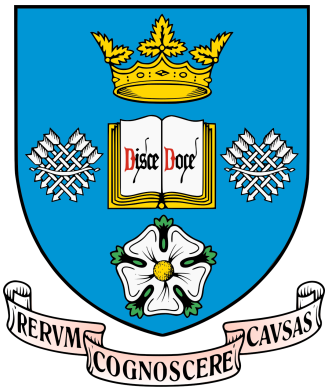
Novel Probabilistic and Bayesian Approaches to Uncertainty Quantification for Operational Modal Analysis.



UNIVERSITY OF SHEFFIELD

A Thesis submitted to the University of Sheffield

for the degree of Doctor of Philosophy in the Faculty of Engineering

by

Brandon John O'Connell

Department of Mechanical Engineering

University of Sheffield

30 October 2024

Acknowledgements

Embarking on a PhD is easy. Putting in the necessary hours, writing a thesis and seeing the PhD through is a long and arduous process; one that I could not have done without the unwavering support, guidance and friendship of so many people of whom here I intend to thank.

First and foremost I would like to thank my supervisor, Dr Tim Rogers. Your all round enthusiasm for research, work ethic and inquisitive nature has been nothing but inspiring. Thank you for pushing me to be the best and for keeping me on track (as best as you could). Not only have you been my mentor, but also a true friend. I would also like to take a moment to thank my second supervisor, Prof. Lizzy Cross. Your guidance and support when needed has been invaluable. I would also like to thank Nikos and Keith, whose module on Advanced Dynamics was the fire that lit the flame for my PhD journey.

Next I would like to thank the Dynamics Research Group (DRG) at the University of Sheffield. A PhD can be a lonely endeavour, but not when you are part of the DRG. I could take up an entire page thanking each and every one of you for your part in my PhD journey but here are a few honourable mentions: Joe, thank you for your friendship and for often allowing me to soundboard off you. Getting to do my PhD alongside you, under the supervision of Tim, has been a lot of fun. To Max, thank you for your mentorship and friendship. You assisted me and advised me on so many occasions throughout my PhD and I am very grateful. And finally, to all of my friends in the DRG I say this — having the opportunity to work with such a lovely, friendly and talented group of intellectuals such as yourselves has, and will continue to be, a pleasure.

To Frances, Charley and Anna. Thank you for being such kind and compassionate friends. You often lent an ear when I needed it most, lifted my spirits and never failed to make me smile.

To Chris and Max. Our yearly trips to Saalfelden in Austria, shredding the slopes and drinking copious amounts of Stiegl, are times I will cherish and never forget. Prost!

To all the musicians in my life. Playing the saxophone throughout my PhD has provided me with so much joy and, more importantly, balance. Being up on stage with each and every one of you is always an absolute pleasure.

To my grandfather, Dr David Burley. Your passion and interest in mathematics inspired me long ago to pursue my own interests in engineering. Thank you for taking the time to read my manuscripts, provide me with feedback, and listen to my ramblings.

To my Mum, Dad, brothers (Joey and Aedán), grandparents, and all my family. Thank you for your unwavering love, support and encouragement to be the best version of myself. You taught me to always push myself out of my comfort zone, never give up, never give in, and to stay true to my values. I would not be the man I am today without you. I love you all.

And finally, a special thank you must go to Nia. Thank you for being patient with me, being by my side and always believing in me — particularly when I had lost faith in myself. For that I will always be grateful.

Abstract

The analysis of dynamic systems is of fundamental importance to countless engineering applications. The default way to analyse linear structural dynamic systems is through a methodology known as modal analysis. Since its introduction over 50 years ago, experimental modal analysis (EMA) has presented engineers with a principled way of recovering the modal properties of dynamic systems from observations of the excitation and structural response. However, the task of EMA becomes difficult when access to, or generation of, a controlled input force is limited. This is often the case for large engineering structures or structures in hard-to-replicate environments. To solve this problem, engineers have developed a new suite of methods known collectively as operational modal analysis (OMA), that rely on the natural excitation of a structure to obtain the modal properties. Over the last 20 years, a broad range of OMA algorithms have become available to the dynamicist, with Stochastic Subspace Identification (SSI) emerging as one of the leading methods across industry and academia owing to its unrivalled performance in many scenarios.

Despite their success, several research challenges remain in the field of OMA. Of these challenges, the handling of uncertainty is of key interest. It is well understood that measured data contain uncertainties that obscure the underlying signal one intends to measure. The effect of these uncertainties on the identification process in OMA algorithms are seldom considered, nay quantified, but are known to result in variations in the modal properties. Variations in recovered modal information that arise from uncertainty can pose a significant risk; modal information is frequently used to inform safety-critical decision-making in areas such as Structural Health Monitoring (SHM).

One way to address this challenge is to develop OMA methodologies capable of quantifying the uncertainty over the modal properties. Access to the uncertainty

provides a level of insight into the variability of the parameters given the available data and chosen methodology. In general Bayesian methods are a popular form of uncertainty quantification that combine a description of ones prior belief, with the likelihood of observed data, to obtain an estimate for the uncertainty as a posterior distribution. Despite some existing approaches to uncertainty quantification for OMA, no Bayesian formulation of the SSI algorithm currently exists in the literature, where posterior distributions over the modal properties are obtainable. In light of this fact, the current thesis aims to address this shortcoming.

At the centre of this work lies a novel probabilistic interpretation of the SSI algorithm. This form is achieved through the direct replacement of the identification mechanism at the core of SSI — canonical correlation analysis (CCA) — with its probabilistic equivalent. This now probabilistic algorithm unlocks the ability to apply Bayesian hierarchical modelling techniques that can better represent noise in the identification process and incorporate prior information. In this work three unique algorithms are presented. The first is a statistically robust SSI algorithm capable of robust identification of the modal parameters when faced with atypical observations in the measured time series. The second is a Bayesian approach to the SSI algorithm. By incorporating weakly informative proper prior information over the identification mechanism (forming Bayesian CCA), posterior distributions over the modal properties can be recovered. The last algorithm is an efficient form of the Bayesian SSI algorithm, achieved using stochastic approximation techniques, known to speed up the computation by reducing the number of operations over the entire dataset.

The algorithms and results contained within this thesis take positive steps towards viewing OMA as a problem in probabilistic and Bayesian inference; one where uncertainty can be principally considered and quantified, to provide a more informed description of a dynamic system's behaviour from observed data. Such descriptions provide additional information about a system and its measurement that can ultimately facilitate more-informed and confident decision-making where modal information is to be considered.

Publications and Awards

Journal Papers

B. J. O'Connell, T. J. Rogers (In preparation). A robust Bayesian approach to stochastic subspace identification

B. J. O'Connell, M. D. Champneys, T. J. Rogers (2025). A new perspective on Bayesian operational modal analysis, *Mechanical Systems and Signal Processing*

B. J. O'Connell, E. J. Cross, T. J. Rogers (2024). A robust probabilistic approach to stochastic subspace identification, *Journal of Sound and Vibration*

J. H. Mclean, M. R. Jones, **B. J. O'Connell**, A. E. Maguire and T. J. Rogers (2023). Physically meaningful uncertainty quantification in probabilistic wind turbine power curve models, *Structural Health Monitoring*

Conference Papers

B. J. O'Connell, T. J. Rogers. Exploring physically meaningful prior distributions for OMA, *11th International Operational Modal Analysis Conference*, Rennes FR (2025)

B. J. O'Connell, T. J. Rogers. On improving the efficiency of Bayesian stochastic subspace identification, *10th International Operational Modal Analysis Conference*, Naples IT (2024)

B. J. O'Connell, M. D. Champneys, E. J. Cross, T. J. Rogers. A novel variational Bayesian approach to stochastic subspace identification, *Uncertainty in Computation*, Athens GR (2023)

B. J. O'Connell, E. J. Cross, U. T. Tygesen, T. J. Rogers. Bayesian canonical correlations for stochastic subspace identification, *International Symposium on Modal Analysis and Uncertainty in Structural Dynamics*, Leuven BE (2022)

B. J. O'Connell, E. J. Cross, T. J. Rogers. Robust probabilistic canonical correlations for stochastic subspace identification, *9th International Operational Modal Analysis Conference*, Vancouver CA (2022)

Awards

10th International Operational Modal Analysis Conference, **Young Researcher's Best Paper Award** - 2024

9th International Operational Modal Analysis Conference, **Best Paper Award** - 2022

Contents

1

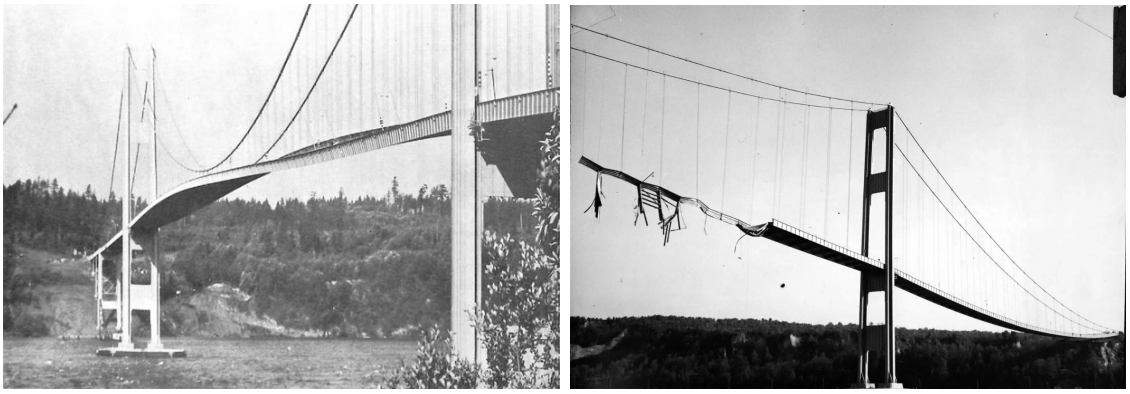
Introduction

1.1 Why are engineers interested in dynamics?

All structures move; they are (often despite our best efforts) *dynamic*. Understanding and describing what is known as the *dynamic behaviour* of these structures can provide one of the most useful and powerful tools in modern engineering. A lack of awareness or understanding of this behaviour can lead to serious consequences; perhaps history can speak for itself.

The most commonly used (some may say overused) example of extreme dynamic failure taught in undergraduate engineering courses across the globe, is that of the Tacoma Narrows bridge. In July 1940, the Tacoma Narrows bridge in Washington USA disastrously, and arguably quite spectacularly, tore itself apart after demonstrating extreme dynamic behaviour as a result of interactions with severe winds [1]. Large aerodynamic forces interacting with the bridge deck led to the excitation of a previously unseen torsional (twisting) mode which, when left to resonate uncontrol-

lably[1], ultimately led to the bridge's untimely demise. The photographs in Figure 1.1 show the torsional motion during (1.1a) and the destruction in the immediate aftermath (1.1b) of the disaster.



(a) (b)

Figure 1.1: The Tacoma Narrows bridge, Washington, US. Photographs from Washington Department for Transport [1]

Access to a description of the dynamic behaviour is therefore, an incredibly useful asset; whether one intends to mitigate (or generate) dynamic behaviour. Across a broad spectrum of engineering disciplines including aerospace, mechanical, civil, and energy, the dynamics of systems play a crucial role in the design, maintenance and continued performance of critical engineering infrastructure.

System identification is the field of engineering concerned with the recovery of a mathematical description of a system's dynamics, from observed data [2, 3]. These mathematical descriptions act as a form of functional mapping that describe how a system will respond to a given input. Originating[2] in the control community in the 1960s as a way of understanding, describing and implementing active control strategies for dynamic systems, the mathematics and methods thereof have since been used to explore other dynamic problems in fluid mechanics, thermodynamics and structural dynamics. In the realm of structural dynamics (for linear systems),

[1]The resonant behaviour was attributed to a phenomenon known as torsional flutter, where the motion of the bridge with the aerodynamic forces led to self-induced forcing behaviour.

[2]Arguably, system identification has existed for a very long time, in some form or another, and has always been of interest to humans. Be it Gauss' work on parameter estimation in the 1800s, Galileo's study of planetary orbits in the late 1500s or even the primitive caveman deciding if dropping a big rock off a cliff will hit the woolly mammoth below; system identification has played a crucial role throughout history in aiding humans' understanding of the world.

the most common representation is the *modal decomposition*, which is parameterised by the *modal properties*; namely the *natural frequencies*, *damping ratios* and *mode shapes*. The modal properties provide a unique and interpretable way of describing the dynamic motion of a structure using the principle of superposition to provide a compact, decoupled representation of the dynamics. Furthermore, the modal properties are also related to the underlying system properties of mass, stiffness and damping. The many techniques engineers have developed to recover the modal properties from observed data, fall under a branch of system identification known as *modal analysis*.

1.2 Modal analysis

Modal analysis is not a new discipline, having been well-studied over the last half a century. To date, there are a plethora of algorithms available to the modern engineer to apply modal analysis to measured data to recover the modal properties [4]. When observed data are available for both the input and output responses, this data-driven inverse problem is known as *experimental modal analysis* (EMA). Common types of input-output testing for EMA include: roving-hammer, fixed-shaker and shaker-table tests [5]. The process of performing experimental modal testing and subsequent modal identification are now very well-understood and adopted by many.

However, the identification of modal properties becomes a difficult task when access to or generation of measured input information is limited [6]. This is often the case for large-scale structures or for structures in hard-to-replicate environments with complex loading conditions; typical of aerospace, civil and offshore structures. For this reason, recent attention has been focused towards a suite of methods known as *operational modal analysis* (OMA).

1.2.1 Operational modal analysis

OMA is the subset of modal analysis methods concerned with the recovery of modal characteristics in the absence of measured input information; using only measurements of the output response. OMA is sometimes referred to as *output-only analysis* or *in-situ monitoring* given its application. The general premise of OMA is to take advantage of the natural excitation of a structure, that arises from ambient

forcing and operational loads (e.g. waves, wind, turbulence, ground-vibration, human-induced vibration), to replace artificially-induced excitation. This assumption not only removes the costly expense of controlled artificial excitation but also ensures the recovered modal properties are more representative of the actual operating conditions.

Consequently, OMA has gained increased popularity and is an active research field[3]. This is predominantly due to OMA being the only viable testing method for many engineering applications [7], which is especially the case for high-value large-scale assets such as bridges, buildings and offshore structures. Some specific applications of OMA will be covered in Chapter 2.

1.2.2 Applications of operational modal analysis

It is important to note that the dynamic properties recovered through EMA and OMA are seldom used in isolation. Instead, they typically exist as part of wider engineering frameworks that are designed to make decisions. The simplest example is in the design of machines and structures. Consider a modern automobile. During the design phase, it is useful for engineers to understand the dynamic properties of both the individual components and the full assembly to purposefully avoid excitation at natural frequencies; thus reducing the risk of resonant behaviour during normal operation. There are several other applications that employ modal properties for decision-making tasks, these include the fields of: *structural health monitoring* (SHM) [8, 9], digital twins [10] and control strategies for dynamic systems [11, 12], amongst others.

1.2.3 Current challenges

Despite the widespread success of OMA algorithms and the application of modal information across a range of fields, significant active research challenges remain. These include topics such as:

– **Model order selection** — Most EMA and OMA algorithms rely on overfitting and the appropriate selection of the true modes based on some chosen criteria.

[3]OMA is not strictly a new discipline either; it has been present since the early days of modal analysis as 'ambient testing', although the development of new powerful identification techniques has brought OMA back into the spotlight as methods become more reliable [7].

Selecting the right number of modes is a complex challenge, and is often addressed in combination with the problem of manual selection.

- **Manual identification** — To provide more consistent identification results, engineers are now seeking automatic identification procedures that remove any dependency on manual analysis, common in existing methods, which can often lead to variable results [13–17].

- **Non-linearity** — All dynamic systems are non-linear, however, many systems can be treated as linear under certain operating conditions. The greater presence of non-linearity in a dynamic system can make the process of system identification a more challenging task. The reliance on the principle of superposition in traditional modal analysis, means alternative approaches are needed to identify the parameters of the system and the non-linearity [18].

- **Non-physical solutions** — Many existing modal analysis methods can result in the identification of modal properties that are not physically meaningful in the context of real systems. An example here would be negative damping ratios, corresponding to unstable behaviour, of what may be expected to be a stable system.

- **Close modes** — When modes are close in frequency, they are often difficult to distinguish. This poses a significant challenge as key information on the dynamics of a system is not being identified. Researchers are now seeking new methods for identifying systems with close modes [19–21].

- **Sensor placement** — Sensor placement is the problem of choosing where on a structure measurements should be taken. This is important task, as one obviously intends to maximise the amount of useful information contained within any observed data. When considering a dynamic system, if several sensors are placed in locations corresponding to regions of minimal displacement at certain resonance frequencies, these modes will be harder to identify. Therefore, optimising sensor locations is of particular interest [19, 22–24].

- **Outliers and robustness** — Experimentally obtained data can often contain atypical observations and unwanted outliers in the measured time series. Due to the construction of some EMA and OMA algorithms, these outliers cannot be removed before analysis in the usual way. Therefore, researchers desire new methods that maintain statistical robustness and avoid biasing in the

identification procedure. Robust approaches to OMA are discussed in more depth in Chapter 5.

– **Uncertainty and noise** — Response measurements in OMA typically have low signal-to-noise ratios [6]. This is a direct result of their reliance on ambient forcing; the magnitudes of excitation seen in OMA are typically much lower than that of EMA testing schemes. This makes the process of identification a more challenging task, with variability frequently seen in the recovered modal properties. Consequently, many seek to quantify the associated uncertainty of the modal parameters to better understand this variation. Current uncertainty quantification methods for OMA are discussed in more depth in Section 2.4.

As it is well beyond the scope of this thesis to address all of these research challenges, a more focused approach is taken; where only the problems of uncertainty quantification and noise are considered. However, as will be shown later in Chapter 5, by directly addressing the problem of uncertainty, the challenge of performing robust identification can also be addressed.

1.3 Data and uncertainty

Observed data are incredibly useful; they inherently encapsulate all the complexities and nuance of physical systems; containing complex features and information one may struggle to embed within existing models. Many methods (like OMA) rely on measured data to infer correlations, uncover underlying patterns and extract valuable insight.

Nevertheless, observed data are imperfect; they provide a distorted view of a desired output as noise and other unwanted features obscure the otherwise 'perfect' output of the system. In the case of structural dynamics, these features arise from limitations in data acquisition or in the experiment. The data are, therefore, known to contain *uncertainties*.

Sources of uncertainty can often be classified as either being: *Epistemic* or *Aleatoric*[4] [25]. Epistemic uncertainty refers to the uncertainties associated with the model and

[4]Epistemic and Aleatoric uncertainty are not rigid terminologies and can depend on the context of the model and problem.

are the result of incomplete information or missing data. Such uncertainties can arise from: sparse measurements that fail to comprehensively describe a system's behaviour; indirect measurements of quantities of interest; and the impracticality of observing all the necessary parts of a system. Often the tools and mechanisms required to reduce this uncertainty are highly uneconomical or impractical, given physical constraints.

Aleatoric uncertainty describes uncertainties associated with the inherent stochasticity of the identification process; an example of this is measurement noise, which can obscure the true state of a system. This type of uncertainty is considered irreducible because it is a fundamental aspect of the system's behaviour or the data, meaning no amount of additional data or knowledge can fully eliminate it. In contrast, one can try to work around epistemic uncertainty; for instance, it can be reduced with access to more data, or with an improved model.

Considering the dependency of data-based methods on measured data, and that data are inherently uncertain, any models constructed from data must also be uncertain. It is, therefore, of critical importance that any uncertainties are quantified to inform and assist engineering decisions; thus, allowing for a more confident and risk-based approach [26].

1.4 Summary

Evidently, there is still a need for the engineering community to accurately obtain and use modal information; this need is especially true in the context of operational testing of structures and in decision-making tasks like SHM. However, significant research challenges remain in OMA. Of the challenges presented previously, the quantification of uncertainty is of key interest. Given that observed data are inherently uncertain, the identified models and parameters of said models must also be uncertain. In the context of OMA, uncertainty in observed data can result in variability in the identified modal parameters. Therefore, it is vital engineers begin to quantify any associated uncertainties; to enable more-informed and confident decision-making when using modal information.

While the quantification of uncertainty has been previously considered in modal analysis, as will be discussed in Chapter 2, this thesis will attempt to address

the problem for OMA using a new model; one where probabilistic and Bayesian understanding is embedded within the definition of an OMA algorithm. Probabilistic and specifically Bayesian methodologies have the advantage of providing a natural and principled way of combining prior knowledge with observed data, to consider and quantify uncertainty [27]. An introduction to Bayesian uncertainty will be provided in Chapter 3.2. Moreover, the author hypothesises that, by thinking in a Bayesian sense, one can not only consider and quantify uncertainty in OMA, but that doing so may also unlock the capability for solving some of the other current research challenges. This will be discussed in future work.

1.4.1 Aim

Considering the many applications of OMA and the identified research challenge introduced in this chapter, a general aim for this work can be defined. This body of work aims to:

Develop robust, operational modal analysis algorithms, combining existing methods with probabilistic and Bayesian techniques, to extend current capabilities and improve models of dynamic systems, all whilst accounting for uncertainty.

1.4.2 Thesis contributions

In an effort to address the aim outlined above, this thesis presents several novel contributions. These include:

1. **A probabilistic stochastic subspace identification (SSI) algorithm**, derived using the theory of probabilistic canonical correlations, to reformulate the classic algorithm using a latent variable model. The maximum likelihood result of this model is shown to be mathematically equivalent to that of standard SSI. This definition principally leaves the results of SSI unchanged, but provides the necessary mathematical foundation to introduce other probabilistic or hierarchical extensions.

2. **A robust probabilistic formulation of the SSI algorithm**. Using the probabilistic formulation of SSI, this thesis introduces a statistically robust form of SSI by replacing the standard Gaussian noise model with that of a

Student's t-distribution; one better equipped to handle atypical observations (outliers) in data. This new methodology is shown to have similar identification performance to classic SSI when no outliers or atypical observations are present, and improved identification performance in the presence of such measurements. This is demonstrated for a variety of case studies exploring different types of atypical observation including: clipping, random dropout, periodic dropout and multi-channel dropout.

3. **A Bayesian formulation of the SSI algorithm**. Through the addition of prior distributions over the model parameters in probabilistic SSI, this thesis introduces a new Bayesian methodology; one capable of quantifying the uncertainty over the modal properties as posterior distributions. The solution to the Bayesian inference problem is presented using two different schemes: Gibbs sampling, and variational inference. This new methodology is shown to accurately recover the posterior distribution over the modal properties at fixed model orders, with mean values that correspond to the results of traditional SSI. The performance of this method is also shown for multiple model orders using a stabilisation diagram, where the mean values of distributions with low variance correspond to the expected natural frequencies; the distributions with mean values far from the expected values for natural frequency display much larger variance.

4. **An efficient implementation of the Bayesian SSI algorithm**. It is important when presenting new methodologies to consider any possible barriers to their adoption; this is particularly true of the Bayesian SSI algorithm. Performing Bayesian inference typically requires higher computational demand than standard deterministic processes. In order to address this problem, this thesis presents a computationally efficient form of Bayesian SSI, employing mini-batch stochastic variational inference [28] as a principled way of reducing the number of operations on the complete dataset before reaching convergence at a local optimum.

1.4.3 Thesis structure

For the reader's convenience, a brief summary of the chapters is included here.

Chapter 2: This chapter provides context to the problem identified and presents a review and summary of the relevant literature. A range of topics are covered, including: a general introduction to modal analysis, OMA, uncertainty quantification for OMA, and Bayesian approaches to uncertainty quantification for OMA. The chapter concludes with some observations, and definition of the specific research objectives of this work.

Chapter 3: Introduces relevant theory necessary for understanding many of the methods presented in this thesis. These include state space models, probability and statistical methods, Bayesian statistics and inference, canonical correlation analysis (CCA), and the theory of stochastic subspace identification (SSI).

Chapter 4: This chapter presents a novel probabilistic formulation of the SSI algorithm, using the theory of probabilistic projections. By replacing the identification mechanism at the core of SSI, namely CCA, with its probabilistic equivalent, the problem of SSI can instead be considered one of probabilistic inference. This unique perspective lays the necessary mathematical foundation to enable a new suite of SSI-based OMA algorithms, constructed using sophisticated probabilistic and hierarchical techniques. These are explored in the later chapters.

Chapter 5: Using the probabilistic formulation defined in Chapter 4, this chapter introduces a statistically robust formulation of SSI, capable of robust modal identification when presented with atypical observations in the measured time series. Such an approach is only achievable using the new probabilistic viewpoint of SSI. This new method is compared using several simulated case studies containing different types of outlying data, and using data obtained from an in-service structure.

Chapter 6: This chapter is principally concerned with the main objective of performing Bayesian uncertainty quantification over the modal properties. By incorporating a prior structure over the parameters of the probabilistic SSI model, a new Bayesian formulation of SSI is formed. The proposed algorithm is capable of recovering posterior estimates over the modal properties, with Bayesian inference shown using two schemes: Gibbs sampling and variational inference. The uncertainty quantification performance of this algorithm is demonstrated on simulated data and

data from an in-service structure.

Chapter 7: Acknowledges the computational limitations of Bayesian methodologies, and presents one possible solution to making Bayesian SSI a more attractive choice for Bayesian UQ. Here an efficient implementation of Bayesian SSI, using stochastic optimisation techniques, is introduced. Expressions for performing Bayesian SSI using stochastic variational inference are defined, and applied to a simulated dataset. It is demonstrated how application of this stochastic methodology achieves a reduction in the number of mathematical operations using the entire dataset.

Chapter 8: Some concluding statements are offered as well as a summary of the novel contributions of this thesis. Finally, some limitations, challenges and opportunities following the work in this thesis are discussed, with possible avenues for future research identified.

2

Literature Review and Context

2.1 Introduction

This chapter provides context and a review of the current literature and existing approaches to: modal analysis, OMA, the adoption of uncertainty quantification (UQ) in OMA, and Bayesian approaches to UQ for OMA. This is followed by a summary on general observations in relation to the aim of this thesis (as defined in Section 1.4.1) and the chosen direction of this work. Any further literature is reviewed and discussed in the relevant chapters as and when required.

2.2 Modal analysis

Early forms of EMA for structural applications were first applied around the late 1940s in search of a better understanding of aircraft [29]. Subsequent developments in modal analysis were slow until the early 1970s when the commercial availability of Fast-Fourier Transform (FFT) spectral analysers and the rapid evolution of modern

computing, meant extracting this information became less of an arduous task. Since, modal testing equipment and data acquisition has greatly improved, with significant development into a comprehensive suite of algorithms designed to deal with a range of scenarios [6, 30–34]. Consequently, experimental modal testing and analysis has become standard practice in many engineering disciplines.

2.3 Operational modal analysis

Despite their effectiveness in a variety of scenarios, EMA algorithms have certain limitations; most notably their reliance on controlled, measured input information. When considering large-scale infrastructure or structures in hard-to-replicate environments, obtaining accurate input data is not always possible. The size of a structure or the type of loading it may experience, make lab-based testing infeasible. Instead, practitioners have turned to operational modal analysis (OMA) to analyse these types of structures.

OMA relies on the ambient excitation of a structure under normal operating conditions to replace artificial excitation. Such forcing arises from operational loads such as: wave loading on offshore structures, turbulent airflow over an aircraft, and human- or vehicle-induced vibration over a bridge. This alternative way of testing a structure not only removes the costly expense of producing controlled artificial excitation but also ensures the recovered modal properties are representative of the system in its intended environment.

Given the absence of input data, in OMA the input must instead be inferred based on a set of sensible assumptions. In most cases, the ambient excitation is assumed to be modelled by a stochastic quantity with unknown parameters but with known behavior, e.g. stationary white noise with zero mean and unknown covariance [6]. A white noise assumption means there are equal amounts of energy input into the system at all frequencies, i.e. a plot of the spectrum would show a horizontal line with equal magnitude at all frequencies. This assumption is particularly useful. Consider the frequency representation of a structural dynamic system [32],

$$Y(\omega) = H(\omega)X(\omega) \tag{2.1}$$

$X(\omega)$ and $Y(\omega)$ are the input and output spectra, respectively, and $H(\omega)$ is a

matrix of frequency response functions (FRFs) that relate to the underlying system dynamics. Therefore, if $X(\omega)$ is a white noise input of magnitude 1, then the output spectra $Y(\omega)$ are directly equal to $H(\omega)$ and thus representative of the underlying dynamics. However, this assumption is highly idealistic; no dynamic system in a real environment will ever truly experience white noise excitation. Therefore, the system learnt by OMA[1] is that described in Figure 2.1. A limitation of this pseudo-system is that the mass distribution of the true structural system is more difficult to retrieve as OMA does not return mass-normalised mode shapes [33]. Nevertheless, the modal properties identified through OMA provide sufficient information for many identification and decision-making tasks.

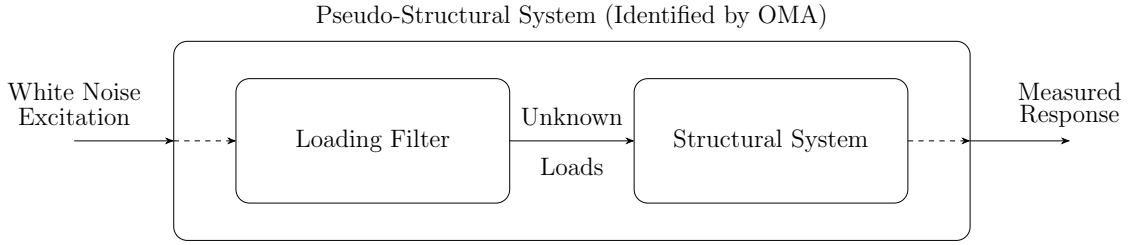


Figure 2.1: Structural system learnt by performing OMA under white noise excitation assumption. Adapted from [6].

Consequently, there have been many successful applications of OMA, including: modal testing of spacecraft [35], in-flight modal analysis of aircraft [36, 37], analysis of floating wind turbines [38], and the in-service modal assessment of a polar research vessel [39]. In particular, OMA has become the primary form of modal testing in the civil engineering community. This is reflected in the abundant literature and case studies on: bridges [40–44], footbridges [45], silos [46], masonry towers [47–49], lighthouses [21], high-rise buildings [50], stadiums [51–53] and dams [15, 54], amongst many others.

2.3.1 Methods

There are many established methods for performing OMA; this section briefly introduces the current landscape, although extensive review papers and textbooks are also available, see [6, 7, 33, 55].

[1]The pseudo-system introduced in Figure 2.1 is the simplest form. Often the measurement system makes up its own dynamic system and the disturbances it generates may not be perfectly white noise. If non-white measurement noise is modelled as white, the coloured noise is modelled as part of the identified system [33].

OMA (and EMA) methods are usually distinguished by the domain in which they operate numerically[2]: *time-domain* or *frequency-domain* [30–32]. Time-domain methods are based on the analysis of time-history responses or correlation functions, whilst frequency domain methods are based on the relation between input and output spectra. Generally, all measured data are obtained in the time-domain and, when required, transformed into the frequency-domain using the *Fourier Transform* [56, 57] or other appropriate algorithm or estimator [58, 59]. Each domain has its own associated merits and limitations (see e.g. [6]), although the main goal of modal analysis (the recovery of the modal properties) is achievable in both. Given the broad nature of OMA methods, this thesis focuses on the development of a time-domain OMA methodology to narrow the scope. Nevertheless, frequency methods are also mentioned here for the reader's benefit.

Frequency-domain

The most common frequency-domain methods in OMA include: peak-picking [31], Frequency-Domain Decomposition (FDD) [60], enhanced-FDD [61] and spatial-FDD [62]. A range of EMA methods can also be suitably adapted for OMA given the assumption of a white noise input such as the Maximum Likelihood Modal Method (MLMM) [63], Poly-reference Least-Squares Complex Frequency-domain (poly-LSCF) [64] or PolyMAX[3] [65, 66] algorithm, amongst others [67]. For a more comprehensive introduction to frequency-domain methods for system identification and OMA, the reader is directed towards [58] and [6].

Time-domain

There is also a comprehensive suite of time-domain approaches to OMA, the most notable methods are introduced here. The *natural excitation technique* (NExT) [68] is a time-domain method that exploits the auto- and cross- correlation functions of the output responses in time history. Although originally for EMA, NExT was naturally extended to OMA for structures under ambient excitation. In [68], it was shown that the correlation functions can be expressed as a sum of decaying sinusoids. Each sinusoid has its own frequency and damping ratio, equivalent to a corresponding structural mode, that can be identified. Therefore, when NExT is combined with an existing time-domain modal analysis algorithm like the subspace-based *eigensystem realisation algorithm* (ERA) [69], *poly-reference least-squares complex exponential* (poly-LSCE) [70], or *extended Ibrahim time-domain method* (EITD) [71], the resulting

[2]Early algorithms were also distinguished by the number of inputs and outputs they were able to analyse [31], although many of the algorithms used today allow multi-input multi-output analysis.

[3]LMS PolyMAX is a trademark of Siemens LMS Ltd.

algorithms can be considered OMA methods. Whilst NExT can be used for modal parameter estimation in OMA, it is important to note that the nature of data in OMA is stochastic while NExT-based methods operate in a deterministic framework [72].

Autoregressive moving average (ARMA) [73, 74] or *autoregressive moving average vector* (ARMAV) [75] procedures in the case of multiple observations and *time-varying ARMA* [76] for non-stationary systems, predict the current value of a time signal based on past values and some prediction error. These methods are usually linked with another system identification method like the prediction error method [77] or instrument variable method [78] amongst others [79]. ARMA methods are typically computationally intensive due to their iterative nature.

Blind source separation (BSS) techniques [80–82] originally emerged in the 1980s as signal processing tools for extracting individual audio sources from recordings, but have recently found their place in structural dynamics as non-parametric OMA methods. In signal separation problems, unknown signals (sources) and their contributions in the resulting mixtures (mixing matrix) are recovered as individual components. Although not initially recognised in earlier works [83–85], this method draws direct parallels with the theory of modal decomposition, where the sources are the modal responses and the mixing matrix the mode shape matrix [81, 86]. Once identified, the natural frequencies and damping ratios can be extracted from the independent components using classic single degree-of-freedom (SDOF) identification techniques. Several BSS techniques now exist in the literature including second-order blind identification (SOBI) [87], sparse component analysis (SCA) [88], time-frequency BSS [89] and AI-driven BSS [90]. A major advantage of BSS methods is their simplicity given their non-parametric form [91].

The *analytical signal method* (ASM) is a hybrid (time-frequency) method of OMA [92]. Using mono-component correlation functions, obtained through the inverse Fourier transform of the *power spectral density* (PSD), it is possible to estimate the natural frequencies and damping ratios, but not the mode shapes. A *time-domain analytical signal method* (TD-ASM) was recently introduced [93], removing the dependency on frequency, although similar limitations with recovering the mode shapes remain.

2.3.2 Stochastic subspace identification

Of the many time-domain algorithms, the most popular is the stochastic subspace identification (SSI) algorithm. The SSI algorithm was first introduced in the 1990's by Van Overschee and De Moor [94, 95] as a way of solving the *stochastic realisation problem* i.e. the identification of an effective stochastic state space model directly from measured data. A theoretical introduction to the state space representation and the stochastic realisation problem is provided later in Section 3.7.2. Principally, SSI recovers the parameters of a stochastic state space model by analysing the projection of shifted output sequences (a block Hankel matrix) onto itself, assuming stationary signals and white noise excitation[4]. This projection, when viewed as the conditional mean of the "future" observations (the bottom half of the block Hankel) given the "past" (the upper half), can be described in terms of covariances and solved for the parameters of the state space using the singular value decomposition (SVD). Although adopted as a new methodology, many people say the SSI algorithm draws similarities with the Ibrahim method and ERA algorithm.

Owing to its effectiveness in many cases, SSI rapidly gained popularity in the control systems community, drawing the attention of researchers in Belgium, Scandinavia and the Netherlands. In the late 1990s, SSI was then picked up in the physical engineering community, in particular structural dynamics, as a principled way of analysing mechanical and civil engineering structures [55, 96–98]. Since then, SSI has become the go-to-algorithm and industry standard for performing time-series OMA in mechanical, aerospace, offshore and civil engineering and is seen as a very reliable method of OMA.

The prominence of SSI is reflected by the abundance of available literature employing SSI to a broad range of problems and in a wide range of scenarios. For example, in [99] SSI was used to accurately estimate the machine-tool dynamics from random cutting excitation in rotary systems. In [100], OMA of an aircraft wing excited by transonic flow was performed using SSI and demonstrated superior identification capabilities over FDD in the case of non-white, non-stationary forcing contaminated by transonic disturbances. A new automated OMA algorithm was developed in [15] and applied to data from concrete dams, combining SSI with SOBI to leverage the

[4]Note that subspace methods like SSI typically do not require stationarity of the noise for consistency, as these methods primarily rely on the structure of the data rather than specific noise properties. However, stationary noise is often assumed for uncertainty quantification to ensure reliable statistical modeling and error estimation.

useful properties of SSI in the analysis. In [101], OMA of a monopole communications tower was conducted, with the SSI results acting as the reference to verify results obtained through other methods. And in [102], SSI was shown to be applicable to data obtained from an offshore floating wind turbine, despite several assumptions for time-series OMA being violated.

In general, there are two main SSI algorithms: covariance-driven SSI (SSI-Cov), and data-driven SSI (SSI-Data) [95, 103]. The SSI-Cov algorithm proceeds by analysing the output cross-covariance matrix of the block Hankel, which can be decomposed in two ways: as the product of the observability matrix and controllability matrix, and through the singular value decomposition (SVD). Using this information, it is possible to recover an estimate for the state matrix from the observability matrix and thus the modal properties through an eigenvalue decomposition. Alternatively, SSI-Data avoids the computation of the covariances by using the QR-decomposition of the block Hankel matrix to project the future output onto the past output. The SVD of the resultant projection can then be used to identify the modal properties in a similar way.

Several variants of SSI exist in the literature. They primarily differ in the weightings of the block Hankel matrix before applying the SVD. The primary weightings are: (1) *canonical correlation* or *canonical variate analysis* (CVA); (2) *principle component* (PC); and (3) *unweighted principle component* (UPC) [95]. Each weighting determines the state-space basis in which the model will be identified; although no significant accuracy difference has been seen between the CVA, PC and UPC implementations [55, 104]. Such weightings can be applied to both SSI-Cov and SSI-Data. The derivation of the CVA-weighted SSI-Cov algorithm is introduced later in Section 3.7.2 and forms the basis for much of the work contained within this thesis.

There have also been several extensions to SSI to aid its applicability in a wider range of scenarios and in an attempt to address some of the research challenges highlighted in Chapter 1. In 1999, Peeters and De Roeck [105] introduced a referenced-based SSI (SSI-Cov/Ref) algorithm that enabled the analysis of larger structures in multiple stages, using reference sensors across the structure to obtain global mode shape estimates, thus reducing the computational effort. This is achieved through the inclusion of reference sensor data in the Hankel matrix construction. Others have looked to improve the robustness of SSI to noisy or corrupted data by providing alternative models that seek to minimise their effects on the identification procedure [20, 106]. In [107], the computational efficiency of SSI for larger datasets is considered,

with randomised matrix algebra suggested as a possible way to address this challenge. In the interest of continuous online monitoring, a range of authors have presented possible recursive SSI algorithms [108–110]. One other key challenge some authors have sought to address is the handling and quantification of uncertainty. This has led to the development of new techniques for performing uncertainty quantification for SSI; these methods will be discussed in more depth in the context of the next section.

2.4 Uncertainty quantification for OMA

As highlighted in the introduction, one of the key challenges in OMA is noise and uncertainty. Multiple sources of uncertainty present in observed data can result in the variability of the recovered modal parameters and the models used to defined them. The impacts of uncertainty on operationally obtained modal identification have been studied in depth [111], whilst a comprehensive list of the associated uncertainties specifically for subspace methods like SSI have been explored by Reynders in [112]. The dependency of other structural dynamics applications on recovered modal information, such as structural health monitoring (SHM), for the purpose of decision-making, necessitates better understanding and assessment of uncertainty in OMA. Consequently, new and alternative approaches to incorporating uncertainty and performing UQ for OMA are of particular value.

Before exploring existing UQ approaches to OMA, it is worth briefly mentioning how different people may choose to represent uncertainty and what forms it can take. Across different texts (in engineering literature and further afield), the representation of 'uncertainty' can differ; it can take many forms not limited to: bounds, confidence intervals, fiducial intervals, variance estimates, and distributions. There is no explicit definition for the "correct" type of uncertainty in any case, however the overall objective of UQ is the same: To obtain a measure of uncertainty to better understand the variability of a given result. The author's preference is to operate within a Bayesian framework where the recovery of *posterior distributions* is ultimately desired, hence the focus of this thesis. An introduction to Bayes theorem and the benefits of posterior uncertainty are introduced later in Section 3.2.

UQ methods for OMA have been studied from both Bayesian and frequentist or non-Bayesian perspectives. Frequentist approaches assume that there exists some true

values for the parameters, but can through statistical analysis recover an estimate for the uncertainty; often referring to the variability arising from the data. Whereas in Bayesian approaches, modal identification is viewed as an inference problem, where prior information and data are combined to determine a probability of the plausibility of many outcomes i.e. a posterior. This section will first introduce existing probabilistic approaches to UQ for OMA, with a specific focus on methods centred around SSI, followed by an exploration of the current Bayesian UQ methods for OMA. Both time- and frequency- domain methods are discussed.

2.4.1 Frequentist uncertainty quantification for OMA

There are several noteworthy contributions present in the literature for frequentist UQ for OMA. One of the earliest examples is the frequency-domain MLMM algorithm [63], which provides an estimate of the modal parameter variance based on an estimated variance of the FRFs. The variance on the modal parameters is computed from a linearization of the Cramèr-Rao lower bound and theoretically achieved by maximum likelihood estimation. This method, although originally a method of UQ for EMA, can be suitably adjusted for output-only identification using PSDs instead of FRFs.

In [113, 114] an alternative approach to obtaining the variance on FRFs was used. This variance was obtained through two different procedures, Monte Carlo and bootstrapping. The Monte Carlo method involves repeated simulation of randomly perturbed datasets based on the statistics of measured FRF data and some assumed variability. Statistical analysis is then performed to obtain statistical distributions over the modal properties. The bootstrap procedure relies on multiple sets of measured FRFs obtained from the same structure. The premise of the algorithm is to randomly sample the FRFs with replacement, forming an ensemble average FRF from which the modal properties can be recovered. Repeating this process, statistical confidence intervals on the modal properties can be estimated. However, in both cases, neither method can account for any bias introduced from the initial FRF estimation.

Since the 1990s, several works explored the asymptotic normality of estimates from subspace algorithms [115–118] and demonstrated their ability to converge to the expected value under the central limit theorem. The theoretical basis shown in these

works has led to a number of subsequent methods for UQ, exploiting these results. In 2007, Pintelon et al. [119] proposed a combination of first-order sensitivity and perturbation techniques on the parameters of a rational polynomial representation of a transfer function to obtain variances on the modal parameters. This was swiftly followed by application to the time-domain method SSI-Cov/Ref by Reynders et al. [112]. The methodologies outlined in [112, 119] estimate a variance or random error, rather than systematic error (bias), of measurements based on the sensitivity of the modal estimates to perturbations in the observed responses. Although presented for SSI, this methodology is also generally applicable for other correlation-based methods such as ERA [120].

Since [112], there have been significant developments in improving and extending perturbation-based UQ for SSI. This is partly because of the existing praised performance of traditional SSI methods in many OMA settings, consistently demonstrating good estimation accuracy and computational robustness. In 2013, Döhler et al. [121] efficiently optimised the method from [112] through rederivation, reducing the computational burden when producing the stabilisation diagram. Moreover, this method was later extended to include multiple-setup measurements [122], input-output analysis [123, 124], UQ for Data-driven SSI [125] and frequency-domain SSI [126]. Consequently, this method has seen application to several practical scenarios, e.g. bridges [42, 127, 128], and has resulted in many other extensions and relevant case studies [129–132]. The outputs of perturbation-based UQ methods for SSI have already proven particularly fruitful, with researchers exploring the use of this uncertainty to address other research challenges such as automated modal identification [17, 129] and model order selection [133].

Other approaches using sensitivity and perturbation techniques have also been presented. In [134, 135], a fast two-step approach to determining the variances on the modal properties using the frequency-domain method PolyMAX was proposed. This approach is based on the linearisation of the sensitivity of the modal parameters to the noise variance, although the noise variance (either explicitly or as a coherence) must be known a priori. Similarly, El-Kafafy et al. [136, 137] introduced a fast MLMM identification technique, using the same approach.

Carden and Mita [138] note that one of the shortcomings of some of these algorithms is the assumption of Gaussian noise on the modal properties or summary of the uncertainty by only the first one or two statistical moments. It has previously been shown that the uncertainty over the modal parameters may exhibit non-normal

distributions [139] specifically when using shorter time-series or analysing systems with lower natural frequencies. It is suggested that variance estimates alone are at times insufficient to describe the uncertainty over the modal parameters and that other techniques may be needed if one was to require greater insight on the underlying distributions. Nevertheless, in the opposing case, with longer time series and higher frequencies, this approximation is generally applicable as the number of observations approaches the infinite limit.

Generally, these frequentist approaches to UQ do not recover a global probability distribution over the parameter/s in question, often only a maximum likelihood estimate and some interpretation of variance; the uncertainty recovered typically describes the ensemble variation in the data, not that arising from the model itself. In contrast, a Bayesian SSI methodology would provide a framework for incorporating prior knowledge or permitting hierarchical structures whilst also recovering a global picture of the uncertainty as posterior distributions. Bayesian approaches can quantify not only the experimental uncertainties, but also any model uncertainty due to inaccuracies in the mathematical definition, describing the entire probability distribution over the parameters in question and offering richer information about the uncertainty. Furthermore, Bayesian methods do not rely on any asymptotic assumptions and performs well with limited data. Consequently, researchers have also looked to address the problem of uncertainty in OMA from a Bayesian viewpoint – these methods will now be explored.

2.4.2 Bayesian uncertainty quantification for OMA

This section introduces the current state of the art Bayesian approaches to OMA (in both time- and frequency- domain), highlighting the associated merits and limitations of existing methods. The major waypoints here include the popular frequency-domain method BAYOMA [140] and more recent alternative approaches, these include: a variational scheme in the time-domain; and two Gibbs sampling approaches (one in the time- and other in the frequency-domain).

Bayesian approaches to output-only system identification have been developed as early as the 1980s [141] but the topic continues to be of significant interest to the engineering community [142–147]. The first approaches labelled Bayesian OMA are much more recent, with the earliest works appearing in the late 2000s. The most cited

Bayesian approach to OMA follows from the work of Au, whose development of a fast Bayesian Fast Fourier transform algorithm (fast-BFFTA) for modal identification [148, 149] removed the computational limitations of the original Bayesian Fast Fourier transform algorithm [150, 151] and subsequently led to a body of work known broadly as Bayesian OMA or 'BAYOMA' [19, 140, 152–154]. BAYOMA can be perceived as the coupling of the fast-BFFTA with a frequency-domain OMA technique, such as FDD [60]. This coupling can be used to recover modal estimates in the form of a most-probable value and a representation of the uncertainty known as the coefficient of variation. These estimates originate from Laplacian approximations of the uncertainty which are used to obtain assumed Gaussian distributions. Since the initial definition of BAYOMA, several extensions have emerged in the literature providing a more general framework to use this methodology, including work on the identification of closely-spaced modes [155–158].

BAYOMA appears to divide some of the research community in its Bayesian definition. In the construction of the algorithm, a flat improper prior is used (see later in Section 3.2.2). Under this assumption, some argue (see [127]) and this author agrees, that the problem fundamentally reduces to a maximum likelihood approach as the prior is not truly valid throughout its entire range. This results in a posterior that is essentially the standardised likelihood, removing the many benefits of a Bayesian formulation. Nevertheless, some in the field adopt BAYOMA as their preferred UQ method of choice [21, 159–161]. Undoubtedly, however your persuasion, BAYOMA does provides a computationally effective way to perform one form of UQ on structures of interest but is seemingly limited in its approximation of the posterior using improper priors.

A few alternative Bayesian OMA approaches have been published in recent years that do employ proper priors. Li and Der Kiureghian [162] proposed a variational Bayes methodology for OMA applied directly to stochastic state space models (SSMs), where joint distributions over the state-transition and observation matrices, and the process noise and measurement error, are calculated analytically. The analytical solutions are then coupled with a first-order Taylor series expansion to recover Gaussian approximations to the distributions over the modal properties. This is necessary given the intractability of the posteriors over the modal properties because of the eigenvalue decomposition involved in their recovery. Alternative assumptions made in their model, distinguish this methodology from much earlier work on variational methods for the identification of SSMs in [163, 164].

Another Bayesian approach was also presented by Li et.al [165]. Their contribution focused on a new form of Bayesian OMA for civil structures under small or moderate seismic excitation. A probabilistic model is first defined, taking advantage of the stochastic state space representation of the equations of motion, with some unmeasured base motion included in the model as a stochastic process. Proper and broadly uninformative priors are then introduced into the model, with the Bayesian inference problem solved using a forward-filtering-backward-sampling algorithm (a Gibbs sampler) to obtain approximate distributions over the modal properties by generated samples.

Yang and Lam [146] recently introduced a time-domain autoregressive (AR) model for performing Bayesian OMA to MDOF systems. From their Bayesian AR model, the most-probable values and variance of the errors over the AR model parameters are obtained. A covariance over the parameters can then be obtained through Laplacian approximation, with the posterior variance over the modal properties estimated as approximately Gaussian through first-order Taylor series expansion.

In 2020, Sedehi et al. [166] presented a hierarchical Bayesian modelling framework for frequency-domain UQ in OMA using hyper-probability distributions. After defining a re-parameterised model of a dynamic system, separating dynamic and non-dynamic parameters, the hyper-distribution over the dynamic properties can be learned. In their paper, this inference is solved through MCMC and shown using Dual Laplacian approximation to recover posterior estimates over the modal properties.

In [167, 168], a new Bayesian OMA method based on modal-component-sampling (Bayes-Mode-ID) was introduced. The method is based on the idea of the modal identification method (Mode-ID) developed by Beck [169–171]. In the paper, a Bayesian posterior probability density function (PDF) over the modal properties, conditioned on the cross-correlations of the observed data, is formed. Bayesian UQ is then conducted by generating samples of the modal parameters in the important regions in the posterior PDF, considering one mode or component at a time. The sampling process is then iterated considering different modal components until convergence. Unlike other methods, the posterior on the modal properties can be obtained without approximating the posterior PDF to be Gaussian in the first instance.

Finally, the most recent algorithm presented was a fast-collapsed Gibbs sampling approach to frequency-domain OMA introduced in 2022 by Dollon et al. [172]. This

alternative proposal performs inference by first defining a hierarchical structure over the modal decomposition of an SDOF linear system. The uncertainty over the parameters of the modal decomposition can then be established using fast Gibbs sampling techniques and the PSD of a system with well-separated modes. Although for SDOF systems, this methodology can be extended to MDOF by analysing each mode separately. Their analysis results in posterior distributions generated from samples that are characterised by a mean and covariance, rather than a most-probable value and coefficient of variation.

2.5 Observations and summary

Considering the existing approaches to UQ for OMA highlighted in this chapter, it is evident there is considerable and continued interest in performing UQ for OMA and developing new reliable methods for doing so. UQ for SSI has seen particular attention, given its existing adoption and many successes in civil and mechanical applications, whilst Bayesian methods are of growing interest, given their inherent ability to encapsulate uncertainties from data and models.

Upon reflection, it appears (as far as this author is aware) that almost all existing approaches to UQ for OMA predominantly perform quantification without changing the underlying OMA algorithm being applied. That is to say any frequentist or Bayesian techniques are applied to the inputs-to or outputs-of OMA algorithms — rather than embedding knowledge of uncertainty within the identification procedure. For example, estimates of uncertainty are found through the statistical analysis or sensitivity of identified parameters, or in the case of BAYOMA, through Bayesian analysis on the inputs, which are propagated through an existing OMA algorithm. There are some recent approaches that are beginning to incorporate knowledge of uncertainty within a model's definition (e.g. [162]). However, these techniques either rely on the modal uncertainty being approximated as Gaussian in their construction or are frequency-domain only techniques.

Based on the literature reviewed in this chapter, there is a motivation for the development of new techniques that can embed knowledge of uncertainty within an algorithms construction and identification mechanism. This author hypothesises that OMA algorithms constructed from a probabilistic viewpoint can provide the necessary framework to better consider noise in data, and more importantly facilitate a Bayesian

form; one where posterior distributions over modal properties are obtainable and hierarchical extensions possible.

Despite the success of SSI as one of the main time-domain OMA algorithms, and the attention it has received in frequentist UQ for OMA, at the time of writing, the author is unaware of any Bayesian formulation of SSI in the literature. As such, this thesis aims to address this gap by introducing a novel probabilistic form of the SSI algorithm, one that can be conveniently extended using hierarchical methods to a statistically robust form and ultimately a Bayesian one.

3

Theory

This chapter introduces key mathematical and technical concepts beneficial for understanding the new methodologies presented in this thesis. Topics include: the dynamics of systems, the stochastic realisation problem, a brief introduction to probability theory, Bayes' rule, and Bayesian inference. For the more familiar reader, the author suggests specifically reviewing the sections on Bayesian inference techniques (Section 3.3), canonical correlation analysis (Section 3.4) and stochastic subspace identification (Section 3.7.2) in the interest of the later Chapters.

3.1 Probability

A brief overview of the basics of probability theory is provided to assist the reader with concepts presented later in the thesis; for further detail, the reader is referred to the following text-books [173–175].

The expression $P(A)$ denotes the *probability* that a given event A is true. A given probability requires that $0 \leq P(A) \leq 1$, such that $P(A) = 0$ implies an event **will almost surely not** happen, and $P(A) = 1$ implies an event **will almost surely** happen, except at measure zero.

Considering the case of non-binary events (more states other than true or false), one can use *random variables*. A discrete random variable, X, can take any value from a finite (or countably infinite) set $\tilde{\mathcal{X}}$; whilst continuous random variables can take *any* value within a given feature-space $\mathcal{X}$. In the context of this thesis, almost all random variables are considered to be continuous. A continuous random variable is associated with a continuous probability density function (pdf) (denoted by a lower-case p). The pdf of a random variable X in the interval $a < X \leq b$ is,

$$P(a < X \leq b) = \int_a^b p(\mathbf{x}_i)\mathrm{d}\mathbf{x}_i \tag{3.1}$$

such that $p(\mathbf{x}_i) \geq 0$, and $\int_{\mathcal{X}} p(\mathbf{x}_i)\mathrm{d}\mathbf{x}_i = 1$.

3.1.1 Some brief probability theory

The probability of the union of two events, the probability of A *or* B occurring, is

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{3.2}$$

$$= P(A) + P(B) \quad \text{(i.f.f A and B are mutually exclusive)} \tag{3.3}$$

The probability of the *joint* occurrence of two events A *and* B, is

$$P(A, B) = P(A \cap B) = P(A|B)P(B) \tag{3.4}$$

where $P(A|B)$ is the conditional probability of an event A happening, *given* that B has already occurred. It is said that event A has been *conditioned* on B (see Equation 3.8). Equation 3.4 is also known as the *product rule*.

Given a joint distribution, $P(A, B)$, specific variables can be *marginalised* out by summing or integrating over all possible states of that variable. This gives the rule of marginalisation,

$$P(A) = \sum_b P(A, B) = \sum_b P(A|B = b)P(B = b) \quad \text{(discrete)} \tag{3.5}$$

$$p(A) = \int_{\infty}^{\infty} p(A|B)p(B)\mathrm{d}B \quad \text{(continuous)} \tag{3.6}$$

This method is often called the *total rule of probability* or *marginalisation rule*. The same method can also be applied to find $p(B)$.

The product rule can be applied multiple times (to continuous or discrete variables) to yield the *chain rule* of probability,

$$p(X_{1:D}) = p(X_1)p(X_2|X_1)P(X_3|X_2, X_1) \dots p(X_D|X_{1:D-1}) \tag{3.7}$$

where $X_{1:D} \triangleq X_1, X_2, X_3, \dots, X_D$

Also from the product rule, the conditional probability of A, given that B has occurred, can be defined as

$$p(A|B) = \frac{p(A, B)}{p(B)} \tag{3.8}$$

3.2 Bayesian statistics

The aim of this thesis is to develop Bayesian UQ methods for OMA applications. This section will introduce the reader to *Bayes' Rule* (or Bayes' Theorem) and briefly introduce the concept of Bayesian statistics. For a more comprehensive introduction to the topic, the reader is directed towards [175] and [176].

Bayesian statistics is a framework of statistical inference where probability is interpreted as a *degree of belief* or *subjective uncertainty* about some unknown quantity. It is grounded in Bayes' Theorem, which provides a formal mechanism for updating probability distributions as new data become available. Unlike frequentist approaches to uncertainty, which interpret probability as the limiting frequency of an event over many instances and typically describe aleatoric uncertainty, Bayesian methods incorporate prior knowledge and observed data under a chosen model to give probability as a measure of epistemic and aleatoric uncertainty.

3.2.1 Bayes' rule

Bayes' theorem, aptly named after Reverend Thomas Bayes[1], describes the probability of an event occurring given some prior knowledge about the conditions of that event. It plays a central role in some modern day machine learning algorithms [173] and forms the basis for a large proportion of the work presented in this thesis.

Bayes' theorem is simple to derive using two key identities from probability theory; those already introduced as the chain rule, which describes the relationship between joint and conditional probabilities,

$$p(A, B) = p(A|B)p(B) \tag{3.9a}$$

$$p(B, A) = p(B|A)p(A) \tag{3.9b}$$

and the law of marginalisation,

$$p(A) = \int p(A|B)p(B)\mathrm{d}B \tag{3.10}$$

[1]Reverend Thomas Bayes (1702-1761) was an English mathematician and Presbyterian minister, whose solution to the problem of inverse probability was published posthumously.

As the chain rule for probabilities can be applied in either order, the derivation of Bayes' rule becomes trivial,

$$p(A, B) = p(B, A) \tag{3.11}$$

$$p(B|A)p(A) = p(A|B)p(B) \tag{3.12}$$

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)} \tag{3.13}$$

This is a sensible point to introduce common terminology for the components of Bayes' rule. By convention, $p(B|A)$ is known as the *posterior*; $p(A|B)$ as the *likelihood*; $p(B)$ as the *prior*; and $p(A)$ as the *evidence* or the *marginal*.

Bayes' theorem is often combined with the law of marginalisation to aid with interpretation (and also computation) to give,

$$p(B|A) = \frac{p(A|B)p(B)}{\int p(A|B)p(B)\mathrm{d}B} \tag{3.14}$$

3.2.2 The prior

The prior probability distribution, often simply called the *prior*, of an uncertain quantity is the probability distribution that encodes ones initial belief or knowledge of a variable, before any data has been observed. The prior is a fundamental part of Bayes' rule and is used to keep the posterior distribution 'in check'. Under limited data, the shape and location of the posterior is predominantly driven by the prior and less by the data. This is logical as, given little data, one would be cautious to deviate from ones prior belief without seeing sufficient evidence to update their model. As the amount of observed data increases, the prior becomes less important and has less influence on the shape and location of the posterior; there is now sufficient evidence to suggest a different model choice than that of the prior.

Figure 3.1 shows an example prior[2] and subsequent posterior distributions for a simple case study with known variance and unknown mean. It is evident from the figure that, as the number of data increase, the posterior hones in on the true value and relies less on the prior.

[2]Note that this would traditionally be classified as a poor prior because of its limited coverage.

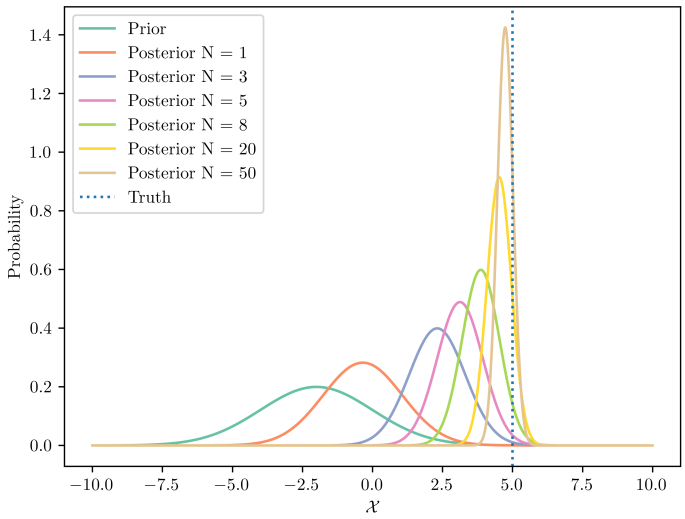


Figure 3.1: An example prior and set of posteriors,given a varying number of data in the likelihood.

In many applications of probabilistic inference, ones prior knowledge can be conveniently expressed through the prior distribution. These are known as *informative priors* and may include distributions such as the Wishart, Gaussian, Beta and Gamma. In some cases, however, there may be little useful information to confidently decide what form the prior distribution should take. Instead, a set of priors known as *uninformative, flat, or diffuse priors* can be used which are intended to have as little effect on the posterior probability as possible. Example priors include the: Jeffreys [177], Box and Tao [178], and Bernardo and Smith [179]. Depending on how well a distribution describes our initial belief, sometimes a prior is said to be *weakly informative*. A weakly informative prior expresses partial information about a variable, informing the analysis towards possible solutions that align with existing knowledge, but without over constraining the model.

It is also necessary for later chapters to make the distinction between *proper* and *improper* priors. A proper prior is defined as a prior whose pdf (over the entire sample space) integrates to 1. A concept basic to the definition of a probability density (see Section 3.1).

$$\int_{-\infty}^{\infty} p(\theta)\mathrm{d}\theta = 1$$

Conversely, the pdf of an improper prior does not integrate to 1,

$$\int_{-\infty}^{\infty} p(\theta)\mathrm{d}\theta = \infty$$

An example of an improper prior is the uniform distribution on an infinite interval. Despite an improper prior, the right choice and combination of priors with an ordinary likelihood can still result in a posterior which is proper. A uniform distribution combined with a normal likelihood would give a standardised likelihood as the posterior; one in which the posterior is dominated by the likelihood. An improper density such as this can be thought of as been valid over some large range of values, but not truly valid throughout its entire range. In the case the prior describes some variable to be constant over a very long interval, with it never being large outside that range, then the posterior is very close to the standardised likelihood; equivalent to the posterior resulting from taking an improper prior over the whole real line [176].

There is often discourse regarding the choice of a uninformative, improper prior in recovering a Bayesian posterior. Problems of this nature essentially reduce to a maximum likelihood approach, as the whole posterior is dominated by the standardised likelihood. In this case, the maximum a-posteriori (MAP) estimate is equivalent to the frequentist MLE [3].

3.2.3 Conjugate priors

The marginal integral (the denominator of Bayes' theorem 3.14) can often be difficult to compute directly [27], so it is common to choose from a restricted class of priors, Π, that ensure the marginal integral exists in closed form; this is usually a prior belonging to the exponential family. This is often done purely for convenience but also ensures the posterior distribution has the same functional form as the prior. Using a closed-form solution of the marginal and therefore the posterior, for a given likelihood, can generally lead to faster computation. These types of priors are conjugate to the likelihood and are aptly named *conjugate priors*. Unfortunately, the available set of conjugate priors is very limited.

[3] This is the reasoning given by some who believe BAYOMA is not a "fully" Bayesian approach.

3.2.4 Posterior

The posterior distribution, as the combination of prior knowledge and evidence, provides a complete representation of the uncertainty about the parameters, after observing data. Unlike point estimates (i.e. the frequentist approach), a Bayesian posterior gives a full probability distribution, offering greater insights into:

- The most likely values (the peak of the posterior distribution)
- The uncertainty or spread of possible values (the variance)
- The shape of the distribution (indicating skewness, multimodality, etc.).

Bayesian methods are also particularly flexible; they can handle complex and challenging problems where traditional methods may struggle. Bayesian methods can easily incorporate:

- Non-linear relationships
- Hierarchical structures (multi-level)
- Missing Data
- Complicated likelihood functions

This makes posterior distributions particularly useful in the field of engineering, where often these problems are encountered. This is also true of the models presented later in this thesis.

3.2.5 Probabilistic graphical models and Bayesian networks

Problems in probability theory with complex joint distributions can be elegantly captured by *probabilistic graphical models* (PGMs) [173, 180]. Fundamentally, PGMs provide a simple graphical representation of more complex joint distributions, exploiting inherent structure to describe them more compactly, in a way that feels intuitive and therefore more easily communicated. In this thesis, only a subset PGMs, known as *Bayesian networks*, are used. Bayesian networks are specifically directed acyclic graphs, where *nodes* represent random variables, shaded nodes represent known quantities (i.e. observed data) and *edges* imply some form of conditional dependency between these variables. An example Bayesian network is shown in Figure 3.2.

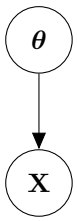


Figure 3.2: Example PGM with some random variable $\mathbf{X}$ and random parameters $\boldsymbol{\theta}$.

The joint probability of the model above can be written as,

$$p(\mathbf{X}, \boldsymbol{\theta}) = p(\mathbf{X}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{3.15}$$

In the situation above, $\boldsymbol{\theta}$ is the *parent node* of $\mathbf{X}$ and $\mathbf{X}$ is the *child node* of $\boldsymbol{\theta}$ such that there is direction to their relationship. A node may have multiple children, or multiple parents.

If multiple random variables share the same probabilistic model or set of attributes, then it is represented via a *plate* model. In plate representation, a template for the repeated structure and shared parameters is encoded into the PGM by drawing the relevant nodes and edges and enclosing them in a box. The box represents an entire stack of identical plates, with a specified range R. An example model is shown in Figure 3.3.

The associated joint in this case can be written as

$$p(\mathbf{X}, \boldsymbol{\theta}) = p(\boldsymbol{\theta}_X) \prod_{r=1}^{R} p(\mathbf{X}_r|\boldsymbol{\theta}_X) \tag{3.16}$$

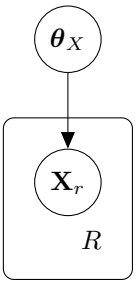


Figure 3.3: Example PGM with some random variable $\mathbf{X}_r$ inside a plate with range R and random parameters outside the plate $\boldsymbol{\theta}$.

One could also envisage a scenario where the variable $\boldsymbol{\theta}$ is inside the plate, encoding a relationship specific for each plate. Furthermore, one could also consider the case where multiple nodes of different types, existing within or outside the plate, are required to define a model.

PGMs can also be used to show hidden or *latent* variables and their relation to some observed variables. This will be the case with models shown later in this thesis. An example model with this architecture is that of probabilistic principal component analysis (PPCA) [181]. This model is shown in Figure 3.4.

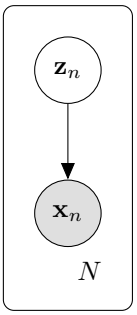


Figure 3.4: PGM describing PPCA

where $\mathbf{z}_n$ is a latent variable and $\mathbf{x}_n$ is an observed variable conditioned on $\mathbf{z}_n$ by the relationship $\mathbf{x}_n|\mathbf{z}_n \sim \mathcal{N}(\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu}, \sigma^2\mathbb{I})$ for N observations. In this case, the parameters $\boldsymbol{\theta} = \{\mathbf{W}, \boldsymbol{\mu}, \sigma\}$ are not random, but are to be determined[4]. The joint of the model shown above can be written as,

$$p(\mathbf{x}, \mathbf{z}) = \prod_{n=1}^{N} p(\mathbf{z}_n)p(\mathbf{x}_n|\mathbf{z}_n) \tag{3.17}$$

[4]The parameters in $\boldsymbol{\theta}$ could be considered random variables themselves, extending this model further. This type of model will be considered later in Chapter 6.

In many problems represented by PGMs, the form of the joint restricts the ability to obtain a closed form maximum likelihood solution, particularly when considering multiple nodes and their conditionals. Therefore, alternative methods of inference can be used to obtain these estimates.

3.3 Estimation and inference

Solving complex problems in probabilistic or Bayesian inference, where exact inference is not possible, requires algorithms capable of estimating parameters or posterior uncertainty. In probabilistic problems this is known as *parameter estimation* or in a Bayesian sense, known as *inference*. The choice and performance of different methods depends on multiple factors, including the model, computational expense, and either a desire for approximate posteriors or more exact solutions. Three methods of estimation and inference used in this thesis are introduced in this section.

3.3.1 Expectation-maximisation

Computing the maximum likelihood estimate (MLE) of parameters of interest in many machine learning and statistical models is a rather trivial task when all the values of all the relevant random variables, are observed i.e., in the presence of complete data. However, if data are missing or latent (hidden) variables are to be learned, then computing a closed form solution for the MLE becomes completely inaccessible [173]. One way to approach this problem is to use a method called the *expectation-maximisation* (EM) algorithm [182].

The premise of the EM algorithm is to alternate between estimating missing values or latent variables in an *expectation step* (E step), and then computing the MLE of the parameters using this 'filled in' data in a *maximisation step* (M step); exploiting the simple fact that if data were fully observed, then finding the MLE would be easy to compute. As the expected values depend on the parameters, and the parameters depend on the expected values, this algorithm is iterated until convergence at the global (or local) maximum [183]. The EM algorithm generally proceeds as follows:

Goal: Find an updated set of parameters $\hat{\theta}$ such that

$$\hat{\theta} = \arg\max_{\theta} \log \left(\sum_{\mathbf{z}} p(\mathbf{x}, \mathbf{z}|\theta) \right)$$

E Step: Compute the expected complete log likelihood,

$$\mathbb{E}_{\mathbf{z}|\mathbf{x},\theta^{(t)}} \left[\log p(\mathbf{x}, \mathbf{z}|\theta^{(t)}) \right] = \sum_{z} \log p(\mathbf{x}, \mathbf{z}|\theta^{(t)}) p(\mathbf{z}|\mathbf{x}, \theta^{(t)})$$

M Step: Solve the maximisation,

$$\theta^{(t+1)} = \arg\max_{\theta} \sum_{z} \log p(\mathbf{x}, \mathbf{z}^{(t)}|\theta) p(\mathbf{z}^{(t)}|\mathbf{x}, \theta)$$

3.3.2 Markov chain Monte Carlo

Markov chain Monte Carlo (MCMC) methods are a powerful class of algorithms used to sample from complex, high-dimensional probability distributions, particularly in Bayesian statistics. When direct sampling is intractable because of the complexity of a posterior distribution, MCMC provides a practical alternative. The premise of MCMC algorithms is to construct a Markov chain whose stationary distribution converges towards a target probability distribution (or density) from which you want to sample i.e. a posterior. After running the algorithm for a sufficient number of iterations (allowing the Markov chain to mix or "forget" its starting point), the states visited by the Markov chain approximate the target distribution. The more steps in the Markov chain, the more closely the approximated distribution will match the desired target distribution [173]. The most prominent MCMC methods are *Metropolis-Hastings* and a special case of the algorithm, *Gibbs sampling*.

3.3.3 Metropolis-Hastings

The Metropolis-Hastings algorithm is an MCMC technique that generates a sample by proposing a new state based on a proposal distribution, and accepting or rejecting that sample given a probability that ensures the Markov chain converges to the target. Specifically, for a current state $\boldsymbol{\theta}$, a new state $\boldsymbol{\theta}'$ is proposed according to

some proposal distribution $q(\boldsymbol{\theta}'|\boldsymbol{\theta})$. The acceptance probability is defined as:

$$\alpha = \min\left(1, \frac{\pi(\boldsymbol{\theta}')q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{\pi(\boldsymbol{\theta})q(\boldsymbol{\theta}'|\mathbf{x})}\right) \tag{3.18}$$

where $\pi(\boldsymbol{\theta})$ represents the target distribution under data up to a constant. If the proposed state is accepted, then the chain can move to $\boldsymbol{\theta}'$; otherwise it remains at $\boldsymbol{\theta}$. This algorithm ensures that the chain explores the space in a way that samples the target distribution properly over time. A key feature of the Metropolis-Hastings algorithm is its flexibility, as the proposal distribution can be chosen in a variety of ways; although this must be done carefully to balance exploration of the relevant space with computational efficiency.

3.3.4 Gibbs sampling

Gibbs sampling is considered a special case of the Metropolis-Hastings algorithm (where $\alpha = 1$) that is particularly useful when trying to sample from high-dimensional probability distributions where direct sampling from the joint distribution is difficult, but sampling from the conditional distributions is more practical. Gibbs sampling is commonly used as a means of Bayesian inference as it is well-adapted to sampling from the posterior distribution of large Bayesian models [173]. Instead of proposing a new state for all the parameters at once, Gibbs sampling updates each parameter in turn, conditioning on the current values of the other parameters. For a target distribution, $\pi(\theta_1, \theta_2, \ldots, \theta_n)$, Gibbs sampling iteratively updates each parameter θ_i by sampling from the full conditional distribution $\pi(\theta_i|\theta_{_i})$ where $\theta_{_i}$ represents all the other variables. Since any proposals are always accepted ($\alpha = 1$), Gibbs sampling can be much more efficient than the Metropolis-Hastings algorithm when the conditional distributions can be easily sampled.

The general implementation of Gibbs sampling is described in Algorithm 1.

When applying the algorithm, the order of the sampling procedure does not matter as long as all variables have been updated to the new state $\tau + 1$, before updating to the state $\tau + 2$. Following successful application of Gibbs sampling, the following hold true (also true of all MCMC methods):

- The samples approximate the joint distribution of all the variables.

Algorithm 1 Gibbs Sampling

initialise: $\mathbf{z}^{(0)} = z_1^{(0)}, \ldots, z_m^{(0)}$ for $p(z_1^{(\tau)}, z_2^{(\tau)}, \ldots, z_m^{(\tau)})$
for $\tau = 1 \rightarrow k$ samples **do**
  Sample $z_1^{(\tau+1)} \sim p(z_1 | z_2^{(\tau)}, z_3^{(\tau)}, \ldots, z_m^{(\tau)})$
  Sample $z_2^{(\tau+1)} \sim p(z_2 | z_1^{(\tau+1)}, z_3^{(\tau)}, \ldots, z_m^{(\tau)})$
  $\vdots$
  Sample $z_j^{(\tau+1)} \sim p(z_j | z_1^{(\tau+1)}, z_2^{(\tau+1)}, \ldots, z_{j-1}^{(\tau+1)}, z_{j+1}^{(\tau)}, \ldots, z_m^{(\tau)})$
  $\vdots$
  Sample $z_m^{(\tau+1)} \sim p(z_m | z_1^{(\tau+1)}, z_2^{(\tau+1)}, \ldots, z_{m-1}^{(\tau+1)})$
end for
return z

- The marginal distribution of any given subset of variables can be approximated by considering only the samples for that subset of variables; ignoring the rest.

- The expected value of any given variable can be approximated by averaging over all the samples for that variable.

As with other MCMC algorithms, Gibbs sampling generates a Markov chain of samples, each of which have some small amount of correlation with its nearby samples. Consequently, care must be taken when observing the posterior distribution. MCMC is started from an arbitrary initial state, yet only when the chain has 'forgotten' where it started, can the resulting distribution reflect the stationary distribution of the chain. Therefore, samples from the beginning of the Markov chain that do not reflect the stationary distribution are instead 'thrown away' and discarded. This process is known as *burn-in* and is applied before analysing the samples [173].

Gibbs sampling can suffer from poor rates of convergence when there is correlation in the posterior. If the joint distribution over the parameters is highly correlated, the algorithm may sometimes struggle to move away from the current state or will do so very slowly, requiring a large number of samples to eventually explore all possible states. There are techniques available to reduce these effects, such as *blocking*, although this is seen as an inherent limitation of Gibbs sampling methods [173]. Alternatively, there are other forms of inference that avoid sampling entirely, and assume a posterior form to make inference computationally more efficient — these include variational methods.

3.3.5 Variational inference

Variational inference (VI) is a powerful form of Bayesian inference and common alternative to MCMC, used to approximate intractable posterior distributions with analytical approximations (see e.g. [27, 173]). The general premise of VI is to approximate the true (intractable) posterior with a simpler, tractable surrogate distribution, chosen from a tractable family of distributions. This task is usually achieved by minimising the *Kullback-Liebler* (KL) *divergence*, a measure of how one probability distribution diverges from another, from the surrogate distribution $q(\theta)$ to the true distribution[5] $p(\theta|\mathcal{D})$. The KL divergence is defined as:

$$\mathbb{KL}(q||p) = \int q(\theta) \log \frac{q(\theta)}{p(\theta|\mathcal{D})} \mathrm{d}\theta \tag{3.19}$$

Minimising this divergence effectively finds the distribution, within the chosen family, that is closest to the target posterior. Direct computation of the KL divergence is challenging because the true posterior $p(\theta|\mathcal{D})$ is intractable. However, VI circumvents this by maximising something known as the *evidence lower bound* (ELBO), an alternative objective function that indirectly minimises the KL divergence [173]. The ELBO can be defined as follows.

The marginal likelihood (or evidence) of the data can be written as:

$$\log p(\mathcal{D}) = \log \int p(\mathcal{D}, \theta) \mathrm{d}\theta \tag{3.20}$$

Using Jensen's inequality (see e.g. [173]), the log of marginal likelihood is lower-bounded by the ELBO:

$$\begin{aligned} \log p(\mathcal{D}) &\geq \mathbb{E}_{q(\theta)} \left[\log p(\mathcal{D}|\theta)\right] - \mathbb{KL}(q(\theta)||p(\theta)) & (3.21) \\ &\geq \mathbb{E}_{q(\theta)} \left[\log p(\mathcal{D}, \theta)\right] - \mathbb{E}_{q(\theta)} \left[\log q(\theta)\right] = \text{ELBO} & (3.22) \end{aligned}$$

Maximising this ELBO can then be achieved by adjusting the parameters of the

[5]The KL divergence is a type of statistical measure of how different one probability distribution, p, differs from another distribution, q. It is important to note that $\mathbb{KL}(p||q) \neq \mathbb{KL}(q||p)$ [173]

approximating distribution $q(\theta)$, to balance the two terms in Equation 3.22:

1. The *expected log likelihood* of the data under the approximate posterior $\mathbb{E}_{q(\theta)}[\log p(\mathcal{D}, \theta)]$, thus encouraging $q(\theta)$ to assign a high probability to values of θ that explain the observed data $\mathcal{D}$.

2. The *entropy* $\mathbb{E}_{q(\theta)}[\log q(\theta)]$ of the surrogate distribution. This term prevents $q(\theta)$ from collapsing too narrowly around a specific value and helps avoid overfitting.

Assuming the surrogate posterior is determined by some set of free parameters ϕ, then the problem of VI reduces to a more familiar optimisation problem which, for certain cases, can be solved using coordinate-ascent and conducted efficiently in a manner similar to the EM algorithm [173]. Compared to MCMC, VI is often more computationally efficient. However, these improvements in efficiency come with a sacrifice: the family of surrogate distributions $q(\theta)$ are often restricted to simpler and more convenient forms (e.g. Gaussian distributions), which might not fully capture the complexity or multimodality of the true posterior. Nevertheless, this simpler form of the posterior makes inference a simpler task.

One of the most popular forms of VI is the *mean-field approximation* [184]. In this approach, the approximate posterior is assumed to be fully factorised over its components, assuming independence between the variables, in the form:

$$q(\theta) = \prod_{j=1}^{J} q_j(\theta_j) \tag{3.23}$$

While this assumption simplifies optimisation significantly, it can sometimes limit the quality of the approximation for specific models where strong correlations between parameters exist and are of interest.

The general implementation of *coordinate-ascent* VI (CAVI) is described in Algorithm 2.

Algorithm 2 Coordinate Ascent Variational Inference (CAVI)

Input: A model $p(\mathcal{D}, \mathbf{z})$ with data $\mathcal{D}$ and latent variables $\mathbf{z}$
Output: A variational density $q(\mathbf{z}) = \prod_{j=1}^{m} q_j(z_j)$
Initialise: Variational factors $q_j(z_j)$
while the evidence lower bound (ELBO) has not converged **do**
  for $j \in \{1, \ldots, m\}$ **do**
    Set $q_j(z_j) \propto \exp\{\mathbb{E}_j \left[\log P(z_j | \mathbf{z}_j, \mathbf{x})\right]\}$
  end for
  Compute $\text{ELBO}(q) = \mathbb{E}\left[\ln p(\mathcal{D}, \mathbf{z})\right] + \mathbb{E}\left[\ln q(\mathbf{z})\right]$
end while
return

When latent variables and the parameters of a model are desired, as will be the case later in Chapter 6, then this form of VI is known as *variational Bayes* (VB). The inclusion of a set of unknown parameters into Algorithm 2 is considered trivial for the VB case.

3.4 Canonical correlation analysis

Canonical correlation analysis (CCA) is a well-established methodology in multivariate statistical analysis; it was first introduced to the linear algebra community in 1875 by Jordan [185] and later to the statistical community by Hotelling in 1936 [186]. The task of CCA is to analyse the mutual dependency between two multivariate datasets $\mathbf{x}$ and $\mathbf{y}$, which can be evaluated by finding an appropriate set of orthogonal basis vectors, $\mathbf{a}$ and $\mathbf{b}$, such that the projected variables $\mathbf{a}^\mathsf{T}\mathbf{x}$ and $\mathbf{b}^\mathsf{T}\mathbf{y}$, are maximally correlated [173]. This correlation, ρ, can be expressed as

$$\rho = \frac{\text{cov}\{\mathbf{a}^\mathsf{T}\mathbf{x}, \mathbf{b}^\mathsf{T}\mathbf{y}\}}{\sqrt{\text{cov}\{\mathbf{a}^\mathsf{T}\mathbf{x}\}}\sqrt{\text{cov}\{\mathbf{b}^\mathsf{T}\mathbf{y}\}}} = \frac{\mathbf{a}^\mathsf{T}\boldsymbol{\Sigma}_{xy}\mathbf{b}}{\sqrt{(\mathbf{a}^\mathsf{T}\boldsymbol{\Sigma}_{xx}\mathbf{a})(\mathbf{b}^\mathsf{T}\boldsymbol{\Sigma}_{yy}\mathbf{b})}} \tag{3.24}$$

One then seeks several pairs of basis vectors that meet the above condition, subject to the constraint that the pairs of transformed variables are themselves independent and uncorrelated from one another. Given two co-occurring zero mean multidimensional

random variables, $\boldsymbol{x} \in \mathbb{R}^{d_1}$, $\boldsymbol{y} \in \mathbb{R}^{d_2}$, with sample covariance $\tilde{\boldsymbol{\Sigma}} = \begin{bmatrix} \tilde{\boldsymbol{\Sigma}}_{xx} & \tilde{\boldsymbol{\Sigma}}_{xy} \\ \tilde{\boldsymbol{\Sigma}}_{yx} & \tilde{\boldsymbol{\Sigma}}_{yy} \end{bmatrix}$, then this task can be achieved through the following maximisation:

$$(\mathbf{a}', \mathbf{b}') = \underset{\mathbf{a},\mathbf{b}}{\arg\max}\ \operatorname{corr}(\mathbf{a}^\mathsf{T}\boldsymbol{x}, \mathbf{b}^\mathsf{T}\boldsymbol{y}) = \underset{\mathbf{a},\mathbf{b}}{\arg\max}\ \frac{\mathbf{a}^\mathsf{T}\tilde{\boldsymbol{\Sigma}}_{xy}\mathbf{b}}{\sqrt{\mathbf{a}^\mathsf{T}\tilde{\boldsymbol{\Sigma}}_{xx}\mathbf{a}\mathbf{b}^\mathsf{T}\tilde{\boldsymbol{\Sigma}}_{yy}\mathbf{b}^\mathsf{T}}} \tag{3.25}$$

where $\tilde{\boldsymbol{\Sigma}}_{xy}$ is the cross covariance between $\boldsymbol{x}$ and $\boldsymbol{y}$, with $\tilde{\boldsymbol{\Sigma}}_{xy} = \tilde{\boldsymbol{\Sigma}}_{yx}^\mathsf{T}$, and $\tilde{\boldsymbol{\Sigma}}_{xx}$ and $\tilde{\boldsymbol{\Sigma}}_{yy}$ are auto covariances. One intuitive explanation of CCA is that it is equivalent to performing principal component analysis (PCA) on two datasets whilst simultaneously maximising the correlation between the two sets of principal components. This maximisation can be computed by solving the following generalised eigenvalue problem:

$$\begin{pmatrix} 0 & \tilde{\boldsymbol{\Sigma}}_{xy} \\ \tilde{\boldsymbol{\Sigma}}_{yx} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} = \lambda \begin{pmatrix} \tilde{\boldsymbol{\Sigma}}_{xx} & 0 \\ 0 & \tilde{\boldsymbol{\Sigma}}_{yy} \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} \tag{3.26}$$

where $\mathbf{a}$ and $\mathbf{b}$ are the eigenvectors or *canonical directions*, and λ is the eigenvalue or *canonical correlation*. In practice, pairs of eigenvalue and eigenvectors are computed from the following singular value decomposition (SVD),

$$\tilde{\boldsymbol{\Sigma}}_{xx}^{-\frac{1}{2}}\tilde{\boldsymbol{\Sigma}}_{xy}\tilde{\boldsymbol{\Sigma}}_{yy}^{-\frac{\mathsf{T}}{2}} = \mathbf{V}_1\boldsymbol{\Lambda}\mathbf{V}_2^\mathsf{T}. \tag{3.27}$$

where $\mathbf{V}_1$ and $\mathbf{V}_2$ are the left and right singular vectors respectively and $\boldsymbol{\Lambda}$ is the diagonal matrix of singular values (the canonical correlations).

3.5 Dynamics of systems

Before exploring new UQ methods for OMA, it is useful to briefly revise the way in which mechanical systems are described and the foundations of linear modal analysis. This theory is well-covered in many undergraduate textbooks (e.g. [5, 31, 32]) but reproduced here in the interest of the reader.

3.5.1 Multiple degrees-of-freedom

A real dynamic system like a wing or bridge, is *spatially distributed*; its motion can be represented by a set of coupled partial differential equations (PDEs). However, the solutions to these equations are difficult if not impossible to obtain. Therefore, it is often assumed that a system can be spatially discretised; the system can instead be truncated into a finite number of lumped masses (with mass m_i) connected by sets of massless springs (with stiffness k_i) and dampers (with damping coefficient c_i) in some suitable configuration. An example of an n_d-degree of freedom (DOF) linear, viscously damped, forced dynamic system is shown in Figure 3.5.

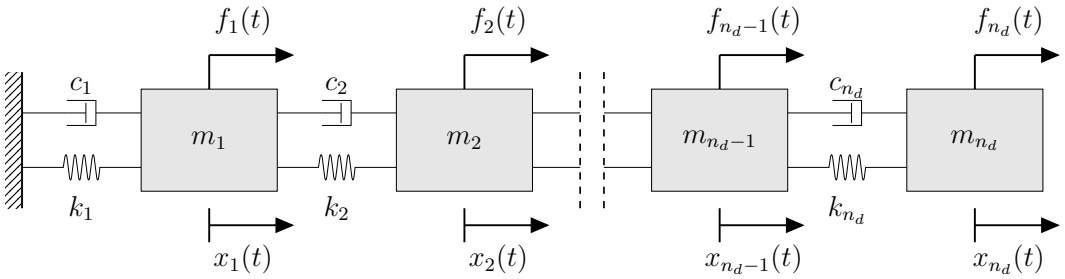


Figure 3.5: Generic forced n_d–DOF linear dynamic system, represented by a lumped set of masses each connected by massless springs and dampers.

Once spatially discretised, this n_d-DOF linear structural dynamic system can, without loss in generality, be entirely represented by a set of coupled heterogeneous linear ODEs of the form

$$\mathbf{M}\ddot{\mathbf{x}} + \mathbf{C}\dot{\mathbf{x}} + \mathbf{K}\mathbf{x} = \mathbf{f} \tag{3.28}$$

where $\mathbf{M}, \mathbf{C}, \mathbf{K}$ are the $n_d \times n_d$ mass, damping and stiffness matrices, respectively, $\mathbf{x}$

is a vector of the system displacements, and $\mathbf{f}$ is a vector of the system excitations.

To solve this set of coupled ODEs, without requiring a complete set of initial states, a convenient way is to decouple them. This can be achieved through the principal of superposition and the *modal decomposition* of linear dynamic systems.

3.5.2 Modal decomposition

The modal decomposition or *modal expansion theorem* of structural dynamic systems provides dynamicists with a convenient way of decoupling the equations of motion (Equation 3.28) to n_d independent (virtual) single DOF systems – these are called the modes. The theorem proposes that, for any given motion, the total response of a dynamic system, $\mathbf{x}(t)$, can be represented as a superposition of the individual responses of the modes of the system. In the general case of an n_d-DOF system, there will be n_d linearly independent modes such that,

$$\mathbf{x}(t) = \boldsymbol{\phi}^{(1)} q_1(t) + \boldsymbol{\phi}^{(2)} q_2(t) + \boldsymbol{\phi}^{(3)} q_3(t) + \cdots + \boldsymbol{\phi}^{(n_d)} q_{n_d}(t) \tag{3.29}$$

or more compactly,

$$\mathbf{x}(t) = \begin{bmatrix} \boldsymbol{\phi}^{(1)} & \dots & \boldsymbol{\phi}^{(n_d)} \end{bmatrix} \begin{Bmatrix} q_1(t) \\ \vdots \\ q_{n_d}(t) \end{Bmatrix} \tag{3.30}$$

where $\boldsymbol{\phi}^{(i)}$ is the i^{th} modal vector or *mode shape*, and $q_i(t)$ is the i^{th} natural coordinate or *modal coordinate*.

To derive these properties, one initially considers the undamped, homogenous case of Equation 3.28 ($\mathbf{C} = \mathbf{0}$, $\mathbf{f} = 0$),

$$\mathbf{M}\ddot{\mathbf{x}} + \mathbf{K}\mathbf{x} = 0 \tag{3.31}$$

The equation above is now a homogenous linear ODE with the nontrivial solution,

$$\mathbf{x} = \boldsymbol{\phi} \cos(\omega t) \tag{3.32}$$

where $\boldsymbol{\phi}$ is a mode shape vector. Substituting into Equation 3.31 gives,

$$-\omega^2 \mathbf{M}\boldsymbol{\phi}\cos(\omega t) + \mathbf{K}\boldsymbol{\phi}\cos(\omega t) = 0 \tag{3.33}$$

which when rearranged and simplified becomes,

$$\mathbf{M}^{-1}\mathbf{K}\boldsymbol{\phi} = \omega^2\boldsymbol{\phi} \tag{3.34}$$

which is now recognisable as a standard eigenvalue problem[6]. The non-zero solution to this problem results in n_d pairs of eigenvalues and eigenvectors (i.e. the modes), where the i^{th} mode is characterised by a natural frequency ω_i and mode shape $\boldsymbol{\phi}^{(i)}$. A key feature of this eigen-decomposition is that the eigenvectors have the useful property of being *weighted orthogonal*. It is this property that will allow the equations of motion to be decoupled, by performing a change of basis in the space of $\mathbb{R}^n$.

Let the vector of generalised coordinates, $\mathbf{x}$, be considered as coming from some modal space $\mathbf{x} = \boldsymbol{\Phi}\mathbf{q}$, where $\boldsymbol{\Phi}$ is the matrix of mode shape vectors and $\mathbf{q}$ is the vector of modal coordinates. Substituting this coordinate system into the (un-damped) heterogeneous equation of motion gives,

$$\mathbf{M}\boldsymbol{\Phi}\ddot{\mathbf{q}} + \mathbf{K}\boldsymbol{\Phi}\mathbf{q} = \mathbf{f} \tag{3.35}$$

After pre-multiplying by $\boldsymbol{\Phi}^{\mathsf{T}}$, one obtains

$$\boldsymbol{\Phi}^{\mathsf{T}}\mathbf{M}\boldsymbol{\Phi}\ddot{\mathbf{q}} + \boldsymbol{\Phi}^{\mathsf{T}}\mathbf{K}\boldsymbol{\Phi}\mathbf{q} = \boldsymbol{\Phi}^{\mathsf{T}}\mathbf{f} \tag{3.36}$$

where $\boldsymbol{\Phi}^{\mathsf{T}}\mathbf{f}$ is the modal force. Given the weighted orthogonality property of the mode shapes,

$$\boldsymbol{\phi}^{(s)\mathsf{T}}\mathbf{M}\boldsymbol{\phi}^{(r)} = \begin{cases} 0, & s \neq r \\ \mathbf{M}_{rr} & s = r \end{cases} \tag{3.37}$$

[6]Note that this is true for well-defined structural systems where the mass matrix can be assumed to be invertible, otherwise this would result in a generalised eigenvalue problem of the form $\mathbf{K}\boldsymbol{\phi} = \omega^2\mathbf{M}\boldsymbol{\phi}$

$$\boldsymbol{\phi}^{(s)\mathsf{T}}\mathbf{K}\boldsymbol{\phi}^{(r)} = \begin{cases} 0, & s \neq r \\ \mathbf{K}_{rr} & s = r \end{cases} \tag{3.38}$$

It becomes apparent that pre- and post- multiplication of the mode shape matrix diagonalises the parameter matrices such that

$$\boldsymbol{\Phi}^{\mathsf{T}}\mathbf{M}\boldsymbol{\Phi} = \begin{bmatrix} \ddots & \\ & \mathbf{M}_i \\ & & \ddots \end{bmatrix} \quad , \quad \boldsymbol{\Phi}^{\mathsf{T}}\mathbf{K}\boldsymbol{\Phi} = \begin{bmatrix} \ddots & \\ & \mathbf{K}_i \\ & & \ddots \end{bmatrix} \tag{3.39}$$

where $\begin{bmatrix} \ddots & \\ & \mathbf{M}_i \\ & & \ddots \end{bmatrix}$ denotes a diagonal matrix.

In the case of proportional (Rayleigh) damping, the damping matrix $\mathbf{C}$ can be expressed as a linear combination of the mass and stiffness matrices $\mathbf{C} = \alpha\mathbf{M} + \beta\mathbf{K}$ and thus, diagonalised in the same way. The general equation of motion for a forced, damped MDOF system can now be expressed as,

$$\begin{bmatrix} \ddots & \\ & \mathbf{M}_i \\ & & \ddots \end{bmatrix} \ddot{\mathbf{q}} + \begin{bmatrix} \ddots & \\ & \mathbf{C}_i \\ & & \ddots \end{bmatrix} \dot{\mathbf{q}} + \begin{bmatrix} \ddots & \\ & \mathbf{K}_i \\ & & \ddots \end{bmatrix} \mathbf{q} = \boldsymbol{\Phi}^{\mathsf{T}}\mathbf{f} \tag{3.40}$$

The equations of motion are now, through a change of basis, fully decoupled into n_d independent single DOF systems.

Moreover, when Equation 3.40 is pre-multiplied by the inverse of the diagonalised mass matrix, the system can also be represented in terms of its *modal properties*: The natural frequencies ω_i, damping ratios ζ_i, and mass normalised mode shapes $\tilde{\boldsymbol{\Phi}}$,

$$\mathbb{I}\,\ddot{\mathbf{q}} + \begin{bmatrix} \ddots & \\ & 2\omega_i\zeta_i \\ & & \ddots \end{bmatrix} \dot{\mathbf{q}} + \begin{bmatrix} \ddots & \\ & \omega_i^2 \\ & & \ddots \end{bmatrix} \mathbf{q} = \tilde{\boldsymbol{\Phi}}^{\mathsf{T}}\mathbf{f} \tag{3.41}$$

where $\mathbb{I}$ is the identity matrix.

This form is perhaps more useful to the engineer as the modal properties afford significant physical intuition. The natural frequencies describe the frequency locations at which a system will experience resonant behaviour; the damping ratios provide insight into how quickly a system will dissipate energy close to resonance; and the mode shapes describe the synchronous motion of the entire system when excited

at each natural frequency. This definition forms the foundation of all linear modal analysis algorithms [31]. If one intends to identify a system from observed data, as is the case in OMA, modal decomposition is assumed to hold true, providing a convenient and interpretable representation of the dynamics that can be more easily recovered. This representation is, of course, only an approximation of a system's true behaviour. Nevertheless, this is generally a very good assumption when considering close to linear systems.

3.6 State space representation

Sets of coupled ODEs can also be conveniently written in an alternative form known as the *state-space representation*; whereby a dynamic system is specified using a set of input, output, and internal variables; related by first-order differential equations. The internal variables are known as *state variables* and evolve in time. A system's current states describe its internal condition at any given time and can be used along with a system's inputs to predict future behaviour, without needing a past history explicitly. States often represent real-world quantities like temperature, price, charge etc. In the case of dynamic systems, the states are the displacement and velocity. The state-space representation of a system provides a convenient and compact way to model and analyse linear systems with multiple inputs and/or outputs; a useful tool when considering large structural dynamic systems with such requirements.

The reader is directed towards the following text books for a more comprehensive introduction to state space methods if required [103, 187]. The following subsection briefly introduces the continuous and discrete forms of the output-only state space model given its relevance and repeated reference throughout this thesis.

3.6.1 Continuous, time-invariant, output-only state space

The continuous state-space representation of a stochastic linear dynamic system of model order r, with l outputs and d state variables takes the following form,

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + w(t) \tag{3.42}$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + v(t) \tag{3.43}$$

where, $\mathbf{x}(t)$ is the state vector $\mathbf{x}(t) \in \mathbb{R}^d$;
$\dot{\mathbf{x}}(t)$ is the derivative of the state vector $\mathbf{x}(t) \in \mathbb{R}^d$;
$\mathbf{y}(t)$ is the output vector $\mathbf{y}(t) \in \mathbb{R}^l$;
$w(t)$ is the process noise, $w(t) \in \mathbb{R}^d$;
$v(t)$ is the measurement noise, $v(t) \in \mathbb{R}^l$;
$\mathbf{A}$ is the continuous state matrix $\mathbf{A} \in \mathbb{R}^{d \times d}$;
$\mathbf{C}$ is the output matrix, $\mathbf{C} \in \mathbb{R}^{l \times d}$;

with noise covariance

$$\mathbb{E}\left[\begin{bmatrix} w(t) \\ v(t) \end{bmatrix}\begin{bmatrix} w^{\mathsf{T}}(t) & v^{\mathsf{T}}(t) \end{bmatrix}\right] = \begin{bmatrix} \mathbf{Q} & \mathbf{S} \\ \mathbf{S}^{\mathsf{T}} & \mathbf{R} \end{bmatrix}\delta(t-s) \tag{3.44}$$

where $\mathbb{E}[\cdot]$ denotes the expectation, $\mathbf{Q}, \mathbf{R}, \mathbf{S}$ represent spectral densities of the noise, and $\delta(t-s)$ is the delta function for any two points in continuous time.

For a multi-degree of freedom structural dynamic system where displacement is observed, the state vector, state matrix and output matrix take the following form,

$$\mathbf{x} = \begin{bmatrix} \dot{\underline{\mathrm{x}}} \\ \underline{\mathrm{x}} \end{bmatrix} \tag{3.45}$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbb{I} \\ -\mathbf{M}^{-1}\mathbf{K} & -\mathbf{M}^{-1}\mathbf{C} \end{bmatrix} \tag{3.46}$$

$$\mathbf{C} = \begin{bmatrix} 0 & 1 \end{bmatrix} \tag{3.47}$$

where $\mathbf{0}$ is a zero matrix, $\mathbb{I}$ is the identity matrix, and $\mathbf{M}, \mathbf{C}, \mathbf{K}$ are matrices of the system parameters as in Equation (3.28). Here $\underline{\mathrm{x}}$ is a vector of displacements and $\dot{\underline{\mathrm{x}}}$ a vector of velocities, distinguishing them from the state vector[7].

Although useful for theoretical or mathematical applications, in practical engineering applications it is common to work with the discrete form of the state space, as observations are obtained in discrete time.

[7]It is perhaps evident that notational conventions here conflict with the notation used in Section 3.5.1. Here the variable $\mathbf{x}$ represents the set of internal states and not the displacement as in 3.5.1. However, the displacement and its derivatives are the internal states of a dynamic system. Hence, the use of $\underline{\mathrm{x}}$ and $\dot{\underline{\mathrm{x}}}$ to make this distinction.

3.6.2 Discrete, time-invariant, output-only state space

Alternatively, the model in (3.42)-(3.43) can be written in discrete time, such that the system evolves in steps with sampling time Δt, such that

$$\mathbf{x}_{k+1} = \mathbf{A}_d \mathbf{x}_k + w_k \tag{3.48}$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + v_k \tag{3.49}$$

where at a given point in a sequence $k\Delta t$ the parameters of the model are defined as follows: $\mathbf{x}_k$ is the state vector, $\mathbf{x}_k \in \mathbb{R}^d$;
$\mathbf{y}_k$ is the output vector, $\mathbf{y}_k \in \mathbb{R}^l$;
$\mathbf{A}_d$ is the discrete state matrix, $\mathbf{A}_d \in \mathbb{R}^{d\times d}$;
$\mathbf{C}$ is the output matrix, $\mathbf{C} \in \mathbb{R}^{l\times d}$;
w_k is the process noise, $w_k \in \mathbb{R}^d$;
v_k is the measurement noise, $v_k \in \mathbb{R}^l$;

with a noise covariance of the form,

$$\mathbb{E}\left[\begin{bmatrix} w_q \\ v_q \end{bmatrix}\begin{bmatrix} w_s^\mathsf{T} & v_s^\mathsf{T} \end{bmatrix}\right] = \begin{bmatrix} \mathbf{Q} & \mathbf{S} \\ \mathbf{S}^\mathsf{T} & \mathbf{R} \end{bmatrix}\delta_{qs} \tag{3.50}$$

where $\mathbb{E}[\cdot]$ denotes the expectation and δ_{qs} is the Kronecker delta for any two samples in time q and s.

The discrete time state matrix $\mathbf{A}_d$ is related to the continuous state matrix by the following expression,

$$\mathbf{A}_d = \text{expm}(\mathbf{A}\Delta t) \tag{3.51}$$

where expm is the matrix exponential and Δt is the sampling time.

3.6.3 Observability

The observability of a system is a measure of how well the internal states of that system can be inferred given knowledge of the external outputs and is obtained using discrete observations from a system [95]. A system is defined as observable i.f.f. the extended observability matrix, for a state dimension of d,

$$\mathcal{O} = \begin{bmatrix} \mathbf{C} \\ \mathbf{C}\mathbf{A}_d \\ \mathbf{C}\mathbf{A}_d^2 \\ \vdots \\ \mathbf{C}\mathbf{A}_d^{d-1} \end{bmatrix} \tag{3.52}$$

has a rank of d. Given the availability of the observability matrix, the discrete state matrix $\mathbf{A}_d$ and output matrix $\mathbf{C}$ can be recovered (see e.g. [95] or [103]).

3.6.4 Recovery of the modal properties from the state matrix

The (continuous) state matrix $\mathbf{A}$ encapsulates the core dynamics of a system and details how the current states (displacement and velocity) will influence the evolution of the states. Using this matrix, the modal parameters of the system can be recovered. If in possession of the discrete time state matrix $\mathbf{A}_d$, the continuous state matrix is related by the expression in 3.51.

From the state matrix, the modal properties of the dynamic system can be obtained by solving the following eigenvalue decomposition,

$$\mathbf{A}\boldsymbol{\psi} = \lambda\boldsymbol{\psi} \tag{3.53}$$

to recover sets of eigenvalues λ and corresponding eigenvectors $\boldsymbol{\psi}$.

Given the discrete nature of the state space model, the respective eigenvalues correspond to a set of discrete time system poles; however, the continuous system poles can be recovered easily through,

$$\boldsymbol{\mu} = \frac{\log(\Lambda)}{\Delta t} \tag{3.54}$$

where Λ is a diagonal matrix containing the set of complex eigenvalues $\{\lambda_i\}_{i=1}^{d}$. Once recovered, the poles can be used to calculate the natural frequencies in Hz,

$$\omega_i = \frac{|\boldsymbol{\mu}_i|}{2\pi} \tag{3.55}$$

where Δt is the sample time, and the damping ratios using,

$$\zeta_i = \frac{-\mathfrak{Re}\{\boldsymbol{\mu}_i\}}{|\boldsymbol{\mu}_i|} \tag{3.56}$$

where $\mathfrak{Re}$ the real part of the complex variable.

Finally, the mode shape matrix $\boldsymbol{\Phi}$ can also be recovered using the recovered output matrix $\mathbf{C}$ and the eigenvectors $\boldsymbol{\Psi}$, such that

$$\boldsymbol{\Phi} = \mathbf{C}\boldsymbol{\Psi} \tag{3.57}$$

In the context of OMA, it is important to note that a state matrix recovered from data does not necessarily take the structure shown in Equation (3.46). However, the eigenmodes of the recovered state matrix are known to relate to the modes of the system in the same way. This is also true of the eigenvalues and the other modal properties. The matrix is said to be ‘similar‘ in a linear algebra sense.

3.7 Stochastic subspace methods

Stochastic subspace methods of system identification have become common place in many control and mechanical settings, as described in Chapter 2. Given the context of this thesis for operational identification from discrete output measurements, only the theory for a stochastic output-only case is considered here.

3.7.1 Output-only stochastic realisation problem

The general stochastic realisation or identification problem, as defined in [95], proceeds as follows. Given N measurements of an output $\mathbf{y}_k \in \mathbb{R}^l$, generated by the unknown stochastic system described by the r^{th} order state space model from Equations (3.48) - (3.49),

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}_d \mathbf{x}_k + w_k \\ \mathbf{y}_k &= \mathbf{C}\mathbf{x}_k + v_k \end{aligned}$$

one intends to determine:

- The order r of the unknown system
- The system matrices $\mathbf{A}_d$ and $\mathbf{C}$

3.7.2 Stochastic subspace identification

The most familiar subspace method for solving this problem in structural dynamics is the *stochastic subspace identification* (SSI) algorithm [95]. As discussed in Section 2.3.2, SSI analyses the projection of shifted output sequences onto itself, assuming white noise excitation and stationarity in the data, but not necessarily the noise. Although stationary noise is assumed when performing uncertainty quantification. This projection, when viewed as the conditional mean of the "future" observations given the "past", can be described in terms of covariances and solved for the parameters of the state space using the singular value decomposition (SVD). This methodology can be used to identify both input-output and output-only systems. In the output-only case, two common methods are available: covariance-driven SSI (SSI-Cov) and data-driven SSI (SSI-Data). This thesis will focus on SSI-Cov, specifically the CVA weighted form. The general derivation of CVA weighted SSI-Cov is given later in this section, however, first some definitions are required.

3.7.3 Definitions

Toeplitz matrices

A *Toeplitz matrix* is a special type of matrix whose entries are constant along each descending diagonal [188]. A matrix $T \in \mathcal{R}^{n \times n}$ is Toeplitz if $T_{(i,j)} = T_{(k,l)} \; \forall \; k-i = l-j$ such that,

$$T = \begin{bmatrix} \tau_1 & \tau_2 & \tau_3 & \ddots \\ \tau_4 & \tau_1 & \tau_2 & \ddots \\ \tau_5 & \tau_4 & \tau_1 & \ddots \\ \ddots & \ddots & \ddots & \ddots \end{bmatrix} \tag{3.58}$$

By extension, a block Toeplitz matrix is one whose block entries follow the structure above.

Hankel matrices

A *Hankel matrix* is a special type of matrix whose entries are constant along each ascending skew-diagonal from left to right. A matrix $H \in \mathcal{R}^{n \times n}$ is Hankel if $H_{(i,j)} = H_{(k,l)} \; \forall \; k + l = k + j$ such that,

$$H = \begin{bmatrix} h_1 & h_2 & h_3 & \iddots \\ h_2 & h_3 & h_4 & \iddots \\ h_3 & h_4 & h_5 & \iddots \\ \iddots & \iddots & \iddots & \iddots \end{bmatrix} \tag{3.59}$$

Similarly, by extension a block Hankel matrix is one whose block entries follow the structure shown above.

3.7.4 Covariance-driven stochastic subspace identification

This section provides a concise overview of the theory underpinning the canonical-variate weighted form of the SSI-Cov algorithm for an output-only case. This derivation follows the description given in Katayama [103]. The reader is directed towards the aforementioned text for a fuller derivation and for specifics.

Recalling the state space theory introduced in Section 3.6 and with a continuation of notation, consider an r^{th} order discrete state space model of a linear dynamic system, equivalent to a mechanical system with n_{d} degrees of freedom, such that $r = 2n_{\text{d}}$.

Output measurements from this process, based on l measurement channels, can be represented by the vector $\{\mathbf{y}_k, k = 0, 1, \ldots, N'\}$ where $N' = N + 2j - 2$ and arranged into a block Hankel matrix of the form,

$$\mathbf{Y}_{0|2j-1} = \left[\begin{array}{cccc} \mathbf{y}_0 & \mathbf{y}_1 & \cdots & \mathbf{y}_{N-1} \\ \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_N \\ \vdots & \vdots & & \vdots \\ \mathbf{y}_{j-1} & \mathbf{y}_j & \cdots & \mathbf{y}_{N+j-2} \\ \hline \mathbf{y}_j & \mathbf{y}_{j+1} & \cdots & \mathbf{y}_{N+j-1} \\ \mathbf{y}_{j+1} & \mathbf{y}_{j+2} & \cdots & \mathbf{y}_{N+j} \\ \vdots & \vdots & & \vdots \\ \mathbf{y}_{2j-1} & \mathbf{y}_{2j} & \cdots & \mathbf{y}_{N+2j-2} \end{array}\right] = \left[\begin{array}{c} \mathbf{Y}_{0|j-1} \\ \hline \mathbf{Y}_{j|2j-1} \end{array}\right] = \left[\begin{array}{c} \mathbf{Y}_p \\ \hline \mathbf{Y}_f \end{array}\right] \tag{3.60}$$

The Hankel matrix contains $2j$ block rows and N columns, with every block consisting of l rows and where $j > 0$ and N is sufficiently large (i.e much larger than $2lj$) and where $j > r$ (target model order) and the number of columns of block matrices is N. The resultant cross-covariance matrix of the future $\mathbf{Y}_f$ with the past $\mathbf{Y}_p$ is therefore given by

$$\tilde{\boldsymbol{\Sigma}} \;=\; \frac{1}{N}\begin{bmatrix}\mathbf{Y}_p \\ \mathbf{Y}_f\end{bmatrix}\begin{bmatrix}\mathbf{Y}_p^{\mathsf{T}} & \mathbf{Y}_f^{\mathsf{T}}\end{bmatrix} \;=\; \begin{bmatrix}\boldsymbol{\Sigma}_{pp} & \boldsymbol{\Sigma}_{pf} \\ \boldsymbol{\Sigma}_{fp} & \boldsymbol{\Sigma}_{ff}\end{bmatrix} \tag{3.61}$$

where $\boldsymbol{\Sigma}_{pf}$ and $\boldsymbol{\Sigma}_{fp}$ are finite block cross-covariance matrices, and $\boldsymbol{\Sigma}_{ff}$, $\boldsymbol{\Sigma}_{pp}$ are finite block auto-covariance matrices respectively. The canonical correlations $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \cdots, \lambda_r)$ between the future and past are the singular values [103], obtained

through the SVD of the following weighted matrix

$$\boldsymbol{\Sigma}_{ff}^{-1/2}\boldsymbol{\Sigma}_{fp}\boldsymbol{\Sigma}_{pp}^{-\mathsf{T}/2} \;=\; \mathbf{V}_1\boldsymbol{\Lambda}\mathbf{V}_2^{\mathsf{T}} \;\backsimeq\; \breve{\mathbf{V}}_1\breve{\boldsymbol{\Lambda}}\breve{\mathbf{V}}_2^{\mathsf{T}} \tag{3.62}$$

where $\boldsymbol{\Sigma}_{ff}^{1/2}\boldsymbol{\Sigma}_{ff}^{\mathsf{T}/2} = \boldsymbol{\Sigma}_{ff}$, such that,

$$\boldsymbol{\Sigma}_{fp} \;\backsimeq\; \boldsymbol{\Sigma}_{ff}^{1/2}\breve{\mathbf{V}}_1\breve{\boldsymbol{\Lambda}}\breve{\mathbf{V}}_2^{\mathsf{T}}\boldsymbol{\Sigma}_{pp}^{\mathsf{T}/2} \tag{3.63}$$

where $\mathbf{V_1}$ and $\mathbf{V}_2$ are the left and right singular vectors, respectively and $\breve{\boldsymbol{\Lambda}}$ neglects sufficiently small singular values (canonical correlations) in $\boldsymbol{\Lambda}$ such that the resultant state vector has the dimension $d = \dim(\breve{\boldsymbol{\Lambda}})$; the supposed model order of the system. The cross-covariance matrix, $\boldsymbol{\Sigma}_{fp}$, can be decomposed into the product of the extended observability ($\mathcal{O}$) and controllability ($\mathcal{C}$) matrices [95], using the relation $\boldsymbol{\Sigma}_{fp} = \mathcal{O}\mathcal{C}$, such that

$$\mathcal{O} \;=\; \boldsymbol{\Sigma}_{ff}^{1/2}\breve{\mathbf{V}}_1\breve{\boldsymbol{\Lambda}}^{1/2} \quad , \quad \mathcal{C} \;=\; \breve{\boldsymbol{\Lambda}}^{1/2}\breve{\mathbf{V}}_2^{\mathsf{T}}\boldsymbol{\Sigma}_{pp}^{\mathsf{T}/2} \tag{3.64}$$

with $\text{rank}(\mathcal{O}) = \text{rank}(\mathcal{C}) = d$. The extended observability matrix can then be used to recover the state $\mathbf{A}_d$ and output $\mathbf{C}$ matrices[8] through,

$$\mathbf{A}_d = \underline{\mathcal{O}}^{\dagger}\overline{\mathcal{O}}, \quad \mathbf{C} = \mathcal{O}(1:q,:) \tag{3.65}$$

where $\underline{\mathcal{O}}$ is the extended observability matrix minus the last block row, $\overline{\mathcal{O}}$ is the extended observability minus the first block row, $\mathcal{O}(1:q,:)$ is the first block row of the extended observability matrix and $\cdot^{\dagger}$ represents the pseudo inverse. Once in possession of $\mathbf{A}$ and $\mathbf{C}$, the modal properties can be duly recovered in the usual way for OMA (see Section 3.6 or [112] and [5]).

[8]The method described here for recovering the system matrices is one of many. See Stochastic Balanced Realisation Algorithm A, Chapter 8 in Katayama [103]

3.7.5 Stabilisation diagrams

In all modal analysis algorithms, including SSI, the true model order is seldom known a priori; one must instead choose an appropriate model order based on evidence. This becomes a difficult task as spurious (non-physical) modes typically appear in estimated models. Therefore, this problem is addressed by using an empirical multi-order process, where a system is identified at multiple (over-specified) model orders and a *stabilisation diagram*[9] is used to distinguish between the true (structural) modes and spurious modes [55]. Typically, the eigenvalues of the eigen decomposition in SSI (or the solutions to the denominator polynomial in frequency-domain rational curve-fitting tasks) are referred to as the *poles* of the system.

In a stabilisation diagram, complex poles that correspond to true structural modes are assumed to be present across increasing models orders — there is sufficient evidence at all model orders that the pole in question accurately describes the dynamics. Therefore, using some pre-defined heuristic criteria on the acceptable *stability* of these poles across multiple model orders, the true modes can be manually separated from the spurious ones. Surprisingly, this manual process is still the current industry standard. However, the automatic identification of these modes is an active research field [13, 16, 189], with a range of approaches being taken including clustering [190] and uncertainty-based methods [17]. Throughout this thesis, stabilisation diagrams are used to explore the effects of increasing model order on the identification procedure. Moreover, later in Chapter 5, modern clustering techniques are also used to automate the analysis of multiple diagrams with ease.

[9]The author believes this choice of terminology often leads to confusion with the stability of system poles referenced in control theory and that the term stabilisation diagram should be replaced with a suitable alternative. Nevertheless, the term stabilisation is used throughout this thesis as not to cause confusion for the familiar dynamicist. Some alternative naming has been suggested [189]; e.g. 'consistency' diagram. However, this also conflicts with other mathematical concepts.

4

Probabilistic Stochastic Subspace Identification

In Chapter 2, the reader was provided with some wider context to the problem of OMA, with a specific focus on OMA methods and in particular: stochastic subspace identification (SSI). Relevant literature surrounding existing approaches to frequentist and Bayesian UQ for OMA was also explored. After reviewing the literature, a gap in the research was identified; specifically the lack of a Bayesian formulation of the SSI algorithm capable of UQ. Consequently, this chapter presents the first step towards achieving this goal — namely the necessary mathematical framework that will later enable a Bayesian approach.

This chapter introduces a novel probabilistic formulation of the (CVA weighted) SSI-Cov algorithm; this form is made possible through the direct replacement of the identification mechanism — canonical correlation analysis (CCA) — with its probabilistic equivalent. It will be demonstrated how the maximum likelihood estimates from this reformulated algorithm are equivalent to that of the original SSI-Cov algorithm, principally leaving the outputs of the algorithm unchanged.

Although seemingly trivial, this now probabilistic representation will ultimately prove advantageous in the pursuit of other probabilistic and Bayesian SSI algorithms; as will be explored in the later chapters.

Earlier in this thesis, existing approaches to UQ for OMA were explored. In both frequentist and Bayesian approaches, current methods predominantly perform statistical or Bayesian analysis on the inputs or outputs of OMA algorithms to determine some form of uncertainty, rather than embedding uncertainty within the identification mechanism. In pursuit of new probabilistic and Bayesian approaches to OMA, it is useful to consider new types of OMA algorithms that embed probabilistic and/or Bayesian understanding in their definition to assist UQ.

One of the key challenges associated with developing probabilistic methodologies is in constructing suitable models that can achieve a desired task or operation; considering factors such as model complexity, model assumptions, or model selection [180]. These challenges make the development of new probabilistic or Bayesian OMA algorithms a non-trivial task. Nevertheless, one can avoid reinventing the proverbial 'OMA wheel', by first considering whether existing approaches can be reimagined probabilistically, given existing knowledge.

Of the many available OMA algorithms, the reader was previously introduced to the SSI-Cov algorithm in Section 3.7.4. The method's familiarity and frequent usage across industry and academia stands as a testament to its reliable identification performance and broad applicability. These benefits make it an attractive candidate when considering which OMA algorithms one may want to adapt to include UQ. However, the defining question is: Does SSI have the potential to incorporate probabilistic and/or Bayesian understanding? And if not, what could be done to enable this?

It is widely known that many stochastic realisation problems (including SSI-Cov) ultimately amount to performing CCA between two datasets that are (assumed to be) doubly infinite; infinite in both rows and columns [95, 103]. Following careful analysis, the problem can be reduced to a semi-infinite matrix problem; where only the number of columns need be very large, whilst the number of block rows sufficiently small. In the case of SSI, the datasets in question are the block Hankel matrices of the future and the past, constructed from the time history responses. This may sound familiar to the reader; recall the description of SSI using canonical correlations in Section 3.7.4.

In 2005, Bach and Jordan presented a probabilistic interpretation of CCA (PCCA) [191], employing the use of a latent variable model[1] to recast the problem of CCA as probabilistic. Given the methods equivalence to the classic formulation of CCA, one can begin to consider whether redefining SSI-Cov using PCCA, could provide the probabilistic OMA algorithm necessary to meet the aims of this work.

In this chapter, the reader will be introduced to a novel probabilistic formulation of the SSI-Cov algorithm, made possible through the direct replacement of CCA with its probabilistic equivalent. The theory of PCCA will be introduced and related to SSI-Cov, helping to define a probabilistic SSI algorithm.

4.1 Probabilistic canonical correlation analysis

At the core of PCCA lies a lower dimensional, unobserved latent space described by the variable $\mathbf{z}_n \in \mathbb{R}^d$ for $n = 1, ..., N$ observations[2]. In the context of state space models, this latent variable can be considered related to the internal states, in the same way that classical SSI can be used to project estimates of the states from the observed data [103]. This latent variable can be transformed through two independent linear mappings $\mathbf{W}^{(1)}$ and $\mathbf{W}^{(2)}$ — note the transformations defined by Bach and Jordan transform variables from the latent space to the data space — and offset with means $\boldsymbol{\mu}^{(1)}$ and $\boldsymbol{\mu}^{(2)}$, to produce a pair of observed variables $\mathbf{x}_n^{(1)} \in \mathbb{R}^{D_1}$ and $\mathbf{x}_n^{(2)} \in \mathbb{R}^{D_2}$. Crucially, as the observed variables are assumed to be generated from a shared latent space, the two datasets can be considered correlated. The full set of observed samples are given by matrices $\mathbf{X}^{(m)} = [\mathbf{x}_1^{(m)}, \dots, \mathbf{x}_N^{(m)}] \in \mathbb{R}^{D_m \times N}$ where $m = 1, 2$ and N is the total number of observations. To simplify notation, $\mathbf{x}_n = [\mathbf{x}_n^{(1)}; \mathbf{x}_n^{(2)}] \in \mathbb{R}^D$ and $\mathbf{X} = [\mathbf{X}^{(1)}; \mathbf{X}^{(2)}] \in \mathbb{R}^{D \times N}$ is the feature-wise row concatenation of the two random variables, where $D = D_1 + D_2$. PCCA is parameterised by the model parameters $\boldsymbol{\theta} = \{\mathbf{W}, \boldsymbol{\Sigma}, \boldsymbol{\mu}, \mathbf{Z}\}$ where the latent variables are denoted by $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_N]$, one for each pair of observed samples, and the remaining variables are the independent model parameters.

[1] This model is similar to that applied by Tipping and Bishop [181] in their earlier development of probabilistic principal component analysis.

[2] Note a distinction between the notation used here to represent the data $\mathbf{x}$ and latent variables $\mathbf{z}$, and that introduced in 3.6 for the output $\mathbf{y}$ and internal states $\mathbf{x}$. This is somewhat purposeful, to stay true to the machine learning and PGM notation, whilst also keeping the definition of the statistical method of probabilistic (and later Bayesian) CCA distinct. For added clarity, when the subscript k will refer to the state space, and subscript n to the data space.

Given this list of assumptions, Bach and Jordan defined the following probabilistic model, shown graphically in Figure 4.1 and given mathematically through Equations (4.1) - (4.3).

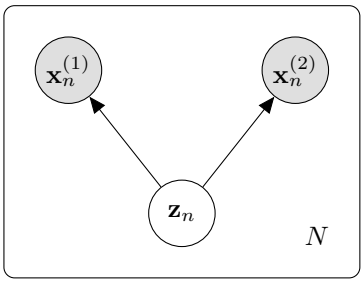


Figure 4.1: Graphical model for the probabilistic interpretation of CCA (PCCA)

$$\mathbf{z}_n \sim \mathcal{N}(0, \mathbb{I}) \tag{4.1}$$

$$\mathbf{x}_n^{(m)}|\mathbf{z}_n \sim \mathcal{N}(\mathbf{W}^{(m)}\mathbf{z}_n + \boldsymbol{\mu}^{(m)}, \boldsymbol{\Sigma}^{(m)}) \tag{4.2}$$

$$\mathbf{x}_n|\mathbf{z}_n \sim \mathcal{N}(\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{4.3}$$

where $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ corresponds to a Gaussian distribution characterised by mean and covariance, $\mathbf{W} = [\mathbf{W}^{(1)}; \mathbf{W}^{(2)}]$, $\boldsymbol{\mu} = [\boldsymbol{\mu}^{(1)}; \boldsymbol{\mu}^{(2)}]$ and $\boldsymbol{\Sigma}$ is a block-diagonal covariance matrix with $\boldsymbol{\Sigma}^{(1)}$ and $\boldsymbol{\Sigma}^{(2)}$ along the diagonal. An isotropic noise model is assumed in the latent space, which enforces independence between the variables whilst imposing a maximum correlation condition. Intuitively, when performing CCA, it is possible to imagine that one is searching for a set of latent variables, from which the observed datasets are linear transformations. Enforcing a shared latent space between two datasets with isotropic covariance imposes the maximum correlation condition from CCA. This condition is seen in the probabilistic model formulation by observing that each dataset is an affine transformation on the latent variables. For further discussion on the equivalence of CCA and this latent variable model the reader is directed to [191].

4.1.1 Maximum likelihood estimates

Using this model, the equivalence between probabilistic CCA and traditional CCA was demonstrated by proving that the MLE for the parameters in Equation (4.2) lead to the results of classical CCA and therefore, contain all the necessary information traditionally obtained through the SVD. The MLE (denoted by $\hat{\cdot}$) of the two weights

are given by,

$$\hat{\mathbf{W}}^{(1)} = \tilde{\mathbf{\Sigma}}_{11}^{1/2}\mathbf{V}_1\mathbf{M}_1 \tag{4.4}$$

$$\hat{\mathbf{W}}^{(2)} = \tilde{\mathbf{\Sigma}}_{22}^{1/2}\mathbf{V}_2\mathbf{M}_2 \tag{4.5}$$

where $\tilde{\mathbf{\Sigma}}_{\bullet}$ are components of the sample variance $\tilde{\mathbf{\Sigma}} = \begin{bmatrix} \tilde{\mathbf{\Sigma}}_{11} & \tilde{\mathbf{\Sigma}}_{12} \\ \tilde{\mathbf{\Sigma}}_{21} & \tilde{\mathbf{\Sigma}}_{22} \end{bmatrix}$, $\mathbf{V}_1$ and $\mathbf{V}_2$ are the left and right singular vectors of $\tilde{\mathbf{\Sigma}}_{11}^{-1/2}\tilde{\mathbf{\Sigma}}_{12}\tilde{\mathbf{\Sigma}}_{22}^{-\mathsf{T}/2}$ respectively, $\mathbf{M}_1\mathbf{M}_2^{\mathsf{T}} = \mathbf{P}_d$ where $\mathbf{P}_d$ is a diagonal matrix of the largest d canonical correlations. The MLE of the mean $\hat{\boldsymbol{\mu}}$ is simply,

$$\hat{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}} \tag{4.6}$$

where $\tilde{\boldsymbol{\mu}}$ is the sample mean. Prior to analysis, the data can be adjusted to reflect a zero mean dataset, simplifying the problem somewhat, although this is not necessary.

As noted by Bach and Jordan [191], amongst all possible solutions, the solution that maximises the conditional entropy of the data given the latent variables occurs when $\mathbf{M}_1 = \mathbf{M}_2 = \mathbf{M}$ is the square root of $\mathbf{P}_d$ i.e. $\mathbf{M} = \mathbf{P}_d^{1/2}\mathbf{R}$, where $\mathbf{R}$ is an arbitrary rotation matrix of size $d \times d$. Accounting for this rotation, Equation (4.5) becomes,

$$\hat{\mathbf{W}}^{(1)} = \tilde{\mathbf{\Sigma}}_{11}^{1/2}\mathbf{V}_1\mathbf{\Lambda}^{1/2}\mathbf{R} \tag{4.7}$$

$$\hat{\mathbf{W}}^{(2)} = \tilde{\mathbf{\Sigma}}_{22}^{1/2}\mathbf{V}_2\mathbf{\Lambda}^{1/2}\mathbf{R} \tag{4.8}$$

However, this arbitrary rotation is often ignored as its omission still results in data being transformed into the relevant subspace, i.e. $\mathbf{W}\mathbf{R}\mathbf{R}^{-1}\mathbf{Z} = \mathbf{W}\mathbf{Z}$, but it can be recovered through a simple post-processing step if required. The method for the recovery of $\mathbf{R}$ is described in the Appendix errata of [192] but is also repeated in Appendix A.4 for completeness.

4.1.2 Expectation-maximisation

A convenient property of latent variable models is that model parameter estimates can also be recovered iteratively using the EM algorithm; recall Section 3.3.1 on the EM algorithm. Despite fully tractable MLE solutions, the EM update equations for PCCA are also available [191] and are given by,

$$\mathbf{W}' = \tilde{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}\mathbf{W}\mathbf{M}(\mathbf{M} + \mathbf{M}\mathbf{W}^{\mathsf{T}}\mathbf{\Sigma}^{-1}\tilde{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}\mathbf{W}\mathbf{M})^{-1} \tag{4.9a}$$

$$\mathbf{\Sigma}' = \begin{bmatrix} (\tilde{\mathbf{\Sigma}} - \tilde{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}\mathbf{W}\mathbf{M}\mathbf{W}'^{\mathsf{T}})_{11} & 0 \\ 0 & (\tilde{\mathbf{\Sigma}} - \tilde{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}\mathbf{W}\mathbf{M}\mathbf{W}'^{\mathsf{T}})_{22} \end{bmatrix} \tag{4.9b}$$

where $\mathbf{M} = (\mathbb{I} + \mathbf{W}^{\mathsf{T}}\mathbf{\Sigma}^{-1}\mathbf{W})^{-1}$, $(\cdot)'$ denotes the update parameter and if not stated, the current parameter estimate is used.

Repeated iteration of these equations will result in the convergence of the EM algorithm to a local maxima; at a point very close to or at the MLE result. The availability of EM update equations for this model, facilitates the complete removal of the SVD from the SSI algorithm. This could have certain computational or analytical advantages. The most interesting of which would be an online recursive form of SSI using the theory of online EM [193]. Although not covered in this thesis, this will be discussed and pursued in future work (see Chapter 8.2).

4.2 Probabilistic SSI

Upon closer inspection of the MLE of the weights from PCCA, and the definition of the extended observability and extended controllability in the context of SSI first introduced in Section 3.7.4, one can begin to draw parallels between the two results; this is of course expected for the CVA weighted form of SSI-Cov.

Unsurprisingly, through substitution of CCA with PCCA in the SSI algorithm, it can be shown that the MLE estimates for the weights $\hat{\mathbf{W}}^{(1)}$ and $\hat{\mathbf{W}}^{(2)}$, given that the two datasets $\mathbf{X}^{(1)} = [\mathbf{x}_1^{(1)}, \ldots, \mathbf{x}_N^{(1)}]$ and $\mathbf{X}^{(2)} = [\mathbf{x}_1^{(2)}, \ldots, \mathbf{x}_N^{(2)}]$ are Hankel matrices of the future $\mathbf{Y}_f$ and past $\mathbf{Y}_p$ respectively, are equivalent to the extended observability matrix and extended controllability matrix transposed respectively, with some arbitrary rotation.

$$\hat{\mathbf{W}}^{(1)} = \boldsymbol{\Sigma}_{ff}^{1/2}\mathbf{V}_1\boldsymbol{\Lambda}^{1/2}\mathbf{R} = \mathcal{O}\mathbf{R} \tag{4.10}$$

$$\hat{\mathbf{W}}^{(2)} = \boldsymbol{\Sigma}_{pp}^{1/2}\mathbf{V}_2\boldsymbol{\Lambda}^{1/2}\mathbf{R} = \mathcal{C}^{\mathsf{T}}\mathbf{R} \tag{4.11}$$

The equivalence of these two methods, although perhaps trivial at first glance, is particularly profound. This now probabilistic SSI algorithm (Prob-SSI) possesses the potential to enable other hierarchical approaches, including a Bayesian SSI algorithm. Before exploring these other methods, however, it is important to address the rotational ambiguity present in the equations above.

Interestingly, in the context of SSI, the rotation does not need to be recovered when considering the maximum chosen model order because the method of recovering the state matrix from the observability matrix removes the rotational effects. Recall Equation 3.65:

$$\mathbf{A}_d = \underline{\mathcal{O}}^{\dagger}\overline{\mathcal{O}}$$

Rewriting the above to include the arbitrary rotation, and the pseudo inverse in full form, gives

$$(\underline{\mathcal{O}}\mathbf{R})^{\dagger}\overline{\mathcal{O}}\mathbf{R} = ((\underline{\mathcal{O}}\mathbf{R})^{\mathsf{T}}\underline{\mathcal{O}}\mathbf{R})^{-1}(\underline{\mathcal{O}}\mathbf{R})^{\mathsf{T}}\overline{\mathcal{O}}\mathbf{R} \tag{4.12}$$

Expanding terms,

$$(\underline{\mathcal{O}}\mathbf{R})^{\dagger}\overline{\mathcal{O}}\mathbf{R} = \mathbf{R}^{-1}(\underline{\mathcal{O}}^{\mathsf{T}}\underline{\mathcal{O}})^{-1}\mathbf{R}^{-\mathsf{T}}\mathbf{R}^{\mathsf{T}}\underline{\mathcal{O}}^{\mathsf{T}}\overline{\mathcal{O}}\mathbf{R} \tag{4.13}$$

Given that $\mathbf{R}^{-\mathsf{T}}\mathbf{R}^{\mathsf{T}} = \mathbb{I}$,

$$(\underline{\mathcal{O}}\mathbf{R})^{\dagger}\overline{\mathcal{O}}\mathbf{R} = \mathbf{R}^{-1}(\underline{\mathcal{O}}^{\mathsf{T}}\underline{\mathcal{O}})^{-1}\underline{\mathcal{O}}^{\mathsf{T}}\overline{\mathcal{O}}\mathbf{R} \tag{4.14}$$

$$= \mathbf{R}^{-1}\underline{\mathcal{O}}^{\dagger}\overline{\mathcal{O}}\mathbf{R} \tag{4.15}$$

$$= \mathbf{R}^{-1}\mathbf{A}_d\mathbf{R} \tag{4.16}$$

The transformation $\mathbf{R}^{-1}\mathbf{A}_d\mathbf{R}$ is as a *similarity* transform. Although this transformation does not necessarily return $\mathbf{A}_d$ to its original form, it represents a change of basis for $\mathbf{A}_d$ which preserves the eigenvalues and transforms the eigenvectors into the new basis. Consequently, this transformed $\mathbf{A}_d$ matrix can still be used to obtain the modal properties in the usual way.

4.3 Results

To verify that both the MLE and EM forms of the Prob-SSI algorithm are consistent with the classic SSI-Cov result, each method was used to analyse data simulated from a simple four degree-of-freedom linear dynamic system. Estimates for the modal properties were then obtained. Assuming a model order equal to the number of degrees-of-freedom. The recovered values for natural frequency and damping ratio from each method are provided in tables 4.1 and 4.2, respectively. The EM procedure was run for 500 iterations, ensuring adequate convergence of the algorithm.

Table 4.1: A comparison of the natural frequencies obtained using SSI-Cov, Prob-SSI (MLE) and Prob-SSI (EM). The values in the table were obtained using the correct model order (a model order of 4).

	Natural Frequency (Hz)			
	1	2	3	4
Truth	2.76369658	7.95774715	12.19197598	14.95567256
SSI-Cov	2.76578389	7.95521844	12.20072807	14.98244897
Prob-SSI (MLE)	2.76578389	7.95521844	12.20072807	14.98244897
Prob-SSI (EM)	2.76578384	7.95521938	12.20072861	14.98244332

Table 4.2: A comparison of the damping ratios obtained using SSI-Cov, Prob-SSI (MLE) and Prob-SSI (EM). The values in the table were obtained using the correct model order (a model order of 4).

	Damping Ratio			
	1	2	3	4
Truth	0.00868241	0.02500000	0.03830222	0.04698463
SSI-Cov	0.00884471	0.02341360	0.04204111	0.05184653
Prob-SSI (MLE)	0.00884471	0.02341360	0.04204111	0.05184653
Prob-SSI (EM)	0.00884468	0.02341346	0.04204116	0.05184743

As expected from the theory, the values for frequency and damping obtained using Prob-SSI (MLE) and standard SSI-Cov are identical, given use of the SVD in both cases to recover the estimates. In the case of Prob-SSI (EM), after 500 iterations of the algorithm, the estimates for frequency and damping can also be seen to be identical to the classic SSI-Cov estimates, up to 6 decimal places. Some minor variation is expected given the nature of the EM algorithm, however the precision observed in this case is sufficient.

4.4 Summary

This chapter presented a new formulation of the SSI-Cov algorithm, redefining it as a problem in probabilistic inference. This new approach was made possible through close alignment of the CVA-weighted SSI-Cov algorithm with the theory of probabilistic projections (i.e. PCCA). The mathematical equivalence between the MLE of the weights (linear transformations of the latent variables), and the observability matrix and controllability transposed, was established.

This unique perspective conveniently lays the necessary mathematical foundation to enable a suite of new SSI-based OMA algorithms, constructed using sophisticated probabilistic and hierarchical techniques, as will be explored in the later Chapters.

5

Robust Probabilistic Stochastic Subspace Identification

In the last chapter, the reader was introduced to a new probabilistic formulation of SSI-Cov, made possible by direct augmentation of the underlying algorithm; redefining the core mathematical construction of SSI-Cov as a latent variable model. This alternative definition leaves the output of SSI-Cov unchanged, but crucially provides the necessary form upon which new probabilistic methods can be constructed. Before exploring a Bayesian formulation, it is worth considering whether any extensions to the Prob-SSI model can be made to address some of the research challenges highlighted in Chapter 1.

Despite the success of OMA algorithms, a typical weakness lies in their handling of atypical observations. In practice, in-situ monitoring can often produce imperfect data containing unwanted features that may present a problem to OMA algorithms; which can severely bias parameter estimates and inevitably leading to the misidentification of a system. These features can arise because of low level excitation, the stochastic nature of the forcing (often not pure white noise) or from unpredicted events that

are often independent of the system being measured. In the case of OMA, these 'events' refer to practically encountered scenarios in testing such as sensor drop-out or sensor clipping.

This misidentification problem is also true of SSI-Cov. Unlike classical approaches to data analysis, where data may be pre-processed to remove outliers and 'cleaned' prior to use, such an approach cannot easily be applied when using SSI-Cov because of its dependence on sequential data, required for the formation of the Hankel matrix. This poses a significant dilemma. During application of SSI, any outliers present remain during the analysis, distorting the measurement of the response and ultimately affecting the identification procedure. Despite this discernible predicament, no simple mechanism currently exists capable of dealing with such anomalies in SSI.

Addressing this challenge, this chapter introduces the first of many new algorithms constructed using the probabilistic formulation of SSI — a statistically-robust SSI algorithm (robust Prob-SSI) — capable of providing a principled and automatic way of handling atypical observations in multi-output time series responses. As will be demonstrated, when confronted with 'corrupted' data, this new approach to modal identification outperforms traditional SSI in a number of scenarios, exhibiting improved pole stability and increased confidence; attributes reflected in the subsequent stabilisation diagrams. Further investigation and application to the well-documented Z24 bridge benchmark is also conducted, with results highlighting similar benefits to identification performance.

5.1 Related work

Although many alternative approaches to OMA have been presented, with an increase in probabilistic identification techniques (see Chapter 2), few directly attempt to formulate statistically robust algorithms capable of handling atypical observations. Most OMA algorithms possess some level of inherent robustness given their dependency on certain mathematical operations, such as the SVD, but in specific cases these algorithms are incapable of dealing with a considerable number of outliers and ensuring statistical robustness, as will be shown for SSI.

Some 'robust' approaches to OMA do exist in literature, however, the term 'robust' is often used to loosely refer to methods capable of reliable performance, rather than statistical robustness. The majority of these methods achieve the stability of the modes over multiple model orders through automated process; using optimised metrics or machine learning [194], often with the inclusion of uncertainty in techniques such as clustering [17]. In contrast, this work achieves statistical robustness through direct augmentation of the underlying algorithm. Furthermore, upon closer inspection of the literature, only a small selection of papers appear to directly address the problem of outliers, including recently developed robust algorithms by Liu et al. on correlation signal subset-based SSI (CoS-SSI) [20] and Goursat et al. [106] on Crystal-Clear SSI (CC-SSI); both concerned with addressing non-stationary and noisy signals.

5.2 Robust probabilistic SSI

Standard probabilistic models can be converted to a more 'robust' form through the replacement of a Gaussian noise model on the data and latent variables to that of a Student's t-distribution [27]. An introduction to the Student's t-distribution is provided in Appendix A.2. Building on the probabilistic interpretation of CCA, Archambeau, Delannay and Verleysen [192] used this alternative noise model to define a statistically robust equivalent of PCCA. This robust model is founded upon two key assumptions. The first is that the observed and latent variables can both be modelled by a Student's t-distribution $\mathcal{S}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ with mean $\boldsymbol{\mu}$, covariance $\boldsymbol{\Sigma}$ and degrees of freedom ν. Recall that the Student's' t-distribution has heavier tails than a typical Gaussian, which are determined by ν. The presence of heavier tails

is preferable as it makes the t-distribution better equipped to handle outliers; it is more likely to capture them within the existing distribution. The second assumption is that an outlier in the feature space must manifest as an outlier in the latent space. Given this set of assumptions, Archambeau et al. presented the following probabilistic model (note a continuation of notation from Section 4.1)

$$\mathbf{z}_n \sim \mathcal{S}(\mathbf{0}, \mathbb{I}_d, \nu) \tag{5.1}$$

$$\mathbf{x}_n^{(m)}|\mathbf{z}_n \sim \mathcal{S}(\mathbf{W}^{(m)}\mathbf{z}_n + \boldsymbol{\mu}^{(m)}, \boldsymbol{\Sigma}^{(m)}, \nu) \tag{5.2}$$

Making the set of latent variables (ν and $\mathbf{z}_n$) explicit, the model can also be written in the following form,

$$u_n \sim \mathcal{G}\left(\frac{\nu}{2}, \frac{\nu}{2}\right) \tag{5.3}$$

$$\mathbf{z}_n|u_n \sim \mathcal{N}\left(\mathbf{0}, u_n^{-1}\mathbb{I}_d\right) \tag{5.4}$$

$$\mathbf{x}_n^{(m)}|\mathbf{z}_n, u_n \sim \mathcal{N}\left(\mathbf{W}^{(m)}\mathbf{z}_n + \boldsymbol{\mu}^{(m)}, u_n^{-1}\boldsymbol{\Sigma}^{(m)}\right) \tag{5.5}$$

$$\mathbf{x}_n|\mathbf{z}_n, u_n \sim \mathcal{N}\left(\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu}, u_n^{-1}\boldsymbol{\Sigma}\right) \tag{5.6}$$

where $\mathcal{G}(\alpha, \beta)$ represents a Gamma distribution, u_n is an additional latent variable and $\mathbf{W}$, $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and ν are as previously defined. This alternative representation of the Student's t is common and arises as result of the t-distribution being viewed as a normal distribution with a covariance weighted by the Gamma distribution. This form is particularly useful, as the benefits of operating with Gaussian distributions is maintained, whilst also providing the necessary randomness in the covariance [27, 173, 175]. The corresponding graphical model to this alternative form is shown in Figure 5.1.

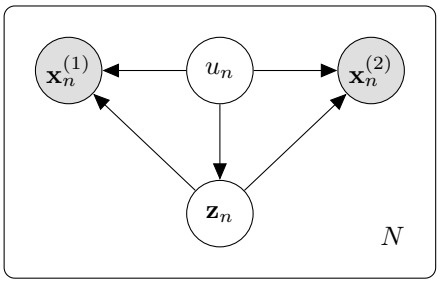


Figure 5.1: Graphical model for robust probabilistic canonical correlation analysis (robust-PCCA)

The same consideration in the model as previously shown (with CCA having a latent

variable model representation) allows for this robust extension to be achieved. The notable difference from which the robustness is generated comes from the inclusion of a *data point dependent* noise scaling. This inclusion allows inference over the influence of each datapoint on the total model. The mapping from the shared latent space is as before in PCCA, except for this modification to the noise process. The inclusion of Student's t- (or equivalently hierarchical Gamma) noise structures into probabilistic models allows inference on the 'importance' of datapoint, where outlying points can effectively be disregarded when u_n becomes large.

5.2.1 Expectation-maximisation

In contrast to PCCA, direct maximisation of the incomplete data log-likelihood $\sum_n \log p(\mathbf{x}_n, \mathbf{z}_n)$ with respect to the parameters is intractable. However, using Equations (5.3) - (5.6) estimates for the variables can be recovered through an iterative scheme. Archambeau et al. employ an EM approach (recall the EM algorithm introduced in Section 3.3.1), maximising the expected complete-log-likelihood [192].

$$\ell(\boldsymbol{\theta}|\mathbf{x}_n, \mathbf{z}_n, u_n) = \sum_{n=1}^{N} \ln p(\mathbf{x}_n, \mathbf{z}_n, u_n|\boldsymbol{\theta}) \tag{5.7}$$

where $\boldsymbol{\theta} = (\boldsymbol{\mu}, \mathbf{W}, \boldsymbol{\Sigma}, \nu)$.

The necessary update equations for the E-step (expectation) and M-step (maximisation) of robust PCCA are given by the Equations (5.8) - (5.15).

The reader may notice certain equations differ from those in the original robust projections manuscript. There appear to be some typographical errors in the original paper, most notably in the definition of the covariance matrix update equation (Equation (37) in [192]). Consequently, the robust PCCA algorithm has been re-derived in A.3 for completeness.

E-Step

$$\bar{u}_n = \frac{D+\nu}{u_n(\mathbf{x}_n-\boldsymbol{\mu})^\mathsf{T}\mathbf{A}^{-1}(\mathbf{x}_n-\boldsymbol{\mu})+\nu} \tag{5.8}$$

$$\ln \tilde{u}_n = \psi\left(\frac{D+\nu}{2}\right) - \ln\left(\frac{u_n(\mathbf{x}_n-\boldsymbol{\mu})^\mathsf{T}\mathbf{A}^{-1}(\mathbf{x}_n-\boldsymbol{\mu})+\nu}{2}\right) \tag{5.9}$$

$$\bar{\mathbf{z}}_n = \mathbf{B}^{-1}\mathbf{W}^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_n-\boldsymbol{\mu}) \tag{5.10}$$

$$\bar{\mathbf{S}}_n = \mathbf{B}^{-1} + \bar{u}_n\bar{\mathbf{z}}_n\bar{\mathbf{z}}_n^\mathsf{T} \tag{5.11}$$

where $\mathbf{A} = \left(\boldsymbol{\Sigma} + \mathbf{W}\mathbf{W}^\mathsf{T}\right)$, $\mathbf{B} = \mathbf{W}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\mathbf{W} + \mathbb{I}_d$, $\bar{u}_n = \mathbb{E}\left[u_n\right]$, $\ln(\tilde{u}_n) = \mathbb{E}\left[\ln(u_n)\right]$, $\bar{\mathbf{z}}_n = \mathbb{E}\left[\mathbf{z}_n\right]$, $\bar{\mathbf{S}}_n = \mathbb{E}\left[u_n\mathbf{z}_n\mathbf{z}_n^\mathsf{T}\right]$ and $\psi(\cdot)$ denotes the digamma function.

Subsequently, the update equations for the parameters in the M-step are given by

M-Step

$$\begin{aligned}\boldsymbol{\Sigma}' = \frac{1}{N}\sum_{n=1}^{N}\Big[&\bar{u}_n(\mathbf{x}_n-\boldsymbol{\mu})(\mathbf{x}_n-\boldsymbol{\mu})^\mathsf{T} - \bar{u}_n(\mathbf{x}_n-\boldsymbol{\mu})(\mathbf{W}\bar{\mathbf{z}}_n)^\mathsf{T} \\ &- \bar{u}_n(\mathbf{W}\bar{\mathbf{z}}_n)(\mathbf{x}_n-\boldsymbol{\mu})^\mathsf{T} + \bar{u}_n\mathbf{W}\bar{\mathbf{S}}_n\mathbf{W}^\mathsf{T}\Big]\end{aligned} \tag{5.12}$$

$$\mathbf{W}' = \frac{\sum_{n=1}^{N}\bar{u}_n(\mathbf{x}_n-\boldsymbol{\mu})\bar{\mathbf{z}}_n^\mathsf{T}}{\sum_{n=1}^{N}\bar{\mathbf{S}}_n} \tag{5.13}$$

$$\boldsymbol{\mu}' = \frac{\sum_{n=1}^{N}\bar{u}_n(\mathbf{x}_n-\mathbf{W}\bar{\mathbf{z}}_n)}{\sum_{n=1}^{N}\bar{u}_n} \tag{5.14}$$

$$0 = 1 + \ln\left(\frac{\nu}{2}\right) - 2\,\psi\left(\frac{\nu}{2}\right) + \frac{1}{N}\sum_{n=1}^{N}\left[\ln(\tilde{u}_n) - \bar{u}_n\right] \tag{5.15}$$

where $(\cdot)'$ denotes an updated parameter and ν' can be found by solving Equation (5.15) through a suitable line search. The convergence of the EM algorithm was monitored using the Q-function, derived using standard theory [27, 183].

It is important to note that despite the presence of heavier tails, in the case of no outliers, it is expected that the mean of the Students' t-distribution will converge to the PCCA maximum likelihood result.

After successful convergence of the EM algorithm, final estimates for the weights can be obtained. These weights, as described by the model, are considered to be the statistically robust transformations from the latent space to the observation space. The commonality between the weights recovered by PCCA and robust PCCA, based on their ability to transform the data to the relevant subspace, grant a similar replacement of CCA in the SSI procedure, now for robust PCCA.

$$\mathbf{W}^{(1)} = \mathcal{O}\mathbf{R} \tag{5.16}$$

$$\mathbf{W}^{(2)} = \mathcal{C}^{\mathsf{T}}\mathbf{R} \tag{5.17}$$

As noted by Archambeau et al., the weights recovered through EM do not account for the rotational ambiguity. Nevertheless, this rotation is irrelevant in the state matrix recovery when analysing the maximum model order, as previously shown in Section 4.1.1.

However, in the specific case of multi-order analysis for robust Prob-SSI, recovery of the rotation matrix is required to formulate the stabilisation diagram in the usual way for SSI-Cov, i.e. truncating the singular values from the maximum chosen model order rather than repeating analysis at multiple model orders. This is required because of the rotational invariance of the recovered observability matrix. The recovery and inclusion of the rotation matrix $\mathbf{R}$, as shown in Equation (5.16) ensures that removing columns from $\mathcal{O}$ is directly equivalent to truncating the singular values in the traditional case. The procedure for recovering $\mathbf{R}$ is shown in A.4.

5.3 Results and discussion

With a mathematical framework for conducting system identification in a statistically robust probabilistic way now established, attention can be directed towards robust identification of modal parameters using experimental data. Three separate case studies are used to demonstrate, evaluate and compare the identification performance of robust Prob-SSI to SSI-Cov. The first study (Section 5.3.1) uses data from a simulated linear MDOF to benchmark the general performance of robust Prob-SSI to standard SSI-Cov under ideal conditions. The variance exhibited by both methods due to different random input forcing is also briefly explored. The second case study (Section 5.3.2) exploits the same simulated MDOF system but now introduces artificial outliers into the measured timeseries. This study is used to test the robustness of identification to outliers under known conditions. A series of smaller case studies are also provided using the same simulated system, demonstrating the effectiveness of robust Prob-SSI to a range of possible outlier scenarios. The final case study (Section 5.3.4) uses data collected from the Z24 bridge benchmark. This study is used to investigate the method's overall applicability to measured data and evaluate its performance against SSI-Cov.

In all tests, robust Prob-SSI was directly compared to SSI-Cov to highlight differences in identification performance. The reader will notice the omission of results from Prob-SSI. As has been shown, the MLE estimates obtained through Prob-SSI are equivalent to those found through SSI-Cov and so their addition was redundant.

5.3.1 Benchmark: Simulated MDOF linear dynamic system

To evaluate the identification performance of robust Prob-SSI, a suitable benchmark was created. Response data was generated using a generic three degree-of-freedom (DOF) linear dynamic system with proportional damping; shown in Figure 5.2 and described by the model parameters in Equation 5.18. The system was subsequently excited using broadband forcing, of order 10^{-2} with a standard deviation of 1, mimicking ambient excitation. The system was simulated at a sample rate of 1×10^3 Hz and generated 8192 samples. The Hankel matrix was constructed using 20 lags; i.e. 10 "past" and 10 "future".

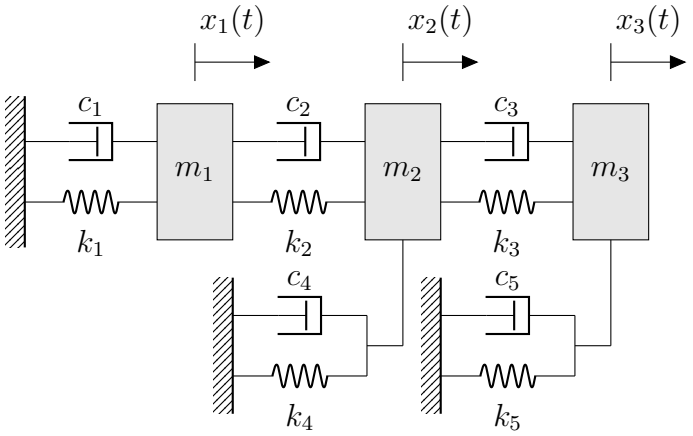


Figure 5.2: Simulated MDOF System

$$\mathbf{M} = \begin{bmatrix} m & 0 & 0 \\ 0 & m & 0 \\ 0 & 0 & m \end{bmatrix}, \quad \mathbf{K} = \begin{bmatrix} 2k & -k & 0 \\ -k & 4k & -0.5k \\ 0 & -0.5k & k \end{bmatrix}, \quad \mathbf{C} = \mathbf{K} \times 10^{-4} \tag{5.18}$$

where,

$$k = k_1 = k_2 = 2k_3 = \frac{2}{3}k_4 = 2k_5, \quad k = 1 \times 10^4 \text{ (N/m)}, \quad m = 10 \text{ (kg)}$$

Before application of robust Prob-SSI, suitable initial conditions are required for the EM algorithm. A small perturbation to the CCA (PCCA maximum likelihood) result was chosen as the initial estimate of the full weight matrix, providing a sensible initialisation point that, one would expect, converges to a solution sooner

and therefore reduces computational expense. The initial estimates for the covariance matrices were randomly sampled from an inverse Wishart distribution $\mathcal{W}^{-1}(\mathbf{K}, \nu)$ where $\nu = D + 2$ and $\mathbf{K} = \mathbb{I}_D$ given D is first dimension of Hankel matrix.

Following analysis of the data, stabilisation diagrams were generated for both methods using standard techniques. These are shown in Figure 5.3. To ensure fair comparison, fixed definitions for the stability criteria were applied to both methods. This ensured that any discrepancies in identified stability arose solely from the method and not through independent changes to specific criteria. The chosen criteria were: a 2% relative change in frequency, a 5% absolute change in damping ratio, and a 98% relative correspondence in the MAC value.

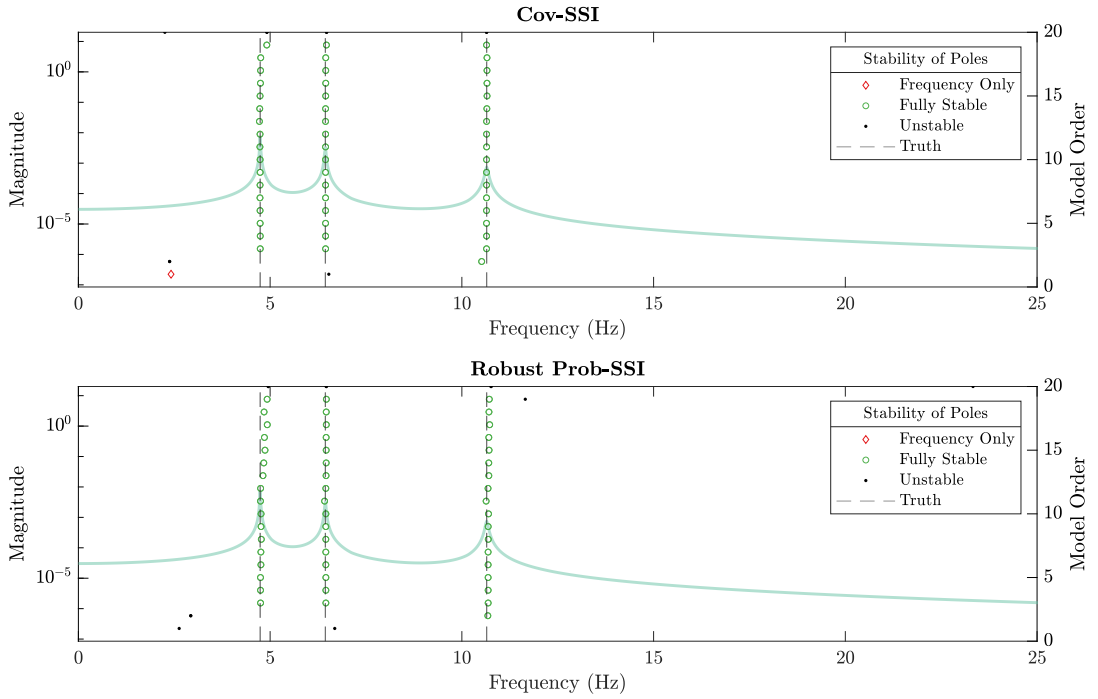


Figure 5.3: Stabilisation diagrams for the 3DOF system, recovered using SSI-Cov (top) and robust Prob-SSI (bottom).

As evident from Figure 5.3, the found stabilisation diagram for Robust Prob-SSI is highly comparable with that of SSI-Cov. Clear columns of fully stable poles are easily identifiable and, as shown in Table 5.1, the estimated values for the natural frequencies and damping ratios, obtained by taking the value at the correct model order, comfortably lie within an acceptable tolerance of the ground truth. The normalised mode shapes, also obtained at the correct model order, are presented in Figure 5.4 with both methodologies being in good agreement with the true modes.

Table 5.1: A comparison of the natural frequencies and damping ratios obtained in the benchmark study. The values in the table were obtained at the correct model order (a model order of 3).

	Natural Frequency (Hz)			Damping Ratio		
	1	2	3	1	2	3
Truth	4.74	6.44	10.65	0.0033	0.0020	0.0015
SSI-Cov	4.74	6.46	10.66	0.0026	0.0052	0.0131
Robust Prob-SSI	4.74	6.45	10.67	0.0031	0.0047	0.0104

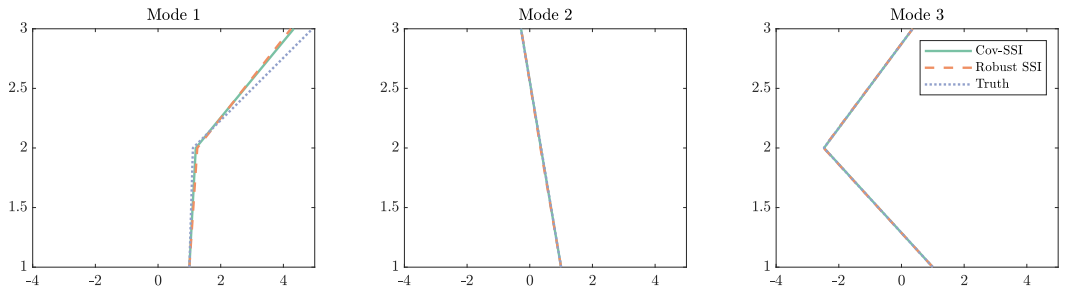


Figure 5.4: Mode shapes obtained using SSI-Cov and robust Prob-SSI compared against the true mode shape, when observed at the correct model order.

Due to stochastic variation in the excitation and, in the case of Robust Prob-SSI, the initialisation of the EM algorithm, the inherent variance shown by both methods was also explored. Assuming the correct number of DOF, both methods were applied to 100 datasets whereby the underlying system dynamics remain unchanged but the random seed needed to generate the random forcing varied between tests. The estimates of the natural frequencies from these tests are shown in Figure 5.5.

As expected, due to the complex nature of the system poles, the recovered estimates exist in conjugate pairs. This is evident from the clear matching of results seen in both tests between poles 1 and 2, poles 3 and 4, and poles 5 and 6. The variance exhibited by robust Prob-SSI is highly similar to that of SSI-Cov for all the recovered poles, when the correct number of degrees of freedom is assumed. Overall, this result demonstrates that robust Prob-SSI is capable of accurately replicating the results of SSI-Cov.

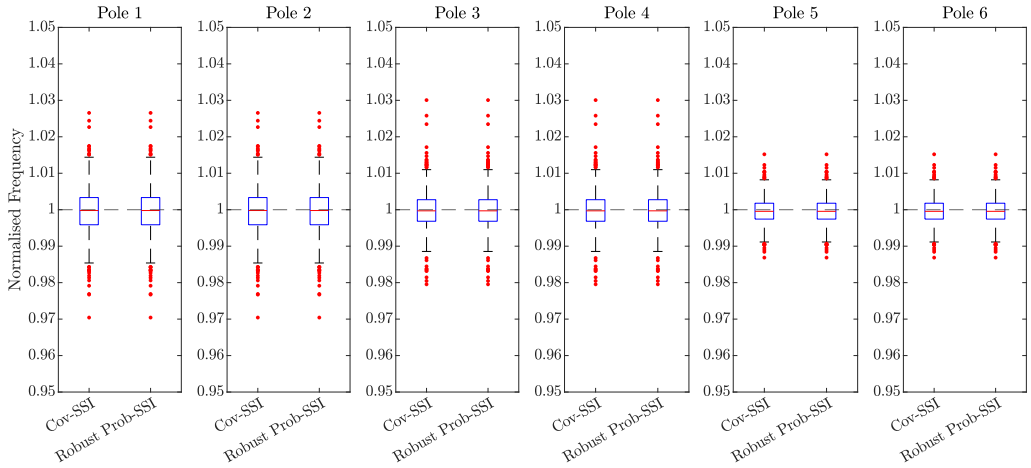


Figure 5.5: Box plots demonstrating the variance seen in the natural frequency estimates obtained using SSI-Cov and robust Prob-SSI for 500 datasets based on the same 3DOF dynamic system but different realisations of the random input forcing. (This is a standard boxplot, whereby the red line — indicates the mean, the limits of the blue box □ indicate the upper and lower quartiles and the red dots • represent datapoints further than 1.5x the inter quartile range away from the mean.)

5.3.2 Corrupted: Simulated MDOF linear dynamic system

Having established that robust Prob-SSI is capable of replicating the results of SSI-Cov for a simulated, 'clean' case, we now turn our attention to the proposed robustness of this method to corrupted datasets. In the context of this work, a 'corrupted' dataset will refer to any dataset containing either artificially generated or naturally occurring, atypical observations (outliers). It is worth briefly noting the definition of outliers in this case. Archambeau et al. assume that an outlier in the latent space will manifest as an outlier in the dataspace. This is a relatively sensible assumption given the linear transformation on the data. However, recall that SSI-Cov relies on components of the cross-covariance matrix and not the exact timeseries. Therefore, if the cross-covariance is insensitive to atypical observations present in the timeseries, there will be no discernible effect on the identification procedure. In the simulated case, this can be explored, but for real datasets, the 'clean' case is unobtainable. Furthermore, the terms 'atypical observations' and 'outliers' will be used interchangeably.

The corrupted dataset used in this study is based on the same linear dynamic MDOF system used in the benchmarking case but now with the inclusion of artificial outliers of a chosen type and dispersion. In the example shown here, outliers were introduced

at random locations in each sensor channel and set to a specified value, plus some small amount of noise. This was intended to mimic random sensor dropout where the measured value is pinned to the lower supply rail of the DAQ unit plus some noise. As one might expect, there are many different types and patterns of outliers that could manifest in experimental data and it would be extremely difficult to generate and analyse all such cases. Nevertheless, for completeness, a selection of corrupted datasets with varying outlier types were also generated and analysed. The results of this series of tests are presented in Section 5.3.3.

In the case shown, the number of outliers was set to 0.1% per channel which, for a dataset with 3 sensor channels and 8192 data points, equated to approximately 24 outliers in total. The response data and outliers are shown in Figure 5.6. To ensure a fair comparison of the two approaches, definitions for the stability criteria were once again fixed. The stabilisation diagrams recovered from this scenario are shown in Figure 5.7.

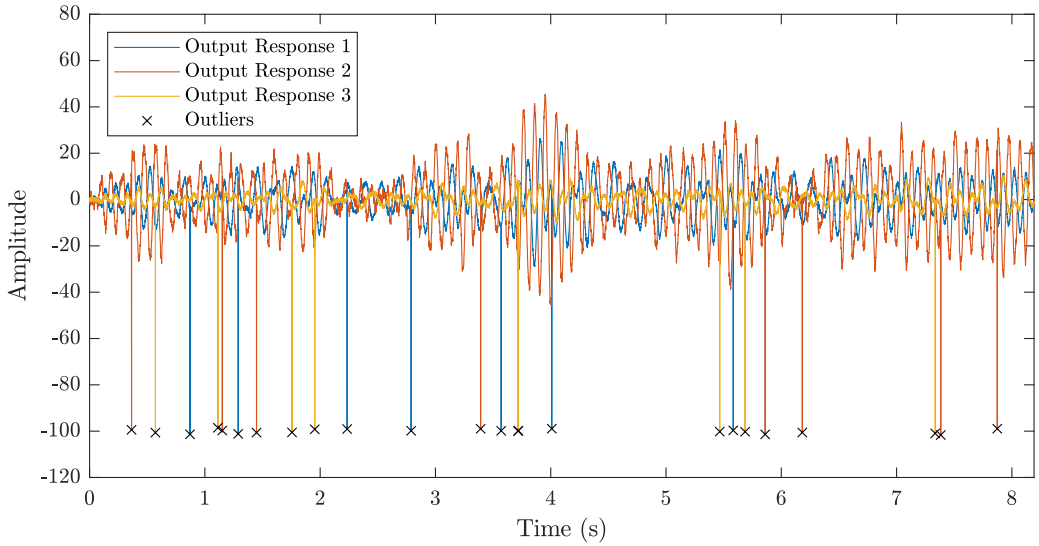


Figure 5.6: Response data from a simulated 3DOF 'corrupted' dataset, containing 0.1% artificially introduced, random outliers in each signal channel. The value of the outliers was set to a specified value plus some small amount of noise, intended to mimic random sensor dropout where the measured value is pinned to the lower supply rail of the DAQ unit with some noise

It is immediately apparent that robust Prob-SSI displays a significant improvement in identification performance over SSI-Cov when confronted with the corrupted dataset. Whilst robust Prob-SSI comfortably identifies all three correct natural frequencies of

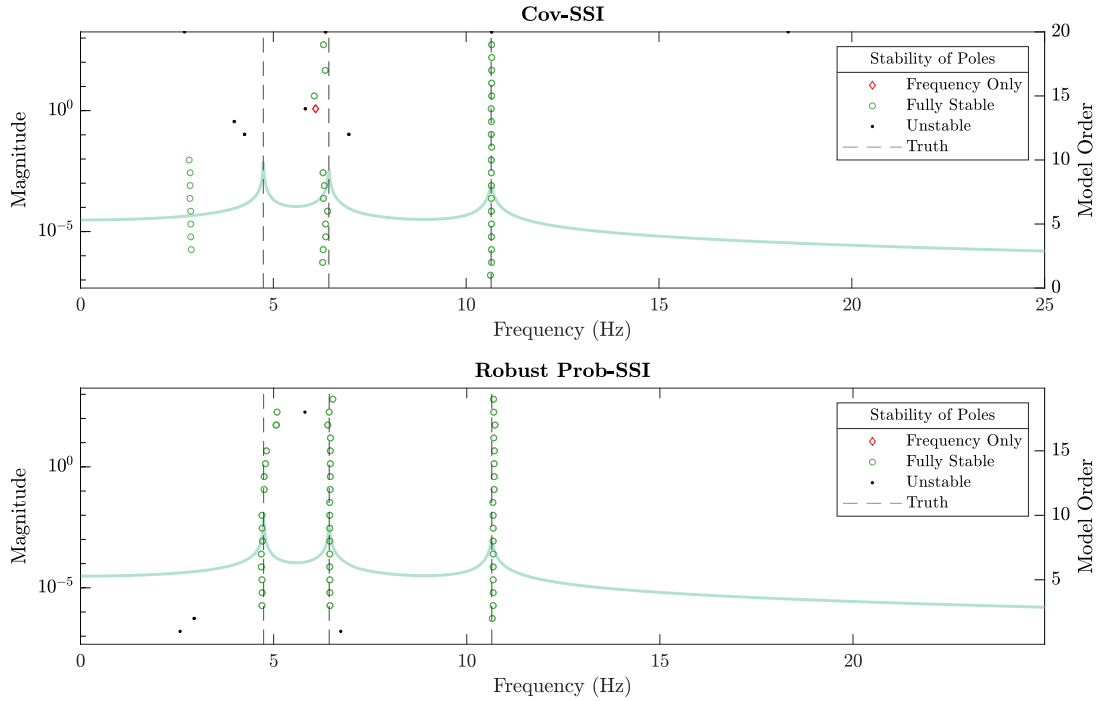


Figure 5.7: Stabilisation diagrams recovered using SSI-Cov (top) and robust Prob-SSI (bottom) using response data from a simulated 3DOF 'corrupted' dataset, containing 0.1% artificially introduced, randomly located outliers (in each channel), set to a specified value, plus some small amount of noise.

the system, traditional SSI-Cov fails to find the first, struggles to accurately identify the second and only successfully identifies the third natural frequency. This failure in identification can be clearly observed as the lack of a column of fully stable poles in the upper frame of Figure 5.7 and in the plot of the normalised mode shapes, see Figure 5.8.

However, it is expected that there will be several instances where, despite the presence of outliers, SSI will continue to perform as expected. Similarly, there will also be cases where robust Prob-SSI will also fail to identify the system. The replacement of SSI with robust Prob-SSI is not a straightforward one. In Section 5.3.3, it is illustrated how increasing the percentage of outliers inevitably leads to a deterioration in performance for both methods. Clearly, although robust Prob-SSI offers some protection to corrupted responses, in the presence of many outliers it will be impossible to identify the dynamics.

As with the benchmark case, the variance of the estimates from both tests was also

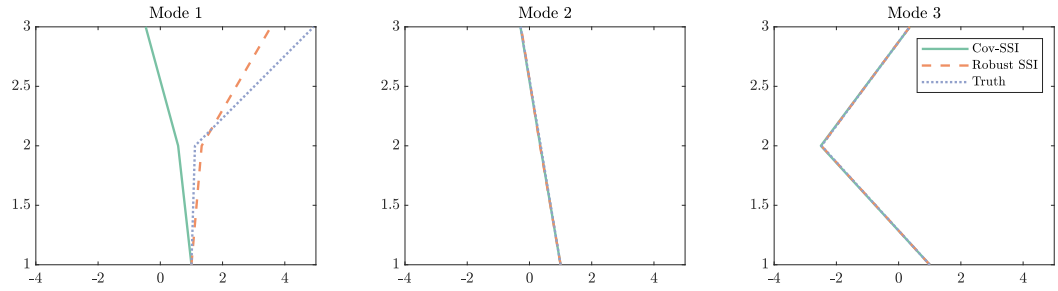


Figure 5.8: Mode shapes obtained using SSI-Cov and robust Prob-SSI, compared against the true mode shape, when observed at the correct model order.

explored, assuming the correct number of DOF. The variance in the resulting natural frequency estimates are shown in Figure 5.9. Traditionally, if a pole has no conjugate, then it can be ignored when calculating the corresponding frequencies and so often only three plots are required. However, here non conjugate poles are retained as they represent an instance of poor identification performance on a given dataset. As clearly shown, the variance in the frequency estimates from SSI-Cov is significantly greater than that of robust Prob-SSI when analysing data containing outliers.

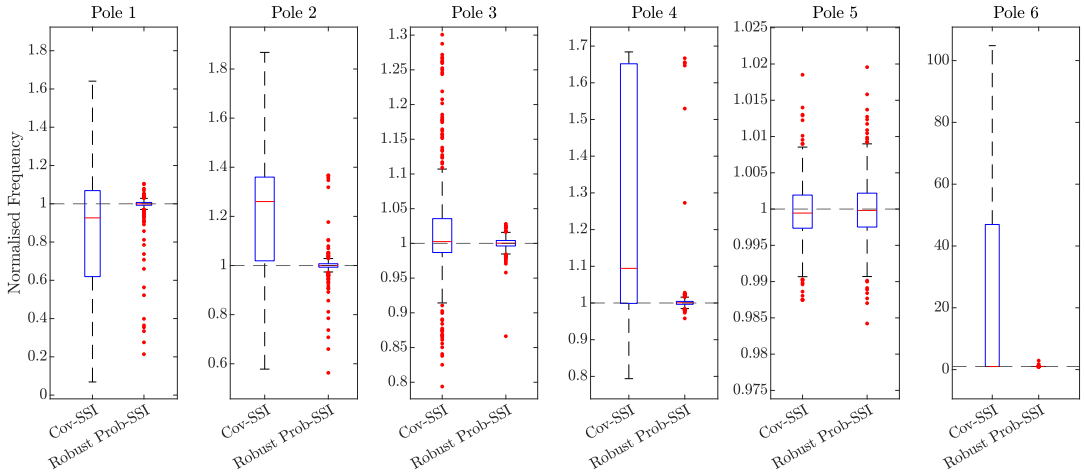


Figure 5.9: A demonstration of the variance seen in the natural frequency estimates when using SSI-Cov and robust Prob-SSI to analyse response data from 100 datasets of the same 3DOF dynamic system with varying input forcing but with artificially induced random outliers (0.1% per channel).(This is a standard boxplot, whereby the red line — indicates the mean, the limits of the blue box □ indicate the upper and lower quartiles and the red dots • represent datapoints further than 1.5x the inter quartile range away from the mean.)

5.3.3 Corrupted: Alternative studies

Employing the same underlying linear dynamic system used in the previous section, a range of alternative outlier cases are also explored. This extended selection of case studies is designed to assess the effects of varying outlier type on identification performance for SSI-Cov and verify the applicability of robust Prob-SSI to other challenging cases.

Periodic block dropout

In this example, the outliers are designed to mimic periodic dropout of a single sensor channel to an 'electrical noise floor', repeating in regular blocks with a duration of 0.01 seconds. The response data with outliers is shown in Figure 5.10. The resulting stabilisation diagrams for SSI-Cov and robust Prob-SSI are shown in Figure 5.11.

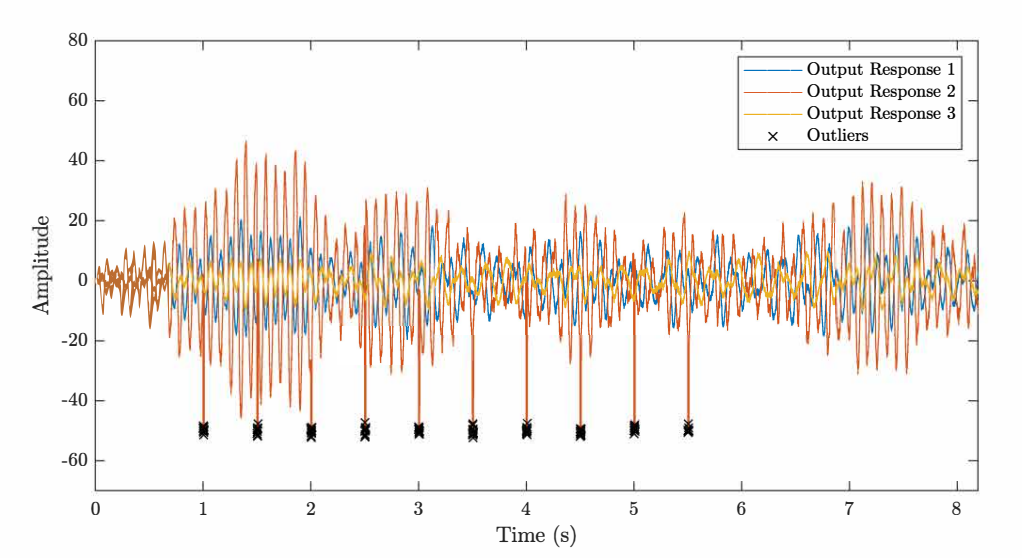


Figure 5.10: Response data from a simulated 3DOF 'corrupted' dataset, containing 0.1% of artificially introduced, periodic blocks of outliers in one channel. The value of the outliers was set to a specified value, plus some small amount of noise.

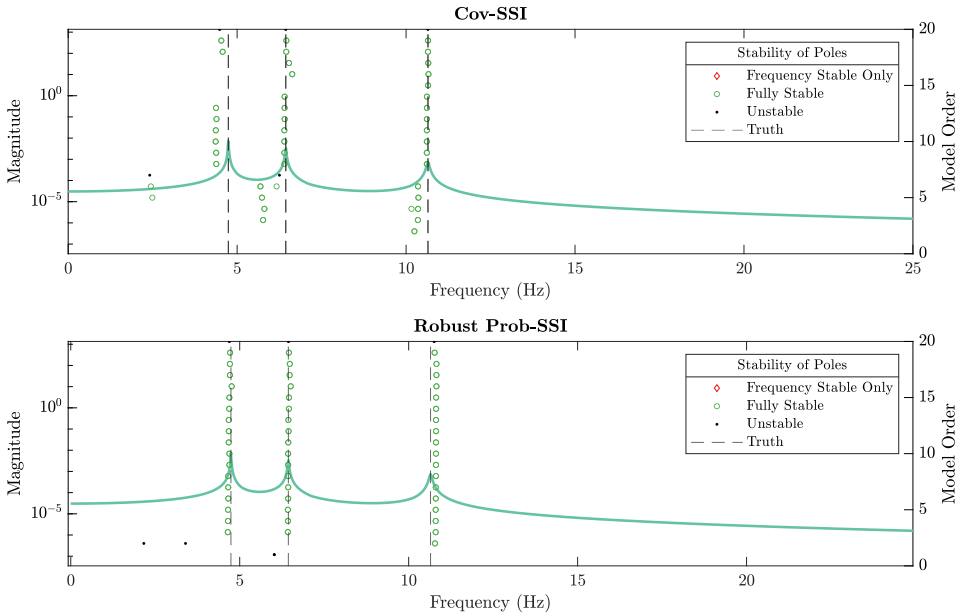


Figure 5.11: Stabilisation diagrams recovered using SSI-Cov (top) and robust Prob-SSI (bottom) for the periodic block dropout case.

The performance of robust Prob-SSI in this case appears to demonstrate a significant improvement over the SSI-Cov result, with greater stability in the identified poles and better identification of the natural frequencies.

Clipping

In this example, data are introduced designed to mimic the clipping of all sensor channels to 80% of the maximum amplitude of the individual signals. Although not strictly outliers, clipping represents an alternative generating mechanism that produces observation not consistent with the physics. The output response and these adjusted observations are shown in Figure 5.12, whilst the stabilisation diagrams for SSI-Cov and robust Prob-SSI are shown in Figure 5.13.

Although demonstrating slightly less stability in the identified modes, particularly at the first natural frequency, robust Prob-SSI can be seen to correctly identify the location of first natural frequency. In comparison, the traditional SSI-Cov algorithm completely misidentifies the first natural frequency, displaying significant stability of the poles but at the wrong frequency location.

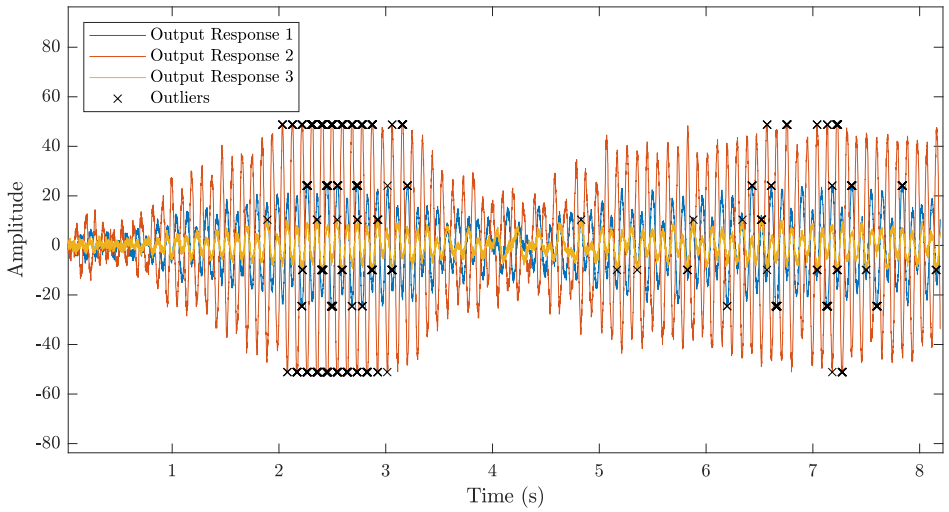


Figure 5.12: Response data from a simulated 3DOF 'corrupted' dataset, where all channels were clipped to 80% of the maximum value of each channel.

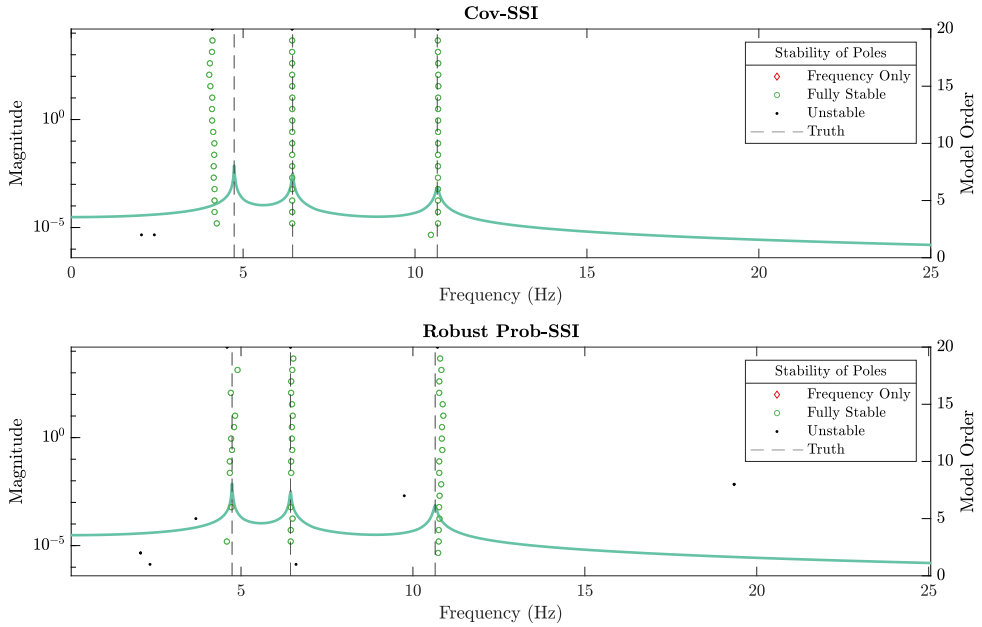


Figure 5.13: Stabilisation diagrams recovered using (a) SSI-Cov and (b) robust Prob-SSI using response data from a simulated 3DOF 'corrupted' dataset, where all channels are clipped to 80% of the maximum value of each channel.

Block dropout to zero

The outliers in this example are designed to mimic the block dropout of a single sensor channel to zero. The dropout starts at 3 seconds and has a duration of 1 second. The corrupted responses are shown in Figure 5.14, whilst the stabilisation diagrams for SSI-Cov and robust Prob-SSI are shown in Figure 5.15.

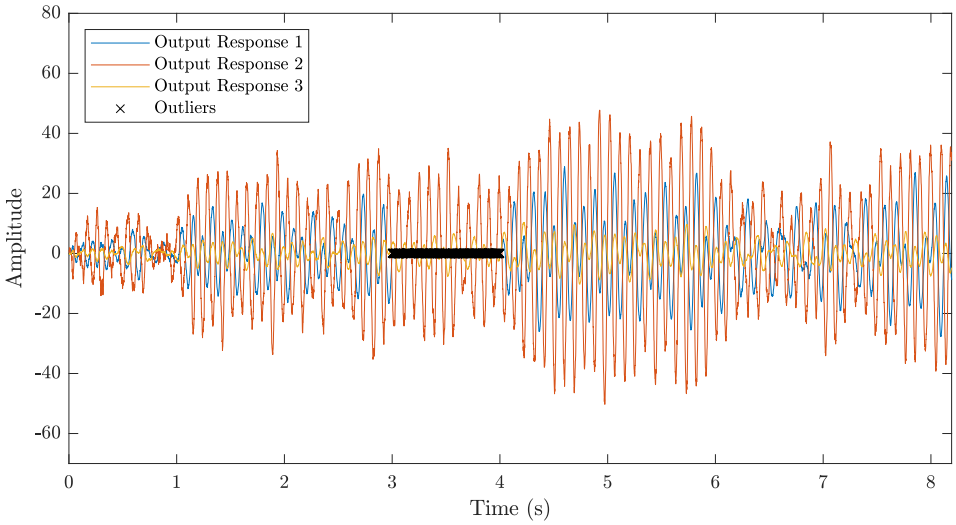


Figure 5.14: Response data from a simulated 3DOF 'corrupted' dataset, where values in a single channel were set to zero amplitude for a 1000 point block in the 8192 point long data series.

In this example, SSI-Cov completely misidentifies the first two natural frequencies, displaying estimates with good stability but at values much higher than the true values. In contrast, robust Prob-SSI displays much better identification performance, with similar stability but with estimated frequencies much closer to the truth.

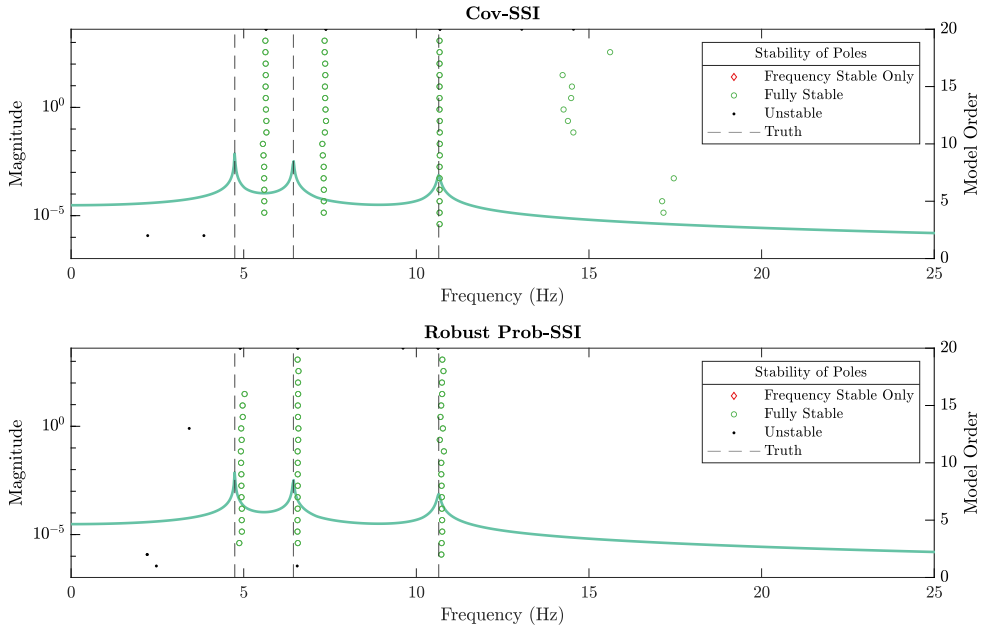


Figure 5.15: Stabilisation diagrams recovered using (a) SSI-Cov and (b) robust Prob-SSI using response data from a simulated 3DOF 'corrupted' dataset, where values in a single channel were set to zero amplitude for a 1000 point block in the 8192 point long data series.

Percentage outlier study

In this study, the effect of changing the percentage of outliers on identification performance is also explored. Using the same outlier scenario from Section 5.3.2, the response data were subjected to an increase of 0.5% random outliers in each channel. The resulting stabilisation diagrams for SSI-Cov and robust Prob-SSI are shown in Figure 5.16.

As can be seen from Figure 5.16, SSI-Cov completely fails to identify the first two natural frequencies of the system, displaying no stable poles around the natural frequencies. In comparison, robust SSI-Cov still manages to correctly identify the first two natural frequencies, producing clear regions of stable poles. The identified poles are not as 'stable' when compared to the 0.1% outlier case shown in Figure 5.7, however this is to be expected.

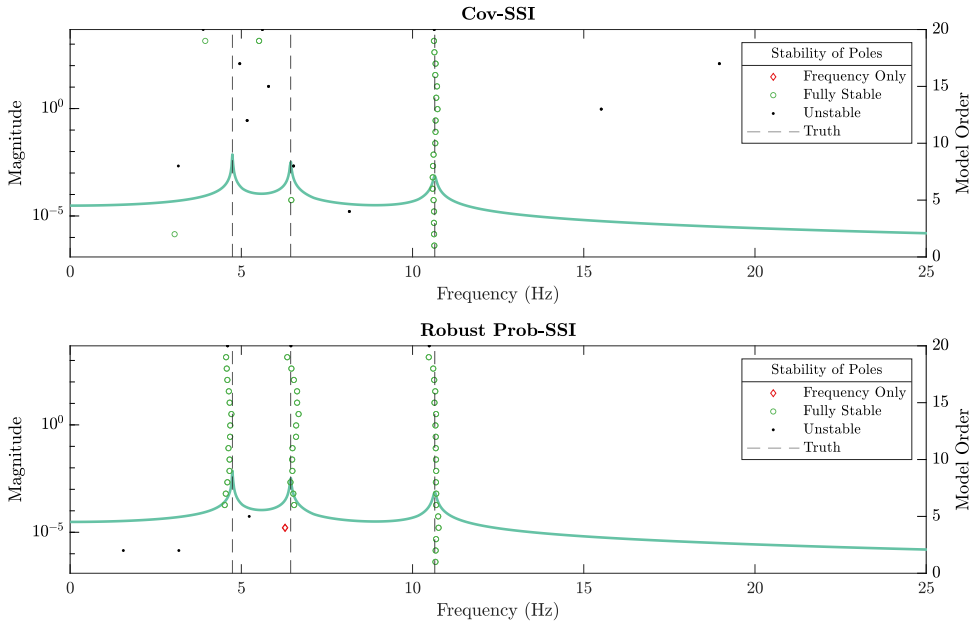


Figure 5.16: Stabilisation diagrams recovered using (a) SSI-Cov and (b) robust Prob-SSI using response data from a simulated 3DOF 'corrupted' dataset, containing 0.5% artificially introduced, randomly located outliers (in each channel), set to a specified value, plus some small amount of noise.

5.3.4 Case study: Z24 bridge

This final study is used to investigate and evaluate the performance of robust Prob-SSI when confronted with a real world example; a subset of the ever dependable Z24 bridge dataset. The Z24 is a (now demolished) post-tensioned concrete two-cell box-girder bridge once located between Bern and Zurich, Switzerland. For more information on the Z24 bridge, access to the dataset and references to the original work, the reader is directed towards the following resources (bwk.kuleuven.be/bwm/z24/z24) [40, 195]. Data collected from the Z24 bridge is often employed for new SHM tasks concerned with temporal changes to the modal properties, induced by damage and/or temperature. Its frequent application across structural dynamic research for benchmarking new system identification approaches to large scale structures, make it a sensible choice for this study.

It is unknown if the Z24 data features any outlying observations or signal corruption. However, multiple output acceleration signals, across multiple tests in the initial stages of testing, appear to demonstrate clipping (see Figure 5.17). It is undetermined

in previous literature if this clipping affects the identification and fundamentally this will remain unknown. However, this interesting feature of the Z24 data makes it highly suitable for our investigation into robust Prob-SSI as a robust alternative to SSI given possible a-typical observations.

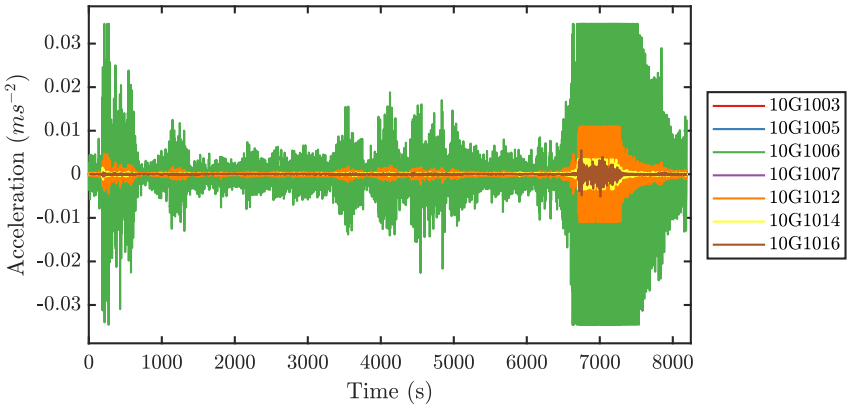


Figure 5.17: A sample of output time series acceleration measurements (the first 8192 points) taken from the Z24 bridge dataset 10G10 (corresponding to data recorded on '17-Jan-1998' at '10:00:00'), that appear to demonstrate over-ranging of the sensors and therefore clipping of the measured signals.

A subset of the Z24 dataset, specifically data folder 07E01 containing acceleration measurements observed on the '06-Dec-1997' at '10:00:00', prior to the induced damage, was initially selected at random for processing. Only the initial segment, the first 8192 points, was used in the analysis. The resulting stabilisation diagrams generated using both SSI methods are shown in Figure 5.18. The frequency range is limited to show the first 4 natural frequencies, as identified in the original works, as these are often the main frequencies used when analysing this dataset.

The first two identified modes are easily observable in both diagrams, given the presence of highly stable poles. The same can also be said of the third and fourth modes, with robust Prob-SSI demonstrating better stability when compared to SSI-Cov over increasing model order. Across all the identified modes, one could argue that robust Prob-SSI identifies stable poles at all of these modes at lower model orders than SSI-Cov and thus, would provide more confidence to the practitioner. There is some speculation in the literature and wider SHM community regarding the existence of a mode at ~ 7 Hz. Interestingly, the robust approach appears to largely ignore the presence of this mode suggesting it may be the result of noise, aptly accounted for in this case by the Student's t-noise model. It is unclear, solely from

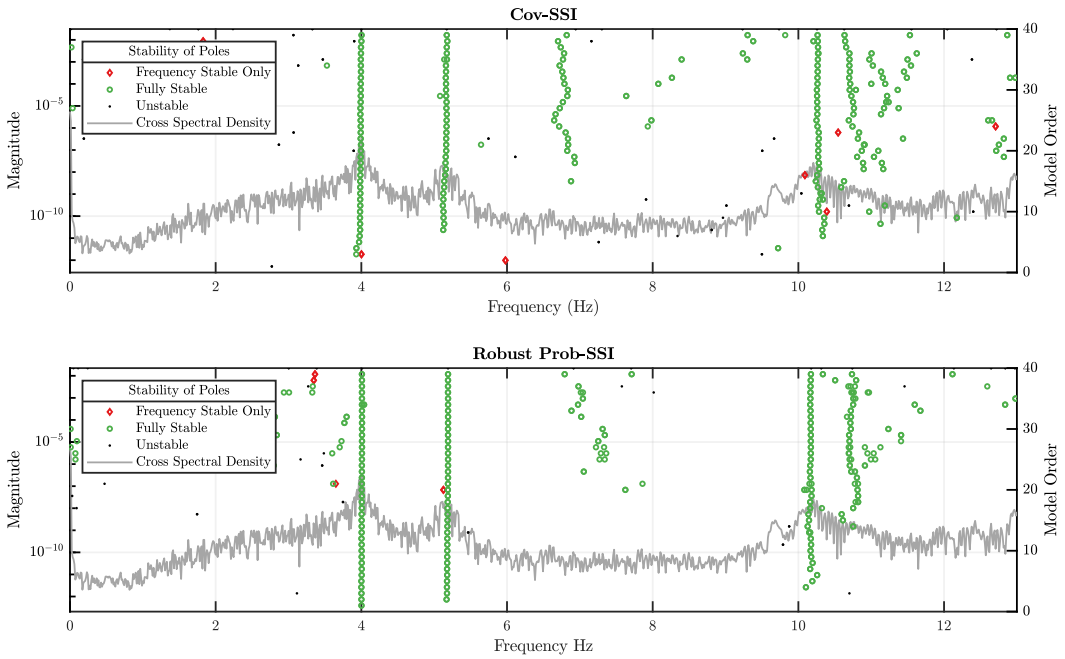


Figure 5.18: Stabilisation diagrams generated using results from SSI-Cov (top) and robust Prob-SSI (bottom) after application to a single subset of the Z24 bridge dataset (07E01); corresponding to data recorded on '06-Dec-1997' at '01:00:00'

the results based on these datasets, whether anything further can be said regarding this potential mode as a result of the robust SSI approach presented here.

This promising initial evidence suggests that robust Prob-SSI exhibits comparatively better identification performance over SSI-Cov. However, a single dataset alone is insufficient to make such a claim. To better assess the overall identification performance to a larger range of cases, SSI-Cov and robust Prob-SSI were applied to the first segment (first 8192 points) of multiple Z24 datafiles, specifically those dated 21st November 1997 14:00:00 through to 25th April 1998 14:00:00, and the resultant poles extracted.

Rather than manually selecting stable poles in each stabilisation diagram, a very time consuming activity, automatic selection of the stable poles was achieved using an implementation of the clustering algorithm DBSCAN. The application of DBSCAN (or derivative thereof, OPTICS) to pole selection in stabilisation diagrams is not a novel one, existing work by Boroschek and Bilbao can be found here [196]. As the performance of any given clustering technique was not the focus of this study, the decision to use DBSCAN clustering over another was arbitrary, however as a very

common mode of clustering and one frequently cited in literature, it may seem the most sensible. For information on the theory and implementation of DBSCAN, the reader is directed here [197].

Akin to the approach used by Boroschek and Bilbao, the distance metric in DBSCAN was chosen to have the following form:

$$dist(p_i, p_j) = \frac{|\omega_i - \omega_j|}{\max(\omega_i, \omega_j)} + (1 - \text{MAC}(\phi_i, \phi_j)) \tag{5.19}$$

where the natural frequencies ω_i, ω_j and mode shapes ϕ_i, ϕ_j correspond to poles p_i and p_j respectively, and the MAC is the modal assurance criterion. The reachability distance was chosen to be 0.005%, with the minimum number of objects set to 25.

Following application of DBSCAN to the Z24 data, centres of the recovered pole clusters were plotted against the corresponding date and time of collection. Figure 5.19 displays the temporal changes to the natural frequency estimates obtained using SSI-Cov and robust Prob-SSI. The first noticeable difference between the two plots in Figure 5.19 is variation in the number of stable poles situated around 7 Hz, corresponding to the supposedly illusive mode. Congruent to the assessment of the single stabilisation diagram, robust Prob-SSI (bottom) appears to find a lower number of stable poles at this frequency across the full range of tests, when compared to SSI-Cov (top). Again, this is likely due to any noise generating this nuisance mode, being accounted for by the Student's t noise model of robust Prob-SSI.

Turning attention to the first two modes, much lower variance is observed in the frequency estimates recovered using robust Prob-SSI across the entire test series compared to SSI-Cov. This reinforces preliminary conclusions drawn from Figure 5.18 regarding the increased stability of recovered poles using robust Prob-SSI. This reduction in variance is also apparent, albeit less obviously, in modes three and four, suggesting robust Prob-SSI is capable of recovering less noisy pole estimates than SSI-Cov. This has several benefits, not least in providing the experienced practioner with more confidence when selecting suitable poles from the stabilisation diagram. This new technique may also prove to be a preferred method of modal parameter recovery for SHM schemes tasked with accurately monitoring temporal changes to modal properties and inferring possible damage. Reducing variance in modal property estimates is of value, in the SHM setting, as it offers a route to reducing false positives and increasing sensitivity in anomaly detection.

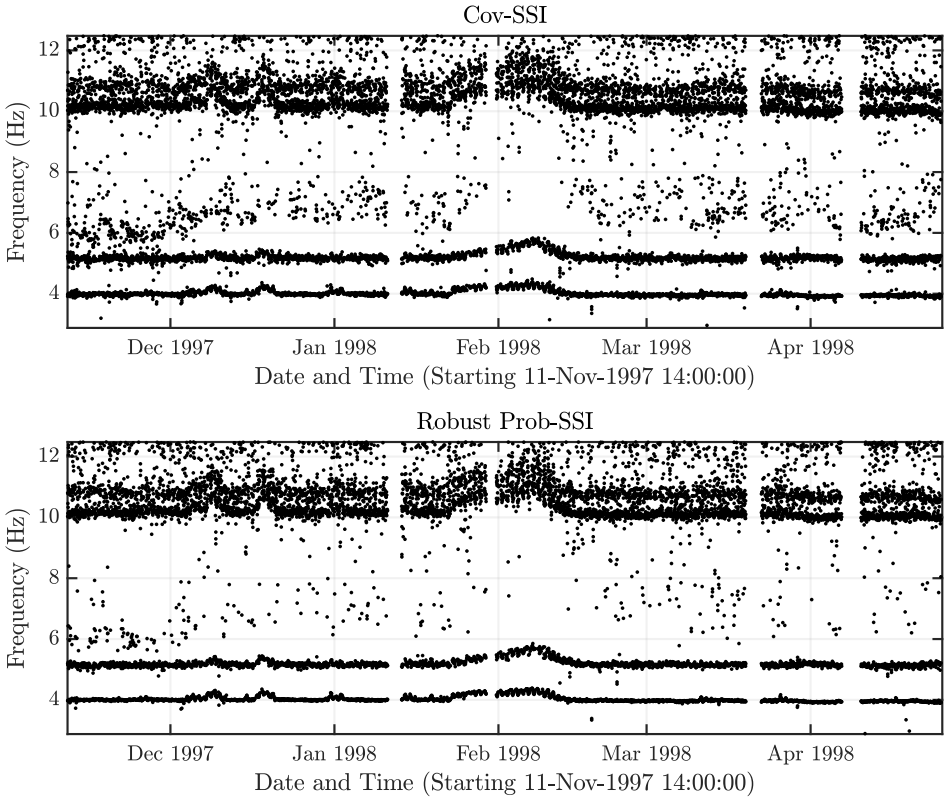


Figure 5.19: Temporal changes to the natural frequency estimates of the Z24 bridge, identified using SSI-Cov (top) and robust Prob-SSI (below). DBSCAN clustering was used to extract the non-spurious, stable poles from each individual dataset. Both methods were applied to the first segment (first 8192 points) of all datafiles ranging from 21st November 1997 14:00:00 to 25th April 1998 14:00:00.

5.4 Summary

In this chapter, the probabilistic SSI algorithm established in Chapter 4 has been extended to a statistically-robust form (robust Prob-SSI), capable of providing a principled and automatic way of handling atypical observations in multi-output time-series responses. One of the key research challenges in OMA identified in the introduction.

This algorithm was constructed by replacing the Gaussian noise model over the data to that of a Student's t-distribution. The inclusion of a Student's t-distributed noise model principally forms robust CCA [192] which, when considered in context of the SSI-Cov algorithm, provides a statistically-robust mechanism for performing system identification.

To investigate the proposed algorithm, the performance was evaluated against a range of case studies. The first study benchmarked robust Prob-SSI against the conventional SSI-Cov algorithm using simulated data for a MDOF linear system. The modal properties identified from both algorithms displayed highly comparable results, with similar variance in the estimates from multiple instances of the random noise. This reinforces the idea that, when presented with data containing no outliers, the two methodologies should provide highly similar results.

The second study evaluated the ability of both methods to resist misidentification when presented with atypical observations in the measured time series. The algorithms were exposed to a series of 'corrupted' datasets containing artificially introduced outliers; designed to mimic a typical forms of anomalous data observed in experimental testing. Subsequent analysis showed robust Prob-SSI outperform conventional SSI in all cases, with the former exhibiting better modal identification performance and higher confidence (greater stability) in the recovered poles over multiple model orders.

The final study in this chapter investigated the application of robust Prob-SSI to data from the Z24 bridge dataset, chosen to better demonstrate identification performance on data indicative of a real system. Analysis was conducted on a single dataset, producing stabilisation diagrams for SSI-Cov and robust Prob-SSI. Results from this analysis showed that robust Prob-SSI was able to locate the natural frequencies and form stable columns of poles at lower model orders than SSI-Cov.

Analysis was also conducted on multiple datasets, obtained periodically for continued monitoring of the Z24. The poles from the resulting stabilisation diagrams were clustered to find the centre of the stable columns of poles, and then plotted in time. From these results, it was shown that robust Prob-SSI shows less variance in the recovered natural frequencies when compared to SSI-Cov, affirming the ability of the robust Prob-SSI to form less noisy estimates of the recovered poles. A reduction in the variance over modal estimates has significant benefits, not limited to assisting SHM strategies that rely on accurate modal information for decision-making tasks.

One potential limitation of this algorithm is that the algorithm is iterative in nature and therefore slower than the traditional SVD-based SSI algorithm. Nevertheless, this can be improved with alternative computation schemes or code optimisation.

Crucially, the work introduced in this chapter has highlighted the first of several new formulations of SSI, made possible through viewing SSI as a problem in probabilistic inference. In the wider context of this thesis, the next logical model choice would be a Bayesian formulation, one where uncertainty quantification is possible. Such an approach is explored in the next chapter.

6

Bayesian Stochastic Subspace Identification

In the last chapter, the reader was introduced to a statistically robust SSI algorithm; a method shown to maintain better identification performance than traditional SSI when presented with atypical observations in time-series responses. This algorithm was the first of many possible extensions to the SSI algorithm, made possible through the novel probabilistic form of SSI (Prob-SSI) introduced in Chapter 4. Although not addressing the problem of uncertainty quantification (UQ), it was demonstrated how this new probabilistic form of SSI allowed for an effective consideration of atypical observations in or distortion of data — a key challenge in OMA — offering a novel approach to robust identification. Linking back to UQ, this chapter addresses the key aim of this thesis: the development of a new Bayesian OMA algorithm, capable of recovering posterior distributions over modal properties.

One of the benefits to the probabilistic form of SSI introduced in Chapter 4, is in its ability to admit an arbitrary prior structure; one that enables the natural extension to a Bayesian formulation. The inclusion of prior distributions over the parameters of the latent variable CCA model extend the problem into one of Bayesian inference, where posterior distributions are obtainable. This model is known as Bayesian

CCA [198, 199]. Through the inclusion of Bayesian CCA in the SSI algorithm, this Chapter will present a new algorithm called Bayesian SSI and demonstrate how the posterior over the observability matrix can be recovered through two different inference schemes: Gibbs sampling and Variational Inference (VI). The identification and UQ performance of Bayesian SSI is then assessed using two case studies. The first is a benchmark study, using data simulated from a linear multi degree-of-freedom system, and the second applies the algorithm to measured data from an in-service structure, specifically the Z24 bridge.

6.1 Bayesian SSI

Following the introduction of probabilistic CCA (PCCA), Wang [198] and Klami and Kaski [199] presented the natural hierarchical extension to the PCCA model and introduced Bayesian CCA. Their model incorporated priors over the model parameters $\boldsymbol{\theta} = \{\mathbf{W}, \boldsymbol{\Sigma}, \boldsymbol{\mu}\}$ and introduced a sparsity inducing prior over the columns of the transformation matrix. This model was later extended for inter-battery factor analysis to include local and global latent variables in [200] and for a mixture of robust CCA models in [201]. The sparsity inducing prior is not included in the Bayesian CCA model described in this paper, nor is it included in later derivations. This alternative model, coupled with the desire for a generalisation of the prior structure (not present in the original works [198, 199]), motivated the need to redefine the model of CCA in this work. The graphical model for the Bayesian CCA problem is shown in Figure 6.1. With a continuation of notation from Chapter 4, the general probabilistic from Equations (4.1) and (4.3) are recalled for the readers benefit, whilst Equations (6.1) - (6.4) define the priors placed over the model parameters.

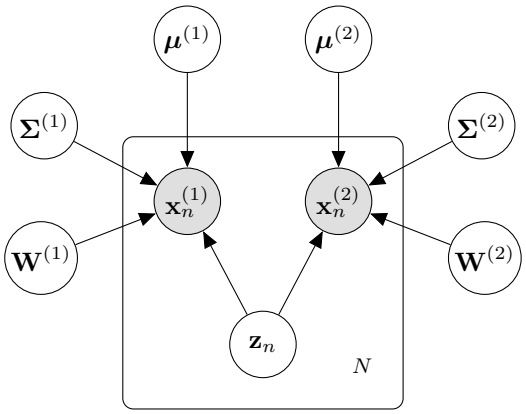


Figure 6.1: Graphical model for the Bayesian interpretation of CCA

$$
\begin{aligned}
\mathbf{z}_n &\sim \mathcal{N}(0, \mathbb{I}) \\
\mathbf{x}_n|\mathbf{z}_n &\sim \mathcal{N}(\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu}, \boldsymbol{\Sigma})
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{w}_i &\sim \mathcal{N}\left(\boldsymbol{\mu}_{\mathbf{w}_i}, \Sigma_{\mathbf{w}_i}\right) && (6.1) \\
\mathbf{W} &\sim \prod_{i=1}^{d} p(\mathbf{w}_i) && (6.2) \\
\boldsymbol{\mu} &\sim \mathcal{N}\left(\boldsymbol{\mu}_{\mu}, \Sigma_{\boldsymbol{\mu}}\right) && (6.3) \\
\boldsymbol{\Sigma} &\sim I\mathcal{W}\left(\mathbf{K}_0, \nu_0\right) && (6.4)
\end{aligned}
$$

where $\mathbf{w}_i$ denotes an individual column of $\mathbf{W}$ where, following on from earlier notation, $\mathbf{w}_i = [\mathbf{w}_i^{(1)}; \mathbf{w}_i^{(2)}]$. The notation $p(\cdot)$ refers to the probability density function in vector or matrix form dependent on the parameters, and $I\mathcal{W}\left(\mathbf{K}, \nu\right)$ corresponds to an inverse Wishart distribution with scale $\mathbf{K}$ and ν degrees of freedom. The precision $\boldsymbol{\Psi} = \boldsymbol{\Sigma}^{-1}$ can also be defined as the Wishart distribution $\mathcal{W}\left(\mathbf{K}^{-1}, \nu\right)$. The prior over $\mathbf{W}$ is constructed assuming independent priors over the columns.

Considering the model above, the full joint likelihood can be expressed as

$$
p(\mathbf{x}, \mathbf{z}, \mathbf{W}, \boldsymbol{\Sigma}, \boldsymbol{\mu}) = p(\mathbf{x}|\mathbf{z}, \mathbf{W}, \boldsymbol{\Sigma}, \boldsymbol{\mu}) p(\mathbf{z}) p(\boldsymbol{\Sigma}) p(\boldsymbol{\mu}) p(\mathbf{W}) \tag{6.5}
$$

Accounting for the independent columns of W, such that $p(\mathbf{w}_i)$ describes each column,

Equation (6.5) becomes

$$p(\mathbf{x}, \mathbf{z}, \mathbf{W}, \boldsymbol{\Sigma}, \boldsymbol{\mu}) = p(\mathbf{x}|\mathbf{z}, \mathbf{W}, \boldsymbol{\Sigma}, \boldsymbol{\mu})p(\mathbf{z})p(\boldsymbol{\Sigma})p(\boldsymbol{\mu}) \prod_{i=1}^{d} p(\mathbf{w}_i) \tag{6.6}$$

Given the commonality between the desired weights in PCCA and the desired weights in Bayesian CCA, based on their ability to transform data to the relevant subspace, it is sensible to assume the posterior distributions over those weights also corresponds to posterior distributions over the observability and controllability matrices in the context of SSI.

$$\mathbf{W}^{(1)} = \mathcal{O}\mathbf{R} \tag{6.7}$$

$$\mathbf{W}^{(2)} = \mathcal{C}^{\mathsf{T}}\mathbf{R} \tag{6.8}$$

Unlike the previous definition in Equations (5.16) - (5.17), the observability and controllability matrices are now distributional quantities that describe posterior distributions. The next section describes how these posterior distributions can be estimated. Once obtained, this posterior uncertainty can then be propagated through a variety of techniques to obtain estimates of the posterior on the modal properties.

6.1.1 Priors

It is useful to consider the meaning behind the priors on $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and $\mathbf{W}$ with a view to later selecting these in the context of Bayesian SSI. The mean $\boldsymbol{\mu}$ refers to the offset of the lagged versions of each channel. For a stationary signal, the mean of the output $\mathbb{E}[y]$ should be constant as $N \to \infty$. Unlike traditional SSI, which requires the data to first be zero mean, the proposed Bayesian approach is capable of learning this mean.

The block covariance of the data, $\boldsymbol{\Sigma}$, describes ones prior belief about the measurement noise on the different columns of the Hankel matrix. One typically expects this to be small in many engineering settings. Access to this prior also allows the practitioner

to consider known coloured noise or correlated noise between the sensor channels or, in the absence of better information, isotropic white noise which matches that seen in regular (maximum likelihood equivalent) SSI.

Finally, the prior over the columns of the weight matrix describes ones belief about the correlation between the parameters in the projection matrix $W = [\mathbf{W}^{(1)}; \mathbf{W}^{(2)}]$. In this case, setting a diagonal covariance structure over the columns of $\mathbf{W}$ is effectively an L_2 regularisation on the parameters, i.e. it penalises very large values and variance in each column of $\mathbf{W}$. This prior would also imply that our initial belief is that all columns of W have a very similar scale of values. In practice one could use a variety of options for choosing this prior, including a normal zero mean prior with identity covariance or an empirical prior based on the observability and controllability obtained from finite element analysis, models or traditional SSI. The authors would note that while the prior chosen in this paper coincides with that from standard (maximum likelihood) SSI, exploring a more informative prior structure going forward is an interesting avenue of research.

6.1.2 Inference schemes

A key aspect of any Bayesian methodology is the recovery of the posterior distributions over the quantities of interest, in this particular case the modal properties of a structural system. However, as is commonly the case, these distributions are not available in closed form given the form of the model presented above. This intractability of the model arises from inability to compute the integral which defines the normalising constant of the posterior, i.e. the marginal likelihood. Therefore, to solve Bayesian inference problems one must turn to approximate or sampling inference schemes.

Earlier in this thesis, the reader was introduced to two potential solutions to this problem. The first solution is to approximate the posterior distribution using MCMC methods to form an empirical representation of the posterior using samples (see Section 3.3.2). Another solution is to use a variational scheme to approximate the posterior using a surrogate distribution, optimising the parameters of said distribution to minimise the KL divergence between the surrogate and the target posterior (see Section 3.3.5).

Within each of these families, a range of possible approaches exist. For the MCMC problem, a Gibbs sampling approach (see Section 3.3.4) is adopted given direct access

to the full conditionals. For the variational scheme, a coordinate-ascent approach variational Bayes (VB) method is used [173].

Both inference schemes have their own set of advantages and disadvantages. Gibbs sampling has the very useful property of converging to the true posterior in the limit of increasing number of samples. However, the computational expense of sampling techniques can be significant when considering large and complex joint distributions, making it unsuitable for some tasks. Alternatively, variational methods only provide an approximation to the posterior but can demonstrate considerably better computational performance. Hence, both inference approaches are shown in this paper with the results compared.

Gibbs sampling

Adopting the Gibbs sampling methodology introduced in Section 3.3.4, the parameter update equations were derived and are provided in Equations (6.9) - (6.12), with the overall algorithm summarised in Algorithm 3. For the interested reader, the full derivation of these updates is provided in Appendix A.5, with the omission of the sparsity inducing prior present in [199] and with the addition of generic mean and covariance priors. The required Gibbs sampling updates for the parameters of the model described in Figure 6.1, are given below. The updated parameters below are written in block formation following previously defined notation. This allows both sets of parameters to be updated in one step. The superscript $\cdot^{\text{new}}$ refers to the updated sample of each parameter given the updated distribution parameters, where $\cdot^{\text{old}}$ would refer to the current parameters. To improve clarity, $\cdot^{\text{old}}$ is omitted. The updated statistical moments are denoted by $\breve{\cdot}$, with the prior statistical moments denoted without.

$\boldsymbol{\Sigma}$ update

$$\boldsymbol{\Sigma}^{\text{new}} \sim I\mathcal{W}\left(\mathbf{K}_0 + \mathbf{K}\,, \nu_0 + N\right) \tag{6.9}$$

where,

$$\mathbf{K} = \sum_{n=1}^{N} (\mathbf{x}_n - \boldsymbol{\mu} - \mathbf{W}\mathbf{z}_n)(\mathbf{x}_n - \boldsymbol{\mu} - \mathbf{W}\mathbf{z}_n)^{\mathsf{T}}$$

$\boldsymbol{\mu}$ update

$$\boldsymbol{\mu}^{\text{new}} \sim \mathcal{N}\left(\breve{\boldsymbol{\mu}}_{\mu}, \breve{\boldsymbol{\Sigma}}_{\mu}\right) \tag{6.10}$$

where,

$$\breve{\boldsymbol{\Sigma}}_{\mu} = (N\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}_{\mu}^{-1})^{-1}$$

and,

$$\breve{\boldsymbol{\mu}}_{\mu} = \breve{\boldsymbol{\Sigma}}_{\mu}\left(\boldsymbol{\Sigma}^{-1}\sum_{n=1}^{N}(\mathbf{x}_n - \mathbf{W}\mathbf{z}_n) + \boldsymbol{\Sigma}_{\mu}^{-1}\boldsymbol{\mu}_{\mu}\right)$$

$\mathbf{w}_i$ update

$$\mathbf{w}_i^{\text{new}} \sim \mathcal{N}\left(\breve{\boldsymbol{\mu}}_{\mathbf{w}_i}, \breve{\boldsymbol{\Sigma}}_{\mathbf{w}_i}\right) \tag{6.11}$$

where,

$$\breve{\boldsymbol{\Sigma}}_{\mathbf{w}_i} = \left(\sum_{n=1}^{N}\mathbf{z}_{i,n}^2\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}\right)^{-1}$$

and,

$$\breve{\boldsymbol{\mu}}_{\mathbf{w}_i} = \breve{\boldsymbol{\Sigma}}_{\mathbf{w}_i}(\boldsymbol{\Sigma}^{-1}\sum_{n=1}^{N}\mathbf{z}_{i,n}\tilde{\mathbf{x}}_n + \boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}\boldsymbol{\mu}_{\mathbf{w}_i})$$

given,

$$\tilde{\mathbf{x}}_n = \mathbf{x}_n - \boldsymbol{\mu} - \mathbf{w}_{\_i}\mathbf{z}_{\_i,n}$$

such that $\mathbf{w}_{\_i}$ corresponds to all columns except the ith column of interest and $\mathbf{z}_{\_i,n}$ corresponds to all rows of $\mathbf{z}_n$ except the ith row.

z update

$$\mathbf{z}_n^{\text{new}} \sim \mathcal{N}\left(\breve{\boldsymbol{\mu}}_{\mathbf{z}}, \breve{\boldsymbol{\Sigma}}_{\mathbf{z}}\right) \tag{6.12}$$

where,

$$\breve{\boldsymbol{\Sigma}}_{\mathbf{z}} = (\mathbf{W}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{W} + \mathbb{I})^{-1}$$

and,

$$\breve{\boldsymbol{\mu}}_{\mathbf{z}} = \breve{\boldsymbol{\Sigma}}_{\mathbf{z}}\mathbf{W}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{x}_n$$

Algorithm 3 Bayesian CCA - Gibbs Sampling (MCMC)

initialise: $\boldsymbol{\Sigma}, \boldsymbol{\mu}, \mathbf{W}, \mathbf{z}$ using the priors given appropriately chosen hyperparameters
for $\tau = 1 \ldots, T$ **do**
  1. Sample $\boldsymbol{\Sigma}^{\text{new}}$ using Equation (6.9)
  2. Sample $\boldsymbol{\mu}^{\text{new}}$ using Equation (6.10)
  for $i = 1 \ldots, d$ **do**
    3. Sample $\mathbf{w}_i^{\text{new}}$ using Equation (6.11)
  end for
  4. Sample $\mathbf{z}_n^{\text{new}}$ using Equation (6.12)
end for
return Samples of $\boldsymbol{\Sigma}, \boldsymbol{\mu}, \mathbf{W}, \mathbf{z}$

Following application of Algorithm 3 for a desired number of samples, the resulting samples of $\mathbf{W}^{(1)}$ (i.e. the observability matrix) can be propagated to obtain samples of the state transition matrix and subsequently propagated further onto the modal properties (see Section 3.6.4), obtaining posteriors as sampled distributional estimates.

Variational Bayes

Owing to limitations with sampling-based approaches, practitioners often turn to approximating algorithms that enable significant computational advantages over sampling methods. In this case Variational Bayes (VB).

An important step in VB is defining the form of the surrogate posterior. Here the mean-field approximation is chosen (see Equation 3.23). Adopting the mean field approximation, the surrogate posterior for Bayesian CCA is chosen to take the following factorised form,

$$q(\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\Psi}, \mathbf{W}) = \prod_{n=1}^{N} q(\mathbf{z}_n) \prod_{i=1}^{d} q(\mathbf{w}_i) q(\boldsymbol{\mu}) q(\boldsymbol{\Psi}) \tag{6.13}$$

where,

$$q(\mathbf{z}_n) \sim \mathcal{N}\left(\mathbf{z}_n|\boldsymbol{\mu}_{\mathbf{z}}, \boldsymbol{\Sigma}_{\mathbf{z}}\right) \tag{6.14}$$

$$q(\boldsymbol{\Psi}) \sim \mathcal{W}\left(\boldsymbol{\Psi}|\mathbf{K}^{-1}, \nu\right) \tag{6.15}$$

$$q(\mathbf{w}_i) \sim \mathcal{N}\left(\mathbf{w}_i|\boldsymbol{\mu}_{\mathbf{w}_i}, \boldsymbol{\Sigma}_{\mathbf{w}_i}\right) \tag{6.16}$$

$$q(\boldsymbol{\mu}) \sim \mathcal{N}\left(\boldsymbol{\mu}|\boldsymbol{\mu}_{\boldsymbol{\mu}}, \boldsymbol{\Sigma}_{\boldsymbol{\mu}}\right) \tag{6.17}$$

given $\boldsymbol{\mu}_{\mathbf{z}}, \boldsymbol{\Sigma}_{\mathbf{z}}, \mathbf{K}^{-1}, \nu, \boldsymbol{\mu}_{\mathbf{w}_i}, \boldsymbol{\Sigma}_{\mathbf{w}_i}, \boldsymbol{\mu}_{\boldsymbol{\mu}}, \boldsymbol{\Sigma}_{\boldsymbol{\mu}}$ are the statistical moments of the surrogate distributions.

Following a VB approach, the update equations for the model parameters were derived and are given in Equations (6.18)-(6.21). The complete VB algorithm is summarised in Algorithm 4, whilst the full derivation of these updates is also provided in Appendix A.6, with the omission of the sparsity inducing prior present in [198] and with generic mean and covariance priors.

The variational update equations for the parameters of the model described in Figure 6.1, are given below. In all equations, $\langle\cdot\rangle$ represents the expected value $\mathbb{E}[\cdot]$ of that variable or combination of variables with respect to all the other parameters. Here $q^{\star}(\cdot)$ refers to the updated surrogate distribution. The updated statistical moments are denoted by $\breve{\cdot}$, with the prior statistical moments denoted without.

$\mathbf{z}_n$ update

$$q^{\star}(\mathbf{z}_n) \sim \mathcal{N}\left(\breve{\boldsymbol{\mu}}_{\mathbf{z}}, \breve{\boldsymbol{\Sigma}}_{\mathbf{z}}\right) \tag{6.18}$$

where

$$\breve{\boldsymbol{\Sigma}}_{\mathbf{z}} = \left(\langle\mathbf{W}^{\mathsf{T}}\boldsymbol{\Psi}\mathbf{W}\rangle + \mathbb{I}\right)^{-1}$$

and

$$\breve{\boldsymbol{\mu}}_{\mathbf{z}} = \breve{\boldsymbol{\Sigma}}_{\mathbf{z}}\langle\mathbf{W}\rangle^{\mathsf{T}}\langle\boldsymbol{\Psi}\rangle(\mathbf{x}_n - \langle\boldsymbol{\mu}\rangle)$$

$\mathbf{w}_i$ update

$$q^{\star}(\mathbf{w}_i) \sim \mathcal{N}\left(\breve{\boldsymbol{\mu}}_{\mathbf{w}_i}, \breve{\boldsymbol{\Sigma}}_{\mathbf{w}_i}\right) \tag{6.19}$$

where

$$\breve{\boldsymbol{\Sigma}}_{\mathbf{w}_i} = \left(\sum_{n=1}^{N} \langle \mathbf{z}_{i,n} \mathbf{z}_{i,n} \rangle \langle \boldsymbol{\Psi} \rangle + \boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1} \right)^{-1}$$

and

$$\breve{\boldsymbol{\mu}}_{\mathbf{w}_i} = \breve{\boldsymbol{\Sigma}}_{\mathbf{w}_i} \left(\langle \boldsymbol{\Psi} \rangle \sum_{n=1}^{N} \tilde{\mathbf{x}}_n \langle \mathbf{z}_{n,i}^{\mathsf{T}} \rangle + \boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1} \boldsymbol{\mu}_{\mathbf{w}_i} \right)$$

given

$$\tilde{\mathbf{x}}_n = \mathbf{x}_n - \langle \boldsymbol{\mu} \rangle - \langle \mathbf{w}_{\_i} \rangle \langle \mathbf{z}_{\_i,n} \rangle$$

such that $\mathbf{w}_{\_i}$ corresponds to all columns except the ith column of interest and $\mathbf{z}_{\_i,n}$ corresponds to all entries of $\mathbf{z}_n$ except the ith entry.

$\boldsymbol{\Psi}$ update

$$q^{\star}(\boldsymbol{\Psi}) \sim \mathcal{W} \left(\breve{\mathbf{K}}^{-1}, \breve{\nu} \right) \tag{6.20}$$

where[1]

$$\breve{\mathbf{K}} = \mathbf{K}_0 + \sum_{n=1}^{N} \left\langle (\mathbf{x}_n - \boldsymbol{\mu} - \mathbf{W}\mathbf{z}_n)(\mathbf{x}_n - \boldsymbol{\mu} - \mathbf{W}\mathbf{z}_n)^{\mathsf{T}} \right\rangle$$

and

$$\breve{\nu} = \nu_0 + N$$

$\boldsymbol{\mu}$ update

$$q^{\star}(\boldsymbol{\mu}) \sim \mathcal{N} \left(\breve{\boldsymbol{\mu}}_{\boldsymbol{\mu}}, \breve{\boldsymbol{\Sigma}}_{\boldsymbol{\mu}} \right) \tag{6.21}$$

where

$$\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\mu}} = (N \langle \boldsymbol{\Psi} \rangle + \boldsymbol{\Sigma}_{\mu}^{-1})^{-1}$$

and

$$\breve{\boldsymbol{\mu}}_{\boldsymbol{\mu}} = \breve{\boldsymbol{\Sigma}}_{\mu} \left(\langle \boldsymbol{\Psi} \rangle \sum_{n=1}^{N} (\mathbf{x}_n - \langle \mathbf{W} \rangle \langle \mathbf{z}_n \rangle) + \boldsymbol{\Sigma}_{\mu}^{-1} \boldsymbol{\mu}_{\mu} \right)$$

[1] Note the need to expand this expectation in implementation, including all the necessary terms using the total law of variance.

Algorithm 4 Bayesian CCA - Coordinate Ascent Variational Inference (CAVI)

initialise: Variational factors $q(\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\Psi}, \mathbf{W})$
while the ELBO has not converged **do**
  Update local variational parameters:
  1. Update moments of $q^\star(\mathbf{z}_n)$ using Equation (6.18)

  Update global variational parameters:
  for $i = 1, \ldots, d$ **do**
    2. Update moments of $q^\star(\mathbf{w}_i)$ using Equation (6.19)
  end for
  3. Update moments of $q^\star(\boldsymbol{\Psi})$ using Equation (6.20)
  4. Update moments of $q^\star(\boldsymbol{\mu})$ using Equation (6.21)

  Compute the ELBO for convergence:
  5. Compute $\text{ELBO}(q) = \mathbb{E}\left[\ln p(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta})\right] + \mathbb{E}\left[\ln q(\mathbf{z}, \boldsymbol{\theta})\right]$
end while
return Surrogate posterior distributions of $\boldsymbol{\Sigma}, \boldsymbol{\mu}, \mathbf{W}, \mathbf{z}$

6.1.3 Overall methodology

Having explored two different forms of inference for the recovery of the posteriors in Bayesian CCA, it is perhaps useful to summarise the general procedure in the context of Bayesian SSI. The following summary gives an overview of the steps needed to conduct Bayesian SSI to a desired dataset.

Given response data, $\mathbf{y} \in \mathbb{R}^{l \times N'}$, with l sensors and data length $N' = N + 2j - 1$:

Step 1. Construct block Hankel matrices of the future, $\mathbf{Y}_f \in \mathbb{R}^{lj \times N}$, and the past, $\mathbf{Y}_p \in \mathbb{R}^{lj \times N}$, recalling that $2j$ is the number of block row lags in the Hankel matrix and l is the number of output channels (the reader is referred back to Section 3.7.4).

Step 2. Apply **Algorithm 1** or **Algorithm 2** with the inputs $\mathbf{x}^{(1)} = \mathbf{Y}_f$ and $\mathbf{x}^{(2)} = \mathbf{Y}_p$, given appropriately chosen priors and convergence criteria[2] to obtain the posteriors over the model parameters.

Step 3. If **Algorithm 1** has been applied, remove a proportion of the initial samples as burn in and continue to Step 4. If **Algorithm 2** has been applied, draw Monte Carlo samples from the posterior distributions over the columns of the weight matrix and reconstruct the full weight matrix to produce a sample of the observability, i.e. $\mathbf{W}^{(1)}$.

Step 4. Using the samples of the observability matrix, calculate the state transition matrix $\mathbf{A}$ and output matrix $\mathbf{C}$ (see Section 3.6), and the modal properties in the usual way for state space models (see Section 3.6.4), for each sample. Considering all the samples, this provides an approximate posterior distribution over each modal property[3].

[2]The convergence criteria, or desired number of samples, in both algorithms must be selected. The choice and specification of these criteria constitutes its own research field and is therefore not explored in depth.

[3]Note that this is not the only way to approximate the posterior uncertainty over the modal properties. There are a range of techniques for approximating the uncertainty, often assuming a Gaussian form.

6.2 Simulated case study

6.2.1 Benchmarking the proposed algorithm

The performance of the proposed Bayesian SSI algorithm, outlined in Section 6.1.3, is benchmarked here by application to simulated data obtained using a numerical model of the four degree-of-freedom shear frame illustrated in Figure 6.2. This structure is identical to that described by Reynders in [125], although a new simulation is used. Using a mass-spring-damper model to represent the dynamics of the system, the mass m_a assigned to each floor is 2 kg and the stiffness k_a applied to the individual columns (springs) on each floor is 2500 N/m, where $a = 1, 2, 3, 4$. The damping in each column is proportional to the associated stiffness, such that $c_a = k_a/1000$. Horizontal forces are assumed to apply to each floor as the inputs $u_a(t)$, whilst the horizontal acceleration of each floor are the corresponding outputs $y_a(t)$.

The equations of motion of the shear frame in continuous time can be defined as the ordinary differential equation

$$\mathbf{M}\ddot{\mathbf{x}} + \mathbf{C}\dot{\mathbf{x}} + \mathbf{K}\mathbf{x} = \mathbf{f} \tag{6.22}$$

In this example, the forcing $\mathbf{f}$ is assumed to be represented by a white noise process in continuous time acting on each floor, formally the derivative of a Brownian motion. This definition results in a linear stochastic differential equation which can be discretised exactly using Van Loan discretisation [202] The forcing white noise process is chosen with a spectral density of 5×10^{-5} (m s^{-2})^2 Hz^{-1} whilst some Gaussian measurement noise is added to each channel output $y_a(t)$, sampling from a zero-mean normal distribution with a standard deviation of 0.05 m s^{-2}. Data are then simulated using a sampling frequency of 50 Hz with length $N = 2^{16}$.

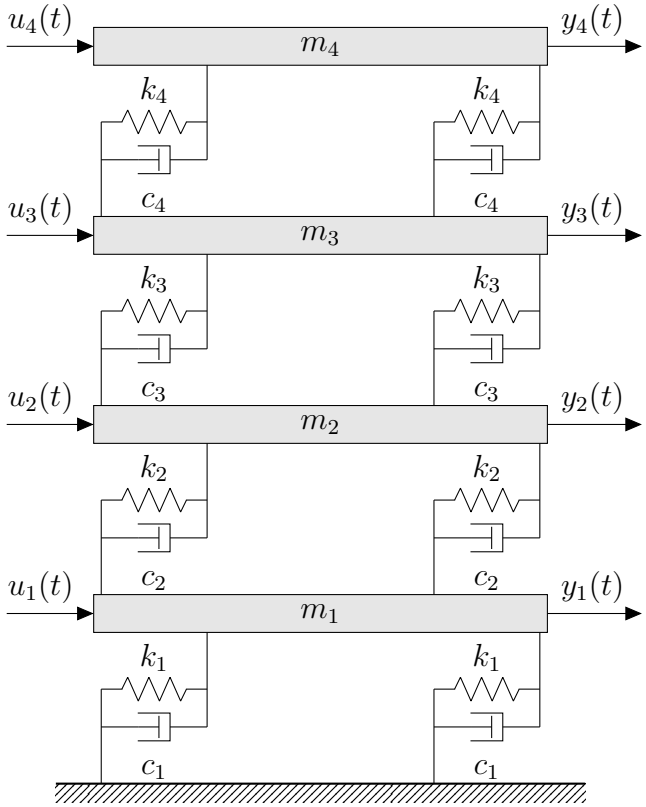


Figure 6.2: Four-story shear building model used to simulate responses given a white noise input on each floor

Based on the model described in Figure 6.2, the matrix coefficients of Equation (6.22) are,

$$\mathbf{M} = \begin{bmatrix} m_1 & 0 & 0 & 0 \\ 0 & m_2 & 0 & 0 \\ 0 & 0 & m_3 & 0 \\ 0 & 0 & 0 & m_4 \end{bmatrix}, \quad \mathbf{K} = 2 \begin{bmatrix} k_1 + k_2 & -k_2 & 0 & 0 \\ -k_2 & k_2 + k_3 & -k_3 & 0 \\ 0 & -k_3 & k_3 + k_4 & -k_4 \\ 0 & 0 & -k_4 & k_4 \end{bmatrix}, \quad \mathbf{C} = \mathbf{K}/1000$$

Priors

Before performing inference, attention must first be paid towards establishing sensible priors for the model. Given the chosen structural model presented in Section 6.2.1, the following prior structure was chosen

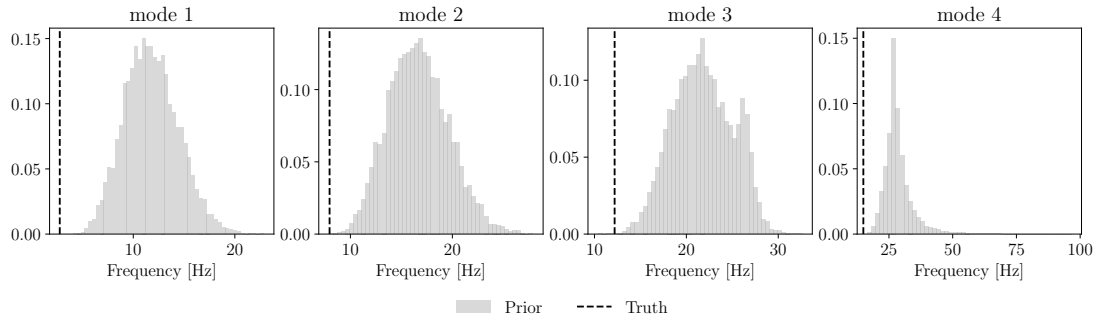
$$\mathbf{w}_i^{(m)} \sim \mathcal{N}(\mathbf{0}, \sigma_{\mathbf{w}_i}\mathbb{I}) \tag{6.23}$$

$$\boldsymbol{\mu}^{(m)} \sim \mathcal{N}(\mathbf{0}, \sigma_{\boldsymbol{\mu}}\mathbb{I}) \tag{6.24}$$

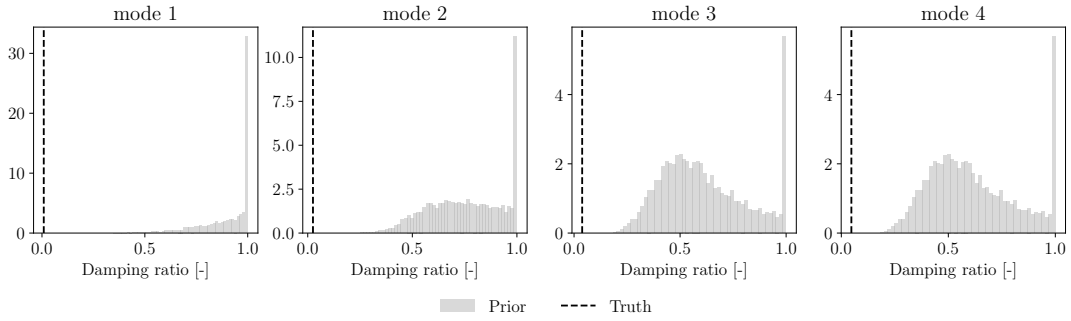
$$\boldsymbol{\Sigma}^{(m)} \sim I\mathcal{W}(\mathbf{K}_0, \nu_0) \tag{6.25}$$

where the independent columns $\mathbf{w}_i$ share the same prior definition. Although it has been shown in the literature that the columns of the observability matrix (the singular vectors) are not strictly independent [203] a posteriori, independent priors on the columns does not necessarily restrict the columns of the posterior to be independent. The remaining hyperparameters, $\boldsymbol{\sigma}_\mu = \sigma_{\mathbf{w}_i} = 1$, $\mathbf{K}_0 = 1 \times 10^2$, $\nu_0 = D + 2$, were chosen to be weakly informative and proper, to provide sufficient flexibility to the model. Samples from the chosen prior over the weight matrices were then used to generate samples of the observability matrix and subsequently propagated through an eigenvalue decomposition to obtain estimates for the posterior distributions over the modal properties. The prior distributions over the modal properties are presented in Figure 6.3.

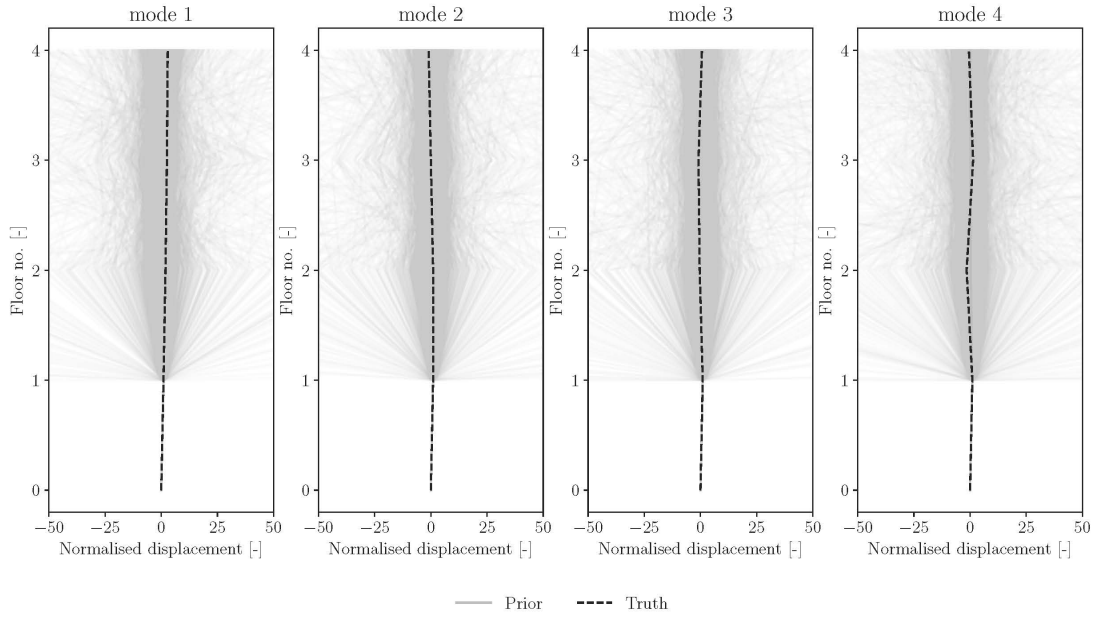
It is useful to briefly discuss the general shape of the priors on the modal parameters. Despite Gaussian assumptions in the model, specifically on the weight prior, the propagated uncertainty results in distributions that are a mix of what appear to be non-Gaussian and even multimodal in nature. The authors believe this is likely caused by the size of the prior variance coupled with the non-linear transformation in the eigenvalue decomposition, which is most evident when considering larger variances.



(a)



(b)



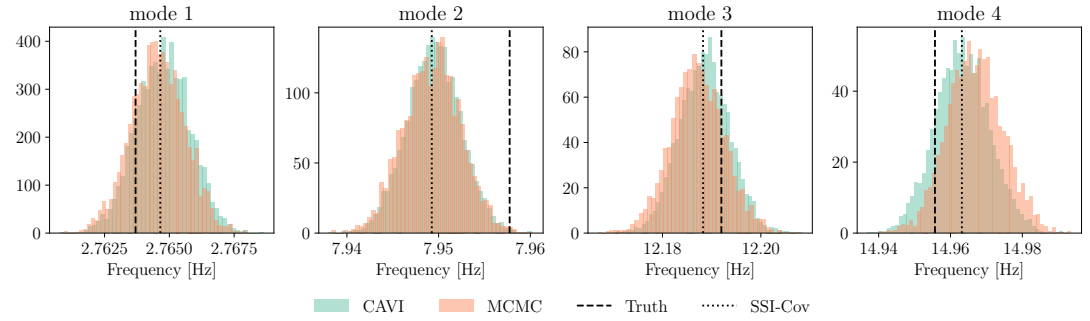
(c)

Figure 6.3: Prior distributions over the natural frequencies (a), damping ratios (b) and mode shapes (c), obtained using propagated Monte Carlo samples of the priors as defined by Equations 6.23 - 6.4.

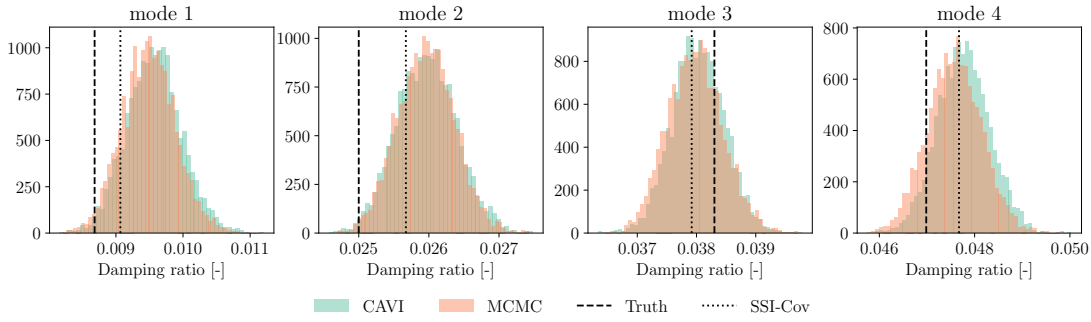
Posteriors

Assuming the correct number of modes (four modes – corresponding to a state space dimension of eight) and 40 time lags in the construction of the Hankel matrix, the Gibbs sampling implementation of Bayesian SSI (Algorithm 3) was used to recover posterior distributions over the modal properties of the simulated structure. 5000 samples were drawn, with the first 20% removed as 'burn in' to remove transients in the Markov chain [173]. The posteriors over the modal properties are shown in Figure 6.4. Similar to the Gibbs sampling case, assuming the correct number of modes, data were analysed using the VI implementation of Bayesian SSI (Algorithm 4). After the recovery of the closed form posterior over the observability, 4000 samples were drawn and propagated. The resulting posteriors over the modal properties are also presented in Figure 6.4.

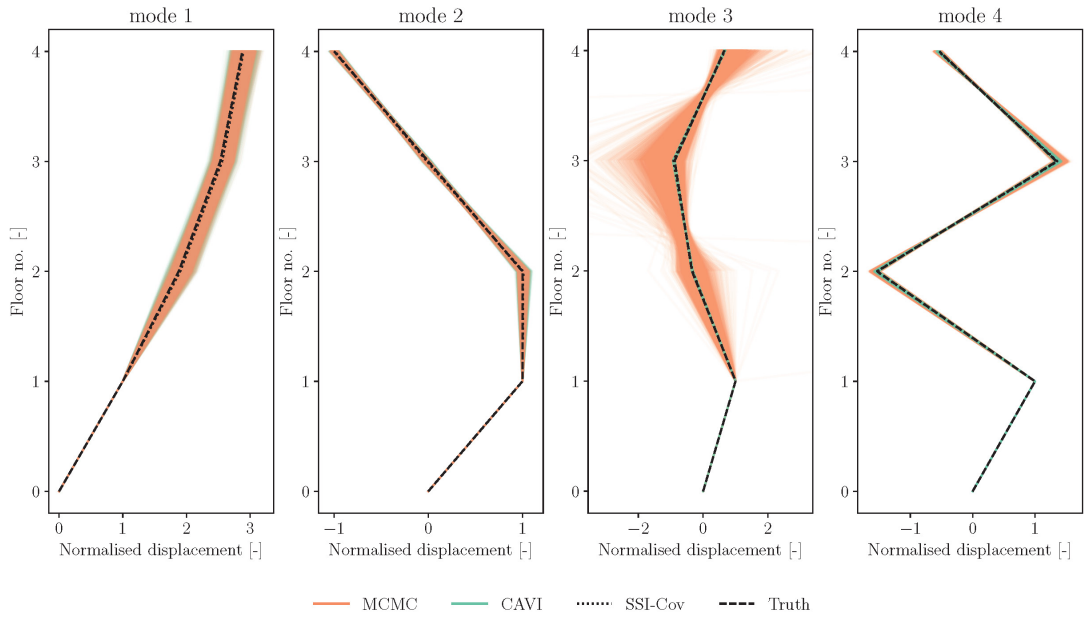
On interpreting Figure 6.4, one can conclude that the surrogate posteriors over the modal parameters obtained using VB closely align with the true posteriors obtained through Gibbs sampling. This closeness is evident from the aligned expected values and the shape of the posteriors, which correspond well. Furthermore, the posterior estimates of both schemes converge toward the SSI-Cov result. This convergence is largely expected as, given the use of weakly informative priors and enough data, the maximum-a-posteriori (MAP) will be close to the MLE, i.e. SSI-Cov. Some minor differences can be seen. The first is in the mean estimates of first and second damping ratios compared to the SSI-Cov result. This misalignment is a consequence of the priors, which cause slight bias in the posterior, more so than the damping ratio estimates for the third and fourth modes. The second difference can be seen in the variance of the third mode shape in Figure 6.4c. A non-Gaussian posterior recovered from the Gibbs scheme explains this difference. This phenomenon originates from application of the eigenvalue decomposition to obtain the mode shapes. However, this difference is more likely a result of an underprediction of the variance using VI. Under or overprediction of the variance is a common characteristic of VI schemes which can occur when some conditional distributions are factorised out of the model during its construction [173]. The omission of these conditionals is a result of the independence assumptions in the surrogate posterior definition. Possible misalignment of the variance is one of several trade-offs that needs to be considered when prioritising the computational efficiency of variational methods over convergence to the true posterior guaranteed by MCMC sampling.



(a)



(b)

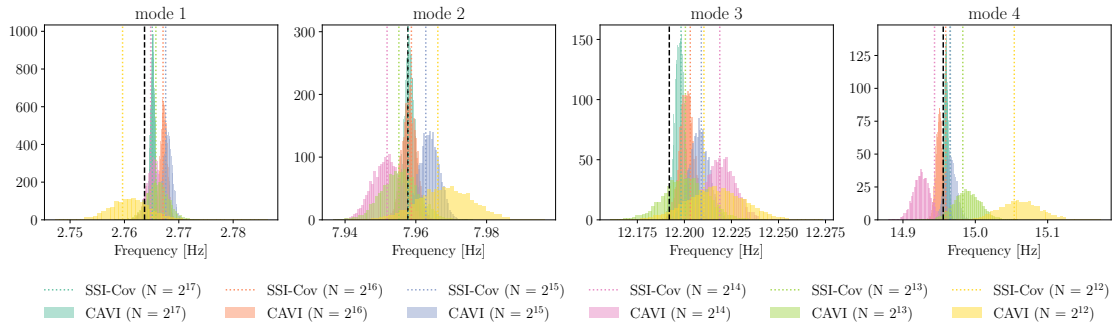


(c)

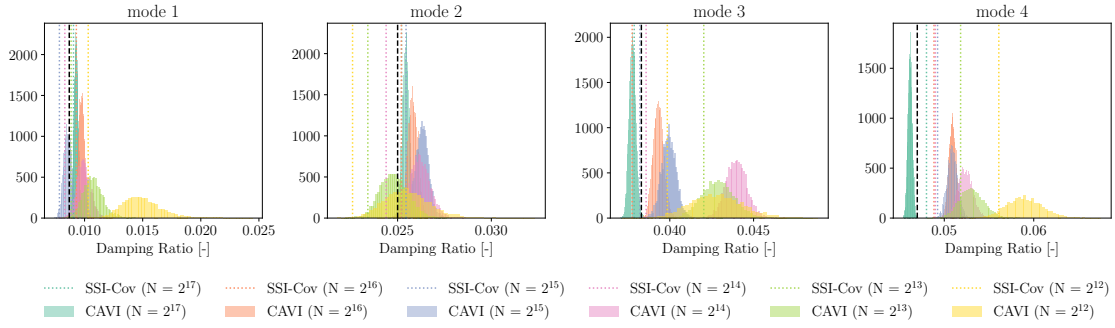
Figure 6.4: Identified posterior distributions over the recovered natural frequencies (a), damping ratios (b) and mode shapes (c), obtained using the Gibbs sampling and VB implementations of Bayesian SSI to response data simulated from a 4DOF linear dynamic system (see Figure 6.23).

6.2.2 Influence of data length on variance

The influence of data length on the variance of the modal posteriors is also investigated. Using the same system parameters and priors described in the Section 6.2.1, the data length was varied from $2^{17} - 2^{12}$ in decreasing powers of two, with the range selected solely for demonstrative purposes. The results for the natural frequencies and the damping ratios are displayed in Figures 6.5a and 6.5b, respectively. As is perhaps expected, the variance of the posterior estimates of both frequency and damping ratio reduce given increasing amounts of data, meanwhile the means of the distributions also converge towards the true values from the model. The influence of data length and perhaps other factors (lags in the Hankel matrix) on the recovered uncertainty has interesting implications and could be used to reduce the amount of data collection and storage required in OMA, for a desired level of confidence. This poses an interesting question on choosing the right amount of data for a given decision-making task, which will be considered in future work.



(a)



(b)

Figure 6.5: Identified posterior distributions over the natural frequencies (a) and damping ratio (b) estimates for each mode, obtained using Bayesian SSI with varying data length N. The true values are represented by **- -**

6.2.3 Posterior variance vs sample variance

In this section, the posterior variance over the natural frequencies recovered using Bayesian SSI on a single dataset is compared to the variance from multiple SSI estimates across the same data. Multiple subsets of the larger dataset in the form of a sliding window, were analysed with SSI to provide some indication as to the variance in the data. Using an original dataset of length $N = 2^{14}$, 5000 smaller datasets were generated using a sliding window, moving one timestep at a time, such that each smaller dataset were $N = 2^{14} - 5000$ in length. The number of lags was chosen to be 40 with all other hyperparameters reflecting those shown in Section 6.2.1. Figure 6.6 shows the Bayesian posterior from Bayesian SSI, and the sample distribution of the 5000 SSI tests, over the normalised natural frequencies for all four modes.

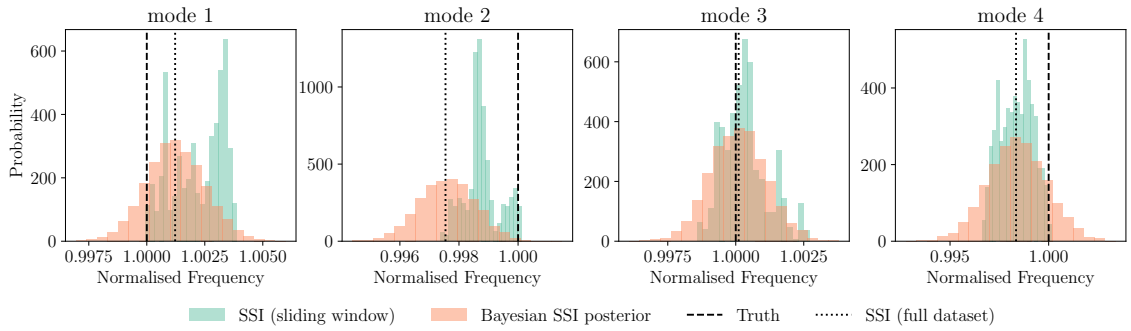


Figure 6.6: Plot of the posterior distribution obtained using Bayesian SSI on a single dataset, presented as a histogram of samples and the SSI result for the same dataset. Also plotted is the sample distribution of results obtained by applying SSI to 5000 subsets of the data, obtained by applying a one timestep sliding window. The resulting distribution is also represented by a histogram of the samples.

The results shown in Figure 6.6 suggest that the variance found through multiple SSI analyses is broadly similar to that of the Bayesian posterior although not identical, displaying slightly inflated variances. This is largely expected due to the fundamentally different assumptions about the uncertainty. Furthermore, the mean of the posteriors align with the SSI estimates obtained from the full available data.

6.2.4 Variability study

In this section, the posterior means of the natural frequencies recovered using Bayesian SSI are compared to estimates recovered using SSI for multiple data generated with different realisations of the noise to study the variability of each algorithm. Each dataset was chosen to have length $N = 2^{14}$ and similarly, the number of lags was chosen to be 40 with all other hyperparameters reflecting those in Section 6.2.1, with the exception of $\sigma^2_{\mathbf{w}_i} = 10^2$. Figure 6.7 shows the posterior means obtained on 500 datasets using Bayesian SSI, and the equivalent estimates from SSI, for the normalised natural frequencies at all four modes.

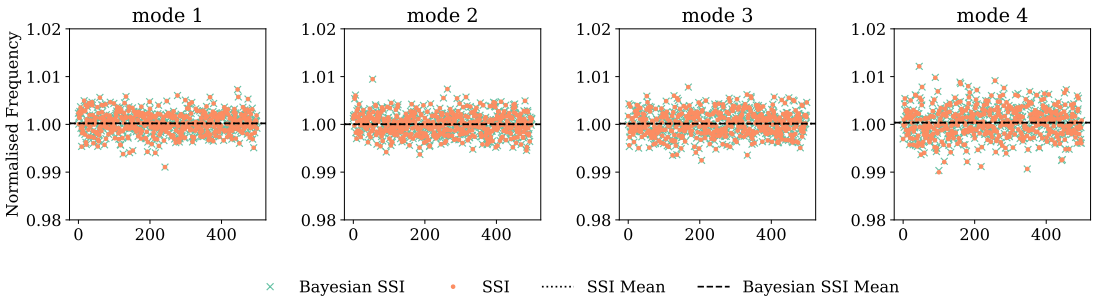


Figure 6.7: Scatter plot of the posterior means obtained using Bayesian SSI, and the corresponding SSI estimates, for 500 datasets with different realisations of the noise. Also plotted are the sample means from both cases.

Inspecting the results in Figure 6.7, it appears that the performance of Bayesian SSI (under weakly informative priors) is very close to that of classic SSI under sufficient data, with the mean from multiple realisations of the noise averaging the truth, and with comparable sample variances in each case. In the case of low data regimes or lower prior variance, however, one may expect the mean of the estimates to align closer to the prior, as is the case in Bayesian methods.

6.3 Case study: Z24 bridge

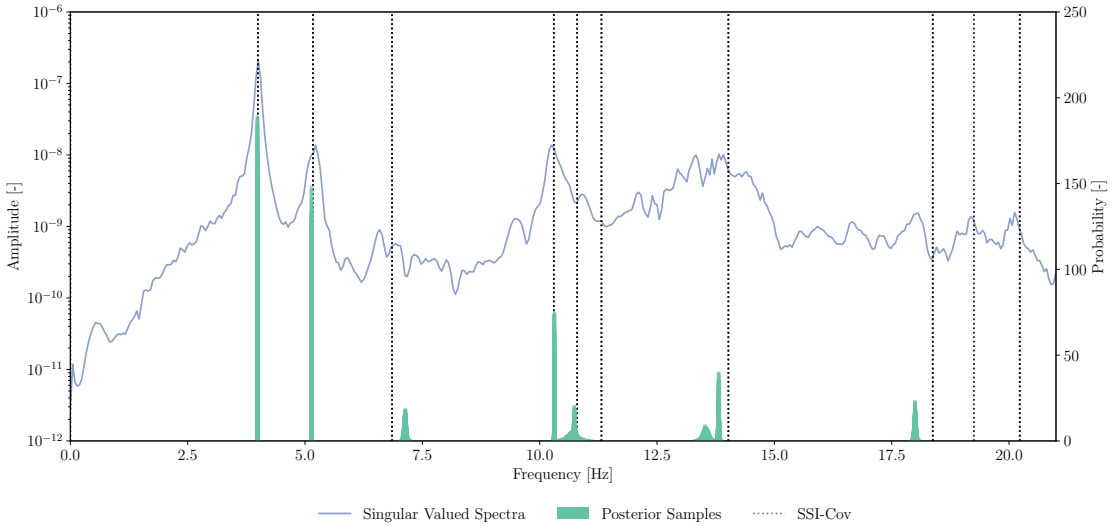
To demonstrate applicability to real world systems, the proposed Bayesian SSI algorithm was used to analyse vibration data obtained from the Z24 bridge, first introduced in Chapter 5. The Z24-Bridge has become a standard and familiar benchmark when demonstrating vibration-based damage diagnosis methods, new system identification techniques, and increasingly for UQ tasks in structural dynamics. This familiarity also makes it a sensible choice for this study. As noted in Chapter 5, more information on the Z24 dataset can be found at the following resources (bwk.kuleuven.be/bwm/z24) [195, 204].

This purpose of this case study was to evaluate the behaviour of the new Bayesian SSI algorithm when confronted with measured data. Analysis is only conducted using the variational approach, since the high computation time of Gibbs sampler was deemed impractical. Results at two different single model orders are shown, followed by a stabilisation diagram to assess the convergence of the modal properties at different model orders.

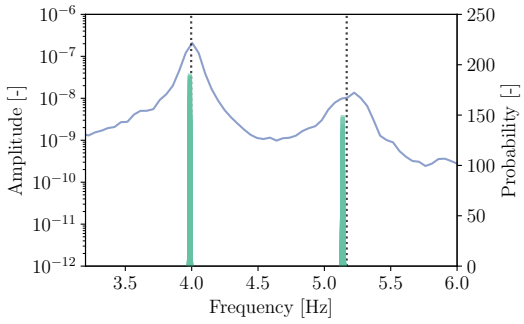
Data, corresponding to acceleration measurements obtained on the 24th December 1997, were used for processing. This dataset comes from the long-term continuous monitoring test set, and in this case comprises data from seven accelerometers corresponding to one side of the Z24 bridge. The first segment of the data (8192 data points at $f_s = 100$ Hz) was formulated into a Hankel matrix with 40 lags (corresponding to 20 past lags and 20 future lags). The priors of the Bayesian SSI model were left unchanged from those used in the previous numerical simulation (see Section 6.2.1) except for the precision hyperparameter K_0, which was changed to the identity providing some additional flexibility to the model.

6.3.1 Single model order

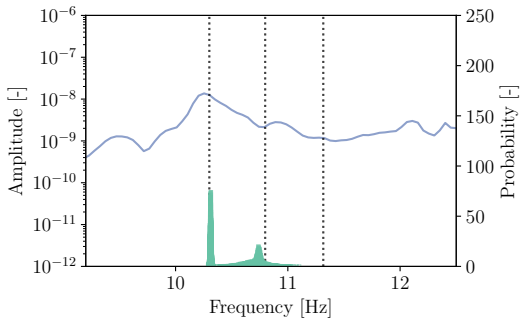
Conducting analysis at model orders of 10 and 30, chosen solely for demonstrative purposes, the posterior distributions of the modal characteristics were obtained using the variational implementation of Bayesian SSI. The posteriors over the natural frequencies, represented by histograms of the samples and overlaid over the sum of the singular valued spectrum spectra from each channel, are shown in Figures 6.8 and 6.9.



(a)



(b)

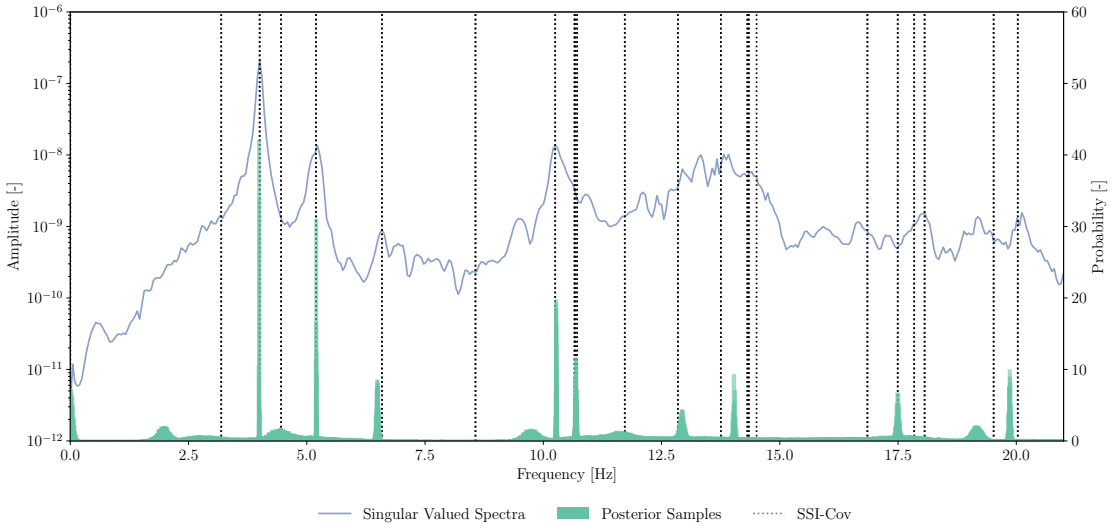


(c)

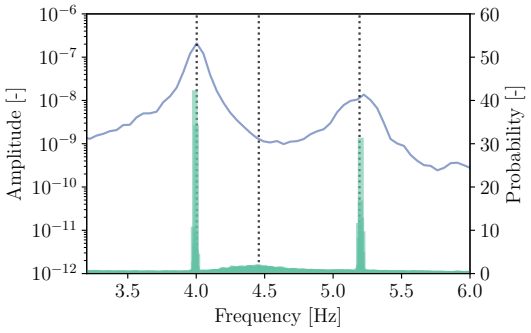
Figure 6.8: The first singular spectrum of the Z24 data overlaid with the posteriors of the natural frequencies, represented by histograms of the natural frequency samples recovered using Bayesian SSI at a model order of 10 is given in (a). Regions of interest are shown in (b) and (c), with axes limits chosen to highlight the differences in the mean and variance of the histograms in different regions.

It is evident from both Figure 6.8a and Figure 6.9a that histograms with means centred around the apparent natural frequencies (peaks in the singular spectrum) demonstrate significantly lower variance than histograms corresponding to frequencies likely more spurious in nature. This lower variance can be seen more clearly in the enhanced Figures 6.8b, 6.8c, 6.9b and 6.9c and is most noticeable around 4 Hz, 5 Hz, 10-11 Hz.

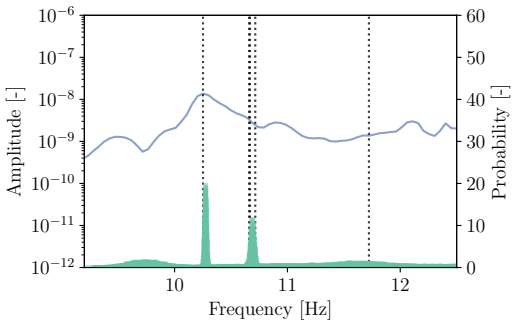
In contrast, larger variances are observed in regions where there is typically less



(a)



(b)



(c)

Figure 6.9: The first singular spectrum of the Z24 data overlaid with the posteriors of the natural frequencies, represented by histograms of the natural frequency samples recovered by Bayesian SSI at a model order of 30, is given in (a). Regions of interest are shown in (b) and (c), with axes limits chosen to highlight the differences in the mean and variance of the histograms in different regions.

evidence of a natural frequency. Focusing specifically on the case where the model order is 30, larger variances can be seen at 4.5 Hz, 9.75 Hz and 11.75 Hz. This obvious variability in the variance for each identified frequency suggests that there is a higher probability or belief in some frequencies to describe the data over others.

6.3.2 Stabilisation diagram

As the true number of modes is unknown, it is standard practice in OMA to construct a stabilisation diagram (see Section 3.7.5). Stabilisation diagrams are a tool used by the practitioner to determine the "physical" poles of a system, helping to decide the nature of the identified complex eigenvalues i.e. spurious or real in a physical context. The decision of which is often governed by some heuristic criteria on the stability or consistency of the modal parameters across a range of model orders.

After conducting Bayesian SSI at multiple model orders, the resulting stabilisation diagram for the Z24 is shown in Figure 6.10. The samples are plotted with transparency such that darker regions indicate areas with a higher density of samples. A second stabilisation diagram is also provided in Figure 6.11, using kernel density estimation (KDE) fits over the samples to visualise the uncertainty. This provides a plot similar to the histogram representations of the posteriors in Figures 6.8 and 6.9, but at multiple model orders. The inclusion of uncertainty in stabilisation diagrams is not an entirely new concept, having already been represented, and proven beneficial in assisting model order selection [17], although no agreed form has yet been defined. In this plot, non-conjugate poles have been removed, however no stability metrics have been defined, purely to observe the raw form of the stabilisation diagram.

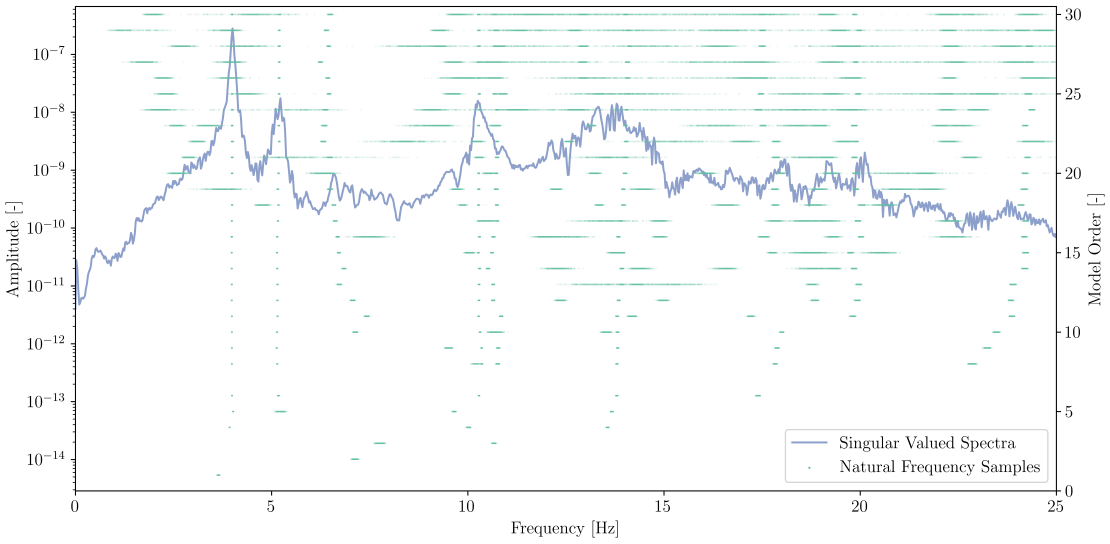


Figure 6.10: Stabilisation diagram of the Z24 bridge, constructed using the natural frequency samples obtained using Bayesian SSI at a range of model orders. Also shown is the first singular spectrum.

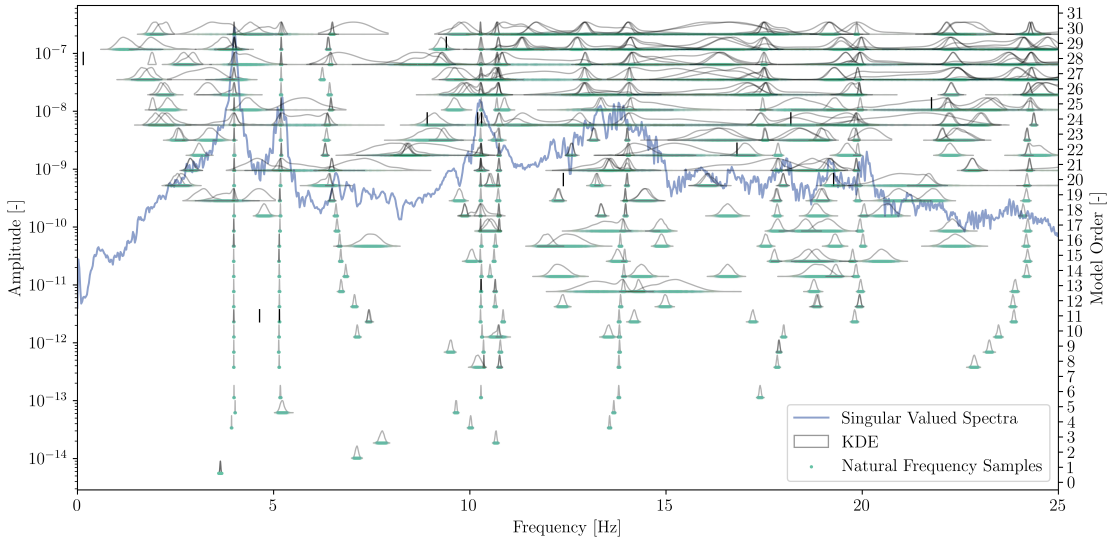


Figure 6.11: Stabilisation diagram of the Z24 bridge, constructed using the natural frequency samples obtained using Bayesian SSI at a range of model orders. The kernel density estimation (KDE) of the samples is shown, where each KDE is fitted using samples for that model order and order of identified mode.

Considering the distributions of the posteriors in Figure 6.10, one can clearly see that peaks in the singular spectrum, typically indicative of a natural frequency, coincide with the posterior means and are consistent across multiple model orders. This is true for multiple peaks, most distinctly around 4 Hz, 5 Hz and 10-11 Hz. The variance of the distributions in these areas is also low across a broad range of model orders, especially when compared to more spurious frequency estimates away from natural frequencies. Thus indicating a higher level of confidence in the ability of certain frequencies to describe the dynamics. Another conspicuous observation is the general increase in variance at higher model orders, particularly in the more spurious estimates, as moving to higher model orders grant greater flexibility with numerous ways for the model to explain the data.

6.4 Summary

In this chapter, it has been demonstrated how the probabilistic SSI algorithm introduced in Chapter 4 can be transformed from the classical maximum likelihood approach to one of Bayesian inference through the inclusion of an arbitrary prior structure; principally forming Bayesian CCA. Specifically, it has been shown how the inclusion of priors over the parameters of the latent variable model in the context of SSI enables the posterior over the observability matrix, and by extension the modal properties, to be obtained.

This new approach was presented with two possible means of performing Bayesian inference: Markov chain Monte Carlo (MCMC) and variational Bayes (VB). The two solutions for Bayesian SSI were then benchmarked using simulated data from a MDOF linear dynamic system. The prior distributions over the modal parameters were shown to be proper and weakly informative, but with sufficient flexibility to the model. Then followed the recovered posteriors, which demonstrated good agreement with each another and convergence of the posterior mean towards the conventional SSI result. The effects of data length on the recovered posteriors was also explored. It was shown that, in the case of increasing data length, the posterior mean trends toward the truth whilst the variance of the posterior decreases. A result largely expected given the theory of SSI and Bayes' rule. Finally, the practicality and applicability of the algorithm was then illustrated using data obtained from an in-service structure, the Z24 bridge. Recovery of the modal parameters and associated posterior uncertainty was then shown at two different model orders, with poles likely referring to physical natural frequencies demonstrating lower variance than that of spurious estimates. Recovered modal information was then also analysed across multiple model orders, forming a stabilisation diagram. The results of this diagram showed that posteriors with means centred around the apparent natural frequencies display much lower variance across all model orders when compared to more spurious frequency estimates.

Access to the posterior uncertainty over the modal parameters via this algorithm has significant benefits. The availability of this uncertainty information means it can be propagated further to assist other important research challenges in OMA, such as model order selection, automatic OMA and sensor placement. In SHM frameworks, one could foresee the inclusion of modal posterior uncertainties as a valuable addition

to damage assessment and decision-making; providing new information one could envisage altering decision boundaries and classification models.

There are some potential limitations to this approach. Despite a successful implementation using variational Bayes (VB), there are many more computations required than the traditional SVD-based SSI, lending itself to higher computational requirements. This is important consideration for practitioners and could effect its adoption into the wider field of OMA. Nevertheless, there are techniques one can use to address this problem, with one possible solution introduced in the following Chapter.

7

Efficient Bayesian Stochastic Subspace Identification

Despite their many advantages, there is often a reluctance to adopt Bayesian methodologies[1], with one factor being their higher computational requirements particularly for high dimensional data. In the case of the Bayesian SSI algorithm introduced in Chapter 6, this problem is even more relevant given the comparatively higher speed of the singular value decomposition (SVD) in the classic SSI algorithm (in practice) when compared to that of Gibbs sampling or variational inference (VI). Computational efficiency is therefore important when considering the application of Bayesian SSI as a suitable alternative to traditional techniques and as a competitive choice for uncertainty quantification. This Chapter presents one possible solution to this problem — a batch stochastic variational inference [28] (SVI) form of CCA. The derivations and preliminary results shown here present results from the conference paper entitled "On improving the efficiency of Bayesian stochastic subspace identification" [206] written by the author.

[1]The author recommends reading the April fools blog-turned-article: 'Objections to Bayesian Statistics' by Andrew Gelman [205]

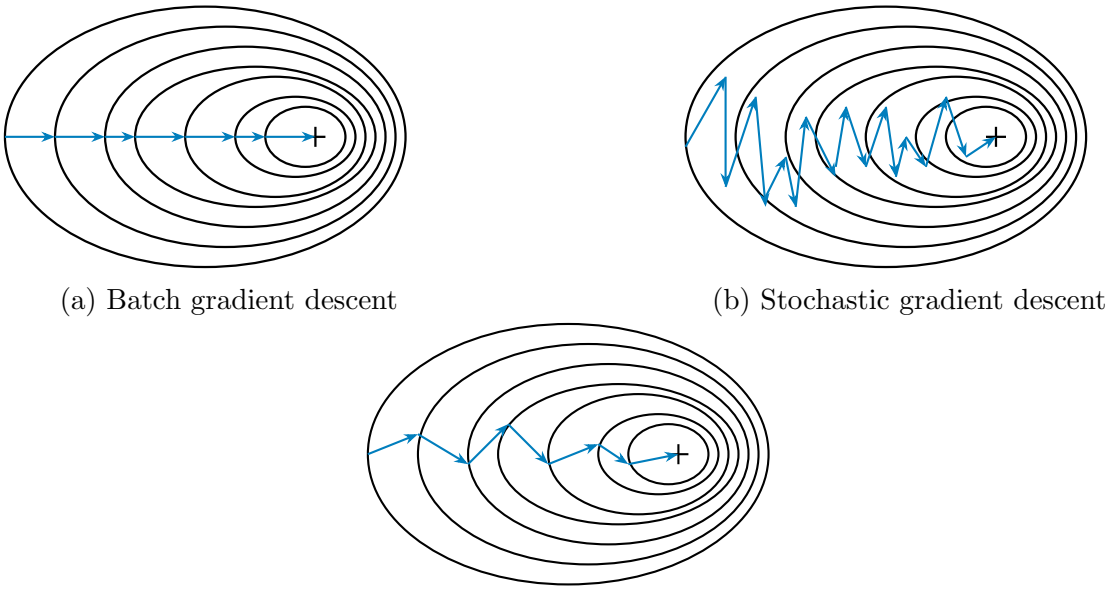
7.1 Introduction

A recurring problem in machine learning and optimisation tasks is that large amounts of data are necessary for good generalisation of models, but large datasets are more computationally expensive, in memory or in number of mathematical operations [207]. In gradient-based optimisation methods for example, such as *gradient descent* (GD), the entire dataset must evaluated in each step before the parameters of a model are updated. For very large datasets, the associated computation time of such processes can make certain methods unfeasible for their intended application.

A very popular approach to addressing these computational challenges is to process data in *batches*. Batch or *mini-batch* computation works by partitioning a larger dataset into a series of smaller batches, comprised of uniformly drawn samples. Parameters are then updated after processing each individual batch, rather than after seeing all the data. Such techniques rely on the assumption that traditional methods can be considered an expectation. Therefore, this expectation can be approximately estimated using a smaller set of samples in a stochastic way [207]. These techniques are very popular in deep learning applications. The mechanism of mini-batch processing can be visualised in a convenient way when thinking about GD.

There are three types of GD: (1) batch GD; (2) stochastic GD; and (3) mini-batch GD. In batch GD, there is only a single batch, which comprises of all the data. This is equivalent to the traditional GD algorithm in that all the data is analysed before an update can be made. An example of the convergence of batch GD is shown in Figure 7.1a. In contrast, a single batch in stochastic GD consists of just a single datapoint. An example convergence path for stochastic GD is shown in Figure 7.1b. Stochastic GD is much quicker than batch GD, but provides a rougher estimate of the gradient.

In mini-batch GD, a single batch contains multiple (uniformly drawn) samples of the data, which are then used to update the parameters. An example convergence path for mini-batch GD is shown in Figure 7.1c. Mini-batch GD is less direct in achieving the desired optimisation than batch GD, but retains the computational advantages of a more stochastic approach.



(a) Batch gradient descent

(b) Stochastic gradient descent

(c) Mini-batch gradient descent

Figure 7.1: Example convergence paths for different batch forms of gradient descent.

In the context of Bayesian SSI and co-ordinate ascent variational inference (CAVI), a very similar mini-batch processing methodology can be applied to reduce the computational limitations of Bayesian methodologies. Before formally introducing this method, it is useful to briefly recall aspects of variational Bayes (VB) from Section 3.3.5 and the VB approach to Bayesian SSI introduced in Chapter 6, with a continuation of the notation into this chapter.

7.2 Recap: Bayesian SSI

Recall that the task of VB is to approximate intractable posterior distributions with surrogate posterior distribution. This is achieved by minimising the Kullback-Leibler (KL) divergence from the surrogate to the true posterior distribution: Maximising the evidence lower bound (ELBO). The ELBO ($\mathcal{L}$) is defined as the sum of the expected log of the joint and the entropy of the variational distribution,

$$\mathcal{L}(\boldsymbol{\phi}, \lambda) = \sum_{n=1}^{N} \mathbb{E}_{q(\mathbf{z}_n, \boldsymbol{\theta})} \left[\log p(\mathbf{x}_n, \mathbf{z}_n, \boldsymbol{\theta}) - \log q_{\boldsymbol{\phi}, \lambda}(\mathbf{z}_n, \boldsymbol{\theta}) \right] \tag{7.1}$$

where ϕ are known as the *local* variational parameters (parameters that describe the

distribution over $\mathbf{z}$), λ are *global* variational parameters (parameters that describe the distributions over $\boldsymbol{\theta}$) and N is the number of observations.

Using the *mean-field* approximation, in which each latent variable is independent and governed by their own set of parameters, the surrogate posteriors take the following factorised form[2], first introduced in Equation 6.13:

$$q(\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\Psi}, \mathbf{W}) = \prod_{n=1}^{N} q(\mathbf{z}_n) \prod_{i=1}^{d} q(\mathbf{w}_i) q(\boldsymbol{\mu}) q(\boldsymbol{\Psi})$$

with the following chosen surrogate posteriors

$$q(\mathbf{z}_n) = \mathcal{N}\left(\mathbf{z}_n | \breve{\boldsymbol{\mu}}_{\mathbf{z}_n}, \breve{\boldsymbol{\Sigma}}_{\mathbf{z}_n}\right) \tag{7.2}$$

$$q(\boldsymbol{\Psi}) = \mathcal{W}\left(\boldsymbol{\Psi} | \breve{\mathbf{K}}, \breve{\boldsymbol{\nu}}\right) \tag{7.3}$$

$$q(\mathbf{w}_i) = \mathcal{N}\left(\mathbf{w}_i | \breve{\boldsymbol{\mu}}_{\mathbf{w}_i}, \breve{\boldsymbol{\Sigma}}_{\mathbf{w}_i}\right) \tag{7.4}$$

$$q(\boldsymbol{\mu}) = \mathcal{N}\left(\boldsymbol{\mu} | \breve{\boldsymbol{\mu}}_{\boldsymbol{\mu}}, \breve{\boldsymbol{\Sigma}}_{\boldsymbol{\mu}}\right) \tag{7.5}$$

where $\phi_n = \{\breve{\boldsymbol{\mu}}_{\mathbf{z}_n}, \breve{\boldsymbol{\Sigma}}_{\mathbf{z}_n}\}$ are the local variational parameters and $\lambda = \{\breve{\boldsymbol{\mu}}_{\mathbf{w}_i}, \breve{\boldsymbol{\Sigma}}_{\mathbf{w}_i}, \breve{\boldsymbol{\mu}}_{\boldsymbol{\mu}}, \breve{\boldsymbol{\Sigma}}_{\boldsymbol{\mu}}, \breve{\mathbf{K}}^{-1}, \breve{\boldsymbol{\nu}}\}$ are the global variational parameters.

The mean field approximation allows the local and global parameters of the surrogate posteriors to be determined which, for VB, is achieved using coordinate ascent on the gradient of the ELBO. The VB algorithm first optimises the local parameters given all the datapoints, and then re-estimates the global parameters, iterating until convergence of the ELBO.

[2]Note the equation recalled here is written in terms of precision, $\boldsymbol{\Psi}$

7.3 Stochastic variational inference

One challenge with traditional VI and VB methods is that the optimisation requires the ELBO be computed after considering the entire dataset, which can be computationally expensive for large data. Consider the problem of Bayesian CCA for data with N observations and recall the form of the ELBO as:

$$\mathcal{L}(\boldsymbol{\phi}, \lambda) = \mathbb{E}_{q(\mathbf{z},\boldsymbol{\theta})} \left[\log p(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}) - \log q(\mathbf{z}, \boldsymbol{\theta})\right] \tag{7.6}$$

It is clear that, to compute the ELBO, the expectation term must be taken over all the data, which becomes prohibitive as N grows sufficiently large. One possible way to address this problem is to use a method known as Stochastic Variational Inference (SVI) [28]. SVI uses stochastic approximation techniques to replace the full-data ELBO optimisation with a stochastic approximation. Instead of computing the expectation over all data at once, SVI uniformly subsamples the data and computes the expectation, making it much more computationally feasible. The computation of the ELBO now becomes:

$$\mathcal{L}(\boldsymbol{\phi}, \lambda) = \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}_{q(\mathbf{z}_n,\boldsymbol{\theta})} \left[\log p(\mathbf{x}_n, \mathbf{z}_n, \boldsymbol{\theta}) - \log q(\mathbf{z}_n, \boldsymbol{\theta})\right] \tag{7.7}$$

In this case, the batch provides a noisy estimate for the ELBO, which allows the optimisation of the variational parameters to be performed efficiently through stochastic gradient ascent.

To perform the optimisation, SVI uses stochastic gradient ascent to form noisy estimates of the *natural* gradients of the ELBO. The term 'natural' here refers to an alternative canonical form of the exponential family of distributions that can be used to simplify computation. An introduction to the natural representation, including natural parameters η, and sufficient statistics, can be found in e.g. [173]. SVI first proceeds by finding the local variational parameters ϕ for a single datapoint. A set of intermediate global parameters $\hat{\lambda}$ are then computed from the natural gradient of the ELBO as though that datapoint were repeated N times. These intermediate parameters are then used to update the current estimate of the global parameters $\lambda^{(t-1)}$ according to,

$$\lambda^{(t)} = (1 - \rho_t)\lambda^{(t-1)} + \rho_t \hat{\lambda} \tag{7.8}$$

where ρ is a decreasing step size [28]. It can be seen that this approach is equivalent to a stochastic optimisation scheme for λ. These steps are then repeated, sampling the dataset multiple times until the desired convergence.

Hoffman [28] also showed that SVI can be extended to a 'mini-batch' algorithm, in a similar way to stochastic GD, to improve the algorithm's stability and reduce the variance in the estimates of $\hat{\lambda}$. In the mini-batch case, batches of S uniformly random points are subsampled from the dataset and considered in each step. The local variational parameters are once again computed for each of the data points in the batch, and the intermediate global parameters $\hat{\lambda}_s$ are computed in the same way. However, before updating the global parameters, the intermediate global estimates are averaged over the batch, such that

$$\lambda^{(t)} = (1 - \rho_t)\lambda^{(t-1)} + \frac{\rho_t}{S} \sum_{s=1}^{S} \hat{\lambda}_s \tag{7.9}$$

Using this type of approach has significant benefits, such as amortising any computational expense associated with updating the global parameters using all the data at once (as is true in the case of Bayesian SSI, incurring this cost less frequently) and helping the algorithm find a better variational posterior.

While SVI offers substantial advantages, there are some limitations that need to be considered. In particular, the reliance on noisy updates can sometimes lead to slower convergence compared to traditional VI methods. Particular care needs to be taken when tuning the step size to ensure stability and efficient convergence.

7.4 Stochastic variational inference for Bayesian SSI

Using the theory of mini-batch SVI, the following natural parameter updates were derived from the equations in Chapter 6 for the specific task of performing Bayesian CCA for SSI. The updates for the local and global variational parameters are given below and used in conjunction with Equation (7.9) and the methodology in [28] to find the moments of the surrogate distributions for $\mathbf{W}$, $\boldsymbol{\Psi}$, $\boldsymbol{\mu}$ and $\mathbf{z}$.

Local variational parameters

Update z

$$\eta_z = \begin{bmatrix} \langle \mathbf{W} \rangle^{\mathsf{T}} \langle \boldsymbol{\Psi} \rangle (\mathbf{x}_n - \langle \boldsymbol{\mu} \rangle) \\ -\frac{1}{2} \breve{\boldsymbol{\Sigma}}_{\mathbf{z}}^{-1} \end{bmatrix} \tag{7.10}$$

where $\breve{\boldsymbol{\Sigma}}_{\mathbf{z}}^{-1} = \langle \mathbf{W}^{\mathsf{T}} \boldsymbol{\Psi} \mathbf{W} \rangle + \mathbb{I}$

Global variational parameters

Update $\boldsymbol{\Psi}$

$$\eta_{\Psi} = \begin{bmatrix} -\frac{1}{2} \breve{\mathbf{K}}^{-1} \\ \frac{\nu_0 + N - D - 1}{2} \end{bmatrix} \tag{7.11}$$

where $\breve{\mathbf{K}}^{-1} = \mathbf{K}_0^{-1} + N \left\langle (\mathbf{x}_n - \boldsymbol{\mu} - \mathbf{W}\mathbf{z}_n)(\mathbf{x}_n - \boldsymbol{\mu} - \mathbf{W}\mathbf{z}_n)^{\mathsf{T}} \right\rangle$

Update $\boldsymbol{\mu}$

$$\eta_z = \begin{bmatrix} N \langle \boldsymbol{\Psi} \rangle (\mathbf{x}_n - \langle \mathbf{W} \rangle \langle \mathbf{z}_n \rangle) + \boldsymbol{\Sigma}_{\mu}^{-1} \boldsymbol{\mu}_{\mu} \\ -\frac{1}{2} \breve{\boldsymbol{\Sigma}}_{\boldsymbol{\mu}}^{-1} \end{bmatrix} \tag{7.12}$$

where $\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\mu}}^{-1} = N \langle \boldsymbol{\Psi} \rangle + \boldsymbol{\Sigma}_{\mu}^{-1}$

Update $\mathbf{w}_i$

$$\eta_{w_i} = \begin{bmatrix} N\langle \mathbf{z}_n \rangle \langle \boldsymbol{\Psi} \rangle (\mathbf{x}_n - \langle \boldsymbol{\mu} \rangle) \\ -\frac{1}{2} \breve{\boldsymbol{\Sigma}}_{\mathbf{w}_i}^{-1} \end{bmatrix} \tag{7.13}$$

where $\breve{\boldsymbol{\Sigma}}_{\mathbf{w}_i}^{-1} = N\langle \mathbf{z}_{i,n} \mathbf{z}_{i,n} \rangle \langle \boldsymbol{\Psi} \rangle + \boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}$ and $\tilde{\mathbf{x}}_n = \mathbf{x}_n - \langle \boldsymbol{\mu} \rangle - \langle \mathbf{w}_{\_i} \rangle \langle \mathbf{z}_{\_i,n} \rangle$ such that $\mathbf{w}_{\_i}$ corresponds to all columns except the ith column of interest and $\mathbf{z}_{\_i,n}$ corresponds to all rows of $\mathbf{z}_n$ except the ith row.

7.5 Results and discussion

To compare the performance of the VI and SVI inference schemes for Bayesian SSI, response data from a generic three-degree of freedom linear dynamic system, described by the modal properties $\omega = \{10.54, 16.35, 24.34\}$ and $\zeta = \{0.0051, 0.0076, 0.0033\}$, were generated given a simple white noise excitation. The system was simulated at a sample rate of 1000 Hz, generating 16 384 datapoints. SSI-Cov, Bayesian SSI (VI) and Bayesian SSI (mini-batch SVI) were then applied assuming a larger than true model order of 20 (10 unique modes) and a lag size of 40. A simulated system was used to allow comparison to a known ground truth.

The following weakly informative, proper priors were chosen, $\mathbf{w}_i \sim \mathcal{N}(\mathbf{0}, \mathbb{I})$, $\boldsymbol{\mu} \sim \mathcal{N}(\mathbf{0}, \mathbb{I} \times 10^{-3})$, $\boldsymbol{\Psi} \sim \mathcal{W}(\mathbb{I} \times 10^{1}, D + 1)$, providing the model with sufficient flexibility. The priors on the mean and variance were lightly constrained given assumed prior knowledge of a zero mean and low measurement noise. Histograms of the prior modal properties, using samples drawn and propagated from the prior observability matrix, are presented in Figure 7.2.

Given the implementations of both algorithms have yet to be optimised for a fair wall-time comparison, the required number of necessary repeats (i.e. the number of times the entire dataset was analysed) to reach convergence of the evidence lower-bound (ELBO) was instead used as an initial indicator of performance. For the batch SVI analysis, a batch size of $S = 1024$ was chosen, equivalent to 16 batches, with the common step size function $\rho = s^{-k}$ where s is the current batch number and k is the forgetting rate, chosen as $k = 0.95$. Both methods were then applied, with VI converging after 5 iterations (5 full sweeps of the data) and batch SVI converging after 48 batches, equivalent to 3 full sweeps of the data.

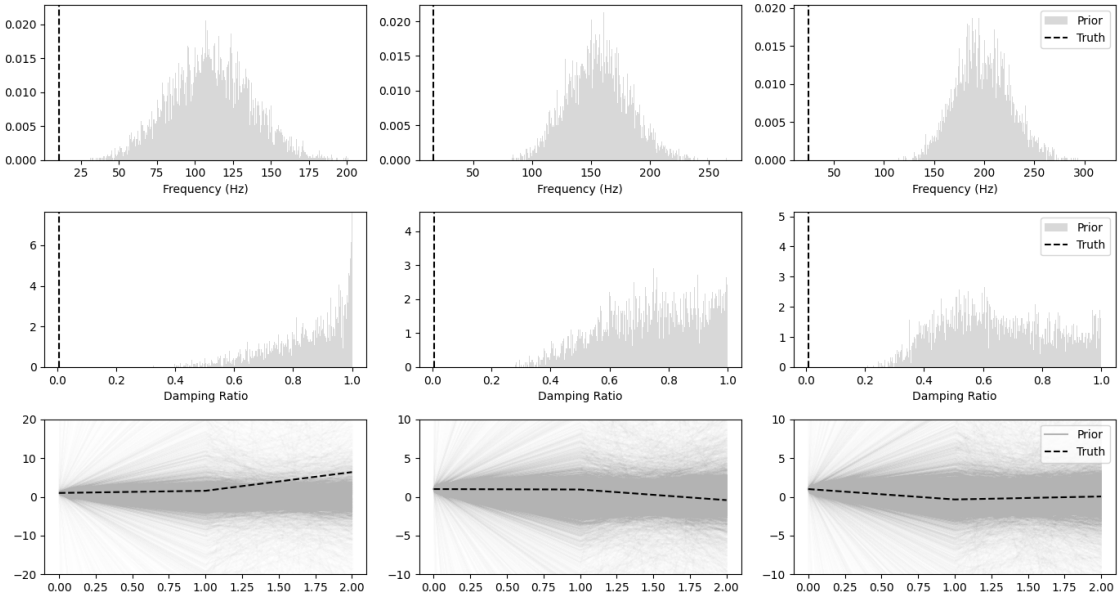


Figure 7.2: Approximate prior distributions over the modal properties as histograms, propagated from samples of the prior distribution over the observability matrix. (top to bottom: natural frequency, damping ratio, normalised mode shape)

As can be seen from Figure 7.3, the approximate posteriors of both algorithms are in relatively good agreement with one another, and with the results from SSI-Cov. This is expected as the posterior mean should converge to the maximum-a-posteriori estimate. Furthermore, the posteriors from batch SVI demonstrate lower overall variance, despite two fewer full sweeps of the data. This is believed to be a result of the stochastic nature of batch SVI, which is often hypothesised to reduce possible stagnation at saddle points.

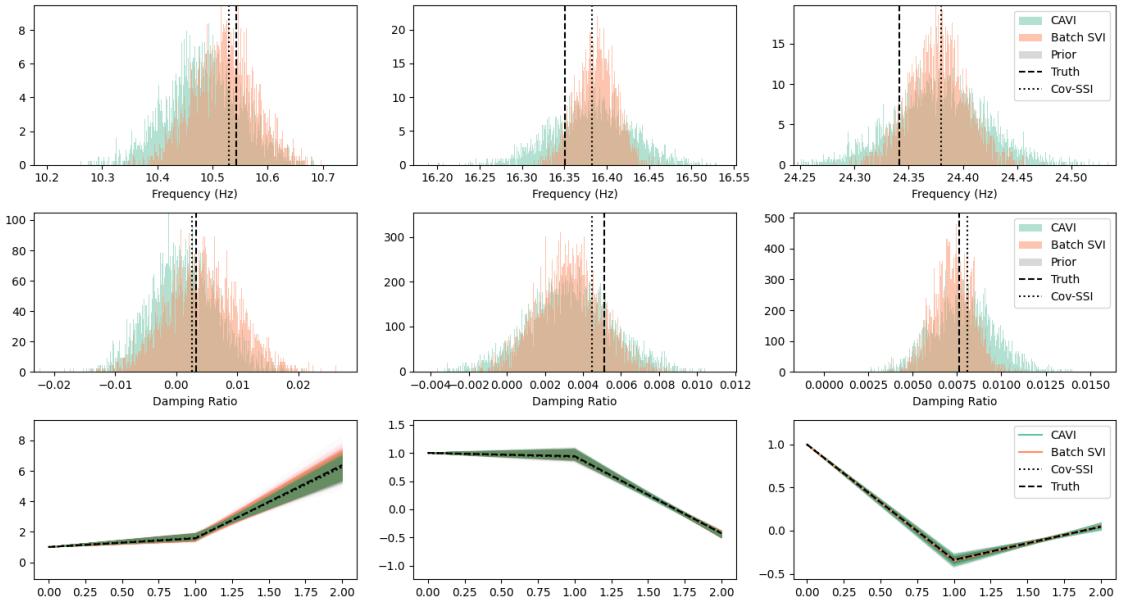


Figure 7.3: Posterior distributions over the modal properties as histograms, made up of samples drawn from the closed form surrogate posterior over the observability matrix found using CAVI and mini-batch SVI. These samples are then propagated onto the modal properties. (top to bottom: natural frequency, damping ratio, normalised mode shape)

7.6 Summary

Given the higher computational requirements associated with Bayesian methods, a problem also true of the Bayesian SSI algorithm introduced in Chapter 6, this chapter introduced an alternative implementation of the Bayesian SSI algorithm, using stochastic approximation techniques, as one possible solution to this challenge. In this new algorithm, the classic variational inference (VI) approach to Bayesian CCA is replaced with that of mini-batch stochastic variational inference (SVI), derived in this Chapter for the CCA problem for the first time. This SVI scheme benefits from the properties of stochastic optimisation, reducing the number operations on the full dataset before reaching convergence.

To assess the identification performance of the VI and mini-batch SVI implementations of Bayesian SSI, analysis was performed on data simulated from a simple MDOF linear system. The results of this test showed that mini-batch SVI was able to successfully reach convergence to a reasonable posterior, very comparable to that of VI, in fewer operations on the full dataset (3 repeats or epochs) than traditional

VI (5 epochs). Although this demonstrative mathematical improvement to computational performance is promising, in practice the compute time or "wall-time" of both algorithms in their current research implementation remained very similar. Nevertheless, mathematically speaking, the stochastic formulation is known to decrease the computational expense and it is believed with further code optimisation and efficient matrix operations, the SVI implementation can be greatly improved to achieve a much lower compute time.

As well as addressing the implementation, there are some other important factors to consider in completing this body of work. These include studying the effects of the hyperparameters, i.e. the batch size and forgetting rate, on the performance of the algorithm, the effects of too short a batch size on the underlying assumptions of SSI, and alternative prior elicitation in an effort to significantly reduce the time to convergence.

8

Conclusions

The field of operational modal analysis (OMA) continues to mature and develop. Its growth can be attributed to an increasing desire for access to modal information of operational structures in hard-to-replicate environments. Modal properties are frequently used as key features in safety-critical or cost-sensitive decision-making tasks; this is true of adjacent fields like structural health monitoring (SHM) and digital twins; thus making accurate modal information an important commodity. Consequently, the ability to confidently and reliably recover modal information is of critical value.

Despite frequent use across industry and academia, existing approaches to OMA are not without their limitations and challenges. In the introduction to this thesis, the reader was introduced to a range of key research challenges present in the field of OMA. One may recall these as:

- Model order selection
- Manual identification
- Non-linearity
- Non-physical solutions
- Close modes
- Sensor placement
- Outliers and robustness
- Uncertainty and noise.

Of all the above challenges, the quantification of uncertainty was identified as being of particular interest. It is widely accepted that measured data contain uncertainties that can obscure the underlying signal one intends to measure. The reliance of OMA on sequences of observed data, known to be inherently uncertain, therefore implies any identified models using said data must also be uncertain. The effects of uncertainty on the identification process have been shown to result in variations in the recovered modal properties. However, uncertainty is seldom considered nor quantified. This possible oversight is concerning. Variations in modal information that arise from uncertainty or misidentification can pose a significant risk i.e. the misidentification of damage in SHM. Therefore, access to a measure of the uncertainty can provide a practitioner with additional useful information regarding the identified parameters, enabling more informed and confident decision-making. OMA algorithms capable of providing this information are, therefore, highly sought; hence the aim of this work.

The current thesis aimed to develop new probabilistic OMA algorithms that expand our current understanding and capabilities, expanding the increasing list of uncertainty quantification methods in OMA. The present author also prefaced that perhaps, by thinking probabilistically, it may be possible to address the problem of uncertainty whilst also providing a necessary framework to tackle some of the remaining research challenges.

8.1 Summary

This section summarises the Chapters of this thesis and highlights the major contributions of this work.

- In Chapter 2 the reader was provided with some preliminary context to help frame the problem of OMA. A range of methodologies were briefly introduced, with emphasis on the stochastic subspace identification (SSI) algorithm. This was followed by a review and summary of relevant literature, discussing the current approaches to uncertainty quantification in OMA, and Bayesian approaches to uncertainty quantification in OMA. This chapter concluded by presenting the lack of a Bayesian approach to SSI in the literature, with the view to address this problem in this thesis.

- Following a comprehensive review of the literature, Chapter 3 introduced relevant theory for mathematical concepts and techniques referenced throughout this thesis. This included: the theory of modal analysis and OMA, some brief probability theory, an introduction to Bayes' rule, parameter estimation and inference, and the theory of SSI.

- Chapter 4 introduced a novel probabilistic formulation of the SSI-Cov algorithm. It was shown that canonical correlation analysis (CCA) acts as the core identification mechanism at the heart of the canonical-variate weighted form of the SSI-Cov algorithm. By replacing this mechanism with the probabilistic model of CCA, the SSI-Cov algorithm was principally reformulated into a probabilistic one. The maximum likelihood estimates of this alternate form were demonstrated to mathematically match that of the traditional SVD-based algorithm. Although the results of the algorithm remain unchanged, the ability to consider SSI probabilistically presented a unique opportunity to consider hierarchical extensions that may address other challenges in OMA.

- Using the probabilistic form of SSI, Chapter 5 introduced a statistically robust form of the SSI algorithm (robust Prob-SSI), by considering the data being modelled with Student's t-distributed noise model, known to be statistically robust to outliers. The inclusion of this noise model formally defined the existing problem of robust CCA, which when considered in the context of SSI, provides a statistically robust identification mechanism needed for performing robust

OMA. This new robust algorithm was first benchmarked against traditional SSI using data from a simulated study, demonstrating similar performance over a number of repeated tests. The robustness of this algorithm was then demonstrated on a series of case studies using data simulated with varying types of artificially-induced outliers. In all cases, robust Prob-SSI demonstrated an improvement in identification performance over SSI-Cov, with stable columns of poles in the stabilisation diagrams. The proposed algorithm was then tested using data obtained from an in-service structure; the Z24 bridge. Results from this test reflected that of the simulated case, with robust Prob-SSI displaying more stable identification than SSI-Cov in a single stabilisation diagram and over a range of datasets.

- In Chapter 6, the probabilistic SSI model introduced in Chapter 4 was extended to one of Bayesian inference through the inclusion of an arbitrary prior structure; principally changing the identification mechanism from CCA to that of Bayesian CCA. It was shown how the inclusion of priors over the parameters of the latent variable model in the context of SSI enables the posterior over the observability matrix, and by extension the modal properties, to be obtained. This new approach was presented with two methods of performing the inference: Markov chain Monte Carlo (MCMC) and variational inference (VI). Both forms of Bayesian SSI were benchmarked using simulated data from a MDOF linear dynamic system and displayed comparable identification performance, whilst the posterior mean converged to the SSI result. It was shown how the uncertainty quantified over the natural frequencies, represented as histograms, have lower variance near the natural frequencies and larger variance at more spurious estimates. Similar variance estimates were also obtained following analysis of the Z24 bridge dataset, with columns of stable poles in the stabilisation diagram displaying much lower variance across multiple model orders when compared to more spurious estimates.

- Acknowledging the computational limitations of the proposed Bayesian methodology, Chapter 7 introduced an efficient form of Bayesian SSI using stochastic approximation; employing the use of mini-batch stochastic variational inference (SVI). The SVI scheme benefits from the properties of stochastic optimisation methods, reducing the number of operations on the full dataset before achieving convergence. The performance of the proposed algorithm was assessed and compared to that of the VI form of Bayesian SSI using data simulated from a

MDOF linear system. Results showed that mini-batch SVI was able to successfully achieve convergence to a reasonable posterior in fewer mathematical operations on the entire dataset. Despite a mathematical improvement, in practice the "wall-time" of both implementations remained similar. However, it was noted that with further code optimisation, the SVI implementation can be improved to achieve a lower compute time.

8.2 Limitations, challenges and future research

Several limitations, challenges and opportunities for research have been identified from the work and methodologies introduced in this thesis. These are identified herein and discussed openly in the interest of transparency and in the context of future research.

Enhanced modal identification

The ability to formulate SSI as a probabilistic model is invaluable. This new formulation permits hierarchical extensions that have the potential to enhance the identification capabilities of the SSI algorithm; the reader was introduced to two of these extensions in Chapters 5 and 6. Beyond these two approaches, one can perhaps envisage a collection of new SSI algorithms, each constructed using a hierarchical model designed to achieve a specific task. One fitting example, and perhaps the next logical algorithm in progression, would be an SSI algorithm that combines the work of Chapters 5 and 6. The resulting probabilistic graphical model for this algorithm is shown in Figure 8.1.

The author hypothesises that an SSI algorithm constructed using such a model would be capable of jointly providing robust modal identification and Bayesian uncertainty quantification over the modal parameters. Another possible avenue for future work, taking inspiration from the work of population-based SHM [9], is the extension of Bayesian SSI to perform OMA on multiple homogenous structures simultaneously. Such an approach could provide a population-based approach to OMA, allowing useful information to be transferred between structures during identification. Evidently, there is sufficient scope to explore different hierarchical extensions and their effects on the identification mechanism, to solve other research challenges.

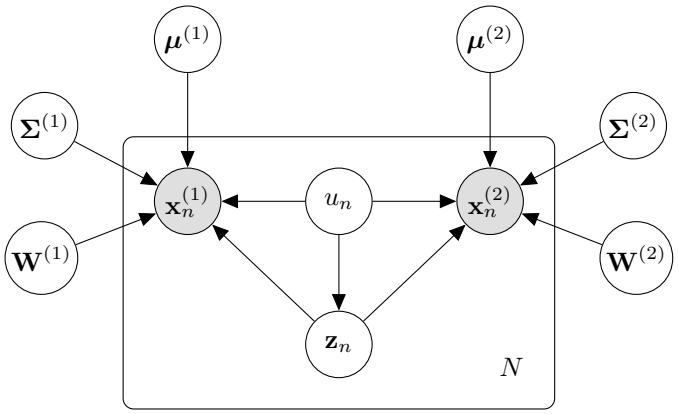


Figure 8.1: A probabilistic graphical model for robust Bayesian CCA.

Priors

One limitation identified of the Bayesian methodology introduced in Chapter 6 arises directly as a result of its dependence on the theory of SSI. When determining the modal properties from data, the classic SVD-based SSI algorithm can result in modal estimates that are not physically meaningful given the context, i.e. negative damping ratios consistent with non-physical attributes to a supposedly stable dynamic system. Mathematically, the construction of the original SSI algorithm does not enforce any restriction that damping ratios be positive. Considering the Bayesian form of SSI is principally constructed using the theory of SSI, Bayesian SSI (under the current choice of priors) also suffers from this same limitation. Consequently, posterior densities over the damping ratios can include negative estimates; this is suboptimal.

Nevertheless, one major advantage of a Bayesian methodology is in the ability to choose and appropriately specify the priors to reflect ones initial belief and embed existing knowledge. In the case of the Bayesian SSI algorithm introduced in this thesis, prior distributions are placed over the columns of the weights (the observability and controllability). This choice of prior follows from the initial definition of Bayesian CCA but is perhaps not the most suitable; under these assumptions the aforementioned limitation still holds. However, one could imagine there being a set of priors that can provide the necessary constraints to guarantee physically meaningful estimates.

The ability to alter and extend the prior structure as described above has an exhaustive list of possibilities; full scope of which is left to the creativity of the reader. However, the present author believes automatic identification and model order selection could also be achieved from an appropriate prior structure.

Computationally efficient implementations

Another important consideration is the computational requirements of the proposed Bayesian and probabilistic SSI methodologies. The ability to carry out OMA quickly and effectively is highly desirable; any computational bottlenecks could limit the adoption of Bayesian SSI (and any extensions thereof) in application. Whilst the results from Chapter 7 were particularly promising, there are many ways to address computational burden. One alternative approach to be explored would be to consider the use of an empirical prior. Initialising the optimisation closer to a solution would significantly reduce the time to reach convergence. In this case of Bayesian or robust Prob-SSI, the SSI result may be an appropriate choice.

Another possible avenue for future work would be to take advantage of the iterative nature of the probabilistic, robust and Bayesian implementations introduced in this thesis. Although perhaps seen as a key limitation in contrast to the SVD-based SSI algorithm, EM and VI inference schemes have the ability to update the parameters using batches of new data, without needing to recompute the solution with the entire dataset. Whilst possibly providing a way to reduce the computational burden, it is believed such an approach may provide a natural way to perform the proposed algorithms in an online recursive way, as new data is observed.

Uncertainty-based metrics

One final consideration, and one of significant interest as new uncertainty-based OMA algorithms are developed, is how uncertain modal parameters are compared. Standard metrics, such as the modal assurance criterion, may not be the most suitable for assessing modal parameters with posterior uncertainty. A new set of metrics may need to be defined that can appropriately describe the similarity of uncertain modes or stability of poles in a stabilisation diagram. Some work has already begun on this topic (see e.g. Gres et. al. [208]), however there is scope for new metrics that utilise the Bayesian posterior form of the uncertainty. Finally, as different methodologies for

uncertainty quantification quantify different types of uncertainty, questions remain unanswered regarding how structures with parameters obtained through different uncertainty quantification methods can be principally compared.

8.2.1 Concluding remarks

To quote the words of Rune Brincker, from his 2006 article in the magazine of Sound and Vibration [209], OMA is not another "flash in the pan"; the study of OMA continues to be of great importance and is still an active research field with several open research challenges. The key contribution of this thesis has been the introduction of novel probabilistic and Bayesian approaches, focussed on the SSI methodology, to OMA for the purpose of uncertainty quantification and robust modal identification. These contributions, and the identified avenues for future work, highlight a strong potential for further powerful extensions to the probabilistic and Bayesian interpretation of SSI presented here. The author hopes that focus and attention on OMA research will continue long into the future and that the work contained within this thesis will provide part of the foundation for new probabilistic and Bayesian approaches. In short, this thesis is only the beginning of a much wider exploration of OMA and specifically SSI as a probabilistic latent variable problem.

A

Appendices

A.1 Gaussian distribution

Univariate

The Gaussian or Normal distribution is a type of continuous probability distribution for a real-valued random variable and is part of the exponential family of distributions. If a random variable, X, is distributed according to a univariate Gaussian distribution, the following notation is used:

$$X \sim \mathcal{N}\left(\mu\,, \sigma^2\right) \tag{A.1}$$

where μ is a scalar describing the *mean* or the *expectation* of the distribution,

$$\mu \triangleq \mathbb{E}\left[X\right] \tag{A.2}$$

where $\mathbb{E}[\cdot]$ denotes the expectation, and σ^2 is the *variance* or "spread" of the distribution

$$\sigma^2 = \mathbb{V}[X] \triangleq \mathbb{E}\left[(X-\mu)^2\right] = \mathbb{E}\left[X^2\right] - \mu^2 \tag{A.3}$$

The *standard deviation*, σ is the defined as

$$\sigma = \text{std}[X] \triangleq \sqrt{\mathbb{V}[X]} \tag{A.4}$$

which is useful as it shares the same units as X itself.

The general form of the pdf for a univariate Gaussian is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \tag{A.5}$$

Figure A.1 presents multiple examples of Gaussian pdfs, each with varying values for mean and variance.

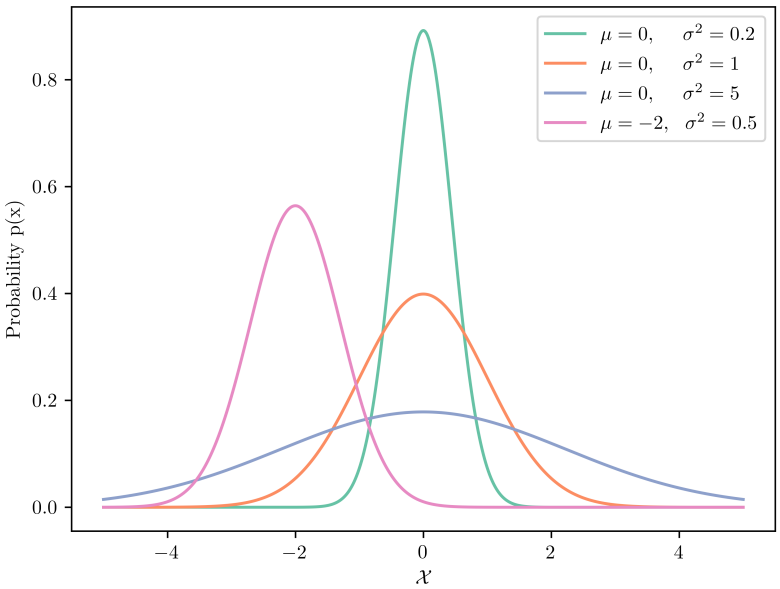


Figure A.1: Gaussian probability density functions

Multivariate

The univariate Gaussian can be generalised for vector-valued random variables to give the *multivariate Gaussian (normal) distribution*.

The multivariate distribution of a k-dimensional random vector $\mathbf{X} = [X_1, \ldots, X_k]^\mathsf{T}$ is written as:

$$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{A.6}$$

where $\boldsymbol{\mu}$ is the k-dimensional *mean vector*

$$\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}] = [\mathbb{E}[X_1], \ldots, \mathbb{E}[X_k]]^\mathsf{T} \tag{A.7}$$

and $\boldsymbol{\Sigma}$ is the $k \times k$ *covariance matrix* such that

$$\boldsymbol{\Sigma}_{i,j} = \mathbb{E}\left[(X_i - \mu_i)(X_j - \mu_j)^\mathsf{T}\right] = \mathrm{Cov}[X_i, X_j] \tag{A.8}$$

where $i, j = 1 \ldots k$. The inverse of the covariance matrix is called the *precision* and is denoted by $\boldsymbol{\Psi} = \boldsymbol{\Sigma}^{-1}$.

The general form of the pdf for a multivariate Gaussian distribution is described by

$$f(\mathbf{x}) = (2\pi)^{-k/2}|\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \tag{A.9}$$

A plot of an example bivariate Gaussian joint density is shown in Figure A.2.

It is also useful to define the log-form of the Gaussian, given its convenience when considering the joint of multiple exponential distributions.

$$\log(f(\mathbf{x})) = -\frac{k}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \tag{A.10}$$

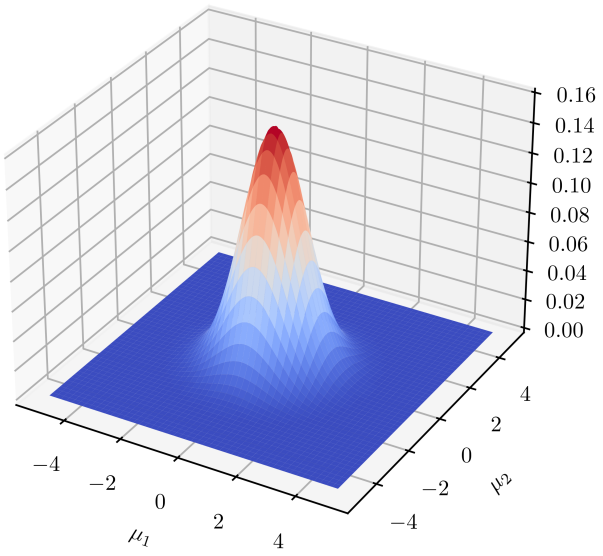


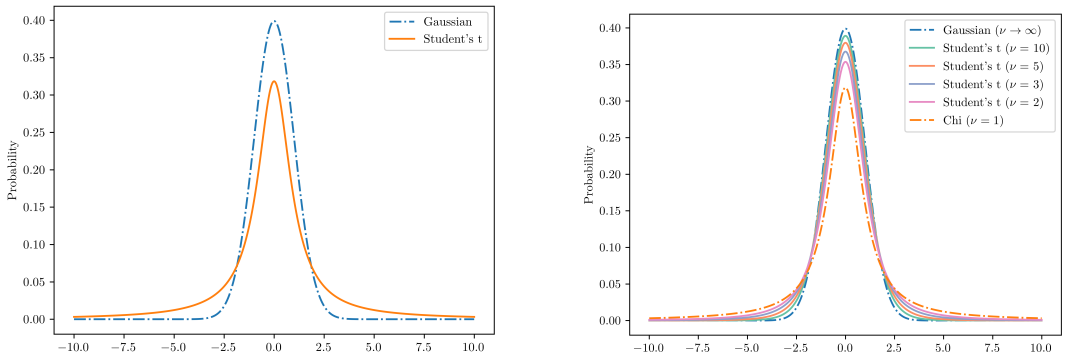
Figure A.2: Bivariate Gaussian probability density function

A.2 Student's t-distribution

One downside of the Gaussian distribution is that it can be quite sensitive to *outliers*. A statistically robust alternative to the Gaussian is the *Student's t-distribution*[1]. The Student's t-distribution is a continuous probability distribution that generalises the Gaussian distribution. Unlike the Gaussian distribution, a Student's t-distribution has heavier tails, where the amount of probability mass in said tails is controlled by a parameter ν (see Figure A.3a). This extra probability mass allows the Student's t to represent outliers in the distribution, with less effect on the mean and variance of the distribution.

In the case where $\nu = 1$, the t-distribution becomes the standard Cauchy distribution, which has very large tails. In the opposing case, when $\nu \to \infty$, it becomes the standard normal distribution, with very thin tails. This is shown more clearly in Figure A.3b.

[1]The distribution gets its name from a 1908 paper in Biometrika written by William Sealy Gosset, who published under the pseudonym "Student" [210] to remain anonymous at the bequest of his employer. Gosset worked at the Guinness brewery in Dublin, Ireland and was interested in the problem of small sample sizes. This is not something the author can relate to, as many 'samples' of Guinness were consumed throughout their PhD.



(a) An example of Gaussian and Student's t-distributions

(b) Student t-distributions with varying degrees of freedom (ν)

The Student's t-distribution is typically represented with the following notation

$$\mathcal{S}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) \tag{A.11}$$

where μ and Σ represent mean and covariance (similar to a Gaussian) and the parameter ν is known as the *degrees of freedom*.

The general form of the pdf for a univariate Student's t-distribution is

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma(\frac{\nu}{2})}\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} \tag{A.12}$$

where $\Gamma(\cdot)$ is the Gamma function. For the multivariate case, the pdf is

$$f(\mathbf{x}) = \frac{\Gamma\left(\frac{\nu+k}{2}\right)}{\Gamma(\frac{\nu}{2})\nu^{k/2}\pi^{k/2}\boldsymbol{\Sigma}^{1/2}}\left(1 + \frac{1}{\nu}(\mathbf{x}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)^{-\frac{\nu+k}{2}} \tag{A.13}$$

A.3 Robust probabilistic CCA derivation

Given the model definition of the original problem in Section 4.1 and the extended robust probabilistic CCA problem in Section 5.2, this Appendix contains the derivation for the EM update equations for robust probabilistic CCA. These are provided because of typographical errors in the original manuscript for the algorithm [192]. In the original robust projections paper, the normal distributions are defined using the precision matrix, rather than covariance matrix. Note that here only covariance matrices are used.

To derive the necessary update equations for the model parameters and an appropriate Q-function needed to monitor convergence of the EM algorithm, the log-likelihood function must first be defined:

$$\ell(\boldsymbol{\theta}) = \sum_{n=1}^{N} \ln p(\mathbf{x}_n, \mathbf{z}_n, u_n | \boldsymbol{\theta}) \tag{A.14}$$

where $\boldsymbol{\theta} = (\boldsymbol{\mu}, \mathbf{W}, \boldsymbol{\Sigma}, \nu)$.

Writing the log-likelihood in full,

$$\begin{aligned} \ell(\boldsymbol{\theta}) = \frac{1}{N} \sum_{n=1}^{N} \Bigg[& -\frac{D}{2} \ln (2\pi) - \frac{1}{2} \ln |u_n^{-1} \mathbb{I}_d| - \frac{1}{2} (\mathbf{z}_n - \mathbf{0})^{\mathsf{T}} (u_n^{-1} \mathbb{I}_d)^{-1} (\mathbf{z}_n - \mathbf{0}) \\ & + \frac{\nu}{2} \ln \left(\frac{\nu}{2} \right) - \ln \left(\Gamma \left(\frac{\nu}{2} \right) \right) + \frac{\nu}{2} \ln (u_n) - \frac{\nu}{2} u_n - \frac{D}{2} \ln (2\pi) \\ & - \frac{1}{2} \ln |u_n^{-1} \boldsymbol{\Sigma}| - \frac{1}{2} \big(\mathbf{x}_n - (\mathbf{W} \mathbf{z}_n + \boldsymbol{\mu}) \big)^{\mathsf{T}} (u_n^{-1} \boldsymbol{\Sigma})^{-1} \big(\mathbf{x}_n - (\mathbf{W} \mathbf{z}_n + \boldsymbol{\mu}) \big) \Bigg] \end{aligned} \tag{A.15}$$

The log matrix identity $\ln|aX| = D\ln(a) + \ln|X|$ can be used to expand and simplify the terms $\ln|(u_n\mathbb{I}_d)|$ and $\ln|u_n\boldsymbol{\Sigma}^{-1}|$, resulting in the form

$$\begin{aligned}\ell(\boldsymbol{\theta}) = \frac{1}{N}\sum_{n=1}^{N}\Bigg[&-D\ln(2\pi) + D\ln(u_n) - \frac{1}{2}u_n\mathbf{z}_n^\mathsf{T}\mathbf{z}_n + \frac{\nu}{2}\ln\left(\frac{\nu}{2}\right)\\ &- \ln\left(\Gamma\left(\frac{\nu}{2}\right)\right) + \frac{\nu}{2}\ln(u_n) - \frac{\nu}{2}u_n - \frac{1}{2}\ln|\boldsymbol{\Sigma}|\\ &- \frac{1}{2}u_n(\mathbf{x}_n-\boldsymbol{\mu})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_n-\boldsymbol{\mu}) + \frac{1}{2}u_n(\mathbf{x}_n-\boldsymbol{\mu})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\mathbf{W}\mathbf{z}_n)\\ &+ \frac{1}{2}u_n(\mathbf{W}\mathbf{z}_n)^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_n-\boldsymbol{\mu}) - \frac{1}{2}u_n(\mathbf{W}\mathbf{z}_n)^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\mathbf{W}\mathbf{z}_n)\Bigg]\end{aligned} \tag{A.16}$$

Q function derivation

Given the definition of the log likelihood in Equation A.16, the Q-function [173], useful for monitoring convergence, and subsequently the EM update equations, can be derived.

$$\begin{aligned}\mathcal{Q} = \frac{1}{N}\sum_{n=1}^{N}\Bigg[&-D\ln(2\pi) + D\mathbb{E}\left[\ln(u_n)\right] - \frac{1}{2}\mathbb{E}\left[u_n\right]\mathbb{E}\left[\mathbf{z}_n^\mathsf{T}\mathbf{z}_n\right] + \frac{\nu}{2}\ln\left(\frac{\nu}{2}\right)\\ &- \ln\left(\Gamma\left(\frac{\nu}{2}\right)\right) + \frac{\nu}{2}\mathbb{E}\left[\ln(u_n)\right] - \frac{\nu}{2}\mathbb{E}\left[u_n\right] - \frac{1}{2}\ln|\boldsymbol{\Sigma}|\\ &- \frac{1}{2}\mathbb{E}\left[u_n\right](\mathbf{x}_n-\boldsymbol{\mu})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_n-\boldsymbol{\mu}) + \frac{1}{2}\mathbb{E}\left[u_n\right](\mathbf{x}_n-\boldsymbol{\mu})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\mathbf{W}\mathbb{E}\left[\mathbf{z}_n\right])\\ &+ \frac{1}{2}\mathbb{E}\left[u_n\right](\mathbf{W}\mathbb{E}\left[\mathbf{z}_n\right])^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_n-\boldsymbol{\mu}) - \frac{1}{2}\mathbb{E}\left[u_n(\mathbf{W}\mathbf{z}_n)^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\mathbf{W}\mathbf{z}_n)\right]\Bigg]\end{aligned} \tag{A.17}$$

Let the expectations of the latent parameters be denoted by

$$\mathbb{E}\left[u_n\right] = \bar{u}_n \tag{A.18}$$

$$\mathbb{E}\left[\ln(u_n)\right] = \ln\tilde{u}_n \tag{A.19}$$

$$\mathbb{E}\left[\mathbf{z}_n\right] = \bar{\mathbf{z}}_n \tag{A.20}$$

$$\mathbb{E}\left[u_n\mathbf{z}_n\mathbf{z}_n^\mathsf{T}\right] = \bar{\mathbf{S}}_n \tag{A.21}$$

Using the joint posterior distributions of the latent scale variables and the latent vectors, exact solutions for the expectations in Equations (A.18) - (A.21) can be recovered.

If

$$p(\mathbf{x}_n|u_n) = \int p(\mathbf{x}_n|\mathbf{z}_n, u_n)p(\mathbf{z}_n|u_n)\mathrm{d}\mathbf{z} \tag{A.22}$$

$$= \mathcal{N}\left(\boldsymbol{\mu}\,, u_n^{-1}\left(\boldsymbol{\Sigma} + \mathbf{W}\mathbf{W}^{\mathsf{T}}\right)\right) \tag{A.23}$$

where $\mathbf{A} = \left(\boldsymbol{\Sigma} + \mathbf{W}\mathbf{W}^{\mathsf{T}}\right)$, and

$$p(u_n|\mathbf{x}_n) \propto p(\mathbf{x}_n|u_n)p(u_n) \tag{A.24}$$

then

$$p(u_n|\mathbf{x}_n) = \mathcal{G}\left(u_n\middle|\frac{D+\nu}{2}\,, \frac{u_n(\mathbf{x}_n - \boldsymbol{\mu})^{\mathsf{T}}\mathbf{A}^{-1}(\mathbf{x}_n - \boldsymbol{\mu})}{2} + \frac{\nu}{2}\right) \tag{A.25}$$

The equivalent $\mathbb{E}\left[u_n\right]$ of the Gamma distribution gives

$$\bar{u}_n = \frac{\alpha}{\beta} = \frac{D+\nu}{u_n(\mathbf{x}_n - \boldsymbol{\mu})^{\mathsf{T}}\mathbf{A}^{-1}(\mathbf{x}_n - \boldsymbol{\mu}) + \nu} \tag{A.26}$$

whilst

$$\ln \tilde{u}_n = \psi\left(\frac{D+\nu}{2}\right) - \ln\left(\frac{u_n(\mathbf{x}_n - \boldsymbol{\mu})^{\mathsf{T}}\mathbf{A}^{-1}(\mathbf{x}_n - \boldsymbol{\mu}) + \nu}{2}\right) \tag{A.27}$$

is the equivalent expectation of the log Gamma distribution.

In a similar way,

$$p(\mathbf{z}_n|\mathbf{x}_n, u_n) = \propto p(\mathbf{x}_n|\mathbf{z}_n, u_n)p(\mathbf{z}_n|u_n) \tag{A.28}$$

$$= \mathcal{N}\left(\mathbf{B}^{-1}\mathbf{W}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_n - \boldsymbol{\mu})\,, (u_n\mathbf{B})^{-1}\right) \tag{A.29}$$

$$\bar{\mathbf{z}}_n = \mathbf{B}^{-1}\mathbf{W}^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_n - \boldsymbol{\mu}) \tag{A.30}$$

where $\mathbf{B} = \mathbf{W}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\mathbf{W} + \mathbb{I}_d$

$$\bar{\mathbf{S}}_n = \mathbf{B}^{-1} + \bar{u}_n\bar{\mathbf{z}}_n\bar{\mathbf{z}}_n^\mathsf{T} \tag{A.31}$$

The EM update equations can be recovered by taking the derivative of the Q-function, with respect to the various parameters $\boldsymbol{\theta}$, and setting the resulting equations to zero and solving for the parameter in question.

Covariance Update

$$\begin{aligned}\frac{\partial \mathcal{Q}}{\partial \boldsymbol{\Sigma}^{-1}} =& \frac{1}{N}\sum_{n=1}^{N}\Bigg[\frac{1}{2}\boldsymbol{\Sigma}^\mathsf{T} - \frac{1}{2}\bar{u}_n(\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{x}_n - \boldsymbol{\mu})^\mathsf{T} + \frac{1}{2}\bar{u}_n(\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{W}\bar{\mathbf{z}}_n)^\mathsf{T} \\ &+ \frac{1}{2}\bar{u}_n(\mathbf{W}\bar{\mathbf{z}}_n)(\mathbf{x}_n - \boldsymbol{\mu})^\mathsf{T} + \frac{1}{2}\bar{u}_n\mathbf{W}\bar{\mathbf{S}}_n\mathbf{W}^\mathsf{T}\Bigg]\end{aligned} \tag{A.32}$$

setting $\frac{\partial \mathcal{Q}}{\partial \boldsymbol{\Sigma}^{-1}} = 0$ and rearranging, note that due to the symmetry of $\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}^\mathsf{T} = \boldsymbol{\Sigma}$, the following update equation for the covariance can be defined.

$$\begin{aligned}\boldsymbol{\Sigma} =& \frac{1}{N}\sum_{n=1}^{N}\Bigg[\bar{u}_n(\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{x}_n - \boldsymbol{\mu})^\mathsf{T} - \bar{u}_n(\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{W}\bar{\mathbf{z}}_n)^\mathsf{T} \\ &- \bar{u}_n(\mathbf{W}\bar{\mathbf{z}}_n)(\mathbf{x}_n - \boldsymbol{\mu})^\mathsf{T} + \bar{u}_n\mathbf{W}\bar{\mathbf{S}}_n\mathbf{W}^\mathsf{T}\Bigg]\end{aligned} \tag{A.33}$$

Weight Update

$$\frac{\partial \mathcal{Q}}{\partial \mathbf{W}} = \frac{1}{N}\sum_{n=1}^{N}\Bigg[\bar{u}_n\boldsymbol{\Sigma}^{-1}(\mathbf{x}_n - \boldsymbol{\mu})\bar{\mathbf{z}}_n^\mathsf{T} - \boldsymbol{\Sigma}^{-1}\mathbf{W}\bar{\mathbf{S}}_n\Bigg] \tag{A.34}$$

letting $\frac{\partial \mathcal{Q}}{\partial \mathbf{W}} = 0$ and rearranging,

$$\mathbf{W} = (\sum_{n=1}^{N} \bar{u}_n(\mathbf{x}_n - \boldsymbol{\mu})\bar{\mathbf{z}}_n^{\mathsf{T}})(\sum_{n=1}^{N} \bar{\mathbf{S}}_n)^{-1} \tag{A.35}$$

Mean Update

$$\frac{\partial \mathcal{Q}}{\partial \boldsymbol{\mu}} = \frac{1}{N}\sum_{n=1}^{N}\left[\bar{u}_n\boldsymbol{\Sigma}^{-1}(\mathbf{x}_n - \boldsymbol{\mu}) - \bar{u}_n\boldsymbol{\Sigma}^{-1}\mathbf{W}\bar{\mathbf{z}}_n\right] \tag{A.36}$$

letting $\frac{\partial \mathcal{Q}}{\partial \boldsymbol{\mu}} = 0$ and expanding,

$$\sum_{n=1}^{N}\left[\bar{u}_n\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right] = \sum_{n=1}^{N}\left[\bar{u}_n\boldsymbol{\Sigma}^{-1}\mathbf{x}_n - \bar{u}_n\boldsymbol{\Sigma}^{-1}\mathbf{W}\bar{\mathbf{z}}_n\right] \tag{A.37}$$

premultiplying by $\boldsymbol{\Sigma}$ and rearranging for $\boldsymbol{\mu}$ gives

$$\sum_{n=1}^{N}\left[\bar{u}_n\boldsymbol{\mu}\right] = \sum_{n=1}^{N}\left[\bar{u}_n(\mathbf{x}_n - \mathbf{W}\bar{\mathbf{z}}_n)\right] \tag{A.38}$$

$$\boldsymbol{\mu} = \frac{\sum_{n=1}^{N}\bar{u}_n(\mathbf{x}_n - \mathbf{W}\bar{\mathbf{z}}_n)}{\sum_{n=1}^{N}\bar{u}_n} \tag{A.39}$$

ν Update

$$\frac{\partial \mathcal{Q}}{\partial \nu} = \frac{1}{N}\sum_{n=1}^{N}\left[\frac{1}{2} + \frac{1}{2}\ln\left(\frac{\nu}{2}\right) - \psi\left(\frac{\nu}{2}\right) + \frac{1}{2}\ln(\tilde{u}_n) - \frac{1}{2}\bar{u}_n\right] \tag{A.40}$$

letting $\frac{\partial \mathcal{Q}}{\partial \nu} = 0$ and rearranging,

$$0 = 1 + \ln\left(\frac{\nu}{2}\right) - 2\,\psi\left(\frac{\nu}{2}\right) + \frac{1}{N}\sum_{n=1}^{N}\left[\ln(\tilde{u}_n) - \bar{u}_n\right] \tag{A.41}$$

The maximum likelihood solution of ν can be found through solving the above equation using line search.

A.4 Recovery of rotation matrix R

The following theory for the recovery of the rotation matrix in robust PCCA summarises the work presented by Archambeau et al. in Appendix A of [192] and the accompanying errata of Appendix A. for the same paper.

Having defined

$$\mathbf{B}_1 = \breve{\mathbf{W}}^{(1)^\mathsf{T}} \boldsymbol{\Sigma}^{(1)^{-1}} \breve{\mathbf{W}}^{(1)} + \mathbb{I}_d \tag{A.42}$$

$$\mathbf{B}_2 = \breve{\mathbf{W}}^{(2)^\mathsf{T}} \boldsymbol{\Sigma}^{(2)^{-1}} \breve{\mathbf{W}}^{(2)} + \mathbb{I}_d \tag{A.43}$$

The matrix $\mathbf{R}_1$ contains the eigenvectors of

$$\mathbf{J}_1 = \left(\mathbb{I}_d - \mathbf{B}_1^{-1}\right)^{\frac{1}{2}} \left(\mathbb{I}_d - \mathbf{B}_2^{-1}\right) \left(\mathbb{I}_d - \mathbf{B}_1^{-1}\right)^{\frac{1}{2}} \tag{A.44}$$

with $\tilde{\Upsilon}^2$ corresponding eigenvalues. Similarly $\mathbf{R}_2$ contains the eigenvectors of

$$\mathbf{J}_2 = \left(\mathbb{I}_d - \mathbf{B}_2^{-1}\right)^{\frac{1}{2}} \left(\mathbb{I}_d - \mathbf{B}_1^{-1}\right) \left(\mathbb{I}_d - \mathbf{B}_2^{-1}\right)^{\frac{1}{2}} \tag{A.45}$$

with the same eigenvalues $\tilde{\Upsilon}^2$.

Given the above, it can finally be shown that the canonical directions are

$$\mathbf{U}_d^{(1)} = \boldsymbol{\Sigma}_{11}^{-1} \breve{\mathbf{W}}^{(1)} \left(\mathbb{I}_d - \mathbf{B}_1^{-1}\right)^{-\frac{1}{2}} \mathbf{R}_1 \tag{A.46}$$

$$\mathbf{U}_d^{(2)} = \boldsymbol{\Sigma}_{22}^{-1} \breve{\mathbf{W}}^{(2)} \left(\mathbb{I}_d - \mathbf{B}_2^{-1}\right)^{-\frac{1}{2}} \mathbf{R}_2 \tag{A.47}$$

A.5 Bayesian CCA: Gibbs sampler derivation

This appendix provides a full derivation for the update equations provided in Algorithm 3, aligning with the Bayesian CCA formulation presented in Klami and Kaski [199] but without the sparsity prior and with the inclusion of generic priors.

Given the model defined in Figure 6.1, the joint can be written as

$$\mathcal{J} = p(\mathbf{x}, \mathbf{z}, \mathbf{W}, \boldsymbol{\Sigma}, \boldsymbol{\mu}) = p(\mathbf{x}|\mathbf{z}, \mathbf{W}, \boldsymbol{\Sigma}, \boldsymbol{\mu})p(\mathbf{z})p(\boldsymbol{\Sigma})p(\boldsymbol{\mu})p(\mathbf{W}) \tag{A.48}$$

Assuming independent columns of W, the joint can also be defined as

$$\mathcal{J} = p(\mathbf{x}|\mathbf{z}, \mathbf{W}, \boldsymbol{\Sigma}, \boldsymbol{\mu})p(\mathbf{z})p(\boldsymbol{\Sigma})p(\boldsymbol{\mu}) \prod_{i=1}^{d} p(\mathbf{w}_i) \tag{A.49}$$

Taking the natural log of the joint, the individual components of the expression are thus given by

$$\ln \mathcal{J}_{p(\mathbf{x}|\mathbf{z},\mathbf{W},\boldsymbol{\Sigma},\boldsymbol{\mu})} \propto -\frac{N}{2} \ln |\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{n=1}^{N} (\mathbf{x}_n - (\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu}))^{\mathsf{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{x}_n - (\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu})) \tag{A.50}$$

$$\ln \mathcal{J}_{p(\mathbf{z})} \propto -\frac{N}{2} \ln |\mathbb{I}| - \frac{1}{2} \sum_{n=1}^{N} (\mathbf{z}_n - 0)^{\mathsf{T}} \mathbb{I}^{-1} (\mathbf{z}_n - 0) = -\frac{1}{2} \sum_{n=1}^{N} \mathbf{z}_n^{\mathsf{T}} \mathbf{z}_n \tag{A.51}$$

$$\ln \mathcal{J}_{p(\boldsymbol{\Sigma})} \propto \frac{\nu_0}{2} \ln |\mathbf{K}_0| - \frac{\nu_0 D}{2} \ln(2) - \ln \left(\Gamma \left(\frac{\nu_0}{2} \right) \right) - \frac{\nu_0 + D + 1}{2} \ln |\boldsymbol{\Sigma}| - \frac{1}{2} \mathrm{Tr} \left(\mathbf{K}_0 \boldsymbol{\Sigma}^{-1} \right) \tag{A.52}$$

$$\ln \mathcal{J}_{p(\boldsymbol{\mu})} \propto -\frac{1}{2} \ln |\boldsymbol{\Sigma}_\mu| - \frac{1}{2} \sum_{n=1}^{N} (\boldsymbol{\mu} - \boldsymbol{\mu}_\mu)^{\mathsf{T}} \boldsymbol{\Sigma}_\mu^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_\mu) \tag{A.53}$$

$$\ln \mathcal{J}_{p(\mathbf{w}_i)} \propto -\frac{1}{2} \ln |\boldsymbol{\Sigma}_{\mathbf{w}_i}| - \frac{1}{2} \sum_{n=1}^{N} (\mathbf{w}_i - \boldsymbol{\mu}_{\mathbf{w}_i})^{\mathsf{T}} \boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1} (\mathbf{w}_i - \boldsymbol{\mu}_{\mathbf{w}_i}) \tag{A.54}$$

where

$$\prod_{i=1}^{d} p(\mathbf{w}_i) = \exp \left\{ \sum_{i=1}^{d} \ln p(\mathbf{w}_i) \right\} \tag{A.55}$$

Using the standard theory of Gibbs MCMC sampling the update equations for the model parameters can be constructed. The premise of Gibbs sampling is to sample each variable in turn, conditioned on the values of the other variables in the joint distribution to obtain a sample from the conditional distribution that is sought. This can be achieved by considering the log of the joint with respect to the parameter of interest, establishing the new conditional distribution, and evaluating that function given the current estimates for all the necessary model parameters.

The following subsections summarise the derivation of the individual updates. Even though a specified order is given in Algorithm 3, the nature of Gibbs sampling means that sampling from the conditionals can be conducted in any order. As such, the equations here are written without specifying which parameters are current or previous estimates, to simplify notation for clarity.

Sample $\boldsymbol{\Sigma}$

Collecting the terms of the log joint pertaining to $\boldsymbol{\Sigma}$

$$\begin{aligned} \ln p(\boldsymbol{\Sigma}^{\text{new}}) \propto -\frac{N}{2} \ln |\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{n=1}^{N} (\mathbf{x}_n - (\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu}))^{\mathsf{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{x}_n - (\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu})) \\ - \frac{\nu_0 + D + 1}{2} \ln |\boldsymbol{\Sigma}| - \frac{1}{2} \text{Tr} \left(\mathbf{K}_0 \boldsymbol{\Sigma}^{-1} \right) \end{aligned} \tag{A.56}$$

Using $\mathbf{x}^{\mathsf{T}} \mathbf{A} \mathbf{x} = \text{trace}(\mathbf{A} \mathbf{x} \mathbf{x}^{\mathsf{T}})$,

$$\begin{aligned} \ln p(\boldsymbol{\Sigma}^{\text{new}}) \propto -\frac{N}{2} \ln |\boldsymbol{\Sigma}| - \frac{1}{2} \text{Tr} \left(\boldsymbol{\Sigma}^{-1} \sum_{n=1}^{N} (\mathbf{x}_n - \boldsymbol{\mu} - \mathbf{W}\mathbf{z}_n)(\mathbf{x}_n - \boldsymbol{\mu} - \mathbf{W}\mathbf{z}_n)^{\mathsf{T}} \right) \\ - \frac{\nu_0 + D + 1}{2} \ln |\boldsymbol{\Sigma}| - \frac{1}{2} \text{Tr} \left(\mathbf{K}_0 \boldsymbol{\Sigma}^{-1} \right) \end{aligned} \tag{A.57}$$

Combining terms, the familiar inverse Wishart form can be obtained

$$\ln p(\boldsymbol{\Sigma}^{\text{new}}) \propto -\frac{\nu_0 + N + D + 1}{2} \ln |\boldsymbol{\Sigma}| \\ - \frac{1}{2}\text{Tr}\left(\left(\mathbf{K}_0 + \sum_{n=1}^{N}(\mathbf{x}_n - \boldsymbol{\mu} - \mathbf{W}\mathbf{z}_n)(\mathbf{x}_n - \boldsymbol{\mu} - \mathbf{W}\mathbf{z}_n)^{\mathsf{T}}\right)\boldsymbol{\Sigma}^{-1}\right) \quad \text{(A.58)}$$

Therefore, a new sample of $\boldsymbol{\Sigma}$ can be drawn from the found conditional distribution

$$\boldsymbol{\Sigma}^{\text{new}} \sim I\mathcal{W}\left(\mathbf{K}_0 + \mathbf{K}\,, \nu_0 + N\right) \quad \text{(A.59)}$$

where

$$\mathbf{K} = \sum_{n=1}^{N}(\mathbf{x}_n - \boldsymbol{\mu} - \mathbf{W}\mathbf{z}_n)(\mathbf{x}_n - \boldsymbol{\mu} - \mathbf{W}\mathbf{z}_n)^{\mathsf{T}} \quad \text{(A.60)}$$

Sample $\boldsymbol{\mu}$

Collecting the terms of the log joint pertaining to $\boldsymbol{\mu}$

$$\ln p(\boldsymbol{\mu}^{\text{new}}) \propto -\frac{1}{2}\sum_{n=1}^{N}\left\{(\mathbf{x}_n - (\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu}))^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_n - (\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu}))\right\} \\ - \frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_\mu)^{\mathsf{T}}\boldsymbol{\Sigma}_\mu^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_\mu) \quad \text{(A.61)}$$

Expanding and ignoring terms not containing $\boldsymbol{\mu}$,

$$\ln p(\boldsymbol{\mu}^{\text{new}}) \propto -\frac{1}{2}\sum_{n=1}^{N}\left\{\boldsymbol{\mu}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_n - \mathbf{W}\mathbf{z}_n) - (\mathbf{x}_n - \mathbf{W}\mathbf{z}_n)^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right\} \\ - \frac{1}{2}(\boldsymbol{\mu}^{\mathsf{T}}\boldsymbol{\Sigma}_\mu^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}_\mu^{\mathsf{T}}\boldsymbol{\Sigma}_\mu^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}^{\mathsf{T}}\boldsymbol{\Sigma}_\mu^{-1}\boldsymbol{\mu}_\mu) \quad \text{(A.62)}$$

Setting $\hat{\boldsymbol{\Sigma}}_{\mu}^{-1} = N\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}_{\mu}^{-1}$, and in the knowledge that $\mathbb{I} = \hat{\boldsymbol{\Sigma}}_{\mu}^{-1}\hat{\boldsymbol{\Sigma}}_{\mu}$, the log form of a Gaussian can be reached

$$\ln p(\boldsymbol{\mu}^{\text{new}}) \propto -\frac{1}{2}\left\{\boldsymbol{\mu}^{\mathsf{T}}\hat{\boldsymbol{\Sigma}}_{\mu}^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}^{\mathsf{T}}\hat{\boldsymbol{\Sigma}}_{\mu}^{-1}\hat{\boldsymbol{\Sigma}}_{\mu}(\boldsymbol{\Sigma}^{-1}\sum_{n=1}^{N}(\mathbf{x}_n - \mathbf{W}\mathbf{z}_n) + \boldsymbol{\Sigma}_{\mu}^{-1}\boldsymbol{\mu}_{\mu}) - \left(\sum_{n=1}^{N}(\mathbf{x}_n - \mathbf{W}\mathbf{z}_n)^{\mathsf{T}}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\mu}_{\mu}\boldsymbol{\Sigma}_{\mu}^{-1}\right)\boldsymbol{\mu}^{\mathsf{T}}\right\} \quad \text{(A.63)}$$

Therefore, a new sample of $\boldsymbol{\mu}$ can be drawn from the conditional distribution defined by

$$\boldsymbol{\mu}^{\text{new}} \sim \mathcal{N}\left(\hat{\boldsymbol{\mu}}_{\mu}, \hat{\boldsymbol{\Sigma}}_{\mu}\right) \quad \text{(A.64)}$$

where

$$\hat{\boldsymbol{\Sigma}}_{\mu} = (N\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}_{\mu}^{-1})^{-1} \quad \text{(A.65)}$$

and

$$\hat{\boldsymbol{\mu}}_{\mu} = \hat{\boldsymbol{\Sigma}}_{\mu}(\boldsymbol{\Sigma}^{-1}\sum_{n=1}^{N}(\mathbf{x}_n - \mathbf{W}\mathbf{z}_n) + \boldsymbol{\Sigma}_{\mu}^{-1}\boldsymbol{\mu}_{\mu}) \quad \text{(A.66)}$$

Sample $\mathbf{w}_i$

Collecting the terms of the log joint pertaining to $\mathbf{W}$

$$\ln p(\mathbf{w}_i^{\text{new}}) \propto -\frac{1}{2}\sum_{n=1}^{N}\left\{(\mathbf{x}_n - (\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu}))^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_n - (\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu}))\right\} - \frac{1}{2}\sum_{i=1}^{d}\left\{(\mathbf{w}_i - \boldsymbol{\mu}_{\mathbf{w}_i})^{\mathsf{T}}\boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}(\mathbf{w}_i - \boldsymbol{\mu}_{\mathbf{w}_i})\right\} \quad \text{(A.67)}$$

As the intention is to update only one column of $\mathbf{W}$ at a time, i.e the independent column $\mathbf{w}_i$, the above equation can be separated as follows

$$\ln p(\mathbf{w}_i^{\text{new}}) \propto -\frac{1}{2}\sum_{n=1}^{N}\left\{(\mathbf{x}_n - \boldsymbol{\mu} - \mathbf{w}_{\_i}\mathbf{z}_{\_i,n} - \mathbf{w}_i\mathbf{z}_{i,n})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_n - \boldsymbol{\mu} - \mathbf{w}_{\_i}\mathbf{z}_{\_i,n} - \mathbf{w}_i\mathbf{z}_{i,n})\right\} - \frac{1}{2}(\mathbf{w}_i - \boldsymbol{\mu}_{\mathbf{w}_i})^\mathsf{T}\boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}(\mathbf{w}_i - \boldsymbol{\mu}_{\mathbf{w}_i}) \quad \text{(A.68)}$$

where the negative indices notation, $\cdot_{\_i}$, refers to every column/row of except the one specified. Letting $\tilde{\mathbf{x}}_n = \mathbf{x}_n - \boldsymbol{\mu} - \mathbf{w}_{\_i}\mathbf{z}_{\_i,n}$, this equation can be simplified to

$$\ln p(\mathbf{w}_i^{\text{new}}) \propto -\frac{1}{2}\sum_{n=1}^{N}\left\{(\tilde{\mathbf{x}}_n - \mathbf{w}_i\mathbf{z}_{i,n})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\tilde{\mathbf{x}}_n - \mathbf{w}_i\mathbf{z}_{i,n})\right\} - \frac{1}{2}(\mathbf{w}_i - \boldsymbol{\mu}_{\mathbf{w}_i})^\mathsf{T}\boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}(\mathbf{w}_i - \boldsymbol{\mu}_{\mathbf{w}_i}) \quad \text{(A.69)}$$

and expanded to give

$$\ln p(\mathbf{w}_i^{\text{new}}) \propto -\frac{1}{2}\sum_{n=1}^{N}\left\{\tilde{\mathbf{x}}_n^\mathsf{T}\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{x}}_n - (\mathbf{w}_i\mathbf{z}_{i,n})^\mathsf{T}\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{x}}_n - \tilde{\mathbf{x}}_n^\mathsf{T}\boldsymbol{\Sigma}^{-1}\mathbf{w}_i\mathbf{z}_{i,n} + (\mathbf{w}_i\mathbf{z}_{i,n})^\mathsf{T}\boldsymbol{\Sigma}^{-1}\mathbf{w}_i\mathbf{z}_{i,n}\right\} - \frac{1}{2}\left(\mathbf{w}_i^\mathsf{T}\boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}\mathbf{w}_i - \mathbf{w}_i^\mathsf{T}\boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}\boldsymbol{\mu}_{\mathbf{w}_i} - \boldsymbol{\mu}_{\mathbf{w}_i}^\mathsf{T}\boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}\mathbf{w}_i + \boldsymbol{\mu}_{\mathbf{w}_i}^\mathsf{T}\boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}\boldsymbol{\mu}_{\mathbf{w}_i}\right) \quad \text{(A.70)}$$

As $\mathbf{z}_{i,n}$ is a scalar, it can be manipulated as such, rearranging to give

$$\ln p(\mathbf{w}_i^{\text{new}}) \propto -\frac{1}{2}\left\{\sum_{n=1}^{N}\{\mathbf{z}_{i,n}\mathbf{z}_{i,n}\}\mathbf{w}_i^\mathsf{T}\boldsymbol{\Sigma}^{-1}\mathbf{w}_i - \mathbf{w}_i^\mathsf{T}\boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}\mathbf{w}_i - \sum_{n=1}^{N}\mathbf{z}_{i,n}\mathbf{w}_i^\mathsf{T}\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{x}}_n - \sum_{n=1}^{N}\mathbf{z}_{i,n}\tilde{\mathbf{x}}_n^\mathsf{T}\boldsymbol{\Sigma}^{-1}\mathbf{w}_i - \mathbf{w}_i^\mathsf{T}\boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}\boldsymbol{\mu}_{\mathbf{w}_i} - \boldsymbol{\mu}_{\mathbf{w}_i}^\mathsf{T}\boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}\mathbf{w}_i\right\} \quad \text{(A.71)}$$

setting $\hat{\boldsymbol{\Sigma}}_{\mathbf{w}_i}^{-1} = \sum_{n=1}^{N}\{\mathbf{z}_{i,n}\mathbf{z}_{i,n}\}\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}$, and in the knowledge that $\mathbb{I} = \hat{\boldsymbol{\Sigma}}_{\mathbf{w}_i}^{-1}\hat{\boldsymbol{\Sigma}}_{\mathbf{w}_i}$, the following log normal form can be reached

$$\ln p(\mathbf{w}_i^{\text{new}}) \propto -\frac{1}{2}\left\{\mathbf{w}_i^\mathsf{T}\hat{\boldsymbol{\Sigma}}_{\mathbf{w}_i}^{-1} - \mathbf{w}_i^\mathsf{T}\hat{\boldsymbol{\Sigma}}_{\mathbf{w}_i}^{-1}\hat{\boldsymbol{\Sigma}}_{\mathbf{w}_i}\left(\sum_{n=1}^{N}\mathbf{z}_{i,n}\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{x}}_n + \boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}\boldsymbol{\mu}_{\mathbf{w}_i}\right) - \left(\sum_{n=1}^{N}\mathbf{z}_{i,n}\tilde{\mathbf{x}}_n^\mathsf{T}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\mu}_{\mathbf{w}_i}^\mathsf{T}\boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}\right)\hat{\boldsymbol{\Sigma}}_{\mathbf{w}_i}\hat{\boldsymbol{\Sigma}}_{\mathbf{w}_i}^{-1}\mathbf{w}_i\right\} \quad \text{(A.72)}$$

Therefore, a new sample of $\mathbf{w}_i$ can be drawn from the conditional distribution defined by

$$\mathbf{w}_i^{\text{new}} \propto \mathcal{N}\left(\hat{\boldsymbol{\mu}}_{\mathbf{w}_i}, \hat{\boldsymbol{\Sigma}}_{\mathbf{w}_i}\right) \tag{A.73}$$

where

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{w}_i} = \left(\sum_{n=1}^{N}\{\mathbf{z}_{i,n}\mathbf{z}_{i,n}\}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}\right)^{-1} \tag{A.74}$$

and

$$\hat{\boldsymbol{\mu}}_{\mathbf{w}_i} = \hat{\boldsymbol{\Sigma}}_{\mathbf{w}_i}(\boldsymbol{\Sigma}^{-1}\sum_{n=1}^{N}\mathbf{z}_{i,n}\tilde{\mathbf{x}}_n + \boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}\boldsymbol{\mu}_{\mathbf{w}_i}) \tag{A.75}$$

Since each column of $\mathbf{W}$ is sampled independently, two methods of sampling can be used. Either: new columns are used in the following column updates, or the columns of the weight matrix are all sampled using the previous value for $\mathbf{W}$ before finally updating the full matrix.

Sample z

Collecting the terms of the log joint pertaining to $\mathbf{z}$

$$\ln p(\mathbf{z}_n^{\text{new}}) \propto -\frac{1}{2}\sum_{n=1}^{N}\left\{(\mathbf{x}_n - (\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu}))^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_n - (\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu}))\right\} - \frac{1}{2}\sum_{n=1}^{N}\mathbf{z}_n^{\mathsf{T}}\mathbf{z}_n \tag{A.76}$$

which can be written as

$$\ln p(\mathbf{z}_n^{\text{new}}) \propto -\frac{1}{2}\sum_{n=1}^{N}\left\{(\mathbf{x}_n - \boldsymbol{\mu} - \mathbf{W}\mathbf{z}_n)^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_n - \boldsymbol{\mu} - \mathbf{W}\mathbf{z}_n) + \mathbf{z}_n^{\mathsf{T}}\mathbb{I}\mathbf{z}_n\right\} \tag{A.77}$$

Expanding and ignoring any terms not containing $\mathbf{z}_n$,

$$\begin{aligned}\ln p(\mathbf{z}_n^{\text{new}}) \propto -\frac{1}{2}\sum_{n=1}^{N}\{&-\mathbf{x}_n^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{W}\mathbf{z}_n - \mathbf{z}_n^{\mathsf{T}}\mathbf{W}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{x}_n \\ &+\mathbf{z}_n^{\mathsf{T}}\mathbf{W}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \mathbf{z}_n^{\mathsf{T}}\mathbf{W}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{W}\mathbf{z}_n + \mathbf{z}_n^{\mathsf{T}}\mathbb{I}\mathbf{z}_n\}\end{aligned} \tag{A.78}$$

letting $\hat{\boldsymbol{\Sigma}}_{\mathbf{z}}^{-1} = (\mathbf{W}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{W} + \mathbb{I})$, and in the knowledge that $\mathbb{I} = \hat{\boldsymbol{\Sigma}}_{\mathbf{z}}^{-1}\hat{\boldsymbol{\Sigma}}_{\mathbf{z}}$, the log form

of a Gaussian can be reached

$$\ln p(\mathbf{z}_n^{\text{new}}) \propto -\frac{1}{2}\sum_{n=1}^{N}\Big\{\mathbf{z}_n^{\mathsf{T}}\hat{\boldsymbol{\Sigma}}_{\mathbf{z}}^{-1}\mathbf{z}_n - (\mathbf{x}_n-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{W}\hat{\boldsymbol{\Sigma}}_{\mathbf{z}}\hat{\boldsymbol{\Sigma}}_{\mathbf{z}}^{-1}\mathbf{z}_n - \mathbf{z}_n^{\mathsf{T}}\hat{\boldsymbol{\Sigma}}_{\mathbf{z}}^{-1}\hat{\boldsymbol{\Sigma}}_{\mathbf{z}}\mathbf{W}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_n-\boldsymbol{\mu})\Big\} \quad \text{(A.79)}$$

Therefore, a new sample of $\mathbf{z}$ can be drawn from the conditional distribution defined by

$$\mathbf{z}_n^{\text{new}} \sim \mathcal{N}\left(\hat{\boldsymbol{\mu}}_{\mathbf{z}}, \hat{\boldsymbol{\Sigma}}_{\mathbf{z}}\right) \quad \text{(A.80)}$$

where

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{z}} = (\mathbf{W}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{W} + \mathbb{I})^{-1} \quad \text{(A.81)}$$

and

$$\hat{\boldsymbol{\mu}}_{\mathbf{z}} = \hat{\boldsymbol{\Sigma}}_{\mathbf{z}}\mathbf{W}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{x}_n \quad \text{(A.82)}$$

A.6 Bayesian CCA: Variational Bayes derivation

This appendix provides a full derivation for the update equations provided in Algorithm 4. These update equations differ with those presented by Wang [198] in that these equations assume independent columns of $\mathbf{W}$, matching the definition in [199], rather than independent rows. Either definition (row or column) is suitable however one common approach was chosen here. Unlike [198], this derivation also includes generic priors on the model parameters.

Given the model defined in Figure 6.1, the joint for Bayesian CCA can be defined as

$$p(\mathbf{x}, \mathbf{z}, \mathbf{W}, \boldsymbol{\Sigma}, \boldsymbol{\mu}) = p(\mathbf{x}|\mathbf{z}, \mathbf{W}, \boldsymbol{\Sigma}, \boldsymbol{\mu})p(\mathbf{z})p(\boldsymbol{\Sigma})p(\boldsymbol{\mu})p(\mathbf{W}) \tag{A.83}$$

Assuming independent columns of W, as stated in the model, and using precision $\boldsymbol{\Psi}$, the joint can be written as:

$$p(\mathbf{x}|\mathbf{z}, \mathbf{W}, \boldsymbol{\Psi}, \boldsymbol{\mu})p(\mathbf{z})p(\boldsymbol{\Psi})p(\boldsymbol{\mu}) \prod_{i=1}^{d} p(\mathbf{w}_i) \tag{A.84}$$

with the same components of the log joint as shown in the A.5 derivation.

The general premise of VI is to select a suitable approximation from a tractable family of distributions and try to make the approximation as close to the true (intractable) posterior as possible, usually by minimising the KL divergence. Assuming the surrogate posterior is determined by some free parameters, this problem reduces the inference to an optimisation problem.

The following subsections summarise the derivation of the individual parameter updates needed to perform this optimisation. Even though a specified order is given in Algorithm 4, the VI scheme means updating the parameters can be conducted in any order, as long as the order remains fixed.

Using a mean field approximation, the surrogate posterior is assumed to take the following factorised form

$$q(\mathbf{z}, \mathbf{W}, \boldsymbol{\Psi}, \boldsymbol{\mu}) = q(\mathbf{z})q(\boldsymbol{\Psi})q(\boldsymbol{\mu}) \prod_{i=1}^{d} q(\mathbf{w}_i) \tag{A.85}$$

where

$$q(\mathbf{z}_n) \sim \mathcal{N}\left(\mathbf{z}_n|\breve{\boldsymbol{\mu}}_{\mathbf{z}}, \breve{\boldsymbol{\Sigma}}_{\mathbf{z}}\right) \tag{A.86}$$

$$q(\boldsymbol{\Psi}) \sim \mathcal{W}\left(\boldsymbol{\Psi}|\breve{\mathbf{K}}^{-1}, \breve{\nu}\right) \tag{A.87}$$

$$q(\mathbf{w}_i) \sim \mathcal{N}\left(\mathbf{w}_i|\breve{\boldsymbol{\mu}}_{\mathbf{w}_i}, \breve{\boldsymbol{\Sigma}}_{\mathbf{w}_i}\right) \tag{A.88}$$

$$q(\boldsymbol{\mu}) \sim \mathcal{N}\left(\boldsymbol{\mu}|\breve{\boldsymbol{\mu}}_{\boldsymbol{\mu}}, \breve{\boldsymbol{\Sigma}}_{\boldsymbol{\mu}}\right) \tag{A.89}$$

Given this factorised form, coordinate ascent VI (CAVI) [173] updates for the parameters of the model can be found using

$$q^\star(\phi_k) = \mathbb{E}_{\phi_-}[\mathcal{J}] = \mathbb{E}_{\phi_-}[p(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta})] \tag{A.90}$$

where $\phi = \{\mathbf{z}, \boldsymbol{\theta}\}$, $\boldsymbol{\theta} = \{\boldsymbol{\Psi}, \boldsymbol{\mu}, \mathbf{W}\}$, ϕ_- denotes all elements of ϕ except the kth element being updated and $q^\star(\phi_k)$ refers to the updated surrogate posterior.

Working with the log of the joint, this can also be expressed as

$$\ln q^\star(\phi_k) = \mathbb{E}_{\phi_-}[\ln p(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta})] + \text{const} \tag{A.91}$$

Update z

Collecting the terms of the log joint pertaining to $\mathbf{z}$ and substituting into Equation (A.91),

$$\ln q^\star(\mathbf{z}_n) \propto \mathbb{E}_{\phi_-}\left[-\frac{1}{2}\sum_{n=1}^{N}\left\{(\mathbf{x}_n - (\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu}))^\mathsf{T}\boldsymbol{\Psi}(\mathbf{x}_n - (\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu}))\right\} - \frac{1}{2}\sum_{n=1}^{N}\mathbf{z}_n^\mathsf{T}\mathbf{z}_n\right] \tag{A.92}$$

The expectation $\mathbb{E}_{\phi_{\text{-}}}[\bullet]$ is rewritten using $\langle\bullet\rangle_{\phi_{\text{-}}}$, such that

$$\ln q^{\star}(\mathbf{z}_n) \propto \left\langle -\frac{1}{2}\sum_{n=1}^{N}\left\{(\mathbf{x}_n - (\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu}))^{\mathsf{T}}\boldsymbol{\Psi}(\mathbf{x}_n - (\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu})) + \mathbf{z}_n^{\mathsf{T}}\mathbf{z}_n\right\}\right\rangle_{\mathbf{W},\,\boldsymbol{\mu},\,\boldsymbol{\Psi}} \tag{A.93}$$

The subscript notation for the expectation is written once at the start of each update derivation with the relevant parameters and then omitted for clarity.

Expanding and ignoring terms not containing $\mathbf{z}_n$, similar to the Gibbs sampling derivation, this reduces to

$$\ln q^{\star}(\mathbf{z}_n) \propto \left\langle -\frac{1}{2}\sum_{n=1}^{N}\left\{\mathbf{z}_n^{\mathsf{T}}\left(\mathbf{W}^{\mathsf{T}}\boldsymbol{\Psi}\mathbf{W} + \mathbb{I}\right)\mathbf{z}_n - (\mathbf{x}_n - \boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Psi}\mathbf{W}\mathbf{z}_n \right.\right. \\ \left.\left. -\mathbf{z}_n^{\mathsf{T}}\mathbf{W}^{\mathsf{T}}\boldsymbol{\Psi}(\mathbf{x}_n - \boldsymbol{\mu})\right\}\right\rangle \tag{A.94}$$

Setting $\breve{\boldsymbol{\Sigma}}_{\mathbf{z}}^{-1} = \langle\mathbf{W}^{\mathsf{T}}\boldsymbol{\Psi}\mathbf{W}\rangle + \mathbb{I}$, and given $\mathbb{I} = \breve{\boldsymbol{\Sigma}}_{\mathbf{z}}^{-1}\breve{\boldsymbol{\Sigma}}_{\mathbf{z}}$, the log form of a Gaussian can be reached

$$\ln q^{\star}(\mathbf{z}_n) \propto -\frac{1}{2}\sum_{n=1}^{N}\left\{\mathbf{z}_n^{\mathsf{T}}\breve{\boldsymbol{\Sigma}}_{\mathbf{z}}^{-1}\mathbf{z}_n - (\mathbf{x}_n - \langle\boldsymbol{\mu}\rangle)^{\mathsf{T}}\langle\boldsymbol{\Psi}\rangle\langle\mathbf{W}\rangle\breve{\boldsymbol{\Sigma}}_{\mathbf{z}}\breve{\boldsymbol{\Sigma}}_{\mathbf{z}}^{-1}\mathbf{z}_n \right. \\ \left. -\mathbf{z}_n^{\mathsf{T}}\breve{\boldsymbol{\Sigma}}_{\mathbf{z}}^{-1}\breve{\boldsymbol{\Sigma}}_{\mathbf{z}}\langle\mathbf{W}\rangle^{\mathsf{T}}\langle\boldsymbol{\Psi}\rangle(\mathbf{x}_n - \langle\boldsymbol{\mu}\rangle)\right\} \tag{A.95}$$

Exponentiating, the surrogate posterior of $\mathbf{z}$ therefore has the following Gaussian form

$$q^{\star}(\mathbf{z}_n) \sim \mathcal{N}\left(\breve{\boldsymbol{\mu}}_{\mathbf{z}}, \breve{\boldsymbol{\Sigma}}_{\mathbf{z}}\right) \tag{A.96}$$

where

$$\breve{\boldsymbol{\Sigma}}_{\mathbf{z}} = \left(\langle\mathbf{W}^{\mathsf{T}}\boldsymbol{\Psi}\mathbf{W}\rangle + \mathbb{I}\right)^{-1} \tag{A.97}$$

and

$$\breve{\boldsymbol{\mu}}_{\mathbf{z}} = \breve{\boldsymbol{\Sigma}}_{\mathbf{z}}\langle\mathbf{W}\rangle^{\mathsf{T}}\langle\boldsymbol{\Psi}\rangle(\mathbf{x}_n - \langle\boldsymbol{\mu}\rangle) \tag{A.98}$$

Update $\boldsymbol{\Psi}$

Collecting the terms of the log joint pertaining to $\boldsymbol{\Psi}$ and substituting into Equation (A.91),

$$\ln q^{\star}(\boldsymbol{\Psi}) \propto \left\langle -\frac{1}{2}\sum_{n=1}^{N}\left\{\ln(|\boldsymbol{\Psi}|^{-1}) + (\mathbf{x} - (\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu}))^{\mathsf{T}}\boldsymbol{\Psi}(\mathbf{x} - (\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu}))\right\} + \frac{\nu_0 - D + 1}{2}\ln|\boldsymbol{\Psi}| - \frac{1}{2}\mathrm{tr}(\boldsymbol{\Psi}\mathbf{K}_0)\right\rangle_{\mathbf{W},\ \boldsymbol{\mu},\ \mathbf{z}_n} \tag{A.99}$$

Simplifying and rearranging terms, a log Wishart form can be reached such that

$$\ln q^{\star}(\boldsymbol{\Psi}) \propto -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Psi}\left[\mathbf{K}_0 + \sum_{n=1}^{N}\left\langle(\mathbf{x} - (\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu}))(\mathbf{x} - (\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu}))^{\mathsf{T}}\right\rangle\right]\right) + \frac{\nu_0 - D + 1 + N}{2}\ln|\boldsymbol{\Psi}| \tag{A.100}$$

Exponentiating, the surrogate posterior of $\boldsymbol{\Psi}$ is then defined by,

$$q^{\star}(\boldsymbol{\Psi}) \sim \mathcal{W}\left(\breve{\mathbf{K}}^{-1}, \breve{\nu}\right) \tag{A.101}$$

where

$$\breve{\mathbf{K}} = \mathbf{K}_0 + \sum_{n=1}^{N}\left\langle(\mathbf{x}_n - \boldsymbol{\mu} - \mathbf{W}\mathbf{z}_n)(\mathbf{x}_n - \boldsymbol{\mu} - \mathbf{W}\mathbf{z}_n)^{\mathsf{T}}\right\rangle \tag{A.102}$$

and

$$\breve{\nu} = \nu_0 + N \tag{A.103}$$

Update $\boldsymbol{\mu}$

Collecting the terms of the log joint pertaining to $\boldsymbol{\mu}$ and substituting into Equation (A.91),

$$\ln q^{\star}(\boldsymbol{\mu}) \propto \left\langle -\frac{1}{2}\sum_{n=1}^{N}\left\{(\mathbf{x}-(\mathbf{W}\mathbf{z}_n+\boldsymbol{\mu}))^{\mathsf{T}}\boldsymbol{\Psi}(\mathbf{x}-(\mathbf{W}\mathbf{z}_n+\boldsymbol{\mu}))\right\} - \frac{1}{2}\boldsymbol{\mu}^{\mathsf{T}}\boldsymbol{\Sigma}_{\boldsymbol{\mu}}^{-1}\boldsymbol{\mu} \right. \\ \left. -\frac{1}{2}\boldsymbol{\mu}_{\boldsymbol{\mu}}^{\mathsf{T}}\boldsymbol{\Sigma}_{\boldsymbol{\mu}}^{-1}\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\mu}^{\mathsf{T}}\boldsymbol{\Sigma}_{\boldsymbol{\mu}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\mu}} \right\rangle_{\mathbf{W},\ \mathbf{z}_n,\ \boldsymbol{\Psi}} \tag{A.104}$$

Simplifying,

$$\ln q^{\star}(\boldsymbol{\mu}) \propto -\frac{1}{2}\left\{\boldsymbol{\mu}^{\mathsf{T}}(N\langle\boldsymbol{\Psi}\rangle+\boldsymbol{\Sigma}_{\boldsymbol{\mu}}^{-1})\boldsymbol{\mu} - \boldsymbol{\mu}^{\mathsf{T}}\sum_{n=1}^{N}\langle\boldsymbol{\Psi}\rangle\left(\mathbf{x}_n-\langle\mathbf{W}\rangle\langle\mathbf{z}_n\rangle\right) \right. \\ \left. -\sum_{n=1}^{N}(\mathbf{x}_n-\langle\mathbf{W}\rangle\langle\mathbf{z}_n\rangle)^{\mathsf{T}}\langle\boldsymbol{\Psi}\rangle\boldsymbol{\mu} + \boldsymbol{\mu}_{\boldsymbol{\mu}}^{\mathsf{T}}\boldsymbol{\Sigma}_{\boldsymbol{\mu}}^{-1}\boldsymbol{\mu} + \boldsymbol{\mu}^{\mathsf{T}}\boldsymbol{\Sigma}_{\boldsymbol{\mu}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\mu}}\right\} \tag{A.105}$$

letting $\breve{\boldsymbol{\Sigma}}_{\mu}^{-1} = N\langle\boldsymbol{\Psi}\rangle + \boldsymbol{\Sigma}_{\boldsymbol{\mu}}^{-1}$, and given $\mathbb{I} = \breve{\boldsymbol{\Sigma}}_{\mu}^{-1}\breve{\boldsymbol{\Sigma}}_{\mu}$,

$$\ln q^{\star}(\boldsymbol{\mu}) \propto -\frac{1}{2}\left\{\boldsymbol{\mu}^{\mathsf{T}}\breve{\boldsymbol{\Sigma}}_{\mu}^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}^{\mathsf{T}}\breve{\boldsymbol{\Sigma}}_{\mu}^{-1}\breve{\boldsymbol{\Sigma}}_{\mu}\left(\sum_{n=1}^{N}\langle\boldsymbol{\Psi}\rangle\left(\mathbf{x}_n-\langle\mathbf{W}\rangle\langle\mathbf{z}_n\rangle\right)+\boldsymbol{\Sigma}_{\boldsymbol{\mu}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\mu}}\right) \right. \\ \left. -\left(\sum_{n=1}^{N}(\mathbf{x}_n-\langle\mathbf{W}\rangle\langle\mathbf{z}_n\rangle)^{\mathsf{T}}\langle\boldsymbol{\Psi}\rangle+\boldsymbol{\mu}_{\boldsymbol{\mu}}^{\mathsf{T}}\boldsymbol{\Sigma}_{\boldsymbol{\mu}}^{-1}\right)\breve{\boldsymbol{\Sigma}}_{\mu}\breve{\boldsymbol{\Sigma}}_{\mu}^{-1}\boldsymbol{\mu}\right\} \tag{A.106}$$

Exponentiating, the surrogate posterior of $\boldsymbol{\mu}$ thus has the following form

$$q^{\star}(\boldsymbol{\mu}) \sim \mathcal{N}\left(\breve{\boldsymbol{\mu}}_{\boldsymbol{\mu}}, \breve{\boldsymbol{\Sigma}}_{\boldsymbol{\mu}}\right) \tag{A.107}$$

where

$$\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\mu}} = (N\langle\boldsymbol{\Psi}\rangle + \boldsymbol{\Sigma}_{\mu}^{-1})^{-1} \tag{A.108}$$

and

$$\breve{\boldsymbol{\mu}}_{\boldsymbol{\mu}} = \breve{\boldsymbol{\Sigma}}_{\mu} \left(\langle \boldsymbol{\Psi} \rangle \sum_{n=1}^{N} (\mathbf{x}_n - \langle \mathbf{W} \rangle \langle \mathbf{z}_n \rangle) + \boldsymbol{\Sigma}_{\mu}^{-1} \boldsymbol{\mu}_{\mu} \right) \tag{A.109}$$

Update $\mathbf{w}_i$

Collecting the terms of the log joint pertaining to $\mathbf{w}_i$ and substituting into Equation (A.91),

$$\ln q^{\star}(\mathbf{w}_i) \propto \left\langle -\frac{1}{2} \sum_{n=1}^{N} \left\{ (\mathbf{x} - (\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu}))^{\mathsf{T}} \boldsymbol{\Psi} (\mathbf{x} - (\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu})) \right\} - \frac{1}{2} \mathbf{w}_i^{\mathsf{T}} \boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1} \mathbf{w}_i \right. \\ \left. -\frac{1}{2} \boldsymbol{\mu}_{\mathbf{w}_i}^{\mathsf{T}} \boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1} \mathbf{w}_i - \frac{1}{2} \mathbf{w}_i^{\mathsf{T}} \boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1} \boldsymbol{\mu}_{\mathbf{w}_i} \right\rangle_{\boldsymbol{\mu},\ \mathbf{z}_n,\ \boldsymbol{\Psi}} \tag{A.110}$$

To construct the update for a single column, the full weight matrix must be separated into the column of interest and the remaining columns. This is achieved by defining

$$\tilde{\mathbf{x}}_n = \mathbf{x}_n - \boldsymbol{\mu} - \mathbf{w}_{\_i} \mathbf{z}_{\_i,n} \tag{A.111}$$

where $\mathbf{w}_{\_i}$ corresponds to all columns except the ith column of interest and $\mathbf{z}_{\_i,n}$ corresponds to all rows of $\mathbf{z}_n$ except the ith row. Using this definition, Equation A.110 becomes

$$\ln q^{\star}(\mathbf{w}_i) \propto \left\langle -\frac{1}{2} \sum_{n=1}^{N} \left\{ (\tilde{\mathbf{x}}_n - \mathbf{w}_i \mathbf{z}_{i,n})^{\mathsf{T}} \boldsymbol{\Psi} (\tilde{\mathbf{x}}_n - \mathbf{w}_i \mathbf{z}_{i,n}) \right\} - \frac{1}{2} \mathbf{w}_i^{\mathsf{T}} \boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1} \mathbf{w}_i \right. \\ \left. -\frac{1}{2} \boldsymbol{\mu}_{\mathbf{w}_i}^{\mathsf{T}} \boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1} \mathbf{w}_i - \frac{1}{2} \mathbf{w}_i^{\mathsf{T}} \boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1} \boldsymbol{\mu}_{\mathbf{w}_i} \right\rangle \tag{A.112}$$

As $\mathbf{z}_{i,n}$ is scalar, it can be manipulated as one. Rearranging and combining terms,

this gives

$$\ln q^{\star}(\mathbf{w}_i) \propto -\frac{1}{2}\left\langle \mathbf{w}_i^{\mathsf{T}}\left(\boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}+\sum_{n=1}^{N}\mathbf{z}_{i,n}\mathbf{z}_{i,n}\boldsymbol{\Psi}\right)\mathbf{w}_i - \mathbf{w}_i^{\mathsf{T}}\left(\sum_{n=1}^{N}\mathbf{z}_{i,n}\boldsymbol{\Psi}\tilde{\mathbf{x}}_n+\boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}\boldsymbol{\mu}_{\mathbf{w}_i}\right) - \left(\sum_{n=1}^{N}\tilde{\mathbf{x}}_n^{\mathsf{T}}\boldsymbol{\Psi}\mathbf{z}_{i,n}+\boldsymbol{\mu}_{\mathbf{w}_i}^{\mathsf{T}}\boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}\right)\mathbf{w}_i\right\rangle \quad \text{(A.113)}$$

Considering the expectations,

$$\ln q^{\star}(\mathbf{w}_i) \propto -\frac{1}{2}\left\{ \mathbf{w}_i^{\mathsf{T}}\left(\boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}+\sum_{n=1}^{N}\left\langle\mathbf{z}_{i,n}\mathbf{z}_{i,n}\right\rangle\left\langle\boldsymbol{\Psi}\right\rangle\right)\mathbf{w}_i - \mathbf{w}_i^{\mathsf{T}}\left(\sum_{n=1}^{N}\left\langle\mathbf{z}_{i,n}\right\rangle\left\langle\boldsymbol{\Psi}\right\rangle\tilde{\mathbf{x}}_n+\boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}\boldsymbol{\mu}_{\mathbf{w}_i}\right) - \left(\sum_{n=1}^{N}\tilde{\mathbf{x}}_n^{\mathsf{T}}\left\langle\boldsymbol{\Psi}\right\rangle\left\langle\mathbf{z}_{i,n}\right\rangle+\boldsymbol{\mu}_{\mathbf{w}_i}^{\mathsf{T}}\boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}\right)\mathbf{w}_i\right\} \quad \text{(A.114)}$$

letting $\breve{\boldsymbol{\Sigma}}_{\mathbf{w}_i}^{-1} = \left(\boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}+\sum_{n=1}^{N}\left\langle\mathbf{z}_{i,n}\mathbf{z}_{i,n}\right\rangle\left\langle\boldsymbol{\Psi}\right\rangle\right)$, and as $\mathbb{I} = \breve{\boldsymbol{\Sigma}}_{\mathbf{w}_i}^{-1}\breve{\boldsymbol{\Sigma}}_{\mathbf{w}_i}$,

$$q^{\star}(\mathbf{w}_i) \sim \mathcal{N}\left(\breve{\boldsymbol{\mu}}_{\mathbf{w}_i}, \breve{\boldsymbol{\Sigma}}_{\mathbf{w}_i}\right) \quad \text{(A.115)}$$

where

$$\breve{\boldsymbol{\Sigma}}_{\mathbf{w}_i} = (\sum_{n=1}^{N}\langle\mathbf{z}_{i,n}\mathbf{z}_{i,n}\rangle\langle\boldsymbol{\Psi}\rangle + \boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1})^{-1} \quad \text{(A.116)}$$

and

$$\breve{\boldsymbol{\mu}}_{\mathbf{w}_i} = \breve{\boldsymbol{\Sigma}}_{\mathbf{w}_i}\left(\langle\boldsymbol{\Psi}\rangle\sum_{n=1}^{N}\tilde{\mathbf{x}}_n\langle\mathbf{z}_{n,i}^{\mathsf{T}}\rangle) + \boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1}\boldsymbol{\mu}_{\mathbf{w}_i}\right) \quad \text{(A.117)}$$

Bibliography

[1] O. H Ammann, T von Karman, and G. B Woodruff. Tacoma Narrows Bridge - Failure of the Bridge. Informational Works, United States Federal Works Agency, Tacoma, WA, USA, 1941.

[2] J. P Norton. *An Introduction to Identification*. Dover Publications, Mineola, NY, USA, 1st ed., 2009.

[3] L Ljung. *System Identification: Theory for the User*. Prentice-Hall Information and System Sciences Series. Prentice Hall, Upper Saddle River, NJ, USA, 2nd ed., 2012.

[4] R. J Allemang and D. L Brown. Experimental Modal Analysis Methods. In R Allemang and P Avitabile, editors, *Handbook of Experimental Structural Dynamics*. Springer, New York, NY, USA, 1st ed., 2020.

[5] D. J Ewins. *Modal Testing: Theory, Practice and Application*. Number 10 in Mechanical Engineering Research Studies Engineering Dynamics Series. Research Studies Press, Baldock, Hertfordshire, England, 2nd ed., 2000.

[6] R Brincker and C. E Ventura. *Introduction to Operational Modal Analysis*. John Wiley & Sons, Ltd, Chichester, West Sussex, UK, 1st ed., 2015.

[7] R Brincker and P. H Kirkegaard. Special issue on Operational Modal Analysis. *Mechanical Systems and Signal Processing*, 24(5):1209–1212, 2010.

[8] C. R Farrar and K Worden. *Structural Health Monitoring: A Machine Learning Perspective*. John Wiley & Sons, Ltd, Chichester, West Sussex, UK, 1st ed., 2013.

[9] K Worden, L. A Bull, P Gardner, J Gosliga, T. J Rogers, E. J Cross, E Papatheou, W Lin, and N Dervilis. A brief introduction to recent developments in Population-Based Structural Health Monitoring. *Frontiers in Built Environment*, 6:146, 2020.

[10] D. J Wagg, P Gardner, R. J Barthorpe, and K Worden. On key technologies for realising digital twins for structural dynamics applications. In R Barthorpe, editor, *Model Validation and Uncertainty Quantification*, volume 3, pages 267–272, Reno, NV, USA, 2020. Springer.

[11] I. M Díaz, E Pereira, and P Reynolds. Integral resonant control scheme for cancelling human-induced vibrations in light-weight pedestrian structures. *Structural Control and Health Monitoring*, 19(1):55–69, 2012.

[12] C Jones, P Reynolds, and A Pavic. Vibration serviceability of stadia structures subjected to dynamic crowd loads: A literature review. *Journal of Sound and Vibration*, 330(8):1531–1566, 2011.

[13] B Peeters, J Lau, J Lanslots, and V. D Auweraer. Automatic Modal Analysis – Myth or Reality? *Sound and Vibration*, Computer-Aided Engineering, 2008.

[14] M. M Rosso, A Aloisio, J Parol, G. C Marano, and G Quaranta. Intelligent automatic operational modal analysis. *Mechanical Systems and Signal Processing*, 201:110669, 2023.

[15] J Li, T Bao, and C. E Ventura. An automated operational modal analysis algorithm and its application to concrete dams. *Mechanical Systems and Signal Processing*, 168:108707, 2022.

[16] C Rainieri, F Magalhaes, and F Ubertini. Automated operational modal analysis and its applications in structural health monitoring. *Shock and Vibration*, 2019:1–3, 2019.

[17] J Priou, S Gres, M Perrault, L Guerineau, and M Döhler. Automated uncertainty-based extraction of modal parameters from stabilization diagrams. In *9th International Operational Modal Analysis Conference*, Vancouver, BC, Canada, 2022.

[18] K Worden and G. R Tomlinson. *Nonlinearity in Structural Dynamics: Detection, Identification and Modelling*. CRC Press, Boca Raton, FL, USA, 1st ed., 2019.

[19] S.-K Au. *Operational Modal Analysis*. Springer Nature, Singapore, 2017.

[20] F Liu, J Wu, F Gu, and A. D Ball. An introduction of a robust OMA method: CoS-SSI and its performance evaluation through the simulation and a case study. *Shock and Vibration*, 2019:1–14, 2019.

[21] J. M. W Brownjohn, A Raby, S.-K Au, Z Zhu, X Wang, A Antonini, A Pappas, and D D'Ayala. Bayesian operational modal analysis of offshore rock lighthouses: Close modes, alignment, symmetry and uncertainty. *Mechanical Systems and Signal Processing*, 133:Kirk, 2019.

[22] J Zhang, K Maes, G De Roeck, E Reynders, C Papadimitriou, and G Lombaert. Optimal sensor placement for multi-setup modal analysis of structures. *Journal of Sound and Vibration*, 401:214–232, 2017.

[23] P. H Kirkegaard. *Optimal Design of Experiments for Parametric Identification of Civil Engineering Structures*. PhD thesis, Dept. of Building Technology and Structural Engineering, Aalborg University, Denmark, 1991.

[24] A Schulze, J Zierath, S.-E Rosenow, R Bockhahn, R Rachholz, and C Woernle. Optimal sensor placement for modal testing on wind turbines. *Journal of Physics: Conference Series*, 753:072031, 2016.

[25] R. C Smith. *Uncertainty Quantification: Theory, Implementation, and Applications*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2013.

[26] A. J Hughes, R. J Barthorpe, N Dervilis, C. R Farrar, and K Worden. A probabilistic risk-based decision framework for structural health monitoring. *Mechanical Systems and Signal Processing*, 150:107339, 2021.

[27] C. M Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, New York, NY, USA, 2006.

[28] M. D Hoffman, D. M Blei, C Wang, and J Paisley. Stochastic variational inference. *Journal of Machine Learning Research*, 14:1303–1347, 2013.

[29] C. C Kennedy and C. D. P Pancu. Use of Vectors in Vibration Measurement and Analysis. *Journal of the Aeronautical Sciences*, 14(11):603–625, 1947.

[30] J. M. M Silva and N. M. M Maia, editors. *Modal Analysis and Testing*. Springer, Dordrecht, Netherlands, 1999.

[31] N. M. M Maia and J. M. M Silva, editors. *Theoretical and Experimental Modal Analysis*. Number 9 in Mechanical Engineering Research Studies. Research Studies Press, Taunton, Somerset, England, 1997.

[32] J He and Z.-F Fu. *Modal Analysis*. Butterworth-Heinemann, Oxford, England, 2001.

[33] E Reynders. System Identification Methods for (Operational) Modal Analysis: Review and Comparison. *Archives of Computational Methods in Engineering*, 19(1):51–124, 2012.

[34] C Rainieri and G Fabbrocino. *Operational Modal Analysis of Civil Engineering Structures: An Introduction and Guide for Applications*. Springer, New York, 2014.

[35] G. H James, T. T Cao, V. A Fogt, R. L Wilson, and T. J Bartkowicz. Extraction of Modal Parameters From Spacecraft Flight Data. In T Proulx, editor, *Advanced Aerospace Applications*, volume 1, pages 307–319. Springer, New York, NY, USA, 2011.

[36] C Kocan. A comparative study on in-flight modal identification of an aircraft using time- and frequency-domain techniques. *Journal of Vibration and Control*, 26(21-22):1920–1934, 2020.

[37] J. V Covioli and G Coppotelli. On the operational modal analysis techniques for the estimate of modal parameters of aircraft structures during flying vibration tests. In *American Institute of Aeronautics and Astronautics (AIAA) Scitech 2020 Forum*, Orlando, FL, USA, 2020.

[38] F Pimenta, D Ribeiro, A Román, and F Magalhães. Modal properties of floating wind turbines: Analytical study and operational modal analysis of an utility-scale wind turbine. *Engineering Structures*, 301:117367, 2024.

[39] C Van Zijl, K Soal, R Volkmar, Y Govers, M Böswald, and A Bekker. The use of operational modal analysis and mode tracking for insight into polar vessel operations. *Marine Structures*, 79:103043, 2021.

[40] J Maeck, B Peeters, and G. D Roeck. Damage identification on the Z24 bridge using vibration monitoring. *Smart Materials and Structures*, 10(3):512–517, 2001.

[41] J Brownjohn, F Magalhaes, E Caetano, and A Cunha. Ambient vibration re-testing and operational modal analysis of the Humber Bridge. *Engineering Structures*, 32(8):2003–2018, 2010.

[42] M Döhler, F Hille, L Mevel, and W Rücker. Structural health monitoring with statistical methods during progressive damage test of S101 Bridge. *Engineering Structures*, 69:183–193, 2014.

[43] K Maes and G Lombaert. Monitoring railway bridge KW51 before, during, and after retrofitting. *Journal of Bridge Engineering*, 26(3):04721001, 2021.

[44] F Magalhães, A Cunha, and E Caetano. Vibration based structural health monitoring of an arch bridge: From automated OMA to damage detection. *Mechanical Systems and Signal Processing*, 28:212–228, 2012.

[45] E Reynders, D Degrauwe, G De Roeck, F Magalhães, and E Caetano. Combined experimental-operational modal testing of footbridges. *Journal of Engineering Mechanics*, 136(6):687–696, 2010.

[46] D Dooms, G Degrande, G De Roeck, and E Reynders. Finite element modelling of a silo based on experimental modal analysis. *Engineering Structures*, 28(4): 532–542, 2006.

[47] C Gentile and A Saisi. Ambient vibration testing of historic masonry towers for structural identification and damage assessment. *Construction and Building Materials*, 21(6):1311–1321, 2007.

[48] G Milani and F Clementi. Advanced seismic assessment of four masonry bell towers in Italy after operational modal analysis (OMA) identification. *International Journal of Architectural Heritage*, 15(1):157–186, 2021.

[49] F Ubertini, G Comanducci, and N Cavalagli. Vibration-based structural health monitoring of a historic bell-tower using output-only measurements and multivariate statistical analysis. *Structural Health Monitoring*, 15(4):438–457, 2016.

[50] K Zhou, Q.-S Li, L.-H Zhi, X.-L Han, and K Xu. Investigation of modal parameters of a 600-m-tall skyscraper based on two-year-long structural health monitoring data and five typhoons measurements. *Engineering Structures*, 274: 115162, 2023.

[51] B Peeters, H Van Der Auweraer, F Vanhollebeke, and P Guillaume. Operational modal analysis for estimating the dynamic properties of a stadium structure during a football game. *Shock and Vibration*, 14(4):283–303, 2007.

[52] F Magalhães, E Caetano, and Á Cunha. Operational modal analysis and finite element model correlation of the Braga Stadium suspended roof. *Engineering Structures*, 30(6):1688–1698, 2008.

[53] M Vanali and A Cigada. Long term operational modal analysis of a stadium grandstand to structural health monitoring purposes. In *IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems*, pages 103–109, Crema, Italy, 2009.

[54] G. R Darbre, C. A. M De Smet, and C Kraemer. Natural frequencies measured from ambient vibration response of the arch dam of Mauvoisin. *Earthquake Engineering & Structural Dynamics*, 29(5):577–586, 2000.

[55] B Peeters and G De Roeck. Stochastic system identification for operational modal analysis: A review. *Journal of Dynamic Systems, Measurement, and Control*, 123(4):659–667, 2001.

[56] J. W Cooley and J. W Tukey. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19(90):297–301, 1965.

[57] J. W Cooley. The re-discovery of the fast Fourier transform algorithm. *Mikrochimica Acta*, 93(1-6):33–45, 1987.

[58] R Pintelon and J Schoukens. *System Identification: A Frequency Domain Approach*. John Wiley & Sons, Ltd, Hoboken, N.J, USA, 2nd ed., 2012.

[59] A. W Phillips and R. J Allemang. Frequency Response Function Estimation. In R Allemang and P Avitabile, editors, *Handbook of Experimental Structural Dynamics*, pages 1–90. Springer, New York, NY, USA, 2020.

[60] 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, 2001.

[61] R Brincker, C. E Ventura, and P Andersen. Damping estimation by frequency domain decomposition. In *Proceedings of 29th International Modal Analysis Conference (IMAC-XIX)*, volume 1, pages 698–703, Orlando, Florida, 2001. Society for Experimental Mechanics.

[62] L Zhang, T Wang, and Y Tamura. A frequency–spatial domain decomposition (FSDD) method for operational modal analysis. *Mechanical Systems and Signal Processing*, 24(5):1227–1239, 2010.

[63] P Guillaume, P Verboven, and S Vanlanduit. Frequency-domain maximum likelihood identification of modal parameters with confidence intervals. In *Proceedings of ISMA2023 International Conference on Noise and Vibration Engineering*, pages 359–366, Leuven, Belgium, 1998.

[64] P Guillaume, P Verboven, S Vanlanduit, H Van der Auweraer, and B Peeters. A poly-reference implementation of the least-squares complex frequency-domain estimator. In *Proceedings of the 21st International Modal Analysis Conference (IMAC-XXI)*, volume 21, Kissimmee, FL, 2003. Society for Experimental Mechanics.

[65] B Peeters, G Lowet, H Van Der Auweraer, and J Leuridan. A New Procedure for Modal Parameter Estimation. *Sound and Vibration*, Structural Analysis, 2004.

[66] B Peeters, H Van Der Auweraer, P Guillaume, and J Leuridan. The PolyMAX frequency-domain method: A new standard for modal parameter estimation? *Shock and Vibration*, 11(3-4):395–409, 2004.

[67] S. D. R Amador and R Brincker. A robust poly-reference frequency-domain identification method to extract dynamic properties from vibration data. *Communications Engineering*, 2(1):73, 2023.

[68] G. H James III, T. G Carrie, and J. P Lauffer. The natural excitation technique (NExT) for modal parameter extraction From operating wind turbines. Technical Report SAND92-1666, Sandia National Laboratories, United States Department for Energy, Albuquerque, NM, USA, 1993.

[69] J.-N Juang and R. S Pappa. An eigensystem realization algorithm for modal parameter identification and model reduction. *Journal of Guidance, Control, and Dynamics*, 8(5):620–627, 1985.

[70] H Vold, J Kundrat, G. T Rocklin, and R Russell. A multi-input modal estimation algorithm for mini-computers. In *SAE Transactions*, pages 815–821. JSTOR, 1982.

[71] S. R Ibrahim and E. C Mikulcik. The experimental determination of vibration parameters from time responses. In *46th Shock and Vibration Symposium*, pages 187–196, San Diego, California, USA, 1975.

[72] M Ghalishooyan and A Shooshtari. Operational modal analysis techniques and their theoretical and practical aspects: A comprehensive review and introduction. In *Proceedings of the 6th International Operational Modal Analysis Conference (IOMAC)*, Gijon, Spain, 2015.

[73] P Andersen. *Identification of Civil Engineering Structures Using Vector ARMA Models*. PhD thesis, Department of Building Technology and Structural Engineering, Aalborg University, Denmark, 1997.

[74] Y. L Pi and N. C Mickleborough. Modal identification of vibrating structures using ARMA model. *Journal of Engineering Mechanics*, 115(10):2232–2250, 1989.

[75] P Andersen, R Brincker, and P. H Kirkegaard. Theory of covariance equivalent ARMAV models of civil engineering structures. In *Proceedings of the 14th International Modal Analysis Conference (IMAC-XIV)*, Dearborn, Michigan, USA, 1996. Society for Experimental Mechanics.

[76] M Bertha and J.-C Golinval. Identification of non-stationary dynamical systems using multivariate ARMA models. *Mechanical Systems and Signal Processing*, 88:166–179, 2017.

[77] L Ljung. Prediction error estimation methods. *Circuits, Systems, and Signal Processing*, 21(1):11–21, 2002.

[78] T Söderström and P Stoica. Instrumental variable methods for system identification. *Circuits, Systems, and Signal Processing*, 21(1):1–9, 2002.

[79] K Petsounis and S Fassois. Parametric time-domain methods for the identification of vibrating structures - A critical comparison and assessment. *Mechanical Systems and Signal Processing*, 15(6):1031–1060, 2001.

[80] J.-F Cardoso. Blind signal separation: Statistical principles. *Proceedings of the IEEE*, 86(10):2009–2025, Oct./1998.

[81] F Poncelet, G Kerschen, J.-C Golinval, and D Verhelst. Output-only modal analysis using blind source separation techniques. *Mechanical Systems and Signal Processing*, 21(6):2335–2358, 2007.

[82] W Zhou and D Chelidze. Blind source separation based vibration mode identification. *Mechanical Systems and Signal Processing*, 21(8):3072–3087, 2007.

[83] J Antoni, L Garibaldi, S Marchesiello, and M Sidhamed. New separation techniques for output-only modal analysis. *Shock and Vibration*, 11(3-4): 227–242, 2004.

[84] J Antoni. Blind separation of vibration components: Principles and demonstrations. *Mechanical Systems and Signal Processing*, 19(6):1166–1180, 2005.

[85] J Antoni and S Braun. Special issue on blind source separation: Editorial. *Mechanical Systems and Signal Processing*, 19(6):1163–1165, 2005.

[86] G Kerschen, F Poncelet, and J.-C Golinval. Physical interpretation of independent component analysis in structural dynamics. *Mechanical Systems and Signal Processing*, 21(4):1561–1575, 2007.

[87] S McNeill and D Zimmerman. A framework for blind modal identification using joint approximate diagonalization. *Mechanical Systems and Signal Processing*, 22(7):1526–1548, 2008.

[88] Y Xu, J. M Brownjohn, and D Hester. Enhanced sparse component analysis for operational modal identification of real-life bridge structures. *Mechanical Systems and Signal Processing*, 116:585–605, 2019.

[89] X.-J Yao, T.-H Yi, and S.-W Zhao. Blind modal identification for decentralized sensor network by modified sparse component analysis in frequency-domain subspace. *Engineering Structures*, 269:114794, 2022.

[90] I. A Hernández-González, E García-Macías, G Costante, and F Ubertini. AI-driven blind source separation for fast operational modal analysis of structures. *Mechanical Systems and Signal Processing*, 211:111267, 2024.

[91] A Sadhu, S Narasimhan, and J Antoni. A review of output-only structural mode identification literature employing blind source separation methods. *Mechanical Systems and Signal Processing*, 94:415–431, 2017.

[92] A Di. Matteo, C Masnata, S Russotto, C Bilello, and A Pirrotta. A novel identification procedure from ambient vibration data. *Meccanica*, 56(4):797–812, 2021.

[93] S Russotto, A Di Matteo, C Masnata, and A Pirrotta. OMA: From research to engineering applications. In C Rainieri, G Fabbrocino, N Caterino, F Ceroni, and M. A Notarangelo, editors, *Civil Structural Health Monitoring*, volume 156, pages 903–920. Springer International Publishing, Cham, 2021.

[94] P Van Overschee and B De Moor. Subspace algorithms for the stochastic identification problem. *Automatica*, 29(3):649–660, 1993.

[95] P Van Overschee and B De Moor. *Subspace Identification for Linear Systems*. Springer, Boston, MA, USA, 1996.

[96] P Kirkegaard and P Andersen. State space identification of civil engineering structures from output measurements. In *Proceedings of the 15th International Modal Analysis Conference (IMAC-XV)*, volume 1, pages 889–895, Florida, USA, 1997. Society for Experimental Mechanics.

[97] B Peeters and G De Roeck. Stochastic subspace system identification of a steel transmitter mast. In *Proceedings of the International Modal Analysis Conference - IMAC*, volume 1, pages 130–136, California, US., 1998.

[98] L Hermans and H Van Der Auweraer. Modal testing and analysis of structures under operational conditions: Industrial applications. *Mechanical Systems and Signal Processing*, 13(2):193–216, 1999.

[99] B Li, H Cai, X Mao, J Huang, and B Luo. Estimation of CNC machine–tool dynamic parameters based on random cutting excitation through operational modal analysis. *International Journal of Machine Tools and Manufacture*, 71: 26–40, 2013.

[100] E Neu, F Janser, A. A Khatibi, C Braun, and A. C Orifici. Operational modal analysis of a wing excited by transonic flow. *Aerospace Science and Technology*, 49:73–79, 2016.

[101] J. A Jimenez Capilla, S.-K Au, J. M. W Brownjohn, and E Hudson. Ambient vibration testing and operational modal analysis of monopole telecoms structures. *Journal of Civil Structural Health Monitoring*, 11(4):1077–1091, 2021.

[102] F Pimenta and F Magalhães. First Results from Operational Modal Analysis of a Floating Offshore Wind Turbine. In M. P Limongelli, P. F Giordano,

S Quqa, C Gentile, and A Cigada, editors, *Experimental Vibration Analysis for Civil Engineering Structures*, volume 432, pages 111–120. Springer Nature, Switzerland, 2023.

[103] T Katayama. *Subspace Methods for System Identification: A Realisation Approach.* Communications and Control Engineering. Springer, Berlin, London, 2005.

[104] L Zhang, R Brincker, and P Andersen. An overview of operational modal analysis: Major development and issues. In *Proceedings of the 1st International Operational Modal Analysis Conference (IOMAC)*, pages 179–190, Copenhagen, Denmark, 2005.

[105] B Peeters and G De Roeck. Reference-based stochastic subspace identification for output-only modal analysis. *Mechanical Systems and Signal Processing*, 13 (6):855–878, 1999.

[106] M Goursat, M Döhler, L Mevel, and P Andersen. Crystal Clear SSI for Operational Modal Analysis of Aerospace Vehicles. In T Proulx, editor, *Structural Dynamics, Volume 3*, Conference Proceedings of the Society for Experimental Mechanics Series, pages 1421–1430, New York, NY, 2011. Springer.

[107] Q Wei, L Shen, B Kövesdi, L Dunai, and M Cao. A lightweight stochastic subspace identification-based modal parameters identification method of time-varying structural systems. *Journal of Sound and Vibration*, 570:118092, 2024.

[108] I Goethals, L Mevel, A Benveniste, and B. D Moor. Recursive output only subspace identification for in-flight flutter monitoring. In *Proceedings of the 22nd International Modal Analysis Conference (IMAC-XXII)*, Dearborn, Michigan, USA, 2004.

[109] C.-H Loh, J.-H Weng, Y.-C Liu, P.-Y Lin, and S.-K Huang. Structural damage diagnosis based on on-line recursive stochastic subspace identification. *Smart Materials and Structures*, 20(5):055004, 2011.

[110] Ping Lin, Nanxiong Zhang, and Bin Ni. On-line modal parameter monitoring of bridges exploiting multi-core capacity by recursive stochastic subspace identification method. In *2008 American Control Conference*, pages 632–637, Seattle, WA, 2008. IEEE.

[111] K Ciloglu, Y Zhou, F Moon, and A. E Aktan. Impacts of Epistemic Uncertainty in Operational Modal Analysis. *Journal of Engineering Mechanics*, 138(9): 1059–1070, 2012.

[112] E Reynders, R Pintelon, and G De Roeck. Uncertainty bounds on modal parameters obtained from stochastic subspace identification. *Mechanical Systems and Signal Processing*, 22(4):948–969, 2008.

[113] C. R Farrar, S. W Doebling, and P. J Cornwell. Comparison study of modal parameter confidence intervals computed using the Monte Carlo and bootstrap techniques. In *Proceedings of the 16th International Modal Analysis Conference (IMAC-XVI)*, volume 2, pages 936–944, 1998.

[114] S. W Doebling and C. R Farrar. Estimation of statistical distributions for modal parameters identified from averaged frequency response function data. *Journal of Vibration and Control*, 7(4):603–624, 2001.

[115] M Deistler, K Peternell, and W Scherrer. Consistency and relative efficiency of subspace methods. *Automatica*, 31(12):1865–1875, 1995. Trends in System Identification.

[116] D Bauer, M Deistler, and W Scherrer. Consistency and asymptotic normality of some subspace algorithms for systems without observed inputs. *Automatica*, 35(7):1243–1254, 1999.

[117] A Chiuso and G Picci. The asymptotic variance of subspace estimates. *Journal of Econometrics*, 118(1):257–291, 2004. Contributions to econometrics, time series analysis, and systems identification: a Festschrift in honor of Manfred Deistler.

[118] D Bauer. Asymptotic properties of subspace estimators. *Automatica*, 41(3): 359–376, 2005. Data-Based Modelling and System Identification.

[119] R Pintelon, P Guillaume, and J Schoukens. Uncertainty calculation in (operational) modal analysis. *Mechanical Systems and Signal Processing*, 21(6): 2359–2373, 2007.

[120] X.-B Lam and L Mevel. Uncertainty quantification for eigensystem-realization-algorithm, a class of subspace system identification. *IFAC Proceedings Volumes*, 44(1):6529–6534, 2011.

[121] M Döhler and L Mevel. Efficient multi-order uncertainty computation for stochastic subspace identification. *Mechanical Systems and Signal Processing*, 38(2):346–366, 2013.

[122] M Döhler, X.-B Lam, and L Mevel. Uncertainty quantification for modal parameters from stochastic subspace identification on multi-setup measurements. *Mechanical Systems and Signal Processing*, 36(2):562–581, 2013.

[123] P Mellinger, M Döhler, and L Mevel. Variance estimation of modal parameters from output-only and input/output subspace-based system identification. *Journal of Sound and Vibration*, 379:1–27, 2016.

[124] S Greś, M Döhler, N.-J Jacobsen, and L Mevel. Uncertainty quantification of input matrices and transfer function in input/output subspace system identification. *Mechanical Systems and Signal Processing*, 167:108581, 2022.

[125] E. P Reynders. Uncertainty quantification in data-driven stochastic subspace identification. *Mechanical Systems and Signal Processing*, 151:107338, 2021.

[126] E Reynders and K Maes. Uncertainty quantification of modal characteristics identified from frequency-domain stochastic subspace identification. *Procedia Engineering*, 199:996–1001, 2017.

[127] E Reynders, K Maes, G Lombaert, and G De Roeck. Uncertainty quantification in operational modal analysis with stochastic subspace identification: Validation and applications. *Mechanical Systems and Signal Processing*, 66–67:13–30, 2016.

[128] M Döhler, P Andersen, and L Mevel. Variance computation of modal parameter estimates from UPC subspace identification. In *Proceedings of the 7th International Operational Modal Analysis Conference (IOMAC)*, Ingolstadt, Germany, 2017.

[129] S Pereira, E Reynders, F Magalhães, Á Cunha, and J. P Gomes. The role of modal parameters uncertainty estimation in automated modal identification, modal tracking and data normalization. *Engineering Structures*, 224:111208, 2020.

[130] S Gres and M Döhler. Uncertainty propagation in subspace methods for operational modal analysis under misspecified model orders. In *International Conference on Noise and Vibration Engineering, and Uncertainty Structural Dynamics - ISMA-USD 2022*, Leuven, Belgium, 2022.

[131] S Greś, R Riva, C. Y Süleyman, P Andersen, and M. M Łuczak. Uncertainty quantification of modal parameter estimates obtained from subspace identification: An experimental validation on a laboratory test of a large-scale wind turbine blade. *Engineering Structures*, 256:114001, 2022.

[132] K Zhou, L.-H Zhi, J.-F Wang, X Hong, K Xu, and Z.-R Shu. An improved stochastic subspace modal identification method considering uncertainty quantification. *Structures*, 51:1083–1094, 2023.

[133] S Greś and M Döhler. Model order selection for uncertainty quantification in subspace-based OMA of Vestas V27 blade. In M. P Limongelli, P. F Giordano, S Quqa, C Gentile, and A Cigada, editors, *Experimental Vibration Analysis for Civil Engineering Structures*, volume 433, pages 43–52. Springer Nature, Switzerland, 2023.

[134] T De Troyer, P Guillaume, R Pintelon, and S Vanlanduit. Fast calculation of confidence intervals on parameter estimates of least-squares frequency-domain estimators. *Mechanical Systems and Signal Processing*, 23(2):261–273, 2009.

[135] T De Troyer, P Guillaume, and G Steenackers. Fast variance calculation of polyreference least-squares frequency-domain estimates. *Mechanical Systems and Signal Processing*, 23(5):1423–1433, 2009.

[136] M El-kafafy, P Guillaume, T De Troyer, and B Peeters. A frequency-domain maximum likelihood implementation using the modal model formulation. *IFAC Proceedings Volumes*, 45(16):179–184, 2012.

[137] M El-kafafy, T De Troyer, B Peeters, and P Guillaume. Fast maximum-likelihood identification of modal parameters with uncertainty intervals: A modal model-based formulation. *Mechanical Systems and Signal Processing*, 37(1-2):422–439, 2013.

[138] E. P Carden and A Mita. Challenges in developing confidence intervals on modal parameters estimated for large civil infrastructure with stochastic subspace identification. *Structural Control and Health Monitoring*, 18:53–78, 2011.

[139] P Barney, C Ferregut, L Perez, N Hunter, and T Paez. Statistical validation of system models. In *Proceedings of the 30th Hawaii International Conference on System Sciences*, volume 5, pages 501–510, Wailea, HI, USA, 1997. IEEE Comput. Soc. Press.

[140] S.-K Au, F.-L Zhang, and Y.-C Ni. Bayesian operational modal analysis: Theory, computation, practice. *Computers & Structures*, 126:3–14, 2013.

[141] V Peterka. Bayesian system identification. *Automatica*, 17(1):41–53, 1981.

[142] T Rogers, K Worden, and E Cross. On the application of Gaussian process latent force models for joint input-state-parameter estimation: With a view to Bayesian operational identification. *Mechanical Systems and Signal Processing*, 140:106580, 2020.

[143] T. J Rogers. *Towards Bayesian System Identification: With Application to SHM of Offshore Structures*. PhD thesis, University of Sheffield, Sheffield, United Kingdom, 2019.

[144] J. L Beck. Bayesian system identification based on probability logic. *Structural Control and Health Monitoring*, 17(7):825–847, 2010.

[145] Y Huang, J. L Beck, and H Li. Bayesian system identification based on hierarchical sparse Bayesian learning and Gibbs sampling with application to structural damage assessment. *Computer Methods in Applied Mechanics and Engineering*, 318:382–411, 2017.

[146] J.-H Yang and H.-F Lam. An innovative Bayesian system identification method using autoregressive model. *Mechanical Systems and Signal Processing*, 133: 106289, 2019.

[147] G Pillonetto and L Ljung. Full Bayesian identification of linear dynamic systems using stable kernels. *Proceedings of the National Academy of Sciences*, 120(18):e2218197120, 2023.

[148] 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.

[149] 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.

[150] K.-V Yuen, L. S Katafygiotis, and J. L Beck. Spectral density estimation of stochastic vector processes. *Probabilistic Engineering Mechanics*, 17(3):265–272, 2002.

[151] K.-V Yuen and L. S Katafygiotis. Bayesian fast Fourier transform approach for modal updating using ambient data. *Advances in Structural Engineering*, 6 (2):81–95, 2003.

[152] S.-K Au. Connecting Bayesian and frequentist quantification of parameter uncertainty in system identification. *Mechanical Systems and Signal Processing*, 29:328–342, 2012.

[153] S.-K Au. Insights on the Bayesian spectral density method for operational modal analysis. *Mechanical Systems and Signal Processing*, 66–67:1–12, 2016.

[154] S.-K Au, J. M Brownjohn, and J. E Mottershead. Quantifying and managing uncertainty in operational modal analysis. *Mechanical Systems and Signal Processing*, 102:139–157, 2018.

[155] Y.-C Zhu and S.-K Au. Bayesian operational modal analysis with asynchronous data, part I: Most probable value. *Mechanical Systems and Signal Processing*, 98:652–666, 2018.

[156] Y.-C Zhu and S.-K Au. Bayesian operational modal analysis with asynchronous data, Part II: Posterior uncertainty. *Mechanical Systems and Signal Processing*, 98:920–935, 2018.

[157] Y.-C Zhu and S.-K Au. Bayesian data driven model for uncertain modal properties identified from operational modal analysis. *Mechanical Systems and Signal Processing*, 136:106511, 2020.

[158] Z Zhu, S.-K Au, B Li, and Y.-L Xie. Bayesian operational modal analysis with multiple setups and multiple (possibly close) modes. *Mechanical Systems and Signal Processing*, 150:107261, 2021.

[159] J Mao, X Su, H Wang, and J Li. Automated Bayesian operational modal analysis of the long-span bridge using machine-learning algorithms. *Engineering Structures*, 289:116336, 2023.

[160] J. M. W Brownjohn, S.-K Au, Y Zhu, Z Sun, B Li, J Bassitt, E Hudson, and H Sun. Bayesian operational modal analysis of Jiangyin Yangtze River Bridge. *Mechanical Systems and Signal Processing*, 110:210–230, 2018.

[161] W.-J Yan and L. S Katafygiotis. A two-stage fast Bayesian spectral density approach for ambient modal analysis. Part II: Mode shape assembly and case studies. *Mechanical Systems and Signal Processing*, 54–55:156–171, 2015.

[162] B Li and A Der Kiureghian. Operational modal identification using variational Bayes. *Mechanical Systems and Signal Processing*, 88:377–398, 2017.

[163] M. J Beal. *Variational Algorithms for Approximate Bayesian Inference*. PhD thesis, University of London, 2003.

[164] 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, Orlando, FL, USA, 2011. IEEE.

[165] B Li, A Der Kiureghian, and S.-K Au. A Gibbs sampling algorithm for structural modal identification under seismic excitation. *Earthquake Engineering & Structural Dynamics*, 47(14):2735–2755, 2018.

[166] O Sedehi, L. S Katafygiotis, and C Papadimitriou. Hierarchical Bayesian operational modal analysis: Theory and computations. *Mechanical Systems and Signal Processing*, 140:106663, 2020.

[167] J.-H Yang, H.-F Lam, and J. L Beck. Bayes-Mode-ID: A Bayesian modal-component-sampling method for operational modal analysis. *Engineering Structures*, 189:222–240, 2019.

[168] H.-F Lam, J.-H Yang, and J. L Beck. Bayesian operational modal analysis and assessment of a full-scale coupled structural system using the Bayes-Mode-ID method. *Engineering Structures*, 186:183–202, 2019.

[169] J. L Beck. *Determining Models of Structures from Earthquake Records*. PhD thesis, California Institute of Technology, 1979.

[170] J. L Beck and P. C Jennings. Structural identification using linear models and earthquake records. *Earthquake Engineering & Structural Dynamics*, 8(2): 145–160, 1980.

[171] S. D Werner, J. L Beck, and M. B Levine. Seismic response evaluation of Meloland Road Overpass using 1979 imperial Valley earthquake records. *Earthquake Engineering & Structural Dynamics*, 15(2):249–274, 1987.

[172] Q Dollon, J Antoni, A Tahan, M Gagnon, and C Monette. A fast collapsed Gibbs sampler for frequency domain operational modal analysis. *Mechanical Systems and Signal Processing*, 173:108985, 2022.

[173] K. P Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, Cambridge, Massachusetts, first ed., 2012.

[174] J. A Rice. *Mathematical Statistics and Data Analysis*. Duxbury Press, Belmont, Calif, 2nd ed., 2002.

[175] A Gelman. *Bayesian Data Analysis*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press, Boca Raton, 3rd ed., 2014.

[176] P. M Lee. *Bayesian Statistics: An Introduction*. Wiley, Chichester, West Sussex, UK, 4th ed., 2012.

[177] H Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186(1007):453–461, 1946.

[178] G. E Box and G. C Tiao. *Bayesian Inference in Statistical Analysis*. John Wiley & Sons, Ltd, 1 ed., 1992.

[179] J. M Bernardo and A. F. M Smith. *Bayesian Theory*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd, Chichester, West Sussex, UK, reprint ed., 2004.

[180] D Koller and N Friedman. *Probabilistic Graphical Models: Principles and Techniques*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, 2009.

[181] M. E Tipping and C. M Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 61 (Part 3):611–622, 1999.

[182] 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.

[183] C Wu, C Yang, H Zhao, and J Zhu. On the convergence of the EM algorithm: A data-adaptive analysis, 2017.

[184] M Opper and D Saad. *Advanced Mean Field Methods: Theory and Practice*. Neural Information Processing. MIT Press, Cambridge, Massachusetts (US), 2001.

[185] C Jordan. Essai sur la géométrie à n dimensions. *Bulletin de la Société Mathématique de France*, 2:103–174, 1875.

[186] H Hotelling. Relations between two sets of variates. *Biometrika*, 28(3/4): 321–375, 1936.

[187] B Friedland. *Control System Design: An Introduction to State-Space Methods*. Dover Publications, Inc, Mineola, New York, dover edition, unabridged republication ed., 2005.

[188] G. H Golub and C. F Van Loan. *Matrix computations*. Johns Hopkins studies in the mathematical sciences. The Johns Hopkins University Press, Baltimore, fourth edition ed., 2013.

[189] R. J Allemang, D. L Brown, and A. W Phillips. Survey of modal techniques applicable to autonomous/semi-autonomous parameter identification. In *Proceedings of ISMA2010 International Conference on Noise and Vibration Engineering*, Leuven, Belgium, 2010.

[190] R Sbarra and G Coppotelli. DBSCAN-Based Approach for the Automatic Estimate of the Modal Parameters. In C Rainieri, C Gentile, and M Aenlle López, editors, *Proceedings of the 10th International Operational Modal Analysis Conference (IOMAC)*, volume 514, pages 618–630. Springer Nature Switzerland, Cham, 2024.

[191] F. R Bach and M. I Jordan. A probabilistic interpretation of canonical correlation analysis. Technical Report 688, University of California, Berkeley, 2005.

[192] C Archambeau, N Delannay, and M Verleysen. Robust probabilistic projections. In *Proceedings of the 23rd International Conference on Machine Learning - ICML '06*, pages 33–40, Pittsburgh, Pennsylvania, 2006. ACM Press.

[193] O Cappé and E Moulines. On-line expectation–maximization algorithm for latent data models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71(3):593–613, 2009.

[194] P Cheema, M. M Alamdari, G Vio, F Zhang, and C Kim. Infinite mixture models for operational modal analysis: An automated and principled approach. *Journal of Sound and Vibration*, 491:115757, 2021.

[195] J Maeck and G De Roeck. Description of Z24 benchmark. *Mechanical Systems and Signal Processing*, 17(1):127–131, 2003.

[196] R. L Boroschek and J. A Bilbao. Interpretation of stabilization diagrams using density-based clustering algorithm. *Engineering Structures*, 178:245–257, 2019.

[197] M Ester, H.-P Kriegel, J Sander, and X Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining*, 1996.

[198] C Wang. Variational Bayesian approach to canonical correlation analysis. *IEEE Transactions on Neural Networks*, 18(3):905–910, 2007.

[199] A Klami and S Kaski. Local dependent components. In *Proceedings of the 24th International Conference on Machine Learning*, pages 425–432, Corvalis Oregon USA, 2007. ACM.

[200] A Klami, S Virtanen, and S Kaski. Bayesian Canonical Correlation Analysis. *Journal of Machine Learning Research*, 14((2013)):965–1003, 2013.

[201] J Viinikanoja, A Klami, and S Kaski. Variational Bayesian Mixture of Robust CCA Models. In *Machine Learning and Knowledge Discovery in Databases*, volume 6323, pages 370–385. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[202] S Särkkä and A Solin. *Applied Stochastic Differential Equations*. Cambridge University Press, 1 ed., 2019.

[203] J Liu, X Liu, and X Ma. First-order perturbation analysis of singular vectors in singular value decomposition. In *2007 IEEE/SP 14th Workshop on Statistical Signal Processing*, pages 532–536, 2007.

[204] L Mevel, M Goursat, and M Basseville. Stochastic supspace-based structureal identification and damage detection and localisation - application to the Z24 bridge benchmark. *Mechanical Systems and Signal Processing*, 17(1):143–151, 2003.

[205] A Gelman. Objections to Bayesian statistics. *Bayesian Analysis*, 3(3), 2008.

[206] B. J O'Connell and T. J Rogers. On improving the efficiency of bayesian stochastic subspace identification. In C Rainieri, C Gentile, and M Aenlle López,

editors, *Proceedings of the 10th International Operational Modal Analysis Conference (IOMAC 2024)*, pages 609–617. Springer Nature Switzerland, 2024.

[207] I Goodfellow, Y Bengio, and A Courville. *Deep Learning*. Adaptive Computation and Machine Learning. MIT press, Cambridge, MA, 2016.

[208] S Greś, M Döhler, and L Mevel. Uncertainty quantification of the modal assurance criterion in operational modal analysis. *Mechanical Systems and Signal Processing*, 152:107457, 2021.

[209] R Brincker and N Møller. Operational modal analysis – a new technique to explore. *Sound and Vibration*, Editorial, 2006.

[210] Student. The probable error of a mean. *Biometrika*, 6(1):1, 1908.

$$q(\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\Psi}, \mathbf{W}, \mathbf{u}) = \prod_{n=1}^{N} q(\mathbf{z}_n) q(\mathbf{u}_n) \prod_{i=1}^{d} q(\mathbf{w}_i) q(\boldsymbol{\mu}) q(\boldsymbol{\Psi})$$

Bayesian SSI will return ...

Disney, please don't sue me.