equivalent* static measurements at a set of predetermined locations. Such a mapping can be achieved by any suitable spatial interpolation function $\boldsymbol{\Omega}$. For example, the structural measurements $\boldsymbol{y}$ at, say, a vector of locations $\boldsymbol{s}^o$ can be mapped to equivalent measurements $\boldsymbol{y}^\alpha$ at a vector of locations $\boldsymbol{s}^\alpha$ via the use of the structural mode shape matrix $\boldsymbol{\Phi}$ as follows:

$$\boldsymbol{y} = \boldsymbol{\Omega}\boldsymbol{y}^\alpha = \boldsymbol{\Phi}^o \underbrace{(\boldsymbol{\Phi}^\alpha)^{-1}\,\boldsymbol{y}^\alpha}_{\boldsymbol{y}^m}. \tag{5.3}$$

Here $\boldsymbol{\Phi}^o$ and $\boldsymbol{\Phi}^\alpha$ are mode shape matrices defined at the vector of locations $\boldsymbol{s}^o$ and $\boldsymbol{s}^\alpha$ respectively, and $\boldsymbol{y}^m$ is the vector of modal responses. One can understand the mapping in Eq.(5.3) as first converting $\boldsymbol{y}^\alpha$ to modal response $\boldsymbol{y}^m$ by the operation $(\boldsymbol{\Phi}^\alpha)^{-1}\boldsymbol{y}^\alpha$ and then converting $\boldsymbol{y}^m$ back to $\boldsymbol{y}$ using $\boldsymbol{\Phi}^o\boldsymbol{y}^m$. As such, $\boldsymbol{\Omega} = \boldsymbol{\Phi}^o(\boldsymbol{\Phi}^\alpha)^{-1}$ is a valid spatial interpolator which helps to reduce the mobile sensor measurements at various locations to equivalent static measurements at a set of virtual static locations (VSLs).

Based on this concept, Matarazzo and Pakzad [46] introduced the truncated physical model (TPM), a time-varying SSM, that enables estimation with the stacked data matrix. The TPM models the structure with DoFs defined at a set of user-chosen VSLs. The total number of VSLs, denoted by n_α, is typically much less than the number of mobile sensing nodes (i.e., locations where mobile sensor measurements are recorded). The vector of spatial locations associated with the VSLs is denoted by $\boldsymbol{s}^\alpha$, and the vector of the mobile sensors' locations at each time instant t_k is denoted by $\boldsymbol{s}_k^o \in \mathbb{R}^{n_o}$. Mathematically, the TPM is expressed as follows:

$$\boldsymbol{x}_{k+1}^\alpha = \mathbf{A}^\alpha \boldsymbol{x}_k^\alpha + \boldsymbol{w}_k^\alpha \tag{5.4a}$$

$$\boldsymbol{y}_k = \boldsymbol{\Omega}_k \underbrace{(\mathbf{G}^\alpha \boldsymbol{x}_k^\alpha)}_{\boldsymbol{y}_k^\alpha} + \boldsymbol{v}_k. \tag{5.4b}$$

Here, $\boldsymbol{x}_k^\alpha \in \mathbb{R}^{2n_\alpha}$, $\mathbf{A}^\alpha \in \mathbb{R}^{2n_\alpha \times 2n_\alpha}$, $\boldsymbol{w}_k^\alpha \in \mathbb{R}^{2n_\alpha}$, $\boldsymbol{y}_k \in \mathbb{R}^{n_o}$, $\boldsymbol{\Omega}_k \in \mathbb{R}^{n_o \times n_\alpha}$, $\mathbf{G}^\alpha \in \mathbb{R}^{n_\alpha \times 2n_\alpha}$ and $\boldsymbol{v}_k \in \mathbb{R}^{n_o}$. The VSL states $\boldsymbol{x}_k^\alpha$ represent the physical displacements and velocities at the selected VSLs. The matrix $\mathbf{G}^\alpha$ relates the VSL states $\boldsymbol{x}_k^\alpha$ to equivalent VSL outputs $\boldsymbol{y}_k^\alpha$, i.e., $\boldsymbol{y}_k^\alpha = \overline{\mathbf{G}}^\alpha \boldsymbol{x}_k^\alpha$. The measurement equation of the TPM embeds a spatial interpolator $\boldsymbol{\Omega}_k \in \mathbb{R}^{n_o \times n_\alpha}$. At each time instant t_k, $\boldsymbol{\Omega}_k$ enables the mapping mobile sensor outputs $\boldsymbol{y}_k$ to equivalent outputs $\boldsymbol{y}_k^\alpha$ at the user-selected VSLs. These spatial interpolators are referred as to mode shape regression matrices [21], after their regression-of-mode-shapes' form in Eq.(5.3). However, in practice, the sequence of exact mode shape regression (MSR) matrices $\boldsymbol{\Omega}_{1:N} = \{\boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_N\}$ are unknown prior to modal identification as they are dependent on the structural mode shapes themselves. Therefore, a sinc-function-based MSR was proposed for use [21, 119]:

$$\boldsymbol{\Omega}_k = \left[\operatorname{sinc}\left(\tfrac{1}{\Delta s^\alpha}\left(\boldsymbol{s}_k^o - s_1^\alpha\right)\right) \quad \operatorname{sinc}\left(\tfrac{1}{\Delta s^\alpha}\left(\boldsymbol{s}_k^o - s_2^\alpha\right)\right) \quad \ldots \quad \operatorname{sinc}\left(\tfrac{1}{\Delta s^\alpha}\left(\boldsymbol{s}_k^o - s_{n_\alpha}^\alpha\right)\right)\right] \tag{5.5}$$

$$\boldsymbol{s}^{\alpha} = \begin{bmatrix} s_1^{\alpha} & s_2^{\alpha} & \dots & s_{n_\alpha}^{\alpha} \end{bmatrix}$$
$$\boldsymbol{s}_k^{o} = \begin{bmatrix} s_{1,k}^{o} & s_{2,k}^{o} & \dots & s_{n_{ms},k}^{o} \end{bmatrix}^T.$$

Here s_i^{α} is the location of the ith VSL (n_α in total) and Δs^{α} is the distance between the VSLs. To achieve good estimation accuracy with sinc function-based spatial interpolation, the VSLs need to be uniformly spaced throughout the structure such that Δs^{α} remains constant [46]. Furthermore, that the spacing of the VSLs must be chosen so as to avoid spatial aliasing. It must be mentioned that the mode shapes are only estimated at the user-selected VSLs. Thus, by regulating the number of VSLs one can control the resolution of the estimated mode shapes.

Matarazzo and Pakzad [21] proposed the STRIDEX algorithm, a modified EM algorithm, for modal identification with the TPM. A single run of STRIDEX with the TPM yielded maximum likelihood point estimates of natural frequencies, damping ratios and mode shape ordinates at the VSLs. However, the STRIDEX formulation lacks robustness; more specifically, it suffers from numerical invertibility issues[1] whenever the MSR matrices are non-square leading to errors in identification. To restrict the MSR matrices to square forms, the authors [21] enforced a minimum model size criterion, following which the number of VSLs, the number of modes, and the number of mobile sensors were set equal to each other. Since the number of VSLs must be large to achieve high-resolution mode shape estimates, the number of mobile sensors must be equally large to satisfy the model size criterion. To avoid this impracticality, the authors [21] constructed multiple TPMs using several non-overlapping subsets of VSLs, each subset having the same number of VSLs as the number of mobile sensors. STRIDEX was then run on each TPM separately to obtain local mode shapes. These non-overlapping local mode shapes were then merged to obtain high resolution global mode shapes.

The two main drawbacks of the TPM-based STRIDEX formulation concern its limited flexibility and robustness which arise due to: (1) the TPM models the structure in physical space which is typically higher than the modal space and requires higher model order which can be computationally demanding, and (2) the minimum model size criterion not only

[1] An error in the STRIDEX update equations was identified during this study

negatively impacts the flexibility of the STRIDEX formulation but also adds the complexity of creating multiple VSL subsets and running STRIDEX multiple times.

To address these limitations, an improved modal estimation framework for stacked sensor data is introduced. In particular, a modal state model (MSM) is presented which admits lower model orders (compared to TPM), and the three computational algorithms, EM, VB and GS, are modified accordingly to facilitate modal parameter estimation with the MSM. The advantages of the new framework are as follows:

1. The MSM is more flexible than the TPM as it does not require modelling the structure in physical space. The MSM can maintain a lower model order (than that required by TPM) without sacrificing any relevant information needed for identification.

2. The constraint of having the number of VSLs equal to the number of mobile sensors is relaxed. There is no need to create multiple smaller subsets of VSLs and the identification of complete mode shapes at all VSLs can be obtained in a single run.

3. By removing the need to create multiple VSL subsets, the framework circumvents the problem of scale ambiguity which can arise during the merging of non-overlapping partial mode shapes into global mode shapes.

5.2.1 Proposed Modal State Model

In this section, the modal-state model is introduced that allows identification with the stacked data matrix. Transforming the physical VSL states $\boldsymbol{x}_k^\alpha \in \mathbb{R}^{2n_\alpha}$ to modal states $\overline{\boldsymbol{x}}_k \in \mathbb{R}^{2n_m}$ using a transformation matrix $\mathbf{T} = \begin{bmatrix} \boldsymbol{\Phi}^\alpha & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi}^\alpha \end{bmatrix}$ such that $\boldsymbol{x}_k^\alpha = \mathbf{T}\overline{\boldsymbol{x}}_k$, one obtains the following LGSSM

$$\begin{aligned} \overline{\boldsymbol{x}}_{k+1} &= \left(\mathbf{T}^{-1}\mathbf{A}^\alpha\mathbf{T}\right)\overline{\boldsymbol{x}}_k + \mathbf{T}^{-1}\boldsymbol{w}_k^\alpha \\ \boldsymbol{y}_k &= \boldsymbol{\Omega}_k\left(\mathbf{G}^\alpha\mathbf{T}\right)\overline{\boldsymbol{x}}_k + \boldsymbol{v}_k. \end{aligned} \tag{5.6}$$

Typically, $n_m < n_\alpha$, thereby further reducing the state dimension of the above SSM in comparison to the TPM. The resulting LGSSM is represented by the following equations

$$\overline{\boldsymbol{x}}_{k+1} = \overline{\mathbf{A}}\overline{\boldsymbol{x}}_k + \overline{\boldsymbol{w}}_k \tag{5.7a}$$

$$\boldsymbol{y}_k = \boldsymbol{\Omega}_k \overline{\mathbf{G}} \bar{\boldsymbol{x}}_k + \boldsymbol{v}_k \tag{5.7b}$$

where $n_s = 2n_m$, $\bar{\boldsymbol{x}}_k \in \mathbb{R}^{n_s}$, $\overline{\mathbf{A}} \in \mathbb{R}^{n_s \times n_s}$, $\overline{\boldsymbol{w}}_k \in \mathbb{R}^{n_s}$ are the variables of the process equation and $\boldsymbol{y}_k \in \mathbb{R}^{n_o}, \boldsymbol{\Omega}_k \in \mathbb{R}^{n_o \times n_\alpha}, \overline{\mathbf{G}} \in \mathbb{R}^{n_\alpha \times n_s}, \boldsymbol{v}_k \in \mathbb{R}^{n_o}$ are the variables of the observation equation. Note that $\overline{\mathbf{A}} = \mathbf{T}^{-1}\mathbf{A}^\alpha\mathbf{T}$ and $\overline{\mathbf{G}} = \mathbf{G}^\alpha\mathbf{T}$ and $\overline{\boldsymbol{w}}_k = \mathbf{T}^{-1}\boldsymbol{w}_k^\alpha$. The VSL states $\bar{\boldsymbol{x}}_k$ represent the modal displacements and modal velocities and no longer carry any physical meaning. The SSM described by Eq.(5.7) is henceforth referred to as the MSM.

The MSM imbibes its probabilistic description from the distributions of $\overline{\boldsymbol{w}}_k$ and $\boldsymbol{v}_k$ as follows:

$$p\left(\bar{\boldsymbol{x}}_{k+1} \mid \bar{\boldsymbol{x}}_k, \overline{\mathbf{A}}, \overline{\mathbf{Q}}\right) = \mathcal{N}\left(\overline{\mathbf{A}}\bar{\boldsymbol{x}}_k, \overline{\mathbf{Q}}\right), \quad k = 1, \ldots, N \tag{5.8a}$$

$$p\left(\boldsymbol{y}_k \mid \bar{\boldsymbol{x}}_k, \overline{\mathbf{G}}, \mathbf{R}\right) = \mathcal{N}\left(\boldsymbol{\Omega}_k\overline{\mathbf{G}}\bar{\boldsymbol{x}}_k, \mathbf{R}\right), \quad k = 1, \ldots, N \tag{5.8b}$$

$$p\left(\bar{\boldsymbol{x}}_1 \mid \bar{\boldsymbol{\mu}}_1, \overline{\mathbf{V}}_1\right) = \mathcal{N}\left(\bar{\boldsymbol{\mu}}_1, \overline{\mathbf{V}}_1\right). \tag{5.8c}$$

Note that the sequence of MSR matrices, $\{\boldsymbol{\Omega}_k\}_{1:N}$, are assumed to be deterministically known, and are determined using the VSLs and the sensors' positions (stored in the sensor-position matrix). The joint distribution over the sequence of states and observations is given by

$$p\left(\overline{\boldsymbol{X}}, \boldsymbol{Y} \mid \overline{\mathbf{A}}, \overline{\mathbf{G}}, \overline{\mathbf{Q}}, \mathbf{R}, \bar{\boldsymbol{\mu}}_1, \overline{\mathbf{V}}_1\right) = p\left(\bar{\boldsymbol{x}}_1 \mid \bar{\boldsymbol{\mu}}_1, \overline{\mathbf{V}}_1\right) \prod_{k=1}^{N} p\left(\bar{\boldsymbol{x}}_{k+1} \mid \bar{\boldsymbol{x}}_k, \overline{\mathbf{A}}, \overline{\mathbf{Q}}\right) p\left(\boldsymbol{y}_k \mid \bar{\boldsymbol{x}}_k, \overline{\mathbf{G}}, \mathbf{R}\right) \tag{5.9}$$

where $\overline{\boldsymbol{X}} = \bar{\boldsymbol{x}}_{1:N+1}$ and $\boldsymbol{Y} = \boldsymbol{y}_{1:N}$. The above equation expresses the joint distribution of variables $\overline{\boldsymbol{X}}$ and $\boldsymbol{Y}$ given the parameters $\boldsymbol{\theta} = \left\{\overline{\mathbf{A}}, \overline{\mathbf{G}}, \overline{\mathbf{Q}}, \mathbf{R}\right\}$ (with auxiliary parameters $\bar{\boldsymbol{\mu}}_1, \overline{\mathbf{V}}_1$) of the MSM as the product of Gaussian marginal and conditional distributions over $\overline{\boldsymbol{X}}$ and $\boldsymbol{Y}$ respectively.

In the following sections, three computational algorithms, namely EM, VB and GS, are employed for modal identification with stacked mobile sensor data matrix. Due to the presence of time-varying MSR terms in the observation equation of the MSM, all the three algorithms are modified and new update rules are formulated to allow inference with the proposed MSM.

5.3 ML estimation of MSM via EM

This section concerns the maximum likelihood estimation of the MSM, in particular, the application of EM to obtain ML estimates of the MSM parameters. The objective then lies in finding the parameters $\boldsymbol{\theta} = \left\{\overline{\mathbf{A}}, \overline{\mathbf{G}}, \overline{\mathbf{Q}}, \mathbf{R}\right\}$ that maximizes the likelihood of the MSM given the measured mobile sensor observations in the form of stacked data matrix.

Due to the presence of a time-varying MSR term in the MSM, the E-step and M-step of the the EM algorithm require modification. To derive the required modifications, it is helpful to consider the complete data log-likelihood of the MSM as shown below:

$$\begin{aligned}
\overline{\mathcal{L}}(\boldsymbol{\theta}) &= \ln\left(p\left(\overline{\boldsymbol{X}}, \boldsymbol{Y} \mid \boldsymbol{\theta}\right)\right) \\
&= \ln\left(p\left(\overline{\boldsymbol{x}}_1 \mid \overline{\boldsymbol{\mu}}_1, \overline{\mathbf{V}}_1\right) \prod_{k=1}^{N} p\left(\overline{\boldsymbol{x}}_{k+1} \mid \overline{\boldsymbol{x}}_k, \overline{\mathbf{A}}, \overline{\mathbf{Q}}\right) \prod_{k=1}^{N} p\left(\boldsymbol{y}_k \mid \overline{\boldsymbol{x}}_k, \boldsymbol{\Omega}_k, \overline{\mathbf{G}}, \mathbf{R}\right)\right) \\
&= -\frac{n_s}{2}\ln 2\pi - \frac{1}{2}\ln\left|\overline{\mathbf{V}}_1\right| - \frac{1}{2}\left(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{\mu}}_1\right)^T \overline{\mathbf{V}}_1^{-1}\left(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{\mu}}_1\right) \\
&\quad - \frac{n_s}{2}\ln 2\pi - \frac{N}{2}\ln\left|\overline{\mathbf{Q}}\right| - \frac{1}{2}\sum_{k=1}^{N}\left(\overline{\boldsymbol{x}}_{k+1} - \overline{\mathbf{A}}\overline{\boldsymbol{x}}_k\right)^T \overline{\mathbf{Q}}^{-1}\left(\overline{\boldsymbol{x}}_{k+1} - \overline{\mathbf{A}}\overline{\boldsymbol{x}}_k\right) \\
&\quad - \frac{n_o N}{2}\ln 2\pi - \frac{N}{2}\ln|\mathbf{R}| - \frac{1}{2}\sum_{k=1}^{N}\left(\boldsymbol{y}_k - \boldsymbol{\Omega}_k\overline{\mathbf{G}}\overline{\boldsymbol{x}}_k\right)^T \mathbf{R}^{-1}\left(\boldsymbol{y}_k - \boldsymbol{\Omega}_k\overline{\mathbf{G}}\overline{\boldsymbol{x}}_k\right).
\end{aligned} \tag{5.10}$$

The conditional expectation of the complete-data log-likelihood function given the observation sequence $\boldsymbol{Y}$ and parameter $\boldsymbol{\theta}^{(j)}$ is defined as

$$\overline{\mathcal{F}}\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(j)}\right) = \left\langle \overline{\mathcal{L}}(\boldsymbol{\theta}) \mid \boldsymbol{Y}, \boldsymbol{\theta}^{(j)} \right\rangle. \tag{5.11}$$

After simplifying, $\overline{\mathcal{F}}\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(j)}\right)$ can be written as

$$\begin{aligned}\overline{\mathcal{F}}\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(j)}\right) = &-\frac{1}{2}\ln\left|\overline{\mathbf{V}}_1\right| - \frac{N}{2}\ln\left|\overline{\mathbf{Q}}\right| - \frac{N}{2}\ln|\mathbf{R}| \\ &-\frac{1}{2}\mathrm{tr}\left\{\overline{\mathbf{V}}_1{}^{-1}\left[\hat{\overline{\mathbf{V}}}_{1|N} + \left(\hat{\overline{\boldsymbol{x}}}_{1|N} - \overline{\boldsymbol{\mu}}_1\right)\left(\hat{\overline{\boldsymbol{x}}}_{1|N} - \overline{\boldsymbol{\mu}}_1\right)^T\right]\right\} \\ &-\frac{1}{2}\sum_{k=1}^{N}\mathrm{tr}\left\{\overline{\mathbf{Q}}^{-1}\left[\mathbf{S}_{\bar{f}\bar{f},k} - \overline{\mathbf{A}}\mathbf{S}_{\bar{f}\bar{x},k}^T - \mathbf{S}_{\bar{f}\bar{x},k}\overline{\mathbf{A}}^T + \overline{\mathbf{A}}\mathbf{S}_{\bar{x}\bar{x},k}\overline{\mathbf{A}}^T\right]\right\} \\ &-\frac{1}{2}\sum_{k=1}^{N}\mathrm{tr}\left\{\mathbf{R}^{-1}\left[\mathbf{S}_{yy,k} - \boldsymbol{\Omega}_k\overline{\mathbf{G}}\mathbf{S}_{y\bar{x},k}^T - \mathbf{S}_{y\bar{x},k}\overline{\mathbf{G}}^T\boldsymbol{\Omega}_k^T + \boldsymbol{\Omega}_k\overline{\mathbf{G}}\mathbf{S}_{\bar{x}\bar{x},k}\overline{\mathbf{G}}^T\boldsymbol{\Omega}_k^T\right]\right\}\end{aligned} \tag{5.12}$$

where the expectations of sufficient statistics are defined as following:

$$\mathbf{S}_{\bar{f}\bar{f}} = \sum_{k=1}^{N}\mathbf{S}_{\bar{f}\bar{f},k} = \sum_{k=1}^{N}\left\langle\overline{\boldsymbol{x}}_{k+1}\overline{\boldsymbol{x}}_{k+1}^T\right\rangle \tag{5.13a}$$

$$\mathbf{S}_{\bar{f}\bar{x}} = \sum_{k=1}^{N}\mathbf{S}_{\bar{f}\bar{x},k} = \sum_{k=1}^{N}\left\langle\overline{\boldsymbol{x}}_{k+1}\overline{\boldsymbol{x}}_{k}^T\right\rangle \tag{5.13b}$$

$$\mathbf{S}_{\bar{x}\bar{x}} = \sum_{k=1}^{N}\mathbf{S}_{\bar{x}\bar{x},k} = \sum_{k=1}^{N}\left\langle\overline{\boldsymbol{x}}_{k}\overline{\boldsymbol{x}}_{k}^T\right\rangle \tag{5.13c}$$

$$\mathbf{S}_{y\bar{x}} = \sum_{k=1}^{N}\mathbf{S}_{yx,k} = \sum_{k=1}^{N}\boldsymbol{y}_k\left\langle\boldsymbol{x}_{k}^T\right\rangle \tag{5.13d}$$

$$\mathbf{S}_{yy} = \sum_{k=1}^{N}\mathbf{S}_{yy,k} = \sum_{k=1}^{N}\boldsymbol{y}_k\boldsymbol{y}_{k}^T. \tag{5.13e}$$

The derivations of the E-step and M-steps of the modified EM algorithm are presented next.

5.3.1 Derivation of E-step

The expected values of the states given by Eq.(5.13) can be obtained using the Kalman filter and the RTS smoother, however due to the time-varying nature of the MSR, an

equivalent observation matrix $\widetilde{\overline{\mathbf{G}}}_k = \mathbf{\Omega}_k \overline{\mathbf{G}}$ must be used at each time-step.

Kalman filter for MSM : Do for $k = 1, \ldots, N$

$$\boldsymbol{e}_k = \boldsymbol{y}_k - \widetilde{\overline{\mathbf{G}}}_k \hat{\bar{\boldsymbol{x}}}_{k|k-1} \tag{5.14a}$$

$$\mathbf{S}_k = \widetilde{\overline{\mathbf{G}}}_k \hat{\overline{\mathbf{V}}}_{k|k-1} \widetilde{\overline{\mathbf{G}}}_k^T + \mathbf{R} \tag{5.14b}$$

$$\mathbf{K}_k = \hat{\overline{\mathbf{V}}}_{k|k-1} \widetilde{\overline{\mathbf{G}}}_k^T \mathbf{S}_k^{-1} \tag{5.14c}$$

$$\hat{\bar{\boldsymbol{x}}}_{k|k} = \hat{\bar{\boldsymbol{x}}}_{k|k-1} + \mathbf{K}_k \boldsymbol{e}_k \tag{5.14d}$$

$$\hat{\overline{\mathbf{V}}}_{k|k} = \hat{\overline{\mathbf{V}}}_{k|k-1} - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^T \tag{5.14e}$$

$$\hat{\bar{\boldsymbol{x}}}_{k+1|k} = \overline{\mathbf{A}} \hat{\bar{\boldsymbol{x}}}_{k|k} \tag{5.14f}$$

$$\hat{\overline{\mathbf{V}}}_{k+1|k} = \overline{\mathbf{A}} \hat{\overline{\mathbf{V}}}_{k|k} \overline{\mathbf{A}} + \mathbf{Q}. \tag{5.14g}$$

RTS smoother for MSM : Do for $k = N, \ldots, 1$

$$\mathbf{N}_k = \hat{\mathbf{V}}_{k|k} \overline{\mathbf{A}}^T \left(\hat{\mathbf{V}}_{k+1|k} \right)^{-1} \tag{5.15a}$$

$$\hat{\boldsymbol{x}}_{k|N} = \hat{\boldsymbol{x}}_{k|k} + \mathbf{N}_k \left(\hat{\boldsymbol{x}}_{k+1|N} - \hat{\boldsymbol{x}}_{k+1|k} \right) \tag{5.15b}$$

$$\hat{\mathbf{V}}_{k|N} = \hat{\mathbf{V}}_{k|k} + \mathbf{N}_k \left(\hat{\mathbf{V}}_{k+1|N} - \hat{\mathbf{V}}_{k+1|k} \right) \mathbf{N}_k^T \tag{5.15c}$$

$$\hat{\mathbf{V}}_{k+1,k|N} = \hat{\mathbf{V}}_{k+1|N} \mathbf{N}_k^T. \tag{5.15d}$$

Using the results from the RTS smoother, the expectations in Eq.(5.13) can be obtained:

$$\begin{aligned}
\mathbf{S}_{y\bar{x},k} &= \boldsymbol{y}_k \langle \boldsymbol{x}_k \rangle = \boldsymbol{y}_k \hat{\bar{\boldsymbol{x}}}_{k|N} \\
\mathbf{S}_{\bar{f}\bar{f},k} &= \left\langle \overline{\boldsymbol{x}}_{k+1} \overline{\boldsymbol{x}}_{k+1}^T \right\rangle = \hat{\bar{\boldsymbol{x}}}_{k+1|N} \hat{\bar{\boldsymbol{x}}}_{k+1|N}^T + \hat{\overline{\mathbf{V}}}_{k+1|N} \\
\mathbf{S}_{\bar{f}\bar{x},k} &= \left\langle \overline{\boldsymbol{x}}_{k+1} \overline{\boldsymbol{x}}_k^T \right\rangle = \hat{\bar{\boldsymbol{x}}}_{k+1|N} \hat{\bar{\boldsymbol{x}}}_{k|N}^T + \hat{\overline{\mathbf{V}}}_{k+1,k|N} \\
\mathbf{S}_{\bar{x}\bar{x},k} &= \left\langle \overline{\boldsymbol{x}}_k \overline{\boldsymbol{x}}_k^T \right\rangle = \hat{\bar{\boldsymbol{x}}}_{k|N} \hat{\bar{\boldsymbol{x}}}_{k|N}^T + \hat{\overline{\mathbf{V}}}_{k|N}.
\end{aligned} \tag{5.16}$$

The log-likelihood at the jth EM iteration can be calculated using the innovations from the Kalman filter as

$$\hat{\bar{\mathcal{F}}}^{(j)} = -\frac{1}{2} \sum_{k=1}^{N} \left(\ln |\mathbf{S}_k| + \boldsymbol{e}_k^T {\mathbf{S}_k}^{-1} \boldsymbol{e}_k \right). \tag{5.17}$$

5.3.2 Derivation of M-step

Maximizing $\overline{\mathcal{F}}\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(j)}\right)$ of Eq.(5.12)) with respect to the MSM parameters $\boldsymbol{\theta}$ at the jth iteration constitutes the M-step of EM. The M-step maximizations are obtained in closed form by setting the gradient $\frac{\partial \overline{\mathcal{F}}\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)}\right)}{\partial \boldsymbol{\theta}} = \mathbf{0}$ for each $\boldsymbol{\theta} = \left\{\overline{\mathbf{A}}, \overline{\mathbf{G}}, \overline{\mathbf{Q}}, \mathbf{R}, \bar{\boldsymbol{\mu}}_1, \overline{\mathbf{V}}_1\right\}$. The update equations for $\overline{\mathbf{A}}^{(j+1)}, \overline{\mathbf{Q}}^{(j+1)}$ and $\bar{\boldsymbol{\mu}}_1^{(j+1)}, \overline{\mathbf{V}}_1^{(j+1)}$ end up being

$$\overline{\mathbf{A}}^{(j+1)} = \left(\bar{\mathbf{S}}_{fx}\right)\left(\bar{\mathbf{S}}_{fx}\right)^{-1} \tag{5.18a}$$

$$\overline{\mathbf{Q}}^{(j+1)} = \frac{1}{N}\left(\bar{\mathbf{S}}_{ff} - \bar{\mathbf{S}}_{fx}\overline{\mathbf{A}}^{(j+1)^T} - \overline{\mathbf{A}}^{(j+1)}\bar{\mathbf{S}}_{fx}^T + \overline{\mathbf{A}}^{(j+1)}\bar{\mathbf{S}}_{xx}\overline{\mathbf{A}}^{(j+1)^T}\right) \tag{5.18b}$$

$$\begin{aligned} \bar{\boldsymbol{\mu}}_1^{(j+1)} &= \hat{\bar{\boldsymbol{x}}}_{1|N} \\ \overline{\mathbf{V}}_1^{(j+1)} &= \hat{\overline{\mathbf{V}}}_{1|N}. \end{aligned} \tag{5.18c}$$

The update equations of $\overline{\mathbf{G}}^{(j+1)}$ and $\mathbf{R}^{(j+1)}$ get modified due to the MSR term and are obtained as follows:

$$\bar{\boldsymbol{g}}^{(j+1)} = \left(\sum_{k=1}^{N}\left(\bar{\mathbf{S}}_{xx,k} \otimes \left(\boldsymbol{\Omega}_k^T \mathbf{R}^{(j)^{-1}} \boldsymbol{\Omega}_k\right)\right)\right)^{-1} \operatorname{vec}\left(\sum_{k=1}^{N} \boldsymbol{\Omega}_k^T \mathbf{R}^{(j)^{-1}} \bar{\mathbf{S}}_{yx,k}\right) \tag{5.19a}$$

$$\overline{\mathbf{G}}^{(j+1)} = \operatorname{mat}\left(\bar{\boldsymbol{g}}^{(j+1)}\right) \tag{5.19b}$$

$$\mathbf{R}^{(j+1)} = \frac{1}{N}\sum_{k=1}^{N}\left(\bar{\mathbf{S}}_{yy,k} - \boldsymbol{\Omega}_k\overline{\mathbf{G}}^{(j+1)}\bar{\mathbf{S}}_{yx,k}^T - \bar{\mathbf{S}}_{yx,k}\overline{\mathbf{G}}^{(j+1)^T}\boldsymbol{\Omega}_k^T + \boldsymbol{\Omega}_k\overline{\mathbf{G}}^{(j+1)}\bar{\mathbf{S}}_{xx,k}\overline{\mathbf{G}}^{(j+1)^T}\boldsymbol{\Omega}_k^T\right) \tag{5.19c}$$

where $\bar{\boldsymbol{g}}^{(j+1)}$ is the vectorized form of $\overline{\mathbf{G}}^{(j+1)}$ and $\operatorname{mat}(\cdot)$ is an operator that reshapes vectors to matrices of relevant dimensions. In this case, $\bar{\boldsymbol{g}}^{(j+1)}$ is a vector of size $(n_\alpha n_s \times 1)$, and $\operatorname{mat}\left(\bar{\boldsymbol{g}}^{(j+1)}\right)$ has the same size as $\overline{\mathbf{G}}$. Note that the update expression for $\bar{\boldsymbol{g}}^{(j+1)}$ differs from the corresponding STRIDEX update equation [21] in the sense that the update of $\bar{\boldsymbol{g}}^{(j+1)}$ in Eq.(5.19) depends on $\mathbf{R}^{(j)}$. The derivations of the above expressions have been provided in the Appendix E.1.

The convergence criterion and the initialization procedure for EM using the MSM follow similarly as discussed in sections 4.4.4 and 4.4.3.

A complete list of steps for applying the EM algorithm to the stacked data case is provided in Algorithm 4.

Algorithm 4 Modal parameter estimation using EM with stacked data approach

Input: $\boldsymbol{Y}$, $\overline{\mathbf{A}}^{(0)}$, $\overline{\mathbf{G}}^{(0)}$, $\overline{\mathbf{Q}}^{(0)}$, $\mathbf{R}^{(0)}$, $\bar{\boldsymbol{\mu}}_1^{(0)}$, $\overline{\mathbf{V}}_1^{(0)}$, ϵ_{tol}, $\{\boldsymbol{s}_k^o\}_{1:N}$

Choose the uniformly spaced locations of VSLs, $\boldsymbol{s}^\alpha$

Build the sequence of mode shape regression matrices $\{\boldsymbol{\Omega}\}_{1:N}$ using Eq.(5.5)

E-step:

1. Use Eqs.(5.14) and (5.15) for Kalman filter and RTS smoother recursions
2. Use Eq.(5.17) to compute $\hat{\overline{\mathcal{F}}}^{(j)}$
3. Calculate the expectations of sufficient statistics using Eqs.(5.13) and (5.16)

while not converged **do**

1. **M**-step: Use Eqs.(5.18) and (5.19) to update parameters
   $\rightarrow \overline{\mathbf{A}}^{(j+1)}, \overline{\mathbf{G}}^{(j+1)}, \overline{\mathbf{Q}}^{(j+1)}, \mathbf{R}^{(j+1)}, \bar{\boldsymbol{\mu}}_1^{(j+1)}, \overline{\mathbf{V}}_1^{(j+1)}$
2. **E**-step:
   (a) Use Eqs.(5.14) and (5.15) for Kalman filter and RTS smoother recursions
   (b) Use Eq.(5.17) to compute $\hat{\overline{\mathcal{F}}}^{(j+1)}$
   (c) Calculate the expectations of sufficient statistics using Eqs.(5.13) and (5.16)
3. Use Eq.(4.20) to check if convergence criterion is satisfied

end while

Return $\hat{\overline{\mathbf{A}}}_{\text{ML}} \leftarrow \overline{\mathbf{A}}^{(final)}$, $\hat{\overline{\mathbf{G}}}_{\text{ML}} \leftarrow \overline{\mathbf{G}}^{(final)}$

Extract modal frequencies, damping ratios and un-normalized mode shapes from $\hat{\overline{\mathbf{A}}}_{\text{ML}}, \hat{\overline{\mathbf{G}}}_{\text{ML}}$ using Appendix B.1

Output: $\hat{f}_i, \hat{\xi}_i, \hat{\boldsymbol{\phi}}_i^{un}$ for all modes $i = 1, 2, \ldots$

5.4 Bayesian inference of MSM via VB

This section concerns the application of VB to the MSM in Eq.(5.7) and how the VB for MSM reduces to a combination of parameter update equations with an augmented Kalman filter and RTS smoothing algorithm. The joint distribution of the measurements, latent

states and the parameters of the MSM satisfies the following relationship

$$p\left(\boldsymbol{Y}, \overline{\boldsymbol{X}}, \overline{\mathbf{A}}, \overline{\mathbf{Q}}, \overline{\mathbf{G}}, \mathbf{R}\right) = p\left(\boldsymbol{Y}, \overline{\boldsymbol{X}} \mid \overline{\mathbf{A}}, \overline{\mathbf{Q}}, \overline{\mathbf{G}}, \mathbf{R}\right) p\left(\overline{\mathbf{A}}, \overline{\mathbf{Q}}, \overline{\mathbf{G}}, \mathbf{R}\right) \tag{5.20}$$

with

$$p\left(\overline{\boldsymbol{x}}_1 \mid \overline{\boldsymbol{\mu}}_1, \overline{\mathbf{V}}_1\right) = \mathcal{N}\left(\overline{\boldsymbol{\mu}}_1, \overline{\mathbf{V}}_1\right) = \frac{\exp\left(-\frac{1}{2}(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{\mu}}_1)^T \overline{\mathbf{V}}_1^{-1}(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{\mu}}_1)\right)}{(2\pi)^{n_s/2}\left|\overline{\mathbf{V}}_1\right|^{1/2}} \tag{5.21a}$$

$$p\left(\overline{\boldsymbol{x}}_{k+1} \mid \overline{\boldsymbol{x}}_k, \overline{\mathbf{A}}, \overline{\mathbf{Q}}\right) = \mathcal{N}\left(\overline{\mathbf{A}}\overline{\boldsymbol{x}}_k, \overline{\mathbf{Q}}\right) = \frac{\exp\left(-\frac{1}{2}(\overline{\boldsymbol{x}}_{k+1} - \overline{\mathbf{A}}\overline{\boldsymbol{x}}_k)^T \overline{\mathbf{Q}}^{-1}(\overline{\boldsymbol{x}}_{k+1} - \overline{\mathbf{A}}\overline{\boldsymbol{x}}_k)\right)}{(2\pi)^{n_s/2}\left|\overline{\mathbf{Q}}\right|^{1/2}} \tag{5.21b}$$

$$p\left(\boldsymbol{y}_k \mid \overline{\boldsymbol{x}}_k, \overline{\mathbf{G}}, \mathbf{R}\right) = \mathcal{N}\left(\overline{\mathbf{G}}\overline{\boldsymbol{x}}_k, \mathbf{R}\right) = \frac{\exp\left(-\frac{1}{2}(\boldsymbol{y}_k - \boldsymbol{\Omega}_k\overline{\mathbf{G}}\overline{\boldsymbol{x}}_k)^T \mathbf{R}^{-1}(\boldsymbol{y}_k - \boldsymbol{\Omega}_k\overline{\mathbf{G}}\overline{\boldsymbol{x}}_k)\right)}{(2\pi)^{n_o/2}\left|\mathbf{R}\right|^{1/2}} \tag{5.21c}$$

An approximate factorized prior distribution of the parameters of the MSM is assumed as follows:

$$p\left(\overline{\mathbf{A}}, \overline{\mathbf{Q}}, \overline{\mathbf{G}}, \mathbf{R}\right) = p\left(\overline{\mathbf{A}}\right) p\left(\overline{\mathbf{Q}}\right) p\left(\overline{\mathbf{G}}\right) p\left(\mathbf{R}\right) \tag{5.22}$$

where the matrices $\overline{\mathbf{A}}$ and $\overline{\mathbf{G}}$ are no longer assumed to be conditionally dependent on $\overline{\mathbf{Q}}$ and $\mathbf{R}$ respectively, unlike in previous chapter. The prior distributions of $\overline{\mathbf{A}}$ and $\overline{\mathbf{G}}$ are now defined by multivariate normal distributions over their vectorized forms (i.e. $\bar{\boldsymbol{a}} = \text{vec}\left(\overline{\mathbf{A}}\right)$ and $\bar{\boldsymbol{g}} = \text{vec}\left(\overline{\mathbf{G}}\right)$) as follows:

$$p\left(\overline{\mathbf{A}}\right) = p\left(\bar{\boldsymbol{a}}\right) = \mathcal{N}\left(\boldsymbol{\mu}_{\bar{\boldsymbol{a}}}, \boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}\right) = \frac{\exp\left(-\frac{1}{2}\left(\bar{\boldsymbol{a}} - \boldsymbol{\mu}_{\bar{\boldsymbol{a}}}\right)^T \boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{-1}\left(\bar{\boldsymbol{a}} - \boldsymbol{\mu}_{\bar{\boldsymbol{a}}}\right)\right)}{(2\pi)^{n_s^2/2}\left|\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}\right|^{\frac{1}{2}}} \tag{5.23a}$$

$$p\left(\overline{\mathbf{G}}\right) = p\left(\bar{\boldsymbol{g}}\right) = \mathcal{N}\left(\boldsymbol{\mu}_{\bar{\boldsymbol{g}}}, \boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}\right) = \frac{\exp\left(-\frac{1}{2}\left(\bar{\boldsymbol{g}} - \boldsymbol{\mu}_{\bar{\boldsymbol{g}}}\right)^T \boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}^{-1}\left(\bar{\boldsymbol{g}} - \boldsymbol{\mu}_{\bar{\boldsymbol{g}}}\right)\right)}{(2\pi)^{n_s n_\alpha/2}\left|\boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}\right|^{\frac{1}{2}}} \tag{5.23b}$$

where $\mathcal{N}(\cdot)$ is the *Gaussian* (Normal) distribution. The prior distributions of $\overline{\mathbf{Q}}$ and $\mathbf{R}$ are given by

$$p\left(\overline{\mathbf{Q}}\right) = \mathcal{IW}\left(d_{\overline{\mathbf{Q}}}, \mathbf{D}_{\overline{\mathbf{Q}}}\right) = \frac{\left|\mathbf{D}_{\overline{\mathbf{Q}}}\right|^{d_{\overline{\mathbf{Q}}}/2}\left|\overline{\mathbf{Q}}\right|^{-\left(d_{\overline{\mathbf{Q}}}+n_s+1\right)/2}\exp\left(-\frac{1}{2}\,\text{tr}\left\{\mathbf{D}_{\overline{\mathbf{Q}}}\overline{\mathbf{Q}}^{-1}\right\}\right)}{2^{n_s d_{\overline{\mathbf{Q}}}/2}\;\;\Gamma_{n_s}\left(\frac{d_{\overline{\mathbf{Q}}}}{2}\right)} \tag{5.23c}$$

$$p\left(\mathbf{R}\right) = \mathcal{IW}\left(d_{\mathbf{R}}, \mathbf{D}_{\mathbf{R}}\right) = \frac{\left|\mathbf{D}_{\mathbf{R}}\right|^{d_{\mathbf{R}}/2} \left|\mathbf{R}\right|^{-(d_{\mathbf{R}}+n_o+1)/2} \exp\left(-\frac{1}{2}\operatorname{tr}\left\{\mathbf{D}_{\mathbf{R}}\mathbf{R}^{-1}\right\}\right)}{2^{n_o d_{\mathbf{R}}/2}\ \Gamma_{n_o}\left(\frac{d_{\mathbf{R}}}{2}\right)} \tag{5.23d}$$

where $\mathcal{IW}(\cdot)$ is the *Inverse Wishart* distribution. The set of prior hyperparameters is denoted by $\boldsymbol{\pi}_h = \left\{\boldsymbol{\mu}_{\bar{\boldsymbol{a}}}, \boldsymbol{\mu}_{\bar{\boldsymbol{g}}}, \boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}, \boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}, \mathbf{D}_{\overline{\mathbf{Q}}}, \mathbf{D}_{\mathbf{R}}, d_{\overline{\mathbf{Q}}}, d_{\mathbf{R}}\right\}$. The next step is to assume some approximate factorized form of the true posterior distribution using the variational distribution $q\left(\cdot\right)$ over the unobserved variables $\left\{\overline{\boldsymbol{X}}, \overline{\mathbf{A}}, \overline{\mathbf{G}}, \overline{\mathbf{Q}}, \mathbf{R}\right\}$ which leads to a tractable bound:

$$p\left(\overline{\mathbf{A}}, \overline{\mathbf{Q}}, \overline{\mathbf{G}}, \mathbf{R}, \overline{\boldsymbol{X}} \mid \boldsymbol{Y}\right) \approx q\left(\bar{\boldsymbol{a}}\right) q\left(\overline{\mathbf{Q}}\right) q\left(\bar{\boldsymbol{g}}\right) q\left(\mathbf{R}\right) q\left(\overline{\boldsymbol{X}}\right). \tag{5.24}$$

It is to be noted that this factorization amongst the parameters $\left\{\overline{\mathbf{A}}, \overline{\mathbf{G}}, \overline{\mathbf{Q}}, \mathbf{R}\right\}$ falls out of the initial factorization assumed in Eq.(5.22). The optimum form of the approximate posteriors is found by taking functional derivatives of the variational free energy with respect to each distribution over the parameters and the latent state variables.

The variational free energy [107] is maximized by setting:

$$q^{(j)}\left(\overline{\boldsymbol{X}}\right) \propto \exp\left(\left\langle \ln p\left(\boldsymbol{Y}, \overline{\boldsymbol{X}}, \overline{\mathbf{A}}, \overline{\mathbf{Q}}, \overline{\mathbf{G}}, \mathbf{R}\right)\right\rangle_{q^{(j)}(\bar{\boldsymbol{a}})q^{(j)}\left(\overline{\mathbf{Q}}\right)q^{(j)}(\bar{\boldsymbol{g}})q^{(j)}(\mathbf{R})}\right) \tag{5.25a}$$

$$q^{(j+1)}\left(\bar{\boldsymbol{a}}\right) \propto \exp\left(\left\langle \ln p\left(\boldsymbol{Y}, \overline{\boldsymbol{X}}, \overline{\mathbf{A}}, \overline{\mathbf{Q}}, \overline{\mathbf{G}}, \mathbf{R}\right)\right\rangle_{q^{(j)}\left(\overline{\mathbf{Q}}\right)q^{(j)}(\bar{\boldsymbol{g}})q^{(j)}(\mathbf{R})q^{(j)}\left(\overline{\boldsymbol{X}}\right)}\right) \tag{5.25b}$$

$$q^{(j+1)}\left(\bar{\boldsymbol{g}}\right) \propto \exp\left(\left\langle \ln p\left(\boldsymbol{Y}, \overline{\boldsymbol{X}}, \overline{\mathbf{A}}, \overline{\mathbf{Q}}, \overline{\mathbf{G}}, \mathbf{R}\right)\right\rangle_{q^{(j)}(\bar{\boldsymbol{a}})q^{(j)}\left(\overline{\mathbf{Q}}\right)q^{(j)}(\mathbf{R})q^{(j)}\left(\overline{\boldsymbol{X}}\right)}\right) \tag{5.25c}$$

$$q^{(j+1)}\left(\overline{\mathbf{Q}}\right) \propto \exp\left(\left\langle \ln p\left(\boldsymbol{Y}, \overline{\boldsymbol{X}}, \overline{\mathbf{A}}, \overline{\mathbf{Q}}, \overline{\mathbf{G}}, \mathbf{R}\right)\right\rangle_{q^{(j)}(\bar{\boldsymbol{a}})q^{(j)}(\bar{\boldsymbol{g}})q^{(j)}(\mathbf{R})q^{(j)}\left(\overline{\boldsymbol{X}}\right)}\right) \tag{5.25d}$$

$$q^{(j+1)}\left(\mathbf{R}\right) \propto \exp\left(\left\langle \ln p\left(\boldsymbol{Y}, \overline{\boldsymbol{X}}, \overline{\mathbf{A}}, \overline{\mathbf{Q}}, \overline{\mathbf{G}}, \mathbf{R}\right)\right\rangle_{q^{(j)}(\bar{\boldsymbol{a}})q^{(j)}\left(\overline{\mathbf{Q}}\right)q^{(j)}(\bar{\boldsymbol{g}})q^{(j)}\left(\overline{\boldsymbol{X}}\right)}\right) \tag{5.25e}$$

where $\left\langle f(x)\right\rangle_{q(x)} := \int f(x)q(x)dx$ and the jth variational distributions are defined as

$$q^{(j)}\left(\bar{\boldsymbol{a}}\right) = \mathcal{N}\left(\boldsymbol{\mu}_{\bar{\boldsymbol{a}}}^{(j)}, \boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{(j)}\right) = \frac{\exp\left(-\frac{1}{2}\left(\bar{\boldsymbol{a}} - \boldsymbol{\mu}_{\bar{\boldsymbol{a}}}^{(j)}\right)^T \boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{-1}\left(\bar{\boldsymbol{a}} - \boldsymbol{\mu}_{\bar{\boldsymbol{a}}}^{(j)}\right)\right)}{(2\pi)^{n_s^2/2}\left|\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{(j)}\right|^{\frac{1}{2}}} \tag{5.26a}$$

$$q^{(j)}\left(\bar{\boldsymbol{g}}\right) = \mathcal{N}\left(\boldsymbol{\mu}_{\bar{\boldsymbol{g}}}^{(j)}, \boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}^{(j)}\right) = \frac{\exp\left(-\frac{1}{2}\left(\bar{\boldsymbol{g}} - \boldsymbol{\mu}_{\bar{\boldsymbol{g}}}^{(j)}\right)^T \boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}^{-1}\left(\bar{\boldsymbol{g}} - \boldsymbol{\mu}_{\bar{\boldsymbol{g}}}^{(j)}\right)\right)}{(2\pi)^{n_s n_\alpha/2}\left|\boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}^{(j)}\right|^{\frac{1}{2}}} \tag{5.26b}$$

$$q^{(j)}\left(\overline{\mathbf{Q}}\right)=\mathcal{IW}\left(d_{\overline{\mathbf{Q}}}^{(j)},\mathbf{D}_{\overline{\mathbf{Q}}}^{(j)}\right)=\frac{\left|\mathbf{D}_{\overline{\mathbf{Q}}}^{(j)}\right|^{d_{\overline{\mathbf{Q}}}^{(j)}/2}\left|\overline{\mathbf{Q}}\right|^{-\left(d_{\overline{\mathbf{Q}}}^{(j)}+n_s+1\right)/2}\exp\left(-\frac{1}{2}\operatorname{tr}\left\{\mathbf{D}_{\overline{\mathbf{Q}}}^{(j)}\overline{\mathbf{Q}}^{-1}\right\}\right)}{2^{n_s d_{\overline{\mathbf{Q}}}^{(j)}/2}\ \Gamma_{n_s}\left(\frac{d_{\overline{\mathbf{Q}}}^{(j)}}{2}\right)} \tag{5.26c}$$

$$q^{(j)}\left(\mathbf{R}\right)=\mathcal{IW}\left(d_{\mathbf{R}}^{(j)},\mathbf{D}_{\mathbf{R}}^{(j)}\right)=\frac{\left|\mathbf{D}_{\mathbf{R}}^{(j)}\right|^{d_{\mathbf{R}}^{(j)}/2}\left|\mathbf{R}\right|^{-\left(d_{\mathbf{R}}^{(j)}+n_o+1\right)/2}\exp\left(-\frac{1}{2}\operatorname{tr}\left\{\mathbf{D}_{\mathbf{R}}^{(j)}\mathbf{R}^{-1}\right\}\right)}{2^{n_o d_{\mathbf{R}}^{(j)}/2}\ \Gamma_{n_o}\left(\frac{d_{\mathbf{R}}^{(j)}}{2}\right)}. \tag{5.26d}$$

The VB algorithm works by iteratively updating the set of hyperparameters

$$\boldsymbol{\theta}_h^{(j)}=\left\{\boldsymbol{\mu}_{\bar{\boldsymbol{a}}}^{(j)},\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{(j)},\boldsymbol{\mu}_{\bar{\boldsymbol{g}}}^{(j)},\boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}^{(j)},d_{\overline{\mathbf{Q}}}^{(j)},\mathbf{D}_{\overline{\mathbf{Q}}}^{(j)},d_{\mathbf{R}}^{(j)},\mathbf{D}_{\mathbf{R}}^{(j)}\right\} \tag{5.27}$$

of the variational distributions in Eq.(5.26) in an effort to maximize the variational free energy with each update. Note that producing the updating rules only requires computation of the variational free energy derivatives.

5.4.1 Evaluation of variational expectation

To derive the update equations for the VBE and VBM steps, it is necessary to evaluate the expectation of the log joint PDF of the hidden MSM variables (i.e., the states and parameters) with respect to the variational PDFs. One can use Eqs.(5.20) and (5.23) to arrive at the following expression for the expected value of the log joint PDF:

$$\begin{aligned}
&\left\langle \ln\left(p\left(\boldsymbol{Y},\overline{\boldsymbol{X}},\overline{\mathbf{A}},\overline{\mathbf{Q}},\overline{\mathbf{G}},\mathbf{R}\right)\right)\right\rangle_{q^{(j)}(\bar{\boldsymbol{a}})q^{(j)}\left(\overline{\mathbf{Q}}\right)q^{(j)}(\bar{\boldsymbol{g}})q^{(j)}(\mathbf{R})q^{(j)}\left(\overline{\boldsymbol{X}}\right)}\\
\propto&-\frac{1}{2}\left\langle(\overline{\boldsymbol{x}}_1-\overline{\boldsymbol{\mu}}_1)^T\overline{\mathbf{V}}_1^{-1}(\overline{\boldsymbol{x}}_1-\overline{\boldsymbol{\mu}}_1)\right\rangle_{q^{(j)}(\overline{\boldsymbol{x}}_1)}\\
&-\frac{1}{2}\left\langle\operatorname{tr}\left\{\overline{\mathbf{Q}}^{-1}\left(\mathbf{S}_{\bar{f}\bar{f}}^{(j)}-\overline{\mathbf{A}}\mathbf{S}_{\bar{f}\bar{x}}^{(j)^T}-\mathbf{S}_{\bar{f}\bar{x}}^{(j)}\overline{\mathbf{A}}^T+\overline{\mathbf{A}}\mathbf{S}_{\bar{x}\bar{x}}^{(j)}\overline{\mathbf{A}}^T\right)\right\}\right\rangle_{q^{(j)}(\bar{\boldsymbol{a}})q^{(j)}\left(\overline{\mathbf{Q}}\right)}\\
&-\frac{1}{2}\sum_{k=1}^{N}\left[\left\langle\operatorname{tr}\left\{\mathbf{R}^{-1}\left(\mathbf{S}_{yy,k}^{(j)}-\boldsymbol{\Omega}_k\overline{\mathbf{G}}\mathbf{S}_{y\bar{x},k}^{(j)^T}-\mathbf{S}_{y\bar{x},k}^{(j)}\overline{\mathbf{G}}^T\boldsymbol{\Omega}_k^T+\boldsymbol{\Omega}_k\overline{\mathbf{G}}\mathbf{S}_{\bar{x}\bar{x},k}^{(j)}\overline{\mathbf{G}}^T\boldsymbol{\Omega}_k^T\right)\right\}\right\rangle_{q^{(j)}(\bar{\boldsymbol{g}})q^{(j)}(\mathbf{R})}\right]
\end{aligned}$$

$$
\begin{aligned}
&-\frac{1}{2}\left\langle \bar{\boldsymbol{a}}^T\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{-1}\bar{\boldsymbol{a}} - \bar{\boldsymbol{a}}^T\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{-1}\boldsymbol{\mu}_{\bar{\boldsymbol{a}}} - \boldsymbol{\mu}_{\bar{\boldsymbol{a}}}^T\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{-1}\bar{\boldsymbol{a}} + \boldsymbol{\mu}_{\bar{\boldsymbol{a}}}^T\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{-1}\boldsymbol{\mu}_{\bar{\boldsymbol{a}}}\right\rangle_{q^{(j)}(\bar{\boldsymbol{a}})} \\
&-\frac{1}{2}\left\langle \bar{\boldsymbol{g}}^T\boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}^{-1}\bar{\boldsymbol{g}} - \bar{\boldsymbol{g}}^T\boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}^{-1}\boldsymbol{\mu}_{\bar{\boldsymbol{g}}} - \boldsymbol{\mu}_{\bar{\boldsymbol{g}}}^T\boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}^{-1}\bar{\boldsymbol{g}} + \boldsymbol{\mu}_{\bar{\boldsymbol{g}}}^T\boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}^{-1}\boldsymbol{\mu}_{\bar{\boldsymbol{g}}}\right\rangle_{q^{(j)}(\bar{\boldsymbol{g}})} \\
&-\frac{d_{\overline{\mathbf{Q}}}+N+n_s+1}{2}\left\langle \ln\left|\overline{\mathbf{Q}}\right|\right\rangle_{q^{(j)}(\overline{\mathbf{Q}})} - \frac{1}{2}\left\langle \mathrm{tr}\left\{\mathbf{D}_{\overline{\mathbf{Q}}}\overline{\mathbf{Q}}^{-1}\right\}\right\rangle_{q^{(j)}(\overline{\mathbf{Q}})} \\
&-\frac{d_{\mathbf{R}}+N+n_o+1}{2}\left\langle \ln|\mathbf{R}|\right\rangle_{q^{(j)}(\mathbf{R})} - \frac{1}{2}\left\langle \mathrm{tr}\left\{\mathbf{D}_{\mathbf{R}}\mathbf{R}^{-1}\right\}\right\rangle_{q^{(j)}(\mathbf{R})}
\end{aligned}
\tag{5.28}
$$

where the matrices $\mathbf{S}^{(j)}_{\bar{x}\bar{x},k}$, $\mathbf{S}^{(j)}_{\bar{f}\bar{x},k}$, $\mathbf{S}^{(j)}_{\bar{f}\bar{f},k}$, $\mathbf{S}^{(j)}_{y\bar{x},k}$ and $\mathbf{S}_{yy,k}$ are the expectations of sufficient statistics, given as:

$$
\mathbf{S}^{(j)}_{\bar{x}\bar{x},k} = \left\langle \bar{\boldsymbol{x}}_k\bar{\boldsymbol{x}}_k^T\right\rangle_{q^{(j)}\left(\overline{\boldsymbol{X}}\right)}, \quad \mathbf{S}^{(j)}_{\bar{x}\bar{x}} = \sum_{k=1}^{N}\mathbf{S}^{(j)}_{\bar{x}\bar{x},k} \tag{5.29a}
$$

$$
\mathbf{S}^{(j)}_{\bar{f}\bar{x},k} = \left\langle \bar{\boldsymbol{x}}_{k+1}\bar{\boldsymbol{x}}_k^T\right\rangle_{q^{(j)}\left(\overline{\boldsymbol{X}}\right)}, \quad \mathbf{S}^{(j)}_{\bar{f}\bar{x}} = \sum_{k=1}^{N}\mathbf{S}^{(j)}_{\bar{f}\bar{x},k} \tag{5.29b}
$$

$$
\mathbf{S}^{(j)}_{\bar{f}\bar{f},k} = \left\langle \bar{\boldsymbol{x}}_{k+1}\bar{\boldsymbol{x}}_{k+1}^T\right\rangle_{q^{(j)}\left(\overline{\boldsymbol{X}}\right)}, \quad \mathbf{S}^{(j)}_{\bar{f}\bar{f}} = \sum_{k=1}^{N}\mathbf{S}^{(j)}_{\bar{f}\bar{f},k} \tag{5.29c}
$$

and

$$
\mathbf{S}^{(j)}_{y\bar{x},k} = \left\langle \boldsymbol{y}_k\bar{\boldsymbol{x}}_k^T\right\rangle_{q^{(j)}\left(\overline{\boldsymbol{X}}\right)}, \quad \mathbf{S}^{(j)}_{y\bar{x}} = \sum_{k=1}^{N}\mathbf{S}^{(j)}_{y\bar{x},k} \tag{5.30a}
$$

$$
\mathbf{S}_{yy,k} = \left\langle \boldsymbol{y}_k\boldsymbol{y}_k^T\right\rangle_{q^{(j)}\left(\overline{\boldsymbol{X}}\right)}, \quad \mathbf{S}_{yy} = \sum_{k=1}^{N}\mathbf{S}^{(j)}_{yy,k}. \tag{5.30b}
$$

The procedure to calculate the expectations of sufficient statistics in Eq.(5.29) and (5.30) are described in the next section.

5.4.2 Derivation of the VBE-step: Evaluating $q^{(j)}\left(\overline{\boldsymbol{X}}\right)$

The goal of the VBE step is to evaluate the variational distribution $q^{(j)}\left(\overline{\boldsymbol{X}}\right)$ given the measurements $\boldsymbol{Y}$ and the hyperparameters $\boldsymbol{\theta}_h^{(j)} = \left\{\boldsymbol{\mu}_{\bar{\boldsymbol{a}}}^{(j)}, \boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{(j)}, \boldsymbol{\mu}_{\bar{\boldsymbol{g}}}^{(j)}, \boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}^{(j)}, d_{\overline{\mathbf{Q}}}^{(j)}, \mathbf{D}_{\overline{\mathbf{Q}}}^{(j)}, d_{\mathbf{R}}^{(j)}, \mathbf{D}_{\mathbf{R}}^{(j)}\right\}$ at

the jth VB iteration. The expectations over the states are then used to compute the expectations of sufficient statistics, as given in Eqs.(5.29) and (5.30). This is achieved by first defining a modified MSM and then employing Kalman filter and RTS smoother on the modified MSM for state estimation, similar to what has already been discussed in Section 4.5.2 of the previous chapter. A modified MSM is defined with the following parameters:

$$\widetilde{\overline{\mathbf{A}}} = \boldsymbol{\mu}_{\overline{\mathbf{A}}}^{(j)}, \quad \widetilde{\overline{\mathbf{Q}}} = \frac{\mathbf{D}_{\overline{\mathbf{Q}}}^{(j)}}{d_{\overline{\mathbf{Q}}}^{(j)} - n_s - 1} \tag{5.31a}$$

$$\overline{\overline{\mathbf{G}}} = \boldsymbol{\mu}_{\overline{\mathbf{G}}}^{(j)}, \quad \overline{\mathbf{R}} = \frac{\mathbf{D}_{\mathbf{R}}^{(j)}}{d_{\mathbf{R}}^{(j)} - n_o - 1} \tag{5.31b}$$

$$\overline{\mathbf{V}}^{\bar{\boldsymbol{a}}} = \left\langle \left(\overline{\mathbf{A}} - \widetilde{\overline{\mathbf{A}}}\right)^T \widetilde{\overline{\mathbf{Q}}}^{-1} \left(\overline{\mathbf{A}} - \widetilde{\overline{\mathbf{A}}}\right) \right\rangle_{q^{(j)}(\bar{\boldsymbol{a}})} \tag{5.31c}$$

$$\overline{\mathbf{V}}_k^{\bar{\boldsymbol{g}}} = \left\langle \left(\overline{\mathbf{G}} - \overline{\overline{\mathbf{G}}}\right)^T \boldsymbol{\Omega}_k^T \overline{\mathbf{R}}^{-1} \boldsymbol{\Omega}_k \left(\overline{\mathbf{G}} - \overline{\overline{\mathbf{G}}}\right) \right\rangle_{q^{(j)}(\bar{\boldsymbol{g}})} \tag{5.31d}$$

$$\mathbf{U}_k = \texttt{chol}(\overline{\mathbf{V}}^{\bar{\boldsymbol{a}}} + \overline{\mathbf{V}}_k^{\bar{\boldsymbol{g}}}) \tag{5.31e}$$

$$\widetilde{\boldsymbol{y}}_k = \begin{bmatrix} \boldsymbol{y}_k \\ \mathbf{0}_{n_s} \end{bmatrix}, \qquad \widetilde{\overline{\mathbf{G}}}_k = \begin{bmatrix} \boldsymbol{\Omega}_k \overline{\overline{\mathbf{G}}} \\ \mathbf{U}_k \end{bmatrix}, \qquad \widetilde{\overline{\mathbf{R}}} = \begin{bmatrix} \overline{\mathbf{R}} & \mathbf{0}_{n_o \times n_s} \\ \mathbf{0}_{n_o \times n_s}^T & \mathbf{I}_{n_s \times n_s} \end{bmatrix}. \tag{5.31f}$$

In Eq.(5.31e), '`chol`' stands for upper triangular Cholesky decomposition. Further, the computations of $\overline{\mathbf{V}}^{\bar{\boldsymbol{a}}}$ and $\overline{\mathbf{V}}^{\bar{\boldsymbol{g}}}$ in Eqs.(5.31c) and (5.31d) are provided in Appendix E.2. The Kalman filter and the RTS smoother can be then run with the following recursive equations:

Kalman filter for modified MSM : Do for $k = 1, \ldots, N$

$$\widetilde{\boldsymbol{e}}_k = \widetilde{\boldsymbol{y}}_k - \widetilde{\overline{\mathbf{G}}}_k \hat{\bar{\boldsymbol{x}}}_{k|k-1} \tag{5.32a}$$

$$\widetilde{\mathbf{H}}_k = \widetilde{\overline{\mathbf{G}}}_k \hat{\overline{\mathbf{V}}}_{k|k-1} \widetilde{\overline{\mathbf{G}}}_k^T + \widetilde{\overline{\mathbf{R}}} \tag{5.32b}$$

$$\widetilde{\mathbf{K}}_k = \hat{\overline{\mathbf{V}}}_{k|k-1} \widetilde{\overline{\mathbf{G}}}_k^T \widetilde{\mathbf{H}}_k^{-1} \tag{5.32c}$$

$$\hat{\bar{\boldsymbol{x}}}_{k|k} = \hat{\bar{\boldsymbol{x}}}_{k|k-1} + \widetilde{\mathbf{K}}_k \widetilde{\boldsymbol{e}}_k \tag{5.32d}$$

$$\hat{\overline{\mathbf{V}}}_{k|k} = \hat{\overline{\mathbf{V}}}_{k|k-1} - \widetilde{\mathbf{K}}_k \widetilde{\mathbf{H}}_k \widetilde{\mathbf{K}}_k^T \tag{5.32e}$$

$$\hat{\bar{\boldsymbol{x}}}_{k+1|k} = \widetilde{\overline{\mathbf{A}}} \hat{\bar{\boldsymbol{x}}}_{k|k} \tag{5.32f}$$

$$\hat{\overline{\mathbf{V}}}_{k+1|k} = \widetilde{\overline{\mathbf{A}}}\hat{\overline{\mathbf{V}}}_{k|k}\widetilde{\overline{\mathbf{A}}} + \widetilde{\overline{\mathbf{Q}}}. \tag{5.32g}$$

RTS smoother for MSM : Do for $k = N, \ldots, 1$

$$\widetilde{\mathbf{N}}_k = \hat{\overline{\mathbf{V}}}_{k|k}\widetilde{\overline{\mathbf{A}}}^T \left(\hat{\overline{\mathbf{V}}}_{k+1|k}\right)^{-1} \tag{5.32h}$$

$$\hat{\bar{\boldsymbol{x}}}_{k|N} = \hat{\bar{\boldsymbol{x}}}_{k|k} + \widetilde{\mathbf{N}}_k \left(\hat{\bar{\boldsymbol{x}}}_{k+1|N} - \hat{\bar{\boldsymbol{x}}}_{k+1|k}\right) \tag{5.32i}$$

$$\hat{\overline{\mathbf{V}}}_{k|N} = \hat{\overline{\mathbf{V}}}_{k|k} + \widetilde{\mathbf{N}}_k \left(\hat{\overline{\mathbf{V}}}_{k+1|N} - \hat{\overline{\mathbf{V}}}_{k+1|k}\right) \widetilde{\mathbf{N}}_k^T \tag{5.32j}$$

$$\hat{\overline{\mathbf{V}}}_{k+1,k|N} = \hat{\overline{\mathbf{V}}}_{k+1|N}\widetilde{\mathbf{N}}_k^T. \tag{5.32k}$$

The $\overline{\boldsymbol{\mu}}_1^{(j+1)}$ and $\overline{\mathbf{V}}_1^{(j+1)}$ for the $(j+1)$th VB iteration are initialized as

$$\overline{\boldsymbol{\mu}}_1^{(j+1)} = \hat{\bar{\boldsymbol{x}}}_{1|N}, \quad \overline{\mathbf{V}}_1^{(j+1)} = \hat{\overline{\mathbf{V}}}_{1|N}. \tag{5.33}$$

Using the outputs from the RTS smoother, the expectations of sufficient statistics defined in Eqs.(5.29) and (5.30) can be computed

$$\mathbf{S}_{\bar{x}\bar{x},k}^{(j)} = \left\langle \bar{\boldsymbol{x}}_k \bar{\boldsymbol{x}}_k^T \right\rangle_{q^{(j)}(\overline{\boldsymbol{X}})} = \hat{\bar{\boldsymbol{x}}}_k \hat{\bar{\boldsymbol{x}}}_k^T + \hat{\overline{\mathbf{V}}}_{k|N} \tag{5.34a}$$

$$\mathbf{S}_{\bar{f}\bar{x},k}^{(j)} = \left\langle \bar{\boldsymbol{x}}_{k+1} \bar{\boldsymbol{x}}_k^T \right\rangle_{q^{(j)}(\overline{\boldsymbol{X}})} = \hat{\bar{\boldsymbol{x}}}_{k+1} \hat{\bar{\boldsymbol{x}}}_k^T + \hat{\overline{\mathbf{V}}}_{k+1,k|N} \tag{5.34b}$$

$$\mathbf{S}_{\bar{f}\bar{f},k}^{(j)} = \left\langle \bar{\boldsymbol{x}}_{k+1} \bar{\boldsymbol{x}}_{k+1}^T \right\rangle_{q^{(j)}(\overline{\boldsymbol{X}})} = \hat{\bar{\boldsymbol{x}}}_{k+1} \hat{\bar{\boldsymbol{x}}}_{k+1}^T + \hat{\overline{\mathbf{V}}}_{k+1|N} \tag{5.34c}$$

$$\mathbf{S}_{y\bar{x},k}^{(j)} = \left\langle \boldsymbol{y}_k \bar{\boldsymbol{x}}_k^T \right\rangle_{q^{(j)}(\overline{\boldsymbol{X}})} = \boldsymbol{y}_k \hat{\bar{\boldsymbol{x}}}_k^T \tag{5.34d}$$

$$\mathbf{S}_{yy,k} = \left\langle \boldsymbol{y}_k \boldsymbol{y}_k^T \right\rangle_{q^{(j)}(\overline{\boldsymbol{X}})} = \boldsymbol{y}_k \boldsymbol{y}_k^T. \tag{5.34e}$$

5.4.3 Derivation of VBM-step: Computing $q^{(j+1)}(\boldsymbol{\theta})$

This section presents the update rules for the hyperparameters of the variational distributions $q^{(j+1)}(\bar{\boldsymbol{a}})$, $q^{(j+1)}(\overline{\mathbf{Q}})$, $q^{(j+1)}(\bar{\boldsymbol{g}})$ and $q^{(j+1)}(\mathbf{R})$, given the set of prior hyperparameters $\boldsymbol{\pi}_h = \left\{\boldsymbol{\mu}_{\bar{a}}, \boldsymbol{\Sigma}_{\bar{a}}, \boldsymbol{\mu}_{\bar{g}}, \boldsymbol{\Sigma}_{\bar{g}}, d_{\overline{\mathbf{Q}}}, \mathbf{D}_{\overline{\mathbf{Q}}}, d_{\mathbf{R}}, \mathbf{D}_{\mathbf{R}}\right\}$ (refer Eq.(5.23)) and the expectations of sufficient statistics for the jth iteration (refer Eq.(5.34)). The details of the derivations of the update rules can be found in Appendix E.3.

Update for $q^{(j+1)}(\bar{\boldsymbol{a}})$

The hyperparameters of $q^{(j+1)}(\bar{\boldsymbol{a}})$ are updated using the following:

$$\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{(j+1)} = \left(\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{-1} + \mathbf{S}_{\bar{x}\bar{x}}^{(j)} \otimes \boldsymbol{\mu}_{\overline{\mathbf{Q}}^{-1}}^{(j)}\right)^{-1} \tag{5.35a}$$

$$\boldsymbol{\mu}_{\bar{\boldsymbol{a}}}^{(j+1)} = \boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{(j+1)} \left(\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{-1}\boldsymbol{\mu}_{\bar{\boldsymbol{a}}} + \mathrm{vec}\left(\boldsymbol{\mu}_{\overline{\mathbf{Q}}^{-1}}^{(j)}\mathbf{S}_{\bar{f}\bar{x}}^{(j)}\right)\right) \tag{5.35b}$$

$$\boldsymbol{\mu}_{\overline{\mathbf{A}}}^{(j+1)} = \mathrm{mat}\left(\boldsymbol{\mu}_{\bar{\boldsymbol{a}}}^{(j+1)}\right) \tag{5.35c}$$

where $\boldsymbol{\mu}_{\overline{\mathbf{Q}}^{-1}}^{(j)} = d_{\overline{\mathbf{Q}}}^{(j)}\mathbf{D}_{\overline{\mathbf{Q}}}^{(j)^{-1}}$, and $\mathrm{mat}(\cdot)$ defines a reshape operator that reshapes vectors to matrices of relevant dimensions. In Eq.(5.35c), it reshapes a vector of dimension $n_s n_s$ to a matrix of dimension $n_s \times n_s$.

Update for $q^{(j+1)}(\overline{\mathbf{Q}})$

The hyperparameters update equations for $q^{(j+1)}(\overline{\mathbf{Q}})$ are:

$$\mathbf{M}_{\overline{\mathbf{A}}}^{(j+1)} = \boldsymbol{\mu}_{\overline{\mathbf{A}}}^{(j+1)}\mathbf{S}_{\bar{x}\bar{x}}^{(j)}\boldsymbol{\mu}_{\overline{\mathbf{A}}}^{(j+1)^T} + \sum_{p=1}^{n_s}\sum_{r=1}^{n_s}\left(\mathbf{S}_{\bar{x}\bar{x}}^{(j)}\right)_{pr}\boldsymbol{\Pi}_{pr}^{\bar{\boldsymbol{a}}} \tag{5.36a}$$

$$\mathbf{D}_{\overline{\mathbf{Q}}}^{(j+1)} = \mathbf{D}_{\overline{\mathbf{Q}}} + \mathbf{S}_{\bar{f}\bar{f}}^{(j)} - \boldsymbol{\mu}_{\overline{\mathbf{A}}}^{(j+1)}\mathbf{S}_{\bar{f}\bar{x}}^{(j)^T} - \mathbf{S}_{\bar{f}\bar{x}}^{(j)}\boldsymbol{\mu}_{\overline{\mathbf{A}}}^{(j+1)^T} + \mathbf{M}_{\overline{\mathbf{A}}}^{(j+1)} \tag{5.36b}$$

$$d_{\overline{\mathbf{Q}}}^{(j+1)} = d_{\overline{\mathbf{Q}}} + N \tag{5.36c}$$

where $\left(\mathbf{S}_{\bar{x}\bar{x}}^{(j)}\right)_{pr}$ is the (p,r)th element of matrix $\mathbf{S}_{\bar{x}\bar{x}}^{(j)}$, and $\boldsymbol{\Pi}_{pr}^{\bar{\boldsymbol{a}}} \in \mathbb{R}^{n_s \times n_s}$ is the (p,r)th block matrix of $\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{(j+1)}$. The block matrix representation of $\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{(j+1)}$ is shown below

$$\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{(j+1)} = \begin{bmatrix} \boldsymbol{\Pi}_{11}^{\bar{\boldsymbol{a}}} & \boldsymbol{\Pi}_{12}^{\bar{\boldsymbol{a}}} & \dots & \boldsymbol{\Pi}_{1n_s}^{\bar{\boldsymbol{a}}} \\ \boldsymbol{\Pi}_{21}^{\bar{\boldsymbol{a}}} & \boldsymbol{\Pi}_{22}^{\bar{\boldsymbol{a}}} & \dots & \boldsymbol{\Pi}_{2n_s}^{\bar{\boldsymbol{a}}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Pi}_{n_s1}^{\bar{\boldsymbol{a}}} & \boldsymbol{\Pi}_{n_s2}^{\bar{\boldsymbol{a}}} & \dots & \boldsymbol{\Pi}_{n_sn_s}^{\bar{\boldsymbol{a}}} \end{bmatrix}. \tag{5.37}$$

Update for $q^{(j+1)}(\bar{\boldsymbol{g}})$

The hyperparameters of $q^{(j+1)}(\bar{\boldsymbol{g}})$ are updated as:

$$\boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}^{(j+1)} = \left(\boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}^{-1} + \sum_{k=1}^{N}\left(\mathbf{S}_{\bar{x}\bar{x},k}^{(j)} \otimes \left(\boldsymbol{\Omega}_k^T \boldsymbol{\mu}_{\mathbf{R}^{-1}}^{(j)} \boldsymbol{\Omega}_k\right)\right)\right)^{-1} \tag{5.38a}$$

$$\boldsymbol{\mu}_{\bar{\boldsymbol{g}}}^{(j+1)} = \boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}^{(j+1)}\left(\boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}^{-1}\boldsymbol{\mu}_{\bar{\boldsymbol{g}}} + \sum_{k=1}^{N}\mathrm{vec}\left(\boldsymbol{\Omega}_k^T \boldsymbol{\mu}_{\mathbf{R}^{-1}}^{(j)} \mathbf{S}_{y\bar{x},k}^{(j)}\right)\right) \tag{5.38b}$$

$$\boldsymbol{\mu}_{\overline{\mathbf{G}}}^{(j+1)} = \mathrm{mat}\left(\boldsymbol{\mu}_{\bar{\boldsymbol{g}}}^{(j+1)}\right) \tag{5.38c}$$

where $\boldsymbol{\mu}_{\mathbf{R}^{-1}}^{(j)} = d_{\mathbf{R}}^{(j)}\mathbf{D}_{\mathbf{R}}^{(j)^{-1}}$ and the $\mathrm{mat}(\cdot)$ operator in Eq.(5.38c) reshapes a vector of dimension $n_s n_\alpha$ to a matrix of dimension $n_\alpha \times n_s$.

Update for $q^{(j+1)}(\mathbf{R})$

The hyperparameters of $q^{(j+1)}(\mathbf{R})$ are updated using the following rules:

$$\mathbf{M}_{\overline{\mathbf{G}},k}^{(j+1)} = \sum_{p=1}^{n_s}\sum_{r=1}^{n_s}\left(\mathbf{S}_{\bar{x}\bar{x},k}^{(j)}\right)_{pr}\boldsymbol{\Pi}_{pr}^{\bar{\boldsymbol{g}}} + \boldsymbol{\mu}_{\overline{\mathbf{G}}}^{(j+1)}\mathbf{S}_{\bar{x}\bar{x},k}^{(j)}\boldsymbol{\mu}_{\overline{\mathbf{G}}}^{(j+1)^T} \tag{5.39a}$$

$$\mathbf{D}_{\mathbf{R}}^{(j+1)} = \mathbf{D}_{\mathbf{R}} + \sum_{k=1}^{N}\left(\mathbf{S}_{yy,k}^{(j)} - \boldsymbol{\Omega}_k\boldsymbol{\mu}_{\overline{\mathbf{G}}}^{(j+1)}\mathbf{S}_{y\bar{x},k}^{(j)^T} - \mathbf{S}_{y\bar{x},k}^{(j)}\boldsymbol{\mu}_{\overline{\mathbf{G}}}^{(j+1)^T}\boldsymbol{\Omega}_k^T + \boldsymbol{\Omega}_k\mathbf{M}_{\overline{\mathbf{G}},k}^{(j+1)}\boldsymbol{\Omega}_k^T\right) \tag{5.39b}$$

$$d_{\mathbf{R}}^{(j+1)} = d_{\mathbf{R}} + N \tag{5.39c}$$

where $\left(\mathbf{S}_{\bar{x}\bar{x},k}^{(j)}\right)_{pr}$ is the (p,r)th element of matrix $\mathbf{S}_{\bar{x}\bar{x},k}^{(j)}$, and $\boldsymbol{\Pi}_{pr}^{\bar{\boldsymbol{g}}} \in \mathbb{R}^{n_\alpha \times n_\alpha}$ is the (p,r)th block matrix of $\boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}^{(j+1)}$. $\boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}^{(j+1)}$ can be represented using block matrices as shown below

$$\boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}^{(j+1)} = \begin{bmatrix} \boldsymbol{\Pi}_{11}^{\bar{\boldsymbol{g}}} & \boldsymbol{\Pi}_{12}^{\bar{\boldsymbol{g}}} & \dots & \boldsymbol{\Pi}_{1n_s}^{\bar{\boldsymbol{g}}} \\ \boldsymbol{\Pi}_{21}^{\bar{\boldsymbol{g}}} & \boldsymbol{\Pi}_{22}^{\bar{\boldsymbol{g}}} & \dots & \boldsymbol{\Pi}_{2n_s}^{\bar{\boldsymbol{g}}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Pi}_{n_s1}^{\bar{\boldsymbol{g}}} & \boldsymbol{\Pi}_{n_s2}^{\bar{\boldsymbol{g}}} & \dots & \boldsymbol{\Pi}_{n_sn_s}^{\bar{\boldsymbol{g}}}. \end{bmatrix} \tag{5.40}$$

5.4.4 Initialization and convergence criterion for VB

The hyperparameters of the prior distributions $\boldsymbol{\pi}_h = \left\{\boldsymbol{\mu}_{\bar{\boldsymbol{a}}}, \boldsymbol{\mu}_{\bar{\boldsymbol{g}}}, \boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}, \boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}, d_{\overline{\mathbf{Q}}}, d_{\mathbf{R}}, \mathbf{D}_{\overline{\mathbf{Q}}}, \mathbf{D}_{\mathbf{R}}\right\}$ are set as

$$\boldsymbol{\mu}_{\bar{\boldsymbol{a}}} = \mathrm{vec}\left(\overline{\mathbf{A}}^{\mathrm{SSI}}\right), \qquad \boldsymbol{\mu}_{\bar{\boldsymbol{g}}} = \mathrm{vec}\left(\overline{\mathbf{G}}^{\mathrm{SSI}}\right) \tag{5.41a}$$

$$\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}} = 10^{15} \times \mathbf{I}_{n_s^2 \times n_s^2}, \qquad \boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}} = 10^{15} \times \mathbf{I}_{(n_s n_\alpha)\times(n_s n_\alpha)} \tag{5.41b}$$

$$\mathbf{D}_{\overline{\mathbf{Q}}} = \mathbf{I}_{n_s \times n_s}, \qquad \mathbf{D}_{\mathbf{R}} = \mathbf{I}_{n_o \times n_o} \tag{5.41c}$$

$$d_{\overline{\mathbf{Q}}} = n_s + 2, \qquad d_{\mathbf{R}} = n_o + 2. \tag{5.41d}$$

The means of prior distributions over $\bar{\boldsymbol{a}}$ and $\bar{\boldsymbol{g}}$ are set equal to their vectorized forms of the SSI estimates as shown in Eq.(5.41) and large values ($\sim 10^{15}$) of covariances are assigned to make the prior sufficiently flat (and non-informative). Furthermore, the hyperparameters of the initial variational distributions are set equal to the hyperparameters of the prior distributions, i.e. $\boldsymbol{\theta}_h^{(0)} = \boldsymbol{\pi}_h$.

Typically, the VB iterations are stopped based on the change in the variational free energy between successive VB iterations. However, the numerical evaluation of variational free energy involves a high dimensional integral which is cumbersome to compute in this case. It is to be noted that the calculation of the variational free energy is not necessary for running the VB algorithm. Instead of variational free energy, the conditional log-likelihood is used as a stopping criterion for VB which also increases monotonically with each VB iteration. This is easily computed using the sequence of innovations $\widetilde{\boldsymbol{e}}_{1:N}$ and innovation covariance matrices $\widetilde{\mathbf{H}}_{1:N}$, obtained as outputs in the Kalman filtering step. Denoting $\widetilde{\mathcal{L}}\left(\boldsymbol{\theta}^{(j)}\right)$ as the conditional log-likelihood at the jth VB iteration, the stopping criterion is described by

$$\frac{2\left(\widetilde{\mathcal{L}}\left(\boldsymbol{\theta}^{(j+1)}\right) - \widetilde{\mathcal{L}}\left(\boldsymbol{\theta}^{(j)}\right)\right)}{\widetilde{\mathcal{L}}\left(\boldsymbol{\theta}^{(j+1)}\right) + \widetilde{\mathcal{L}}\left(\boldsymbol{\theta}^{(j)}\right)} \leq \epsilon_{tol}. \tag{5.42}$$

Typical values of the tolerance threshold ϵ_{tol} are in the range of $10^{-4} \sim 10^{-6}$. The log-likelihood $\mathcal{L}(\boldsymbol{\theta})$ is computed as [104, 120]

$$\widetilde{\mathcal{L}}(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{k=1}^{N} \ln\left|\widetilde{\mathbf{H}}_k\right| - \frac{1}{2}\sum_{k=1}^{N} \widetilde{\boldsymbol{e}}_k^T \widetilde{\mathbf{H}}_k \widetilde{\boldsymbol{e}}_k \tag{5.43}$$

where $|\cdot|$ is the determinant operator, and $\widetilde{\boldsymbol{e}}_k$ and $\widetilde{\mathbf{H}}_k$ are obtained as outputs from the Kalman filtering step, shown in Eqs.(5.32a) and (5.32b) respectively. Once the stopping criterion is reached, the VB iterations are stopped and the hyperparameters $\boldsymbol{\theta}_h^* = \left\{\overline{\boldsymbol{\mu}}_{\overline{\boldsymbol{a}}}^*, \overline{\boldsymbol{\mu}}_{\overline{\boldsymbol{g}}}^*, \boldsymbol{\Sigma}_{\overline{\boldsymbol{a}}}^*, \boldsymbol{\Sigma}_{\overline{\boldsymbol{g}}}^*, \mathbf{D}_{\overline{\mathbf{Q}}}^*, \mathbf{D}_{\mathbf{R}}^*, d_{\overline{\mathbf{Q}}}^*, d_{\mathbf{R}}^*\right\}$ from the final iteration are treated as the converged hyperparameters governing the posterior distribution of the parameters $\boldsymbol{\theta} = \left\{\overline{\mathbf{A}}, \overline{\mathbf{G}}, \overline{\mathbf{Q}}, \mathbf{R}\right\}$.

5.4.5 Posterior distributions of modal parameters from VB

The posterior distribution of modal parameters are obtained following a procedure similar to that outlined in Section 4.5.6 where a first-order Taylor series expansion is used to compute closed-form posterior distributions of the modal parameters. The modal parameters are expressed approximately as linear equations, as demonstrated in Eq.(4.51). Thus, the ith identified modal parameter will follow a multivariate normal distribution:

$$\ln f_i\left(\overline{\mathbf{A}}\right) \sim \mathcal{N}\left(\mu_{\ln f_i},\ \sigma^2_{\ln f_i}\right) \tag{5.44a}$$

$$\ln \xi_i\left(\overline{\mathbf{A}}\right) \sim \mathcal{N}\left(\mu_{\ln \xi_i},\ \sigma^2_{\ln \xi_i}\right) \tag{5.44b}$$

$$\boldsymbol{\phi}_i^{un}\left(\overline{\mathbf{A}}, \overline{\mathbf{G}}\right) \sim \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{\phi}_i^{un}},\ \boldsymbol{\Sigma}_{\boldsymbol{\phi}_i^{un}}\right) \tag{5.44c}$$

where

$$\mu_{\ln f_i} = \ln f_i\left(\boldsymbol{\mu}_{\overline{\mathbf{A}}}^*\right) \tag{5.45a}$$

$$\mu_{\ln \xi_i} = \ln \xi_i\left(\boldsymbol{\mu}_{\overline{\mathbf{A}}}^*\right) \tag{5.45b}$$

$$\boldsymbol{\mu}_{\boldsymbol{\phi}_i^{un}} = \boldsymbol{\phi}_i^{un}\left(\boldsymbol{\mu}_{\overline{\mathbf{A}}}^*, \boldsymbol{\mu}_{\overline{\mathbf{G}}}^*\right) \tag{5.45c}$$

$$\sigma^2_{\ln f_i} = \mathbf{J}_{\ln f_i, \overline{\mathbf{A}}}\ \boldsymbol{\Sigma}_{\overline{\boldsymbol{a}}}^*\ \mathbf{J}_{\ln f_i, \overline{\mathbf{A}}}^T \tag{5.45d}$$

$$\sigma^2_{\ln \xi_i} = \mathbf{J}_{\ln \xi_i, \overline{\mathbf{A}}}\ \boldsymbol{\Sigma}_{\overline{\boldsymbol{a}}}^*\ \mathbf{J}_{\ln \xi_i, \overline{\mathbf{A}}}^T \tag{5.45e}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\phi}_i^{un}} = \mathbf{J}_{\boldsymbol{\phi}_i^{un}, \overline{\mathbf{A}}}\ \boldsymbol{\Sigma}_{\overline{\boldsymbol{a}}}^*\ \mathbf{J}_{\boldsymbol{\phi}_i^{un}, \overline{\mathbf{A}}}^T + \mathbf{J}_{\boldsymbol{\phi}_i^{un}, \overline{\mathbf{G}}}\ \boldsymbol{\Sigma}_{\overline{\boldsymbol{g}}}^*\ \mathbf{J}_{\boldsymbol{\phi}_i^{un}, \overline{\mathbf{G}}}^T \tag{5.45f}$$

$$\mathbf{J}_{\ln f_i, \overline{\mathbf{A}}} = \left.\frac{\partial \ln f_i\left(\overline{\mathbf{A}}\right)}{\partial \mathrm{vec}\left(\overline{\mathbf{A}}\right)}\right|_{\overline{\mathbf{A}} = \boldsymbol{\mu}_{\overline{\mathbf{A}}}^*}, \quad \mathbf{J}_{\ln \xi_i, \overline{\mathbf{A}}} = \left.\frac{\partial \ln f_i\left(\overline{\mathbf{A}}\right)}{\partial \mathrm{vec}\left(\overline{\mathbf{A}}\right)}\right|_{\overline{\mathbf{A}} = \boldsymbol{\mu}_{\overline{\mathbf{A}}}^*} \tag{5.45g}$$

$$\mathbf{J}_{\boldsymbol{\phi}_i^{un},\overline{\mathbf{A}}} = \left.\frac{\partial \boldsymbol{\phi}_i^{un}\left(\overline{\mathbf{A}},\overline{\mathbf{G}}\right)}{\partial \mathrm{vec}\left(\overline{\mathbf{A}}\right)}\right|_{\overline{\mathbf{A}}=\boldsymbol{\mu}^*_{\overline{\mathbf{A}}}}, \quad \mathbf{J}_{\boldsymbol{\phi}_i^{un},\overline{\mathbf{G}}} = \left.\frac{\partial \boldsymbol{\phi}_i^{un}\left(\overline{\mathbf{A}},\overline{\mathbf{G}}\right)}{\partial \mathrm{vec}\left(\overline{\mathbf{G}}\right)}\right|_{\overline{\mathbf{G}}=\boldsymbol{\mu}^*_{\overline{\mathbf{G}}}}. \tag{5.45h}$$

It is noteworthy to mention that the mode shapes follow multivariate normal distributions whereas the logarithms of modal frequencies and damping ratios follow normal distributions. To convert to the lognormal distributions of the modal frequencies and damping ratios to equivalent normal distribution, the transformation formulas as given in Eq.(4.55) are used.

The complete list of steps for applying the VB algorithm to the stacked data case is provided in Algorithm 5.

Algorithm 5 Modal parameter estimation with VB using stacked data

Input: $\boldsymbol{Y}$, ϵ_{tol}, VSLs $\boldsymbol{s}^{\alpha}$ and sensing nodes $\{\boldsymbol{s}_k^o\}_{1:N}$

Build the sequence of mode shape regression matrices $\{\boldsymbol{\Omega}\}_{1:N}$ using Eq.(5.5)

Set prior hyperparameter $\boldsymbol{\pi}_h$ using Eq.(5.41), and initialize $\boldsymbol{\theta}_h^{(0)} = \boldsymbol{\pi}_h$

VB E-step:

1. Set $\overline{\mathbf{A}}^{(0)}$, $\overline{\mathbf{G}}^{(0)}$, $\overline{\mathbf{Q}}^{(0)}$ and $\mathbf{R}^{(0)}$ to the mean values of their respective distributions
2. Use Eq.(5.14) and (5.15) for Kalman filter and RTS smoother recursions
3. Calculate the expectations of sufficient statistics using Eqs.(5.29) and (5.30)

while not converged **do**

1. **VB M**-step: Use Eqs.(5.35), (5.38), (5.36), (5.39) and (5.33) to update hyperparameters $\rightarrow \boldsymbol{\mu}_{\overline{\mathbf{A}}}^{(j+1)}, \boldsymbol{\mu}_{\overline{\mathbf{G}}}^{(j+1)}, \boldsymbol{\Sigma}_{\overline{\boldsymbol{a}}}^{(j+1)}, \boldsymbol{\Sigma}_{\overline{\boldsymbol{g}}}^{(j+1)}, \mathbf{D}_{\overline{\mathbf{Q}}}^{(j+1)}, \mathbf{D}_{\mathbf{R}}^{(j+1)}, d_{\overline{\mathbf{Q}}}^{(j+1)}, d_{\mathbf{R}}^{(j+1)}, \overline{\boldsymbol{\mu}}_1^{(j+1)}, \overline{\mathbf{V}}_1^{(j+1)}$
2. **VB E**-step:
   (a) Run Kalman filter and RTS smoother using Eqs.(5.31)– (5.32)
   (b) Calculate the expectations of sufficient statistics using Eqs.(5.29) and (5.30)
3. Compute variational free energy $\mathcal{F}_{ve}^{(j+1)}$
4. Use Eq.(4.50) to check if convergence criterion is satisfied

end while

Return $\boldsymbol{\mu}_{\overline{\mathbf{A}}}^{*} \leftarrow \boldsymbol{\mu}_{\overline{\mathbf{A}}}^{(final)}$, $\boldsymbol{\mu}_{\overline{\mathbf{G}}}^{*} \leftarrow \boldsymbol{\mu}_{\overline{\mathbf{G}}}^{(final)}$, $\mathbf{D}_{\overline{\mathbf{Q}}}^{*} \leftarrow \mathbf{D}_{\overline{\mathbf{Q}}}^{(final)}$, $\mathbf{D}_{\mathbf{R}}^{*} \leftarrow \mathbf{D}_{\mathbf{R}}^{(final)}$, $\boldsymbol{\Sigma}_{\overline{\boldsymbol{a}}}^{*} \leftarrow \boldsymbol{\Sigma}_{\overline{\boldsymbol{a}}}^{(final)}$, $\boldsymbol{\Sigma}_{\overline{\boldsymbol{g}}}^{*} \leftarrow \boldsymbol{\Sigma}_{\overline{\boldsymbol{g}}}^{(final)}$, $d_{\overline{\mathbf{Q}}}^{*} \leftarrow d_{\overline{\mathbf{Q}}}^{(final)}$, $d_{\mathbf{R}}^{*} \leftarrow d_{\mathbf{R}}^{(final)}$

Set $\overline{\mathbf{A}} = \boldsymbol{\mu}_{\overline{\mathbf{A}}}^{*}$ and $\overline{\mathbf{G}} = \boldsymbol{\mu}_{\overline{\mathbf{G}}}^{*}$. Then extract the modal parameters i.e. frequency f_i, damping ratio ξ_i, un-normalized mode shape $\boldsymbol{\phi}_i^{un}$, using Appendix B.1

for $f_i, \xi_i, \boldsymbol{\phi}_i^{un}$ **do**

Compute the sensitivities using Eq.(B.6) and Appendix B.3

Compute the means and covariances of the posterior distribution of the identified modal parameters using Eqs.(5.44) and Eq.(4.55)

end for

Output: $\mu_{f_i}, \mu_{\xi_i}, \boldsymbol{\mu}_{\boldsymbol{\phi}_i^{un}}, \sigma_{f_i}^2, \sigma_{\xi_i}^2, \boldsymbol{\Sigma}_{\boldsymbol{\phi}_i^{un}}$ for all modes $i = 1, 2, \ldots$

5.5 Bayesian inference of MSM via GS

In this section, the GS algorithm is derived for the MSM. First, the joint distribution of the unknown parameters $\boldsymbol{\theta} = \left\{\overline{\mathbf{A}}, \overline{\mathbf{G}}, \overline{\mathbf{Q}}, \mathbf{R}\right\}$, the hidden states $\overline{\boldsymbol{X}}$ and the measurements $\boldsymbol{Y}$ can be written as:

$$p\left(\boldsymbol{Y}, \overline{\boldsymbol{X}}, \overline{\mathbf{A}}, \overline{\mathbf{G}}, \overline{\mathbf{Q}}, \mathbf{R}\right) = p\left(\overline{\mathbf{A}}\right) p\left(\overline{\mathbf{Q}}\right) p\left(\overline{\mathbf{G}}\right) p\left(\mathbf{R}\right) p\left(\bar{\boldsymbol{x}}_1\right) \tag{5.46}$$

$$\times \prod_{k=1}^{N} p\left(\bar{\boldsymbol{x}}_{k+1} \mid \bar{\boldsymbol{x}}_k, \overline{\mathbf{A}}, \overline{\mathbf{Q}}\right) p\left(\boldsymbol{y}_k \mid \bar{\boldsymbol{x}}_k, \overline{\mathbf{G}}, \mathbf{R}\right). \tag{5.47}$$

Given the observed measurements $\boldsymbol{Y}$, the required conditional distributions can be derived and samples can be drawn from the Markov chain in the following cyclic fashion

1. Given $\boldsymbol{\theta}^{(j)} = \left\{\overline{\mathbf{A}}^{(j)}, \overline{\mathbf{G}}^{(j)}, \overline{\mathbf{Q}}^{(j)}, \mathbf{R}^{(j)}\right\}$ and observed measurements $\boldsymbol{Y}$, a sequence of hidden state variables are sampled according to

$$\overline{\boldsymbol{X}}^{(j)} \sim p\left(\overline{\boldsymbol{X}} \mid \boldsymbol{Y}, \boldsymbol{\theta}^{(j)}\right), \tag{5.48}$$

2. Given $\overline{\boldsymbol{X}}^{(j)}$, the unknown parameters $\boldsymbol{\theta}^{(j+1)}$ are sampled according to

$$\boldsymbol{\theta}^{(j+1)} \sim p\left(\boldsymbol{\theta} \mid \boldsymbol{Y}, \overline{\boldsymbol{X}}^{(j)}\right). \tag{5.49a}$$

5.5.1 Forward filtering backward sampling for states

The state sequence $\overline{\boldsymbol{X}}$ is sampled following the *forward filtering backward sampling* strategy elaborated in Wills et al. [92]. The strategy consists of two parts: forward filtering and backward sampling. The forward filtering part sequentially predicts and updates the states and state-error covariances using the Kalman filter, obtaining $p\left(\bar{\boldsymbol{x}}_k \mid \boldsymbol{Y}, \boldsymbol{\theta}\right)$ for all $k = 1, \ldots, N$. Using the results from forward filtering, the backward sampling strategy simulates state vectors from time $k = N + 1$ to $k = 1$. The joint distribution of $\bar{\boldsymbol{x}}_{k+1}$ and $\bar{\boldsymbol{x}}_k$ conditioned on observed data $\boldsymbol{Y}$ [65] can be written as

$$p\left(\bar{\boldsymbol{x}}_{k+1}, \bar{\boldsymbol{x}}_k \mid \boldsymbol{Y}\right) = \mathcal{N}\left(\begin{bmatrix}\bar{\boldsymbol{x}}_{k+1} \\ \bar{\boldsymbol{x}}_k\end{bmatrix} \middle| \tilde{\boldsymbol{m}}_1, \tilde{\mathbf{P}}_1\right) \tag{5.50}$$

where

$$\begin{aligned}\tilde{\boldsymbol{m}}_1 &= \begin{bmatrix} \hat{\bar{\boldsymbol{x}}}_{k+1|N} \\ \hat{\bar{\boldsymbol{x}}}_{k|k} + \mathbf{N}_k \left(\hat{\bar{\boldsymbol{x}}}_{k+1|N} - \overline{\mathbf{A}} \hat{\bar{\boldsymbol{x}}}_{k|k} \right) \end{bmatrix} \\ \tilde{\mathbf{P}}_1 &= \begin{bmatrix} \hat{\bar{\mathbf{V}}}_{k+1|N} & \hat{\bar{\mathbf{V}}}_{k+1|N} \mathbf{N}_k^T \\ \mathbf{N}_k \hat{\bar{\mathbf{V}}}_{k+1|N}^T & \hat{\bar{\mathbf{V}}}_{k|k} + \mathbf{N}_k \left(\hat{\bar{\mathbf{V}}}_{k+1|N} - \hat{\bar{\mathbf{V}}}_{k+1|k} \right) \mathbf{N}_k^T \end{bmatrix} \end{aligned} \tag{5.51}$$

The matrix $\mathbf{N}_k = \hat{\bar{\mathbf{V}}}_{k|k} \overline{\mathbf{A}}^T \left(\hat{\bar{\mathbf{V}}}_{k+1|k} \right)^{-1}$ is the smoother gain matrix at the kth time instant. Using the property of the multivariate normal distribution, the conditional distribution of $\bar{\boldsymbol{x}}_k$ given $\bar{\boldsymbol{x}}_{k+1}$ follows a multivariate normal distribution (refer to Eq. (8.10) in [65])

$$p\left(\bar{\boldsymbol{x}}_k \mid \bar{\boldsymbol{x}}_{k+1}, \boldsymbol{Y} \right) = \mathcal{N}\left(\bar{\boldsymbol{x}}_k \mid \tilde{\boldsymbol{m}}_2, \tilde{\mathbf{P}}_2 \right) \tag{5.52}$$

where

$$\begin{aligned} \tilde{\boldsymbol{m}}_2 &= \hat{\bar{\boldsymbol{x}}}_{k|k} + \mathbf{N}_k \left(\bar{\boldsymbol{x}}_{k+1} - \overline{\mathbf{A}} \hat{\bar{\boldsymbol{x}}}_{k|k} \right) \\ \tilde{\mathbf{P}}_2 &= \hat{\bar{\mathbf{V}}}_{k+1|k} - \mathbf{N}_k \hat{\bar{\mathbf{V}}}_{k+1|k} \mathbf{N}_k^T. \end{aligned} \tag{5.53}$$

Thus, given $\overline{\mathbf{A}}^{(j)}, \overline{\mathbf{G}}^{(j)}, \overline{\mathbf{Q}}^{(j)}, \mathbf{R}^{(j)}$ and $\boldsymbol{Y}$, one can generate the sequence $\overline{\boldsymbol{X}}^{(j)}$.

5.5.2 Sampling unknown parameters

The posterior distribution of the parameters $\boldsymbol{\theta} = \left\{ \overline{\mathbf{A}}, \overline{\mathbf{G}}, \overline{\mathbf{Q}}, \mathbf{R} \right\}$ given a sample of the state sequence $\overline{\boldsymbol{X}}$ and the measurements $\boldsymbol{Y}$ is given by

$$p\left(\boldsymbol{\theta} \mid \overline{\boldsymbol{X}}, \boldsymbol{Y} \right) \propto p\left(\overline{\boldsymbol{X}}, \boldsymbol{Y} \mid \boldsymbol{\theta} \right) p\left(\boldsymbol{\theta} \right). \tag{5.54}$$

It turns out that even though the parameters were assumed independent (i.e. a fully factorized prior distribution was assumed over the parameters), they become dependent in their joint posterior distributions. Hence, first the joint posterior distribution is written out and then the fully conditional distributions of the parameters are derived to facilitate Gibbs sampling. Using the joint probability distribution of complete data (Eq.(5.20)), one

can write the joint posterior distribution over the parameters as:

$$
\begin{aligned}
p\Big(\overline{\mathbf{A}},\overline{\mathbf{Q}},\overline{\mathbf{G}},\mathbf{R} \mid \boldsymbol{Y},\overline{\boldsymbol{X}}^{(j)}\Big) &\propto p\left(\boldsymbol{Y},\overline{\boldsymbol{X}}^{(j)} \mid \overline{\mathbf{A}},\overline{\mathbf{Q}},\overline{\mathbf{G}},\mathbf{R}\right) p\left(\overline{\mathbf{A}},\overline{\mathbf{Q}},\overline{\mathbf{G}},\mathbf{R}\right) \\
&\propto \quad \frac{d_{\overline{\mathbf{Q}}}+N+n_s+1}{2}\ln\left|\overline{\mathbf{Q}}^{-1}\right| + \frac{d_{\mathbf{R}}+N+n_o+1}{2}\ln\left|\mathbf{R}^{-1}\right| \\
&\quad - \frac{1}{2}\operatorname{tr}\left\{\overline{\mathbf{Q}}^{-1}\left(\mathbf{P}^{(j)}_{\bar{f}\bar{f}} - \overline{\mathbf{A}}\left(\mathbf{P}^{(j)}_{\bar{f}\bar{x}}\right)^T - \mathbf{P}^{(j)}_{\bar{f}\bar{x}}\overline{\mathbf{A}}^T + \overline{\mathbf{A}}\mathbf{P}^{(j)}_{\bar{x}\bar{x}}\overline{\mathbf{A}}^T\right)\right\} \\
&\quad - \frac{1}{2}\operatorname{tr}\left\{\mathbf{R}^{-1}\sum_{k=1}^{N}\left(\mathbf{P}^{(j)}_{yy,k} - \boldsymbol{\Omega}_k\overline{\mathbf{G}}\left(\mathbf{P}^{(j)}_{y\bar{x},k}\right)^T - \mathbf{P}^{(j)}_{y\bar{x},k}\overline{\mathbf{G}}^T\boldsymbol{\Omega}_k^T + \boldsymbol{\Omega}_k\overline{\mathbf{G}}\mathbf{P}^{(j)}_{\bar{x}\bar{x},k}\overline{\mathbf{G}}^T\boldsymbol{\Omega}_k^T\right)\right\} \\
&\quad - \frac{1}{2}\operatorname{tr}\left\{\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{-1}\left(\boldsymbol{\mu}_{\bar{\boldsymbol{a}}}\boldsymbol{\mu}_{\bar{\boldsymbol{a}}}^T - \bar{\boldsymbol{a}}\boldsymbol{\mu}_{\bar{\boldsymbol{a}}}^T - \boldsymbol{\mu}_{\bar{\boldsymbol{a}}}\bar{\boldsymbol{a}}^T + \bar{\boldsymbol{a}}\bar{\boldsymbol{a}}^T\right)\right\} \\
&\quad - \frac{1}{2}\operatorname{tr}\left\{\boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}^{-1}\left(\boldsymbol{\mu}_{\bar{\boldsymbol{g}}}\boldsymbol{\mu}_{\bar{\boldsymbol{g}}}^T - \bar{\boldsymbol{g}}\boldsymbol{\mu}_{\bar{\boldsymbol{g}}}^T - \boldsymbol{\mu}_{\bar{\boldsymbol{g}}}\bar{\boldsymbol{g}}^T + \bar{\boldsymbol{g}}\bar{\boldsymbol{g}}^T\right)\right\} \\
&\quad - \frac{1}{2}\operatorname{tr}\left\{\overline{\mathbf{Q}}^{-1}\mathbf{D}_{\overline{\mathbf{Q}}}\right\} - \frac{1}{2}\operatorname{tr}\left\{\mathbf{R}^{-1}\mathbf{D}_{\mathbf{R}}\right\}
\end{aligned}
\tag{5.55}
$$

where $\mathbf{P}^{(j)}_{y\bar{x},k}, \mathbf{P}^{(j)}_{yy,k}, \mathbf{P}^{(j)}_{\bar{x}\bar{x},k}, \mathbf{P}^{(j)}_{\bar{x}\bar{x}}, \mathbf{P}^{(j)}_{\bar{f}\bar{x}}, \mathbf{P}^{(j)}_{\bar{f}\bar{f}}$ are the expectations of sufficient statistics at the jth sampling iteration of Markov chain described as

$$
\mathbf{P}^{(j)}_{y\bar{x},k} = \boldsymbol{y}_k\left(\bar{\boldsymbol{x}}^{(j)}_k\right)^T, \qquad \mathbf{P}_{yy,k} = \boldsymbol{y}_k\boldsymbol{y}_k^T \tag{5.56a}
$$

$$
\mathbf{P}^{(j)}_{\bar{x}\bar{x}} = \sum_{k=1}^{N}\mathbf{P}^{(j)}_{\bar{x}\bar{x},k} = \sum_{k=1}^{N}\bar{\boldsymbol{x}}^{(j)}_k\left(\bar{\boldsymbol{x}}^{(j)}_k\right)^T \tag{5.56b}
$$

$$
\mathbf{P}^{(j)}_{\bar{f}\bar{x}} = \sum_{k=1}^{N}\bar{\boldsymbol{x}}^{(j)}_{k+1}\left(\bar{\boldsymbol{x}}^{(j)}_k\right)^T, \qquad \mathbf{P}^{(j)}_{\bar{f}\bar{f}} = \sum_{k=1}^{N}\bar{\boldsymbol{x}}^{(j)}_{k+1}\left(\bar{\boldsymbol{x}}^{(j)}_{k+1}\right)^T. \tag{5.56c}
$$

Using the $\boldsymbol{\theta}^{(j)} = \left(\overline{\mathbf{A}}^{(j)}, \overline{\mathbf{G}}^{(j)}, \overline{\mathbf{Q}}^{(j)}, \mathbf{R}^{(j)}\right)$ and $\overline{\boldsymbol{X}}^{(j)}$ from the jth step of the Markov chain, the parameters $\boldsymbol{\theta}^{(j+1)}$ can be sampled from the following conditional posterior distributions:

- Sample $\left(\bar{\boldsymbol{a}}^{(j+1)} \mid \overline{\mathbf{Q}}^{(j)}, \overline{\boldsymbol{X}}^{(j)}\right) \sim \mathcal{N}\left(\widetilde{\boldsymbol{\mu}_{\bar{\boldsymbol{a}}}}^{(j+1)}, \widetilde{\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}}^{(j+1)}\right)$ and set $\overline{\mathbf{A}}^{(j+1)} = \operatorname{mat}\left(\bar{\boldsymbol{a}}^{(j+1)}\right)$, where

$$
\widetilde{\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}}^{(j+1)} = \left(\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{-1} + \mathbf{P}^{(j)}_{\bar{x}\bar{x}} \otimes \overline{\mathbf{Q}}^{(j)^{-1}}\right)^{-1} \tag{5.57a}
$$

$$
\widetilde{\boldsymbol{\mu}_{\bar{\boldsymbol{a}}}}^{(j+1)} = \widetilde{\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}}^{(j+1)}\left(\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{-1}\boldsymbol{\mu}_{\bar{\boldsymbol{a}}} + \operatorname{vec}\left(\overline{\mathbf{Q}}^{(j)^{-1}}\mathbf{P}^{(j)}_{\bar{f}\bar{x}}\right)\right). \tag{5.57b}
$$

- Sample $\left(\bar{\boldsymbol{g}}^{(j+1)} \mid \mathbf{R}^{(j)}, \overline{\boldsymbol{X}}^{(j)}\right) \sim \mathcal{N}\left(\widetilde{\boldsymbol{\mu}_{\bar{\boldsymbol{g}}}}^{(j+1)}, \widetilde{\boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}}^{(j+1)}\right)$ and set $\overline{\mathbf{G}}^{(j+1)} = \operatorname{mat}\left(\bar{\boldsymbol{g}}^{(j+1)}\right)$, where

$$\widetilde{\boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}}^{(j+1)} = \left(\boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}^{-1} + \sum_{k=1}^{N}\left(\mathbf{P}_{\bar{x}\bar{x},k}^{(j)} \otimes \left(\boldsymbol{\Omega}_k^T {\mathbf{R}^{(j)}}^{-1} \boldsymbol{\Omega}_k\right)\right)\right)^{-1} \tag{5.58a}$$

$$\widetilde{\boldsymbol{\mu}_{\bar{\boldsymbol{g}}}}^{(j+1)} = \widetilde{\boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}}^{(j+1)}\left(\boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}^{-1}\boldsymbol{\mu}_{\bar{\boldsymbol{g}}} + \operatorname{vec}\left({\mathbf{R}^{(j)}}^{-1}\mathbf{P}_{y\bar{x}}^{(j)}\right)\right). \tag{5.58b}$$

- Sample $\left(\overline{\mathbf{Q}}^{(j+1)} \mid \boldsymbol{X}^{(j)}, \overline{\mathbf{A}}^{(j+1)}\right) \sim \mathcal{IW}\left(\widetilde{d_{\overline{\mathbf{Q}}}}^{(j+1)}, \widetilde{\mathbf{D}_{\overline{\mathbf{Q}}}}^{(j+1)}\right)$, where

$$\widetilde{\mathbf{D}_{\overline{\mathbf{Q}}}}^{(j+1)} = \mathbf{D}_{\overline{\mathbf{Q}}} + \mathbf{P}_{\bar{f}\bar{f}}^{(j)} - \overline{\mathbf{A}}^{(j+1)}\left(\mathbf{P}_{\bar{f}\bar{x}}^{(j)}\right)^T - \mathbf{P}_{\bar{f}\bar{x}}^{(j)}\left(\overline{\mathbf{A}}^{(j+1)}\right)^T + \overline{\mathbf{A}}^{(j+1)}\mathbf{P}_{\bar{x}\bar{x}}^{(j)}\left(\overline{\mathbf{A}}^{(j+1)}\right)^T \tag{5.59a}$$

$$\widetilde{d_{\overline{\mathbf{Q}}}}^{(j+1)} = d_{\overline{\mathbf{Q}}} + N. \tag{5.59b}$$

- Sample $\left(\mathbf{R}^{(j+1)} \mid \boldsymbol{X}^{(j)}, \overline{\mathbf{G}}^{(j+1)}\right) \sim \mathcal{IW}\left(\widetilde{d_{\mathbf{R}}}^{(j+1)}, \widetilde{\mathbf{D}_{\mathbf{R}}}^{(j+1)}\right)$, where

$$\begin{aligned}\widetilde{\mathbf{D}_{\mathbf{R}}}^{(j+1)} = \mathbf{D}_{\mathbf{R}} &+ \sum_{k=1}^{N}\left(\mathbf{P}_{yy,k}^{(j)} - \boldsymbol{\Omega}_k\overline{\mathbf{G}}^{(j+1)}\left(\mathbf{P}_{y\bar{x},k}^{(j)}\right)^T - \mathbf{P}_{y\bar{x},k}^{(j)}\left(\overline{\mathbf{G}}^{(j+1)}\right)^T\boldsymbol{\Omega}_k^T\right) \\ &+ \sum_{k=1}^{N}\left(\boldsymbol{\Omega}_k\overline{\mathbf{G}}^{(j+1)}\mathbf{P}_{\bar{x}\bar{x},k}^{(j)}\left(\overline{\mathbf{G}}^{(j+1)}\right)^T\boldsymbol{\Omega}_k^T\right)\end{aligned} \tag{5.60a}$$

$$\widetilde{d_{\mathbf{R}}}^{(j+1)} = d_{\mathbf{R}} + N. \tag{5.60b}$$

Thus, given the jth sample of parameters $\boldsymbol{\theta}^{(j)}$ and state sequence $\overline{\boldsymbol{X}}^{(j)}$, the updated parameters $\boldsymbol{\theta}^{(j+1)}$ can be sampled from their corresponding conditional distributions with the updated hyperparameters expressed in Eqs.(5.57), (5.58), (5.59) and (5.60). For the purposes of implementation, first $\bar{\boldsymbol{a}}^{(j+1)}$ and $\bar{\boldsymbol{g}}^{(j+1)}$) are sampled from the respective multivariate Gaussian distributions using $\overline{\mathbf{Q}}^{(j)}$, $\mathbf{R}^{(j)}$ and $\boldsymbol{X}^{(j)}$, and reshaped into matrices $\overline{\mathbf{A}}^{(j+1)}$ and $\overline{\mathbf{G}}^{(j+1)}$. Next using $\overline{\mathbf{A}}^{(j+1)}, \overline{\mathbf{G}}^{(j+1)}$ and $\boldsymbol{X}^{(j)}$, the parameters $\overline{\mathbf{Q}}^{(j+1)}, \mathbf{R}^{(j+1)}$ are sampled from their respective inverse Wishart distributions.

5.5.3 Initialization of GS

In Gibbs sampling, the desired posterior distributions are obtained after the Markov chain reaches its stationary distribution. The number of samples needed for a Markov chain to reach a stationary distribution depends to a great extent on the amount of observed data. The lesser the amount of observed data, the greater the number of samples required to reach stationarity, thus longer becomes the Markov chain. The GS is started with $\boldsymbol{\theta}_h^{(0)} = \left\{\boldsymbol{\mu}_{\bar{\boldsymbol{a}}}^{(0)}, \boldsymbol{\mu}_{\bar{\boldsymbol{g}}}^{(0)}, \boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{(0)}, \boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}^{(0)}, \mathbf{D}_{\overline{\mathbf{Q}}}^{(0)}, \mathbf{D}_{\mathbf{R}}^{(0)}, d_{\overline{\mathbf{Q}}}^{(0)}, d_{\mathbf{R}}^{(0)}\right\}$ as the initial estimate along with specification of the prior hyperparameters $\boldsymbol{\pi}_h = \left\{\boldsymbol{\mu}_{\bar{\boldsymbol{a}}}, \boldsymbol{\mu}_{\bar{\boldsymbol{g}}}, \boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}, \boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}, d_{\overline{\mathbf{Q}}}, d_{\mathbf{R}}, \mathbf{D}_{\overline{\mathbf{Q}}}, \mathbf{D}_{\mathbf{R}}\right\}$. The initial estimates and the prior parameters are set in the same way as done for VB in Section 5.4.4.

5.5.4 Computation of posterior statistics

The first few samples of each Markov chain are discarded as burn-in samples to remove transient behavior in the initial phases of the Markov chain. Typically, the samples of a Markov chain are correlated; to minimize the correlation between samples one may choose to subsample the Markov chain. It must be noted that, after discarding the burn-in samples and subsampling the Markov chain, one extracts the modal parameters from the retained sample pairs $\left(\overline{\mathbf{A}}^{(j)}, \overline{\mathbf{G}}^{(j)}\right)$, $j = 1, \ldots, J$, obtaining J sets of modal parameters. For grouping the modes, a k-means clustering is adopted using three features: frequencies, damping ratios and normalized mode shapes. The clusters with physical modes appear more consistently and these are used for computing the modal distributions.

The complete list of steps for applying the VB algorithm to the stacked data case is provided in Algorithm 6.

Algorithm 6 Modal parameter estimation with GS using stacked data

Input: $\boldsymbol{Y}$, ϵ_{tol}, VSLs $\boldsymbol{s}^{\alpha}$ and sensing nodes $\{\boldsymbol{s}_k^o\}_{1:N}$, chain length, burn-in samples, subsampling period

Build the sequence of mode shape regression matrices $\{\boldsymbol{\Omega}\}_{1:N}$ using Eq.(5.5)

Set prior hyperparameter $\boldsymbol{\pi}_h$ using Eq.(5.41) and initialize $\boldsymbol{\theta}_h^{(0)} = \boldsymbol{\pi}_h$

for sample $j = 1 : J$ **do**

1. **Forward filtering backward sampling for states**
   (a) Use Eq.(5.14) for Kalman filter recursions
   (b) Calculate smoother gains $\{\mathbf{N}\}_{1:N}$ using Eq.(5.15a)
   (c) Generate sequence $\overline{\boldsymbol{X}}^{(j)}$ using Eqs.(5.52) and (5.53)
2. **Sampling parameters**
   (a) Compute expectations of sufficient statistics using Eq.(5.56)
   (b) Using $\overline{\mathbf{Q}}^{(j)}$ and expectations of sufficient statistics, update the distributional hyperparameters of $\overline{\mathbf{A}}$ following Eq.(5.57). Draw sample $\bar{\boldsymbol{a}}^{(j+1)}$ and reshape to $\overline{\mathbf{A}}^{(j+1)}$
   (c) Using $\overline{\mathbf{A}}^{(j+1)}$ and expectations of sufficient statistics, update the distributional hyperparameters of $\overline{\mathbf{Q}}$ using Eq.(5.59). Draw sample $\overline{\mathbf{Q}}^{(j+1)}$
   (d) Using $\mathbf{R}^{(j)}$ and expectations of sufficient statistics, update the distributional hyperparameters of $\bar{\boldsymbol{g}}$ using Eq.(5.58). Draw sample $\bar{\boldsymbol{g}}^{(j+1)}$ and reshape to $\overline{\mathbf{G}}^{(j+1)}$
   (e) Using $\overline{\mathbf{A}}^{(j+1)}$ and expectations of sufficient statistics, update the distributional hyperparameters of $\mathbf{R}$ using Eq.(5.60). Draw sample $\mathbf{R}^{(j+1)}$
   (f) Use Appendix B.1 to extract the modal parameters $f_i, \xi_i, \boldsymbol{\phi}_i^{un}$ using matrices $\overline{\mathbf{A}}^{(j+1)}$ and $\overline{\mathbf{G}}^{(j+1)}$

end for

for $f_i, \xi_i, \boldsymbol{\phi}_i^{un}$ **do**

Compute the mean and variance estimates of the samples of modal parameters (post subsampling and discarding the initial burn-in samples)

end for

Output: $\mu_{f_i}, \mu_{\xi_i}, \boldsymbol{\mu}_{\boldsymbol{\phi}_i^{un}}, \sigma_{f_i}^2, \sigma_{\xi_i}^2, \boldsymbol{\Sigma}_{\boldsymbol{\phi}_i^{un}}$ for all modes $i = 1, 2, \ldots$

5.6 Numerical Demonstration

This section presents the performance of the EM, VB and GS for modal parameter estimation on a simply supported beam following the stacked data approach to mobile sensing. The simply supported beam is assumed to have the following properties: length $L = 20$m, flexural rigidity $3.2 \times 10^8 \text{Nm}^2$, and mass per unit length 2000kg/m. The beam is modelled using 4000 finite (beam) elements of length equal to 0.005m. The natural frequencies for the first five modes of the beam are 1.57, 6.28, 14.14, 25.14 and 39.28 Hz. The damping ratios for the first five modes are set to 1% of the critical damping. The numerical response data are generated using the following:

- Gaussian white noise inputs with zero means and variances 10^4N^2 were applied at 99 equidistant points on the beam (each point separated by 0.2m). A frequency cutoff at 45 Hz was applied to the Gaussian white noise input signals with the objective to excite only the first five modes.
- Sampling period $\Delta t = 0.01$s, total time duration $T = 200$s, total time samples $N = 20000$.
- Static sensor acceleration responses were obtained at 3999 vertical degrees of freedom of the beam using the modal superposition method (including only the first five modes).
- Employing n_{ms} mobile sensors, a stacked matrix of size $n_{ms} \times N$ was created by selecting a specific subset of the noise contaminated static sensor data. The subset corresponds to the static sensor data that coincided with the path of the mobile sensors.
- Zero mean Gaussian white noise sequences with RMS values equal to 20% of the RMS of the corresponding "true" acceleration signals of the stacked matrix were added to obtain noisy measurements.

The MSM underpinning the three algorithms has to be specified with a model order as well as the number of VSLs. A model order of 14 is set and 19 equidistant VSLs are chosen

along the length of the beam as shown in Figure 5.2a. Note the mode shape ordinates are estimated only at these VSLs. The motivation to choose equidistant VSLs stems from the fact that the *sinc*-function based MSR matrices are most accurate when the separation between VSLs is constant. It must be emphasized that in the proposed framework, the number of VSLs can be chosen irrespective of the choice of the number of mobile sensors and/or the model order of the MSM. This is one of the features that contrasts the proposed approach from the TPM approach presented in [21]. Furthermore, the physical modes of the structure are identified based on the two successive conditions as mentioned in the previous chapter, that is, (a) they should correspond to eigenvalues (of $\overline{\mathbf{A}}$) that appear in complex conjugate pairs, and (b) they should have positive damping ratios below 5%. The three algorithms (i.e., EM, VB and GS), employed to estimate the modal parameters from the stacked data matrix, are run with the following specifications:

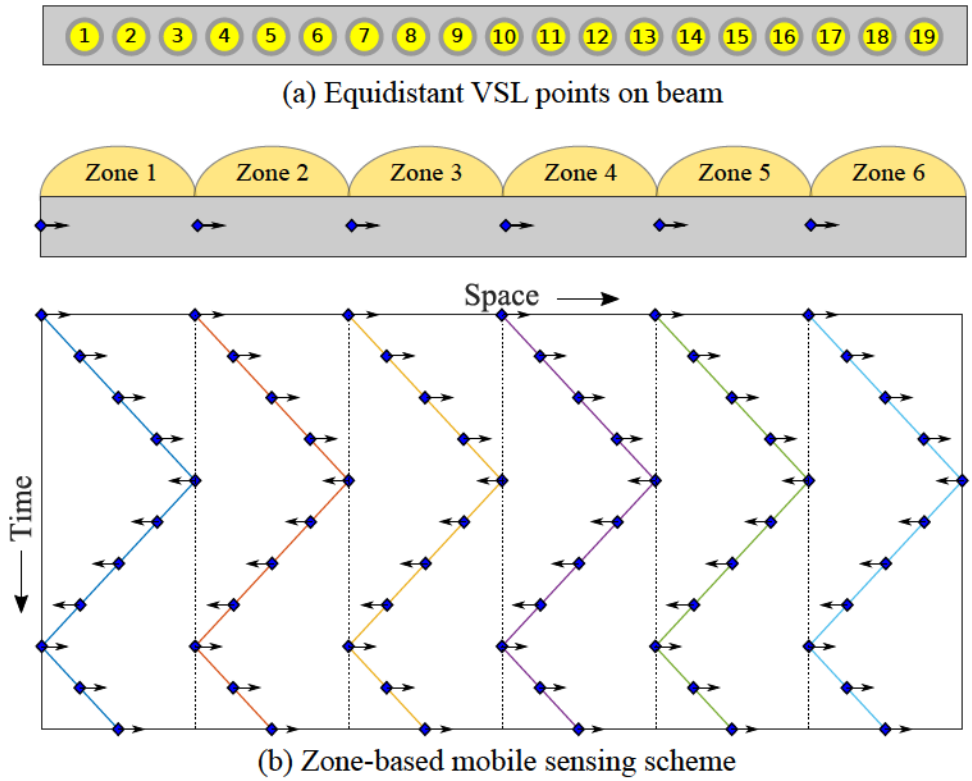
- Parameter estimates of $\overline{\mathbf{A}}^{\text{SSI}}$ and $\overline{\mathbf{G}}^{\text{SSI}}$, with model order equal to 14, were provided as initial estimates; these initial estimates corresponded to frequencies $f^{(0)}$, damping ratios $\xi^{(0)}$ and mode shapes $\boldsymbol{\phi}^{(0)}$ having errors in the range of 10%, 40% and 30%, about their corresponding true values, respectively. The initial errors are introduced to validate the robustness and efficiency of the employed algorithms.

- A tolerance value of $\epsilon_{tol} = 5 \times 10^{-6}$ is set as a stopping threshold for iterations of both EM and VB.

- For GS, a Markov chain of length 3500 (samples) is run. Initial 500 samples of the chain are discarded as burn-in and the rest of the chain is sub-sampled with a period of 3 samples to obtain 1000 samples.

- The remaining parameters related to initialization are set in accordance with the initialization guidelines mentioned in sections 4.4.3, 5.4.4 and 5.5.3.

A mobile sensor network comprising six mobile sensors, each moving with a constant velocity of 0.5m/s, scans 3999 equidistant points on the beam. Figure 5.2b illustrates the mobile sensor network of six moving sensors. The spatial domain of the beam is divided into six coverage zones corresponding to the six mobile sensors. Each mobile sensor scans



(a) Equidistant VSL points on beam

(b) Zone-based mobile sensing scheme

Figure 5.2: (a) 19 equidistant VSLs are selected on a beam where the mode shape estimates are desired, and (b) a network of six mobile sensors, each sensor moving at same speed, scans designated coverage zones on the beam following a back-and-forth movement.

its designated zone with continuous back-and forth motions. In the total duration of 200s, each mobile sensor in the six-mobile-sensor-network is able to scan its designated coverage zone 30 times.

Table 5.1 and 5.2 list the identified modal frequencies and damping ratios using the three methods. Note that EM provides only point estimates whereas VB and GS provides the posterior distributions from which the mean, SD and CoV for each modal parameter can be calculated. The estimates of frequencies from all three methods are quite consistent, although slight overestimation can be seen in the identified frequencies. The CoV of the identified frequencies using both VB and GS are substantially smaller than that of the damping ratios. It is also noted that the estimated CoVs of the identified modal

frequencies and damping ratios from VB are typically smaller than those of GS, with a maximum difference of around 15%. Comparing the CoVs of identified modal frequencies and damping ratios obtained across different modes, it can be seen that the uncertainty is the highest in the first mode and decreases with higher modes.

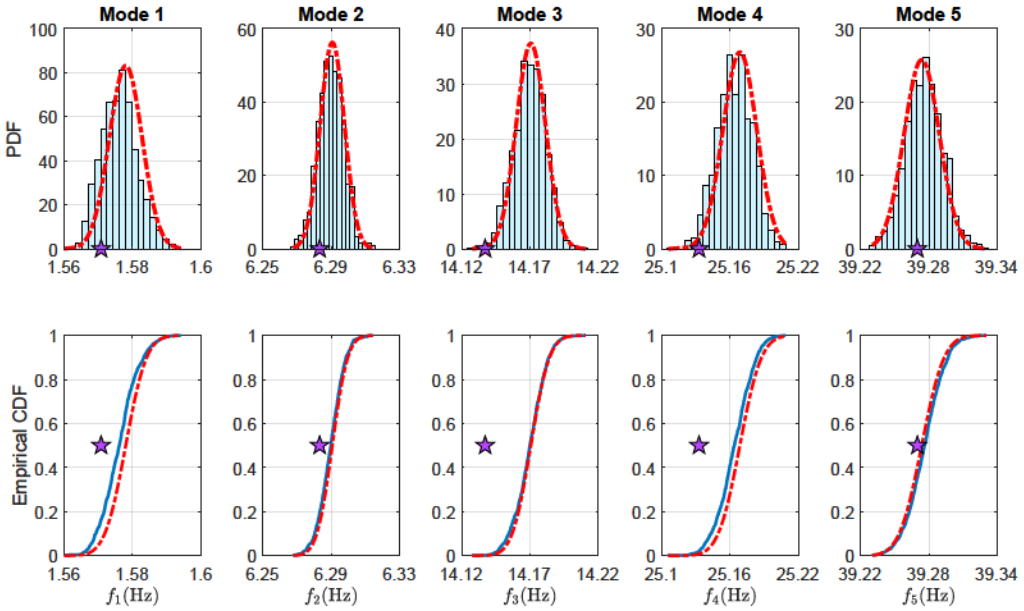
Table 5.1: Identified modal frequencies (in Hz) using EM, VB and GS.

Mode	f_{true}	$f^{(0)}$	EM	VB		GS	
				Mean	CoV (%)	Mean	CoV (%)
1	1.571	1.737	1.577	1.578	0.304	1.576	0.336
2	6.283	6.918	6.292	6.291	0.112	6.290	0.118
3	14.137	12.727	14.177	14.171	0.076	14.170	0.080
4	25.133	27.624	25.161	25.169	0.059	25.164	0.061
5	39.270	35.335	39.274	39.274	0.039	39.276	0.041

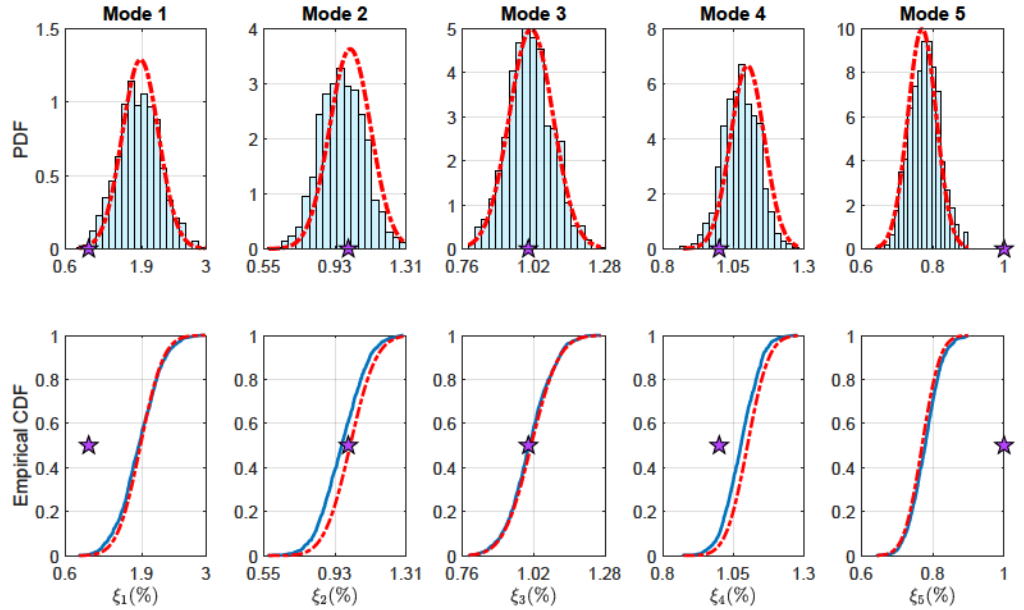
Table 5.2: Identified modal damping ratios (in %) using EM, VB and GS.

Mode	ξ_{true}	$\xi^{(0)}$	EM	VB		GS	
				Mean	CoV (%)	Mean	CoV (%)
1	1.000	0.712	1.611	1.877	16.493	1.865	18.828
2	1.000	1.105	0.980	1.001	10.897	0.974	12.371
3	1.000	0.607	1.043	1.011	7.922	1.010	7.920
4	1.000	0.856	1.075	1.100	5.461	1.088	5.563
5	1.000	1.352	0.762	0.770	5.202	0.781	5.132

Comparison of the PDFs and CDFs of the identified modal frequencies and damping ratios are provided in Figure 5.3. The distributions from VB and GS agree quite well.



(a) PDFs and empirical CDFs of identified modal frequencies



(b) PDFs and empirical CDFs of identified modal damping ratios.

Figure 5.3: Probability distributions of identified modal frequencies and damping ratios. Solid line: GS; dotted line: variational Bayes; star: true value.

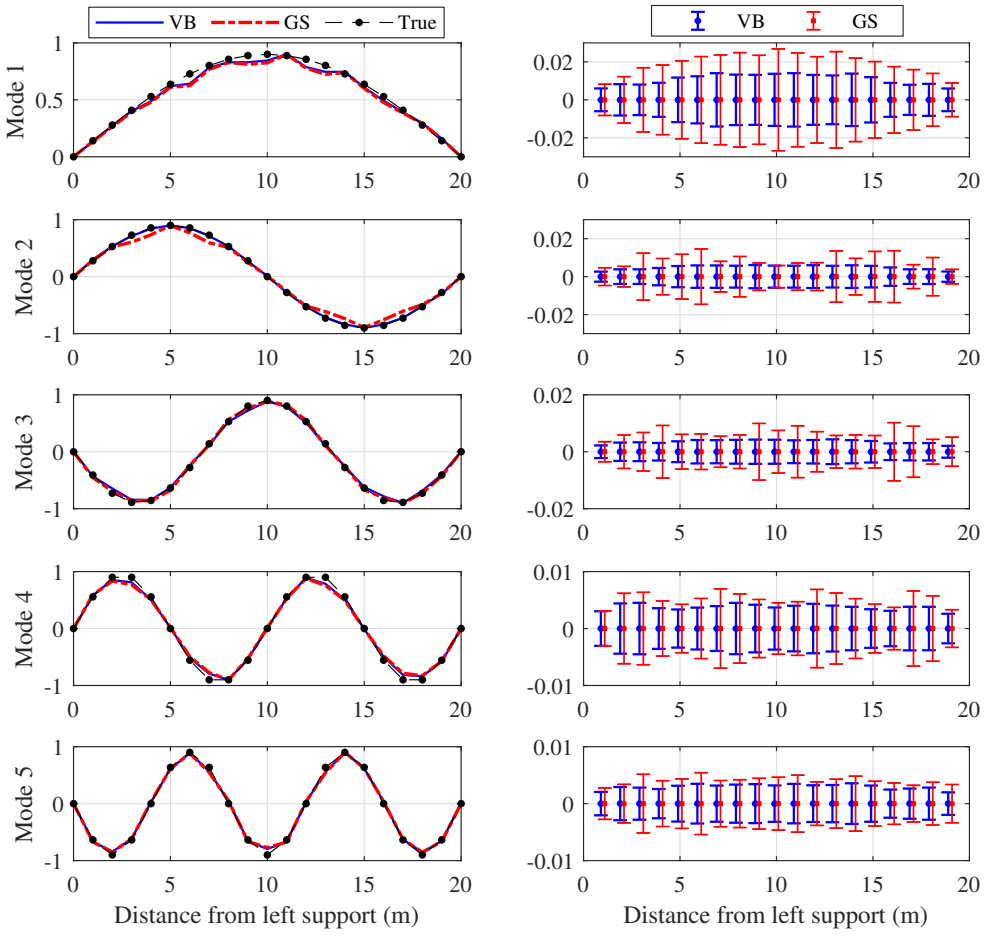
Table 5.3: Comparison of Modal Assurance Criterion (MAC) values of mode shapes identified using EM, VB and GS.

Mode	True	Initial	EM	VB	GS
1	1.000	0.957	0.998	0.998	0.998
2	1.000	0.976	0.999	0.999	0.995
3	1.000	0.982	0.998	0.998	0.998
4	1.000	0.966	0.997	0.997	0.995
5	1.000	0.975	0.995	0.996	0.995

Regarding the mode shape estimates, the MAC values of the mode shapes at the 19 VSLs as estimated by EM, VB and GS are tabulated in Table 5.3. It is seen that the three algorithms, after starting from poor initial mode shape estimates, are able to yield good final mode shape estimates (with MAC values close to 0.99).

The means and the SDs of the mode shapes estimated at the VSLs are plotted in Figure 5.4. The left panel plots the mean values, and it can seen that the mean estimates of the mode shapes from both VB and GS are consistent. The right panel of figure plots the SDs of the mode shape ordinates corresponding to their mean values in the left panel. Comparing the SDs of the estimated mode shapes across different modes, a trend of lower SDs associated with higher modes is seen for this particular example. This is similar to the previously discussed trend of CoVs associated with the identified modal frequencies and damping ratios. It is presumed that a higher sensitivity of higher modes to accelerations result in their lower uncertainties. Finally it is noted that at certain mode shape ordinates across different modes, the SDs from VB exhibit considerable underestimation compared to that from GS i.e. the VB estimates of SD can vary between 40-60% of that from GS. This is particularly prominent for mode 1 where at all 19 VSLs, the uncertainties are underestimated.

The log-likelihood convergence plots of the three algorithms i.e. EM, VB and GS, are shown in Figure 5.5. It can be seen that all three algorithms converge to similar likelihood



(a) Mean (b) SD

Figure 5.4: The left panel shows the mean values and the right panel shows the standard deviations (SDs) associated with the mode shape ordinates at the VSLs, estimated using VB and GS.

values.

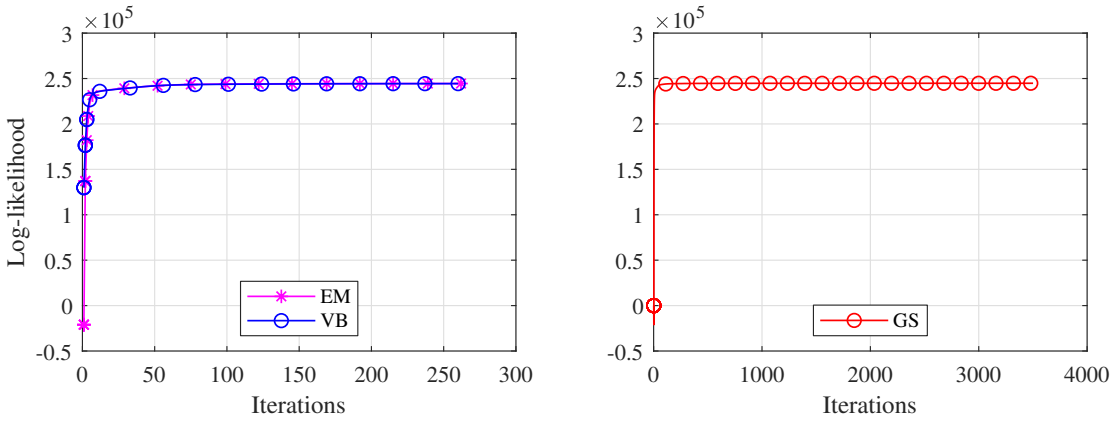


Figure 5.5: Convergence of EM, VB and GS.

Regarding the computational time[2] taken for running the algorithms for modal parameter estimation, the EM and VB algorithms took much less time than the GS as listed in Table 5.4.

Table 5.4: Comparison of computational time taken for modal parameter estimation using EM, VB and GS; computations were done on a SHARCNET server node with Intel E5-2683V4 CPU @2.1GHz processor and 16GB RAM.

Algorithm	Time (hrs)
EM	1.42
VB	2.10
GS	16.57

To this end, the performance of modal parameter estimation using EM, VB and GS on a stacked data matrix acquired using a network of six mobile sensors has been assessed. The following observations can be made:

[2]The computations were done on SHARCNET [118] server nodes

1. The three algorithms are able to provide good final estimates of modal parameters despite starting from poor initial estimates.

2. The point estimates of modal parameters from EM are found to be very close to the mean estimates of the modal parameters from VB.

3. The uncertainty predicted by VB is typically lower than that predicted by GS. While the uncertainty output from VB for the estimated modal frequencies and damping ratios are underestimated by 15-20% of that from GS, the underestimation in the case of identified modal shapes may go up to be around 40-60%.

4. Computationally, the GS takes much longer than VB and EM.

5.6.1 Assessment of estimation performance with different number of VSLs

A desirable feature of the MSM is that it provides flexibility to control the resolution of the mode shapes by adjusting the number of VSLs. This can be achieved irrespective of the model size or the number of mobile sensors. With a single dataset recorded by a group of n_{ms} mobile sensors, the MSM can extract mode shapes at any number of spatial points (i.e. the VSLs), as determined by the user. This section illustrates this feature of the MSM using four different cases consisting of different number of VSLs while keeping the number of mobile sensors constant ($n_{ms} = 6$), as shown in Table 5.5. In each case, the VSLs are selected symmetrically on the beam with uniform spacing between them.

Table 5.5: Different cases of VSLs used with dataset recorded with six mobile sensors ($n_{ms} = 6$).; the VSLs are selected with uniform spacing spanning the length ($L = 20$ m) of the simply supported beam.

Case	**Set of locations** (spanning $L = 20$m)	**No. of VSLs** (n_α)	**Spacing**
`VSL9`	$\{2\text{m}, 4\text{m}, \ldots, 16\text{m}, 18\text{m}\}$	9	2 m
`VSL19`	$\{1\text{m}, 2\text{m}, \ldots, 18\text{m}, 19\text{m}\}$	19	1 m
`VSL39`	$\{0.5\text{m}, 1\text{m}, \ldots, 19\text{m}, 19.5\text{m}\}$	39	0.5 m
`VSL79`	$\{0.25\text{m}, 0.5\text{m}, \ldots, 19.5\text{m}, 19.75\text{m}\}$	79	0.25 m

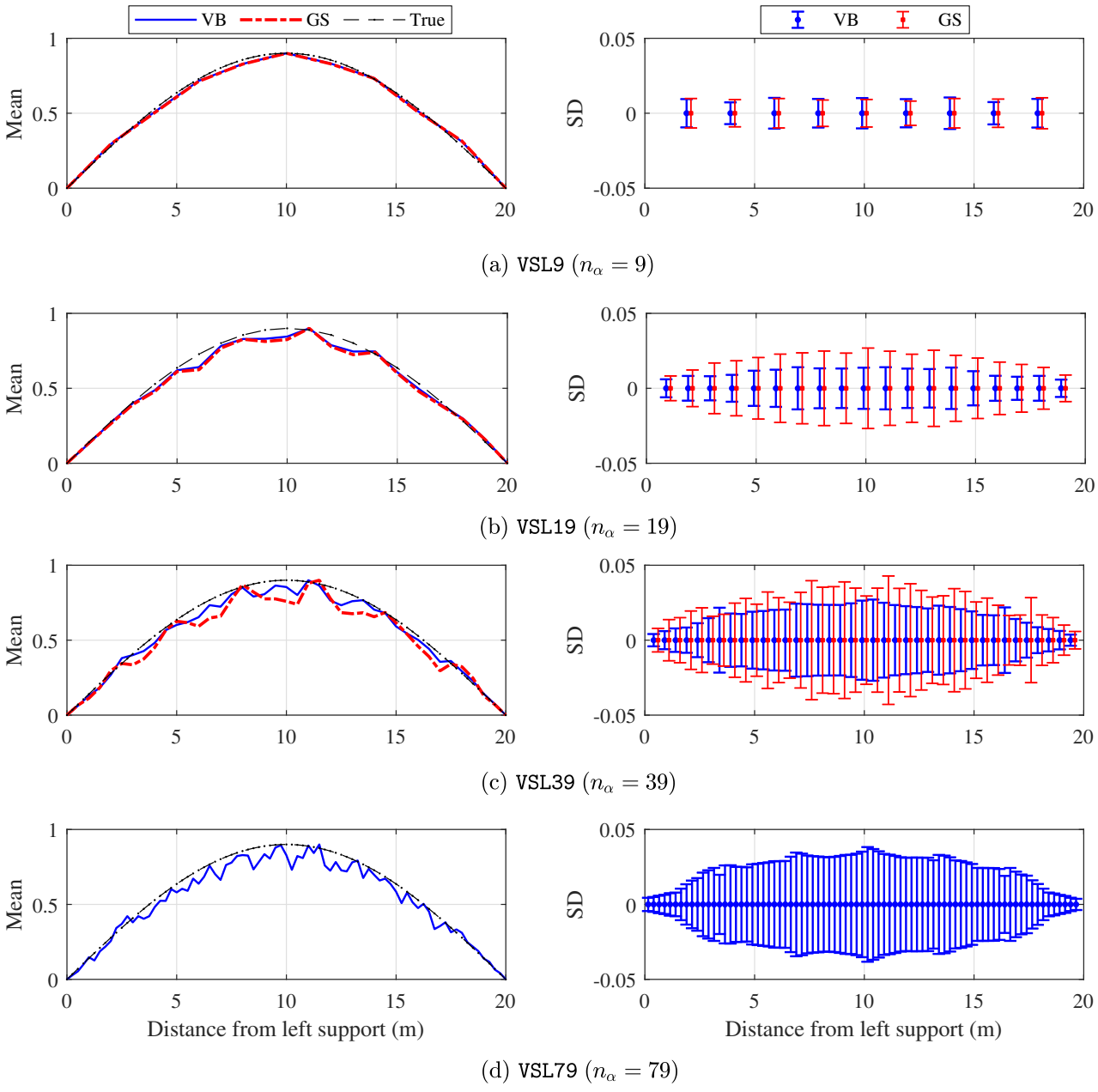
First, it is noted that the computational time required by VB and GS for parameter estimation increases as the number of VSLs increases, as listed in Table 5.6. This is due to the increase in the size of the observation matrix (i.e., $\overline{\mathbf{G}}$) with an increase in the number of VSLs. Specifically, the estimation of MSM parameters for the case `VSL79` (with 79 VSLs) using the GS can take an inordinate amount of computational time.

Table 5.6: Comparison of computational time (in hrs) taken by VB and GS; computations were done on a SHARCNET server node with Intel E5-2683V4 CPU @2.1GHz processor and 16GB RAM.

Algorithm	Time used (in hrs)			
	`VSL9`	`VSL19`	`VSL39`	`VSL79`
VB	1.53	2.10	3.69	8.21
GS	9.63	16.57	28.04	85.06

The means and SDs of the mode shapes estimated by VB and GS, corresponding to the four cases of different VSLs, are plotted for the same three representative modes (i.e., modes 1, 2 and 5) in Figures 5.6, 5.7 and 5.8 respectively. It is expected that a lower number

of VSLs would be unable to capture an accurate representation of the higher modes. This is confirmed in the left topmost plot in Figure 5.8 where the estimation with 9 VSLs points (case `VSL9`) provides a poor representation of the fifth mode shape. Similarly, the use of more VSLs improves the representation of the fifth mode shape as shown in the left bottommost plot in Figure 5.8. Nevertheless, the improved representation is associated with a trade-off in the accuracy of mode shape estimates. That is, as the number of VSLs increase, the mode shape estimates tend to become more noisy. This is clear in Figure 5.6, where the first mode shape becomes increasingly poor with increasing numbers of VSLs. Moreover, the first modal estimate from GS for the case `VSL79` was unable to converge to correct values (and hence omitted in Figure 5.6), implying that GS would need increasingly more samples to converge as the number of VSLs increase.



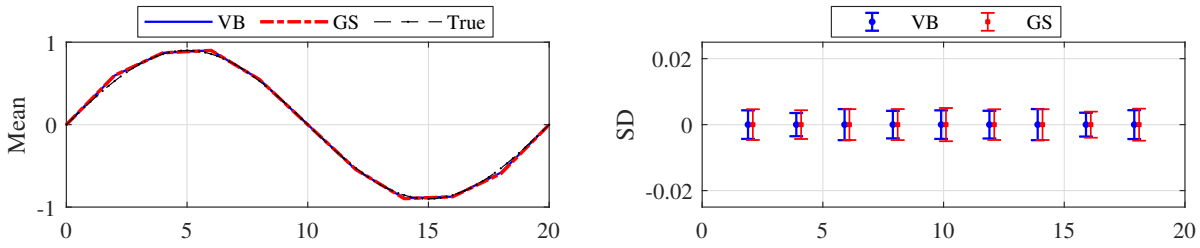
(a) `VSL9` ($n_\alpha = 9$)

(b) `VSL19` ($n_\alpha = 19$)

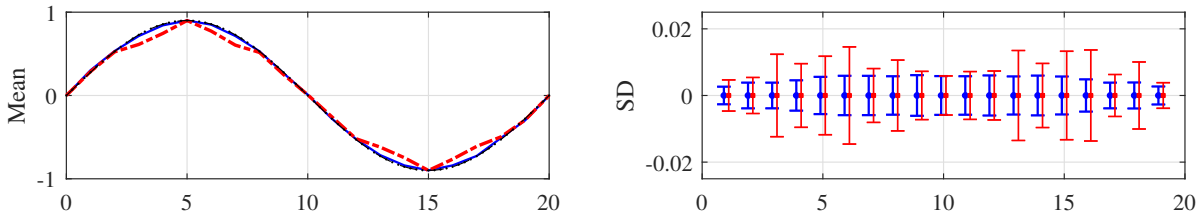
(c) `VSL39` ($n_\alpha = 39$)

(d) `VSL79` ($n_\alpha = 79$)

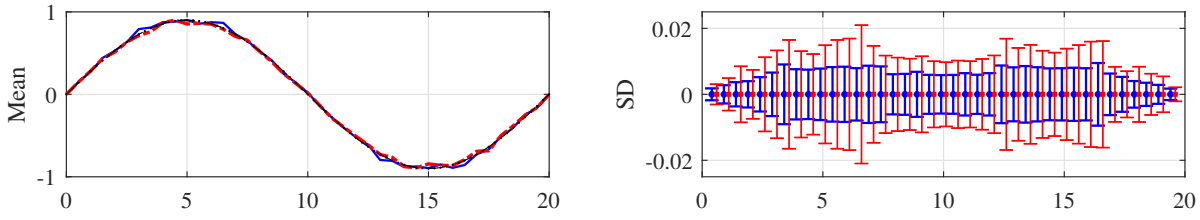
Figure 5.6: Comparison of the means and the SDs of the estimated first mode shape obtained from VB and GS for the four different cases `VSL9`, `VSL19`, `VSL39` and `VSL79`; $n_{ms} = 6$. Note GS did not converge for `VSL79` and hence omitted from plotting.



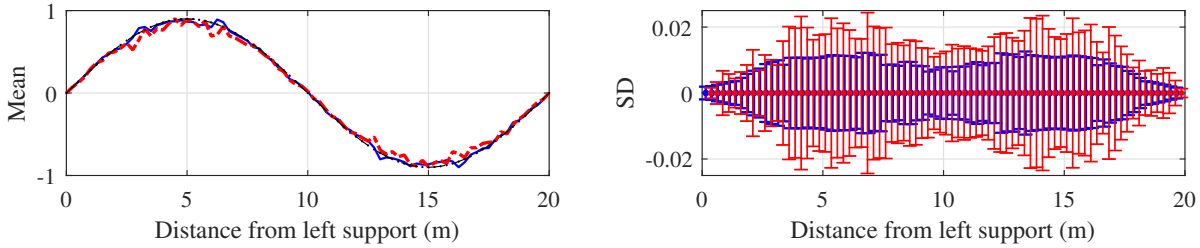
(a) `VSL9` ($n_\alpha = 9$)



(b) `VSL19` ($n_\alpha = 19$)

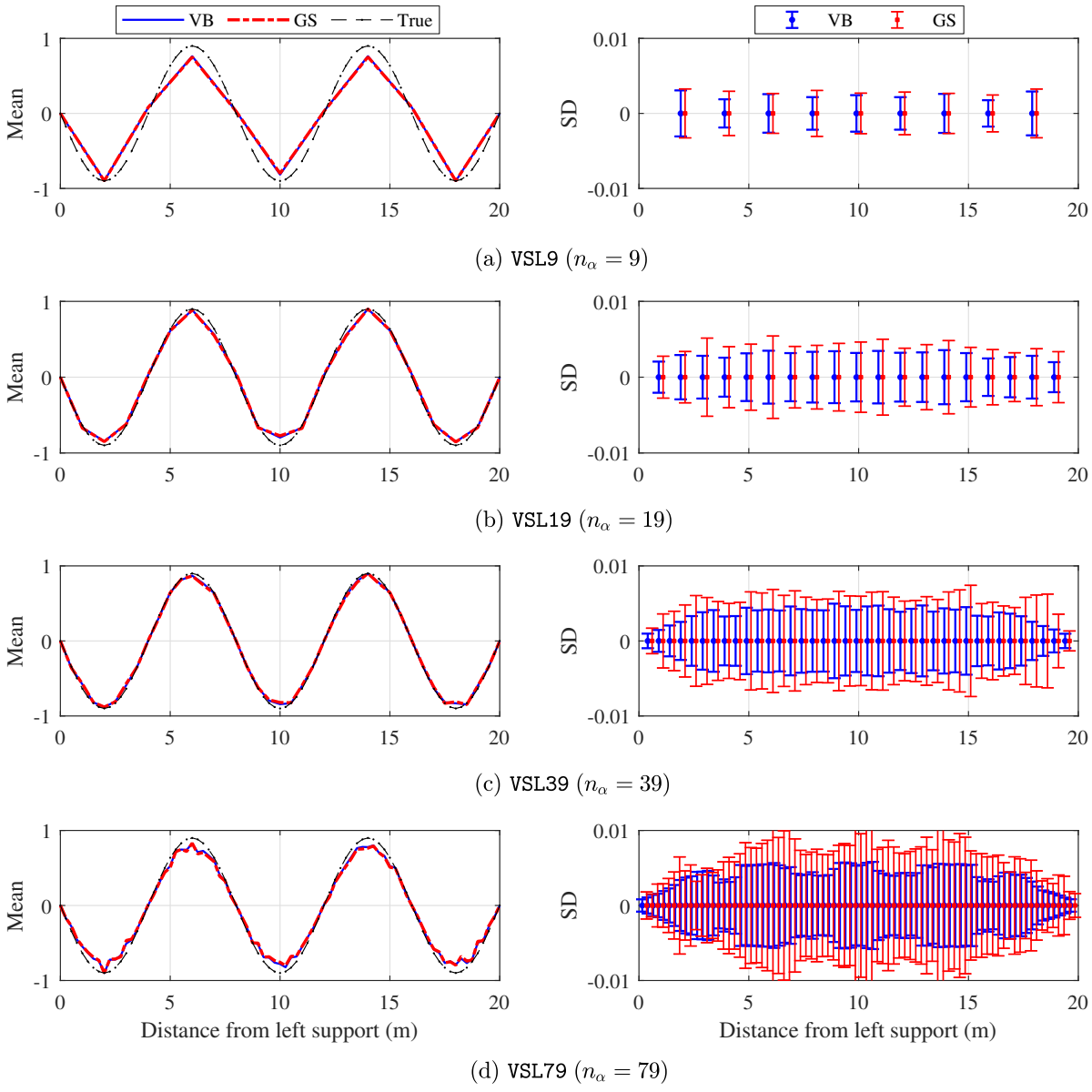


(c) `VSL39` ($n_\alpha = 39$)



(d) `VSL79` ($n_\alpha = 79$)

Figure 5.7: Comparison of the means and the SDs of the estimated second mode shape obtained from VB and GS for the four different cases `VSL9`, `VSL19`, `VSL39` and `VSL79`; $n_{ms} = 6$.



(a) `VSL9` ($n_\alpha = 9$)

(b) `VSL19` ($n_\alpha = 19$)

(c) `VSL39` ($n_\alpha = 39$)

(d) `VSL79` ($n_\alpha = 79$)

Figure 5.8: Comparison of the means and the SDs of the estimated fifth mode shape obtained from VB and GS for the four different cases `VSL9`, `VSL19`, `VSL39` and `VSL79`; $n_{ms} = 6$.

The decrease in the accuracy of mode shape estimates with increase in the number of

VSLs is related to the normalized sinc function spatial interpolator, particularly to its effective interpolation bandwidth – the spatial region where the sinc function interpolator has non-negligible amplitude. Figure 5.9 shows normalized sinc functions constructed at the mid-span for cases `VSL9` and `VSL79`. When the number of VSLs is small (e.g. case `VSL9`), the effective interpolation bandwidth of the sinc function is quite large, enabling interpolation from a larger number of neighbouring samples and increasing accuracy in mode shape estimation. In contrast to that, for greater number of VSLs (e.g. case `VSL79`) the sinc function decays very fast decreasing its effective interpolation bandwidth; thus only a few neighbouring samples contributes to the interpolation and therefore reduces the accuracy of estimated mode shapes.

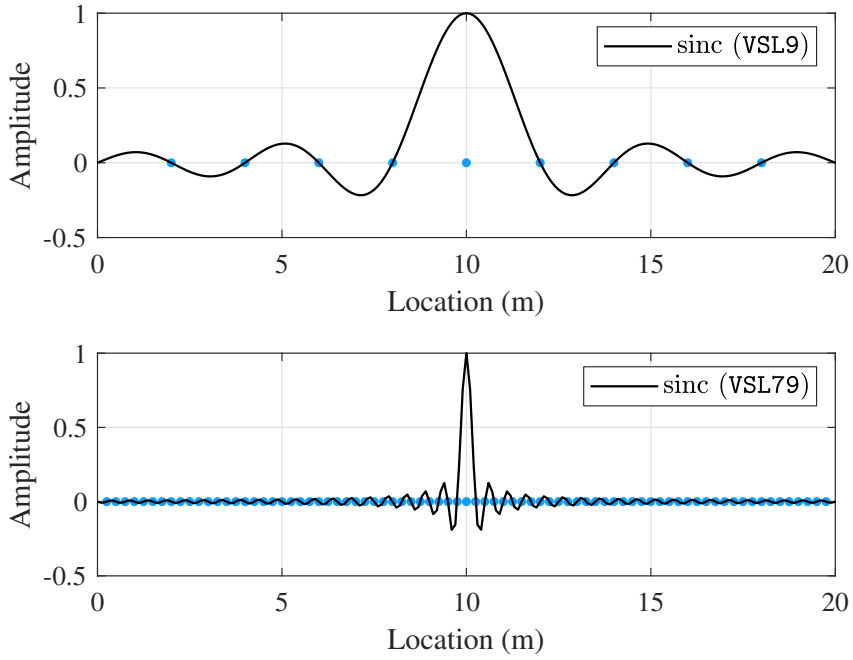


Figure 5.9: Comparison of effective interpolation bandwidth of a normalized sinc function drawn for a VSL at mid-span for cases `VSL9` (top) and `VSL79` (bottom).

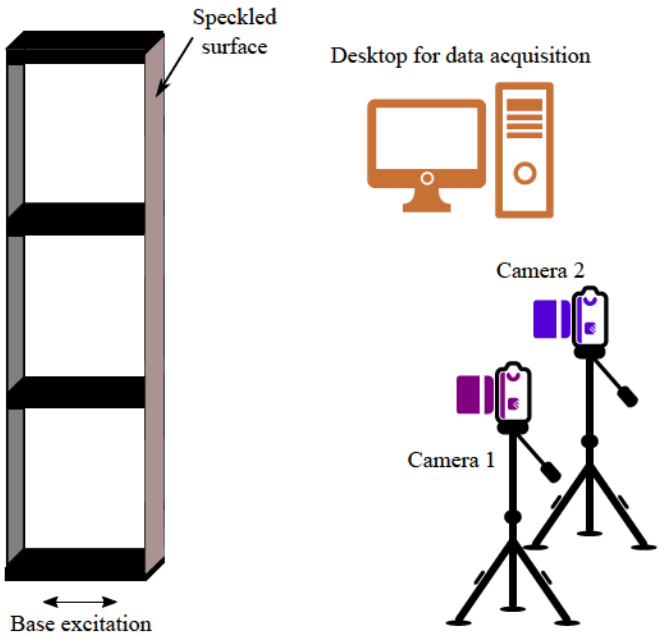
It is found that mode shape estimates from VB are more accurate than those from GS as seen from Figures 5.6, 5.7 and 5.8 for the first, second and fifth mode shapes respectively. Additionally, the uncertainty in the mode shape estimates increase as the number of VSLs

increases. This is because as more information is extracted from the same dataset, the uncertainty in the information increases.

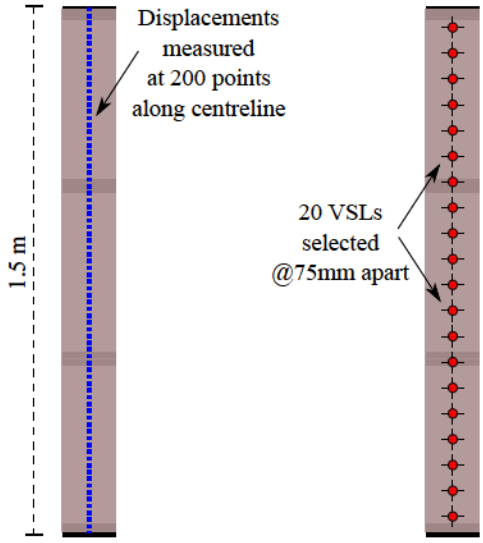
5.7 Experimental verification

In this section, the proposed framework of MSM-based Bayesian inference is experimentally assessed for modal parameter identification following the stacked data approach to mobile sensing. The 3-storey bench-scale test frame, described in Section 4.8, was used for testing. However, instead of a few static accelerometers gathering data at a few locations, a 3D digital image correlation (DIC) technique was used to get displacement responses at 200 locations along the height of the 3-storey test frame. Since a distinct advantage of the proposed framework is that it allows incorporating a large number of mobile sensing nodes, a 3D-DIC setup was found well suited for this purpose as it could measure static responses at a large number of locations and an appropriate mobile sensor dataset could be extracted from it subsequently. The setup for the 3D-DIC experiment is illustrated in Figure 5.10. Two digital cameras were used to measure displacement responses at 200 equidistant locations along the height of the 3-storey building frame. A sampling frequency of 30Hz was used for 3D-DIC test and a total of 7200 samples of displacement measurements were obtained at the 200 locations. The averaged power spectral density of the 200 static displacement measurements is shown in Figure 5.12. The first three modal frequencies were found to be 2.20, 6.56 and 9.99Hz respectively. The third mode had very low energy and its corresponding peak in the PSD was not so distinct. Within the frequency band of 10Hz, two additional peaks are observed at 4.40Hz and 8.80Hz respectively which are found to be spurious modes. The displacement responses were low-pass filtered with a cutoff of 8Hz using a 5th order Butterworth filter for use in modal identification.

A stacked dataset employing three mobile sensors is simulated using the static sensor dataset. A zone-based mobile sensing scheme is followed – in which the height of the frame is divided into three coverage zones corresponding to three mobile sensors – a mobile sensor dataset is then created by selecting the subset of the static sensor responses that coincided with the virtual mobile sensor paths. For modal identification with the proposed approach,



(a) 3D-DIC experimental setup



(b) Measurement locations and selection of VSLs

Figure 5.10: Schematic for modal identification using 3D-DIC experiment.

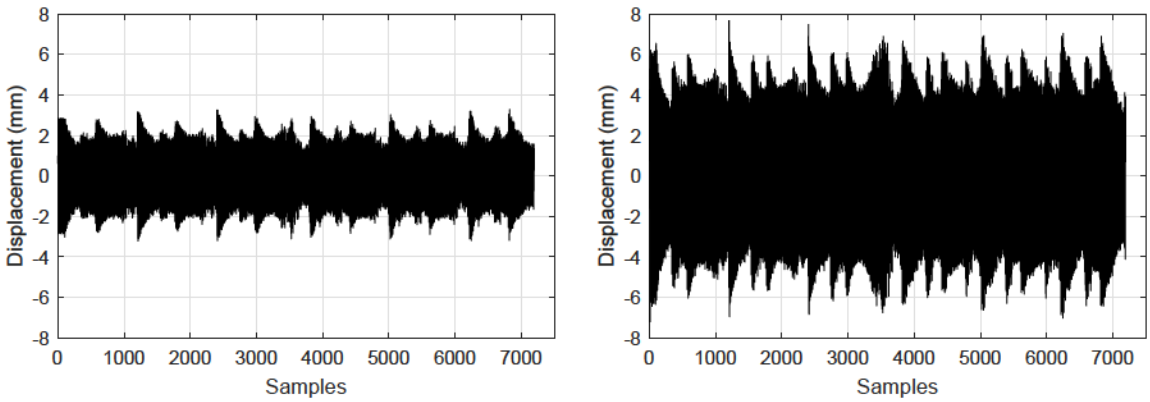


Figure 5.11: Sample displacement time histories from DIC at middle and top of the test frame.

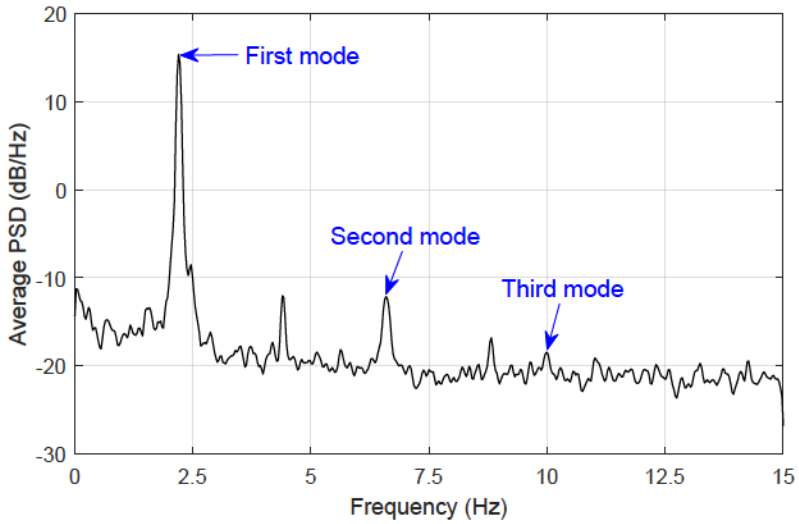


Figure 5.12: Averaged power spectral density of the displacement measurements from DIC.

a model order of 30 is chosen for the MSM and twenty equally spaced VSLs at 75mm apart are selected on the beam. Modal identification results obtained from using SSI (with a model order of 75) using static displacements at the locations of the VSLs are treated as baseline results for modal identification.

The computational algorithms EM, VB and GS are initialized with $\mathbf{A}^{(0)}$ and $\mathbf{G}^{(0)}$ matrices set equal to the estimates from SSI. For EM and VB the default stopping threshold of 5×10^{-6} is used. For GS, a Markov chain of 2000 samples was simulated where the first 1000 samples were discarded as burn-in and the rest 1000 samples (without subsampling) were used for the computation of posterior distributions. The modal parameters identified with the computational algorithms using the stacked dataset are then compared to the baseline results from SSI.

The frequency and damping estimates obtained from EM, VB and GS are listed in Tables 5.7 and 5.8 respectively. The frequency estimates from SSI using static sensor data and that from EM, VB and GS using mobile sensor data are in good agreement. The damping estimates from SSI using static sensor data is found to differ from those obtained with EM, VB and GS, especially for the second mode. The damping estimates from EM, VB and GS for the first modes are found to be pretty consistent with each other, while for the second mode the VB damping estimate seem to lower than that obtained using EM and GS.

Table 5.7: Identified modal frequencies (in Hz) using EM, VB and GS.

Mode	f^{SSI}	$f^{(0)}$	EM	VB			GS		
				Mean	SD ($\times 10^{-2}$)	CoV (%)	Mean	SD ($\times 10^{-2}$)	CoV (%)
1	2.199	2.311	2.203	2.203	0.166	0.075	2.204	0.167	0.076
2	6.561	6.328	6.592	6.583	0.659	0.101	6.581	0.755	0.115

Table 5.8: Identified modal damping ratios (in %) using EM, VB and GS.

Mode	ξ^{SSI}	EM	VB			GS		
			Mean	SD ($\times 10^{-2}$)	CoV (%)	Mean	SD ($\times 10^{-2}$)	CoV (%)
1	0.133	0.224	0.224	8.304	37.071	0.173	7.052	40.763
2	0.206	0.899	0.991	10.080	10.172	0.778	12.300	15.810

The mean values and the SDs of the two mode shapes estimated using VB and GS are illustrated in Figure 5.13. The mean estimates of the first mode shape from VB and GS agree quite well with the baseline first mode shape estimated using SSI at the VSLs. However, the VB and GS estimates of the second mode shape from are seen to differ from that using SSI. Also, the second mode shape has large uncertainty compared to the first mode shape.

The MAC values of the modes shapes obtained using EM, VB and GS are listed in Table 5.9. The MAC values are obtained with respect to the baseline mode shapes from SSI.

Table 5.9: Modal Assurance Criterion (MAC) values of mode shapes identified using EM, VB and GS; MAC values calculated with respect to mode shape estimates from SSI.

Mode	EM	VB	GS
1	0.986	0.994	0.994
2	0.918	0.935	0.927

Computationally, the VB worked much faster taking roughly 4 hours in comparison to GS which took 10 hours, both executed on a SHARCNET [118] server node with 16GB RAM and Intel E5-2683V4 CPU @2.1GHz processor. It is to be mentioned that only one run of the MSM produced modal ordinates at 20 VSLs which is advantageous compared to

the TPM [21] that would need multiple runs to estimate mode shapes at the same number of VSLs.

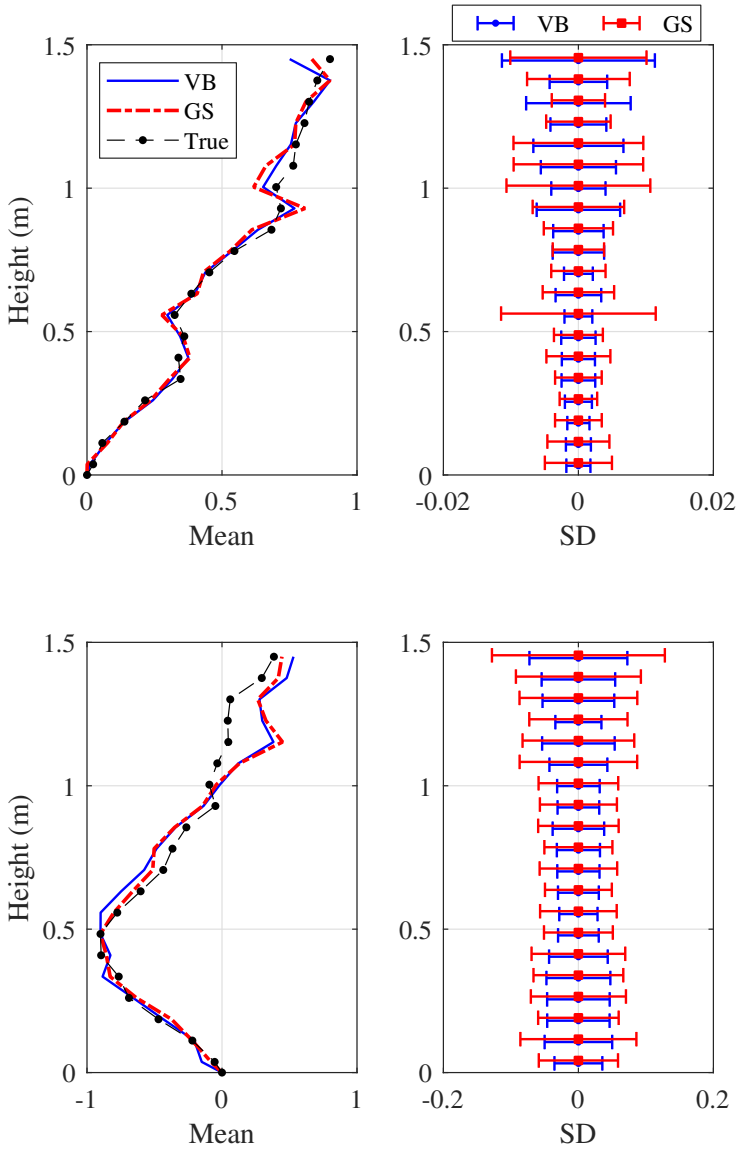


Figure 5.13: Means (left panel) and SDs (right panel) of two mode shapes obtained with VB and GS using the mobile sensor dataset is compared with the mode shape estimates obtained with SSI using the static sensor dataset.

5.8 Summary

In this chapter, the problem of output-only modal parameter estimation using in-motion mobile sensors is posed from a complete data perspective. The time-series data recorded by each mobile sensor are concatenated to form a stacked data matrix; each entry in the stacked data matrix is tagged by its corresponding position of record – stored in a sensor-position matrix. The stacked data matrix is featured by spatial discontinuities which makes it incompatible with the conventional system identification methods. In this study, a modal state model (MSM) is proposed to conduct modal parameter estimation using the stacked data matrix. The MSM is based on the concept of converting the mobile sensor data to equivalent static sensor data at some user-chosen set of virtual locations on the structure. A sequence of sinc function based mode shape regression (MSR) matrices enables interpolation of the mobile sensor entries from various locations to these VSLs on the structure. By regulating the number of VSLs used in the MSM, the resolution of the estimated mode shapes can be controlled.

A Bayesian inference framework is combined with the proposed MSM for estimating the posterior distribution of modal parameters via the estimation of MSM parameters. Three computational algorithms, namely the EM, VB and GS, are modified to enable inference with the MSM. The performance of the modal parameter estimation and uncertainty quantification is studied using numerical simulations on a simply supported Euler-Bernoulli beam. The posterior means of the modal frequencies and damping ratios obtained all three algorithms are found to be similar, but that for the mode shapes are found to better from EM and VB. The VB however tends to underestimate the uncertainty in the estimated modal parameters compared to GS. In a comparison study, keeping the number of mobile sensors constant, the number of VSLs used in MSM are varied to investigate its effect on modal parameter estimation. While the posterior means of the modal frequencies and damping ratios did not vary much, it is found that selecting a small number of VSLs produces lower resolution mode shape estimates while choosing a very high number produces high resolution but noisy estimates of mode shapes. Moreover, the uncertainty in the mode shape estimates also increase with an increase in the number of VSLs. Computationally, the EM and VB algorithms take similar computational time, however, the computational

demand of GS is usually very high. Finally, an experimental study was undertaken on a laboratory scale 3-storey building frame. A mobile sensor dataset was created from the dense static displacements (obtained from DIC) which corresponded to a displacement dataset collected using three mobile sensors. A MSM with 20 VSLs was employed to carry out Bayesian inference using the mobile sensor dataset and the estimation results were found to be comparable to that obtained by conventional techniques using static sensor data.

Chapter 6

Extracting contact-point responses for suppression of carrier-vehicle dynamics

In the previous chapters, it was assumed that the mobile sensors were able to directly measure the bridge responses at vehicle-bridge contact points. In other words, the carrier-vehicles (i.e., the mobile platforms carrying the sensor) were assumed inertialess and the vehicle dynamics affecting the responses measured by the mobile sensors were ignored. However, in practice, the carrier-vehicle can introduce additional dynamics to the sensor measurements – it acts a low pass filter to the bridge responses at contact points. This low-pass filtering phenomena can mask the bridge dynamics and prevent successful identification of modal features. Therefore, a procedure to suppress the effect of vehicle dynamics in the measurements prior to their use in modal identification is desired. This chapter proposes a technique that attempts to reduce the effect of vehicle dynamics via the recovery of contact-point responses.

The contact-point response is defined as the response of the vehicle at the point of contact of the vehicle with the bridge. As will be evident, this response acts as a base-excited input to the carrier-vehicle whose response is measured by the sensors on the top of the carrier-vehicle. The problem of recovering contact-point responses from the vehicle

response measurements is cast as an input reconstruction problem. A Gaussian process latent force model (GPLFM)-based Kalman filtering approach is adopted to estimate the unknown contact-point response. The main reason to use GPLFM over some of the popular input estimation techniques [121–124] is that it adds stability to the inverse-estimation procedure and thereby provides drift-free input and state estimates [125]. It is assumed that the knowledge of the dynamic characterization of the carrier-vehicle is available in the form of either a transfer function or a state-space model. Upon recovery of contact-point responses, they can be used as outputs in bridge modal identification using any of the methods presented in the Chapters 4 and 5.

This chapter is organized as follows. First, a mathematical model for the coupled dynamics of a sensor-instrumented vehicle crossing a bridge is presented. Next, a relationship between the vehicle response and the contact-point response is derived where it is shown that the contact-point response can be considered as a base-excited input to the vehicle dynamical system. To recover the base-excited input (i.e., the contact-point response), a GPLFM approach is proposed and the procedure to estimate the base-excited input is presented. Finally, the effectiveness of the proposed approach is assessed through a numerical case study involving a sensor-instrumented-vehicle moving over a bridge.

6.1 Carrier-vehicle-bridge interaction model

In this study, the carrier-vehicle is modelled as a linear sprung mass damper with a single DoF (SDoF) moving at speed v as shown in Figure 6.1 and the bridge deck is modelled as a simply supported Euler-Bernoulli beam. Note that the carrier-vehicle could also be modelled using multiple DoFs (such as a quarter car model).

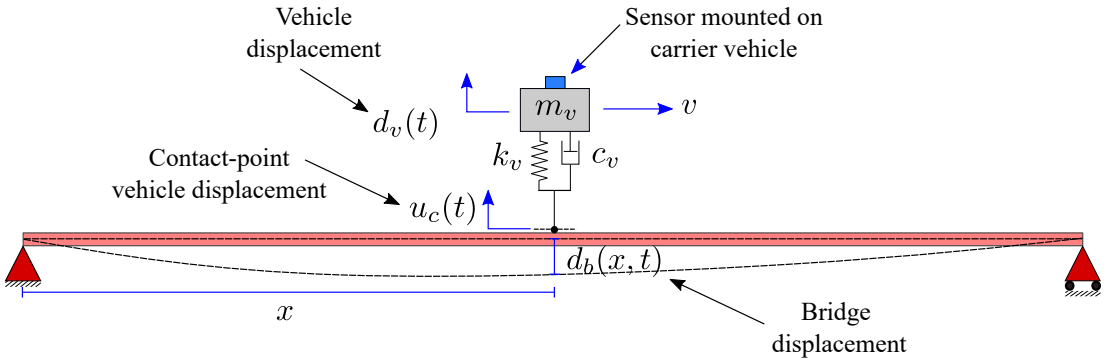


Figure 6.1: Vehicle-bridge interaction.

6.1.1 Vehicle equation of motion

The equation of motion for a linear SDoF carrier-vehicle moving on a smooth bridge deck can be written as

$$m_v\ddot{d}_v(t) + c_v\dot{d}_v(t) + k_v d_v(t) = c_v\dot{d}_b(vt,t) + k_v d_b(vt,t) \tag{6.1}$$

where m_v, k_v, c_v are the mass, stiffness and damping of the vehicle respectively, $d_v(t)$ is the vertical displacement of the vehicle, $d_b(vt,t)$ is the vertical bridge displacement at the spatial location $x = vt$ at time t.

6.1.2 Bridge equation of motion

The governing equation for flexural vibrations of a bridge deck (modelled as Euler-Bernoulli beam) under the influence of a moving vehicle can be written as

$$EI\frac{\partial^4 d_b(x,t)}{\partial x^4} + \rho_b\frac{\partial^2 d_b(x,t)}{\partial t^2} + \mu_b\frac{\partial d_b(x,t)}{\partial t} = -f(t)\delta(x-vt) \tag{6.2}$$

where EI is the flexural rigidity of the bridge deck, ρ_b is the mass per unit length, μ_b is the viscous damping parameter, $f(t)$ is the interaction force between the force and vehicle with $f(t) = m_v g + m_v\ddot{d}_v(t)$; g is the acceleration due to gravity and $\delta(\cdot)$ is the Dirac's delta function.

The displacement response of the bridge deck can be expressed using modal superposition as

$$d_b(x,t) = \sum_{r=1}^{n_m} \phi_r(x)\eta_r(t) \tag{6.3}$$

where n_m is the number of participating modes. $\phi_r(x) = \sqrt{\frac{2}{\rho_b L}} \sin \frac{r\pi x}{L}$ is the rth vibration mode shape, $\eta_r(t)$ is the rth modal response and L is the span of the bridge.

Substituting Eq.(6.3) in Eq.(6.2) and applying orthogonality conditions of vibration modes yields

$$\ddot{\eta}_r(t) + 2\zeta_r\omega_r\dot{\eta}_r(t) + \omega_r^2\eta_r(t) = -f(t)\phi_r(vt) \tag{6.4}$$

where ω_r and ζ_r are the rth modal frequency and damping ratios respectively given by relations: $\omega_r^2 = \frac{EI}{\rho_b}\left(\frac{r\pi}{L}\right)^4$ and $\zeta_r = \frac{\mu_b}{2\rho\omega_r}$.

The bridge deck is typically subjected to ambient excitations resulting from on-going traffic, ground motion, wind excitation, etc. For simplicity, the ambient excitation is considered through support excitations modelled as white noise time series applied at the right and left bridge supports, denoted by $\ddot{d}_r(t)$ and $\ddot{d}_l(t)$, respectively. The equation of motion for the ith vibration mode of the deck subject to moving vehicle and support excitations can then be written as

$$\ddot{\eta}_r(t) + 2\zeta_r\omega_r\dot{\eta}_r(t) + \omega_r^2\eta_r(t) = -f(t)\phi_r(vt) - P_r(t) \tag{6.5}$$

where $P_r(t) = \ddot{d}_l(t)\int_0^L \rho_b\left(1-\frac{x}{L}\right)\phi_r(x)dx + \ddot{d}_r(t)\int_0^L \rho\frac{x}{L}\phi_r(x)dx$.

Combining Eqs.(6.1) and (6.5), the equation of motion of the combined vehicle-bridge system is obtained in the matrix form as

$$\mathbf{M}_a(t)\ddot{\boldsymbol{d}}_a(t) + \mathbf{C}_a(t)\dot{\boldsymbol{d}}_a(t) + \mathbf{K}_a(t)\boldsymbol{d}_a(t) = \boldsymbol{f}_a(t) \tag{6.6}$$

where

$$\mathbf{M}_a(t) = \begin{bmatrix} 1 & 0 & \dots & 0 & m_v\phi_1(vt) \\ 0 & 1 & \dots & 0 & m_v\phi_2(vt) \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \dots & 1 & m_v\phi_n(vt) \\ 0 & 0 & \dots & 0 & m_v \end{bmatrix}$$

$$\mathbf{C}_a(t) = \begin{bmatrix} 2\zeta_1\omega_1 & 0 & \dots & 0 & 0 \\ 0 & 2\zeta_2\omega_2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \dots & 2\zeta_{n_m}\omega_{n_m} & 0 \\ -c_v\phi_1(vt) & -c_v\phi_2(vt) & \dots & -c_v\phi_{n_m}(vt) & c_v \end{bmatrix}$$

$$\mathbf{K}_a(t) = \begin{bmatrix} \omega_1^2 & 0 & \dots & 0 & 0 \\ 0 & \omega_2^2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \dots & \omega_{n_m}^2 & 0 \\ -c_v\dot{\phi}_1(vt) - k_v\phi_1(vt) & -c_v\dot{\phi}_2(vt) - k_v\phi_2(vt) & \dots & -c_v\dot{\phi}_{n_m}(vt) - k_v\phi_{n_m}(vt) & k_v \end{bmatrix}$$

$$\boldsymbol{d}_a(t) = \begin{bmatrix} \eta_1(t) \\ \eta_2(t) \\ \vdots \\ \eta_{n_m}(t) \\ d_v(t) \end{bmatrix} \quad \boldsymbol{f}_a(t) = \begin{bmatrix} -\phi_1(vt)m_v g - P_1(t) \\ -\phi_2(vt)m_v g - P_2(t) \\ \vdots \\ -\phi_{n_m}(vt)m_v g - P_{n_m}(t) \\ 0 \end{bmatrix}.$$

Eq.(6.6) represents an linear ordinary differential equation with time-varying coefficient matrices and can be solved using Newmark-Beta explicit method [126]. Upon solving Eq.(6.6), one obtains the dynamic response of the moving vehicle, which in practice, would represent the measured mobile sensor responses.

6.2 Contact-point responses as base excitation to the carrier-vehicle

The contact-point displacements, denoted by $u_c(t)$, are displacements of the vehicle at the contact point of the vehicle with the bridge, and act as an estimator of the bridge responses $d_b(x,t)$ at $x = vt$. Therefore, replacing $d_b(vt,t)$ with $u_c(t)$ in Eq.(6.1) and rewriting it in relative coordinates $d_{v,rel}(t) = d_v(t) - u_c(t)$, one obtains

$$m_v\ddot{d}_{v,rel}(t) + c_v\dot{d}_{v,rel}(t) + k_v d_{v,rel}(t) = -m_v\ddot{u}_c(t) \tag{6.7}$$

This form of Eq.(6.7) resembles the equation of motion of a SDoF dynamical system subject to a base-excited input, which means that the contact-point acceleration $\ddot{u}_c(t)$ acts as a base-excitation to the carrier-vehicle. Note that the carrier-vehicle typically records the absolute vehicle acceleration response $\ddot{d}_v(t)$.

With known vehicle dynamics, Eq.(6.7) can be written in the state space form as

$$\dot{\boldsymbol{x}}_v(t) = \mathbf{A}_{vc}\boldsymbol{x}_v(t) + \mathbf{B}_{vc}\ddot{u}_c(t) \tag{6.8a}$$

$$y_v(t_k) = \ddot{d}_v(t_k) = \mathbf{G}_v\boldsymbol{x}_v(t_k) \tag{6.8b}$$

where $\boldsymbol{x}_v(t)$ represent the vehicle states and $y_v(t_k)$ is the absolute acceleration response measured by the sensor on the vehicle. The terms in Eq.(6.8) are given by

$$\boldsymbol{x}_v(t) = \begin{bmatrix} d_{v,rel}(t) \\ \dot{d}_{v,rel}(t) \end{bmatrix}, \quad \mathbf{A}_{vc} = \begin{bmatrix} 0 & 1 \\ -\frac{k_v}{m_v} & -\frac{c_v}{m_v} \end{bmatrix}, \quad \mathbf{B}_{vc} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \quad \mathbf{G}_v = \begin{bmatrix} -\frac{k_v}{m_v} & -\frac{c_v}{m_v} \end{bmatrix} \tag{6.9}$$

It is noted that there is no direct feedthrough term in Eq.(6.8b) because the absolute acceleration is measured (this term gets cancelled in the observation equation Eq.(6.8b)). As well, the state space matrices $\mathbf{A}_{vc}$, $\mathbf{B}_{vc}$ and $\mathbf{G}_v$ are subject to similarity transformation [127], that is, for a measured input-output pair there could possibly be infinite pairs of $\mathbf{A}_{vc}$, $\mathbf{B}_{vc}$ and $\mathbf{G}_v$ satisfying Eq.(6.8), and hence the triplet $\{\mathbf{A}_{vc}, \mathbf{B}_{vc}, \mathbf{G}_v\}$ in Eq.(6.9) is just one possible pair.

Now, given noisy vehicle acceleration measurements $y_v(t_k)$ and the state matrices $\mathbf{A}_{vc}$, $\mathbf{B}_{vc}$ and $\mathbf{G}_v$ for the carrier-vehicle – identified *apriori* using some system identification

technique – the goal is to infer the contact-point acceleration response $\ddot{u}_c(t_k)$ at discrete time points $k = 1, \ldots, N$. To achieve this goal, a GPLFM with Kalman filter is adopted as the GPLFM-based input estimation has desirable properties such as observability, stable inversion and resistance to low-frequency drifts in input and state estimation [125].

6.3 The Gaussian process latent force approach

In this section[1], a Gaussian process latent force approach is formulated to reconstruct the contact-point accelerations from the vehicle acceleration measurements with the knowledge of vehicle SSM. The Gaussian process latent force model (GPLFM) [125, 128, 129] acts as a hybrid *grey-box* model that augments the mechanistic model representing a physical system with data-driven non-parametric Gaussian processs (GPs); the GPs are used to represent the inputs exciting the physical system.

6.3.1 Construction of GPLFM in state space

GPs [130] are a popular class of stochastic processes that provide a paradigm for specifying probability distributions over functions, so each random draw from a GP is a function from a functional family defined by a covariance function. In this study, the Matérn family of covariance functions [131, 132] is chosen, and a GP with zero mean function and a stationary covariance function with smoothness parameter 5/2 is used to model the contact-point acceleration response. This is typically denoted as follows:

$$\ddot{u}_c(t) \sim \mathcal{GP}\left(0, \kappa_{5/2}(\tau; \boldsymbol{\theta}_{cf})\right) \tag{6.10}$$

where $\kappa_{5/2}(\tau; \boldsymbol{\theta}_{cf})$ represents a stationary Matérn covariance function governed by hyper-parameters $\boldsymbol{\theta}_{cf} = \{\varsigma, l\}$

$$\kappa_{5/2}(\tau) = \varsigma^2 \left(1 + \frac{\sqrt{5}\tau}{l} + \frac{5\tau^2}{3l^2}\right) \exp\left(-\frac{\sqrt{5}\tau}{l}\right) \tag{6.11}$$

[1]The material presented in this section was developed in collaboration with Dr. Chakraborty. For details, refer to the Statement of Contributions

Here, ς^2 and l are positive hyperparameters that denote signal variance and lengthscale respectively. This class of covariance functions is convenient for modelling narrowband process and is hence deemed suitable for this study. Several other forms of stationary covariance functions also exists, the details of which can be found in [130].

Hartikainen and Särkkä [133] showed that it is possible to convert a stationary covariance function into a LTI SSM following a spectral factorization. Following this, the Matérn covariance function in Eq.(6.11) can be converted to a GP SSM which outputs the desired $\ddot{u}_c(t)$:

$$\begin{aligned} \dot{\boldsymbol{z}}(t) &= \mathbf{F}_{cf}\boldsymbol{z}(t) + \mathbf{L}_{cf}w(t), \\ \ddot{u}_c(t) &= \mathbf{H}_{cf}\boldsymbol{z}(t), \end{aligned} \tag{6.12}$$

where $\boldsymbol{z}(t)$ represent the state vector of GP SSM and the parameter matrices are given as

$$\mathbf{F}_{cf} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a^3 & -3a^2 & -3a \end{bmatrix}, \ \mathbf{L}_{cf} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \ \mathbf{H}_{cf} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \tag{6.13}$$

In Eq.(6.12), $w(t)$ represents a scalar zero-mean stationary Gaussian white noise process with power spectral density $\sigma_w = 400\sqrt{5}\varsigma^2/\left(3l^5\right)$ and $a = \sqrt{5}/l$.

The GP SSM (in Eq.(6.12)) can be combined with the system SSM (in Eq.(6.8)) to yield an augmented SSM, which is referred to as the GPLFM

$$\begin{bmatrix} \dot{\boldsymbol{x}}_v(t) \\ \dot{\boldsymbol{z}}(t) \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{A}_{vc} & \mathbf{B}_{vc}\mathbf{H}_{cf} \\ \mathbf{0} & \mathbf{F}_{cf} \end{bmatrix}}_{\mathbf{F}_{ac}} \underbrace{\begin{bmatrix} \boldsymbol{x}_v(t) \\ \boldsymbol{z}(t) \end{bmatrix}}_{\boldsymbol{x}_c^a(t)} + \underbrace{\begin{bmatrix} \mathbf{0} \\ \mathbf{L}_{cf} \end{bmatrix}}_{\mathbf{L}_{ac}} w(t), \tag{6.14a}$$

$$y_v(t_k) = \underbrace{\begin{bmatrix} \mathbf{G}_v & \mathbf{0} \end{bmatrix}}_{\mathbf{G}_{ac}} \begin{bmatrix} \boldsymbol{x}_v(t_k) \\ \boldsymbol{z}(t_k) \end{bmatrix} + \widetilde{v}_k, \tag{6.14b}$$

where $\widetilde{v}_k$ represents measurement noise modelled as zero-mean Gaussian white noise with variance R. In shorthand notation, the GPLFM in Eq.(6.14) can be represented as

$$\begin{aligned} \dot{\boldsymbol{x}}_c^a(t) &= \mathbf{F}_{ac}\boldsymbol{x}_c^a(t) + \boldsymbol{w}(t) \\ y_v(t_k) &= \mathbf{G}_{ac}\boldsymbol{x}_c^a(t_k) + \widetilde{v}_k. \end{aligned} \tag{6.15}$$

where $\boldsymbol{w}(t) = \mathbf{L}_{cf} w(t)$. The state space formulation makes the GPLFM amenable for joint estimation of latent states and inputs via Kalman filtering.

6.3.2 Joint posterior inference of latent states and inputs

To be able to employ Kalman filter, the continuous-time form of the GPLFM in Eq.(6.15) is converted to discrete-time form:

$$\begin{aligned} \boldsymbol{x}^a_{k+1} &= \mathbf{F}_{ad} \boldsymbol{x}^a_k + \boldsymbol{w}_k, \\ y_{v,k} &= \mathbf{G}_{ad} \boldsymbol{x}^a_k + \widetilde{v}_k, \end{aligned} \tag{6.16}$$

where $y_{v,k} = y_v(t_k)$ and the state matrices $\mathbf{F}_{ad} = \exp(\mathbf{F}_{ac}\Delta t)$ and $\mathbf{G}_{ad} = \mathbf{G}_{ac}$. Furthermore, $\boldsymbol{w}_k$ is a zero-mean Gaussian white noise vector representing the discrete-time form of $\boldsymbol{w}(t)$ whose covariance is given by

$$\mathbf{Q}_d = \int_0^{\Delta t} \boldsymbol{\Psi}_a(\Delta t - \tau)\, \mathbf{L}_{cf}\, \sigma_w\, \mathbf{L}_{cf}^T\, \boldsymbol{\Psi}_a(\Delta t - \tau)^T d\tau, \tag{6.17}$$

where $\boldsymbol{\Psi}_a(\tau) = \exp(\mathbf{F}_{ac}\tau)$ is the matrix exponential of the state-transition matrix. The integral in Eq.(6.17) is solved using matrix fraction decomposition (see [134, 135] for implementation details).

Once the hyperparameters $\boldsymbol{\theta}_{cf}$ are determined, the matrices $\mathbf{F}_{ad}$, $\mathbf{G}_{ad}$ and $\mathbf{Q}_d$ can be determined. Thereafter, the posterior distribution of the vehicle states and the inputs can be estimated with a classical Kalman filter.

6.3.3 Extraction of contact-point response

The input sequence $\{\ddot{u}_{c,k}\}_{1:N}$ inferred from GPLFM-based Kalman filter correspond to the contact-point acceleration. An even better estimate of the contact-point acceleration can be obtained by using the estimated vehicle states, since the system states are often estimated with better accuracy than the input. The procedure to obtain the contact-point response from the estimated vehicle states is described as follows. First, the discrete-time form of the transition equation of Eq.(6.8) is written as

$$\boldsymbol{x}_{v,k+1} = \mathbf{A}_{vd} \boldsymbol{x}_{v,k} + \mathbf{B}_{vd} \ddot{u}_{c,k} \tag{6.18}$$

(6.19)

with the matrices $\mathbf{A}_{vd}$ and $\mathbf{B}_{vd}$ defined as

$$\mathbf{A}_{vd} = \exp\left(\mathbf{A}_{vc}\Delta t\right) \quad \mathbf{B}_{vd} = \left[\mathbf{A}_{vd} - \mathbf{I}\right]\mathbf{A}_{vc}^{-1}\mathbf{B}_{vc}.$$

Next, from the GPLFM inference, the time sequence of the estimated vehicle states $\{\hat{\boldsymbol{x}}_v\}_{1:N+1}$ are obtained. Using Eq.(6.18), the contact-point responses $\{\ddot{u}_{c,k}\}_{1:N}$ can be derived by solving a linear over-determined set of equations as follows:

$$\ddot{u}_{c,k} = \mathbf{B}_{vd}^{\dagger}\left(\hat{\boldsymbol{x}}_{v,k+1} - \mathbf{A}_{vd}\hat{\boldsymbol{x}}_{v,k}\right). \tag{6.20}$$

The term $\mathbf{B}_{vd}^{\dagger} = \left(\mathbf{B}_{vd}^{T}\mathbf{B}_{vd}\right)^{-1}\mathbf{B}_{vd}^{T}$ represents the Moore-Penrose pseudo-inverse for inverting non-square matrices.

6.3.4 Hyperparameter optimization for GP covariance function

The GPLFM needs the knowledge of the vehicle model parameter matrices and the covariance function hyperparameters. The knowledge of the vehicle state space matrices is already assumed to be known *a priori* (either from FE model or from system identification), and therefore the augmented state estimation results will depend only on the parameters of the chosen covariance functions modelling the latent input. In general, the hyperparameters are optimized based on the measurement sensor data. Typical hyperparameters include lengthscale and signal variance for a standard family of covariance functions. The hyperparameters can be estimated in different ways, including maximization of marginal likelihood [130], maximum a posteriori [136], and MCMC methods [137]. In this study, the optimized hyperparameters are obtained by maximizing the likelihood function based on the measurements. Maximum likelihood estimates of the hyperparameters (i.e. signal variance and lengthscale) of the covariance function can be obtained by minimizing the negative log-likelihood (or maximizing the log-likelihood) of the measurements as follows:

$$\hat{\boldsymbol{\theta}}_{cf} = \arg\min_{\boldsymbol{\theta}_{cf}} \left[\sum_{k=1}^{N}\left(\log\det\mathbf{S}_k + \boldsymbol{e}_k^{T}\mathbf{S}_k\boldsymbol{e}_k\right)\right] \tag{6.21}$$

The expressions for $\boldsymbol{e}_k$ and $\mathbf{S}_k$ are provided in Equations F.2a and F.2b respectively. The minimization can be done using optimization tools such as MATLAB's built-in functions `fminunc` or `fmincon`. It is noteworthy to mention that maximum likelihood optimization is susceptible to local minima; to avoid this one may need to start the optimization from different initial points.

An algorithm depicting the steps involved in the proposed methodology is shown in Algorithm 7.

Algorithm 7 Contact-point response reconstruction with GPLFM

Input: $\mathbf{A}_{vc}$, $\mathbf{B}_{vc}$, $\mathbf{G}_v$ and vehicle acceleration time history $\left\{\ddot{d}_{v,k}\right\}_{1:N}$

1. Choose a Matérn covariance function with smoothness parameter $\frac{5}{2}$ and initialize the hyperparameters $\boldsymbol{\theta}_{cf}$ (refer Eq.(6.11))
2. Convert the covariance function into equivalent continuous-time GP SSM and obtain $\mathbf{F}_{cf}$, $\mathbf{L}_{cf}$, $\mathbf{H}_{cf}$ and σ_w (refer Eq.(6.12) and (6.13)).
3. Construct the continuous-time augmented SSM matrices $\mathbf{F}_{ac}$ and $\mathbf{G}_{ac}$, as shown in Eq.(6.14)
4. Select initial state and noise covariances for Kalman filter recursions
5. Compute optimum hyperparameter $\hat{\boldsymbol{\theta}}_{cf}$ by maximum likelihood optimization (refer Section 6.3.4)
6. Use $\hat{\boldsymbol{\theta}}_{cf}$ to compute $\mathbf{F}_{ad}$, $\mathbf{G}_{ad}$, $\mathbf{Q}_d$ as in Eq.(6.16) and Eq.(6.17)
7. Use $\mathbf{F}_{ad}$, $\mathbf{G}_{ad}$, $\mathbf{Q}_d$ for jointly estimating the input and the vehicle states with Kalman filter (see Appendix F)
8. Use Eq.(6.20) and the estimated vehicle states to estimate the contact-point responses.

Output: Time sequence of contact-point acceleration $\{\ddot{u}_{c,k}\}_{1:N}$

6.4 Numerical study

A numerical bridge example (used by Yang and Chen [42]) is adopted to evaluate the performance of the proposed approach. Eq.(6.6) is used for forward simulation of the vehicle and bridge contact-point responses with the following parameters:

- Bridge properties: $L = 30\text{m}$, $\rho_b = 10700\text{kg/m}$, $E = 29.43\text{GPa}$, $I = 0.494\text{m}^4$.
- The first four modal frequencies of the bridge are 2.03, 8.14, 18.31 and 32.55Hz, respectively. Modal damping ratio is equal to 1% of the critical damping for the first four bridge modes.
- The amplitude of bridge support excitation at both left and right supports is 0.02m/s^2.
- The properties of the carrier-vehicle are chosen as $m_v = 100\text{kg}$, $k_v = 150\text{kN/m}$. The modal frequency of the carrier-vehicle is 6.16Hz. Furthermore, a modal damping ratio equal to 10% of its critical damping is assumed.
- The moving speed of the carrier-vehicle is $v = 2\text{m/s}$. Mobile sensor data is collected at a sampling rate of 100Hz. The data is collected for the time duration in which the carrier-vehicle crosses the entire bridge span once.

The (noise-free) dynamic response of the vehicle and the contact-point response of the bridge is shown in Figure 6.2. From the power spectral density plot in Figure 6.3, it can be seen that there is a noticeable presence of the vehicle dynamics; there is a significant peak at the vehicle modal frequency of 6.16Hz and the energy in the third and fourth modes of the bridge are suppressed to a considerable extent in the vehicle response. This confirms that the presence of vehicle dynamics alters the vehicle response from the true contact-point response which could prevent accurate identification of the (higher) modal features.

The reason for suppression of the third and fourth bridge modes can be explained by observing the dynamic amplification factor for the carrier-vehicle, illustrated in Figure 6.4. The dynamic amplification factor provides an idea of how the vehicle dynamics modifies the input. The shaded region in the plot encompasses the frequency spectrum where the amplification factor is below 1 and approaches zero asymptotically. This implies that any bridge mode lying in the shaded region would get suppressed significantly. Since the third and fourth bridge modes lie in the shaded region their prominence is significantly reduced in the measured vehicle response.

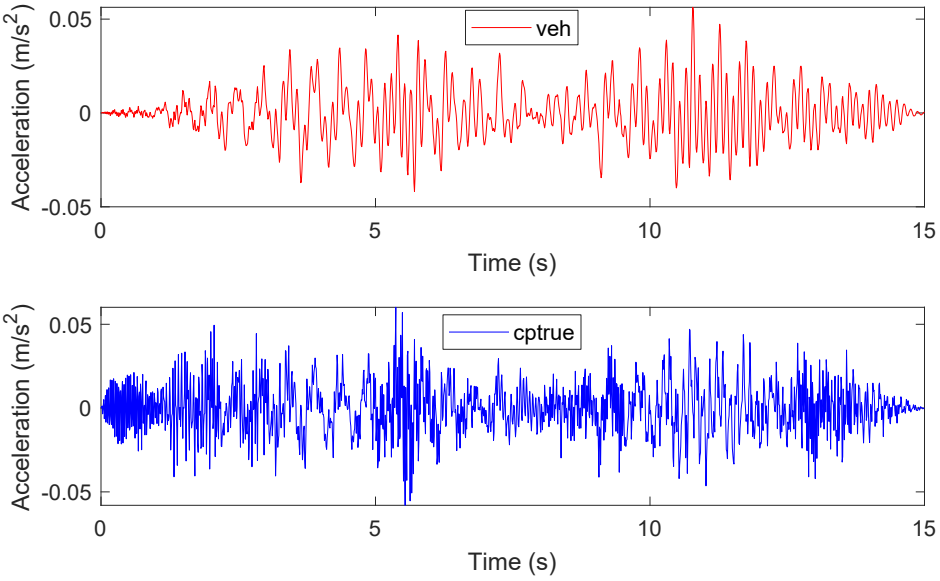


Figure 6.2: Vehicle dynamic response (top) and the true contact-point response (bottom).

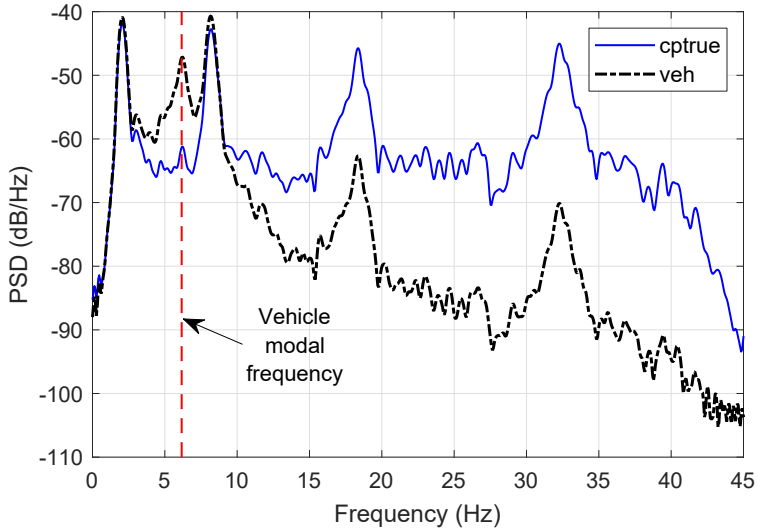


Figure 6.3: PSDs of vehicle response and true contact-point response.

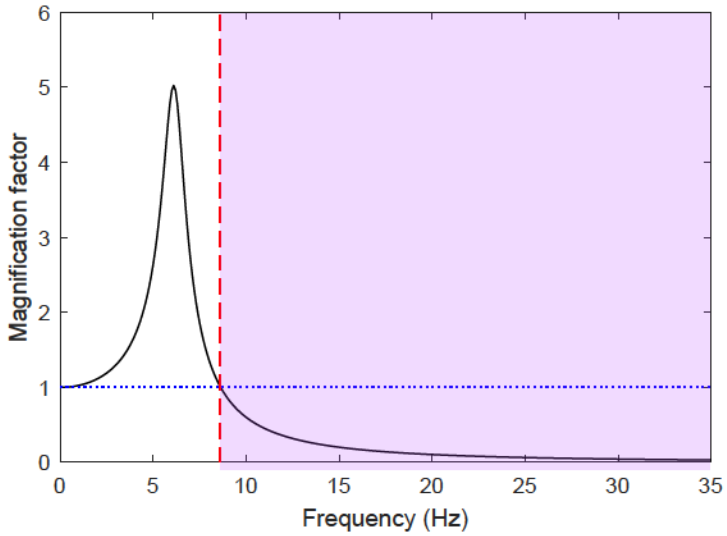


Figure 6.4: Plot of dynamic amplification factor of the carrier-vehicle.

Figure 6.5 shows a comparison of the true and reconstructed contact-point responses; the reconstruction is facilitated by the use of GPLFM – with a Matèrn covariance function as mentioned in Section 6.3 – on noise-free vehicle responses. The reconstructed contact-point responses are consistent with the true contact-point to an acceptable degree. A constant delay in the order of 0.01s is noticed in the reconstructed responses compared to the true values, however the delay being constant is unlikely to pose any problems in output-only modal identification.

Figure 6.6 plots the PSDs of the true and reconstructed contact-point responses. It can be seen that the reconstructed contact-point response has not only suppressed the effect of vehicle modal frequencies but also enhanced the prominence of the third and fourth modes considerably. Thus, obtaining contact-point responses is deemed useful in improving the visibility of (higher) modal features of the bridge.

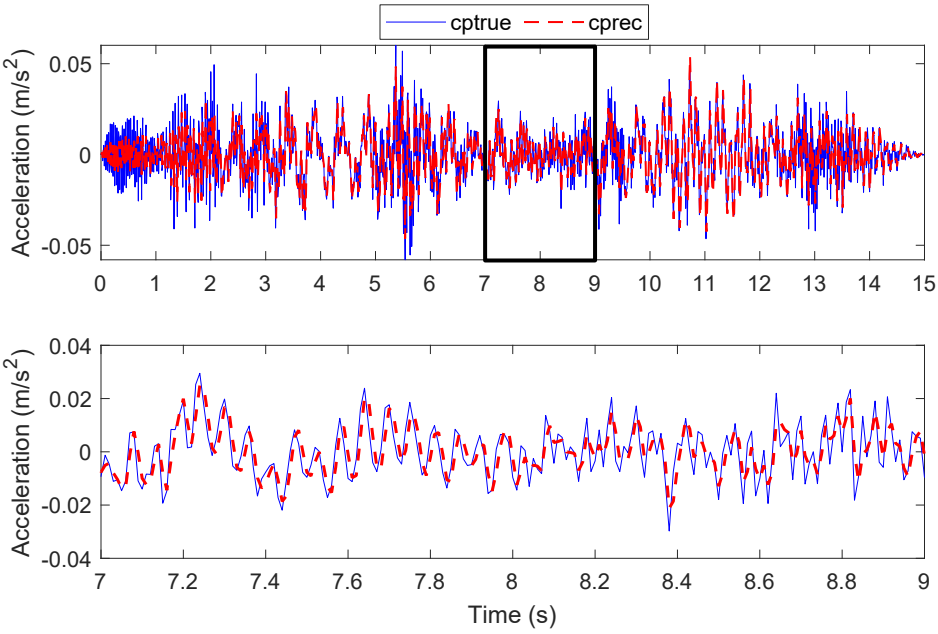


Figure 6.5: Comparison of true and reconstructed contact-point responses.

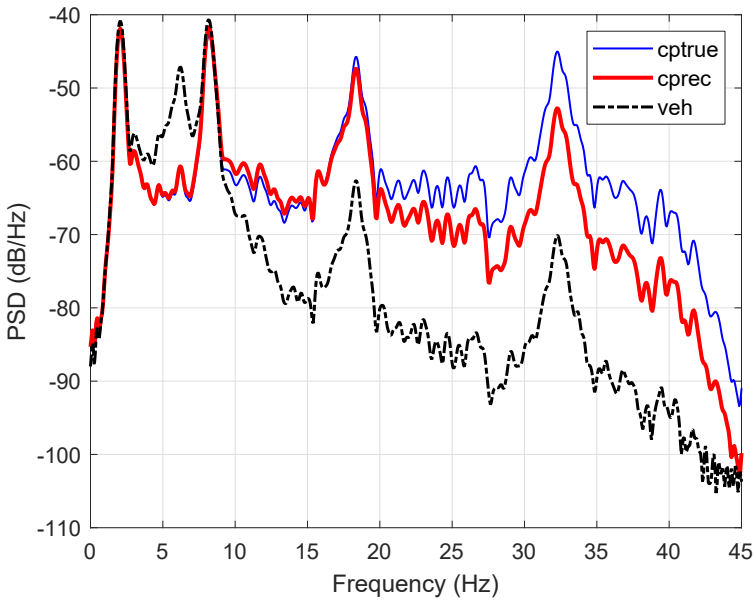


Figure 6.6: PSDs of true and reconstructed contact-point responses.

6.4.1 Effect of measurement noise

In practice, the measured responses will contain some level of sensor noise which will distort the true sensor responses. Keeping all vehicle parameters unchanged, two sensor noise scenarios of (i) 5% and (ii) 10% are considered. The noisy sensor signal is numerically calculated as

$$y_v^{noisy} = y_v^{true} + \frac{\texttt{noise}\%}{100} \times \texttt{RMS}\left(y_v^{true}\right) \times \texttt{WGN} \tag{6.22}$$

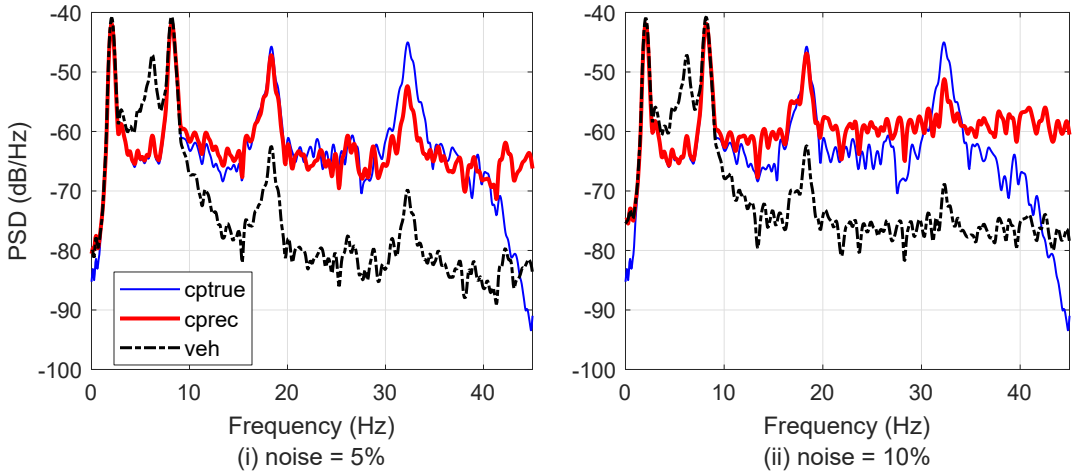
where, y_v^{true} is the true calculated vehicle acceleration, $\texttt{noise}\%$ is the noise percentage, $\texttt{RMS}(y_v^{true})$ is the RMS of y_v^{true} and $\texttt{WGN}$ is a vector of standard Gaussian white noise. For the two noise scenarios, comparisons of the time-series and the PSDs of the true and reconstruction contact-point responses are shown in Figure 6.7. As expected, higher noise leads to poorer reconstruction of contact-point responses. In particular, it is noticed that the fourth mode (which is farthest from the vehicle modal frequency) undergoes some distortion due to sensor noise which indicates that identification of higher modes may be difficult under high sensor noise.

6.4.2 Effect of vehicle damping

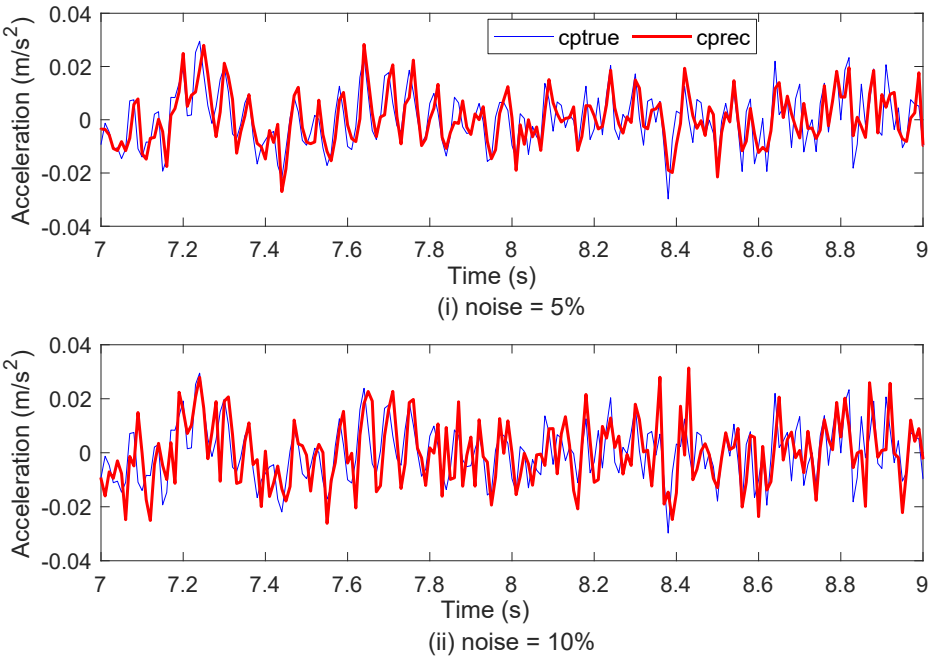
Assuming noise-free sensor measurements and holding all other vehicular parameters constant, the effect of vehicle damping on the reconstructed contact-point response is investigated using two different cases of vehicle modal damping ratios of (i) 5% and (ii) 20%, illustrated in Figure 6.8. The two cases correspond to a low (5%) and a high (20%) modal damping for the sensor-instrumented vehicle. Observing the PSD in Figure 6.8a, it is seen that the energies in the third and the fourth modes improve with more damping. Further, inspecting Figure 6.8b it is found that lower vehicle damping introduces more delay in the reconstructed contact-point responses compared to higher vehicle damping.

6.4.3 Effect of vehicle speed

Assuming a smooth bridge deck, the effect of vehicle speed on the vehicle response is considered; all other vehicle parameters are held constant. Vehicle responses are simulated

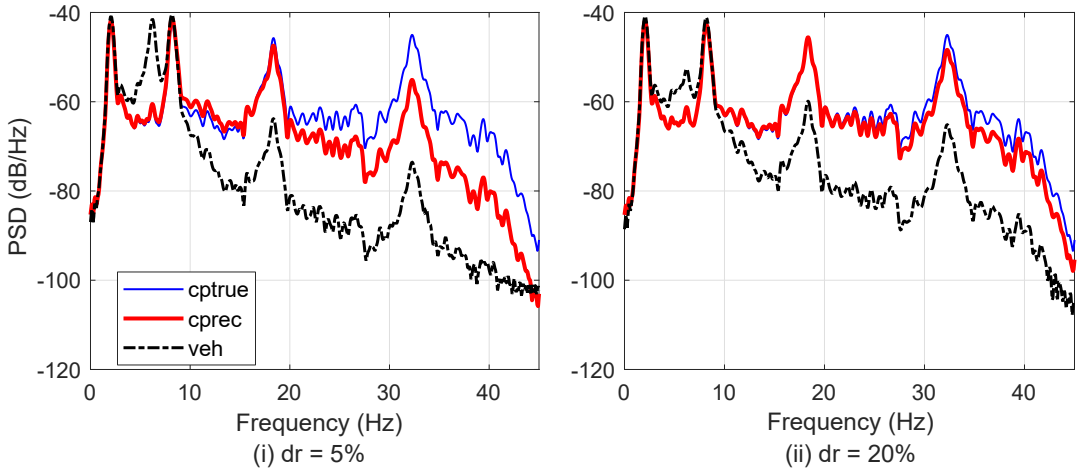


(a) PSD of reconstructed contact-point responses

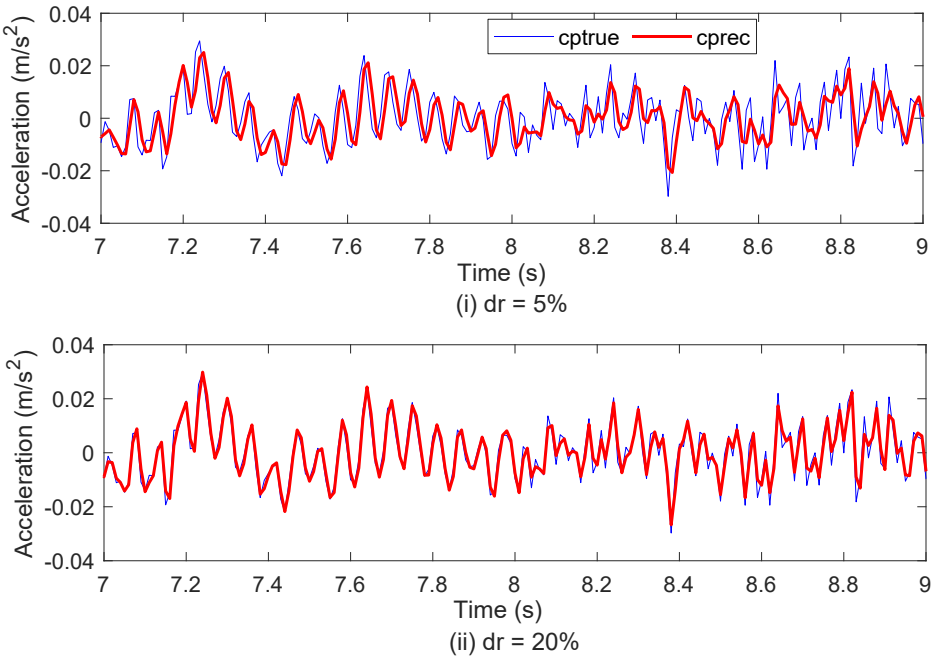


(b) Reconstructed contact-point responses (zoomed in between 7-9 s)

Figure 6.7: Effect of sensor noise with noise percentages 5% and 10%.



(a) PSD of reconstructed contact-point responses



(b) Reconstructed contact-point responses (zoomed in between 7-9 s)

Figure 6.8: Effect of vehicle damping with vehicle damping ratios of 5% and 20%.

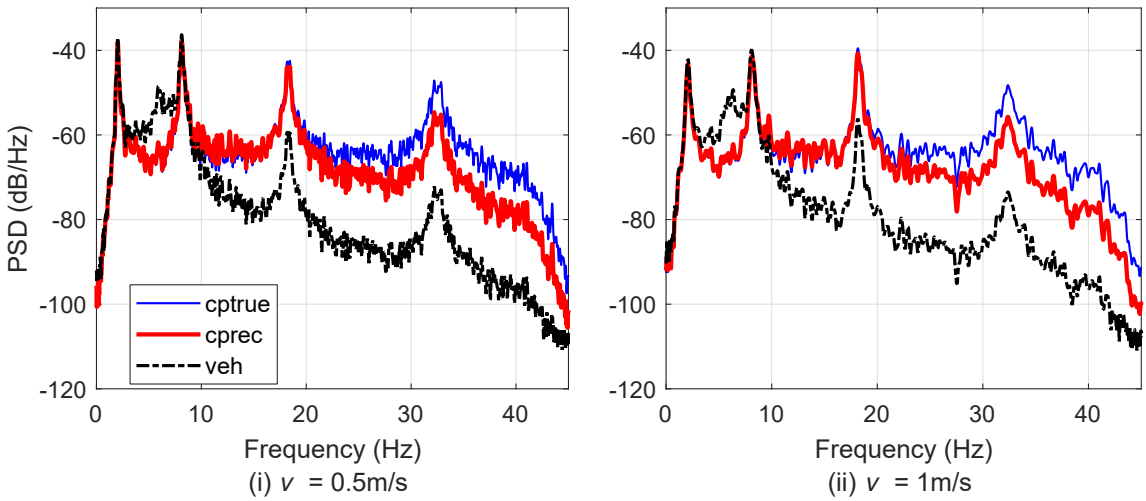


Figure 6.9: Effect of vehicle speed on the PSD of reconstructed contact-point responses.

using two cases of vehicle speeds (i) 0.5m/s and (ii) 1m/s. Note the corresponding time duration of simulation for crossing the bridge span once for the two cases are 60s and 30s, respectively. The PSDs of the reconstructed contact-point responses for the two cases are shown in Figure 6.9. The speed of the vehicle is found to have not much effect on the vehicle response and the reconstructed contact-point response, although, in cases where the bridge deck is rough lower vehicle speeds are reported to be more effective for extracting modal features [38].

6.4.4 Effect of vehicle mass

For this, two cases of vehicle mass are considered: (i) 50kg and (ii) 150kg. The vehicle modal frequencies corresponding to the two cases are 8.72Hz and 5.03Hz respectively. The PSDs of the reconstructed contact-point responses for the two cases are shown in Figure 6.10. For the first case – with vehicle mass equal to 50kg – the vehicle modal frequency overlaps closely with the second bridge modal frequency, and hence the second bridge frequency gets amplified. This causes a slight change in the shape of the PSD at the second modal peak, which may lead to errors in identification of second modal bridge damping from vehicle response. It is noticed, however, that the GPLFM recovered contact-point

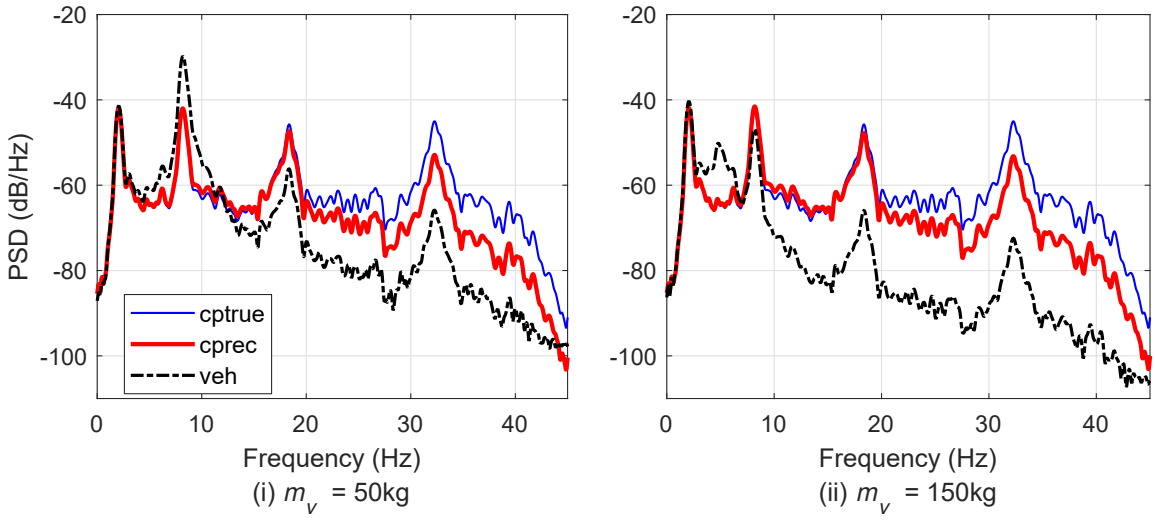


Figure 6.10: Effect of vehicle mass on the PSD of reconstructed contact-point responses.

response produces a good match with the true PSD at the second bridge mode. For a higher mass of 150 kg, the vehicle modal frequency becomes even smaller furthering the suppression of third and fourth modes, nevertheless, the recovered contact-point response is still able to enhance the suppressed modes to a considerable extent.

6.5 Summary

In this chapter, the idea of recovering contact-point response from the carrier-vehicle response is proposed using an input reconstruction technique. The contact-point response is shown to act as an base-excited input to the moving carrier-vehicle, and is estimated using a GPLFM-based Kalman filtering approach, given the knowledge of vehicle dynamical characteristics. The estimated contact-point responses can be used for modal identification using the techniques presented in the previous chapters.

The GPLFM-based Kalman filtering methodology presented is suitable for any bridge-vehicle interaction problem that includes the effect of damping of the bridge and the vehicle, and is also generalizable to multi-DoF models of carrier-vehicles. Using several numerical case studies, it is shown that the recovered contact-point responses can serve to not only

cancel out the effect of vehicle modal frequency but also enhance the suppressed higher bridge modes.

Chapter 7

Concluding Remarks

7.1 Summary of contributions

In this thesis, several theoretical and algorithmic challenges associated with modal parameter identification using mobile sensors are addressed. Two broad strategies of mobile sensing, re-configurable mobile sensing and in-motion mobile sensing, are considered. Specific challenges faced with processing data from the two mobile sensing strategies are discussed, and numerical algorithms facilitating modal parameter identification (and uncertainty quantification) with the mobile sensor data are developed. The following are the main contributions of this thesis:

1. The feasibility of identifying high resolution mass-normalized mode shapes using a single *re-configurable* mobile actuator-sensor pair is investigated. The single mobile actuator-and-sensor corresponds to a case of minimal instrumentation required for modal identification. A high resolution mode shape identification methodology, based on pseudo-modal responses, has been proposed using two different schemes of re-configurable mobile actuation-sensing (i.e., collocated and non-collocated actuation-sensing). The performance of the proposed approach has been studied using numerical simulations as well as laboratory scale experiments. The major findings from this study are as follows:

(a) When the mobile actuator and mobile sensor form two separate units and are not always collocated, the mode shape components can be identified uniquely.

(b) When the mobile actuator and sensor are collocated (i.e, when they act as a single unit), only mode shapes squares can be extracted and signs of mode shapes are needed to be assigned using some prior knowledge (e.g. from an approximate FE model).

(c) In practice, due to presence of noise, some component(s) of the mode shape squares near the nodes of mode shapes may turn out to be negative; in such cases additional tests with different points of collocation may be required.

(d) Based on the numerical and experimental results, it is found that the proposed methodology may be considered as a feasible alternative to traditional schemes employing many static sensors for identifying the high resolution mode shapes of a structure.

2. A Bayesian posterior inference framework is established for output-only modal analysis using *in-motion* mobile sensors, following a time-domain state space modelling approach. Two perspectives of processing in-motion mobile sensor data has been considered: missing (incomplete) data and stacked (complete) data. Three computational algorithms, Expectation Maximization (EM), Variational Bayes (VB) and Gibbs Sampler (GS), are derived estimate modal parameters from mobile sensor data. Numerical and bench-scale experimental studies have been used to validate the respective approaches.

   For identification with the missing data approach, the mobile sensor data is treated as equivalent static sensor data with missing entries – at locations which do not coincide with the mobile sensor paths. The findings from the modal parameter estimation using the missing data approach are as follows:

   (a) The GS-based multiple imputation procedure for parameter estimation provides reliable posterior distributions of modal parameters but suffer from heavy computational burden.

(b) Both EM and VB are computationally very efficient compared to GS; EM provides point estimates only whereas VB provides posterior distributions of modal parameters.

(c) The point estimates of modal parameters from EM are similar to the corresponding mean estimates from VB.

(d) The VB and GS mean estimates of modal parameters are quite consistent but VB underestimates the uncertainty in the estimated modal parameters as compared to that from GS. In presence of large proportions of missing data ($> 50\%$), the underestimation in uncertainty for modal frequencies and damping ratios is found to be in range of 5-25% while that for mode shapes can be between 40-60%.

(e) The proportion of missing entries is a function of number of sensing nodes and the number of employed mobile sensors; it increases with the number of sensing nodes and decreases with the number of mobile sensors. The GS algorithm is found superior to VB in capturing the uncertainty of the modal parameters as a function of missing data.

For identification with the stacked data approach, a time-varying Modal State Model (MSM) is derived. The MSM embeds a sinc function-based deterministic interpolator that converts mobile sensor data collected at different locations of a structure to equivalent static sensor data at some user-chosen virtual static locations (VSLs); the mode shape ordinates are estimated only at these VSLs. The findings from the modal parameter estimation using the stacked data approach are as follows:

(a) The posterior means of modal parameters are found to be more accurate for EM and VB compared to GS. However, the uncertainties in the modal parameters are underestimated by VB compared to GS. In particular for mode shapes, the uncertainty in scenarios with large number of VSLs and a few mobile sensors can be severely underestimated (by around 40-70%).

(b) The number of VSLs governs the size of the MSM and the resolution of estimates mode shapes. More VSLs imply larger size of the MSM and higher resolution of the estimated mode shapes.

(c) The computational time required for posterior inference depends on the size of MSM through the number of the VSLs. As the number of VSLs grows, the inference procedure takes longer time.

(d) The computational time required by GS is quite large compared to that of VB and EM i.e., the GS takes on an average eight times longer than EM or VB. For cases with large number of VSLs, the GS could take very long time and may be impractical to use.

(e) For a given dataset, increasing the number of VSLs increases the resolution of the estimated mode shapes, however also increases the uncertainty in the estimated mode shape ordinates.

3. An input identification framework is established to recover contact-point responses from responses captured by sensors on top of a carrier-vehicle, given the knowledge of dynamic characteristics of the carrier-vehicle. The recovery of the contact-point responses is facilitated by employing a Gaussian process latent force model (GPLFM) with Kalman filter. The main findings of this study are as follows:

   (a) The contact-point responses act as base-excited ground input to the moving vehicle(s).

   (b) The carrier-vehicle responses gets altered from the true contact-point responses due to vehicle dynamics – the higher modes of the bridge gets suppressed and the fundamental vehicle modal frequency gets introduced in the vehicle responses.

   (c) The GPLFM with Kalman filtering provides a generalizable framework for recovering the contact-point responses.

   (d) Based on a series of numerical simulations of a linear single DoF vehicle moving over a Euler-Bernoulli beam, it is found that the contact-point responses are more effective for modal parameter identification than vehicle responses as they cancel out the vehicle frequency and enhance the presence of higher bridge modes.

7.2 Directions for future study

Based on the research work proposed in this thesis, a couple of research directions can be pursued for extending the framework:

1. **Improvement of VB covariance underestimation**: Although VB provides a fast approach to posterior inference, the posterior covariances tends to be underestimated in comparison to GS, sometimes drastically. Recent research on using full-rank approach [138] and linear response technique [139] with VB address this issue and provide a correction to the covariance underestimation. Incorporation of these corrections into the current framework may improve the covariance estimation from VB while still preserving its computational efficiency.

2. **Extension to direct structural parameter estimation**: The paradigm for modal parameter estimation using in-motion mobile sensors can be extended to direct estimation of a structural FE model parameters such as elemental stiffnesses or masses from in-motion mobile sensor data. However, the convenience of closed-form update expressions may be unlikely in such cases; the GS will have to replaced by MH sampler and the updating step of VB will involve numerical optimization.

3. **Separation of road roughness for contact-point response estimation**: The road roughness profile of a pavement can present difficulties in accurate identification of contact-point responses, particularly when the speed of the mobile sensors is high. Extension of the GPLFM framework to account for road roughness has not been considered in the thesis and can be taken up as a future work.

References

[1] ASCE report card for America's infrastructure, 2017. URL https://www.infrastructurereportcard.org/.

[2] Canadian Infrastructure Report Card, 2016. URL http://canadianinfrastructure.ca/en/index.html.

[3] B. A. Zárate and J. M. Caicedo. Finite element model updating: multiple alternatives. *Engineering Structures*, 30(12):3724–3730, 2008.

[4] H. Schlune, M. Plos, and K. Gylltoft. Improved bridge evaluation through finite element model updating using static and dynamic measurements. *Engineering Structures*, 31(7):1477–1485, 7 2009.

[5] M. Alyami, M. Rouainia, and S. M. Wilkinson. Numerical analysis of deformation behaviour of quay walls under earthquake loading. *Soil Dynamics and Earthquake Engineering*, 29(3):525–536, 3 2009.

[6] J. Y. Kim, E. Yu, D. Y. Kim, and S.-D. Kim. Calibration of analytical models to assess wind-induced acceleration responses of tall buildings in serviceability level. *Engineering Structures*, 31(9):2086–2096, 9 2009.

[7] V. Racic, A. Pavic, and J. M. W. Brownjohn. Experimental identification and analytical modelling of human walking forces: Literature review. *Journal of Sound and Vibration*, 326(1-2):1–49, 9 2009.

[8] N. Galati, A. Nanni, J. G. Tumialan, and P. H. Ziehl. In-situ evaluation of two concrete slab systems I: Load determination and loading procedure. *Journal of Performance of Constructed Facilities*, 22(4):207–216, 8 2008.

[9] A. Teughels and G. De Roeck. Structural damage identification of the highway bridge Z24 by FE model updating. *Journal of Sound and Vibration*, 278(3):589–610, 12 2004.

[10] E. H. Stehmeyer and D. C. Rizos. Considering dynamic soil structure interaction effects on seismic isolation retrofit efficiency and the importance of natural frequency ratio. *Soil Dynamics and Earthquake Engineering*, 28(6):468–479, 6 2008.

[11] C.-P. Fritzen and P. Kraemer. Self-diagnosis of smart structures based on dynamical properties. *Mechanical Systems and Signal Processing*, 23(6):1830–1845, 8 2009.

[12] M. Friswell and J. E. Mottershead. *Finite element model updating in structural dynamics*, volume 38. Springer Science and Business Media, 2013.

[13] J. P. Lynch and K. J. Loh. A summary review of wireless sensors and sensor networks for structural health monitoring. *The Shock and Vibration Digest*, 38(2):91–129, 3 2006.

[14] S. N. Pakzad, G. L. Fenves, S. Kim, and D. E. Culler. Design and implementation of scalable wireless sensor network for structural monitoring. *Journal of Infrastructure Systems*, 14(1):89–101, 3 2008.

[15] J. A. Rice, K. Mechitov, S.-H. Sim, T. Nagayama, S. Jang, R. Kim, B. F. Spencer, G. Agha, and Y. Fujino. Flexible smart sensor framework for autonomous structural health monitoring. *Smart Structures and Systems*, 5(6):423–438, 2010.

[16] D. Zhu, X. Yi, Y. Wang, K.-M. Lee, and J. Guo. A mobile sensing system for structural health monitoring: design and validation. *Smart Materials and Structures*, 19(5):055011, 5 2010.

[17] D. Zhu, J. Guo, C. Cho, Y. Wang, and K.-M. Lee. Wireless mobile sensor network for the system identification of a space frame bridge. *IEEE/ASME Transactions on Mechatronics*, 17(3):499–507, 6 2012.

[18] J. M. Caicedo and J. Marulanda. Fast mode identification technique for online monitoring. *Structural Control and Health Monitoring*, 18:416–429, 2011.

[19] J. Marulanda, J. M. Caicedo, and P. Thomson. Modal identification using mobile sensors under ambient excitation. *Journal of Computing in Civil Engineering*, 31(2): 04016051, 3 2017.

[20] T. J. Matarazzo and S. N. Pakzad. Structural identification for mobile sensing with missing observations. *Journal of Engineering Mechanics*, 142(5):04016021, 5 2016.

[21] T. J. Matarazzo and S. N. Pakzad. Scalable structural modal identification using dynamic sensor network data with STRIDEX. *Computer-Aided Civil and Infrastructure Engineering*, 33(1):4–20, 1 2018.

[22] K. Goorts, S. Phillips, A. Ashasi-Sorkhabi, and S. Narasimhan. Structural control using a deployable autonomous control system. *International Journal of Intelligent Robotics and Applications*, 1(3):306–326, 9 2017.

[23] R. Brincker, L. Zhang, and P. Andersen. Modal identification of output-only systems using frequency domain decomposition. *Smart Materials and Structures*, 10(3):441–445, 6 2001.

[24] B. Peeters and G. De Roeck. Stochastic system identification for operational modal analysis: A review. *Journal of Dynamic Systems, Measurement, and Control*, (4): 659, 2001.

[25] J.-N. Juang. *Applied system identification*. Prentice Hall, 1994.

[26] J. M. W. Brownjohn. Structural health monitoring of civil infrastructure. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851):589–622, 2 2007.

[27] A. de Stefano. Structural identification and health monitoring on the historical architectural heritage. *Key Engineering Materials*, 347:37–54, 9 2007.

[28] C. R. Farrar, S. W. Doebling, and D. A. Nix. Vibrationbased structural damage identification. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 359(1778):131–149, 1 2001.

[29] T. Pothisiri and K. D. Hjelmstad. Structural Damage Detection and Assessment from Modal Response. *Journal of Engineering Mechanics*, 129(2):135–145, 2 2003.

[30] J. Gang, M. K. Sain, and B. F. Spencer. Frequency domain system identification for controlled civil engineering structures. *IEEE Transactions on Control Systems Technology*, 13(6):1055–1062, 11 2005.

[31] V. Gattulli, M. Lepidi, and F. Potenza. Seismic protection of frame structures via semi-active control: modeling and implementation issues. *Earthquake Engineering and Engineering Vibration*, 8(4):627–645, 12 2009.

[32] P. Van Overschee and B. L. De Moor. *Subspace identification for linear systems: Theory-Implementation -Applications*. Springer Science and Business Media, 2012.

[33] R. D. Nayeri, F. Tasbihgoo, M. Wahbeh, J. P. Caffrey, S. F. Masri, J. P. Conte, and A. Elgamal. Study of time-domain techniques for modal parameter identification of a long suspension bridge with dense sensor arrays. *Journal of Engineering Mechanics*, 135(7):669–683, 7 2009.

[34] M. Meo and G. Zumpano. On the optimal sensor placement techniques for a bridge structure. *Engineering Structures*, 27(10):1488–1497, 8 2005.

[35] M. Chang and S. N. Pakzad. Optimal sensor placement for modal identification of bridge systems considering number of sensing nodes. *Journal of Bridge Engineering*, 19(6):04014019, 6 2014.

[36] G. H. James, T. G. Carne, and J. P. Lauffer. The natural excitation technique (NExT) for modal parameter extraction from operating structures. *Modal Analysis-*

the International Journal of Analytical and Experimental Modal Analysis, 10(4):260, 1995.

[37] A. Malekjafarian, P. J. McGetrick, and E. J. OBrien. A review of indirect bridge monitoring using passing vehicles. *Shock and Vibration*, 2015, 2015.

[38] Y. B. Yang and J. P. Yang. State-of-the-art review on modal identification and damage detection of bridges by moving test vehicles. *International Journal of Structural Stability and Dynamics*, 18(02):1850025, 2 2018.

[39] Y. B. Yang, C. W. Lin, and J. D. Yau. Extracting bridge frequencies from the dynamic response of a passing vehicle. *Journal of Sound and Vibration*, 272(3-5): 471–493, 2004.

[40] C. W. Lin and Y. B. Yang. Use of a passing vehicle to scan the fundamental bridge frequencies: An experimental verification. *Engineering Structures*, 27(13):1865–1878, 2005.

[41] Y. B. Yang and K. C. Chang. Extraction of bridge frequencies from the dynamic response of a passing vehicle enhanced by the emd technique. *Journal of sound and vibration*, 322(4-5):718–739, 2009.

[42] Y. B. Yang and W.-F. Chen. Extraction of bridge frequencies from a moving test vehicle by stochastic subspace identification. *Journal of Bridge Engineering*, 21(3): 04015053, 2015.

[43] A. González, E. J OBrien, and P. J. McGetrick. Identification of damping in a bridge using a moving instrumented vehicle. *Journal of Sound and Vibration*, 331 (18):4115–4131, 2012.

[44] Y. B. Yang, Y. C. Li, and K. C. Chang. Constructing the mode shapes of a bridge from a passing vehicle: a theoretical study. *Smart Structures and Systems*, 13(5): 797–819, 2014.

[45] A. Malekjafarian and E. J. OBrien. Identification of bridge mode shapes using short time frequency domain decomposition of the responses measured in a passing vehicle. *Engineering Structures*, 81:386–397, 2014.

[46] T. J. Matarazzo and S. N. Pakzad. Truncated physical model for dynamic sensor networks with applications in high-resolution mobile sensing and BIGDATA. *Journal of Engineering Mechanics*, 142(5):04016019, 5 2016.

[47] M. Horner, K. Koser, K. Korneva, and T.J. Matarazzo. A wireless mobile sensor platform for structural health monitoring. In *Proceedings of the Joint 6th International Conference on Advances in Experimental Structural Engineering and 11th International Workshop on Advanced Smart Materials and Smart Structures Technology*, pages 1–7, 2015.

[48] Y. B. Yang, B. Zhang, Y. Qian, and Y. Wu. Contact-point response for modal identification of bridges by a moving test vehicle. *International Journal of Structural Stability and Dynamics*, 18(05):1850073, 2018.

[49] Armen Der Kiureghian and Ove Ditlevsen. Aleatory or epistemic? does it matter? *Structural Safety*, 31(2):105–112, 2009.

[50] J. L. Beck and L. S. Katafygiotis. Updating models and their uncertainties. i: Bayesian statistical framework. *Journal of Engineering Mechanics*, 124(4):455–461, 1998.

[51] K.-V. Yuen and L. S. Katafygiotis. Bayesian time–domain approach for modal updating using ambient data. *Probabilistic Engineering Mechanics*, 16(3):219–231, 2001.

[52] L. S. Katafygiotis and K.-V. Yuen. Bayesian spectral density approach for modal updating using ambient data. *Earthquake engineering and structural dynamics*, 30 (8):1103–1123, 2001.

[53] S.-K. Au. Fast Bayesian FFT method for ambient modal identification with separated modes. *Journal of Engineering Mechanics*, 137(3):214–226, 2011.

[54] W.-J. Yan and L. S. Katafygiotis. A two-stage fast Bayesian spectral density approach for ambient modal analysis. Part I: posterior most probable value and uncertainty. *Mechanical Systems and Signal Processing*, 54:139–155, 2015.

[55] S.-K. Au. Fast bayesian ambient modal identification in the frequency domain, part i: Posterior most probable value. *Mechanical Systems and Signal Processing*, 26: 60–75, 2012.

[56] S.-K. Au. Fast Bayesian ambient modal identification in the frequency domain, Part II: posterior uncertainty. *Mechanical Systems and Signal Processing*, 26:76–90, 2012.

[57] B. Li and A. Der Kiureghian. Operational modal identification using variational Bayes. *Mechanical Systems and Signal Processing*, 88:377–398, 5 2017.

[58] B. Li, A. Der Kiureghian, and S.-K. Au. A Gibbs sampling algorithm for structural modal identification under seismic excitation. *Earthquake Engineering and Structural Dynamics*, 47(14):2735–2755, 2018.

[59] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 3rd edition, 2013.

[60] R. J. A. Little and D. B. Rubin. *Statistical analysis with missing data*. John Wiley and Sons, Inc., 2nd edition, 2002.

[61] L. Tierney and J. B. Kadane. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82–86, 3 1986.

[62] J. K. Ghosh, M. Delampady, and T. Samanta. *An introduction to Bayesian analysis: Theory and methods*. Springer Science and Business Media, 2007.

[63] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.

[64] R. H. Shumway and D. S. Stoffer. *Time series analysis and its applications: with R examples*. Springer, 2017.

[65] S. Särkkä. *Bayesian filtering and smoothing*, volume 3. Cambridge University Press, 2013.

[66] C. F. J. Wu. On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11(1):95–103, 3 1983.

[67] R. Sundberg. Maximum Likelihood Theory for Incomplete Data from an Exponential Family, 1974.

[68] F. J. Cara, J. Carpio, J. Juan, and Enrique Alarcón. An approach to operational modal analysis using the expectation maximization algorithm. *Mechanical Systems and Signal Processing*, 31:109–129, 8 2012.

[69] F. J. Cara, J. Juan, and E. Alarcón. Estimating the modal parameters from multiple measurement setups using a joint state space model. *Mechanical Systems and Signal Processing*, 43(1-2):171–191, 2 2014.

[70] T. J. Matarazzo and S. N. Pakzad. STRIDE for structural identification using expectation maximization: Iterative output-only method for modal identification. *Journal of Engineering Mechanics*, 142(4), 2016.

[71] X.-L. Meng and D. B. Rubin. Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm. *Journal of the American Statistical Association*, 86 (416):899, 12 1991.

[72] K. B. Petersen, Ole Winther, and L. K. Hansen. On the slow convergence of EM and VBEM in low-noise linear models. *Neural Computation*, 17(9):1921–1926, 9 2005.

[73] G. McLachlan and T. Krishnan. *The EM algorithm and extensions*, volume 382. John Wiley and Sons, 2007. ISBN 0470191600.

[74] M. J. Beal. *Variational algorithms for approximate Bayesian inference*. PhD thesis, University College London, 2003.

[75] M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(12):1–305, 2008.

[76] D. J. C. Mackay. *Information theory, inference and learning algorithms*. Cambridge University Press, Cambridge, UK, 2003.

[77] S. Kullback and R. A. Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

[78] H. Attias. A variational Bayesian framework for graphical models. In *Advances in neural information processing systems*, pages 209–215, 2000.

[79] Z. Ghahramani and M. J. Beal. Propagation algorithms for variational Bayesian learning. In *Advances in neural information processing systems*, pages 507–513, 2001.

[80] F. Guo, X. Wang, K. Fan, T. Broderick, and D. B. Dunson. Boosting variational inference. *arXiv preprint arXiv:1611.05559*, 2016.

[81] A. C. Miller, N. J. Foti, and R. P. Adams. Variational boosting: Iteratively refining posterior approximations. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 2420–2429, 2017.

[82] R. E. Turner and M. Sahani. Two problems with variational expectation maximisation for time-series models. In D. Barber, A. T. Cemgil, and S. Chiappa, editors, *Bayesian Time Series Models*. Cambridge University Press, 2011.

[83] B. Wang and D. M. Titterington. Inadequacy of interval estimates corresponding to variational bayesian approximations. In *AISTATS*. Barbados, 2005.

[84] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng. *Handbook of Markov chain Monte Carlo*. Chapman and Hall/CRC Boca Raton, 2011.

[85] C. Robert and G. Casella. *Monte Carlo statistical methods*. Springer Science and Business Media, 2013.

[86] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 6 1953.

[87] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97, 4 1970.

[88] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):721–741, 11 1984.

[89] B. Ninness and S. Henriksen. Bayesian system identification via Markov chain Monte Carlo techniques. *Automatica*, 46(1):40–51, 1 2010.

[90] C. Andrieu, A. Doucet, and R. Holenstein. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):269–342, 6 2010.

[91] J. S. Rosenthal. Optimal proposal distributions and adaptive MCMC. In *Handbook of Markov Chain Monte Carlo*, volume 4, chapter 4. Chapman and Hall/CRC Boca Raton, 2011.

[92] A. Wills, T. B. Schön, F. Lindsten, and B. Ninness. Estimation of linear systems using a Gibbs sampler. *IFAC Proceedings Volumes*, 45(16):203–208, 7 2012.

[93] D. Tcherniak, S. Chauhan, and M. H. Hansen. Applicability limits of operational modal analysis to operational wind turbines. In *Structural Dynamics and Renewable Energy, Volume 1*, pages 317–327. Springer, 2011.

[94] S. Mukhopadhyay, H. Luş, and R. Betti. Modal parameter based structural identification using inputoutput data: minimal instrumentation and global identifiability issues. *Mechanical Systems and Signal Processing*, 45(2):283–301, 2014.

[95] R. J. Allemang and D. L. Brown. A correlation coefficient for modal vector analysis. In *Proceedings of the 1st international modal analysis conference*, volume 1, pages 110–116. Orlando: Union College Press, 1982.

[96] J. M. M. e Silva and N. M. M. Maia. *Modal analysis and testing*, volume 363. Springer Science and Business Media, 2012.

[97] E. M. Hernandez and N. R. Polanco. A lower bound for the variance of frequency and damping ratio identified from noisy vibration measurements. *Structural Control and Health Monitoring*, 23(1):5–19, 1 2016.

[98] S.-K. Au and B. Li. Posterior uncertainty, asymptotic law and Cramér-Rao bound. *Structural Control and Health Monitoring*, 25(3):e2113, 3 2018.

[99] G. Kerschen and J. C. Golinval. Experimental modal analysis, 2014. URL http://www.ltas-vis.ulg.ac.be/cmsms/uploads/File/Mvibr_notes.pdf. Last accessed: 2017-09-14.

[100] F. Vicario, M. Q. Phan, R. Betti, and R. W. Longman. Outputonly observer/Kalman filter identification (O3KID). *Structural Control and Health Monitoring*, 22(5):847–872, 2015.

[101] P. Andersen. *Identification of Civil Engineering Structures using Vector ARMA Models*. PhD thesis, Aalborg University, Denmark, 1997.

[102] P. Van Overschee and B. De Moor. N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, 30(1):75–93, 1 1994.

[103] R. H. Shumway and D. S. Stoffer. An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, 3(4):253–264, 7 1982.

[104] S. Gibson and B. Ninness. Robust maximum-likelihood estimation of multivariable dynamic systems. *Automatica*, 41(10):1667–1682, 10 2005.

[105] K. P. Murphy. Conjugate Bayesian analysis of the Gaussian distribution. Technical report, 2007.

[106] C. M Bishop. *Pattern recognition and machine learning*. Springer, 2006.

[107] D. Ostwald, E. Kirilina, L. Starke, and F. Blankenburg. A tutorial on variational Bayes for latent linear stochastic time-series models. *Journal of Mathematical Psychology*, 60:1–19, 6 2014.

[108] D. Barber and S. Chiappa. Unified inference for variational Bayesian linear Gaussian state-space models. In *Advances in Neural Information Processing Systems*, pages 81–88, 2007.

[109] M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, volume 55. Courier Corporation, 1965.

[110] T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley-Interscience, 2006.

[111] K. Fujimoto, A. Satoh, and S. Fukunaga. System identification based on variational Bayes method and the invariance under coordinate transformations. In *IEEE Conference on Decision and Control and European Control Conference*, pages 3882–3888. IEEE, 12 2011.

[112] J. Schafer. *Analysis of incomplete multivariate data*, volume 72 of *CRC Monographs on Statistics and Applied Probability*. Chapman and Hall, 8 1997.

[113] S. Van Buuren. *Flexible imputation of missing data*. Chapman and Hall/CRC, 2018.

[114] B. Li. *Uncertainty quantification in vibration-based structural health monitoring using Bayesian statistics*. PhD thesis, UC Berkeley, 2016.

[115] M. West and J. Harrison. Bayesian forecasting and dynamic models. Springer-Verlag, 1997.

[116] A. Gelman and D. B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472, 11 1992.

[117] S. P. Brooks and A. Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455, 12 1998.

[118] Shared Hierarchical Academic Research Computing Network (SHARCNET) and Compute/Calcul Canada. http://www.sharcnet.ca, 2019.

[119] S. O. R. Moheimani, D. Halim, and A. J. Fleming. *Spatial control of vibration: theory and experiments*, volume 10. World scientific, 2003.

[120] L. Ljung. *System identification: Theory for the user*. Pearson Education, 1999.

[121] S. Gillijns and B. De Moor. Unbiased minimum-variance input and state estimation for linear discrete-time systems with direct feedthrough. *Automatica*, 43(5):934–937, 2007.

[122] E. Lourens, E. Reynders, G. De Roeck, G. Degrande, and G. Lombaert. An augmented Kalman filter for force identification in structural dynamics. *Mechanical Systems and Signal Processing*, 27:446–460, 2012.

[123] F. Naets, J. Cuadrado, and W. Desmet. Stable force identification in structural dynamics using Kalman filtering and dummy-measurements. *Mechanical Systems and Signal Processing*, 50:235–248, 2015.

[124] S. E. Azam, E. Chatzi, and C. Papadimitriou. A dual Kalman filter approach for state estimation via output-only acceleration measurements. *Mechanical Systems and Signal Processing*, 60:866–886, 2015.

[125] R. Nayek, S. Chakraborty, and S. Narasimhan. A Gaussian process latent force model for joint input-state estimation in linear structural systems. *Mechanical Systems and Signal Processing*, 128:497–530, 2019.

[126] H. P. Gavin. Lecture notes on numerical integration for structural dynamics, 2018.

[127] F. M. Callier and C. A. Desoer. *Linear system theory*. Springer Science and Business Media, 2012.

[128] S. Särkkä, M. A. Álvarez, and N. D. Lawrence. Gaussian process latent force models for learning and stochastic control of physical systems. *IEEE Transactions on Automatic Control*, 2018.

[129] M. A. Alvarez, D. Luengo, and N. D. Lawrence. Linear latent force models using Gaussian processes. *IEEE Transactions on Pattern analysis and Machine intelligence*, 35(11):2693–2705, 2013.

[130] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, USA, 2006.

[131] B. Matérn. Spatial variation: Stochastic models and their applications to some problems in forest surveys and other sampling investigations. *Meddelanden från Statens Skogsforskningsinstitut*, 49:1–144, 1960.

[132] P. Whittle. On stationary processes in the plane. *Biometrika*, pages 434–449, 1954.

[133] J. Hartikainen and S. Särkkä. Kalman filtering and smoothing solutions to temporal Gaussian process regression models. In *Machine Learning for Signal Processing (MLSP), 2010 IEEE International Workshop on*, pages 379–384. IEEE, 2010.

[134] S. Särkkä. *Recursive Bayesian inference on stochastic differential equations*. PhD thesis, Helsinki University of Technology, 2006.

[135] Simo Särkkä and Arno Solin. *Applied stochastic differential equations*, volume 10. Cambridge University Press, 2019.

[136] K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

[137] M. Filippone, M. Zhong, and M. Girolami. A comparative evaluation of stochastic-based inference methods for gaussian process models. *Machine Learning*, 93(1): 93–114, 2013.

[138] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 4 2017.

[139] R. J. Giordano, T. Broderick, and M. I. Jordan. Linear response methods for accurate covariance estimates from mean field variational Bayes. In *Advances in Neural Information Processing Systems*, pages 1441–1449, 2015.

[140] Vlach Jiří and Kishore Singhal. *Computer methods for circuit analysis and design*. Springer Science and Business Media, 1983.

[141] M. Gupta and S. Srivastava. Parametric bayesian estimation of differential entropy and relative entropy. *Entropy*, 12(4):818–843, 2010.

[142] K. B. Petersen and M. S. Pedersen. The matrix cookbook. Technical report, 2012. URL https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf. Last accessed: 2019-05-27.

[143] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35, 3 1960.

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

APPENDICES

Appendix A

List of Publications

The following is a list of journal publications and conference papers resulting from the work contained in this thesis:

Publications

1. **R. Nayek**, S. Chakraborty and S. Narasimhan. A Gaussian process latent force model for joint input-state estimation in linear structural systems. *Mechanical Systems and Signal Processing* 128: 497-530, 2019.

2. **R. Nayek**, S. Mukhopadhyay and S. Narasimhan. Mass normalized mode shape identification of bridge structures using a single actuator-sensor pair. *Structural Control and Health Monitoring* 25(11), 2018.

Conference Proceedings - Full paper

1. **R. Nayek**, S. Mukhopadhyay and S. Narasimhan. Modal Identification Using a roving actuator and a fixed sensor. *Proceedings of the 36th International Modal Analysis Conference (IMAC)*, Orlando, FL. USA, 2018.

2. **R. Nayek**, S. Mukhopadhyay and S. Narasimhan. Bridge model updating and health monitoring using a single mobile actuator-sensor pair. *Proceedings of the 9th International Conference on Inverse Problems in Engineering (ICIPE)*, Waterloo, ON. CA, 2017.

Appendix B

Calculation of modal parameter and their derivatives

B.1 Modal parameter extraction from discrete-time SSM matrices

The modal parameters of a SSM are invariant under a similarity transformation and can be uniquely determined from the measured data. The modal parameters can be extracted by applying the following steps [57]:

1. Apply eigenvalue decomposition on $\mathbf{A} \rightarrow \mathbf{A} = \boldsymbol{\Psi}\boldsymbol{\Lambda}\boldsymbol{\Psi}^{-1}$;
2. Select the complex pairs of modes, and partition the corresponding eigenvalues and the eigenvector matrices

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\lambda} & \mathbf{0} \\ \mathbf{0} & \overline{\boldsymbol{\lambda}} \end{bmatrix}, \boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\psi} & \overline{\boldsymbol{\psi}} \end{bmatrix} \tag{B.1}$$

3. For each $\lambda_i, \boldsymbol{\psi}_i$, extract the ith modal frequency and modal damping ratio

$$\lambda_{ci} = \ln \lambda_i / \Delta t, \quad f_i = \text{ abs}\left(\lambda_{ci}\right)/2\pi, \quad \xi_i = -\text{Re}(\lambda_{ci})/\text{ abs}\left(\lambda_{ci}\right) \tag{B.2}$$

where Δt is the sampling time period, λ_i is the ith diagonal element of $\boldsymbol{\lambda}$ and $\boldsymbol{\psi}_i$ is the ith column vector of $\boldsymbol{\Psi}$

4. Define a matrix
$$\boldsymbol{\varphi}_i = \begin{bmatrix} \text{Re}(\boldsymbol{\psi}_i) & \text{Im}(\boldsymbol{\psi}_i) \end{bmatrix} \tag{B.3}$$

5. Take the generalized eigenvalue decomposition:
$$\mathbf{P}_1 \boldsymbol{\varphi}_i^T \mathbf{G}^T \mathbf{G} \boldsymbol{\varphi}_i \mathbf{P}_1 \hat{\boldsymbol{z}} = \lambda_{\min} \mathbf{P}_2 \boldsymbol{\varphi}_i^T \mathbf{G}^T \mathbf{G} \boldsymbol{\varphi}_i \mathbf{P}_2 \hat{\boldsymbol{z}} \tag{B.4}$$
where $\mathbf{P}_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and $\mathbf{P}_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, $\lambda_{\min}$ is the smallest eigenvalue and $\hat{\boldsymbol{z}}$ is the corresponding eigenvector

6. The ith un-normalized real-valued mode shape is
$$\boldsymbol{\phi}_i^{un} = \mathbf{G} \boldsymbol{\varphi}_i \mathbf{P}_2 \hat{\boldsymbol{z}} \tag{B.5}$$

B.2 Calculation of Jacobian matrices

The entries of the Jacobian matrices in Eq.(4.51) are calculated by differentiating the modal parameters with respect to each element a_{mn} (of $\mathbf{A}$) and g_{mn} (of $\mathbf{G}$) using chain rule:

$$\begin{aligned} \frac{\partial \ln f_i(\mathbf{A})}{\partial a_{mn}} &= \frac{\partial \ln f_i(\mathbf{A})}{\partial f_i(\mathbf{A})} \times \frac{\partial f_i(\lambda_{ci})}{\partial \lambda_{ci}} \times \frac{\partial \lambda_{ci}}{\partial \lambda_i} \times \frac{\partial \lambda_i}{\partial a_{mn}} \\ &= \frac{1}{f_i} \begin{bmatrix} \frac{\text{Re}(\lambda_{ci})}{2\pi \text{ abs}(\lambda_{ci})} & \frac{\text{Im}(\lambda_{ci})}{2\pi \text{ abs}(\lambda_{ci})} \end{bmatrix} \begin{bmatrix} \frac{\text{Re}(\lambda_{ci})}{\Delta t \text{ abs}(\lambda_i)^2} & \frac{\text{Im}(\lambda_i)}{\Delta t \text{ abs}(\lambda_{ci})^2} \\ -\frac{\text{Im}(\lambda_i)}{\Delta t \text{ abs}(\lambda_i)^2} & \frac{\text{Re}(\lambda_i)}{\Delta t \text{ abs}(\lambda_i)^2} \end{bmatrix} \begin{bmatrix} \frac{\text{Re}(\lambda_i)}{\partial a_{mn}} \\ \frac{\text{Im}(\lambda_i)}{\partial a_{mn}} \end{bmatrix} \end{aligned} \tag{B.6a}$$

$$\begin{aligned} \frac{\partial \ln \xi_i(\mathbf{A})}{\partial a_{mn}} &= \frac{\partial \ln \xi_i(\mathbf{A})}{\partial \xi_i(\mathbf{A})} \times \frac{\partial \xi_i(\lambda_{ci})}{\partial \lambda_{ci}} \times \frac{\partial \lambda_{ci}}{\partial \lambda_i} \times \frac{\partial \lambda_i}{\partial a_{mn}} \\ &= \frac{1}{f_i} \begin{bmatrix} -\frac{\text{Im}(\lambda_{ci})}{\text{abs}(\lambda_{ci})^3} & \frac{\text{Re}(\lambda_{ci})\text{Im}(\lambda_{ci})}{\text{abs}(\lambda_{ci})^3} \end{bmatrix} \begin{bmatrix} \frac{\text{Re}(\lambda_{ci})}{\Delta t \text{ abs}(\lambda_i)^2} & \frac{\text{Im}(\lambda_i)}{\Delta t \text{ abs}(\lambda_{ci})^2} \\ -\frac{\text{Im}(\lambda_i)}{\Delta t \text{ abs}(\lambda_i)^2} & \frac{\text{Re}(\lambda_i)}{\Delta t \text{ abs}(\lambda_i)^2} \end{bmatrix} \begin{bmatrix} \frac{\text{Re}(\lambda_i)}{\partial a_{mn}} \\ \frac{\text{Im}(\lambda_i)}{\partial a_{mn}} \end{bmatrix} \end{aligned} \tag{B.6b}$$

$$\frac{\partial \boldsymbol{\phi}_i^{un}(\mathbf{A}, \mathbf{G})}{\partial a_{mn}} = \mathbf{G} \begin{bmatrix} \frac{\partial \text{Re}(\boldsymbol{\psi}_i)}{\partial a_{mn}} & \frac{\partial \text{Im}(\boldsymbol{\psi}_i)}{\partial a_{mn}} \end{bmatrix} \mathbf{P}_2 \hat{\boldsymbol{z}} + \mathbf{G} \boldsymbol{\psi}_i \mathbf{P}_2 \frac{\partial \hat{\boldsymbol{z}}}{\partial a_{mn}} \tag{B.6c}$$

$$\frac{\partial \boldsymbol{\phi}_i^{un}(\mathbf{A},\mathbf{G})}{\partial g_{mn}} = \boldsymbol{e}_m \boldsymbol{e}_n^T \boldsymbol{\psi}_i \mathbf{P}_2 \hat{\boldsymbol{z}} + \mathbf{G}\boldsymbol{\psi}_i \mathbf{P}_2 \frac{\partial \hat{\boldsymbol{z}}}{\partial g_{mn}} \tag{B.6d}$$

Computation of the partial derivatives of eigenvalues and eigenvectors is described next.

B.3 Partial derivatives of eigenvalue and eigenvector

This material in this section mostly follows from [57] and [140]. Consider the generalized eigenvalue problem

$$\mathbf{A}(\nu)\,\boldsymbol{\phi} = \lambda \mathbf{B}(\nu)\,\boldsymbol{\phi} \tag{B.7}$$

where $\mathbf{A}(\nu), \mathbf{B}(\nu) \in \mathbb{R}^{n\times n}$. The matrix $\mathbf{H} = \mathbf{A}(\nu) - \lambda \mathbf{B}(\nu)$ is singular (i.e. $|\mathbf{H}| = 0$). Consider the LU decomposition of $\mathbf{H}$,

$$\mathbf{PH} = \mathbf{LU} \tag{B.8}$$

where $\mathbf{P}$ is a permutation matrix, $\mathbf{L}$ is a lower triangular matrix with unit entries along the diagonal and $\mathbf{u}$ is an upper triangular matrix. Using the property of determinant, one can write

$$|\mathbf{H}| = |\mathbf{L}|\,|\mathbf{U}| = |\mathbf{U}| \tag{B.9}$$

where the fact $|\mathbf{L}| = 1$ is used. Since $|\mathbf{H}| = 0$, it means $|\mathbf{U}| = 0$, implying the presence of a zero element in the diagonal of $\mathbf{U}$. Assuming that this zero element resides in the last column, i.e. $u_{nn} = 0$, one can obtain the derivative of $u_{nn}(\nu, \lambda(\nu))$ with respect to the parameter ν as

$$\begin{aligned} &\frac{\partial u_{nn}}{\partial \lambda}\frac{d\lambda}{d\nu} + \frac{\partial u_{nn}}{\partial \nu} = 0 \\ \implies &\frac{d\lambda}{d\nu} = -\frac{\frac{\partial u_{nn}}{\partial \nu}}{\frac{\partial u_{nn}}{\partial \lambda}} \end{aligned} \tag{B.10}$$

To find the terms $\frac{\partial u_{nn}}{\partial \lambda}$ and $\frac{\partial u_{nn}}{\partial \nu}$, differentiate Eq. (B.8) with respect to an arbitrary parameter h

$$\mathbf{P}\frac{\partial \mathbf{H}}{\partial h} = \frac{\partial \mathbf{L}}{\partial h}\mathbf{U} + \mathbf{L}\frac{\partial \mathbf{U}}{\partial h} \tag{B.11}$$

Define two vectors $\boldsymbol{u}$ and $\boldsymbol{z}$ as follows:

$$\boldsymbol{u} \in \mathbb{C}^n, \quad \text{s.t. } \mathbf{L}^H \boldsymbol{u} = \boldsymbol{e}_n \tag{B.12a}$$

$$\boldsymbol{z} \in \mathbb{C}^n, \quad \text{s.t. } \mathbf{U}\boldsymbol{z} = \mathbf{0} \text{ and } z_n = 1 \tag{B.12b}$$

where $\boldsymbol{e}_n$ is the nth column of an $n \times n$ identity matrix and H represents conjugate transpose. Pre- and post-multiplying Eq. (B.11) by $\boldsymbol{u}^H$ and $\boldsymbol{z}$ yields

$$\boldsymbol{u}^H \mathbf{P} \frac{\partial \mathbf{H}}{\partial h} \boldsymbol{z} = \boldsymbol{u}^H \frac{\partial \mathbf{L}}{\partial h} \mathbf{U}\boldsymbol{z} + \boldsymbol{u}^H \mathbf{L} \frac{\partial \mathbf{U}}{\partial h} \boldsymbol{z} = \frac{\partial u_{nn}}{\partial h} \tag{B.13}$$

By setting h equal to λ and ν, one can obtain $\frac{\partial u_{nn}}{\partial \lambda}$ and $\frac{\partial u_{nn}}{\partial \nu}$, and then $\frac{d\lambda}{d\nu}$ is obtained from Eq. (B.10).

For the partial derivative of eigenvectors, one starts by taking derivative of $\mathbf{H}(\nu, \lambda(\nu))\boldsymbol{\phi}(\nu)$

$$\left(\frac{\partial \mathbf{H}}{\partial \nu} + \frac{\partial \mathbf{H}}{\partial \lambda} \frac{\partial \lambda}{\partial \nu} \right) \boldsymbol{\phi} + \mathbf{H} \frac{\partial \boldsymbol{\phi}}{\partial \nu} = 0 \tag{B.14}$$

Inserting Eq. (B.8) for $\mathbf{H}$ and rearranging the above equation leads to the following linear equation

$$\mathbf{U} \frac{\partial \boldsymbol{\phi}}{\partial \nu} = -\mathbf{L}^{-1}\mathbf{P} \left(\frac{\partial \mathbf{H}}{\partial \nu} + \frac{\partial \mathbf{H}}{\partial \lambda} \frac{\partial \lambda}{\partial \nu} \right) \boldsymbol{\phi} \tag{B.15}$$

where $\mathbf{U}$ and the terms on the right hand side are all known from previous steps. However, $\mathbf{U}$ is rank-deficient, and hence not readily invertible. To allow for solution of Eq. (B.15), a constraint on the normalization of eigenvectors are used i.e.

$$\boldsymbol{\phi}^H \boldsymbol{\phi} = \text{constant} \tag{B.16}$$

Differentiating this constraint with respect to ν

$$\boldsymbol{\phi}^H \frac{\partial \boldsymbol{\phi}}{\partial \nu} = 0 \tag{B.17}$$

Combining Eqs. (B.17) and (B.15) yields

$$\begin{bmatrix} \mathbf{U}_{(1:n-1,:)} \\ \boldsymbol{\phi}^H \end{bmatrix} \frac{\partial \boldsymbol{\phi}}{\partial \nu} = -\mathbf{L}^{-1}\mathbf{P} \left(\frac{\partial \mathbf{H}}{\partial \nu} + \frac{\partial \mathbf{H}}{\partial \lambda} \frac{\partial \lambda}{\partial \nu} \right) \boldsymbol{\phi} \tag{B.18}$$

where $\mathbf{U}_{(1:n-1,:)}$ denotes the first $n-1$ rows of matrix $\mathbf{U}$. Now the matrix on the left hand side becomes square and full-rank; hence $\partial\boldsymbol{\phi}/\partial\nu$ can be uniquely determined.

- **Calculation of $\partial\lambda/\partial a_{ij}$**

Define $\mathbf{H}_1 = \mathbf{A} - \lambda\mathbf{I}$, and one can write $\frac{\partial\mathbf{H}_1}{\partial\lambda} = -\mathbf{I}$ and $\frac{\partial\mathbf{H}_1}{\partial a_{ij}} = \boldsymbol{e}_i\boldsymbol{e}_j^T$. Using $\mathbf{H}_1$ in place of $\mathbf{H}$ gives

$$\frac{\partial u_{nn}}{\partial\lambda} = -\boldsymbol{u}^H\mathbf{P}\boldsymbol{z}, \quad \frac{\partial u_{nn}}{\partial a_{ij}} = \boldsymbol{u}^H\mathbf{P}\boldsymbol{e}_i\boldsymbol{e}_j^T\boldsymbol{z} = (\boldsymbol{u}^H\mathbf{P})_i z_j \tag{B.19}$$

Using Eq. (B.10),the following is obtained

$$\frac{\partial\lambda}{\partial a_{ij}} = \frac{(\boldsymbol{u}^H\mathbf{P})_i z_j}{\boldsymbol{u}^H\mathbf{P}\boldsymbol{z}} \tag{B.20}$$

which can be written in a compact matrix form as

$$\frac{\partial\lambda}{\partial\mathrm{vec}\,(\mathbf{A})} = \frac{\boldsymbol{z}^T\otimes(\boldsymbol{u}^H\mathbf{P})}{\boldsymbol{u}^H\mathbf{P}\boldsymbol{z}} \tag{B.21}$$

- **Calculation of $\partial\boldsymbol{\phi}/\partial a_{ij}$**

$$\begin{bmatrix}\mathbf{U}_{(1:n-1,:)}\\ \boldsymbol{\phi}^H\end{bmatrix}\frac{\partial\boldsymbol{\phi}}{\partial a_{ij}} = -\mathbf{L}^{-1}\mathbf{P}\left(\frac{\partial\mathbf{H}_1}{\partial a_{ij}} + \frac{\partial\mathbf{H}_1}{\partial\lambda}\frac{\partial\lambda}{\partial a_{ij}}\right)\boldsymbol{\phi} \tag{B.22}$$

Inserting the known terms from Eqs. (B.19) and (B.20) in the above equation gives

$$\begin{bmatrix}\mathbf{U}_{(1:n-1,:)}\\ \boldsymbol{\phi}^H\end{bmatrix}\frac{\partial\boldsymbol{\phi}}{\partial a_{ij}} = \mathbf{L}^{-1}\mathbf{P}\left(\frac{\partial\lambda}{\partial a_{ij}}\mathbf{I} - \boldsymbol{e}_i\boldsymbol{e}_j^T\right)\boldsymbol{\phi} \tag{B.23}$$

which can be written into a compact matrix form as

$$\begin{bmatrix}\mathbf{U}_{(1:n-1,:)}\\ \boldsymbol{\phi}^H\end{bmatrix}\frac{\partial\boldsymbol{\phi}}{\partial\mathrm{vec}\,(\mathbf{A})} = \mathbf{L}^{-1}\mathbf{P}\left(\frac{\partial\lambda}{\partial\mathrm{vec}\,(\mathbf{A})}\otimes\boldsymbol{\phi} - \boldsymbol{\phi}^T\otimes\mathbf{I}\right) \tag{B.24}$$

Solving Eq. (B.23) yields $\partial\boldsymbol{\phi}/\partial a_{ij}$.

- **Calculation of $\partial\lambda_{\min}/\partial a_{mn}$**

Define $\mathbf{H}_2 = \mathbf{P}_1\boldsymbol{\varphi}_i^T\mathbf{G}^T\mathbf{G}\boldsymbol{\varphi}_i\mathbf{P}_1 - \lambda_{\min}\mathbf{P}_2\boldsymbol{\varphi}_i^T\mathbf{G}^T\mathbf{G}\boldsymbol{\varphi}_i\mathbf{P}_2$, then using $\mathbf{H}_2$ in place of $\mathbf{H}$ in Eq. (B.8)

$$\frac{\partial\mathbf{H}_2}{\partial\lambda_{\min}} = -\mathbf{P}_2\boldsymbol{\varphi}_i^T\mathbf{G}^T\mathbf{G}\boldsymbol{\varphi}_i\mathbf{P}_2 \tag{B.25}$$

and

$$\begin{aligned}\frac{\partial\mathbf{H}_2}{\partial a_{mn}} =& \mathbf{P}_1\frac{\partial\boldsymbol{\varphi}_i}{\partial a_{mn}}^T\mathbf{G}^T\mathbf{G}\boldsymbol{\varphi}_i\mathbf{P}_1 + \mathbf{P}_1\boldsymbol{\varphi}_i^T\mathbf{G}^T\mathbf{G}\frac{\partial\boldsymbol{\varphi}_i}{\partial a_{mn}}\mathbf{P}_1 \\ &- \lambda_{\min}\left(\mathbf{P}_2\frac{\partial\boldsymbol{\varphi}_i}{\partial a_{mn}}^T\mathbf{G}^T\mathbf{G}\boldsymbol{\varphi}_i\mathbf{P}_2 + \mathbf{P}_2\boldsymbol{\varphi}_i^T\mathbf{G}^T\mathbf{G}\frac{\partial\boldsymbol{\varphi}_i}{\partial a_{mn}}\mathbf{P}_2\right)\end{aligned} \tag{B.26}$$

With that, one can write using Eq. (B.13)

$$\frac{\partial u_{22}}{\partial\lambda_{\min}} = -\boldsymbol{u}^H\mathbf{P}\mathbf{P}_2\boldsymbol{\varphi}_i^T\mathbf{G}^T\mathbf{G}\boldsymbol{\varphi}_i\mathbf{P}_2\boldsymbol{z} \tag{B.27a}$$

$$\frac{\partial u_{22}}{\partial a_{mn}} = \boldsymbol{u}^H\mathbf{P}\frac{\partial\mathbf{H}_2}{\partial a_{mn}}\boldsymbol{z} \tag{B.27b}$$

and thus one obtains using Eq. (B.27) and Eq. (B.10)

$$\frac{\partial\lambda_{\min}}{\partial a_{mn}} = -\frac{\frac{\partial u_{22}}{\partial a_{mn}}}{\frac{\partial u_{22}}{\partial\lambda_{\min}}} = \frac{\boldsymbol{u}^H\mathbf{P}\frac{\partial\mathbf{H}_2}{\partial a_{mn}}\boldsymbol{z}}{\boldsymbol{u}^H\mathbf{P}\mathbf{P}_2\boldsymbol{\varphi}_i^T\mathbf{G}^T\mathbf{G}\boldsymbol{\varphi}_i\mathbf{P}_2\boldsymbol{z}} \tag{B.28}$$

- **Calculation of $\partial\hat{\boldsymbol{z}}/\partial a_{mn}$**

For the sensitivity of eigenvector $\hat{\boldsymbol{z}}$, one can write using Eq. (B.18)

$$\begin{aligned}&\begin{bmatrix}\mathbf{U}_{(1,:)}\\ \hat{\boldsymbol{z}}^H\end{bmatrix}\frac{\partial\hat{\boldsymbol{z}}}{\partial a_{mn}} = -\mathbf{L}^{-1}\mathbf{P}\left(\frac{\partial\mathbf{H}_2}{\partial a_{mn}} + \frac{\partial\mathbf{H}_2}{\partial\lambda_{\min}}\frac{\partial\lambda_{\min}}{\partial a_{mn}}\right)\hat{\boldsymbol{z}} \\ \Longrightarrow\ &\begin{bmatrix}\mathbf{U}_{(1,:)}\\ \hat{\boldsymbol{z}}^H\end{bmatrix}\frac{\partial\hat{\boldsymbol{z}}}{\partial a_{mn}} = -\mathbf{L}^{-1}\mathbf{P}\left(\frac{\partial\mathbf{H}_2}{\partial a_{mn}} - \mathbf{P}_2\boldsymbol{\varphi}_i^T\mathbf{G}^T\mathbf{G}\boldsymbol{\varphi}_i\mathbf{P}_2\frac{\partial\lambda_{\min}}{\partial a_{mn}}\right)\hat{\boldsymbol{z}}\end{aligned} \tag{B.29}$$

Solving Eq. (B.29) yields $\partial\hat{\boldsymbol{z}}/\partial a_{mn}$.

- **Calculation of** $\partial\lambda_{\min}/\partial g_{mn}$ and $\partial\hat{\boldsymbol{z}}/\partial g_{mn}$

Taking partial derivative of $\mathbf{H}_2$ with respect to g_{mn}, one gets

$$\begin{aligned}\frac{\partial \mathbf{H}_2}{\partial g_{mn}} =& \mathbf{P}_1\boldsymbol{\varphi}_i^T\boldsymbol{e}_n\boldsymbol{e}_m^T\mathbf{G}\boldsymbol{\varphi}_i\mathbf{P}_1 + \mathbf{P}_1\boldsymbol{\varphi}_i^T\mathbf{G}^T\boldsymbol{e}_m\boldsymbol{e}_n^T\boldsymbol{\varphi}_i\mathbf{P}_1 \\ &- \lambda_{\min}\left(\mathbf{P}_2\boldsymbol{\varphi}_i^T\boldsymbol{e}_n\boldsymbol{e}_m^T\mathbf{G}\boldsymbol{\varphi}_i\mathbf{P}_2 + \mathbf{P}_2\boldsymbol{\varphi}_i^T\mathbf{G}^T\boldsymbol{e}_m\boldsymbol{e}_n^T\boldsymbol{\varphi}_i\mathbf{P}_2\right)\end{aligned} \tag{B.30}$$

Therefore, with $\frac{\partial u_{22}}{\partial g_{mn}} = \boldsymbol{u}^H\mathbf{P}\frac{\partial \mathbf{H}_2}{\partial g_{mn}}\boldsymbol{z}$

$$\frac{\partial \lambda_{\min}}{\partial g_{mn}} = -\frac{\frac{\partial u_{22}}{\partial g_{mn}}}{\frac{\partial u_{22}}{\partial \lambda_{\min}}} = \frac{\boldsymbol{u}^H\mathbf{P}\frac{\partial \mathbf{H}_2}{\partial g_{mn}}\boldsymbol{z}}{\boldsymbol{u}^H\mathbf{P}\mathbf{P}_2\boldsymbol{\varphi}_i^T\mathbf{G}^T\mathbf{G}\boldsymbol{\varphi}_i\mathbf{P}_2\boldsymbol{z}} \tag{B.31}$$

and the sensitivity of the eigenvector with respect to g_{mn} is given by

$$\begin{bmatrix}\mathbf{U}_{(1,:)} \\ \hat{\boldsymbol{z}}^H\end{bmatrix}\frac{\partial \hat{\boldsymbol{z}}}{\partial g_{mn}} = -\mathbf{L}^{-1}\mathbf{P}\left(\frac{\partial \mathbf{H}_2}{\partial g_{mn}} - \mathbf{P}_2\boldsymbol{\varphi}_i^T\mathbf{G}^T\mathbf{G}\boldsymbol{\varphi}_i\mathbf{P}_2\frac{\partial \lambda_{\min}}{\partial g_{mn}}\right)\hat{\boldsymbol{z}} \tag{B.32}$$

Solving Eq. (B.32) yields $\partial\hat{\boldsymbol{z}}/\partial g_{mn}$.

Appendix C

Calculation of expectations over missing observations

1. **Derivation of $\langle \boldsymbol{y}_k \rangle$**

   The distribution of $\boldsymbol{y}_k$ conditioned on $\boldsymbol{x}_k$ can be written

$$\begin{bmatrix} \boldsymbol{y}_k^{obs} \mid \boldsymbol{x}_k \\ \boldsymbol{y}_k^{miss} \mid \boldsymbol{x}_k \end{bmatrix} = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\upsilon}_k^{(1)} \mathbf{G} \boldsymbol{x}_k \\ \boldsymbol{\upsilon}_k^{(2)} \mathbf{G} \boldsymbol{x}_k \end{bmatrix}, \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{21}^T \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{bmatrix} \right)$$

   Using the property of conditional distributions for multivariate normal distribution, one can obtain

$$\begin{aligned} \left(\boldsymbol{y}_k^{obs} \mid \boldsymbol{y}_k^{obs}, \boldsymbol{x}_k \right) &= \boldsymbol{\upsilon}_k^{(1)} \boldsymbol{y}_k, \\ \left(\boldsymbol{y}_k^{miss} \mid \boldsymbol{y}_k^{obs}, \boldsymbol{x}_k \right) &\sim \mathcal{N}\left(\ddot{\boldsymbol{\mu}}, \ddot{\boldsymbol{\Sigma}} \right), \text{ where} \\ \ddot{\boldsymbol{\mu}} &= \boldsymbol{\upsilon}_k^{(2)} \mathbf{G} \boldsymbol{x}_k + \mathbf{R}_{21} \mathbf{R}_{11}^{-1} \boldsymbol{\upsilon}_k^{(1)} \left(\boldsymbol{y}_k - \mathbf{G} \boldsymbol{x}_k \right) \\ \ddot{\boldsymbol{\Sigma}} &= \mathbf{R}_{22} - \mathbf{R}_{21} \mathbf{R}_{11}^{-1} \mathbf{R}_{21}^T \end{aligned} \tag{C.1}$$

From this, the expected value of $\boldsymbol{y}_k$ given $\boldsymbol{y}_k^{obs}$ and $\langle \boldsymbol{x}_k \rangle = \hat{\boldsymbol{x}}_{k|N}$ can be written as

$$\begin{aligned} \tilde{\boldsymbol{y}}_k = \left\langle \boldsymbol{y}_k \mid \boldsymbol{y}_k^{obs}, \boldsymbol{x}_k \right\rangle &= \left\langle \left\langle \boldsymbol{y}_k \mid \boldsymbol{y}_k^{obs}, \boldsymbol{x}_k \right\rangle_{y|x} \right\rangle_x \\ &= \left\langle \boldsymbol{y}_k - \mathbf{Z}_k \left(\boldsymbol{y}_k - \mathbf{G}\boldsymbol{x}_k \right) \right\rangle_x \\ &= \boldsymbol{y}_k - \mathbf{Z}_k \left(\boldsymbol{y}_k - \mathbf{G}\hat{\boldsymbol{x}}_{k|N} \right) \\ \text{where } \mathbf{Z}_k &= \mathbf{I} - \mathbf{R}\boldsymbol{\mho}_k^{(1)^T} \mathbf{R}_{11}^{-1} \boldsymbol{\mho}_k^{(1)} \end{aligned} \tag{C.2}$$

2. **Derivation of $\left\langle \boldsymbol{y}_k \boldsymbol{y}_k^T \right\rangle$** Denoting 'COV' as the covariance of a random variable, one can write

$$\begin{aligned} \mathrm{COV}_{y|x} \left[\boldsymbol{y}_k \mid \boldsymbol{y}_k^{obs}, \boldsymbol{x}_k \right] &= \boldsymbol{\mho}_k^{(2)^T} \left(\mathbf{R}_{22} - \mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{R}_{21}^T \right) \boldsymbol{\mho}_k^{(2)} \\ &= \boldsymbol{\mho}_k^{(2)^T} \left(\boldsymbol{\mho}_k^{(2)} \mathbf{R} \boldsymbol{\mho}_k^{(2)^T} - \boldsymbol{\mho}_k^{(2)} \mathbf{R} \boldsymbol{\mho}_k^{(1)^T} \mathbf{R}_{11}^{-1} \boldsymbol{\mho}_k^{(1)} \mathbf{R}^T \boldsymbol{\mho}_k^{(2)^T} \right) \boldsymbol{\mho}_k^{(2)} \\ &= \mathbf{I}_k^{(2)} \left(\mathbf{R} - \mathbf{R}\boldsymbol{\mho}_k^{(1)^T} \mathbf{R}_{11}^{-1} \boldsymbol{\mho}_k^{(1)} \mathbf{R} \right) \mathbf{I}_k^{(2)} \\ &= \mathbf{I}_k^{(2)} \mathbf{Z}_k \mathbf{R} \mathbf{I}_k^{(2)} \end{aligned} \tag{C.3}$$

$$\begin{aligned} \left\langle \boldsymbol{y}_k \boldsymbol{y}_k^T \mid \boldsymbol{y}_k^{obs}, \boldsymbol{x}_k \right\rangle &= \left\langle \left\langle \boldsymbol{y}_k \boldsymbol{y}_k^T \mid \boldsymbol{y}_k^{obs}, \boldsymbol{x}_k \right\rangle_{y|x} \right\rangle_x \\ &= \left\langle \mathrm{COV}_{y|x} \left[\boldsymbol{y}_k \mid \boldsymbol{y}_k^{obs}, \boldsymbol{x}_k \right] + \left\langle \boldsymbol{y}_k \mid \boldsymbol{y}_k^{obs}, \boldsymbol{x}_k \right\rangle_{y|x} \left\langle \boldsymbol{y}_k^T \mid \boldsymbol{y}_k^{obs}, \boldsymbol{x}_k \right\rangle_{y|x} \right\rangle_x \\ &= \left\langle \mathbf{I}_k^{(2)} \mathbf{Z}_k \mathbf{R} \mathbf{I}_k^{(2)} + \left(\boldsymbol{y}_k - \mathbf{Z}_k \left(\boldsymbol{y}_k - \mathbf{G}\boldsymbol{x}_k \right) \right) \left(\boldsymbol{y}_k - \mathbf{Z}_k \left(\boldsymbol{y}_k - \mathbf{G}\boldsymbol{x}_k \right) \right)^T \right\rangle_x \\ &= \mathbf{I}_k^{(2)} \mathbf{Z}_k \mathbf{R} \mathbf{I}_k^{(2)} + \mathrm{COV}_x \left[\boldsymbol{y}_k - \mathbf{Z}_k \left(\boldsymbol{y}_k - \mathbf{G}\boldsymbol{x}_k \right) \right] + \\ &\qquad \left\langle \boldsymbol{y}_k - \mathbf{Z}_k \left(\boldsymbol{y}_k - \mathbf{G}\boldsymbol{x}_k \right) \right\rangle_x \left\langle \boldsymbol{y}_k - \mathbf{Z}_k \left(\boldsymbol{y}_k - \mathbf{G}\boldsymbol{x}_k \right) \right\rangle_x^T \\ &= \mathbf{I}_k^{(2)} \mathbf{Z}_k \mathbf{R} \mathbf{I}_k^{(2)} + \mathbf{I}_k^{(2)} \mathbf{Z}_k \mathbf{G} \hat{\mathbf{V}}_{k|N} \mathbf{G}^T \mathbf{Z}_k^T \mathbf{I}_k^{(2)} + \tilde{\boldsymbol{y}}_k \tilde{\boldsymbol{y}}_k^T \\ &= \mathbf{I}_k^{(2)} \mathbf{Z}_k \left(\mathbf{R} + \mathbf{G} \hat{\mathbf{V}}_{k|N} \mathbf{G}^T \mathbf{Z}_k^T \right) \mathbf{I}_k^{(2)} + \tilde{\boldsymbol{y}}_k \tilde{\boldsymbol{y}}_k^T \\ \text{where } \tilde{\boldsymbol{y}}_k &= \boldsymbol{y}_k - \mathbf{Z}_k \left(\boldsymbol{y}_k - \mathbf{G}\hat{\boldsymbol{x}}_{k|N} \right) \end{aligned} \tag{C.4}$$

3. **Derivation of $\left\langle \boldsymbol{y}_k \boldsymbol{x}_k^T \right\rangle$** Note that the observation equation (**??**) relates $\boldsymbol{y}_k$ to $\boldsymbol{x}_k$ as $\boldsymbol{y}_k = \mathbf{G}\boldsymbol{x}_k + \boldsymbol{v}_k$. Now $\boldsymbol{v}_k$ is independent of $\boldsymbol{x}_k$ and has the property that $\langle \boldsymbol{v}_k \rangle = \mathbf{0}$.

Thus, $\mathrm{COV}\left(\boldsymbol{v}_k, \boldsymbol{x}_k\right) = \mathbf{0}$. This result will be used in the following derivation.

$$\begin{aligned}
\left\langle \boldsymbol{y}_k \boldsymbol{x}_k^T \mid \boldsymbol{y}_k^{obs} \right\rangle &= \mathrm{COV}_x\left[\boldsymbol{y}_k, \boldsymbol{x}_k \mid \boldsymbol{y}_k^{obs}\right] + \left\langle \boldsymbol{y}_k \mid \boldsymbol{y}_k^{obs} \right\rangle_x \left\langle \boldsymbol{x}_k \mid \boldsymbol{y}_k^{obs} \right\rangle_x^T \\
&= \mathrm{COV}_x\left[\boldsymbol{y}_k - \mathbf{Z}_k\left(\boldsymbol{y}_k - \mathbf{G}\boldsymbol{x}_k\right) + \boldsymbol{v}_k^*, \boldsymbol{x}_k \mid \boldsymbol{y}_k^{obs}\right] + \tilde{\boldsymbol{y}}_k \hat{\boldsymbol{x}}_{k|N}^T \\
&= \mathrm{COV}_x\left[\boldsymbol{y}_k, \boldsymbol{x}_k\right] - \mathrm{COV}_x\left[\mathbf{Z}_k \boldsymbol{y}_k, \boldsymbol{x}_k\right] + \mathrm{COV}_x\left[\mathbf{Z}_k \mathbf{G}\boldsymbol{x}_k, \boldsymbol{x}_k\right] + \\
&\qquad \mathrm{COV}_x\left[\boldsymbol{v}_k^*, \boldsymbol{x}_k\right] + \tilde{\boldsymbol{y}}_k \hat{\boldsymbol{x}}_{k|N}^T \\
&= \mathbf{0} - \mathbf{0} + \mathbf{Z}_k \mathbf{G} \hat{\mathbf{V}}_{x|N} + \mathbf{0} + \tilde{\boldsymbol{y}}_k \hat{\boldsymbol{x}}_{k|N}^T \\
&= \mathbf{Z}_k \mathbf{G} \hat{\mathbf{V}}_{x|N} + \tilde{\boldsymbol{y}}_k \hat{\boldsymbol{x}}_{k|N}^T
\end{aligned} \tag{C.5}$$

where $\boldsymbol{v}_k^*$ is a random variable with mean $\mathbf{0}$ and covariance $\ddot{\Sigma} = \mathbf{R}_{22} - \mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{R}_{21}^T$ from Equation (**??**). $\boldsymbol{v}_k^*$ and $\boldsymbol{x}_k$ are independent of each other and thus $\mathrm{COV}\left[\boldsymbol{v}_k^*, \boldsymbol{x}_k\right] = \mathbf{0}$.

Appendix D

Derivation of VBM updates and evaluation of variational free energy for missing data approach

D.1 Derivation of update rules for VBM step

The principle for deriving the update rules of parameters $\mathbf{A}$, $\mathbf{G}$, $\mathbf{Q}$ and $\mathbf{R}$ follows the maximization expressions given by Eq.(4.27). For a parameter set under consideration, the expectation under its own variational distribution is ignored and the expectations with respect to variational distributions of other parameters are evaluated. Additionally, all terms independent of the parameter under consideration are subsumed in a proportionality constant.

(a) **Update for** $q^{(j+1)}(\mathbf{A}, \mathbf{Q})$
Using the log joint likelihood in Eq.(4.30) and ignoring the expectation w.r.t. $q^{(j)}(\mathbf{A}, \mathbf{Q})$, one obtains the following expression as a function of $\mathbf{A}$ and $\mathbf{Q}$

$$
\begin{aligned}
\ln q^{(j+1)}\left(\mathbf{A},\mathbf{Q}\right) \propto & -\frac{1}{2}\operatorname{tr}\left\{\mathbf{Q}^{-1}\left(\mathbf{S}_{ff}^{(j)}-\mathbf{A}\mathbf{S}_{fx}^{(j)^T}-\mathbf{S}_{fx}^{(j)}\mathbf{A}^T+\mathbf{A}\mathbf{S}_{xx}^{(j)}\mathbf{A}^T\right)\right\} \\
& -\frac{1}{2}\operatorname{tr}\left\{\mathbf{Q}^{-1}\left(\boldsymbol{\mu}_{\mathbf{A}}\boldsymbol{\Pi}^{-1}\boldsymbol{\mu}_{\mathbf{A}}^T-\mathbf{A}\boldsymbol{\Pi}^{-1}\boldsymbol{\mu}_{\mathbf{A}}^T-\boldsymbol{\mu}_{\mathbf{A}}\boldsymbol{\Pi}^{-1}\mathbf{A}^T+\mathbf{A}\boldsymbol{\Pi}^{-1}\mathbf{A}^T\right)\right\} \\
& -\frac{d_{\mathbf{Q}}+N+n_s+1}{2}\ln|\mathbf{Q}|-\frac{n_s}{2}\ln|\mathbf{Q}|-\frac{1}{2}\operatorname{tr}\left\{\mathbf{D}_{\mathbf{Q}}\mathbf{Q}^{-1}\right\} \\
\propto & -\frac{n_s}{2}\ln|\mathbf{Q}|-\frac{d_{\mathbf{Q}}+N+n_s+1}{2}\ln|\mathbf{Q}| \\
& -\frac{1}{2}\operatorname{tr}\left\{\left(\mathbf{A}\left(\mathbf{S}_{xx}^{(j)}+\boldsymbol{\Pi}^{-1}\right)\mathbf{A}^T-\mathbf{A}\left(\mathbf{S}_{fx}^{(j)}+\boldsymbol{\mu}_{\mathbf{A}}\boldsymbol{\Pi}^{-1}\right)^T\right)\mathbf{Q}^{-1}\right\} \\
& -\frac{1}{2}\operatorname{tr}\left\{\left(-\left(\mathbf{S}_{fx}^{(j)}+\boldsymbol{\mu}_{\mathbf{A}}\boldsymbol{\Pi}^{-1}\right)\mathbf{A}^T+\mathbf{D}_{\mathbf{Q}}+\mathbf{S}_{ff}^{(j)}+\boldsymbol{\mu}_{\mathbf{A}}\boldsymbol{\Pi}^{-1}\boldsymbol{\mu}_{\mathbf{A}}^T\right)\mathbf{Q}^{-1}\right\}
\end{aligned}
\tag{D.1}
$$

Adding and subtracting $\left(\mathbf{S}_{fx}^{(j)}+\boldsymbol{\mu}_{\mathbf{A}}\boldsymbol{\Pi}^{-1}\right)\left(\mathbf{S}_{xx}^{(j)}+\boldsymbol{\Pi}^{-1}\right)\left(\mathbf{S}_{fx}^{(j)}+\boldsymbol{\mu}_{\mathbf{A}}\boldsymbol{\Pi}^{-1}\right)^T$ inside the trace helps in achieving 'completion-of-squares' and one obtains the following

$$
\begin{aligned}
\ln q^{(j+1)}\left(\mathbf{A},\mathbf{Q}\right) \propto & -\frac{n_s}{2}\ln|\mathbf{Q}|-\frac{d_{\mathbf{Q}}+N+n_s+1}{2}\ln|\mathbf{Q}| \\
& -\frac{1}{2}\operatorname{tr}\Bigg\{\left(\mathbf{A}-\left(\mathbf{S}_{fx}^{(j)}+\boldsymbol{\mu}_{\mathbf{A}}\boldsymbol{\Pi}^{-1}\right)\left(\mathbf{S}_{xx}^{(j)}+\boldsymbol{\Pi}^{-1}\right)^{-1}\right)\left(\mathbf{S}_{xx}^{(j)}+\boldsymbol{\Pi}^{-1}\right)\times \\
& \qquad\left(\mathbf{A}-\left(\mathbf{S}_{fx}^{(j)}+\boldsymbol{\mu}_{\mathbf{A}}\boldsymbol{\Pi}^{-1}\right)\left(\mathbf{S}_{xx}^{(j)}+\boldsymbol{\Pi}^{-1}\right)^{-1}\right)^T\Bigg)\mathbf{Q}^{-1}\Bigg\} \\
\propto & -\frac{1}{2}\operatorname{tr}\left\{\left(\left(\mathbf{A}-\boldsymbol{\mu}_{\mathbf{A}}^{(j+1)}\right)\boldsymbol{\Pi}^{(j+1)^{-1}}\left(\mathbf{A}-\boldsymbol{\mu}_{\mathbf{A}}^{(j+1)}\right)^T\right)\mathbf{Q}^{-1}\right\}
\end{aligned}
\tag{D.2}
$$

where the expressions in the last line are given as

$$\boldsymbol{\Pi}^{(j+1)}=\left(\mathbf{S}_{xx}^{(j)}+\boldsymbol{\Pi}^{-1}\right)^{-1} \tag{D.3a}$$

$$\boldsymbol{\mu}_{\mathbf{A}}^{(j+1)}=\left(\mathbf{S}_{fx}^{(j)}+\boldsymbol{\mu}_{\mathbf{A}}\boldsymbol{\Pi}^{-1}\right)\boldsymbol{\Pi}^{(j+1)} \tag{D.3b}$$

$$d_{\mathbf{Q}}^{(j+1)}=d_{\mathbf{Q}}+N \tag{D.3c}$$

$$\mathbf{D}_{\mathbf{Q}}^{(j+1)}=\mathbf{D}_{\mathbf{Q}}+\mathbf{S}_{ff}^{(j)}+\boldsymbol{\mu}_{\mathbf{A}}\boldsymbol{\Pi}^{-1}\boldsymbol{\mu}_{\mathbf{A}}^T-\boldsymbol{\mu}_{\mathbf{A}}^{(j+1)}\boldsymbol{\Pi}^{(j+1)^{-1}}\boldsymbol{\mu}_{\mathbf{A}}^{(j+1)^T} \tag{D.3d}$$

(b) **Update for $q^{(j+1)}(\mathbf{G},\mathbf{R})$**
The update rule for $q^{(j+1)}(\mathbf{G},\mathbf{R})$ follows similarly as that for $q^{(j+1)}(\mathbf{A},\mathbf{Q})$ and one can derive the following:

$$\begin{aligned}\ln q^{(j+1)}(\mathbf{G},\mathbf{R}) \propto & -\frac{n_o}{2}\ln|\mathbf{R}| - \frac{d_{\mathbf{R}}+N+n_o+1}{2}\ln|\mathbf{R}| \\ & -\frac{1}{2}\mathrm{tr}\Bigg\{\left(\mathbf{G}-\left(\mathbf{S}_{yx}^{(j)}+\boldsymbol{\mu}_{\mathbf{G}}\boldsymbol{\Pi}^{-1}\right)\left(\mathbf{S}_{xx}^{(j)}+\boldsymbol{\Pi}^{-1}\right)^{-1}\right)\left(\mathbf{S}_{xx}^{(j)}+\boldsymbol{\Pi}^{-1}\right)\times \\ & \qquad \left(\mathbf{G}-\left(\mathbf{S}_{yx}^{(j)}+\boldsymbol{\mu}_{\mathbf{G}}\boldsymbol{\Pi}^{-1}\right)\left(\mathbf{S}_{xx}^{(j)}+\boldsymbol{\Pi}^{-1}\right)^{-1}\right)^{T}\Bigg)\mathbf{R}^{-1}\Bigg\} \\ \propto & -\frac{1}{2}\mathrm{tr}\left\{\left(\left(\mathbf{G}-\boldsymbol{\mu}_{\mathbf{G}}^{(j+1)}\right)\boldsymbol{\Pi}^{(j+1)^{-1}}\left(\mathbf{G}-\boldsymbol{\mu}_{\mathbf{G}}^{(j+1)}\right)^{T}\right)\mathbf{R}^{-1}\right\}\end{aligned} \tag{D.4}$$

where the expressions in the last line are given as

$$\boldsymbol{\mu}_{\mathbf{G}}^{(j+1)} = \left(\mathbf{S}_{yx}^{(j)}+\boldsymbol{\mu}_{\mathbf{G}}\boldsymbol{\Pi}^{-1}\right)\boldsymbol{\Pi}^{(j+1)} \tag{D.5a}$$

$$d_{\mathbf{R}}^{(j+1)} = d_{\mathbf{R}}+N \tag{D.5b}$$

$$\mathbf{D}_{\mathbf{R}}^{(j+1)} = \mathbf{D}_{\mathbf{R}}+\mathbf{S}_{yy}^{(j)}+\boldsymbol{\mu}_{\mathbf{A}}\boldsymbol{\Pi}^{-1}\boldsymbol{\mu}_{\mathbf{A}}^{T}-\boldsymbol{\mu}_{\mathbf{A}}^{(j+1)}\boldsymbol{\Pi}^{(j+1)^{-1}}\boldsymbol{\mu}_{\mathbf{A}}^{(j+1)^{T}} \tag{D.5c}$$

D.2 Evaluation of variational free energy

The derivation of the variational free energy mostly follows from [107]. The variational free energy is expressed as the sum of average energy $\mathcal{E}_{av}$ and the entropy $\mathcal{H}$ as shown in Eq.(4.41)

$$\mathcal{F}_{ve}^{(j)} = \mathcal{E}_{av}^{(j)}+\mathcal{H}^{(j)} \tag{D.6}$$

For conciseness, the superscript (j) on variational distribution $q(\cdot)$ has been dropped in the following derivations. The expression for average energy $\mathcal{E}_{av}$ can be written as follows:

$$\mathcal{E}_{av} = \langle \ln\left(p\left(\boldsymbol{Y},\boldsymbol{X},\mathbf{A},\mathbf{Q},\mathbf{G},\mathbf{R}\right)\right)\rangle_{q(\mathbf{A},\mathbf{Q})q(\mathbf{G},\mathbf{R})q(\boldsymbol{X})}$$

$$
\begin{aligned}
&= -\frac{n_s}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{V}_1| - \frac{1}{2}\left\langle (\boldsymbol{x}_1-\boldsymbol{\mu}_1)^T\mathbf{V}_1^{-1}(\boldsymbol{x}_1-\boldsymbol{\mu}_1)\right\rangle_{q(\boldsymbol{x}_1)} \\
&\quad + \sum_{k=1}^{N}\left[-\frac{n_s}{2}\ln 2\pi - \frac{1}{2}\left\langle\ln|\mathbf{Q}|\right\rangle_{q(\mathbf{Q})} - \frac{1}{2}\left\langle(\boldsymbol{x}_{k+1}-\mathbf{A}\boldsymbol{x}_k)^T\mathbf{Q}^{-1}(\boldsymbol{x}_{k+1}-\mathbf{A}\boldsymbol{x}_k)\right\rangle_{q(\mathbf{A},\mathbf{Q})q(\boldsymbol{X})}\right] \\
&\quad + \sum_{k=1}^{N}\left[-\frac{n_o}{2}\ln 2\pi - \frac{1}{2}\left\langle\ln|\mathbf{R}|\right\rangle_{q(\mathbf{R})} - \frac{1}{2}\left\langle(\boldsymbol{y}_k-\mathbf{G}\boldsymbol{x}_k)^T\mathbf{R}^{-1}(\boldsymbol{y}_k-\mathbf{G}\boldsymbol{x}_k)\right\rangle_{q(\mathbf{G},\mathbf{R})q(\boldsymbol{X})}\right] \\
&\quad - \frac{n_s^2}{2}\ln 2\pi - \frac{n_s}{2}\left\langle\ln|\mathbf{Q}|\right\rangle_{q(\mathbf{Q})} - \frac{n_s}{2}\ln|\boldsymbol{\Pi}| - \frac{1}{2}\left\langle\mathrm{tr}\left\{\boldsymbol{\Pi}^{-1}(\mathbf{A}-\boldsymbol{\mu}_\mathbf{A})^T\mathbf{Q}^{-1}(\mathbf{A}-\boldsymbol{\mu}_\mathbf{A})\right\}\right\rangle_{q(\mathbf{A},\mathbf{Q})} \\
&\quad - \frac{n_s n_o}{2}\ln 2\pi - \frac{n_s}{2}\left\langle\ln|\mathbf{R}|\right\rangle_{q(\mathbf{R})} - \frac{n_o}{2}\ln|\boldsymbol{\Pi}| - \frac{1}{2}\left\langle\mathrm{tr}\left\{\boldsymbol{\Pi}^{-1}(\mathbf{G}-\boldsymbol{\mu}_\mathbf{G})^T\mathbf{R}^{-1}(\mathbf{G}-\boldsymbol{\mu}_\mathbf{G})\right\}\right\rangle_{q(\mathbf{G},\mathbf{R})} \\
&\quad + \frac{d_\mathbf{Q}}{2}\ln|\mathbf{D}_\mathbf{Q}| - \frac{d_\mathbf{Q}+n_s+1}{2}\left\langle\ln|\mathbf{Q}|\right\rangle_{q(\mathbf{Q})} - \frac{d_\mathbf{Q}n_s}{2}\ln 2 - \ln\Gamma_{n_s}\left(\frac{d_\mathbf{Q}}{2}\right) - \frac{1}{2}\left\langle\mathrm{tr}\left\{\mathbf{D}_\mathbf{Q}\mathbf{Q}^{-1}\right\}\right\rangle_{q(\mathbf{Q})} \\
&\quad + \frac{d_\mathbf{R}}{2}\ln|\mathbf{D}_\mathbf{R}| - \frac{d_\mathbf{R}+n_o+1}{2}\left\langle\ln|\mathbf{R}|\right\rangle_{q(\mathbf{R})} - \frac{d_\mathbf{R}n_o}{2}\ln 2 - \ln\Gamma_{n_o}\left(\frac{d_\mathbf{R}}{2}\right) - \frac{1}{2}\left\langle\mathrm{tr}\left\{\mathbf{D}_\mathbf{R}\mathbf{R}^{-1}\right\}\right\rangle_{q(\mathbf{R})} \\
&= \frac{d_\mathbf{Q}+N+2n_s+1}{2}\left\langle\ln\left|\mathbf{Q}^{-1}\right|\right\rangle_{q(\mathbf{Q})} + \frac{d_\mathbf{R}+N+n_s+n_o+1}{2}\left\langle\ln\left|\mathbf{R}^{-1}\right|\right\rangle_{q(\mathbf{R})} \\
&\quad - \frac{1}{2}\mathrm{tr}\left\{\mathbf{V}_1^{-1}\left(\left\langle\boldsymbol{x}_1\boldsymbol{x}_1^T\right\rangle_{q(\boldsymbol{x}_1)} - \left\langle\boldsymbol{x}_1\right\rangle_{q(\boldsymbol{x}_1)}\boldsymbol{\mu}_1^T - \boldsymbol{\mu}_1\left\langle\boldsymbol{x}_1^T\right\rangle_{q(\boldsymbol{x}_1)} + \boldsymbol{\mu}_1\boldsymbol{\mu}_1^T\right)\right\} \\
&\quad - \frac{1}{2}\mathrm{tr}\left\{\left\langle\mathbf{Q}^{-1}\right\rangle_{q(\mathbf{Q})}\left(\boldsymbol{\mu}_\mathbf{A}\boldsymbol{\Pi}^{-1}\boldsymbol{\mu}_\mathbf{A}^T + \sum_{k=1}^{N}\left\langle\boldsymbol{x}_{k+1}\boldsymbol{x}_{k+1}^T\right\rangle_{q(\boldsymbol{X})} - \ldots\right.\right. \\
&\qquad\qquad \left.\left.\left\langle\mathbf{A}\right\rangle_{q(\mathbf{A}|\mathbf{Q})}\left(\sum_{k=1}^{N}\left\langle\boldsymbol{x}_k\boldsymbol{x}_{k+1}^T\right\rangle_{q(\boldsymbol{X})} + \boldsymbol{\Pi}^{-1}\boldsymbol{\mu}_\mathbf{A}^T\right)\right)\right\} \\
&\quad - \frac{1}{2}\mathrm{tr}\left\{-\left\langle\mathbf{Q}^{-1}\right\rangle_{q(\mathbf{Q})}\left(\sum_{k=1}^{N}\left\langle\boldsymbol{x}_{k+1}\boldsymbol{x}_k^T\right\rangle_{q(\boldsymbol{X})} + \boldsymbol{\mu}_\mathbf{A}\boldsymbol{\Pi}^{-1}\right)\left\langle\mathbf{A}^T\right\rangle_{q(\mathbf{A}|\mathbf{Q})}\right\} \\
&\quad - \frac{1}{2}\mathrm{tr}\left\{\left\langle\mathbf{Q}^{-1}\left\langle\mathbf{A}\left(\sum_{k=1}^{N}\left\langle\boldsymbol{x}_k\boldsymbol{x}_k^T\right\rangle_{q(\boldsymbol{X})} + \boldsymbol{\Pi}^{-1}\right)\mathbf{A}^T\right\rangle_{q(\mathbf{A}|\mathbf{Q})} + \mathbf{D}_\mathbf{Q}\right\rangle_{q(\mathbf{Q})}\right\}
\end{aligned}
$$

$$
\begin{aligned}
&-\frac{1}{2}\operatorname{tr}\left\{\left\langle \mathbf{R}^{-1}\right\rangle_{q(\mathbf{R})}\left(\boldsymbol{\mu}_{\mathbf{G}}\boldsymbol{\Pi}^{-1}\boldsymbol{\mu}_{\mathbf{G}}^{T}+\sum_{k=1}^{N}\left\langle \boldsymbol{y}_{k}\boldsymbol{y}_{k}^{T}\right\rangle_{q(\boldsymbol{X})}\cdots\right.\right.\\
&\qquad\left.\left.-\left\langle \mathbf{G}\right\rangle_{q(\mathbf{G}|\mathbf{R})}\left(\sum_{k=1}^{N}\left\langle \boldsymbol{x}_{k}\boldsymbol{y}_{k}^{T}\right\rangle_{q(\boldsymbol{X})}+\boldsymbol{\Pi}^{-1}\boldsymbol{\mu}_{\mathbf{G}}^{T}\right)\right)\right\}\\
&-\frac{1}{2}\operatorname{tr}\left\{-\left\langle \mathbf{R}^{-1}\right\rangle_{q(\mathbf{R})}\left(\sum_{k=1}^{N}\left\langle \boldsymbol{y}_{k}\boldsymbol{x}_{k}^{T}\right\rangle_{q(\boldsymbol{X})}+\boldsymbol{\mu}_{\mathbf{G}}\boldsymbol{\Pi}^{-1}\right)\left\langle \mathbf{G}^{T}\right\rangle_{q(\mathbf{G}|\mathbf{R})}\right\}\\
&-\frac{1}{2}\operatorname{tr}\left\{\left\langle \mathbf{R}^{-1}\left\langle \mathbf{G}\left(\sum_{k=1}^{N}\left\langle \boldsymbol{x}_{k}\boldsymbol{x}_{k}^{T}\right\rangle_{q(\boldsymbol{X})}+\boldsymbol{\Pi}^{-1}\right)\mathbf{G}^{T}\right\rangle_{q(\mathbf{G}|\mathbf{R})}+\mathbf{D}_{\mathbf{R}}\right\rangle_{q(\mathbf{R})}\right\}\\
&-\frac{n_{s}+Nn_{s}+Nn_{o}+n_{s}^{2}+n_{s}n_{o}}{2}\ln(2\pi)-\frac{d_{\mathbf{Q}}n_{s}+d_{\mathbf{R}}n_{o}}{2}\ln 2-\ln\Gamma_{n_{s}}\left(\frac{d_{\mathbf{Q}}}{2}\right)\\
&-\ln\Gamma_{n_{o}}\left(\frac{d_{\mathbf{R}}}{2}\right)-\frac{1}{2}\ln\left|\mathbf{V}_{1}\right|-\frac{n_{s}+n_{o}}{2}\ln\left|\boldsymbol{\Pi}\right|+\frac{d_{\mathbf{Q}}}{2}\ln\left|\mathbf{D}_{\mathbf{Q}}\right|+\frac{d_{\mathbf{R}}}{2}\ln\left|\mathbf{D}_{\mathbf{R}}\right|
\end{aligned}
$$

Note $\Gamma(\cdot)$ represents the 'gamma' function [109].
Next, use the following results

- Expected values of unobserved states and measurements

$$
\begin{aligned}
\mathbf{S}_{xx}^{(j)}&=\sum_{k=1}^{N}\left\langle \boldsymbol{x}_{k}\boldsymbol{x}_{k}^{T}\right\rangle_{q(\boldsymbol{X})} \quad \mathbf{S}_{fx}^{(j)}=\sum_{k=1}^{N}\left\langle \boldsymbol{x}_{k+1}\boldsymbol{x}_{k}^{T}\right\rangle_{q(\boldsymbol{X})} \quad \mathbf{S}_{ff}^{(j)}=\sum_{k=1}^{N}\left\langle \boldsymbol{x}_{k+1}\boldsymbol{x}_{k+1}^{T}\right\rangle_{q(\boldsymbol{X})}\\
\mathbf{S}_{yx}^{(j)}&=\sum_{k=1}^{N}\left\langle \boldsymbol{y}_{k}\boldsymbol{x}_{k}^{T}\right\rangle_{q(\boldsymbol{X})} \quad \mathbf{S}_{yy}^{(j)}=\sum_{k=1}^{N}\left\langle \boldsymbol{y}_{k}\boldsymbol{y}_{k}^{T}\right\rangle_{q(\boldsymbol{X})}
\end{aligned}
$$

- $\mathbf{Q}^{-1}$ and $\mathbf{R}^{-1}$ are Wishart distributed, therefore

$$
\begin{aligned}
\left\langle \mathbf{Q}^{-1}\right\rangle_{q(\mathbf{Q})}&=d_{\mathbf{Q}}^{(j)}\mathbf{D}_{\mathbf{Q}}^{(j)^{-1}}\\
\left\langle \mathbf{R}^{-1}\right\rangle_{q(\mathbf{R})}&=d_{\mathbf{R}}^{(j)}\mathbf{D}_{\mathbf{R}}^{(j)^{-1}}\\
\left\langle \ln\left|\mathbf{Q}^{-1}\right|\right\rangle_{q(\mathbf{Q})}&=n_{s}\ln 2+\ln\left|\mathbf{D}_{\mathbf{Q}}^{(j)^{-1}}\right|+\sum_{i=1}^{n_{s}}\gamma\left(\frac{d_{\mathbf{Q}}^{(j)}-i+1}{2}\right)\\
\left\langle \ln\left|\mathbf{R}^{-1}\right|\right\rangle_{q(\mathbf{R})}&=n_{o}\ln 2+\ln\left|\mathbf{D}_{\mathbf{R}}^{(j)^{-1}}\right|+\sum_{i=1}^{n_{o}}\gamma\left(\frac{d_{\mathbf{R}}^{(j)}-i+1}{2}\right)
\end{aligned}
$$

where $\gamma(\cdot)$ represents the 'digamma' function [109].

- Expected values of transition and observation matrices

$$\begin{aligned}
\langle \mathbf{A} \rangle_{q(\mathbf{A}|\mathbf{Q})} &= \boldsymbol{\mu}_{\mathbf{A}}^{(j)} \\
\langle \mathbf{G} \rangle_{q(\mathbf{G}|\mathbf{R})} &= \boldsymbol{\mu}_{\mathbf{G}}^{(j)} \\
\left\langle \mathbf{A}\mathbf{M}\mathbf{A}^T \right\rangle_{q(\mathbf{A}|\mathbf{Q})} &= \mathbf{Q}\ \mathrm{tr}\left\{\mathbf{M}\boldsymbol{\Pi}^{(j)}\right\} + \boldsymbol{\mu}_{\mathbf{A}}^{(j)}\mathbf{M}\boldsymbol{\mu}_{\mathbf{A}}^{(j)^T} \\
\left\langle \mathbf{G}\mathbf{M}\mathbf{G}^T \right\rangle_{q(\mathbf{G}|\mathbf{R})} &= \mathbf{R}\ \mathrm{tr}\left\{\mathbf{M}\boldsymbol{\Pi}^{(j)}\right\} + \boldsymbol{\mu}_{\mathbf{G}}^{(j)}\mathbf{M}\boldsymbol{\mu}_{\mathbf{G}}^{(j)^T}
\end{aligned}$$

So one obtains

$$\begin{aligned}
&\left\langle \mathbf{Q}^{-1} \left\langle \mathbf{A}\left(\mathbf{S}_{xx} + \boldsymbol{\Pi}^{-1}\right)\mathbf{A}^T \right\rangle_{q(\mathbf{A}|\mathbf{Q})} \right\rangle_{q(\mathbf{Q})} \\
&= \left\langle \mathbf{Q}^{-1}\left(\mathbf{Q}\ \mathrm{tr}\left\{\left(\mathbf{S}_{xx} + \boldsymbol{\Pi}^{-1}\right)\boldsymbol{\Pi}^{(j)}\right\} + \boldsymbol{\mu}_{\mathbf{A}}^{(j)}\left(\mathbf{S}_{xx} + \boldsymbol{\Pi}^{-1}\right)\boldsymbol{\mu}_{\mathbf{A}}^{(j)^T}\right)\right\rangle_{q(\mathbf{Q})} \\
&= \mathrm{tr}\left\{\left(\mathbf{S}_{xx} + \boldsymbol{\Pi}^{-1}\right)\boldsymbol{\Pi}^{(j)}\right\} + \left\langle \mathbf{Q}^{-1}\right\rangle_{q(\mathbf{Q})}\boldsymbol{\mu}_{\mathbf{A}}^{(j)}\left(\mathbf{S}_{xx} + \boldsymbol{\Pi}^{-1}\right)\boldsymbol{\mu}_{\mathbf{A}}^{(j)^T} \\
&\left\langle \mathbf{R}^{-1} \left\langle \mathbf{G}\left(\mathbf{S}_{xx} + \boldsymbol{\Pi}^{-1}\right)\mathbf{G}^T \right\rangle_{q(\mathbf{G}|\mathbf{R})} \right\rangle_{q(\mathbf{R})} \\
&= \left\langle \mathbf{R}^{-1}\left(\mathbf{R}\ \mathrm{tr}\left\{\left(\mathbf{S}_{xx} + \boldsymbol{\Pi}^{-1}\right)\boldsymbol{\Pi}^{(j)}\right\} + \boldsymbol{\mu}_{\mathbf{G}}^{(j)}\left(\mathbf{S}_{xx} + \boldsymbol{\Pi}^{-1}\right)\boldsymbol{\mu}_{\mathbf{G}}^{(j)^T}\right)\right\rangle_{q(\mathbf{R})} \\
&= \mathrm{tr}\left\{\left(\mathbf{S}_{xx} + \boldsymbol{\Pi}^{-1}\right)\boldsymbol{\Pi}^{(j)}\right\} + \left\langle \mathbf{R}^{-1}\right\rangle_{q(\mathbf{R})}\boldsymbol{\mu}_{\mathbf{G}}^{(j)}\left(\mathbf{S}_{xx} + \boldsymbol{\Pi}^{-1}\right)\boldsymbol{\mu}_{\mathbf{G}}^{(j)^T}
\end{aligned}$$

Substituting the above results into the derivation of $\mathcal{E}_{av}$, one gets

$$\begin{aligned}
\mathcal{E}_{av} &= \left\langle \ln\left(p\left(\boldsymbol{Y}, \boldsymbol{X}, \mathbf{A}, \mathbf{Q}, \mathbf{G}, \mathbf{R}\right)\right)\right\rangle_{q(\mathbf{A},\mathbf{Q})q(\mathbf{G},\mathbf{R})q(\boldsymbol{X})} \\
&= \frac{d_{\mathbf{Q}} + N + 2n_s + 1}{2}\left(n_s \ln 2 + \ln\left|\mathbf{D}_{\mathbf{Q}}^{(j)^{-1}}\right| + \sum_{i=1}^{n_s}\gamma\left(\frac{d_{\mathbf{Q}}^{(j)} - i + 1}{2}\right)\right) \\
&\quad + \frac{d_{\mathbf{R}} + N + n_s + n_o + 1}{2}\left(n_o \ln 2 + \ln\left|\mathbf{D}_{\mathbf{R}}^{(j)^{-1}}\right| + \sum_{i=1}^{n_o}\gamma\left(\frac{d_{\mathbf{R}}^{(j)} - i + 1}{2}\right)\right) \\
&\quad - \frac{1}{2}\mathrm{tr}\left\{\mathbf{V}_1^{-1}\left(\left\langle \boldsymbol{x}_1\boldsymbol{x}_1^T\right\rangle_{q(\boldsymbol{x}_1)} - \left\langle \boldsymbol{x}_1\right\rangle_{q(\boldsymbol{x}_1)}\boldsymbol{\mu}_1^T - \boldsymbol{\mu}_1\left\langle \boldsymbol{x}_1^T\right\rangle_{q(\boldsymbol{x}_1)} + \boldsymbol{\mu}_1\boldsymbol{\mu}_1^T\right)\right\} \\
&\quad - \frac{1}{2}\mathrm{tr}\left\{d_{\mathbf{Q}}^{(j)}\mathbf{D}_{\mathbf{Q}}^{(j)^{-1}}\left(\mathbf{D}_{\mathbf{Q}} + \boldsymbol{\mu}_{\mathbf{A}}\boldsymbol{\Pi}^{-1}\boldsymbol{\mu}_{\mathbf{A}}^T + \mathbf{S}_{ff}^{(j)} - \boldsymbol{\mu}_{\mathbf{A}}^{(j)}\left(\mathbf{S}_{fx}^{(j)} + \boldsymbol{\mu}_{\mathbf{A}}\boldsymbol{\Pi}^{-1}\right)^T - \ldots\right.\right. \\
&\qquad\qquad \left.\left.\left(\mathbf{S}_{fx}^{(j)} + \boldsymbol{\mu}_{\mathbf{A}}\boldsymbol{\Pi}^{-1}\right)\boldsymbol{\mu}_{\mathbf{A}}^{(j)^T}\right)\right\}
\end{aligned}$$

$$
\begin{aligned}
&- \frac{1}{2}\operatorname{tr}\left\{ d_{\mathbf{Q}}^{(j)} \mathbf{D}_{\mathbf{Q}}^{(j)^{-1}} \boldsymbol{\mu}_{\mathbf{A}}^{(j)} \left(\mathbf{S}_{xx} + \boldsymbol{\Pi}^{-1} \right) \boldsymbol{\mu}_{\mathbf{A}}^{(j)^T} \right\} - \frac{1}{2}\operatorname{tr}\left\{ \left(\mathbf{S}_{xx} + \boldsymbol{\Pi}^{-1} \right) \boldsymbol{\Pi}^{(j)} \right\} \\
&- \frac{1}{2}\operatorname{tr}\left\{ d_{\mathbf{R}}^{(j)} \mathbf{D}_{\mathbf{R}}^{(j)^{-1}} \left(\mathbf{D}_{\mathbf{R}} + \boldsymbol{\mu}_{\mathbf{G}} \boldsymbol{\Pi}^{-1} \boldsymbol{\mu}_{\mathbf{G}}^T + \mathbf{S}_{yy}^{(j)} - \boldsymbol{\mu}_{\mathbf{G}}^{(j)} \left(\mathbf{S}_{yx}^{(j)} + \boldsymbol{\mu}_{\mathbf{G}} \boldsymbol{\Pi}^{-1} \right)^T - \dots \right. \right. \\
&\qquad \left. \left. \left(\mathbf{S}_{yx}^{(j)} + \boldsymbol{\mu}_{\mathbf{G}} \boldsymbol{\Pi}^{-1} \right) \boldsymbol{\mu}_{\mathbf{G}}^{(j)^T} \right) \right\} \\
&- \frac{1}{2}\operatorname{tr}\left\{ d_{\mathbf{R}}^{(j)} \mathbf{D}_{\mathbf{R}}^{(j)^{-1}} \boldsymbol{\mu}_{\mathbf{G}}^{(j)} \left(\mathbf{S}_{xx} + \boldsymbol{\Pi}^{-1} \right) \boldsymbol{\mu}_{\mathbf{G}}^{(j)^T} \right\} - \frac{1}{2}\operatorname{tr}\left\{ \left(\mathbf{S}_{xx} + \boldsymbol{\Pi}^{-1} \right) \boldsymbol{\Pi}^{(j)} \right\} \\
&- \frac{n_s + N n_s + N n_o + n_s^2 + n_s n_o}{2} \ln(2\pi) - \frac{d_{\mathbf{Q}} n_s + d_{\mathbf{R}} n_o}{2} \ln 2 - \ln \Gamma_{n_s}\left(\frac{d_{\mathbf{Q}}}{2} \right) \\
&- \ln \Gamma_{n_o}\left(\frac{d_{\mathbf{R}}}{2} \right) - \frac{1}{2} \ln |\mathbf{V}_1| - \frac{n_s + n_o}{2} \ln |\boldsymbol{\Pi}| + \frac{d_{\mathbf{Q}}}{2} \ln |\mathbf{D}_{\mathbf{Q}}| + \frac{d_{\mathbf{R}}}{2} \ln |\mathbf{D}_{\mathbf{R}}|
\end{aligned}
$$

Concisely, the expression for $\mathcal{E}_{av}^{(j)}$ can be written as

$$
\begin{aligned}
\mathcal{E}_{av}^{(j)} = \frac{1}{2} \Bigg[& - c_3 \ln(2\pi) + (c_4 n_s + c_5 n_o - c_6) \ln 2 - 2 \ln \Gamma_{n_s}\left(\frac{d_{\mathbf{Q}}}{2} \right) - 2 \ln \Gamma_{n_o}\left(\frac{d_{\mathbf{R}}}{2} \right) \\
& - \ln |\mathbf{V}_1| - (n_s + n_o) \ln \left| \boldsymbol{\Pi}^{-1} \right| + d_{\mathbf{Q}} \ln |\mathbf{D}_{\mathbf{Q}}| + d_{\mathbf{R}} \ln |\mathbf{D}_{\mathbf{R}}| \\
& + c_4 \sum_{i=1}^{n_s} \gamma\left(\frac{d_{\mathbf{Q}}^{(j)} - i + 1}{2} \right) + c_5 \sum_{i=1}^{n_o} \gamma\left(\frac{d_{\mathbf{R}}^{(j)} - i + 1}{2} \right) - 2\operatorname{tr}\left\{ \left(\mathbf{S}_{xx} + \boldsymbol{\Pi}^{-1} \right) \boldsymbol{\Pi}^{(j)} \right\} \\
& - c_4 \ln \left| \mathbf{D}_{\mathbf{Q}}^{(j)} \right| - c_5 \ln \left| \mathbf{D}_{\mathbf{R}}^{(j)} \right| - \operatorname{tr}\left\{ \mathbf{V}_1^{-1} \mathbf{E}_1 + d_{\mathbf{Q}}^{(j)} \mathbf{D}_{\mathbf{Q}}^{(j)^{-1}} \mathbf{E}_2 + d_{\mathbf{R}}^{(j)} \mathbf{D}_{\mathbf{R}}^{(j)^{-1}} \mathbf{E}_3 \right\} \quad \Bigg]
\end{aligned}
\tag{D.7}
$$

where

$$
\begin{aligned}
c_3 &= n_s + N n_s + N n_o + n_s^2 + n_s n_o = n_s + (N + n_s)(n_s + n_o) \\
c_4 &= d_{\mathbf{Q}} + N + 2 n_s + 1 \\
c_5 &= d_{\mathbf{R}} + N + n_s + n_o + 1 \\
c_6 &= d_{\mathbf{Q}} n_s + d_{\mathbf{R}} n_o \\
\mathbf{E}_1 &= \left\langle \boldsymbol{x}_1 \boldsymbol{x}_1^T \right\rangle_{q(\boldsymbol{x}_1)^{(j)}} - \left\langle \boldsymbol{x}_1 \right\rangle_{q(\boldsymbol{x}_1)^{(j)}} \boldsymbol{\mu}_1^T - \boldsymbol{\mu}_1 \left\langle \boldsymbol{x}_1^T \right\rangle_{q(\boldsymbol{x}_1)^{(j)}} + \boldsymbol{\mu}_1 \boldsymbol{\mu}_1^T \\
&= \left(\hat{\mathbf{V}}_{1|N} + \hat{\boldsymbol{x}}_{1|N} \hat{\boldsymbol{x}}_{1|N}^T \right) - \hat{\boldsymbol{x}}_{1|N} \boldsymbol{\mu}_1^T - \boldsymbol{\mu}_1 \hat{\boldsymbol{x}}_{1|N}^T + \boldsymbol{\mu}_1 \boldsymbol{\mu}_1^T
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{E}_2 &= \mathbf{D}_{\mathbf{Q}} + \boldsymbol{\mu}_{\mathbf{A}}\boldsymbol{\Pi}^{-1}\boldsymbol{\mu}_{\mathbf{A}}^T + \mathbf{S}_{ff}^{(j)} - \boldsymbol{\mu}_{\mathbf{A}}^{(j)}\left(\mathbf{S}_{fx}^{(j)} + \boldsymbol{\mu}_{\mathbf{A}}\boldsymbol{\Pi}^{-1}\right)^T - \left(\mathbf{S}_{fx}^{(j)} + \boldsymbol{\mu}_{\mathbf{A}}\boldsymbol{\Pi}^{-1}\right)\boldsymbol{\mu}_{\mathbf{A}}^{(j)^T} + \dots \\
&\quad \boldsymbol{\mu}_{\mathbf{A}}^{(j)}\left(\mathbf{S}_{xx}^{(j)} + \boldsymbol{\Pi}^{-1}\right)\boldsymbol{\mu}_{\mathbf{A}}^{(j)^T} \\
\mathbf{E}_3 &= \mathbf{D}_{\mathbf{R}} + \boldsymbol{\mu}_{\mathbf{G}}\boldsymbol{\Pi}^{-1}\boldsymbol{\mu}_{\mathbf{G}}^T + \mathbf{S}_{yy}^{(j)} - \boldsymbol{\mu}_{\mathbf{G}}^{(j)}\left(\mathbf{S}_{yx}^{(j)} + \boldsymbol{\mu}_{\mathbf{G}}\boldsymbol{\Pi}^{-1}\right)^T - \left(\mathbf{S}_{yx}^{(j)} + \boldsymbol{\mu}_{\mathbf{G}}\boldsymbol{\Pi}^{-1}\right)\boldsymbol{\mu}_{\mathbf{G}}^{(j)^T} + \dots \\
&\quad \boldsymbol{\mu}_{\mathbf{G}}^{(j)}\left(\mathbf{S}_{xx}^{(j)} + \boldsymbol{\Pi}^{-1}\right)\boldsymbol{\mu}_{\mathbf{G}}^{(j)^T}
\end{aligned}
\tag{D.8}
$$

Next we calculate the differential entropy term $\mathcal{H}$. For the entropy term

$$
\mathcal{H}^{(j)} = \mathcal{H}\left(q(\mathbf{A}|\mathbf{Q})q(\mathbf{Q})q(\mathbf{G}|\mathbf{R})q(\mathbf{R})q(\boldsymbol{X})\right) \tag{D.9}
$$

Using the additivity property of differential entropy for independent variables ([110], pp. 253), the following can be written

$$
\mathcal{H}^{(j)} = \mathcal{H}\left(q^{(j)}(\mathbf{A}|\mathbf{Q})\right) + \mathcal{H}\left(q^{(j)}(\mathbf{Q})\right) + \mathcal{H}\left(q^{(j)}(\mathbf{G}|\mathbf{R})\right) + \mathcal{H}\left(q^{(j)}(\mathbf{R})\right) + \mathcal{H}\left(q^{(j)}(\boldsymbol{X})\right) \tag{D.10}
$$

and the expression for the first four terms are available as

$$
\mathcal{H}\left(q^{(j)}(\mathbf{A}|\mathbf{Q})\right) = \frac{1}{2}\ln\left|\boldsymbol{\Pi}^{(j)} \otimes \mathbf{Q}^{(j)}\right| + \frac{n_s^2}{2}\left(1 + \ln(2\pi)\right) \tag{D.11a}
$$

$$
\mathcal{H}\left(q^{(j)}(\mathbf{Q})\right) = \ln\Gamma_{n_s}\left(\frac{d_{\mathbf{Q}}^{(j)}}{2}\right) + \frac{d_{\mathbf{Q}}^{(j)} n_s}{2} + \frac{n_s+1}{2}\ln\left|\frac{\mathbf{D}_{\mathbf{Q}}^{(j)}}{2}\right| - \frac{d_{\mathbf{Q}}^{(j)} + n_s + 1}{2}\sum_{i=1}^{n_s}\gamma\left(\frac{d_{\mathbf{Q}}^{(j)} - i + 1}{2}\right) \tag{D.11b}
$$

$$
\mathcal{H}\left(q^{(j)}(\mathbf{G}|\mathbf{R})\right) = \frac{1}{2}\ln\left|\boldsymbol{\Pi}^{(j)} \otimes \mathbf{R}^{(j)}\right| + \frac{n_s n_o}{2}\left(1 + \ln(2\pi)\right) \tag{D.11c}
$$

$$
\mathcal{H}\left(q^{(j)}(\mathbf{R})\right) = \ln\Gamma_{n_o}\left(\frac{d_{\mathbf{R}}^{(j)}}{2}\right) + \frac{d_{\mathbf{R}}^{(j)} n_o}{2} + \frac{n_o+1}{2}\ln\left|\frac{\mathbf{D}_{\mathbf{R}}^{(j)}}{2}\right| - \frac{d_{\mathbf{R}}^{(j)} + n_o + 1}{2}\sum_{i=1}^{n_o}\gamma\left(\frac{d_{\mathbf{R}}^{(j)} - i + 1}{2}\right) \tag{D.11d}
$$

The differential entropy of inverse Wishart distribution can be found in [141].
$\mathcal{H}\left(q^{(j)}(\boldsymbol{X})\right)$ can be evaluated by considering the factorization property of $q(\boldsymbol{X})$ as

$$
q^{(j)}(\boldsymbol{X}) = p\left(\boldsymbol{x}_{1:N+1}|\widetilde{\boldsymbol{y}}_{1:N}\right) = p\left(\boldsymbol{x}_1|\widetilde{\boldsymbol{y}}_{1:N}\right)\prod_{k=1}^{N} p\left(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k, \widetilde{\boldsymbol{y}}_{1:N}\right)
$$

$$= p\left(\boldsymbol{x}_1|\widetilde{\boldsymbol{y}}_{1:N}\right)\prod_{k=1}^{N}\frac{p\left(\boldsymbol{x}_{k+1},\boldsymbol{x}_k|\widetilde{\boldsymbol{y}}_{1:N}\right)}{p\left(\boldsymbol{x}_k|\widetilde{\boldsymbol{y}}_{1:N}\right)}$$

where $\widetilde{\boldsymbol{y}}_k$ is shown in Eq.(4.33). The differential entropy for $q(\boldsymbol{X})$ then becomes

$$\begin{aligned}\mathcal{H}\left(q^{(j)}(\boldsymbol{X})\right) &= \mathcal{H}\left(p\left(\boldsymbol{x}_1|\widetilde{\boldsymbol{y}}_{1:N}\right)\prod_{k=1}^{N}\frac{p\left(\boldsymbol{x}_{k+1},\boldsymbol{x}_k|\widetilde{\boldsymbol{y}}_{1:N}\right)}{p\left(\boldsymbol{x}_k|\widetilde{\boldsymbol{y}}_{1:N}\right)}\right)\\ &= \mathcal{H}\left(p\left(\boldsymbol{x}_1|\widetilde{\boldsymbol{y}}_{1:N}\right)\right)+\sum_{k=1}^{N}\mathcal{H}\left(p\left(\boldsymbol{x}_{k+1},\boldsymbol{x}_k|\widetilde{\boldsymbol{y}}_{1:N}\right)\right)-\sum_{k=1}^{N}\mathcal{H}\left(p\left(\boldsymbol{x}_k|\widetilde{\boldsymbol{y}}_{1:N}\right)\right)\end{aligned}$$

Since the mean and covariances are obtained from KRTS smoother results as

$$\begin{aligned}p\left(\boldsymbol{x}_1|\widetilde{\boldsymbol{y}}_{1:N}\right) &= \mathcal{N}\left(\hat{\boldsymbol{x}}_{1|N},\hat{\mathbf{V}}_{1|N}\right)\\ p\left(\boldsymbol{x}_{k+1},\boldsymbol{x}_k|\widetilde{\boldsymbol{y}}_{1:N}\right) &= \mathcal{N}\left(\begin{bmatrix}\hat{\boldsymbol{x}}_{k+1|N}\\ \hat{\boldsymbol{x}}_{k|N}\end{bmatrix},\begin{bmatrix}\hat{\mathbf{V}}_{k+1|N} & \hat{\mathbf{V}}_{k+1,k|N}\\ \hat{\mathbf{V}}_{k+1,k|N}^T & \hat{\mathbf{V}}_{k|N}\end{bmatrix}\right)\\ p\left(\boldsymbol{x}_k|\widetilde{\boldsymbol{y}}_{1:N}\right) &= \mathcal{N}\left(\hat{\boldsymbol{x}}_{k|N},\hat{\mathbf{V}}_{k|N}\right)\end{aligned}$$

one can write

$$\begin{aligned}\mathcal{H}\left(p\left(\boldsymbol{x}_1|\widetilde{\boldsymbol{y}}_{1:N}\right)\right) &= \frac{1}{2}\ln\left|\hat{\mathbf{V}}_{1|N}\right|+\frac{n_s}{2}(1+\ln(2\pi))\\ \mathcal{H}\left(p\left(\boldsymbol{x}_{k+1},\boldsymbol{x}_k|\widetilde{\boldsymbol{y}}_{1:N}\right)\right) &= \frac{1}{2}\ln\left|\begin{bmatrix}\hat{\mathbf{V}}_{k+1|N} & \hat{\mathbf{V}}_{k+1,k|N}\\ \hat{\mathbf{V}}_{k+1,k|N}^T & \hat{\mathbf{V}}_{k|N}\end{bmatrix}\right|+\frac{n_s^2}{2}(1+\ln(2\pi))\\ \mathcal{H}\left(p\left(\boldsymbol{x}_k|\widetilde{\boldsymbol{y}}_{1:N}\right)\right) &= \frac{1}{2}\ln\left|\hat{\mathbf{V}}_{k|N}\right|+\frac{n_s}{2}(1+\ln(2\pi))\end{aligned}$$

Thus, the differential entropy at the jth iteration can be written as

$$\begin{aligned}\mathcal{H}\left(q^{(j)}(\boldsymbol{X})\right) &= \frac{1}{2}\left[\ln\left|\hat{\mathbf{V}}_{1|N}\right|+\sum_{k=1}^{N}\left(\ln\left|\begin{bmatrix}\hat{\mathbf{V}}_{k+1|N} & \hat{\mathbf{V}}_{k+1,k|N}\\ \hat{\mathbf{V}}_{k+1,k|N}^T & \hat{\mathbf{V}}_{k|N}\end{bmatrix}\right|-\ln\left|\hat{\mathbf{V}}_{k|N}\right|\right)+c_7\right]\\ &= \frac{1}{2}\left[\ln\left|\hat{\mathbf{V}}_{1|N}\right|+\sum_{k=1}^{N}\left(\ln\left|\hat{\mathbf{V}}_{k+1|N}-\hat{\mathbf{V}}_{k+1,k|N}\hat{\mathbf{V}}_{k|N}^{-1}\hat{\mathbf{V}}_{k+1,k|N}^T\right|\right)+c_7\right] \quad \text{(D.12a)}\end{aligned}$$

where the constant c_7 is

$$c_7 = \sum_{k=1}^{N} n_s^2(1+\ln(2\pi)) - \sum_{k=2}^{N} n_s(1+\ln(2\pi)) = \left(Nn_s^2 - (N-1)n_s\right)(1+\ln(2\pi))$$

and the following result of determinant of block matrices was used

$$\ln\left|\begin{bmatrix} \hat{\mathbf{V}}_{k+1|N} & \hat{\mathbf{V}}_{k+1,k|N} \\ \hat{\mathbf{V}}_{k+1,k|N}^T & \hat{\mathbf{V}}_{k|N} \end{bmatrix}\right| = \ln\left|\hat{\mathbf{V}}_{k+1|N} - \hat{\mathbf{V}}_{k+1,k|N}\hat{\mathbf{V}}_{k|N}^{-1}\hat{\mathbf{V}}_{k+1,k|N}^T\right| + \ln\left|\hat{\mathbf{V}}_{k|N}\right|.$$

Appendix E

Derivation of updates for EM and VB in stacked data approach

E.1 Derivation of update rules for M-step

The update equations of EM are obtained by taking gradient of $\overline{\mathcal{F}}\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)}\right)$ in Eq.(5.12) with respect to the parameters $\boldsymbol{\theta}$ at iteration j and then setting them to zero. This operation leads to closed form M-step maximizations. Despite the new likelihood function for the MSM in Equation (5.10), the updating equations for $\overline{\mathbf{A}}^{(j+1)}, \overline{\mathbf{Q}}^{(j+1)}$ and $\overline{\boldsymbol{\mu}}_1^{(j+1)}, \overline{\mathbf{V}}_1^{(j+1)}$ end up being identical to those in Eq.(5.12) and (4.18) as they are not involved in the observation equation. However, the updating formula for $\overline{\mathbf{G}}^{(j+1)}$ and $\mathbf{R}^{(j+1)}$ gets modified due to incorporation of MSR matrices in the observation equation of the MSM.

(a) **Update for $\overline{\mathbf{G}}^{(j+1)}$**

To derive the update rule for $\overline{\mathbf{G}}^{(j+1)}$, take the gradient of $\overline{\mathcal{F}}\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)}\right)$ w.r.t. $\overline{\mathbf{G}}$ and set it to zero

$$\frac{\partial \overline{\mathcal{F}}\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)}\right)}{\partial \overline{\mathbf{G}}} = \frac{\partial}{\partial \overline{\mathbf{G}}} \left\langle -\frac{1}{2}\sum_{k=1}^{N} \left(\boldsymbol{y}_k - \boldsymbol{\Omega}_k \overline{\mathbf{G}} \overline{\boldsymbol{x}}_k\right)^T \mathbf{R}^{-1} \left(\boldsymbol{y}_k - \boldsymbol{\Omega}_k \overline{\mathbf{G}} \overline{\boldsymbol{x}}_k\right) \right\rangle = \mathbf{0}$$

Using the expressions in (70), (71) and (82) from the Matrix Cookbook [142], one obtains

$$\frac{\partial \overline{\mathcal{F}}\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)}\right)}{\partial \overline{\mathbf{G}}} = \sum_{k=1}^{N}\left(\boldsymbol{\Omega}_k^T\mathbf{R}^{-1}\boldsymbol{y}_k\left\langle\bar{\boldsymbol{x}}^T\right\rangle - \boldsymbol{\Omega}_k^T\mathbf{R}^{-1}\boldsymbol{\Omega}_k\overline{\mathbf{G}}\left\langle\bar{\boldsymbol{x}}_k\bar{\boldsymbol{x}}_k^T\right\rangle\right) = \mathbf{0}$$

Vectorizing the above equation and using the formula $\operatorname{vec}(\mathcal{ABC}) = \left(\mathcal{C}^T\otimes\mathcal{A}\right)\operatorname{vec}(\mathcal{B})$

$$\begin{aligned}
&\operatorname{vec}\left(\sum_{k=1}^{N}\left(\boldsymbol{\Omega}_k^T\mathbf{R}^{-1}\boldsymbol{y}_k\left\langle\bar{\boldsymbol{x}}_k^T\right\rangle - \boldsymbol{\Omega}_k^T\mathbf{R}^{-1}\boldsymbol{\Omega}_k\overline{\mathbf{G}}\left\langle\bar{\boldsymbol{x}}_k\bar{\boldsymbol{x}}_k^T\right\rangle\right)\right) = \mathbf{0}\\
\implies &\sum_{k=1}^{N}\operatorname{vec}\left(\boldsymbol{\Omega}_k^T\mathbf{R}^{-1}\boldsymbol{y}_k\left\langle\bar{\boldsymbol{x}}_k^T\right\rangle\right) - \sum_{k=1}^{N}\operatorname{vec}\left(\boldsymbol{\Omega}_k^T\mathbf{R}^{-1}\boldsymbol{\Omega}_k\overline{\mathbf{G}}\left\langle\bar{\boldsymbol{x}}_k\bar{\boldsymbol{x}}_k^T\right\rangle\right) = \mathbf{0}\\
\implies &\sum_{k=1}^{N}\operatorname{vec}\left(\boldsymbol{\Omega}_k^T\mathbf{R}^{-1}\boldsymbol{y}_k\left\langle\bar{\boldsymbol{x}}_k^T\right\rangle\right) - \sum_{k=1}^{N}\left(\left\langle\bar{\boldsymbol{x}}_k\bar{\boldsymbol{x}}_k^T\right\rangle\otimes\left(\boldsymbol{\Omega}_k^T\mathbf{R}^{-1}\boldsymbol{\Omega}_k\right)\right)\operatorname{vec}\left(\overline{\mathbf{G}}\right) = \mathbf{0}
\end{aligned}$$

In the above equations, the individual terms are vectorized and the Kronecker product '$\otimes$' is implemented to extract the $\overline{\mathbf{G}}$ which is sandwiched between two terms. Finally, the update equation for $\overline{\mathbf{G}}$ is obtained as given in Eq.(5.19a)

$$\begin{aligned}
\operatorname{vec}\left(\overline{\mathbf{G}}^{(j+1)}\right) &= \left(\sum_{k=1}^{N}\left(\left\langle\bar{\boldsymbol{x}}_k\bar{\boldsymbol{x}}_k^T\right\rangle\otimes\left(\boldsymbol{\Omega}_k^T{\mathbf{R}^{(j)}}^{-1}\boldsymbol{\Omega}_k\right)\right)\right)^{-1}\operatorname{vec}\left(\sum_{k=1}^{N}\left(\boldsymbol{\Omega}_k^T{\mathbf{R}^{(j)}}^{-1}\boldsymbol{y}_k\left\langle\bar{\boldsymbol{x}}^T\right\rangle\right)\right)\\
\implies \bar{\boldsymbol{g}}^{(j+1)} &= \left(\sum_{k=1}^{N}\left(\bar{\mathbf{S}}_{xx,k}\otimes\left(\boldsymbol{\Omega}_k^T{\mathbf{R}^{(j)}}^{-1}\boldsymbol{\Omega}_k\right)\right)\right)^{-1}\operatorname{vec}\left(\sum_{k=1}^{N}\boldsymbol{\Omega}_k^T{\mathbf{R}^{(j)}}^{-1}\bar{\mathbf{S}}_{yx,k}\right)
\end{aligned} \tag{E.1}$$

(b) **Update for $\mathbf{R}^{(j+1)}$**

Next, to obtain the update rule for $\mathbf{R}^{(j+1)}$, the gradient of the expression in Equation (5.12) with respect to $\mathbf{R}^{-1}$ is set to zero, as shown below

$$\begin{aligned}
\frac{\partial}{\partial\mathbf{R}^{-1}}\Bigg(\frac{N}{2}\log\left|\mathbf{R}^{-1}\right| - \frac{1}{2}\sum_{k=1}^{N}\operatorname{tr}\Big\{\mathbf{R}^{-1}\big(\bar{\mathbf{S}}_{yy,k} - \boldsymbol{\Omega}_k\overline{\mathbf{G}}\bar{\mathbf{S}}_{yx,k}^T - \ldots\\
\bar{\mathbf{S}}_{yx,k}\overline{\mathbf{G}}^T\boldsymbol{\Omega}_k^T + \boldsymbol{\Omega}_k\overline{\mathbf{G}}\bar{\mathbf{S}}_{xx,k}\overline{\mathbf{G}}^T\boldsymbol{\Omega}_k^T\big)\Big\}\Bigg) = \mathbf{0}
\end{aligned}$$

Using the derivative formula (49) and (100) from Matrix Cookbook [142], one gets Eq.(5.19c)

$$\mathbf{R}^{(j+1)} = \frac{1}{N}\sum_{k=1}^{N}\left(\bar{\mathbf{S}}_{yy,k} - \boldsymbol{\Omega}_k\overline{\mathbf{G}}^{(j+1)}\bar{\mathbf{S}}_{yx,k}^T - \bar{\mathbf{S}}_{yx,k}\overline{\mathbf{G}}^{(j+1)^T}\boldsymbol{\Omega}_k^T + \boldsymbol{\Omega}_k\overline{\mathbf{G}}^{(j+1)}\bar{\mathbf{S}}_{xx,k}\overline{\mathbf{G}}^{(j+1)^T}\boldsymbol{\Omega}_k^T\right) \tag{E.2}$$

E.2 Computation of $\overline{\mathbf{V}}^{\bar{\boldsymbol{a}}}$ and $\overline{\mathbf{V}}_k^{\bar{\boldsymbol{g}}}$

The matrices $\overline{\mathbf{V}}^{\bar{\boldsymbol{a}}}$ and $\overline{\mathbf{V}}_k^{\bar{\boldsymbol{g}}}$ represent weighted covariances whose values can be evaluated using $\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}$ and $\boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}$ respectively. The evaluation of both $\overline{\mathbf{V}}^{\bar{\boldsymbol{a}}}$ and $\overline{\mathbf{V}}_k^{\bar{\boldsymbol{g}}}$ follow similar procedure and therefore, only the evaluation of $\overline{\mathbf{V}}_k^{\bar{\boldsymbol{g}}}$ is shown for the sake of brevity. Using the property, $\text{vec}\left(\mathcal{ABC}\right) = \left(\mathcal{C}^T \otimes \mathcal{A}\right)\text{vec}\left(\mathcal{B}\right)$, the following can be written

$$\begin{aligned}\text{vec}\left(\overline{\mathbf{V}}_k^{\bar{\boldsymbol{g}}}\right) &= \left\langle \text{vec}\left(\left(\overline{\mathbf{G}} - \overline{\overline{\mathbf{G}}}\right)^T \boldsymbol{\Omega}_k^T\overline{\mathbf{R}}^{-1}\boldsymbol{\Omega}_k\left(\overline{\mathbf{G}} - \overline{\overline{\mathbf{G}}}\right)\right)\right\rangle_{q^{(j)}(\bar{\boldsymbol{g}})} \\ &= \left\langle \left(\overline{\mathbf{G}} - \overline{\overline{\mathbf{G}}}\right)^T \otimes \left(\overline{\mathbf{G}} - \overline{\overline{\mathbf{G}}}\right)^T\right\rangle_{q^{(j)}(\bar{\boldsymbol{g}})} \text{vec}\left(\boldsymbol{\Omega}_k^T\overline{\mathbf{R}}^{-1}\boldsymbol{\Omega}_k\right)\end{aligned} \tag{E.3}$$

The variational expectation over $\left(\overline{\mathbf{G}} - \overline{\overline{\mathbf{G}}}\right)^T \otimes \left(\overline{\mathbf{G}} - \overline{\overline{\mathbf{G}}}\right)^T$ with respect to $q^{(j)}\left(\bar{\boldsymbol{g}}\right)$ is related to the covariance matrix of $\overline{\mathbf{G}}^T$ at the jth VB iteration. With the knowledge of the covariance matrix of $\bar{\boldsymbol{g}}$ $(= \text{vec}\left(\overline{\mathbf{G}}\right))$ at the jth VB iteration, it is required to obtain the covariance matrix of $\bar{\boldsymbol{g}}'$ $(= \text{vec}\left(\overline{\mathbf{G}}^T\right))$. For this, a permutation matrix $\mathbf{T}_{\bar{\boldsymbol{g}}}$ is defined such that $\bar{\boldsymbol{g}}' = \mathbf{T}_{\bar{\boldsymbol{g}}}\bar{\boldsymbol{g}}$. The covariance matrix for $\bar{\boldsymbol{g}}'$ can then be obtained as $\boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}'}^{(j)} = \mathbf{T}_{\bar{\boldsymbol{g}}}\boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}}^{(j)}\mathbf{T}_{\bar{\boldsymbol{g}}}^T$. Further, $\boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}'}^{(j)}$ can be expressed in terms of block matrices, each of the size $n_\alpha \times n_\alpha$:

$$\boldsymbol{\Sigma}_{\bar{\boldsymbol{g}}'}^{(j)} = \begin{bmatrix} \boldsymbol{\Pi}_{11}^{\bar{\boldsymbol{g}}'} & \boldsymbol{\Pi}_{12}^{\bar{\boldsymbol{g}}'} & \dots & \boldsymbol{\Pi}_{1n_s}^{\bar{\boldsymbol{g}}'} \\ \boldsymbol{\Pi}_{21}^{\bar{\boldsymbol{g}}'} & \boldsymbol{\Pi}_{22}^{\bar{\boldsymbol{g}}'} & \dots & \boldsymbol{\Pi}_{2n_s}^{\bar{\boldsymbol{g}}'} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Pi}_{n_s1}^{\bar{\boldsymbol{g}}'} & \boldsymbol{\Pi}_{n_s2}^{\bar{\boldsymbol{g}}'} & \dots & \boldsymbol{\Pi}_{n_sn_s}^{\bar{\boldsymbol{g}}'} \end{bmatrix} \tag{E.4}$$

Using the block matrices, the expression in Eq.(E.3) can be expressed as a matrix multiplication

$$\text{vec}\left(\overline{\mathbf{V}}_k^{\bar{\boldsymbol{g}}}\right) = \begin{bmatrix}\text{vec}\left(\boldsymbol{\Pi}_{11}^{\bar{\boldsymbol{g}}'}\right) & \text{vec}\left(\boldsymbol{\Pi}_{21}^{\bar{\boldsymbol{g}}'}\right) & \dots & \text{vec}\left(\boldsymbol{\Pi}_{n_sn_s}^{\bar{\boldsymbol{g}}'}\right)\end{bmatrix}\text{vec}\left(\boldsymbol{\Omega}_k^T\overline{\mathbf{R}}^{-1}\boldsymbol{\Omega}_k\right) \tag{E.5}$$

which is useful for numerical implementation. Alternatively, $\overline{\mathbf{V}}_k^{\bar{\boldsymbol{g}}}$ can be written in the summation form as

$$\overline{\mathbf{V}}_k^{\bar{\boldsymbol{g}}} = \sum_{p=1}^{n_s}\sum_{r=1}^{n_s} \left(\boldsymbol{\Omega}_k^T \overline{\mathbf{R}}^{-1} \boldsymbol{\Omega}_k\right)_{pr} \boldsymbol{\Pi}_{pr}^{\bar{\boldsymbol{g}}'} \tag{E.6}$$

Following a similar procedure for $\overline{\mathbf{V}}^{\bar{\boldsymbol{a}}}$, one can define $\bar{\boldsymbol{a}}' = \mathbf{T}_{\bar{\boldsymbol{a}}}\bar{\boldsymbol{a}}$ and obtain $\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}'}^{(j)} = \mathbf{T}_{\bar{\boldsymbol{a}}}\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{(j)}\mathbf{T}_{\bar{\boldsymbol{a}}}^T$. Finally $\overline{\mathbf{V}}^{\bar{\boldsymbol{a}}}$ can be expressed as

$$\overline{\mathbf{V}}^{\bar{\boldsymbol{a}}} = \sum_{p=1}^{n_s}\sum_{r=1}^{n_s} \left(\widetilde{\overline{\mathbf{Q}}}^{-1}\right)_{pr} \boldsymbol{\Pi}_{pr}^{\bar{\boldsymbol{a}}'} \tag{E.7}$$

where $\boldsymbol{\Pi}_{pr}^{\bar{\boldsymbol{a}}'}$ is the (p,r)th block matrix of $\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}'}^{(j)}$

$$\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}'}^{(j)} = \begin{bmatrix} \boldsymbol{\Pi}_{11}^{\bar{\boldsymbol{a}}'} & \boldsymbol{\Pi}_{12}^{\bar{\boldsymbol{a}}'} & \dots & \boldsymbol{\Pi}_{1n_s}^{\bar{\boldsymbol{a}}'} \\ \boldsymbol{\Pi}_{21}^{\bar{\boldsymbol{a}}'} & \boldsymbol{\Pi}_{22}^{\bar{\boldsymbol{a}}'} & \dots & \boldsymbol{\Pi}_{2n_s}^{\bar{\boldsymbol{a}}'} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Pi}_{n_s1}^{\bar{\boldsymbol{a}}'} & \boldsymbol{\Pi}_{n_s2}^{\bar{\boldsymbol{a}}'} & \dots & \boldsymbol{\Pi}_{n_sn_s}^{\bar{\boldsymbol{a}}'} \end{bmatrix} \tag{E.8}$$

This completes the computation of $\overline{\mathbf{V}}^{\bar{\boldsymbol{a}}}$ and $\overline{\mathbf{V}}_k^{\bar{\boldsymbol{g}}}$ used in Eqs.(5.31c) and (5.31d) respectively.

E.3 Derivation of update rules for the VBM step

The principle for deriving the update rules of parameters $\overline{\mathbf{A}}$, $\overline{\mathbf{G}}$, $\overline{\mathbf{Q}}$ and $\mathbf{R}$ follows the maximization expressions given in Eq.(5.25). For a parameter under consideration, the expectation under its own variational distribution is ignored and the expectations with respect to variational distributions of other parameters are evaluated. Additionally, all terms independent of the parameter under consideration are subsumed in a proportionality constant.

(a) **Update for** $q^{(j+1)}(\bar{\boldsymbol{a}})$

Using the log joint likelihood in Eq.(5.28) and ignoring the expectation w.r.t. $q^{(j)}(\bar{\boldsymbol{a}})$, one obtains the following expression as a function of $\bar{\boldsymbol{a}}$

$$\begin{aligned} \ln q^{(j+1)}(\bar{\boldsymbol{a}}) \propto\ & \operatorname{tr}\left\{\left\langle \overline{\mathbf{Q}}^{-1}\right\rangle_{q^{(j)}(\overline{\mathbf{Q}})} \left(\mathbf{S}_{\bar{f}\bar{f}}^{(j)} - \overline{\mathbf{A}}\mathbf{S}_{\bar{f}\bar{x}}^{(j)^T} - \mathbf{S}_{\bar{f}\bar{x}}^{(j)}\overline{\mathbf{A}}^T + \overline{\mathbf{A}}\mathbf{S}_{\bar{x}\bar{x}}^{(j)}\overline{\mathbf{A}}^T\right)\right\} \\ & + \left\{\bar{\boldsymbol{a}}^T\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{-1}\bar{\boldsymbol{a}} - \bar{\boldsymbol{a}}^T\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{-1}\boldsymbol{\mu}_{\bar{\boldsymbol{a}}} - \boldsymbol{\mu}_{\bar{\boldsymbol{a}}}^T\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{-1}\bar{\boldsymbol{a}} + \boldsymbol{\mu}_{\bar{\boldsymbol{a}}}^T\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{-1}\boldsymbol{\mu}_{\bar{\boldsymbol{a}}}\right\} \end{aligned} \tag{E.9}$$

Note that $\overline{\mathbf{A}}$ is the matrix form of $\bar{\boldsymbol{a}}$. Using the vectorization properties

$$\operatorname{tr}\{\mathcal{A}\mathcal{B}\} = \operatorname{vec}\left(\mathcal{A}^T\right)^T \operatorname{vec}(\mathcal{B}) \tag{E.10a}$$

$$\operatorname{tr}\{\mathcal{A}\mathcal{B}\mathcal{C}\mathcal{D}\} = \operatorname{vec}\left(\mathcal{A}^T\right)^T (\mathcal{D}\otimes\mathcal{B}) \operatorname{vec}(\mathcal{C}) \tag{E.10b}$$

and the notation $\boldsymbol{\mu}^{(j)}_{\overline{\mathbf{Q}}^{-1}} \coloneqq \left\langle \overline{\mathbf{Q}}^{-1}\right\rangle_{q^{(j)}(\overline{\mathbf{Q}})} = d^{(j)}_{\overline{\mathbf{Q}}}{\mathbf{D}^{(j)}_{\overline{\mathbf{Q}}}}^{-1}$, one can express Eq.(E.9) as

$$\begin{aligned}
\ln q^{(j+1)}(\bar{\boldsymbol{a}}) \propto\ & \operatorname{vec}\left(\boldsymbol{\mu}^{(j)}_{\overline{\mathbf{Q}}^{-1}}\mathbf{S}^{(j)}_{\bar{f}\bar{f}}\right) - \operatorname{vec}\left(\boldsymbol{\mu}^{(j)}_{\overline{\mathbf{Q}}^{-1}}\mathbf{S}^{(j)}_{\bar{f}\bar{x}}\right)^T \bar{\boldsymbol{a}} - \bar{\boldsymbol{a}}^T \operatorname{vec}\left(\boldsymbol{\mu}^{(j)}_{\overline{\mathbf{Q}}^{-1}}\mathbf{S}^{(j)}_{\bar{f}\bar{x}}\right) \\
& + \bar{\boldsymbol{a}}^T\left(\mathbf{S}^{(j)}_{\bar{x}\bar{x}}\otimes\boldsymbol{\mu}^{(j)}_{\overline{\mathbf{Q}}^{-1}}\right)\bar{\boldsymbol{a}} + \left\{\bar{\boldsymbol{a}}^T\boldsymbol{\Sigma}^{-1}_{\bar{\boldsymbol{a}}}\bar{\boldsymbol{a}} - \bar{\boldsymbol{a}}^T\boldsymbol{\Sigma}^{-1}_{\bar{\boldsymbol{a}}}\boldsymbol{\mu}_{\bar{\boldsymbol{a}}} - \boldsymbol{\mu}^T_{\bar{\boldsymbol{a}}}\boldsymbol{\Sigma}^{-1}_{\bar{\boldsymbol{a}}}\bar{\boldsymbol{a}} + \boldsymbol{\mu}^T_{\bar{\boldsymbol{a}}}\boldsymbol{\Sigma}^{-1}_{\bar{\boldsymbol{a}}}\boldsymbol{\mu}_{\bar{\boldsymbol{a}}}\right\} \\
\propto\ & \bar{\boldsymbol{a}}^T\left(\boldsymbol{\Sigma}^{-1}_{\bar{\boldsymbol{a}}} + \mathbf{S}^{(j)}_{\bar{x}\bar{x}}\otimes\boldsymbol{\mu}^{(j)}_{\overline{\mathbf{Q}}^{-1}}\right)\bar{\boldsymbol{a}} - 2\left(\boldsymbol{\Sigma}^{-1}_{\bar{\boldsymbol{a}}}\boldsymbol{\mu}_{\bar{\boldsymbol{a}}} + \operatorname{vec}\left(\boldsymbol{\mu}^{(j)}_{\overline{\mathbf{Q}}^{-1}}\mathbf{S}^{(j)}_{\bar{f}\bar{x}}\right)\right)^T\bar{\boldsymbol{a}}
\end{aligned} \tag{E.11}$$

Using the argument of 'completion-of-squares' (see Section 8.1.6 in Matrix Cookbook [142]), one obtains the following form

$$\ln q^{(j+1)}(\bar{\boldsymbol{a}}) \propto \left(\bar{\boldsymbol{a}} - \boldsymbol{\mu}^{(j+1)}_{\bar{\boldsymbol{a}}}\right)^T \boldsymbol{\Sigma}^{(j+1)}_{\bar{\boldsymbol{a}}}\left(\bar{\boldsymbol{a}} - \boldsymbol{\mu}^{(j+1)}_{\bar{\boldsymbol{a}}}\right) + C_1 \tag{E.12}$$

where C_1 consists of terms independent of $\bar{\boldsymbol{a}}$. Comparing the above form to that of a multivariate Gaussian distribution, the expressions for hyperparameters $\boldsymbol{\Sigma}^{(j+1)}_{\bar{\boldsymbol{a}}}$, $\boldsymbol{\mu}^{(j+1)}_{\bar{\boldsymbol{a}}}$, $\boldsymbol{\mu}^{(j+1)}_{\overline{\mathbf{A}}}$ are obtained, as depicted in Eq.(5.35).

(b) **Update for** $q^{(j+1)}(\bar{\boldsymbol{g}})$

The derivation follows similarly as that for $q^{(j+1)}(\bar{\boldsymbol{a}})$. One can write

$$\begin{aligned}
\ln q^{(j+1)}(\bar{\boldsymbol{g}}) \propto & \sum_{k=1}^{N}\left(\operatorname{tr}\left\{\left\langle\mathbf{R}^{-1}\right\rangle_{q^{(j)}(\mathbf{R})}\left(\mathbf{S}^{(j)}_{yy,k} - \boldsymbol{\Omega}_k\overline{\mathbf{G}}{\mathbf{S}^{(j)}_{y\bar{x},k}}^T - \mathbf{S}^{(j)}_{y\bar{x},k}\overline{\mathbf{G}}^T\boldsymbol{\Omega}^T_k + \boldsymbol{\Omega}_k\overline{\mathbf{G}}\mathbf{S}^{(j)}_{\bar{x}\bar{x},k}\overline{\mathbf{G}}^T\boldsymbol{\Omega}^T_k\right)\right\}\right) \\
& + \bar{\boldsymbol{g}}^T\boldsymbol{\Sigma}^{-1}_{\bar{\boldsymbol{g}}}\bar{\boldsymbol{g}} - \bar{\boldsymbol{g}}^T\boldsymbol{\Sigma}^{-1}_{\bar{\boldsymbol{g}}}\boldsymbol{\mu}_{\bar{\boldsymbol{g}}} - \boldsymbol{\mu}^T_{\bar{\boldsymbol{g}}}\boldsymbol{\Sigma}^{-1}_{\bar{\boldsymbol{g}}}\bar{\boldsymbol{g}} + \boldsymbol{\mu}^T_{\bar{\boldsymbol{g}}}\boldsymbol{\Sigma}^{-1}_{\bar{\boldsymbol{g}}}\boldsymbol{\mu}_{\bar{\boldsymbol{g}}}
\end{aligned} \tag{E.13}$$

Denoting $\boldsymbol{\mu}^{(j)}_{\mathbf{R}^{-1}} \coloneqq \left\langle\mathbf{R}^{-1}\right\rangle_{q^{(j)}(\mathbf{R})} = d^{(j)}_{\mathbf{R}}{\mathbf{D}^{(j)}_{\mathbf{R}}}^{-1}$ and using the vectorization properties in Eq.(E.10), one arrives at

$$\begin{aligned}
\ln q^{(j+1)}(\bar{\boldsymbol{g}}) \propto\ & \bar{\boldsymbol{g}}^T\left(\boldsymbol{\Sigma}^{-1}_{\bar{\boldsymbol{g}}} + \sum_{k=1}^{N}\left(\mathbf{S}^{(j)}_{\bar{x}\bar{x},k}\otimes\left(\boldsymbol{\Omega}^T_k\boldsymbol{\mu}^{(j)}_{\mathbf{R}^{-1}}\boldsymbol{\Omega}_k\right)\right)\right)\bar{\boldsymbol{g}} \\
& - 2\left(\boldsymbol{\Sigma}^{-1}_{\bar{\boldsymbol{g}}}\boldsymbol{\mu}_{\bar{\boldsymbol{g}}} + \sum_{k=1}^{N}\operatorname{vec}\left(\boldsymbol{\Omega}^T_k\boldsymbol{\mu}^{(j)}_{\mathbf{R}^{-1}}\mathbf{S}^{(j)}_{y\bar{x},k}\right)\right)^T\bar{\boldsymbol{g}}
\end{aligned} \tag{E.14}$$

which can be further arranged in the form of 'completion-of-squares'. Comparing the above form to that of a multivariate Gaussian distribution, the expressions for hyperparameters $\boldsymbol{\Sigma}_{\overline{\boldsymbol{g}}}^{(j+1)}$, $\boldsymbol{\mu}_{\overline{\boldsymbol{g}}}^{(j+1)}$, $\boldsymbol{\mu}_{\overline{\mathbf{G}}}^{(j+1)}$ are obtained, as depicted in Eq.(5.38).

(c) **Update for** $q^{(j+1)}\left(\overline{\mathbf{Q}}\right)$

Ignoring the expectation of the log joint likelihood in Eq.(5.28) with respect to $q^{(j)}\left(\overline{\mathbf{Q}}\right)$, one obtains the following expression as a function of $\overline{\mathbf{Q}}$

$$\begin{aligned}\ln q^{(j+1)}\left(\overline{\mathbf{Q}}\right) \propto\ & \left(d_{\overline{\mathbf{Q}}}+N+n_s+1\right)\ln\left|\overline{\mathbf{Q}}\right|+\operatorname{tr}\left\{\mathbf{D}_{\overline{\mathbf{Q}}}\overline{\mathbf{Q}}^{-1}\right\}\\ &+\operatorname{tr}\left\{\overline{\mathbf{Q}}^{-1}\left(\mathbf{S}_{\bar{f}\bar{f}}^{(j)}-\left\langle\overline{\mathbf{A}}\right\rangle_{q^{(j+1)}(\bar{\boldsymbol{a}})}\mathbf{S}_{\bar{f}\bar{x}}^{(j)^T}-\mathbf{S}_{\bar{f}\bar{x}}^{(j)}\left\langle\overline{\mathbf{A}}\right\rangle_{q^{(j+1)}(\bar{\boldsymbol{a}})}^{T}+\left\langle\overline{\mathbf{A}}\mathbf{S}_{\bar{x}\bar{x}}^{(j)}\overline{\mathbf{A}}^T\right\rangle_{q^{(j+1)}(\bar{\boldsymbol{a}})}\right)\right\}\end{aligned} \tag{E.15}$$

The expectations on the right hand side can be evaluated as follows:

$$\left\langle\overline{\mathbf{A}}\right\rangle_{q^{(j+1)}(\bar{\boldsymbol{a}})}=\boldsymbol{\mu}_{\overline{\mathbf{A}}}^{(j+1)} \tag{E.16a}$$

$$\mathbf{M}_{\overline{\mathbf{A}}}^{(j+1)}\coloneqq\left\langle\overline{\mathbf{A}}\mathbf{S}_{\bar{x}\bar{x}}^{(j)}\overline{\mathbf{A}}^T\right\rangle_{q^{(j+1)}(\bar{\boldsymbol{a}})}=\sum_{p=1}^{n_s}\sum_{r=1}^{n_s}\left(\mathbf{S}_{\bar{x}\bar{x}}^{(j)}\right)_{pr}\boldsymbol{\Pi}_{pr}^{\bar{\boldsymbol{a}}}+\boldsymbol{\mu}_{\overline{\mathbf{A}}}^{(j+1)}\mathbf{S}_{\bar{x}\bar{x}}^{(j)}\boldsymbol{\mu}_{\overline{\mathbf{A}}}^{(j+1)^T} \tag{E.16b}$$

where $\boldsymbol{\Pi}_{pr}^{\bar{\boldsymbol{a}}}$ is the (p,r)th block matrix in $\boldsymbol{\Sigma}_{\bar{\boldsymbol{a}}}^{(j+1)}$ as depicted in Eq.(5.37) and $\left(\mathbf{S}_{\bar{x}\bar{x}}^{(j)}\right)_{pr}$ is the (p,r)th entry of $\mathbf{S}_{\bar{x}\bar{x}}^{(j)}$. Substituting Eq.(E.16) in Eq.(E.15), one obtains the following form

$$\begin{aligned}\ln q^{(j+1)}\left(\overline{\mathbf{Q}}\right) &\propto \left(d_{\overline{\mathbf{Q}}}+N+n_s+1\right)\ln\left|\overline{\mathbf{Q}}\right|\\ &\quad+\operatorname{tr}\left\{\overline{\mathbf{Q}}^{-1}\left(\mathbf{D}_{\overline{\mathbf{Q}}}+\mathbf{S}_{\bar{f}\bar{f}}^{(j)}-\boldsymbol{\mu}_{\overline{\mathbf{A}}}^{(j+1)}\mathbf{S}_{\bar{f}\bar{x}}^{(j)^T}-\mathbf{S}_{\bar{f}\bar{x}}^{(j)}\boldsymbol{\mu}_{\overline{\mathbf{A}}}^{(j+1)^T}+\mathbf{M}_{\overline{\mathbf{A}}}^{(j+1)}\right)\right\}\\ &\propto \left(d_{\overline{\mathbf{Q}}}^{(j+1)}+n_s+1\right)\ln\left|\overline{\mathbf{Q}}\right|+\operatorname{tr}\left\{\overline{\mathbf{Q}}^{-1}\mathbf{D}_{\overline{\mathbf{Q}}}^{(j+1)}\right\}\end{aligned} \tag{E.17}$$

Comparing the above form to that of an inverse Wishart distribution, the expressions for hyperparameters $d_{\overline{\mathbf{Q}}}^{(j+1)}$, $\mathbf{D}_{\overline{\mathbf{Q}}}^{(j+1)}$ are obtained, as depicted in Eq.(5.36).

(d) **Update for** $q^{(j+1)}\left(\mathbf{R}\right)$

Ignoring the expectation of the log joint likelihood in Eq.(5.28) with respect to $q^{(j)}\left(\mathbf{R}\right)$,

one obtains the following expression as a function of $\mathbf{R}$

$$\begin{aligned}\ln q^{(j+1)}(\mathbf{R}) \propto\ & (d_{\mathbf{R}}+N+n_o+1)\ln|\mathbf{R}| + \operatorname{tr}\left\{\mathbf{D}_{\mathbf{R}}\mathbf{R}^{-1}\right\} \\ & + \sum_{k=1}^{N}\operatorname{tr}\left\{\mathbf{R}^{-1}\left(\mathbf{S}^{(j)}_{yy,k} - \boldsymbol{\Omega}_k\left\langle\overline{\mathbf{G}}\right\rangle_{q^{(j+1)}(\bar{\boldsymbol{g}})}\mathbf{S}^{(j)}_{y\bar{x},k}{}^{T} - \mathbf{S}^{(j)}_{y\bar{x},k}\left\langle\overline{\mathbf{G}}\right\rangle^{T}_{q^{(j+1)}(\bar{\boldsymbol{g}})}\boldsymbol{\Omega}_k^T\right)\right\} \\ & + \sum_{k=1}^{N}\operatorname{tr}\left\{\mathbf{R}^{-1}\left(\boldsymbol{\Omega}_k\left\langle\overline{\mathbf{G}}\mathbf{S}^{(j)}_{\bar{x}\bar{x},k}\overline{\mathbf{G}}^T\right\rangle_{q^{(j+1)}(\bar{\boldsymbol{g}})}\boldsymbol{\Omega}_k^T\right)\right\}\end{aligned} \tag{E.18}$$

The terms inside the expectations can be evaluated as follows:

$$\left\langle\overline{\mathbf{G}}\right\rangle_{q^{(j+1)}(\bar{\boldsymbol{g}})} = \boldsymbol{\mu}^{(j+1)}_{\overline{\mathbf{G}}} \tag{E.19a}$$

$$\mathbf{M}^{(j+1)}_{\overline{\mathbf{G}},k} \coloneqq \left\langle\overline{\mathbf{G}}\mathbf{S}^{(j)}_{\bar{x}\bar{x},k}\overline{\mathbf{G}}^T\right\rangle_{q^{(j+1)}(\bar{\boldsymbol{g}})} = \sum_{p=1}^{n_s}\sum_{r=1}^{n_s}\left(\mathbf{S}^{(j)}_{\bar{x}\bar{x},k}\right)_{pr}\boldsymbol{\Pi}^{\bar{\boldsymbol{g}}}_{pr} + \boldsymbol{\mu}^{(j+1)}_{\overline{\mathbf{G}}}\mathbf{S}^{(j)}_{\bar{x}\bar{x},k}\boldsymbol{\mu}^{(j+1)^T}_{\overline{\mathbf{G}}} \tag{E.19b}$$

where $\boldsymbol{\Pi}^{\bar{\boldsymbol{g}}}_{pr}$ is the (p,r)th block matrix in $\boldsymbol{\Sigma}^{(j+1)}_{\bar{\boldsymbol{g}}}$ as depicted in Eq.(5.40) and $\left(\mathbf{S}^{(j)}_{\bar{x}\bar{x},k}\right)_{pr}$ is the (p,r)th entry of $\mathbf{S}^{(j)}_{\bar{x}\bar{x},k}$. Substituting Eq.(E.19) in Eq.(E.18), one obtains the following form

$$\begin{aligned}\ln q^{(j+1)}(\mathbf{R}) \propto\ & (d_{\mathbf{R}}+N+n_o+1)\ln|\mathbf{R}| \\ & + \operatorname{tr}\left\{\mathbf{R}^{-1}\left(\mathbf{D}_{\mathbf{R}} + \sum_{k=1}^{N}\left(\mathbf{S}^{(j)}_{yy,k} - \boldsymbol{\Omega}_k\boldsymbol{\mu}^{(j+1)}_{\overline{\mathbf{G}}}\mathbf{S}^{(j)}_{y\bar{x},k}{}^{T} - \mathbf{S}^{(j)}_{y\bar{x},k}\boldsymbol{\mu}^{(j+1)^T}_{\overline{\mathbf{G}}}\boldsymbol{\Omega}_k^T\right)\right)\right\} \\ & + \operatorname{tr}\left\{\mathbf{R}^{-1}\left(\sum_{k=1}^{N}\left(\boldsymbol{\Omega}_k\mathbf{M}^{(j+1)}_{\overline{\mathbf{G}},k}\boldsymbol{\Omega}_k^T\right)\right)\right\} \\ \propto\ & \left(d^{(j+1)}_{\mathbf{R}}+n_o+1\right)\ln|\mathbf{R}| + \operatorname{tr}\left\{\mathbf{R}^{-1}\mathbf{D}^{(j+1)}_{\mathbf{R}})\right\}\end{aligned} \tag{E.20}$$

Comparing the above form to that of an inverse Wishart distribution, the expressions for hyperparameters $d^{(j+1)}_{\mathbf{R}}$, $\mathbf{D}^{(j+1)}_{\mathbf{R}}$ are obtained, as depicted in Eq.(5.39).

Appendix F

Kalman filter and RTS smoother

Consider a discrete-time time-invariant linear system with additive Gaussian noise

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \mathbf{A}\boldsymbol{x}_k + \boldsymbol{w}_k \\ \boldsymbol{y}_k &= \mathbf{G}\boldsymbol{x}_k + \boldsymbol{v}_k \end{aligned} \tag{F.1}$$

where $\boldsymbol{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ and $\boldsymbol{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ are the process noise and the measurement noise distribution respectively, with zero cross-correlation between $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$.

The Kalman filter [143] and RTS smoother [144] equations for obtaining the smoothed states $\hat{\boldsymbol{x}}_{k|N}$ and covariances $\hat{\mathbf{V}}_{k|N}, \hat{\mathbf{V}}_{k+1,k|N}$ are given below.

Kalman filter : Do for $k = 1, \ldots, N$

$$\boldsymbol{e}_k = \boldsymbol{y}_k - \mathbf{G}\hat{\boldsymbol{x}}_{k|k-1} \tag{F.2a}$$
$$\mathbf{S}_k = \mathbf{G}\hat{\mathbf{V}}_{k|k-1}\mathbf{G}^T + \mathbf{R} \tag{F.2b}$$
$$\mathbf{K}_k = \hat{\mathbf{V}}_{k|k-1}\mathbf{G}^T\mathbf{S}_k^{-1} \tag{F.2c}$$
$$\hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k-1} + \mathbf{K}_k\boldsymbol{e}_k \tag{F.2d}$$
$$\hat{\mathbf{V}}_{k|k} = \hat{\mathbf{V}}_{k|k-1} - \mathbf{K}_k\mathbf{S}_k\mathbf{K}_k^T \tag{F.2e}$$
$$\hat{\boldsymbol{x}}_{k+1|k} = \mathbf{A}\hat{\boldsymbol{x}}_{k|k} \tag{F.2f}$$
$$\hat{\mathbf{V}}_{k+1|k} = \mathbf{A}\hat{\mathbf{V}}_{k|k}\mathbf{A} + \mathbf{Q} \tag{F.2g}$$

Here $\hat{\boldsymbol{x}}_{k|k-1}$ and $\hat{\boldsymbol{x}}_{k|k}$ represent the kth predicted and filtered state estimate respectively, and, $\hat{\mathbf{V}}_{k|k-1}$ and $\hat{\mathbf{V}}_{k|k}$ denote the kth predicted and filtered state error covariance matrices respectively. The Kalman filter recursion is started from an initial state $\hat{\boldsymbol{x}}_{1|0}$ and an initial covariance $\hat{\mathbf{V}}_{1|0}$. $\boldsymbol{e}_k$ and $\mathbf{S}_k$ represent the innovation and the innovation covariance at the kth time step.

Following the filtering step, the (fixed interval) smoothing recursions given by the RTS smoother are computed as follows:

Kalman smoother : Do for $k = N, \ldots, 1$

$$\mathbf{N}_k = \hat{\mathbf{V}}_{k|k}\mathbf{A}^T \left(\hat{\mathbf{V}}_{k+1|k}\right)^{-1} \tag{F.3a}$$

$$\hat{\boldsymbol{x}}_{k|N} = \hat{\boldsymbol{x}}_{k|k} + \mathbf{N}_k \left(\hat{\boldsymbol{x}}_{k+1|N} - \hat{\boldsymbol{x}}_{k+1|k}\right) \tag{F.3b}$$

$$\hat{\mathbf{V}}_{k|N} = \hat{\mathbf{V}}_{k|k} + \mathbf{N}_k \left(\hat{\mathbf{V}}_{k+1|N} - \hat{\mathbf{V}}_{k+1|k}\right) \mathbf{N}_k^T \tag{F.3c}$$

$$\hat{\mathbf{V}}_{k+1,k|N} = \hat{\mathbf{V}}_{k+1|N}\mathbf{N}_k^T \tag{F.3d}$$