Author:
Liu, Y

Title:
Identification of modal parameters based on moving force excitation

Identification of Modal Parameters based on Moving Force Excitation

Yi Liu

A dissertation submitted to the University of Bristol in accordance
with the requirement for award of the degree of Doctor of Philosophy
in the Faculty of Engineering

School of Civil, Aerospace and Mechanical Engineering

October 2020

Abstract

This thesis proposed a two-stage input-output system identification methodology to investigate the possibility of estimating the bridge modal parameters, i.e., natural frequencies, damping ratios, and modal masses, under excitation of a moving vehicle with the simultaneous use of the moving input acceleration signal and the measured bridge acceleration responses. In the first stage, which is an output-only identification problem, the bridge mode shapes are estimated from the measured bridge forced and free vibration acceleration responses only. The obtained estimated bridge mode shapes are then served as a modal basis to re-express the coupled bridge and vehicle geometric coordinates as decoupled modal coordinates. This procedure then leads us to the second stage identification. With both the input and output information for each mode, a series of (Frequency Response Functions) FRFs can be constructed by using the Discrete Fourier Transform (DFT) technique, along with the parametric model of the Accelerance for a linear system, an optimization procedure is performed to extract the natural frequencies, damping ratios, as well as the modal masses simultaneously.

In order to verify the proposed method, a simply supported Euler-Bernoulli beam of known parameters is used as an example. Its responses to a moving load and a quarter car with the influence of the road roughness are calculated numerically, yielding simulated measured accelerations at a series of fixed locations on the structure. With these numerically generated data, the feasibility and efficacy of the proposed two-stage strategy are validated. A comparison work to examine the estimation accuracy and efficiency is conducted by using the free decay response of the bridge.

To make the proposed method more general, both models, i.e., moving load and quarter car traversing the bridge, are nondimensionalised, and they are simulated with and without noise. Upon verifying the proposed method, we discussed the impact factors, which can influence the estimation results. We discovered that the nondimensionalised spatial frequency is the most important one, which, if known beforehand, can give guidance to the identification results.

Apart from the nondimensionalised spatial frequency, we recognised that the mode shapes estimated from the first stage are also an important factor that can influence the identification accuracy. Therefore, we proposed a new concept real-valued one-sided spectral density matrix, which not only gives the same level of accuracy for the estimated natural frequencies and the damping ratios but also generates better mode shape estimation compared to the classical complex-valued two-sided spectral density matrix. This new approach is used throughout this study to give a better estimation of the mode shapes.

Notwithstanding that the proposed two-stage method is the main topic of this study, our research is far beyond the Vehicle-Bridge Interaction (VBI) analysis. One of the major contributions is made to the Frequency Domain Decomposition (FDD) technique, which serves as the output-only technique for the mode shape estimation in the first stage identification of the proposed method. We reinterpreted the FDD using the Principal Component Analysis (PCA) with a Periodogram defined Power Spectral Density (PSD) estimator. With our new theory, the role of the singular values and singular vectors are better defined. And we conclude that the FDD is a good technique to be used to detect the presence of close modes. In order to apply the FDD to extract the bridge mode shapes from the vehicle-induced bridge vibration responses (i.e., nonstationary process) only, a simple case study is designed to assess its ability to deal with a similar kind of nonstationary random process.

While analysing a PSD estimator, which is a positive semidefinite self-adjoint compact operator on a Hilbert space, we introduced the concepts of pure state and mixed states from the quantum mechanics and defined a density operator based on a PSD estimator. With a density operator, we can calculate the corresponding Von Neumann entropy or the purity condition. By plotting the entropy or purity of a density operator against the frequency line, we can obtain a different picture of the behaviour of a system that the FDD fails to provide when two modes are very closely spaced. Particularly, we found that this new picture is mainly affected by three influence factors, i.e., correlation relation between the mode shapes of the two modes, damping ratios of the system, and the measurement noise level, which in turn can affect our mode shape estimation results.

To my beloved parents and my dog!

Acknowledgement

I guess this is the most difficult part for me to write in this thesis because my thanks to those who helped me in this PhD adventure is beyond words.

The appreciation and the opportunity given by Prof. John Macdonald and the University of Bristol for me to come here to pursue my PhD are invaluable to me. Without the guidance, support, and constant encouragement given by Prof. John Macdonald in the last few years, this dissertation would never have been accomplished.

My sincere gratitude also goes to Prof. Rune Brincker, who gave me the recognition of my work on his FDD, drafted the first version of the introduction for my FDD chapter, and shared his idea about the real-valued one-sided spectral density matrix with me. Besides, I also like to express my thanks to Prof. Jonathan Cooper and Dr. Dario Di Maio, as they also gave me helpful suggestions and comments on my work.

Besides, I want to thank Dr. Stephanie Wallis, Shuyin Feng, Dr. Sha Luo, Chizi Hao, and many others for being my friends and supporting me whenever I needed it for the past few years.

Last but not least, I want to thank my parents and my previous little dog for their support and encouragement throughout my studies. My Tufei, if you could hear me from heaven, I want you to know that I love you so much, more than anything in this world, and this thesis is written for you and our parents. You came to me before I came to the UK to do my PhD, but you left me right before I finished it. If I could make a deal with the devil to make you come back to me, then I wish I had never come here to do this PhD.

Yi Liu

Bristol, October 2020

Author's Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ....   DATE: ....

Table of Contents

List of tables

List of figures

1 Introduction

With the goal of assessing the health condition, i.e., automatically identify structural damage, of the structure in real time, Structural Health Monitoring (SHM) has been a hot topic for the last two decades. Thanks to the development of sensing technology and the related signal analysis and interpretation algorithms, SHM has become an important and fast-growing research discipline in many areas, such as civil, mechanical, and aerospace engineering. However, due to the difficulties and uncertainties increased in the presence of large civil structures, the pace of the SHM development on civil structures is relatively slower when compared to the mechanical or aerospace structures [1]. Nevertheless, various different methods and algorithms for estimating damages in civil structures have been proposed, and the majority of them are dynamic-based techniques [1], [2]. It is reckoned that structure properties, such as stiffness, are usually modified by the damage, which is manifested in the measured vibration response of the structure [3]. Thus, damages, either they are invisible or hidden, can be detected through analysing the vibration measurements.

Among all the civil infrastructures, bridges, as a vital link in the transportation system, have undoubtedly received special attention [4], [5]. As a subset of the bridge health monitoring research, the vehicle-induced bridge vibration analysis has been a very popular topic in the past two decades. It is not just an economical alternative to the forced vibration test [6], but also a finer version of the ambient vibration test [6], as the input response is no longer treated as white noise. Depending on the purpose and the proposed procedure, the input force can be reconstructed [7].

1.1 Background

The research upon the Vehicle-Bridge Interaction (VBI) comprises of many aspects, such as force identification or reconstruction [7], [8], [9], [10], [11], [12], identification of the bridge properties [13], [14], [15], or damage detection of the bridge from passing test vehicle [16], [17]. Although each research has a different focus, they all share the same goal that to better design and maintain the bridge structure. This is important because dynamic loads on bridges are ever-increasing, and with the application of new materials and improved designs, modern bridges tend to be lighter and more flexible. As a result, highway bridges are increasingly susceptible to vibration [18].

Particularly, when a vehicle is moving across a bridge, the dynamic loads of which will cause the bridge to vibrate. The service life of the bridge will be reduced by the repeated application of such dynamic

loads, as they contribute to fatigue, surface wear, and cracked concrete, which leads to corrosion problems [18]. As a result, frequent maintenance of the bridge is required because of the deterioration of the bridge condition, which will incur enormous maintenance fees. Thus, in order to build vibration resistant bridges, it is rather important to understand the dynamics of the VBI system [18].

As the name goes, namely, Vehicle-Bridge Interaction system, which implies two subsystems, i.e., the vehicle and the bridge, are coupled with each other due to forces at the contact points. Hence, the performance of one is affected by the performance of the other and vice versa [11]. Between the vehicle wheels and the bridge structure, there is road roughness, and the vehicle is excited by road roughness as well as the dynamic deflection of the bridge [18]. As a result, dynamic wheel loads generated by the vehicle, which in turn excite the bridge causing larger dynamic displacements, and such dynamic displacement will be fed back into the vehicle [18]. This feedback mechanism of interaction forces is the one that couples the dynamic response of the bridge to that of the vehicle [18]. Note that such coupling phenomenon happens, and it is significant, mainly due to the frequency range of vehicle dynamic wheel load (1.5 to 4.5 Hz) overlaps the bridge resonance frequency range (usually below 10 Hz and often as low as 2 to 3 Hz) [18]. Apart from this, the low bridge damping (approximately 2% of critical in the first bending mode) complicate matters further, as it does not significantly reduce the vibration caused by the VBI [18].

The above analysis about the VBI system suggests three research aspects, bridge response only based analysis, vehicle response only based analysis, and vehicle-bridge information-based analysis. In the process of all three analysis aspects, the identification of the vehicle or bridge parameters is the main task, but with the bridge parameters being more important. We may easily measure the vehicle parameters with instruments readily available, but it has always been a challenge for the bridge parameter measurements due to its massive size and special requirements for the measuring procedure and equipment. For instance, we may need to shut down a bridge in operator to do an on-site experiment or use a special instrument to excite a bridge to get a particular mode. Besides, depending on the formulated procedure, the road profile [19], [20], [21] can also be identified.

For the bridge response only based analysis, it is known as the direct methods, which requires a large number of sensors to be installed at different positions on the bridge to monitor the dynamic properties [22]. Under the VBI context, such a direct method can generate an inverse problem, which is normally ill-conditioned when we use the bridge responses to identify the moving loads, for the responses are continuous functions of time defined only at a few spatial coordinates [23]. One of the remedies is to use a regularization method to tackle this ill-conditioning problem, such as the Tikhonov's regularisation method [7]. Apart from the moving loads identification research, the vehicle-induced bridge responses are also used to conduct a damage detection procedure [24], [25], [26], [27], as it is believed that local damages in structure would be more sensitive to excitations nearby, and the use of

moving vehicle as an exciter along a bridge deck can provide equal opportunities for identifying local damages at different locations of the structure [28].

On the other hand, in some cases, for instance, some short or medium span bridges, which form the greatest proportion of bridges in service, such sensor system is not readily accessible, then an indirect approach [19], [21], [32], [33], [34] is favoured. This is indirect approach is the aforementioned second type of VBI analysis, which utilises the dynamic response of a passing vehicle to estimate the dynamic properties of the bridge.

Compared to the direct method, the indirect approach, which aims at reducing the need for any direct installation of equipment on the bridge itself, is less expensive [35], as it only involves a vehicle instrumented with sensors. The dynamic properties of the bridge, such as natural frequencies [36], [37], [38], damping ratios [13], and mode shapes [39], [40] are extracted from the vehicle response only. Because of many advantages of the indirect method possesses, namely, simplicity, economy, efficiency, and mobility, such method has received a lot of attention ever since it was proposed by Yang et al. [41]. Similar to the direct method, many researchers have been investigating the use of the vehicle response to detect bridge damages [16], [42], [43]. Although the indirect approach is promising, and it is now well established with theoretical and experimental investigations. Challenges, such as the influence of the road profile on the vehicle response and the variation of the bridge frequencies under a moving vehicle excitation, need to be overcome before the approach becomes an effective and reliable system [35]. It is because, in an interaction system, the influence of the road profile on the vehicle response can, in turn, reduce the visibility of the bridge frequency peaks. While in bridge damage detection, the variation of the bridge frequencies may mask any frequency changes caused by damage [35].

As for the third approach, which utilises both the vehicle and the bridge information, was developed alongside the development of the modern sensor technology with the purpose of combating the high uncertainty loading problem associated with the system identification and damage detection by using output-only datasets [44]. It is believed that if the vehicle loading was precisely known, then the accuracy of current damage detection methods would be improved. Currently, three sensor systems are adopted to monitor traffic on highways, namely, Weigh-in-Motion Systems (WIMS) [45], [46], video-based sensing system [44], [47], and wireless sensor networks [44], [48].

The most widely used one is the WIMS, which has been used to identify the weight of vehicles crossing a bridge at a specific position. Regardless of its popularity, such systems only measure vehicle loading at a static location on the bridge, which made it inadequate for direct observation of the VBI. This is the same problem shared by the video assisted approach for SHM of highway bridges [49], [14], [15], [50]. In order to capture the dynamic coupling that exists during VBI, the bridge and the vehicle must both be instrumented. Ideally, they need to be monitored within a single monitoring system architecture, which can be achieved by using the wireless sensor networks [44], [51]. Rather, the two subsystems

can be independently monitored, the vehicle and bridge response datasets would need to be combined with accurate time synchronisation [44].

Based on the assumption that the responses of two subsystems can be successfully monitored and synchronised, we proposed a two-stage input-output system identification procedure for the bridge modal parameters estimation. Unlike the two-stage system identification methodology proposed by Kim et al. [51], who used the free decays of the bridge to identify the dynamic characteristics of the bridge in the first stage, then use the vehicle-bridge response data to estimate the moving load in the second stage, the proposed method in this study only focuses on the bridge modal parameters identification. The concerned modal parameters include the natural frequencies, damping ratios, modal masses, and mode shapes. Especially for the modal masses, which cannot be directly estimated with unknown loading, but with the proposed method, it can be estimated.

Note that the mode shapes are identified in the first stage by applying an output-only system identification technique to the bridge output-only responses. The technique we used in this study is the Frequency Domain Decomposition (FDD). With the mode shapes estimated from the FDD, we can then use them to decouple the bridge responses and the vehicle response into modal coordinates. By combining the measured moving force vibrational information and the bridge acceleration response, a series of Frequency Response Functions (FRFs) can be constructed. Then the modal parameters of the bridge can then be estimated through an optimisation procedure with the objective function formed in terms of the theoretical FRF expression, and this procedure forms the second stage identification. Besides, it is worth mentioning that the full-length record of the bridge is used in both stages in this work, which is also different from the work given by Kim et al. [51].

A breakdown of this thesis content is presented in the next section, and the proposed two-stage system identification methodology forms the framework of this study.

1.2 Thesis structure

We have essentially three main objectives in this thesis. The first two objectives, namely the real-valued spectral density matrix and the new theory for the FDD and the close mode analysis via a density operator, are presented in chapter 3 and chapter 4, respectively. While the third objective, which is about the proposed two-stage input-output system identification methodology and its verification, is mainly discussed in chapter 5 and chapter 6.

Since the essence of this thesis is about modal analysis, more precisely, system identification, our discussion starts with the general concept of system identification, and some of the traditional identification techniques for the linear systems with constant coefficients are overviewed in chapter 2. Special attention is given to the Polyreference method and the Data-Driven Stochastic Subspace

Identification (SSI-DATA) method. Some new ideas and new proofs regarding these two methods are given.

In chapter 3, we proposed a new concept, namely, the real-valued one-sided spectral density matrix, which is verified by applying the Polyreference method to a dynamic vibrating system with two modes fixed and another one moving between the two fixed modes. This real-valued one-side spectral density matrix idea is further adopted in one of our case studies in chapter 4, and it is applied in the first stage of the proposed two-stage method to all the simulations in chapter 6.

As for chapter 4, it presents one of the most important work in this thesis, which not only gives a new theoretical background for the FDD technique but also put forwards some new measures in terms of a density operator to estimate the mode shapes of two modes when they are extremely close to each other.

As a nonparametric system identification technique, the FDD has received popularity in the past two decades. However, the poor theoretical background given by Brincker et al. [52], [53] has largely restricted its application. Therefore, in this chapter, we aim to redefine this method and give it a new theoretical background in terms of a power spectral density (PSD) estimator obtained from the Periodogram. Because of this new definition, we can treat a PSD estimator as a covariance matrix in the frequency domain and use the Principal Component Analysis (PCA) to explain it. While by analysing the properties of this PSD estimator on a finite-dimensional Hilbert space, we can extend our discussion to quantum mechanics, from which we introduce a density operator and the corresponding purity and Von Neumann entropy as new measures to estimate the close modes mode shapes.

Two case studies are presented in this chapter. The first one aims to evaluate the capability of the FDD to be used to deal with a certain kind of nonstationary random process. This case study seems irrelevant, but it is an important part which links the whole thesis. Given the contribution of the kinetic energy as the vehicle moves cross the bridge, when we look at a bridge subject to this moving force from rest (before the vehicle enters the bridge) to rest (after the vehicle enters the bridge), the overall time history of the bridge response is nonstationary with finite energy. Therefore, before we use the FDD to estimate the bridge mode shape from the output-only responses of the bridge, we need to show that the FDD can be used to deal with a certain kind of nonstationary process. Besides, it is worth mentioning that the mode shapes estimated from the FDD is not mass normalised, but they are the orthonormal basis vectors of the modal space.

As for the other case study, it focuses on the close mode analysis. We included a comparison study of the proposed method based on a density operator with two time domain methods, namely, the Polyreference method and the SSI-DATA method. In this case study, we mainly work on the frequency region where the FDD fails. However, by analysing a density operator alone, we are unable to estimate the natural frequencies and damping ratios for the close modes. Therefore, we introduce a new procedure, which combines our new measures and the Polyreference, i.e., estimate the mode shapes

first based on a density operator, then use our estimates to decompose the measurements into a series of modal coordinates; after this, we apply the Polyreference to estimates the natural frequencies and damping ratios. In our analysis, we consider 8 different cases, which are the different combinations of three influence factors, i.e., correlation relation of the mode shapes of the two close modes, damping ratio, and noise level. For some combinations, our proposed method can provide us some visual information to make inferences about the system's modal parameters.

While in chapter 5, the theoretical background for the proposed two-stage method is presented with a discussion of the mode shape and its properties. The discussion of the mode shape is important, as we need to decide whether we should use the estimated complex mode shapes (in most cases) or the real part of the estimation to decouple the system. If the complex-valued mode shapes are used directly to decouple the system, while the system has real or close to real mode shape, an extra error will be carried to the second stage identification.

In general, the proposed two-stage method has the following merits. It can be used to identify the modal masses of the system, which can be used to scale the mode shapes for forced response analysis. Due to the inherent problem, none of the current Operational Modal Analysis (OMA) techniques is able to estimate it. Besides, theoretically speaking, it can produce a better natural frequency identification compared to the ones we obtained from the ambient vibration test, as the influence of the input can be separated from the bridge system according to the FRF expression. Also, the experiment set up is simpler and cheaper than the traditional forced vibration test, which is extremely helpful when one needs to conduct dynamic experiments on some structures such as a long-span bridge. Furthermore, because of the moving nature of the force, the excitation of the higher modes of the system can be easily achieved, which makes it comparable to an OMA method. One of the benefits of using OMA is that all DOFs can be excited by assuming white noise input along the structure. Now, with the advantages mentioned above, the OMA methods will be dwarfed by the proposed two-stage method devised for the moving load case. Apart from this, due to the simplicity of the measurement procedure, we can run the vehicle on the bridge multiple times, which allows a statistical analysis of the results.

Finally, in order to validate the proposed method, two vehicle systems, i.e., point moving load and a quarter car traversing a simply supported bridge, are considered in chapter 6. For the quarter car system, the road roughness effect is considered. Both models are nondimensionalised, and they are simulated with and without noise. With the nondimensionalised VBI models, we discovered a dominant impact factor which, if known beforehand, can give guidance to the estimation results. As for chapter 7, it summarises the results of this study and makes recommendations for further research.

1.3 Credit authorship contribution statement

Note that the idea of the real-valued spectral density matrix presented in chapter 3 comes from Prof Rune Brincker, so is the idea of the case study. However, the theoretical explanation was exclusively proposed by myself.

While for the content in chapter 4, except for some overview comments on the original FDD method in the introduction section and the numerical model used in the close mode analysis (which is the same as I used in chapter 3) were suggested by Prof. Brincker. The rest, from sensing the abnormality of the FDD, to establishing my theories, i.e., from modal analysis to quantum mechanics and from the FDD to a density operator; to designing the comparison work; to proposing a new procedure to estimate the modal parameters for the close mode analysis, and to the final figure analysis, are all done by myself independently without anyone's help, so is the nonstationary data analysis. Therefore, I take full responsibility for my work in chapter 4. It must be pointed out that all of his contributions were made on the other work I was working with him regarding the FDD and the real-valued density matrix two years ago. He is not involved in any of the work I presented in chapter 4. Since I used some of his ideas from our previous work, I must pay tribute to his contribution as well. Although I have done almost all the work by myself in chapter 4, Prof. Macdonald gave me some suggestions on figure presenting.

As for my work in chapter 5 and chapter 6, my supervisor Prof. John Macdonald helped me to establish the identification procedure. All the numerical analysis was done under his supervision based on his suggestions.

2 Review and new proofs for Operational Modal Analysis (OMA) methods

2.1 Introduction

System identification plays a significant role in modern science. It deals with the problem of building mathematical models of dynamical systems based on observed data from the systems, which is part of basic scientific methodology [54]. The notation of a system is a broad concept. Generally speaking, a system is an object in which variables of different kinds interact with each other and yield observable signals. Thus, to identify a system, an inevitable procedure is signal processing. However, even if signal processing is vital to system identification, one cannot identify the system without a mathematical model/structure to describe the system. Such a mathematical model may or may not possess physical meaning. This implies the essence of system identification, which is about determining the best mathematical description of the system. Usually, the best representation captures the eigenstate of the system, which reveals the innate properties of the system.

For most cases in Civil, Aerospace, and Mechanical Engineering, the system of interest may be modelled with a clear mathematical structure. For instance, the Linear Time-Invariant (LTI) system, described by linear constant-coefficient differential or difference equations, are often adopted to represent the system.

To find the best model to describe a system, one needs to develop a procedure to realise the model selection process. This procedure is the so-called identification algorithms, which is one of the most challenging parts of system identification. With the careful design and the aid of the computer, the realisation of this process can be achieved automatically and efficiently. In the past few decades, this automated process has gradually become what we have now known as Machine Learning (ML) [55], [56]. Therefore, the core of system identification is to work on some accessible data related to a system, then make inferences about the system based on some descriptive parameters.

Over the past century, different kinds of identification techniques have been proposed in different fields with different names, and the number of the techniques is ever increasing. Clearly, it is impossible to give a comprehensive overview of all the system identification methods, and it is also not our purpose in this chapter. However, lots of techniques are just variations of some more fundamental methods, and this is quite phenomenal in modal analysis. It is worth mentioning that in Civil, Aerospace, and Mechanical Engineering, system identification is usually referred to as modal analysis. Thus, in this

8

chapter, we will first identify some fundamental ideas used in modal analysis. Then introduce some of the time domain Operational Modal Analysis (OMA) techniques which will be used in chapter 3 (i.e., Polyreference method [57]) and chapter 4 (i.e., the Data-Driven Stochastic Subspace Identification (SSI-DATA) method [58]) for the comparison study.

To be more specific, in order to explain some of the fundamental ideas used in modal analysis, an overview of the system identification and modal parameters identification is first presented in the next two subsections. Then we will talk about the identification problem start with a linear differential equation with constant coefficients (section 2.2), followed by the discussion of a linear difference equation and its variations (section 2.3). Having done this, we will present some possible approaches to estimate the system coefficients (section 2.4), and from which we can obtain the modal parameters. Upon discussing those possible ways, we will present some new proofs for the OMA we just mentioned above, especially for the Polyreference method (section 2.4.1.2) and the SSI-DATA method (section 2.4.3.2). Note that those new proofs do not fundamentally change the method but can offer some new insight to OMA.

2.1.1 System identification overview

System identification is a broad term used to describe the process of experimental modelling, which has been developed for many years during which different scholars have a different definition of identification. However, the essence of system identification is the same that it is a method of using input and output data of the system to build a mathematical model of the system [59].

In general, the identification process consists of five important elements or steps, i.e., design an appropriate input signal, use experimental input and output data, select a class of models, construct an error criterion function, and determine a model that fits the data best by optimisation methods [60].

Normally, the quality of system identification depends on the quality of the inputs, which are under the control of the engineer. The input of the system belongs to the system's external stimuli, and it can or cannot be manipulated by the observer. Excitations, such as a continuous sinusoidal input or a transient pulse or chirp signal, are usually controllable in an experiment [61]. For a known system with known input, then the output of the system can be calculated exactly. However, in most cases, this is unrealistic. The system output is constantly subject to disturbances, which may or may not be directly measured by the observer [62]. Usually, both measurement noise and uncontrollable inputs are disturbances. For those unmeasurable disturbances, we can only observe their influence on the output. For instance, the actual loading of wind on a real structure is unmeasurable, so do the temperature loading and different ground conditions [63]. Here, the output of the system is the observable signals of interest.

This classification of the inputs and the disturbances implies another branch in system identification, i.e., OMA, which uses only the output data to estimate the system properties. In terms of the definition of system identification, it generally utilises both input and output data of the system. It is true that an input-output technique would be more accurate, but the input data is not always available. For instance, when the system is excited by the unmeasurable disturbances, the system information can only be projected by using the output data. This is quite common for large infrastructures in Civil Engineering, where the identification of the modal parameters based on Experimental Modal Analysis (EMA) [64] techniques becomes more challenging in the presence of large-scale structures with low frequency ranges.

In both practice and the literature, there is no doubt that LTI systems form the most important class of dynamical systems, as they represent the idealisations of processes encountered in real life. Most importantly, such approximation is often justified, and design considerations based on linear theory lead to good results in many cases [62]. Particularly, for an LTI causal infinite-duration impulse response (IIR) system [65], it can be described by its Impulse Response (or weighting) Function (IRF) $h(\tau)$ as follows,

$$y(t) = \int_{\tau=0}^{\infty} h(\tau) u(t-\tau) \mathrm{d}\tau \tag{2.1}$$

where $u(s)$, $s \leq t$ is the input, and $y(s)$, $s \leq t$ is the corresponding output of the system. Knowing $h(\tau)$ and $u(s)$, we can compute $y(s)$. Thus, the IRF is a complete characterisation of the system. The discrete version of Eq. (2.1) is given by

$$y(n) = \sum_{k=0}^{\infty} h(k) u(n-k), \ n = 0,1,2,\ldots \tag{2.2}$$

For this reason, identifying the impulse response of the system or its corresponding Fourier transform (i.e., Frequency Response Function (FRF)) is the target of the system identification. To determine the IRF or FRF, one needs to estimate the eigenstate of the system or find the roots of the characteristic polynomial of the system. To simplify the problem, the linear constant-coefficient differential or difference equations are widely used to describe the system. Models, such as state-space models and Autoregressive-moving-average (ARMA) models, are commonly seen in the literature. However, determining the best model from the model sets based on these model structures is not an easy task and this where optimisation comes into play. In effect, the optimisation procedure [66] is a key factor in many system identification techniques. System identification is essentially an optimisation problem. Among all the optimisation problems, the least-square problems are widely seen in many identification techniques.

When a set of candidate models has been selected, and it is parametrised as a model structure using a parameter vector, the best model is then founded by determining that parameter vector. Note that, quite often, a model set of interest is noncountable. This leads to the essential step in system identification, namely parameters estimation. For different model structures and identification purposes, the set of parameters that need to be identified will be different, and for different identification methods, they may have different names as well. For instance, when dealing with an engineering problem, the most widely used LTI dynamic system is given by the 2nd order differential equation of motion based on Newton's second law, i.e.,

$$\mathbf{M}\ddot{\mathbf{q}}(t) + \mathbf{C}\dot{\mathbf{q}}(t) + \mathbf{K}\mathbf{q}(t) = \mathbf{f}(t) \tag{2.3}$$

where $\mathbf{M}$ is the mass matrix, while $\mathbf{C}$ and $\mathbf{K}$ denote the damping and stiffness matrices, respectively. They are all symmetric matrices. $\mathbf{q}(t)$ is the output state vector. $\mathbf{f}(t)$ is the force vector that corresponds to each state. In most cases, it is very expensive to identify the mass, damping and stiffness matrices from Eq. (2.3) directly. Instead, we use the modal properties of the system to simplify the identification process by transforming the system matrices into diagonal matrices, then the model parameters to be identified become modal parameters of the system.

Traditional system identification methods, including the least square method, gradient correction method, and maximum likelihood method, and so on, have been developed comparatively perfect [62] . The identification of the model parameters based on these methods are determined by minimising the error criterion function. However, this is not the only way to identify the model parameters. For instance, transient response analysis, correlation analysis, frequency response analysis, etc. [62], are direct techniques that do not require one to select a confined set of possible models in the first place. Such methods are often also called nonparametric since they do not (explicitly) employ a finite-dimensional parameter vector in the search for the best description [54]. The Frequency Domain Decomposition (FDD) technique, which will be discussed in chapter 4, is one such nonparametric system identification technique.

The advantages and disadvantages of traditional system identification methods have been extensively discussed in the literature, e.g., [67] [68]. For example, the least square method has a very simple principle easy to understand and implement, and it converges very fast. However, it is not consistent, there is a deviation, and in order to use it, the input signal must be known. Apart from this, the input signal needs to change relatively and richly, and such a condition can be satisfied in many of the closed-loop control systems. But in some dynamic prediction systems and process control systems, the input signal cannot be obtained or changed [68]. For linear systems, the traditional methods normally generate a satisfactory result. However, it is not the case for nonlinear systems. Therefore, modern methods, such as Artificial Neural Networks (ANN) [56], Genetic Algorithm (GA) [69], [70], Fuzzy Logic (FL) [71],

just to name a few, have been proposed in recent years to deal with problems that conventional methods cannot handle.

2.1.2 Modal parameter identification

The system identification problem in machinery and structures is normally referred to as modal analysis. Normally, the resonant vibration of a structure is characterised by a series of normal modes, while a mode of vibration is usually defined by natural frequency, damping ratio, and mode shape. A set of these parameters is called a modal model for the structure. With these modal parameters, the dynamic properties of a structure can be completely characterised. This is because the poles of the transfer function or the roots of the characteristic polynomial of the system are made up of the combination of the natural frequencies and the damping ratios, while the mode shapes are the modal parameter for a system with multiple degrees of freedom. Typically, the system identification of the structures which satisfy Newton's second law of motion is the process of determining these parameters from experimental data, and this process is the so-called modal parameters estimation.

Estimating these modal parameters is of practical importance. For example, in flutter analysis [72], the lower the natural frequency, the more susceptible the structure is to flutter. What is worse, for the classical flutter, wind can couple the first vertical mode and the first torsional mode together. The closer in frequency those two modes are, the stronger the coupling is, then the stronger the flutter.

As an energy dissipation indicator, damping plays a crucial role in structural dynamics as well. However, it cannot be identified theoretically. When a structure subject to positive damping, it helps to reduce the amplitude of the vibration. But if the structure has negative damping, the system will absorb energy from the external source, which is fatal for a vibrating structure, as it can lead to the collapse of the system. The flutter is actually defined to be the dynamical instability due to negative damping from the wind. Therefore, it is vital to know the fundamental frequency of a vibrating structure, the frequencies for the first vertical mode and the first torsional mode, and their corresponding damping ratios. These parameters do not live independently, and they are interacting with each other all the time, including the modal mass. For a vibrating system, the larger the modal mass, the more difficult for us to excite that mode. Therefore, it is important for us to measure those modal parameters as accurately as we can.

2.1.2.1 Experimental Modal Analysis (EMA)

Typically, as the direct derivative of the concept of system identification, the modal parameters are estimated via modal testing, which is also referred to as EMA [64][73][74], and both the input signal and the output response of the system are measurable. The essence of obtaining the modal parameters in EMA is about curve-fitting the FRF plots in the vicinity of resonance in the frequency domain or evaluating the IRFs in the time domain. Compared to the time domain IRFs, by inspection of FRFs in

the frequency domain to interpret the structure responses is more convenient. As a result, a greater emphasis and utilisation has been made to the frequency domain methods [73].

Depending on how the data are collected or measured, the FRF data sets can be divided into three types [73], namely, Single-input, Single-output (SISO), Single-input Multi-output (SIMO), Multi-input Multi-output (MIMO) [73]. According to this classification of the FRFs, we will have two types of modal analysis methods based on the FRFs, i.e., single-FRF methods, which process one single FRF curve at a time, and multi-FRF methods, which analyse several curves simultaneously. Conventionally, it is time-consuming for the analyst if the simpler method is applied to large datasets. Whereas the more powerful methods (e.g., Polyreference [64][73], a multi-IRF method deal with MIMO data) can be intolerant of the small consistencies in data massed by repeated application of the SISO method [73]. The second stage of our proposed method in this thesis is a single-FRF method, but the FRF is evaluated by using the generalised coordinates.

It should be pointed out that, depending on whether a single mode is to be extracted at a time, both the frequency domain methods and the time domain methods can be further divided into two categories [73], i.e., single degree of freedom (SDOF) methods and multi-degree of freedom (MDOF) methods. Thus, based on these classifications, the second stage of the proposed method in this study should be a single-FRF SDOF method.

Even if frequency domain methods based on the FRFs of the system are predominant in EMA, in the cases of structures with very low natural frequencies and the required time to obtain enough cycles is a problem, the time domain method is superior [73]. The basic tenet of these time domain methods is that, for any IRF or free vibration response function, it can be expressed by a series of complex exponential components [73], that is,

$$h(t) = \sum_i h_i(t) = \sum_i A_i e^{s_i t} \tag{2.4}$$

where A_i is called the modal constant, which is also the amplitude of the IRF, s_i is the i th pole of the system. The properties of each of $h_i(t)$ contains the eigenvalue and eigenvector properties of one mode. Methods derived based on Eq. (2.4) is forms a category of the so-called "complex exponential" methods. In this category, the most widely known methods are the Complex Exponential (CE) method based on Prony's trick [75], the Least-Square Complex Exponential method (LSCE) [64], The Polyreference Complex Exponential method (PRCE) [64], [76], the Ibrahim Time Domain (ITD) method [77].

2.1.2.2 Operational Modal Analysis (OMA)

Having discussed the EMA, in this subsection, we shall look at the OMA. The general assumption of OMA testing technique is that the structure to be tested is being excited by some type of excitation that

has approximately white noise characteristics [63]. Hence, it has energy distributed over a wide frequency range that covers the frequency range of the modal characteristics of the structure.

In general, OMA is deemed as an economical alternative to EMA to estimate the modal parameters of the system. The EMA uses controlled or at least measurable input forces, while the OMA uses the operational forces or ambient excitations. Even if the level of excitation and vibration can be carefully controlled in the forced vibration test, the forced vibration methods can be more complex and expensive than in-operation vibration tests. Besides, for large structures in Civil Engineering, such as long-span bridges and high-rise buildings, such controllable input and output test is difficult to achieve a satisfactory level. Especially for large-scale structures with low frequency ranges, the identification of the modal parameters by EMA techniques becomes more difficult. Also, installing a controlled and measurable exciter on large structures is a complex and sometimes unfeasible task. Quite often, it causes interruptions to the normal operations of structure, as the facility needs to be shut down to conduct the test. To compensate and rectify the shortcomings in EMA, OMA was developed.

The ideas behind OMA techniques seem different from the theories for the forced vibration test. Also, it is claimed that OMA only unitises the output response of the system, which to some extent, violates the definition of system identification that it needs both input and output responses to estimate the system parameters. However, it is not true.

In OMA, most of the traditional time domain techniques are an extension of the EMA time domain techniques, especially the MIMO methods. The measurable forced inputs are merged into the disturbances, and then the unmeasurable disturbances are treated as ambient excitations, which are normally assumed to be uncorrelated Gaussian white noise. This assumption about the inputs is characterised by the Natural Excitation Technique (NExT) [78]. In general, NExT is a method of modal testing that allows the structure to be tested in their ambient environments. Since the ambient excitations cannot be measured, the conventional modal analysis, utilising the FRF or IRF of the system, are not directly applicable. Thus, we utilise the statistical properties of the ambient excitations.

Due to the statistical nature of the input assumed in OMA, the identification problem is a statistical data analysis problem. When we apply the related statistical theorem to the system equation, the presence of the natural excitation can be removed from the right hand side of Eq. (2.3). In the end, what is left behind is the homogenous expression for the system, which has the same poles and zeros as the original system. In practice, this can be easily achieved by calculating the auto- and cross-correlation function between the output responses. By doing so, the influence of the Gaussian white noise input can be removed, as it is an independent identically distributed (i.i.d) process with zero mean and constant covariance, and it is normally assumed to be uncorrelated to the output responses.

Since the correlation function of a stationary process for a system governed by a linear constant-coefficient differential or difference equation can be expressed as summations of decaying sinusoids

provided that all the roots of the characteristic equation of the associated homogenous equation have negative real parts, as a result, we can extract the natural frequencies and the corresponding damping ratios those decaying sinusoids [78], [79]. To some extent, one can say that conventional OMA is about the correlation function analysis.

In the time domain, methods such as the Eigensystem Realisation Algorithm (ERA) [80] has been modified to accommodate the stochastic properties of ambient excitation input. Used in conjunction with NExT, this version of ERA is normally referred to NExT-ERA [79][81][82]. Similar to NExT-ERA, the Covariance-Driven Stochastic Realisation (SSI-COV) method [83] [84] is also widely used in OMA. These two methods are derived based on a state-space model, and the modal parameters are calculated by evaluating the block Hankel matrix formed by the auto- and cross-correlation function of the output responses of the system.

While the SSI-DATA method [85] may have a state-space model base, but the essence of which is different from the NExT-ERA and the SSI-COV method. Such difference will be covered in section 2.4.3. As for the Polyreference method [86], because of the property of the correlation function for a stationary random process, this method is also widely used in OMA, so is the ITD method [64] [87] [88], which works on the free decay responses.

In EMA, to perform the identification, we can either deal with the impulse response of the system in the time domain or the FRF in the frequency domain. The IRF and FRF are Fourier transform pairs. Similarly, in OMA, we can estimate the modal parameters through the correlation function analysis of the output responses of the system in the time domain, or in the frequency domain, we can analyse the Fourier transform of the correlation functions. For stationary random processes, the Fourier transform of the correlation function is normally referred to as the Power Spectral Density (PSD) of the random process.

In the frequency domain, the FDD method, which was originally proposed to deal with the stationary process, is the most popular OMA technique due to its user-friendliness and effectiveness. Over the past 20 years, several versions of it have been developed, such as Enhanced FDD [89] [90], and Frequency-Spatial Domain Decomposition (FSDD) [91]. However, none of these versions of the FDD has exploited its true value. In chapter 4, we will present a new theory to further investigate this method in dealing with nonstationary processes and close modes analysis. Besides, it is worth mentioning that, as an opposite method to the FDD, the Time Domain Decomposition (TDD) method [92], [93] received some attention. However, compared to the FDD, the TDD method is not straightforward and user friendly, as it needs to use a digital band-filter to isolate each target mode in the first place.

Note that this subsection only gives a very brief introduction to OMA, there are many overview papers and books regarding this topic published in the past two decades, for more information, interested reader can refer to references [63], [83], [86], [94]–[97].

2.2 Linear differential equation with constant coefficients

In the last section, we provided a brief overview of system identification and modal analysis. In this section, we want to present some mathematical models which are widely used in modal analysis. Different mathematical models give different representations of a system, which in turn invoke different identification procedures. And those different identification procedures carry some fundamental ideas used in modal analysis.

In the introduction section, we have mentioned that the system of interest in Civil, Aerospace, and Mechanical Engineering is usually modelled by a linear constant-coefficient differential or difference equation, as such approximation is often justified, and design considerations based on linear theory lead to good results in many cases [62]. Therefore, we will restrict our discussion to a linear system with constant coefficients only. We will first introduce the continuous representation of a system, then discuss the discrete expressions in the next section.

2.2.1 Ordinary Differential Equation (ODE) and its solutions

The mathematical expression for a n th order linear differential equation with constant coefficients is often given by

$$x^{(n)} + a_1 x^{(n-1)} + \cdots + a_n x = u(t) \tag{2.5}$$

where a_i are the constant coefficient of the above ODE. $x(t)$ gives the system state variable, while $u(t)$ denotes the external stimuli. If $u(t) \neq 0$, the equation is inhomogeneous. Then the associated homogeneous equation is given by

$$x^{(n)} + a_1 x^{(n-1)} + \cdots + a_n x = 0 \tag{2.6}$$

It is well known that the general solution of Eq. (2.5) can be expressed as

$$x = x_c + x_p \tag{2.7}$$

where x_p is a particular solution to Eq. (2.5), and x_c is complementary solution to the homogeneous equation given by Eq. (2.6).

To find the expression for y given Eq. (2.5), we can utilised the differentiation operator D and write Eq. (2.5) in the form

$$\left(D^{(n)} + a_1 D^{(n-1)} + \cdots + a_n \right) x = u(t) \tag{2.8}$$

Define $p(D) = D^{(n)} + a_1 D^{(n-1)} + \cdots + a_n$, then $p(D)$ is the polynomial operator which allows us to use polynomial algebra to find the solutions of the original ODE given by Eq.(2.5).

Now, given $p(D)$, its characteristic equation is given by

$$p(s) = s^n + a_1 s^{n-1} + \cdots + a_n = 0 \tag{2.9}$$

If there are n distinct roots to $p(s) = 0$, then we can obtain n linearly independent solutions expressed in the exponential form $e^{s_i t}$ to Eq. (2.6).

Since any linear combination of the n linearly independent solutions is a solution to Eq. (2.6), then the general form of the complementary solution can be expressed as

$$y_c = c_1 e^{s_1 t} + \cdots + c_n e^{s_n t} \tag{2.10}$$

where c_i are constants determined by the initial conditions, and s_i is a root of $p(s) = 0$, which can be complex-valued.

If not all s_i are distinct to each other, for instance, s_1 is a k fold root to $p(s) = 0$, then

$$x_c = \left(c_1 + c_2 t + \cdots c_k t^{k-1} \right) e^{s_1 t} + \cdots + c_n e^{s_n t} \tag{2.11}$$

As for the particular solution x_p, it should have the form of $u(t)$ which in general bears no information of the system structure. Since x_p varies with $u(t)$, it is unnecessary for us to derive a particular expression for it in here.

In general, the goal of system identification is to determine the order of system and those constant coefficients a_i provided that the system behaviour can be approximated by Eq. (2.5). When we assume the system is of order 2, then the only task of system identification is to identify the coefficients a_1 and a_2. For a vibration system with multi degrees of freedom, we can decouple the higher order system into a series of the 2nd order systems by using a different set of coordinates. Therefore, if a system has N degrees of freedom, then its corresponding characteristic equation for the whole system is of order $n = 2N$, as for each degree of freedom we need a 2nd ODE to characterise its vibration properties.

However, obtaining those constant coefficients a_i is not our ultimate goal. In modal analysis, we are concerned of the natural frequencies and damping ratios, which are, as we mentioned before, embedded in the roots of the characteristic equation. Unfortunately, in many cases solving a n the order polynomial like Eq. (2.9) directly is challenging in practice. Thus, in the next section, we introduce the companion matrix, which provides us an effective way to find all the roots of the system.

2.2.2 The companion matrix

In order to find all the roots of the polynomial

$$p(s) = s^n + a_1 s^{n-1} + \cdots + a_n \tag{2.12}$$

when $p(s) = 0$, we can first define a companion matrix which contains all the coefficients of $p(s)$,

$$\mathbf{A}_C = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 1 \\ -a_n & -a_{n-1} & \cdots & -a_1 \end{bmatrix} \tag{2.13}$$

in which the first super-diagonal consists entirely of ones and all other elements above the last rows are zeros.

There is a nice property of this companion matrix that, the characteristic polynomial $\det(\mathbf{A}_C - \lambda \mathbf{I})$ of matrix $\mathbf{A}_C$ is the polynomial $(-1)^n p(s)$ [98], where λ is the eigenvalue of the companion matrix. Since s is a dummy variable, we can replace it with λ, then we have

$$(-1)^n p(\lambda) = \det(\mathbf{A}_C - \lambda \mathbf{I}) \tag{2.14}$$

Therefore, instead of solving $p(s) = 0$, we can solve $\det(\mathbf{A}_C - \lambda \mathbf{I}) = 0$. The eigenvalues of the companion matrix are the zeros of the original polynomial, which are also the homogenous solutions of a linear constant-coefficient difference equation.

Furthermore, according to the fundamental theorem of algebra [99], for a univariate polynomial of degree n with real coefficients, it will have n number of roots complex or real. If it has m complex roots, then those complex-valued roots occur as complex conjugate pairs, in this case, m is a multiple of 2. Thus, when n is odd, then $n > m$, which means the polynomial definitely has at least one real root. Whereas n is even, then $n \geq m$, which means all the roots could be complex conjugates.

2.2.3 Continuous time state-space model

The introduction of the companion matrix in the last section not just provided an alternative way to find the roots of the characteristic equation, it also implies that we can rewrite the n th order differential equation into a first order equation by using the companion, that is

$$
\begin{bmatrix} \ddot{x} \\ x^{(3)} \\ \vdots \\ x^{(n)} \end{bmatrix} = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 1 \\ -a_n & -a_{n-1} & \cdots & -a_1 \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \\ \vdots \\ x^{(n-1)} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ u(t) \end{bmatrix} \tag{2.15}
$$

In short,

$$
\dot{\mathbf{x}}(t) = \mathbf{A}_C \mathbf{x}(t) + \mathbf{u}(t) \tag{2.16}
$$

This equation not only has the same form of the state equation in the state-space model, the same as the state matrix, the companion matrix also contains all the information about the system. To see this, we first introduce the general form the state-space model, then use the 2nd order system given by Eq. (2.3) as an example.

The most general state-space representation of a linear system is given by

$$
\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) \\ \mathbf{y}(t) = \mathbf{C}(t)\mathbf{x}(t) + \mathbf{D}(t)\mathbf{u}(t) \end{cases} \tag{2.17}
$$

where $\mathbf{x}(t) \in \mathbb{R}^n$ is the state vector, $\mathbf{u}(t) \in \mathbb{R}^p$ is the input or control vector, and $\mathbf{y}(t) \in \mathbb{R}^q$ is the output vector. Here, n, p, q denotes the number of states, inputs, and outputs, respectively. While $\mathbf{A}(t)$ is the $n \times n$ state or system matrix, $\mathbf{B}(t)$ is the $n \times p$ input matrix, $\mathbf{C}(t)$ is the $q \times n$ output or observation matrix, and $\mathbf{D}(t)$ is the $q \times p$ feedthrough or feedforward matrix.

The first order differential equation given by the first equation in Eq. (2.17) is called the state equation, which tells us how the input affects the internal qualities of a system. Note that the number of states n (in a minimal realisation) is the order of the system. While the second equation is called the output equation. It quantifies how the states affect measurements $\mathbf{y}(t)$, thus $\mathbf{C}(t)$ could be thought of a calibration matrix. While $\mathbf{D}(t)$ which is known as the feedthrough term quantifies the direct effect of the input. The present of the direct feedthrough term means that there is a component of the output that changes instantaneously when the input changes. For a strictly causal system, $\mathbf{D}(t) = 0$. In the general formulation of the state-space model, all the matrices from $\mathbf{A}(t)$ to $\mathbf{D}(t)$ are allowed to be time-variant, while for a LTI system, all those matrices are time invariant.

Now consider the LTI dynamic system given by Eq. (2.3),

$$
\mathbf{M}\ddot{\mathbf{q}}(t) + \mathbf{C}\dot{\mathbf{q}}(t) + \mathbf{K}\mathbf{q}(t) = \mathbf{f}(t) \tag{2.18}
$$

where the mass matrix $\mathbf{M}$, damping matrix $\mathbf{C}$ and the stiffness matrices $\mathbf{K}$ are all $n \times n$ symmetric matrices.

If we define a state vector of dimension $2n \times 1$, that is

$$\mathbf{x}(t) = \begin{bmatrix} \mathbf{q}(t) \\ \dot{\mathbf{q}}(t) \end{bmatrix} \tag{2.19}$$

Its time derivative is given by

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} \dot{\mathbf{q}}(t) \\ \ddot{\mathbf{q}}(t) \end{bmatrix} \tag{2.20}$$

If we pre-multiply $\mathbf{M}^{-1}$ to Eq. (2.18), then rearrange the equation into

$$\begin{bmatrix} \dot{\mathbf{q}}(t) \\ \ddot{\mathbf{q}}(t) \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{I} \\ -\mathbf{M}^{-1}\mathbf{K} & -\mathbf{M}^{-1}\mathbf{C} \end{bmatrix} \begin{bmatrix} \mathbf{q}(t) \\ \dot{\mathbf{q}}(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \mathbf{M}^{-1}\mathbf{f}(t) \end{bmatrix} \tag{2.21}$$

or more simply

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \tag{2.22}$$

where

$$\mathbf{A} = \begin{bmatrix} 0 & \mathbf{I} \\ -\mathbf{M}^{-1}\mathbf{K} & -\mathbf{M}^{-1}\mathbf{C} \end{bmatrix} \tag{2.23}$$

is the system matrix. $\mathbf{u}(t)$ is the $p \times 1$ input vector at p locations, it is associated with the $n \times 1$ input vector $\mathbf{f}(t)$, namely

$$\mathbf{f}(t) = \mathbf{F}\mathbf{u}(t) \tag{2.24}$$

Here, $\mathbf{F}$ is the $n \times p$ input coefficient matrix. Thus, the $n \times p$ input matrix $\mathbf{B}$ is given by

$$\mathbf{B} = \begin{bmatrix} 0 \\ \mathbf{M}^{-1}\mathbf{F} \end{bmatrix} \tag{2.25}$$

In the meantime, the measured responses $\mathbf{y}(t)$ at q physical coordinates are related to the state vector $\mathbf{x}(t)$ through an $q \times 2n$ observation matrix $\mathbf{C}$,

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) \tag{2.26}$$

Together with Eq. (2.22), we have the state-space model for the 2nd order dynamic system, i.e.,

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \\ \quad \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) \end{cases} \tag{2.27}$$

The continuous form of the state-space model is of no practical use in modal analysis, as well usually deal with discrete time signals. Therefore, in section 2.3., we will introduce one of the discrete form of Eq. (2.27), namely linear Gauss- Markov model along with the famous Kalman Filter.

2.2.4 Transfer function and the Frequency Response Function (FRF) of the system

The discussion of the characteristic polynomial of the system establishes the foundation for many time domain methods based on the state-space model and the ARMA models, as those methods all end up with computing the companion matrix to find the poles of the system. However, this is not the only way to estimate the poles.

In order to gain some ideas about the system while we have almost no information about the mathematical structure of the system, we can use some transformation techniques to transfer the input-output or output only responses of the system to another domain, where we can gain some insight about the behaviour of system poles nonparametrically.

One of the well-known transformations is the Laplace transform [64], which works for continuous signals, whereas the so-called z-transform is the discrete counterpart of it [100]. As for the famous Fourier transform, it is just a special case of the Laplace transform evaluated on the imaginary axis. However, due to some nice properties of the Fourier transform and the computation advantage of the Fast Fourier Transform (FFT), the Fourier transform is more widely used in data analysis than the Laplace transform and the z-transform. Nevertheless, we will start our discussion with the Laplace transform, as it is more general than the Fourier transform in solving linear differential equations.

Given a function $h(t)$ of a real variable t, the Laplace transform of $h(t)$ is given by [101]

$$H(s) = \int_{-0}^{\infty} h(t)e^{-st}\mathrm{d}t \tag{2.28}$$

where $H(s) = \mathcal{L}\{h(t)\}$ is a function of s. Here, s is known as the complex frequency, which is in general complex, i.e., $s = \sigma \pm i\omega$.

Note that in the vibration analysis, when a system subjects to damping, we rewrite ω as ω_d. And if we denote ω and ξ as the natural frequency and the corresponding the damping ratio respectively, we can express s as

$$s = -\omega\xi \pm i\omega\sqrt{1-\xi^2} \tag{2.29}$$

where $\omega_d = \omega\sqrt{1-\xi^2}$ is the so-called damped frequency.

Now if we apply the Laplace transform to Eq. (2.5) on both side of the equation, using the time-derivative rule [101], and assume the system is at rest initial, in other words, $\mathcal{L}\{h^{(n)}(t)\} = s^n H(s)$, we will obtain

$$\left(s^n + a_1 s^{n-1} + \cdots + a_n\right)X(s) = U(s) \tag{2.30}$$

or more simply

$$p(s)X(s) = U(s) \tag{2.31}$$

With this expression in the system equation in the Laplace domain, we can define the transfer function of the system $H(s)$ to be

$$H(s) = \frac{X(s)}{U(s)} = \frac{1}{p(s)} \tag{2.32}$$

where $X(s)$ and $U(s)$ represents the Laplace transform of the output and input, respectively.

Now if we apply the Heaviside cover-up method [100] to $H(s)$, we can decompose it into partial factions. Suppose the all the poles are distinct, we can obtain the following form for the transfer function,

$$H(s) = \frac{A_1}{s-s_1} + \cdots + \frac{A_n}{s-s_n} \tag{2.33}$$

where s_i are the system poles, they are also the roots of the characteristic equation. A_i are the coefficients which can be determined by the cover-up method. Note that if and only all the s_i have negative real part, i.e., $\sigma_i < 0$, $\forall i = 1, \ldots, n$, the system is stable.

Moreover, the transfer function $H(s)$ and the impulse response function $h(t)$ of the system are Laplace transform pairs. Thus, if we apply the inverse Laplace transform to $H(s)$, we will end up with Eq. (2.4).

Particularly, if we set $s = i\omega$, then $H(s) = H(i\omega)$ we obtain the complex gain of the system. For the 2nd order system follows the Newton's the law, the equation of motion is given by

$$m\ddot{x} + c\dot{x} + kx = u(t) \tag{2.34}$$

22

where m, c, and k represent the mass, damping, and stiffness coefficient, respectively. And these three coefficients are interrelated by $c = 2m\omega\xi$ and $k = m\omega^2$. Thus, for this system, the complex gain is simply given by

$$H(i\omega) = H(s)\big|_{s=i\omega} = \frac{1}{k - \omega^2 m + ic\omega} \tag{2.35}$$

which is known as the system's Frequency Response Function (FRF). If the system is excited by some sinusoidal input, then the FRF is simply the ratio between the amplitude of the output response and the input response.

Accordingly, we can obtain the amplitude response of the system, which is the absolute value of the FRF, that is

$$\left|H(i\omega)\right| = \frac{1}{\sqrt{\left(k - \omega^2 m\right)^2 + c^2\omega^2}} \tag{2.36}$$

and the phase lag

$$\varphi(\omega) = -\tan^{-1}\left(\frac{c\omega}{k - m\omega^2}\right) \tag{2.37}$$

In practice, when doing system identification, all we have some information about $X(s)$ and $U(s)$. We can calculate the FRF, but the actual structure is unknown. However, if the system can be approximated by Eq. (2.34), then we can make inference about the modal parameters of the system according to Eq. (2.35) through evaluating the ratio between $X(s)$ and $U(s)$.

2.3 Linear difference equation with constant coefficients

As the discrete counterpart of the linear differential equation, the linear difference equation is widely used in Digital Signal Processing (DSP) [100] and Time Series Analysis (TSA) [102]. The basic formulation of the nth order linear difference equation with constant coefficients has the following form

$$x_t = a_1 x_{t-1} + \cdots + a_n x_{t-n} + u_t \tag{2.38}$$

where a_i are the constant coefficients. When the extra term $u_t = 0$, the equation is homogeneous, otherwise it's nonhomogeneous.

Typically, there are many variations of Eq.(2.38) depending on the order of the equation and the conditions of u_t. For instance, when u_t is the uncorrelated white noise with only one term, then Eq. (2.38) denotes a autoregression (AR) model; when u_t is the combination of the current and past error terms, in the meantime, we assume $n=1$, $x_0 =1$, then Eq. becomes a moving average (MA). Apart from this, we can also rewrite Eq. (2.38) into the first order form with a companion matrix.

It is obvious that we are unbale to discuss all the variations of the linear difference equation in this thesis. Therefore, in the next subsection, we will consider the case when u_t is controllable and nonstatistical and give the convolution formula for the input and IRF of a system. Then, we will formally introduce some of the autoregressive-moving-average (ARMA) related models with the conventional notation. After that, we will shift our focus on to the discrete time state-space model. Having done this, we will briefly discuss the state-space model for a stochastic system.

2.3.1 Deterministic input and z-transform

In system identification, with the controllable deterministic input, we can easily establish the relation between the input and the output via the convolution formula of the input and the impulse response function.

As a result, it is common to work with a linear constant-coefficient difference equation of this kind,

$$x_t = -\sum_{i=1}^{N} a_i x_{t-i} + \sum_{j=0}^{M} b_j u_{t-j} \tag{2.39}$$

Or equivalently,

$$\sum_{i=0}^{N} a_i x_{t-i} = \sum_{j=0}^{M} b_j u_{t-j} , \quad a_0 \equiv 1 \tag{2.40}$$

where x_t is the output, while u_t denotes the input. a_i and b_j .are the constant coefficients for the output and input, respectively. N is the order of the system, while M is the order of the input.

In section , we have briefly mentioned a multiplicative relation between the Laplace transform of the input and the transfer function of the system to generate the output response in the Laplace domain, i.e., $X(s)=H(s)U(s)$. Whereas in the time domain, such a simple relation becomes a convolution, which has the form of Eq. (2.1) continuous. For a system characterised by a linear difference equation with an underlying discrete assumption, it also has a transfer function. However, to derive it, we need to use the z-transform, which is defined as follows

$$H(z) = \sum_{k=-\infty}^{\infty} h(k) z^{-k} \tag{2.41}$$

This equation says that if $h(n)$ is the discrete time impulse response function given by the same form of Eq. (2.2), then $H(z)$ gives the corresponding transfer function in the z domain.

If we consider the discrete time formula of Eq. (2.40), i.e.,

$$\sum_{i=0}^{N} a_i x(n-i) = \sum_{j=0}^{M} b_j u(n-j), \quad a_0 \equiv 1 \tag{2.42}$$

and apply the z-transform on both sides of the equation, in the meantime, we assume the system is at rest initially, then we can obtain the expression for $H(z)$, namely

$$H(z) = \frac{X(z)}{U(z)} = \frac{\sum_{j=0}^{M} b_j z^{-j}}{1 + \sum_{i=0}^{N} a_i z^{-i}} \tag{2.43}$$

Note that the same as s, z is also complex-valued. As for the convergence discussion of the $H(z)$, readers can refer to reference [100].

2.3.2 Autoregressive-moving-average (ARMA) models

In time series analysis, a (weakly) stationary stochastic process is usually expressed by ARMA (p,q) models in terms of two polynomials, i.e., autoregression (AR) and the second for the moving average (MA), where p and q represent is the order of the AR part and the MA part, respectively. These models, which set up a foundation for system identification with random inputs which are unobservable, play a significant role in system identification.

Given a time series $\{X_t\}$, the ARMA (p,q) model, is usually used for understanding or predicting future values in $\{X_t\}$. When $q = 0$, we obtain the AR(p) model, which regresses the variables on its own past values; whereas when $p = 0$, we get MA(q) model, which models the error term as a linear combination of the past error terms. Note that for the MA model, one should not be confused with smooth technique moving average.

The study of the ARMA models and their extensions are important subjects in the time series analysis. Especially for the AR model, which is written in the form of a stochastic difference equation, is a direct application of the linear difference equations. Particularly, many OMA time domain techniques are

derived in terms of these models. Thus, it is important for us to understand them before we introduce those techniques.

2.3.2.1 Basic models

For the AR(p) model with order p, it can be given by

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t \tag{2.44}$$

where ϕ_i are parameters, c is a constant, while ε_t denotes white noise. Note that the terms ε_t are generally assumed to be i.i.d sampled from a normal distribution with zero mean, i.e., $\varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$, where σ_ε^2 is the variance.

For the $\mathrm{MA}(q)$ model with order q,

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{2.45}$$

where θ_i are the parameters of the model associated with the white noise error terms. μ is the expected value of $\{X_t\}$, quite often, we assume $\mu = 0$.

For the $\mathrm{ARMA}(p,q)$ model with p AR terms and q MA term, the model is expressed as

$$X_t = c + \varepsilon_t + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{2.46}$$

which is simply the merger of $\mathrm{AR}(p)$ and $\mathrm{MA}(q)$ models. Since the constant c absorbs the means, and adding means and other deterministic trends is easy, so in the following analysis we will be working with the mean zero versions. In other words, we will assume $c = 0$.

2.3.2.2 Stationarity of the random process

The invertibility of the ARMA models is very important. For example, to study an IRF, MA representations may be convenient, whereas to estimate an ARMA model, AR representations will be more convenient, as usually $\{X_t\}$ is observable while ε_i is not. This corresponds to two types of study in system analysis. One works on the causality, using the IRF to describe the evolution of a model's variables in reaction to a shock in one or more variables, in which case, we wish to convert our process into an MA model-based process (ε_t can be an exogenous shock). The other one works on prediction,

and the autoregressions simplify the task. In other words, we use the most recent p cases to predict the future response. In this case, an AR-based model is of great interest.

In the next section, we consider the conditions for a random process expressed by the ARMA models to be stationary, which is associated with the invertibility of the characteristic polynomials of the ARMA models.

2.3.2.3 Invertibility and stationarity of the ARMA processes

To facilitate our investigation of the invertibility and stationarity of the ARMA processes, first, we want to derive the characteristic polynomial for the ARMA models.

The same as the differentiation operator, we introduce the backshift operators, which enable us to present ARMA models in a more concise way. The backshift operator moves the index back one time unit, that is

$$LX_t = X_{t-1} \tag{2.47}$$

and applying it i times, we move the index back i units

$$L^2 X_t = X_{t-2}, \quad \cdots \quad , L^i X_t = X_{t-i} \tag{2.48}$$

Using the backshift operator, we can write the ARMA models as

$$\text{AR}(p): \left(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p\right) X_t = \varepsilon_t \text{ or } \phi(L) X_t = \varepsilon_t \tag{2.49}$$

$$\text{MA}(q): X_t = \left(1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q\right) \varepsilon_t \text{ or } X_t = \theta(L) \varepsilon_t \tag{2.50}$$

$$\text{ARMA}(p,q): \phi(L) X_t = \theta(L) \varepsilon_t \tag{2.51}$$

where the characteristic polynomials are

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p \tag{2.52}$$

and

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q \tag{2.53}$$

Note that here we assume the constants c and μ are equal to zero.

The invertibility property of the ARMA models encompass an important topic in time series analysis, i.e., stationarity. In the ARMA models, the stationarity is characterised by the roots and coefficients of the characteristic polynomials $\phi(z)$ and $\theta(z)$. For instance, when an AR process can be inverted to

an MA process, then the process is stationary. This is possible if and only if $\phi(z) \neq 0$ for all z in a unit circle. It can be seen clearly by evaluating Eq. (2.49). Hence, if all λ_i in

$$1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p = (1 - \lambda_1 L)(1 - \lambda_2 L)\cdots(1 - \lambda_p L) \tag{2.54}$$

have less than one absolute value, then the AR(p) process can be inverted to an MA process, which is given by Eq. (2.50) . This means that with $|\lambda_i| < 1$, $\{\psi_k\}$ ($k = 0, 1, \ldots$) is absolutely summable, where ψ_k are the coefficients of $\psi(L)$, and $\psi(L) = \phi^{-1}(L)$ for the AR-based model. Here, one can see that for the AR(1) process, to ensure the process is invertible and stationary, the condition $|\phi_1| < 1$ must be satisfied.

One the other hand, we can express Eq. (2.54) as

$$(L - y_1)(L - y_2)\cdots(L - y_p) = 0 \tag{2.55}$$

that all roots in Eq. (2.55) lie outside of the unit circle, i.e., $|y_i| > 1$ for all i , is equivalent to the requirement that $|\lambda_i| < 1$. Therefore, one can say that if and only if all roots of the characteristic polynomial $\phi(z)$ lie outside the unit circle.

For the MA process with finite number of MA coefficients, it is always stationary. When $q \to \infty$, in order to guarantee finite mean, we require $\{\theta_i\} \in \ell_1$. ℓ_1 is the space of absolutely summable sequence, $\{\theta_i\} \in \ell_1$ means $\sum|\theta_i| < \infty$. While in order for $\{X_t\}$ to have second moments, then $\{\theta_i\} \in \ell_2$, where ℓ_2 is the space of all square summable sequences, those for which $\sum\theta_i^2 < \infty$. Similarly to the AR process, if $\theta^{-1}(z)$ exists, an MA(q) process is invertible if all roots of $\theta(z)$ lie outside of the unit circle. This is implies that for an MA(1) process to be invertible if and only if $|\theta_1| < 1$.

As for an ARMA process, its stationarity is completely depended on its AR part. It is because a finite MA process is always stationary. So, if we can convert an ARMA process to an MA process, then its stationarity is secured. To satisfy this condition, its autoregressive lag polynomial must be invertible and the polynomials $\phi(z)$ and $\theta(z)$ have no common zeros. Thus, given an invertible ARMA(p, q) process, its MA form is given by

$$X_t = \phi^{-1}(L)\theta(L)\varepsilon_t = \psi(L)\varepsilon_t \tag{2.56}$$

where $\psi(L)$ give the moving average polynomial for the ARMA(p, q) process.

2.3.2.4 AR covariance function and Yule-Walker equations

In this section, we introduce the Yule-Walker equation equations, which are used to compute the AR coefficients. In section 2.4.1.2 , we will refer to these equations to explain the Polyreference method used in OMA.

According to our discussion in the last section, for a stationary AR process given by Eq. (2.49), it has a moving average representation of the process, that is

$$X_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} = \psi(L)\varepsilon_t \tag{2.57}$$

This implies that ε_t is uncorrelated with prior observations X_s, $s < t$.

Since ε_t is assumed to be white noise, $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, then the first moment of X_t equals 0, namely $\mathbb{E}[X_t] = 0$. Hence, for lags $k = 0, 1, 2, \dots$

$$\mathbb{E}\left[X_{t-k}\left(X_t - \phi_1 X_{t-1} - \cdots \phi_p X_{t-p} \right) \right] = \mathbb{E}[X_{t-k}\varepsilon_t] = \delta_k \sigma^2 \tag{2.58}$$

where $\delta_k = 0$ for $k \neq 0$.

For a stationary process, the autocovariance is only the function of the time difference. Thus, if we define $\gamma(k) = \mathbb{E}\left[X_{t+|k|} X_t \right]$, then the above equation can be simplified as follows

$$\delta_k \sigma^2 = \gamma(k) - \sum_{j=1}^{p} \phi_j \gamma(k-j) \tag{2.59}$$

which is the Yule-walker equations. Particularly, when $k = 0$,

$$\sigma^2 = \gamma(0) - \phi_1 \gamma(1) - \cdots \phi_p \gamma(p) \tag{2.60}$$

When $k = 1$

$$0 = \gamma(1) - \phi_1 \gamma(0) - \phi_2 \gamma(1) - \cdots - \phi_p \gamma(p-1) \tag{2.61}$$

From $k = 1$ to $k = p$, we can write p equations that give the coefficients as follows

$$\begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \gamma(3) \\ \vdots \\ \gamma(p) \end{bmatrix} = \begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \cdots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \cdots & \gamma(p-2) \\ \gamma(2) & \gamma(1) & \gamma(0) & \cdots & \gamma(p-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma(p-1) & \gamma(p-2) & \gamma(p-3) & \cdots & \gamma(0) \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \vdots \\ \phi_p \end{bmatrix} \tag{2.62}$$

which is the Yule-Walker equations.

Define the vectors $\boldsymbol{\gamma} = \left(\gamma(1), \quad \cdots, \quad \gamma(p) \right)'$ and $\phi = \left(\phi_1, \quad \cdots, \quad \phi_p \right)'$, and the matrix $\boldsymbol{\Gamma}_p = \left[\gamma(j-k) \right]_{j,k=1,\ldots,p}$. The matrix form of Eq. (2.62) is given by

$$\boldsymbol{\gamma} = \boldsymbol{\Gamma}_p \phi \tag{2.63}$$

where $\boldsymbol{\Gamma}_p$ is a Toeplitz matrix.

Now let us consider a special case AR(1). When $k \neq 0$, from Eq. (2.59), we have

$$\gamma(k) = \phi_1 \gamma(k-1) \tag{2.64}$$

which gives a recursive form of the autocovariance of AR(1) process.

Further, we can obtain

$$\gamma(k) = \phi_1^k \gamma(0) \tag{2.65}$$

Thus, the autocorrelation function for the AR(1) process is given by

$$\rho(k) = \phi_1^k \tag{2.66}$$

This equation implies that for the AR(1) process, the parameter, ϕ_1, can be discerned from the autocorrelation plot. For stationary process, the absolute value of ϕ_1 is required to be less than one. Particularly, when $\phi_1 > 0$, we will have an exponentially decay plot of $\rho(k)$. Whereas, $\phi_1 < 0$, since $\rho(k)$ has an extra $(-1)^k$ factor, it will cause the $\rho(k)$ to oscillate while decaying.

On the other hand, if we invert the AR(1) to an MA process, using Eq. (2.49), we have

$$X_t = \frac{\varepsilon_t}{1 - \phi_1 L} \tag{2.67}$$

According to the Binomial theorem, we obtain

$$X_t = \left(1 + \phi_1 L + \phi_1^2 L^2 + \cdots \right) \varepsilon_t = \sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j} \tag{2.68}$$

where

$$h(k) = \phi_1^k \tag{2.69}$$

is the IRF of the AR(1) process, which has the same format as the autocorrelation function given by Eq. (2.66).

2.3.3 Discrete time state-space model and the Markov parameters

Since the discrete-time state-space model is of practical importance, in this section, we will present the discrete version Eq. (2.27) (the vector form) first, then derive the expression of the corresponding Markov parameters, which is essentially the impulse response function of the system.

To obtain the discrete version of Eq. (2.27), we can either use the Euler's method [64]. The details will be omitted here, instead, we give the result directly

$$\begin{cases} \mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k) \\ \qquad \mathbf{y}(k) = \mathbf{C}\mathbf{x}(k) \end{cases} \tag{2.70}$$

Then with the given initial state $\mathbf{x}(0)$ and the input $\mathbf{u}(k)$, and from which we can derive the Markov parameters [103].

Consider the vector form of Eq.(2.2) of a causal IIR system,

$$\mathbf{y}(n) = \sum_{k=0}^{\infty} \mathbf{h}(k)\mathbf{u}(n-k) \tag{2.71}$$

with the zero-initial condition. Our goal is to derive an explicit expression of $\mathbf{h}(k)$ in term of Eq. (2.70) is the Markov parameters of the system.

Now, if we solve for the observable output $\mathbf{y}(k)$ in terms of the previous inputs in the following fashion in terms of Eq. (2.70)

$$\begin{cases} \mathbf{x}(1) = \mathbf{A}\mathbf{x}(0) + \mathbf{B}\mathbf{u}(0) \\ \qquad \mathbf{y}(0) = \mathbf{C}\mathbf{x}(0) \\ \mathbf{x}(2) = \mathbf{A}\mathbf{x}(1) + \mathbf{B}\mathbf{u}(1) = \mathbf{A}^2\mathbf{x}(0) + \mathbf{A}\mathbf{B}\mathbf{u}(0) + \mathbf{B}\mathbf{u}(1) \\ \qquad \mathbf{y}(1) = \mathbf{C}\mathbf{x}(1) = \mathbf{C}\mathbf{A}\mathbf{x}(0) + \mathbf{C}\mathbf{B}\mathbf{u}(0) \end{cases} \tag{2.72}$$

it yields

$$\begin{cases} \mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k) = \mathbf{A}^{k+1}\mathbf{x}(0) + \sum_{i=0}^{k} \mathbf{A}^{k-i}\mathbf{B}\mathbf{u}(i) \\ \qquad \mathbf{y}(k) = \mathbf{C}\mathbf{A}^k\mathbf{x}(0) + \sum_{i=1}^{k} \mathbf{C}\mathbf{A}^{i-1}\mathbf{B}\mathbf{u}(k-i) \end{cases} \tag{2.73}$$

Suppose the initial conditions is zero, namely, $\mathbf{x}(0) = \mathbf{0}$, we have

$$\begin{cases} \mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k) = \sum_{i=0}^{k} \mathbf{A}^{k-i}\mathbf{B}\mathbf{u}(i) \\ \mathbf{y}(k) = \sum_{i=1}^{k} \mathbf{C}\mathbf{A}^{i-1}\mathbf{B}\mathbf{u}(k-i) \end{cases} \tag{2.74}$$

where for $k < 0$, $\mathbf{u}(k) = 0$.

Now if we compare the observation equation in Eq. (2.73) to Eq. (2.71), then the Markov parameters in Eq. (2.71) is given by

$$\mathbf{h}(k) = \mathbf{C}\mathbf{A}^{k-1}\mathbf{B} \tag{2.75}$$

where the integer k shall start from 1, and $\mathbf{h}(k)$ is a $q \times p$ matrix.

From Eq. (2.75), we can see that the Markov parameter sequence is simply the impulse response of the system. Therefore, they should be unique for a given system [103]. However, since the state vector is coordinate dependent, the state-space matrices are not unique, which will result in a non-unique Markov parameter sequence [103]. This is undesirable. Nevertheless, this does not mean the system's innate property will be different. For instance, the modal parameters of the system should remain the same regardless of the format of the impulse response sequence.

Apart from this, it shall be pointed out that the essential step in Eigensystem Realisation Algorithm (ERA) [80] method is about decomposing the block Hankel matrix formed by the Markov parameters given by Eq. (2.75), which will be explained in section 2.3. One should be aware that Eq. (2.75) gives the Markov parameters for a deterministic system, and its counterpart for the stochastic system has the same format, which is used the NExT-ERA method and the SSI-COV method. Since these two methods are fundamentally similar, we will only discuss the SSI-COV method. As an important extension, in the next subsection, we shall introduce a widely used state-space model for a stochastic system.

2.3.4 Linear Gauss-Markov model and state estimation

In OMA, the basic assumption for modal analysis is that the structures are constantly subjected to some unmeasurable ambient excitations. The same as the introduction of the ARMA models we discussed before in terms of a n th order difference equation, we need to modify the discrete time state-space model to accommodate the statistical properties of the input and the output responses of a system.

In last section, we discussed the discrete version of a state-space model for a deterministic system. In this section, we assume the system is excited by some Gaussian white noise. Typically, the state-space model of such a system is often referred to as a linear Gauss-Markov model, which can be given by

$$\begin{cases} \mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathcal{G}\mathbf{w}(k) \\ \mathbf{y}(k) = \mathbf{C}\mathbf{x}(k) + \mathcal{H}\mathbf{v}(k) \end{cases} \tag{2.76}$$

where, $\mathcal{G}$ is a $n \times p$ process noise to state matrix, while $\mathcal{H}$ is a $q \times p$ measurement noise to output matrix. Here, we suppose both $\mathbf{w}(k)$ and $\mathbf{v}(k)$ are of size $p \times 1$.

Here, $\mathbf{x}(k+1)$ is a $n \times 1$ vector, which only depends on the previous state at time k, is an the n-dimensional hidden state process, and it is corrupted by a (correlated or uncorrelated, here we consider the latter case) state noise process $\mathbf{w}(k)$ [104]. While for the q-dimensional measurement $\mathbf{y}(k)$, which is further corrupted by a measurement noise process $\mathbf{v}(k)$, is subject to a linear transformation of the hidden state $\mathbf{x}(k)$ [104]. Note that the state equation in Eq. (2.76) implies $\mathbf{x}(k+1)$ follows a first-order Markovian dynamics, which describes the state-space evolution of a stochastic dynamical system [104].

When $\mathbf{w}(k)$ and $\mathbf{v}(k)$ are assumed to be uncorrelated Gaussian process with zero mean, then $\mathbf{y}(k)$ is a Gaussian process as well. To be more specific, for a multivariate linear Gaussian system, we would expect

$$\mathbb{E}\left[\mathbf{w}(k)\right] = 0 \text{ and } \mathbb{E}\left[\mathbf{v}(k)\right] = 0 \tag{2.77}$$

with the autocovariance matrices given by

$$\mathbb{E}\left[\mathbf{w}_i(k)\mathbf{w}_j^{\mathbf{T}}(k)\right] = \begin{cases} \mathbf{Q} & i = j \\ 0 & i \neq j \end{cases} \text{ and } \mathbb{E}\left[\mathbf{v}_i(k)\mathbf{v}_j^{\mathbf{T}}(k)\right] = \begin{cases} \mathbf{R} & i = j \\ 0 & i \neq j \end{cases} \tag{2.78}$$

and the cross-covariance matrix given by

$$\mathbb{E}\left[\mathbf{w}(k)\mathbf{v}^{\mathbf{T}}(k)\right] = \mathbf{S} \tag{2.79}$$

where $\mathbb{E}\left[\cdot\right]$ gives the notation for expectation.

In the meantime, we usually assume that noise processes are uncorrelated with the state process and the system output, thus,

$$\mathbb{E}\left[\mathbf{w}(k)\mathbf{x}^{\mathrm{T}}(k)\right]=0 \qquad \mathbb{E}\left[\mathbf{v}(k)\mathbf{x}^{\mathrm{T}}(k)\right]=0$$
$$\mathbb{E}\left[\mathbf{w}(k)\mathbf{y}^{\mathrm{T}}(k)\right]=0 \qquad \mathbb{E}\left[\mathbf{v}(k)\mathbf{y}^{\mathrm{T}}(k)\right]=0 \tag{2.80}$$

The linear Gauss-Markov model has very wide application. The widely used OMA methods, such as NexT-ERA method, the SSI-COV method and the SSI-DATA method, are all derived based on Eq. (2.76). In system control, the estimation of state-space models is normally realised by ways of the Kalman Filter (KF) [105], which is a recursive set of equations to update the estimated parameters. And the idea of updating in the KF is related to the Bayesian approach [106]. Indeed, the theory behind the Kalam Filter is Bayesian. Particularly, the SSI-DATA method is formulated based on the state estimate equation in the KF process. Now consider the following Lemma [105].

Lemma 1. *Suppose X and Y are jointly Gaussian random variables, with $Z=[X^{\mathrm{T}}\ Y^{\mathrm{T}}]^{\mathrm{T}}$, then Z is gaussian with mean $\mu_z \in \mathbb{R}^n$ and covariance matrix $\Sigma_{zz} \in \mathbb{S}^n_{++}$ (i.e., the space of symmetric positive definite $n \times n$ matrices)*

$$\mu_z = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} \quad \text{and} \quad \Sigma_{zz} = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} \tag{2.81}$$

Then the conditional probability distribution of X given Y, i.e., $p_{x|y}(x|y)$ follows a Gaussian distribution with mean and covariance

$$\mathbb{E}\left[x|y\right]=\mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}\left(y-\mu_y\right) \quad \text{and} \quad \Sigma_{xx}-\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} \tag{2.82}$$

Note that the conditional expectation of X given $Y = y$ in Eq. (2.82) is the prime equation of the linear Kalman filter, which is used to estimate the state variable based on the past measurements up to the time instant k. In other words, if $\mathbf{y}(k)$ denotes the measurement process, we can estimate the system state $\mathbf{x}(k)$ in terms of the past observations $\mathbf{y}(0),\mathbf{y}(1),\dots,\mathbf{y}(k)$ via the conditional expectation formula, i.e., $\mathbb{E}\{\mathbf{x}(k)|\mathbf{y}(0),\mathbf{y}(1),\dots,\mathbf{y}(k)\}$ for $k=0,1,2,\dots$. For simplicity, we can use $\mathcal{Y}_k = \{\mathbf{y}(0),\mathbf{y}(1),\dots,\mathbf{y}(k)\}$ to denote the set of the measured quantities. Then the estimate of the system state at time k based on the past observations up till time k is given by $\hat{\mathbf{x}}(k|k)=\mathbb{E}\left[\mathbf{x}(k)|\mathcal{Y}_k\right]$.

For the SSI-DATA method used in OMA, the whole method is realised by computing the conditional expectation of the state variable based on the measurements. In other words, the realisation of the SSI-DATA method is not a Kalman filtering problem, but a state estimation problem. A detailed discussion is presented in section 2.4.3.2.

2.4 Identification of the constant coefficients

There are essentially three important matrices in system identification, i.e., the (block) Hankel matrix, (block) Toeplitz matrix and the (block) companion matrix. To estimate the system parameters from the MDOF system by using traditional time domain SIMO or MIMO methods, these three matrices are almost unavoidable.

We have already met the companion matrix in section 2.2.2, which is used to compute the system poles provided that the system coefficients are known. While for the (block) Hankel matrix and (block) Toeplitz matrix, they are essentially used to compute the system coefficients. Therefore, the main task in this section if to form the (block) Hankel matrix or (block) Toeplitz matrix with the available information.

To be more specific, the (block) Hankel matrix is introduced by using a simple AR model first, based on which we provide a full description of the Polyreference method. Then we formulate the (block) Hankel matrix for linear deterministic system. After that, we show its formulation for the linear stochastic system, important examples, such as the SSI-COV method and SSI-DATA method are investigated in this section.

As for the (block) Toeplitz matrix, which is closely related to the (block) Hankel matrix, is not discussed separately on this occasion. However, it will be mentioned as appropriate. For instance, the Toeplitz matrix associated with the Yule-Walker equations, see Eq. (2.62).

2.4.1 AR coefficients estimation

Essentially, the Polyreference method is an AR-based method, which is initially developed by Vold et al. [107] to solve the MIMO problem. Later on, it was modified by Lingmi Zhang et al [57] to deal with the modal parameters estimation problems based on output-only responses. In reference [86], Zhang updated the theoretical background of the Polyreference method and coined it with the name NExT-PRCE method. The Polyreference method we will be discussing in this thesis in one used in OMA. Instead of the using presenting the existing theoretical background for Polyreference, we provide a more rigorous and easier explanation for it in this section. Before doing that, we need to know how to obtain those AR coefficients first.

2.4.1.1 Homogenous equation of an AR process

For the AR model-based identification approaches, normally the first step is to find the AR coefficients, and this procedure is associated with forming a (block) Hankel matrix. First, let us consider a simple case by looking at the homogenous part of a vector AR model, that is [63]

$$\mathbf{y}(n) = \mathbf{A}_1 \mathbf{y}(n-1) + \mathbf{A}_2 \mathbf{y}(n-2) + \cdots + \mathbf{A}_{na} \mathbf{y}(n-na) \tag{2.83}$$

Where na is the number of AR coefficient matrices, which also implies the order of the system. When we have nc number of channels for $\mathbf{y}(n)$, i.e., $\mathbf{y}(n)$ is a $nc \times 1$ vector. Note that the discrete and matrix form of the AR model we used here is normally referred to as VAR model, where V stands for the vector. Accordingly, we have VARMA models as well. However, in the rest of the context in this thesis, we will not particularly address the "vector" for the ARMA models.

In order to find out the na coefficient matrix, we need to solve at least na linear equations like Eq.(2.83). For instance, given $\mathbf{y}(n)$, $n = na+1, na+2,..., np$ with $np - na$ number of samples, where $np > na$, we have

$$\begin{cases} \mathbf{y}(na+1) = \mathbf{A}_1 \mathbf{y}(na) + \mathbf{A}_2 \mathbf{y}(na-1) + \cdots + \mathbf{A}_{na} \mathbf{y}(1) \\ \mathbf{y}(na+2) = \mathbf{A}_1 \mathbf{y}(na+1) + \mathbf{A}_2 \mathbf{y}(na) + \cdots + \mathbf{A}_{na} \mathbf{y}(2) \\ \qquad\qquad\qquad\vdots \\ \mathbf{y}(np) = \mathbf{A}_1 \mathbf{y}(np-na) + \mathbf{A}_2 \mathbf{y}(np-na+1) + \cdots + \mathbf{A}_{na} \mathbf{y}(np-na) \end{cases} \tag{2.84}$$

Note that to avoid negative time lapses and secure the causal nature of the system we concern, the starting time n should at least start from na. Here we set n from $na+1$ is because we wish to derive a comparable block Hankel matrix with our discussion in the next subsection.

These linear equations can be rewritten into many different matrix forms. To our particular interest, we would like to use the following format

$$\begin{bmatrix} \mathbf{y}(na+1) & \mathbf{y}(na+2) & \cdots & \mathbf{y}(np) \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{A}_{na} & \mathbf{A}_{na-1} & \cdots & \mathbf{A}_1 \end{bmatrix} \begin{bmatrix} \mathbf{y}(1) & \mathbf{y}(2) & \cdots & \mathbf{y}(np-na) \\ \mathbf{y}(2) & \mathbf{y}(3) & \cdots & \mathbf{y}(np-na+1) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}(na) & \mathbf{y}(na+1) & \cdots & \mathbf{y}(np-1) \end{bmatrix} \tag{2.85}$$

where the (block) Hankel matrix is given by

$$\mathbf{H}_{na} = \begin{bmatrix} \mathbf{y}(1) & \mathbf{y}(2) & \cdots & \mathbf{y}(np-na) \\ \mathbf{y}(2) & \mathbf{y}(3) & \cdots & \mathbf{y}(n-na+1) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}(na) & \mathbf{y}(na+1) & \cdots & \mathbf{y}(np-1) \end{bmatrix} \tag{2.86}$$

where $\mathbf{H}_{na}$ is a $(na*nc) \times (np-na)$ matrix.

Define matrix $\mathbf{Y}_{na} = \begin{bmatrix} \mathbf{y}(na+1) & \cdots & \mathbf{y}(np) \end{bmatrix}$, and $\mathbf{A}_{ar} = \begin{bmatrix} \mathbf{A}_{na} & \mathbf{A}_{na-1} & \cdots & \mathbf{A}_1 \end{bmatrix}$. Then we can write Eq. (2.85) into a simpler form

$$\mathbf{Y}_{na} = \mathbf{A}_{ar}\mathbf{H}_{na} \tag{2.87}$$

This equation is called the least square equation for estimation of the matrix polynomial [63]. Note that the subscription represents autoregression.

Suppose the estimate of matrix $\mathbf{A}_{ar}$ is denoted as $\hat{\mathbf{A}}_{ar}$, then by solving Eq. (2.87) we have

$$\mathbf{Y}_{na}^{\mathrm{T}} = \mathbf{H}_{na}^{\mathrm{T}}\mathbf{A}_{ar}^{\mathrm{T}} \quad \Rightarrow \quad \hat{\mathbf{A}}_{ar} = \left(\mathbf{H}_{na}^{\mathrm{T}\dagger}\mathbf{Y}_{na}^{\mathrm{T}}\right)^{\mathrm{T}} = \mathbf{Y}_{na}\mathbf{H}_{na}^{\dagger} \tag{2.88}$$

where $\mathbf{H}_{na}^{\dagger}$ represents the pseudo inverse of $\mathbf{H}_{na}$. Note that this procedure will only work satisfactory if the problem is well overdetermined, thus we must require that [63]

$$np - na \gg na * nc \tag{2.89}$$

It should be pointed out that the solution given by Eq. (2.88) for the homogenous problem imposes no restriction on the roots of the linear difference equation represented by Eq. (2.83), which means Eq. (2.83) can be applied to the nonstationary process.

2.4.1.2 Polyreference method

Now that we know how to use the (block) Hankel matrix to estimate the AR coefficients of a homogeneous equation. In this section, we can use the same technique to find the AR coefficients of a inhomogeneous equation, which is a valid mathematical description of the system in OMA. And this mathematical model forms the basis of the Polyreference method. Therefore, to estimate the AR coefficients of an inhomogeneous equation, we must assume the random input is at least uncorrelated with the output responses.

Consider the following representative equation

$$\mathbf{y}(n) = \mathbf{A}_1\mathbf{y}(n-1) + \mathbf{A}_2\mathbf{y}(n-2) + \cdots + \mathbf{A}_{na}\mathbf{y}(n-na) + \mathbf{w}(n) \tag{2.90}$$

where $\mathbf{w}(n)$ is the white noise process uncorrelated with $\mathbf{y}(n)$.

To obtain the AR coefficient matrices for a stationary process, one way is to use the Yule-Walker's equations we mentioned in section 2.3.2.. However, direct application of Eq. (2.62) is not recommended here, as it will generate a non-causal process. This happens due to the symmetry of the covariances. For a causal system, namely, the current value of the process only depends on the past not the future, the system could be unidentifiable by directly using Eq. (2.62). To avoid this, we can reformulate Eq. (2.90) as follows,

$$\mathbf{y}(n+na+k) = \mathbf{A}_1\mathbf{y}(n+na+k-1) + \cdots + \mathbf{A}_{na}\mathbf{y}(n+k) + \mathbf{w}(n+na+k) \tag{2.91}$$

given $\mathbf{y}(n)$, $n = 1, 2, ..., np$ with np number of samples, the time differences $k = 1, 2, ...,$ $np - na + 1$.

Now the same as we derive the Yule-Walker equations, we first consider the covariance between $\mathbf{y}(n+na+k)$ and $\mathbf{y}(n)$, where $\mathbf{y}(n)$ is a stationary process with zero mean.

$$
\begin{aligned}
&\mathbb{E}\left[\mathbf{y}(n+na+k)\mathbf{y}^{\mathrm{T}}(n)\right] \\
&= \mathbb{E}\left[\mathbf{A}_1\mathbf{y}(n+na+k-1)\mathbf{y}^{\mathrm{T}}(n)\right] + \cdots + \mathbb{E}\left[\mathbf{A}_{na}\mathbf{y}(n+k)\mathbf{y}^{\mathrm{T}}(n)\right] + \mathrm{E}\left[\mathbf{w}(n+na+k)\mathbf{y}^{\mathrm{T}}(n)\right]
\end{aligned}
\tag{2.92}
$$

Since $\mathbb{E}\left[\mathbf{w}(n+na+k)\mathbf{y}^{\mathrm{T}}(n)\right] = 0$, Eq. (2.92) can be simplified as

$$\gamma(na+k) = \mathbf{A}_1\gamma(na+k-1) + \mathbf{A}_1\gamma(na+k-2) + \cdots + \mathbf{A}_{na}\gamma(k) \tag{2.93}$$

where $\gamma(na+k) = \mathbb{E}\left[\mathbf{y}(n+na+k)\mathbf{y}^{\mathrm{T}}(n)\right]$ is the covariance between the two processes. Note that a similar equation was presented by Lingmi Zhang in reference [86], but the formula was transformed from an state-space model.

For $k = 1, 2, ..., np - na + 1$, we can write $np - na + 1$ equations, Eq. (2.93) will give the coefficients through the following equation,

$$
\begin{aligned}
&\left[\gamma(na+1) \quad \gamma(na+2) \quad \cdots \quad \gamma(np)\right] \\
&= \left[\mathbf{A}_{na} \quad \mathbf{A}_{na-1} \quad \cdots \quad \mathbf{A}_1\right]
\begin{bmatrix}
\gamma(1) & \gamma(2) & \cdots & \gamma(np-na) \\
\gamma(2) & \gamma(3) & \cdots & \gamma(np-na+1) \\
\vdots & \vdots & \ddots & \vdots \\
\gamma(na) & \gamma(na+1) & \cdots & \gamma(np-1)
\end{bmatrix}
\end{aligned}
\tag{2.94}
$$

Define matrices $\boldsymbol{\gamma} = \left[\gamma(na+1) \quad \gamma(na+2) \quad \cdots \quad \gamma(np)\right]$, $\mathbf{A}_{\mathrm{ar}} = \left[\mathbf{A}_{na} \quad \mathbf{A}_{na-1} \quad \cdots \quad \mathbf{A}_1\right]$, and

$$
\mathbf{H}_{na} =
\begin{bmatrix}
\gamma(1) & \gamma(2) & \cdots & \gamma(np-na) \\
\gamma(2) & \gamma(3) & \cdots & \gamma(np-na+1) \\
\vdots & \vdots & \ddots & \vdots \\
\gamma(na) & \gamma(na+1) & \cdots & \gamma(np-1)
\end{bmatrix}
\tag{2.95}
$$

where $\mathbf{H}_{na}$ is the block Hankel matrix with na block rows. When we have nc number of channels, then $\gamma(k)$ will be a $nc \times nc$ square matrix. In this case, $\mathbf{H}_{na}$ becomes a $(na*nc) \times (np-na)*nc$ matrix.

Let us write Eq. (2.94) into a simpler form, i.e,,

$$\boldsymbol{\gamma}_{na} = \mathbf{A}_{ar}\mathbf{H}_{na} \tag{2.96}$$

This equation has the same format of Eq. (2.87). Hence, we shall apply the same computing procedure to it as well. In the meantime, the similar requirement is applied to the this formula too, see Eq. (2.89).

With Eq. (2.83) and Eq. (2.93), it is worth mentioning that, although the Polyreference method developed in OMA is different from the PRCE method developed by Vold et al. [107] at first glance, their core is the same. The PRCE method is implemented to deal with the exponential combinations of the impulse responses, which are the homogeneous solutions of the system, whereas the Polyreference method is working on the correlation function. However, we know that the solution of a homogeneous equation is fully determined by the mathematical structure of the system and the corresponding coefficients. Therefore, even if Eq. (2.83) looks different from Eq. (2.93), they have the same pole locations, and the correlation function of the output responses of a system behaves in the same way as the impulse response function of that system. In other words, PRCE method is not fundamentally different from the Polyreference developed in OMA. Based on this inference, we can write its discrete time correlation function into the following format

$$\gamma(n) = \mathbf{b}e^{\lambda n \Delta t} = \mathbf{C}\boldsymbol{\varphi}e^{\lambda n \Delta t} \tag{2.97}$$

where $\mathbf{b}$ is the mode shape vector, $\boldsymbol{\varphi}$ is the state-space mode shape. While $\mathbf{C}$ is the observation matrix, which can help us to pick the mode shape $\mathbf{b}$ from the state-space mode shape $\boldsymbol{\varphi}$ as $\mathbf{b} = \mathbf{C}\boldsymbol{\varphi}$. The pole the system is given by λ. $n\Delta t$ gives the discrete time.

Suppose the AR coefficient matrix $\mathbf{A}_{ar} = \begin{bmatrix} \mathbf{A}_{na} & \mathbf{A}_{na-1} & \cdots & \mathbf{A}_{1} \end{bmatrix}$ is available, we can rewrite Eq. (2.93) into a state-space equation, namely

$$\begin{bmatrix} \gamma(k+2) \\ \gamma(k+3) \\ \gamma(k+4) \\ \vdots \\ \gamma(na+k+1) \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{I} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{I} & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & \mathbf{I} \\ \mathbf{A}_{na} & \mathbf{A}_{na-1} & \mathbf{A}_{na-2} & \cdots & \mathbf{A}_{1} \end{bmatrix} \begin{bmatrix} \gamma(k+1) \\ \gamma(k+2) \\ \gamma(k+3) \\ \vdots \\ \gamma(na+k) \end{bmatrix} \tag{2.98}$$

Define vector $\mathbf{x}(k) = \begin{bmatrix} \gamma(k+1), \ldots, \gamma(na+k) \end{bmatrix}^{\mathrm{T}}$, Eq. (2.98) can be expressed compactly as

$$\mathbf{x}(k+1) = \mathbf{A}_C \mathbf{x}(k) \qquad (2.99)$$

Note that

$$\mathbf{x}(k) = \begin{bmatrix} \gamma(k+1) \\ \gamma(k+2) \\ \vdots \\ \gamma(na+k) \end{bmatrix} = \begin{bmatrix} \mathbf{C}\boldsymbol{\varphi} \\ \mathbf{C}\boldsymbol{\varphi}e^{\lambda\Delta t} \\ \vdots \\ \mathbf{C}\boldsymbol{\varphi}e^{\lambda(na-1)\Delta t} \end{bmatrix} e^{\lambda(k+1)\Delta t} = \boldsymbol{\phi}e^{\lambda(k+1)\Delta t} \qquad (2.100)$$

where $\boldsymbol{\phi}$ is the discrete eigenvector. If we substitute this expression into Eq. (2.100) we get

$$\mathbf{A}_C \boldsymbol{\phi} = e^{\lambda\Delta t} \boldsymbol{\phi} \qquad (2.101)$$

To find the all the poles of the system, we shall apply the EVD to the companion matrix, then,

$$\mathbf{A}_C = \begin{bmatrix} \boldsymbol{\phi}_i \end{bmatrix} \begin{bmatrix} e^{\lambda_i \Delta t} \end{bmatrix} \begin{bmatrix} \boldsymbol{\phi}_i \end{bmatrix}^{-1} \qquad (2.102)$$

with which we can identify the mode shape $\mathbf{b}_i$ from the first past of the eigenvectors, and the corresponding modal parameters can be easily obtained by evaluating $e^{\lambda_i \Delta t}$, namely

$$\lambda_i = \ln\left(e^{\lambda_i \Delta t}\right)\big/\Delta t \qquad (2.103)$$

while λ_i are normally complex values and its complex conjugates are given by

$$\lambda_i, \lambda_i^* = -\omega_i \xi_i \pm i\omega_i \sqrt{1-\xi_i^2} \qquad (2.104)$$

where ω_i and ξ_i are the ith natural frequency and damping ratio, respectively. And they can be easily obtained from the following two equations, namely

$$\omega_i = \sqrt{\lambda_i \lambda_i^*} \quad \text{and} \quad \xi_i = -\frac{\lambda_i + \lambda_i^*}{2\omega_i} \qquad (2.105)$$

Note that when dealing with a system governed by the 2nd order differential equation, na is normally set at 2. Up to now, we have completed our discussion of the Polyreference method used in OMA.

2.4.2 Estimation of the state matrix for a deterministic system

Based on our discussion of the system equations in section 2.2 and 2.3, we know that all the system coefficients can be grouped into a companion matrix. In the state-space model, this companion matrix is given by the state matrix, which implies that finding the system coefficients is equivalent to find the state matrix. Thus, to achieve our goal, we will again try to form a (block) Hankel matrix but based on the Markov parameters.

From Eq. (2.73), we know the solution of $\mathbf{y}(k)$ can be given by the summation of the input multiply the corresponding Markov parameter, i.e.,

$$\mathbf{y}(k) = \sum_{i=1}^{k} \mathbf{CA}^{i-1}\mathbf{Bu}(k-i) = \sum_{i=1}^{k} \mathbf{h}(i)\mathbf{u}(k-i) \tag{2.106}$$

This equation can also be expressed as

$$\mathbf{y}(k) = \sum_{j=0}^{k-1} \mathbf{h}(k-j)\mathbf{u}(j) = \sum_{j=0}^{k-1} \mathbf{CA}^{k-j-1}\mathbf{Bu}(j) \tag{2.107}$$

where $j = k - i$. After l time steps

$$\mathbf{y}(k+l) = \sum_{j=0}^{k+l-1} \mathbf{h}(k+l-j)\mathbf{u}(j) \tag{2.108}$$

In general, the output responses at future times $l = 0, 1, 2, \ldots$, due to the sequence of M past impulses can be computed as follows

$$\mathbf{y}(k+l) = \sum_{j=0}^{-(M-1)} \mathbf{h}(k+l-j)\mathbf{u}(j) \tag{2.109}$$

Note that here we just changed the upper limit of Eq. (2.108) to control the effect of the length of the inputs at each time step of the output.

$$\begin{bmatrix} \mathbf{y}(k) \\ \mathbf{y}(k+1) \\ \vdots \\ \mathbf{y}(k+l-1) \end{bmatrix} = \begin{bmatrix} \mathbf{h}(k) & \mathbf{h}(k+1) & \cdots & \mathbf{h}(k+M-1) \\ \mathbf{h}(k+1) & \mathbf{h}(k+2) & \cdots & \mathbf{h}(k+M) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{h}(k+l-1) & \mathbf{h}(k+l) & \cdots & \mathbf{h}(k+l+M-2) \end{bmatrix} \begin{bmatrix} \mathbf{u}(0) \\ \mathbf{u}(-1) \\ \vdots \\ \mathbf{u}(1-M) \end{bmatrix} \tag{2.110}$$

In short, the above equation is given by

$$\mathbf{y}_l(k) = \mathbf{H}_{lM}(k)\mathbf{u}_M \tag{2.111}$$

where the block-symmetric matrix $\mathbf{H}_{lM}(k)$ is given by

$$\mathbf{H}_{lM}(k) = \begin{bmatrix} \mathbf{CA}^{k-1}\mathbf{B} & \mathbf{CA}^{k}\mathbf{B} & \cdots & \mathbf{CA}^{k+M-2}\mathbf{B} \\ \mathbf{CA}^{k}\mathbf{B} & \mathbf{CA}^{k+1}\mathbf{B} & \cdots & \mathbf{CA}^{k+M-1}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{CA}^{k+l-2}\mathbf{B} & \mathbf{CA}^{k+l-1}\mathbf{B} & \cdots & \mathbf{CA}^{k+l+M-3}\mathbf{B} \end{bmatrix} \tag{2.112}$$

which is our (block) Hankel matrix of Markov parameters. It is a matrix representation of the input-output relationship for the system in the discrete time domain.

Note that $\mathbf{H}_{lM}(k)$ can be further written as three matrices multiplication, which is

$$\mathbf{H}_{lM}(k) = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^{l-1} \end{bmatrix} \mathbf{A}^{k-1} \begin{bmatrix} \mathbf{B} & \mathbf{AB} & \cdots & \mathbf{A}^{M-1}\mathbf{B} \end{bmatrix} \qquad (2.113)$$

For a sufficiently large l, we define

$$\mathcal{O}_l = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^{l-1} \end{bmatrix} \qquad (2.114)$$

which is a $lq \times n$ observability matrix, and

$$\mathcal{C}_M = \begin{bmatrix} \mathbf{B} & \mathbf{AB} & \cdots & \mathbf{A}^{M-1}\mathbf{B} \end{bmatrix} \qquad (2.115)$$

which a $n \times Mp$ controllability matrix.

In short, $\mathbf{H}_{lM}(k)$ is given by

$$\mathbf{H}_{lM}(k) = \mathcal{O}_l \mathbf{A}^{k-1} \mathcal{C}_M \qquad (2.116)$$

which is a $lq \times Mp$ matrix.

When $k = 1$, we have

$$\mathbf{H}_{lM}(1) = \mathcal{O}_l \mathcal{C}_M \qquad (2.117)$$

While $k = 2$, we get

$$\mathbf{H}_{lM}(2) = \mathcal{O}_l \mathbf{A} \mathcal{C}_M \qquad (2.118)$$

Generally, the system realisation in terms of the state-space models begins by forming the generalised Hankel matrix. Because if we can find both $\mathbf{H}_{lM}(1)$ and $\mathbf{H}_{lM}(2)$, we can always find the state matrix $\mathbf{A}$ which contains the system information.

For instance, if we use the Singular Value Decomposition (SVD) to decompose $\mathbf{H}_{lM}(1)$ we get

$$\mathbf{H}_{lM}(1) = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_1 & 0 \\ 0 & \mathbf{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^{\mathrm{T}} \\ \mathbf{V}_2^{\mathrm{T}} \end{bmatrix} = \mathcal{O}_l \mathcal{C}_M \qquad (2.119)$$

where $\mathbf{U}_1 \in \mathbb{R}^{lq \times n}$, $\mathbf{V}_1 \in \mathbb{R}^{Mp \times n}$, and $\boldsymbol{\Sigma}_1 \in \mathbb{R}^{n \times n}$ are the nonzero singular vector matrices and singular value matrix, respectively. The number of nonzero singular values indicates the rank of $\mathbf{H}_{lM}(1)$ which is the order of the system n. Note that n might have different meaning throughout the text, but it will be pointed out appropriately. However, due to the computational and measurement noise, $\boldsymbol{\Sigma}_2$ is usually not identical to zero. Hence, $\boldsymbol{\Sigma}_1$ is computed by setting a threshold to distinguish and discard the insignificant singular values.

So from Eq. (2.119), according to the non-zero singular values, we have

$$\mathcal{O}_l \mathcal{C}_M = \mathbf{U}_1 \boldsymbol{\Sigma}_1 \mathbf{V}_1^{\mathrm{T}} = \left(\mathbf{U}_1 \boldsymbol{\Sigma}_1^{\frac{1}{2}} \right) \left(\boldsymbol{\Sigma}_1^{\frac{1}{2}} \mathbf{V}_1^{\mathrm{T}} \right) \tag{2.120}$$

If there exist a non-singular matrix $\mathbf{T} \in \mathbb{R}^{n \times n}$ corresponds to a state-space change which can always be set equal to the identity matrix, then

$$\mathcal{O}_l \mathbf{T} = \mathbf{U}_1 \boldsymbol{\Sigma}_1^{\frac{1}{2}} \quad \text{and} \quad \mathbf{T}^{-1} \mathcal{C}_M = \boldsymbol{\Sigma}_1^{\frac{1}{2}} \mathbf{V}_1^{\mathrm{T}} \tag{2.121}$$

When $\mathbf{T} = \mathbf{I}$, the state matrix $\mathbf{A}$ can be determined from

$$\mathbf{A} = \mathcal{O}_l^{-1} \mathbf{H}_{lM}(2) \mathcal{C}_M^{-1} = \boldsymbol{\Sigma}_1^{-\frac{1}{2}} \mathbf{U}_1^{-1} \mathbf{H}_{lM}(2) \mathbf{V}_1^{-\mathrm{T}} \boldsymbol{\Sigma}_1^{-\frac{1}{2}} \tag{2.122}$$

By applying the eigenvalue decomposition to the state matrix, we can obtain the model parameters of the system, that is

$$\mathbf{A} = \boldsymbol{\Psi} \boldsymbol{\Lambda} \boldsymbol{\Psi}^{-1} \tag{2.123}$$

where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_i) \in \mathbb{C}^{n \times n}$, $i = 1, \dots, n$, is a diagonal matrix containing the complex eigenvalues, while $\boldsymbol{\Psi} \in \mathbb{C}^{n \times n}$ contains the eigenvectors as columns. Here, λ_i and its complex conjugates are given by Eq. (2.104).

As for the observation matrix $\mathbf{C}$ and the input matrix $\mathbf{B}$, they can be easily obtained from (2.121), the first block entry in $\mathcal{O}_l$ is $\mathbf{C}$ and the first block entry of $\mathcal{C}_M$ is $\mathbf{B}$.

Note that the state vector does not necessarily have a physical meaning. Thus, the eigenvectors of the state matrix $\boldsymbol{\Psi}$ needs to be transferred to the outside world [85]. The observed parts of the system eigenvectors in $\boldsymbol{\Psi}$ give the mode shapes, which are defined as the columns in $\boldsymbol{\Phi} \in \mathbb{C}^{l \times n}$. Therefore, we can obtain the $\boldsymbol{\Phi}$ via the output matrix,

$$\boldsymbol{\Phi} = \mathbf{C} \boldsymbol{\Psi} \tag{2.124}$$

Before we leave this subsection, we want to talk a bit more about the observability and the controllability matrix, as both have a meaning in the state-space model, i.e., the observability matrix is associated with the system input while the controllability matrix is related to the output.

Now, let us consider Eq. (2.70) again, and repeat the substitution in the following fashion,

$$\begin{cases} \mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k) \\ \quad\quad \mathbf{y}(k) = \mathbf{C}\mathbf{x}(k) \end{cases}$$
$$\begin{cases} \mathbf{x}(k+2) = \mathbf{A}\mathbf{x}(k+1) + \mathbf{B}\mathbf{u}(k+1) = \mathbf{A}^2\mathbf{x}(k) + \mathbf{A}\mathbf{B}\mathbf{u}(k) + \mathbf{B}\mathbf{u}(k+1) \\ \quad\quad \mathbf{y}(k+1) = \mathbf{C}\mathbf{x}(k+1) = \mathbf{C}\mathbf{A}\mathbf{x}(k) + \mathbf{C}\mathbf{B}\mathbf{u}(k) \end{cases} \tag{2.125}$$

After $l \geq 0$ steps, we have

$$\begin{cases} \mathbf{x}(k+l) = \mathbf{A}^l\mathbf{x}(k) + \sum_{j=0}^{l-1} \mathbf{A}^{l-j-1}\mathbf{B}\mathbf{u}(k+j) \\ \mathbf{y}(k+l-1) = \mathbf{C}\mathbf{A}^{l-1}\mathbf{x}(k) + \sum_{j=0}^{l-2} \mathbf{C}\mathbf{A}^{l-j-2}\mathbf{B}\mathbf{u}(k+j) \end{cases} \tag{2.126}$$

Note that this iteration result is different from Eq. (2.73) we presented before.

Now let us make a step further to rewrite the above equation into the matrix form as follows

$$\begin{cases} \mathbf{x}(k+l) = \mathbf{A}^l\mathbf{x}(k) + \mathcal{C}_l\mathbf{u}_l(k) \\ \mathbf{y}_l(k) = \mathcal{O}_l\mathbf{x}(k) + \mathcal{T}_l\mathbf{u}_l(k) \end{cases} \tag{2.127}$$

where $\mathbf{u}_l(k)$ and $\mathbf{y}_l(k)$ are defined as column vectors of input and output data going l steps into the future starting with $\mathbf{u}(k)$ and $\mathbf{y}(k)$, respectively,

$$\mathbf{u}_l(k) = \begin{bmatrix} \mathbf{u}(k) \\ \mathbf{u}(k+1) \\ \vdots \\ \mathbf{u}(k+l-1) \end{bmatrix}, \quad \mathbf{y}_l(k) = \begin{bmatrix} \mathbf{y}(k) \\ \mathbf{y}(k+1) \\ \vdots \\ \mathbf{y}(k+l-1) \end{bmatrix} \tag{2.128}$$

Here, $\mathcal{C}_l$ in Eq. (2.127)

$$\mathcal{C}_l = \begin{bmatrix} \mathbf{A}_d^{l-1}\mathbf{B}_d & \cdots & \mathbf{A}_d\mathbf{B}_d & \mathbf{B}_d \end{bmatrix} \tag{2.129}$$

is an $n \times lp$ controllability matrix, but it is the flipped version of Eq. (2.115). While $\mathcal{O}_l$ is a $lq \times n$ observability matrix, which is given exactly by Eq. (2.114). If we put them together, we will have

$$\mathbf{T}_l = \mathcal{O}_l \mathcal{C}_l = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^{l-1} \end{bmatrix} \begin{bmatrix} \mathbf{A}^{l-1}\mathbf{B} & \cdots & \mathbf{AB} & \mathbf{B} \end{bmatrix} = \begin{bmatrix} \mathbf{CA}^{l-1}\mathbf{B} & \mathbf{CA}^{l-2}\mathbf{B} & \cdots & \mathbf{CB} \\ \mathbf{CA}^{l}\mathbf{B} & \mathbf{CA}^{l-1}\mathbf{B} & \cdots & \mathbf{CAB} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{CA}^{2l-1}\mathbf{B}_d & \mathbf{CA}^{2l-2}\mathbf{B} & \cdots & \mathbf{CA}^{l-1}\mathbf{B} \end{bmatrix} \tag{2.130}$$

which becomes a $lq \times lp$ block Toeplitz matrix.

Similar to the block Hankel matrix, we can have

$$\mathbf{T}_l(k) = \mathcal{O}_l \mathcal{C}_l = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^{l-1} \end{bmatrix} \mathbf{A}^{k-1} \begin{bmatrix} \mathbf{A}^{l-1}\mathbf{B} & \cdots & \mathbf{AB} & \mathbf{B} \end{bmatrix} \tag{2.131}$$

where Eq. (2.130) is $\mathbf{T}_l(1)$. From $\mathbf{T}_l(1)$ and $\mathbf{T}_l(2)$ we can identify the state matrix $\mathbf{A}$.

As for $\mathcal{T}_l$, it is also a $lq \times lp$ Toeplitz matrix of the system Markov parameters, which is given by

$$\mathcal{T}_l = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ \mathbf{CB} & 0 & 0 & \cdots & 0 \\ \mathbf{CAB} & \mathbf{CB} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \mathbf{CA}^{l-2}\mathbf{B} & \mathbf{CA}^{l-3}\mathbf{B} & \cdots & \mathbf{CB} & 0 \end{bmatrix} \tag{2.132}$$

2.4.3 Estimation of the state matrix for a stochastic system

In this section, we want to derive the (block) Hankel matrix and the (block) Toeplitz matrix in terms of a stochastic system which is characterised by the state-space model shown as Eq. (2.76). Since we are dealing with a stochastic system, we shall take a statistical approach to deal with the problem. The mean and the covariance (variance) of the stochastic process are considered.

It is worth mentioning that the (block) Hankel matrix plays an important role in the Subspace State-space System Identification (4SID) methods [108], as the measurements that appear in the algorithms are normally organised in the form of (block) Hankel matrices. Unfortunately, as we mentioned earlier, we are unable to give an exhaustive review of all methods and their variation. Therefore, in the section, we will use the Reference-Base Covariance-Driven Stochastic Realisation (SSI-COV) method and the Data-Driven Stochastic Subspace Identification (SSI-DATA) method [58] as examples to demonstrate the main ideas of the 4SID methods.

Suppose we have a $r \times 1$ vector that represents the reference output $\mathbf{y}^{\text{ref}}(k)$ at time instant k, and another $l \times 1$ vector represents our newly measured outputs $\mathbf{y}(k)$, and $l > r$. The same as our discussion in section 2.4.1, we can always gather the output measurements in a block Hankel matrix. Here, we shall do the same. Thus, similar to Eq. (2.86), we can put the reference outputs into the following block Hankel matrix,

$$\mathbf{Y}_p^{\text{ref}} = \mathbf{Y}_{0|i-1}^{\text{ref}} = \frac{1}{\sqrt{j}} \begin{bmatrix} \mathbf{y}^{\text{ref}}(0) & \mathbf{y}^{\text{ref}}(1) & \cdots & \mathbf{y}^{\text{ref}}(j-1) \\ \mathbf{y}^{\text{ref}}(1) & \mathbf{y}^{\text{ref}}(2) & \cdots & \mathbf{y}^{\text{ref}}(j) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}^{\text{ref}}(i-1) & \mathbf{y}^{\text{ref}}(i) & \cdots & \mathbf{y}^{\text{ref}}(i+j-2) \end{bmatrix} \tag{2.133}$$

where p denotes "past", while the subscript of $\mathbf{Y}_{0|i-1}^{\text{ref}}$ represents the first and last element in the first column of the block Hankel matrix. It is obvious that $\mathbf{Y}_p^{\text{ref}}$ is a $ir \times j$ matrix.

Similarly we can put the output measurements into $\mathbf{Y}_f$, where f denotes "future",

$$\mathbf{Y}_f = \mathbf{Y}_{i|2i-1} = \frac{1}{\sqrt{j}} \begin{bmatrix} \mathbf{y}(i) & \mathbf{y}(i+1) & \cdots & \mathbf{y}(i+j-1) \\ \mathbf{y}(i+1) & \mathbf{y}(i+2) & \cdots & \mathbf{y}(i+j) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}(2i-1) & \mathbf{y}(2i) & \cdots & \mathbf{y}(2i+j-2) \end{bmatrix} \tag{2.134}$$

where $\mathbf{Y}_f$ is a $il \times j$ matrix.

Note that the matrix is scaled by $1/\sqrt{j}$, and for either SSI-COV or SSI-DATA method, ideally, the number of the columns j should be as large as possible, i.e., $2i < j \to \infty$. This is because when we compute the sample auto-covariance of $\mathbf{Y}_p^{\text{ref}}$ or sample cross-covariance matrix between $\mathbf{Y}_f$ and $\mathbf{Y}_p^{\text{ref}}$, to ensure the covariance matrix is positive definite, we need both $\mathbf{Y}_f$ and $\mathbf{Y}_p^{\text{ref}}$ to be full row rank matrices.

Having defined these the past and future output block Hankel matrices, now we can combine them together to form a new output block Hankel matrix, that is

$$\mathbf{H}_{2i} = \begin{pmatrix} \mathbf{Y}_p^{\text{ref}} \\ \mathbf{Y}_f \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_{0|i-1}^{\text{ref}} \\ \mathbf{Y}_{i|2i-1} \end{pmatrix} \begin{matrix} \updownarrow ri \ "past" \\ \updownarrow li \ "future" \end{matrix} \in \mathbb{R}^{(r+l)i \times j} \tag{2.135}$$

2.4.3.1 Covariance-Driven Stochastic Realisation (SSI-COV) method

Now, let us consider the following statistical properties of the system by using the relations given by Eq. (2.77) to Eq. (2.80). Define the state covariance matrix $\mathbf{\Omega}$,

$$\mathbf{\Omega}(k+1) = \mathbb{E}\left[\mathbf{x}(k+1)\mathbf{x}^{\mathrm{T}}(k+1)\right] = \mathbf{A}\mathbf{\Omega}(k)\mathbf{A}^{\mathrm{T}} + \mathcal{G}\mathbf{Q}\mathcal{G}^{\mathrm{T}} \tag{2.136}$$

When $\mathbf{\Omega}(k)$ converges to a steady state value then $\mathbf{\Omega}(k+1) = \mathbf{\Omega}(k)$ for large k, then we have the following equation which are independent of time, i.e.,

$$\mathbf{\Omega} = \mathbf{A}\mathbf{\Omega}\mathbf{A}^{\mathrm{T}} + \mathcal{G}\mathbf{Q}\mathcal{G}^{\mathrm{T}} \tag{2.137}$$

is the so-called Lyapunov equation.

When the system is in steady state, we can obtain the corresponding time-independent state-output covariance matrix $\mathbf{G}$ as follows

$$\mathbf{G} = \mathbb{E}\left[\mathbf{x}(k+1)\mathbf{y}^{\mathrm{T}}(k)\right] = \mathbf{A}\mathbf{\Omega}\mathbf{C}^{\mathrm{T}} + \mathcal{G}\mathbf{S}\mathcal{H}^{\mathrm{T}} \tag{2.138}$$

also, the output covariance matrices $\mathbf{\Lambda}_i$,

$$\mathbf{\Lambda}_i = \mathbb{E}\left[\mathbf{y}(k+i)\mathbf{y}^{\mathrm{T}}(k)\right] = \mathbf{A}\mathbb{E}\left[\mathbf{x}(k+i)\mathbf{x}^{\mathrm{T}}(k)\right]\mathbf{C}^{\mathrm{T}} + \mathcal{H}\mathbb{E}\left[\mathbf{v}(k+i)\mathbf{v}^{\mathrm{T}}(k)\right]\mathcal{H} \tag{2.139}$$

when $i=0$,

$$\mathbf{\Lambda}_0 = \mathbf{A}\mathbf{\Omega}\mathbf{C}^{\mathrm{T}} + \mathcal{H}\mathbf{R}\mathcal{H}^{\mathrm{T}} \tag{2.140}$$

when $i \neq 0$, instead of using result from Eq. (2.139) by solving the following equation

$$\mathbf{\Lambda}_i = \mathbf{A}\mathbb{E}\left[\mathbf{x}(k+i)\mathbf{x}^{\mathrm{T}}(k)\right]\mathbf{C}^{\mathrm{T}} \tag{2.141}$$

We want to use the result given by Eq. (2.138) to present $\mathbf{\Lambda}_i$.

Since

$$\begin{aligned}
\mathbf{x}(k+1) &= \mathbf{A}\mathbf{x}(k) + \mathcal{G}\mathbf{w}(k) \\
\mathbf{x}(k+2) &= \mathbf{A}\mathbf{x}(k+1) + \mathcal{G}\mathbf{w}(k+1) \\
&\cdots \\
\mathbf{x}(k+i) &= \mathbf{A}\mathbf{x}(k+i) + \mathcal{G}\mathbf{w}(k+i) = \mathbf{A}^{i-1}\mathbf{x}(k+1) + \mathbf{A}^{i-2}\mathcal{G}\mathbf{w}(k+1) + \cdots + \mathcal{G}\mathbf{w}(k+i)
\end{aligned} \tag{2.142}$$

In the meantime,

$$\mathbf{y}(k) = \mathbf{C}\mathbf{x}(k) + \mathcal{H}\mathbf{v}(k)$$
$$\mathbf{y}(k+1) = \mathbf{C}\mathbf{x}(k+1) + \mathcal{H}\mathbf{v}(k+1)$$
$$\dots \tag{2.143}$$
$$\mathbf{y}(k+i) = \mathbf{C}\mathbf{x}(k+i) + \mathcal{H}\mathbf{v}(k+i)$$
$$= \mathbf{C}\mathbf{A}^{i-1}\mathbf{x}(k+1) + \mathbf{C}\mathbf{A}^{i-2}\mathcal{G}\mathbf{w}(k+1) + \dots + \mathbf{C}\mathcal{G}\mathbf{w}(k+i) + \mathcal{H}\mathbf{v}(k+i)$$

We can substitute $\mathbf{y}(k+i)$ given by the above equation into Eq. (2.139), then we obtain

$$\mathbf{\Lambda}_i = \mathbb{E}\left[\mathbf{y}(k+i)\mathbf{y}^{\mathrm{T}}(k)\right] = \mathbf{C}\mathbf{A}^{i-1}\mathbf{G} \tag{2.144}$$

Now we put the previously derived equations together, we have

$$\mathbf{\Omega} = \mathbf{A}\mathbf{\Omega}\mathbf{A}^{\mathrm{T}} + \mathcal{G}\mathbf{Q}\mathcal{G}^{\mathrm{T}}$$
$$\mathbf{G} = \mathbf{A}\mathbf{\Omega}\mathbf{C}^{\mathrm{T}} + \mathcal{G}\mathbf{S}\mathcal{H}^{\mathrm{T}} \tag{2.145}$$
$$\mathbf{\Lambda}_i = \begin{cases} \mathbf{A}\mathbf{\Omega}\mathbf{C}^{\mathrm{T}} + \mathcal{H}\mathbf{R}\mathcal{H}^{\mathrm{T}} & i = 0 \\ \mathbf{C}\mathbf{A}^{i-1}\mathbf{G} & i \neq 0 \end{cases}$$

Among all four equations in Eq. (2.145), the last one is our key function, as it gives the Markov parameters for the stochastic system. Since $\mathbf{\Lambda}_i$ is obtained based on the correlation between the outputs of the system, which has the same format of the impulse response of the system given by Eq. (2.75), this implies that there is a relationship between the correlation function of the system output responses and the impulse response of the system.

Now let us define the covariance matrices between all outputs and the references as follows

$$\mathbf{\Lambda}_i^{\mathrm{ref}} = \mathbb{E}\left[\mathbf{y}(k+i)\left(\mathbf{y}^{\mathrm{ref}}(k)\right)^{\mathrm{T}}\right] \tag{2.146}$$

which is a $l \times r$ matrix.

For the state-reference output covariance matrix $\mathbf{G}_i^{\mathrm{ref}}$,

$$\mathbf{G}_i^{\mathrm{ref}} = \mathbb{E}\left[\mathbf{x}_s(k+1)\left(\mathbf{y}^{\mathrm{ref}}(k)\right)^{\mathrm{T}}\right] \tag{2.147}$$

which is a $n \times r$ matrix. Further we know that

$$\mathbf{\Lambda}_i^{\mathrm{ref}} = \mathbf{C}\mathbf{A}^{i-1}\mathbf{G}^{\mathrm{ref}} \tag{2.148}$$

Now if we evaluate the cross-covariance between the past and future observations, we can obtain

$$\mathbf{T}_{1|i}^{\mathrm{ref}} = \mathbb{E}\left[\mathbf{Y}_f\left(\mathbf{Y}_p^{\mathrm{ref}}\right)^{\mathrm{T}}\right] \tag{2.149}$$

which is the (block) Toeplitz matrix, and each entry is given by the corresponding Λ_i^{ref}.

If the measurements have zero mean, then $\mathbf{T}_{1|i}^{\text{ref}}$ a cross-covariance matrix between the past and future measurements. For instance, if we evaluate the first row in $\mathbf{Y}_f$ and the second row in $\left(\mathbf{Y}_p^{\text{ref}}\right)^{\text{T}}$, we have

$$
\begin{aligned}
&\mathbb{E}\left[\mathbf{Y}_f\left(1,:\right)\left(\mathbf{Y}_p^{\text{ref}}\left(2,:\right)\right)^{\text{T}}\right] \\
&= \frac{1}{j}\left(\mathbb{E}\left[\mathbf{y}\left(1+i-1\right)\left(\mathbf{y}^{\text{ref}}\left(1\right)\right)^{\text{T}}\right] + \mathbb{E}\left[\mathbf{y}\left(2+i-1\right)\left(\mathbf{y}^{\text{ref}}\left(2\right)\right)^{\text{T}}\right] + \cdots + \mathbb{E}\left[\mathbf{y}\left(j+i-1\right)\left(\mathbf{y}^{\text{ref}}\left(j\right)\right)^{\text{T}}\right]\right) \\
&= \frac{1}{j}\left(\Lambda_{i-1}^{\text{ref}}\left(1\right) + \Lambda_{i-1}^{\text{ref}}\left(2\right) + \cdots + \Lambda_{i-1}^{\text{ref}}\left(j\right)\right) = \Lambda_{i-1}^{\text{ref}}
\end{aligned}
$$

(2.150)

As a result, $\mathbf{T}_{1|i}^{\text{ref}}$ is given by

$$
\mathbf{T}_{1|i}^{\text{ref}} = \begin{bmatrix} \Lambda_i^{\text{ref}} & \Lambda_{i-1}^{\text{ref}} & \cdots & \Lambda_1^{\text{ref}} \\ \Lambda_{i+1}^{\text{ref}} & \Lambda_i^{\text{ref}} & \cdots & \Lambda_2^{\text{ref}} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_{2i-1}^{\text{ref}} & \Lambda_{2i-2}^{\text{ref}} & \cdots & \Lambda_i^{\text{ref}} \end{bmatrix}
$$

(2.151)

which is a $il \times ir$ matrix. From our analysis, we can see that $1/j$ is cancelled out during the averaging process.

In the meantime, we can see that $\mathbf{T}_{1|i}^{\text{ref}}$ can be decomposed as

$$
\mathbf{T}_{1|i}^{\text{ref}} = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^{i-1} \end{bmatrix} \begin{bmatrix} \mathbf{A}^{i-1}\mathbf{G}^{\text{ref}} & \cdots & \mathbf{AG}^{\text{ref}} & \mathbf{G}^{\text{ref}} \end{bmatrix} = \mathcal{O}_i \mathcal{C}_i^{\text{ref}}
$$

(2.152)

where $\mathcal{C}_i^{\text{ref}}$ is the reference reversed extended stochastic controllability matrix, which has size $n \times ir$.

Also, we can obtain a shifted block Toeplitz matrix

$$
\mathbf{T}_{2|i+1}^{\text{ref}} = \mathcal{O}_i \mathbf{A} \mathcal{C}_i^{\text{ref}}
$$

(2.153)

Together with $\mathbf{T}_{1|i}^{\text{ref}}$, we can use the previously mentioned procedure to find the state matrix $\mathbf{A}$, see from Eq. (2.119) to Eq. (2.122).

2.4.3.2 Data-Driven Stochastic Subspace Identification (SSI-DATA) method

Unlike the covariance-based method SSI-COV, the SSI-DATA is built upon the state variable estimation. In section 2.3.4, we have briefly mentioned the state estimation formula, while in this section, we shall see it in action.

From Eq. (2.143) we observe that

$$
\begin{aligned}
\hat{\mathbf{y}}\left(k \mid k\right) &= \mathbf{C}\mathbb{E}\left[\mathbf{x}(k) \mid \mathcal{Y}_k\right] \\
\hat{\mathbf{y}}\left(k+1 \mid k+1\right) &= \mathbf{C}\mathbb{E}\left[\mathbf{x}(k+1) \mid \mathcal{Y}_k\right] = \mathbf{C}\mathbf{A}\mathbb{E}\left[\mathbf{x}(k) \mid \mathcal{Y}_k\right] \\
&\dots \\
\hat{\mathbf{y}}\left(k+i-1 \mid k+i-1\right) &= \mathbf{C}\mathbb{E}\left[\mathbf{x}(k+i-1) \mid \mathcal{Y}_k\right] = \mathbf{C}\mathbf{A}^{i-1}\mathbb{E}\left[\mathbf{x}(k) \mid \mathcal{Y}_k\right]
\end{aligned}
\tag{2.154}
$$

where $\hat{\mathbf{x}}\left(k \mid k\right) = \mathbb{E}\left[\mathbf{x}(k) \mid \mathcal{Y}_k\right]$ is the estimated state variable estimated from the past observations $\mathcal{Y}_k = \left\{\mathbf{y}(0), \mathbf{y}(1), \dots, \mathbf{y}(k)\right\}$.

Now, if we arrange Eq. (2.154) into a block Hankel matrix, we will have

$$
\begin{aligned}
\mathcal{P}_{i \mid 2i-1} &= \begin{bmatrix} \mathbf{C} \\ \mathbf{C}\mathbf{A} \\ \vdots \\ \mathbf{C}\mathbf{A}^{i-1} \end{bmatrix} \begin{bmatrix} \mathbb{E}\left[\mathbf{x}(k) \mid \mathcal{Y}_k\right] & \mathbb{E}\left[\mathbf{x}(k+1) \mid \mathcal{Y}_{k+1}\right] & \cdots & \mathbb{E}\left[\mathbf{x}(k+j-1) \mid \mathcal{Y}_{k+j-1}\right] \end{bmatrix} \\
&= \mathcal{O}_i \hat{\mathbf{X}}_i
\end{aligned}
\tag{2.155}
$$

where

$$
\hat{\mathbf{X}}_i = \begin{bmatrix} \mathbb{E}\left[\mathbf{x}(k) \mid \mathcal{Y}_k\right] & \mathbb{E}\left[\mathbf{x}(k+1) \mid \mathcal{Y}_{k+1}\right] & \cdots & \mathbb{E}\left[\mathbf{x}(k+j-1) \mid \mathcal{Y}_{k+j-1}\right] \end{bmatrix}
\tag{2.156}
$$

is the so-called Kalman filter state sequence [62] or the system state sequence. $\mathcal{P}_{i \mid 2i-1} = \hat{\mathbf{Y}}_f \in \mathbb{R}^{il \times j}$ is the projection matrix, which contains all the estimation of the future observations in terms of the past observations.

As we know that to estimate the system state $\mathbf{x}(k)$ at time k, we can condition on the past k observations, i.e., $\mathbf{y}(0), \mathbf{y}(1), \dots, \mathbf{y}(k)$, then calculate the conditional expectation $\mathbb{E}\left[\mathbf{x}(k) \mid \mathcal{Y}_k\right]$ by using the formula given in Eq. (2.82) provided that the measurements are Gaussian. Following this inference, we can estimate the state sequence $\hat{\mathbf{X}}_i$ via the same formula.

Now suppose the measured responses are jointly Gaussian and have been reprocessed to have zero mean. Then the sample cross-covariance matrix between the future and past observations and the sample auto-covariance matrix of the past observations are given by

$$\boldsymbol{\Sigma}_{fp} = \mathbf{Y}_f \left(\mathbf{Y}_p^{\text{ref}}\right)^{\text{T}} \quad \text{and} \quad \boldsymbol{\Sigma}_{pp} = \mathbf{Y}_p^{\text{ref}} \left(\mathbf{Y}_p^{\text{ref}}\right)^{\text{T}} \tag{2.157}$$

respectively. With Eq. (2.82), we can easily write out the estimated future observations based on the past i observations, then

$$\mathcal{P}_{i|2i-1} = \hat{\mathbf{Y}}_f = \boldsymbol{\Sigma}_{fp} \boldsymbol{\Sigma}_{pp}^{-1} \mathbf{Y}_p^{\text{ref}} \tag{2.158}$$

Since the projection matrix $\mathcal{P}_{i|2i-1}$ is readily available, then apply the SVD to it, that is

$$\mathcal{P}_{i|2i-1} = \mathbf{U}_1 \mathbf{S}_1 \mathbf{V}_1^{\text{T}} \tag{2.159}$$

Here $\mathcal{P}_{i|2i-1}$ is a rank deficient matrix with rank n, which is determined by the rank of the state matrix $\mathbf{A}$. Thus, theoretically, we have $\mathbf{U}_1 \in \mathbb{R}^{il \times n}$, $\mathbf{S}_1 \in \mathbb{R}^{n \times n}$, and $\mathbf{V}_1 \in \mathbb{R}^{n \times j}$. Combining Eq. (2.155) and Eq. (2.159) gives

$$\mathcal{O}_i = \mathbf{U}_1 \mathbf{S}_1^{1/2} \quad \text{and} \quad \hat{\mathbf{X}}_i = \mathcal{O}_i^{\dagger} \mathcal{P}_{i|2i-1} \tag{2.160}$$

Now, it is worth mentioning that the reason why we want to estimate state sequence $\hat{\mathbf{X}}_i$ is because we want to use it to estimate the state matrix $\mathbf{A}$ by using the following relation

$$\hat{\mathbf{X}}_{i+1} = \mathbf{A}\hat{\mathbf{X}}_i \tag{2.161}$$

However, $\hat{\mathbf{X}}_{i+1}$ is still unknown. Fortunately, this can be achieved by analysing the time shifted version of Eq. (2.154), namely

$$\begin{aligned}
\hat{\mathbf{y}}\left(k+1|k+1\right) &= \mathbf{C}\mathbb{E}\left[\mathbf{x}\left(k+1\right) | \mathcal{Y}_{k+1}\right] \\
\hat{\mathbf{y}}\left(k+2|k+2\right) &= \mathbf{C}\mathbf{A}\mathbb{E}\left[\mathbf{x}\left(k+1\right) | \mathcal{Y}_{k+1}\right] \\
&\cdots \\
\hat{\mathbf{y}}\left(k+i-1|k+i-1\right) &= \mathbf{C}\mathbf{A}^{i-2}\mathbb{E}\left[\mathbf{x}\left(k+1\right) | \mathcal{Y}_{k+1}\right]
\end{aligned} \tag{2.162}$$

Similar to Eq. (2.155), we can get the following equation

$$\mathcal{P}_{i-1|2i-1} = \mathcal{O}_{i-1}\hat{\mathbf{X}}_{i+1} \tag{2.163}$$

To determine $\hat{\mathbf{X}}_{i+1}$, we first need to estimate projection matrix $\mathcal{P}_{i-1|2i-1}$ and the observability matrix $\mathcal{O}_{i-1}$. In this case, $\mathcal{O}_{i-1}$ can be obtained by simply deleting the last l rows of $\mathcal{O}_i$. While for $\mathcal{P}_{i-1|2i-1}$, it

is related to a different division of the block Hankel matrix (Eq. (2.135)) regarding the past reference and future output, namely

$$
\mathbf{H}_{2i} = \left(\frac{\mathbf{Y}_p^{\text{ref}+}}{\frac{\mathbf{Y}_{i|i}^{\sim\text{ref}}}{\mathbf{Y}_f^{-}}} \right) = \left(\begin{array}{c} \mathbf{Y}_{0|i}^{\text{ref}} \\ \hline \mathbf{Y}_{i|i}^{\sim\text{ref}} \\ \hline \mathbf{Y}_{i+1|2i-1} \end{array} \right) \begin{array}{l} \updownarrow r(i+1) \\ \updownarrow l-r \\ \updownarrow l(i-1) \end{array}
\tag{2.164}
$$

In this new division, we separate (-) the first block row of the future outputs $\mathbf{Y}_{i|i}$ from $\mathbf{Y}_f$, then split it into two parts, the first r rows are added (+) to the past references, while the remaining $l-r$ rows stay independently. We can do this is because the references are only a subset of the outputs ($r \le l$), and it is necessary to do this, because in order to estimate $\mathcal{P}_{i-1|2i-1}$, we need to use the past $i+1$ observations.

Now the same as Eq. (2.158), we can get the estimate of $\hat{\mathbf{Y}}_f^{-}$, which is also our projection matrix $\mathcal{P}_{i-1|2i-1}$ .by the following expression

$$
\mathcal{P}_{i-1|2i-1} = \hat{\mathbf{Y}}_f^{-} = \mathbf{\Sigma}_{f^- p^+} \mathbf{\Sigma}_{p^+ p^+}^{-1} \mathbf{Y}_p^{\text{ref}+}
\tag{2.165}
$$

where

$$
\mathbf{\Sigma}_{f^- p^+} = \mathbf{Y}_f^{-} \left(\mathbf{Y}_p^{\text{ref}+} \right)^{\text{T}} \quad \text{and} \quad \mathbf{\Sigma}_{p^+ p^+} = \mathbf{Y}_p^{\text{ref}+} \left(\mathbf{Y}_p^{\text{ref}+} \right)^{\text{T}}
\tag{2.166}
$$

With this the estimated $\mathcal{P}_{i-1|2i-1}$, we can easily obtain $\hat{\mathbf{X}}_{i+1}$ by the following equation

Since $\mathcal{O}_{i-1}$ can be obtained by simply deleting the last l rows of $\mathcal{O}_i$, then

$$
\hat{\mathbf{X}}_{i+1} = \mathcal{O}_{i-1}^{\dagger} \mathcal{P}_{i-1|2i-1}
\tag{2.167}
$$

Up to now, we have obtained all the information we need to solve Eq. (2.161) and obtain the state matrix $\mathbf{A}$. However, to get the observable mode shapes, we also need to find the output matrix $\mathbf{C}$. From Eq. (2.155), we see that the first row of $\mathcal{P}_{i|2i-1}$ contains the output matrix $\mathbf{C}$, namely

$$
\mathcal{P}_i = \mathbf{C}\hat{\mathbf{X}}_i
\tag{2.168}
$$

Since both $\mathcal{P}_i$ and $\hat{\mathbf{X}}_i$ are available at this point, we can easily find the output matrix $\mathbf{C}$.

Once we obtained the state matrix $\mathbf{A}$ and the output matric $\mathbf{C}$ via Eq. (2.161) and Eq. (2.168), the modal parameters can be estimated from Eq. (2.123) and Eq. (2.124).

2.5 Summary

In this chapter, we reviewed some basic concepts in system identification, particularly in modal analysis. We started with the basic mathematical model, i.e., linear differential equation with constant coefficients, then followed by the discussion of a linear difference equation with constant coefficients. Upon which we reviewed the impulse response function of the system in the time domain and its corresponding transfer function in the Laplace domain and z domain. Apart from this, we introduced the companion matrix to effectively compute the system poles via the eigenvalue decomposition method. And we also use it to convert a n th order linear differential or linear difference equation into a first order system while the system poles remain the same. With the companion matrix, the state-space expression of a system is introduced naturally. Due to the practical interest, our main discussion was focused on a stochastic system with a mathematical model to represent a stationary AR process. This model can be directly rewritten in the linear difference equation form or be expressed as a state-space model with a state equation and an output equation. When the random input is given by Gaussian white noise, this state-space model is often known as the linear Gauss-Markov model.

With the established mathematical model for the system, we revisited some modal parameters identification methods, namely, the Polyreference method and the SSI-COV method, the SSI-DATA method, to demonstrate the possible ways to estimate the system coefficients, with which we can easily compute the modal parameters we are looking for. Particularly, we provided a clearer and easier explanation for the Polyreference method and the SSI-DATA method based on the Yule Walkers equations and conditional Gaussian distribution, respectively. Besides, we articulated the role of the (block) Hankel matrix and the (block) Toeplitz matrix in obtaining the system coefficients.

In the next chapter, a new concept, i.e., real-valued and one-sided spectral density matrix, is introduced. The Polyreference method we discussed in this chapter is used as the identification technique to extract the modal parameters.

3. Real-valued and one-sided spectral density matrix

3.1 Introduction

In OMA, information is normally extracted from the random responses by calculating correlation functions and/or spectral density functions. If we consider a certain number of channels (sensors) d, then we would arrange the correlation functions in the correlation function (CF) matrix and the Spectral Density (SD) matrix. The CF and SD matrices are then $d \times d$ matrices [63]

It is well-known that the auto SD functions in the SD matrix (the diagonal elements) are real-valued functions, whereas the cross SD functions are complex-valued functions. However, even though the cross-spectral density functions are in fact complex-valued, the complex parts of these functions are normally relatively small compared to the real parts.

Further, when we sum up the SD functions matrices over the total frequency band to obtain the covariance matrix of the initial random responses based on the Parseval's theorem, this process will remove the imaginary parts of the SD matrix. It is because any imaginary part of the SD matrix contribution from the positive frequency band will be cancelled out due to the corresponding contribution from the negative frequency band.

As a result, it is natural to investigate a simple hypothesis stating: The imaginary parts of the SD matrix do not have any physical importance and thus can be removed. It could be beneficial when the system properties are nonchanging over the operation.

Therefore, in this chapter, we will investigate this hypothesis based on a comparative simulation study. Note that the proposed real-valued SD matrix defined in this chapter will be one-sided, i.e., only the positive frequency components are kept. However, it is different from the one-sided spectrum generated mechanically by multiplying 2 to the original spectrum then discard the negative frequency components, as it does not satisfy the Parseval's theorem.

Thus, in the first section of this chapter, we will briefly describe the basic idea about the real-valued one-sided spectral SD matrix. While in the second section, a digital sinusoidal generator with three active modes and two noise modes will be designed to test our idea. The results will be compared to the results produced by the classical complex-valued two-sided SD matrix. The Polyreference [63], [109] is adopted to estimate the modal parameters from the three types of the SD matrices. Lastly, we will

present a simple explanation of using the real-valued SD matrix to estimate the modal parameters based on Periodogram.

3.2 Real-valued and one-sided spectral density matrix

3.2.1 Parseval's theorem

Given a signal $x(n)$ with N sample points, the discrete Fourier transform (DFT) of it is given by

$$\tilde{x}(f_k) = \sum_{n=0}^{N-1} x(n) e^{-i2\pi f_k n}, \quad k = 0, 1, \ldots, N-1 \tag{3.1}$$

where the frequency $f_k = k/T$ are the Fourier frequencies, $T = N\Delta t$ is recording length, Δt is the sampling interval in the time domain. For the inverse DFT (iDFT) we have

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{x}(f_k) e^{i2\pi f_k n} \tag{3.2}$$

where $1/N$ is the sampling rate.

Now consider a new signal $z(n)$ which is the combination of the two signals, i.e., $z(n) = x(n) y(n)$, we can obtain the DFT of this signal by substituting it into Eq. (3.1), from which we obtain

$$\tilde{z}(f_j) = \sum_{n=0}^{N-1} x(n) y(n) e^{-i2\pi f_j n}, \quad j = 0, 1, \ldots, N-1 \tag{3.3}$$

Now if we substitute the iDFT of $x(n)$ into the above equation, we shall arrive at the following expression

$$\tilde{z}(f_j) = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{x}(f_k) \sum_{n=0}^{N-1} y(n) e^{-i2\pi(f_j - f_k)n} = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{x}(f_k) \tilde{y}(f_j - f_k) \tag{3.4}$$

when $j = 0$, $f_j = 0$, from Eq. (3.3)

$$\sum_{n=0}^{N-1} x(n) y(n) = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{x}(f_k) \tilde{y}(-f_k) \tag{3.5}$$

where $\tilde{y}(f)$ represents the Fourier transform of $y(n)$. For real signals, we have $\tilde{y}(-f) = \tilde{y}^*(f)$.

For a special case, when $x(n) = y(n)$, Eq. (3.5) becomes

$$\sum_{n=0}^{N-1} x(n)x(n) = \frac{1}{N}\sum_{k=0}^{N-1} \tilde{x}(f_k)\tilde{x}^*(f_k) \tag{3.6}$$

which is the Parseval's theorem for signal $x(n)$ given in the DFT form. Thus, given by Eq. (3.6), the Parseval's theorem says that the sum of the square of a function is equal to the sum of the square of its transform.

Note that the energy signals, which usually have finite length, $\left|\tilde{x}(f_k)\right|^2 = \tilde{x}(f_k)\tilde{x}^*(f_k)$, is known as the Energy Spectral Density (ESD); while for the stationary random processes, which usually have infinite energy but finite power, $\left|\tilde{x}(f_k)\right|^2$ is called the Power Spectral Density (PSD).

Practically, we normally deal with random measurements, so it is worth mentioning the Parseval's theorem for the random data.

For random measurements, we must consider the ensemble average of the measurements. Given Eq. (3.6), we can take expectation of both sides of the equation, that is

$$\sum_{n=0}^{N-1} \mathbb{E}\left[x(n)x(n)\right] = \frac{1}{N}\sum_{k=0}^{N-1} \mathbb{E}\left[\tilde{x}(f_k)\tilde{x}^*(f_k)\right] \tag{3.7}$$

3.2.2 Covariance matrix and spectral density matrix

Suppose we have a set of stationary random processes $\{X(n)\}$ collected from d locations (i.e., d measured degrees of freedom (DOF's)), each of which has N sample points. For convenience, we can group them into a data matrix $\mathbf{X} = \begin{bmatrix} \mathbf{x}(0) & \mathbf{x}(2) & \cdots & \mathbf{x}(N-1) \end{bmatrix}$, where $n = 0,1,\ldots,N-1$, $\mathbf{x}(n)$ is a $d \times 1$ vector, and $\mathbf{X}$ is a $d \times N$ matrix. Then the classical correlation function is given by

$$\mathbf{R}(m) = \mathbb{E}\left[\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n-m)\right], \quad m = 0,\pm 1,\ldots \tag{3.8}$$

where m denotes the time lag. When $m = 0$,

$$\mathbf{R}(0) = \mathbb{E}\left[\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\right] \tag{3.9}$$

If the Fourier transform of $\mathbf{x}(n)$ is given by $\tilde{\mathbf{x}}(f_k)$, according to Eq. (3.7), we shall have

$$\sum_{n=0}^{N-1} \mathbb{E}\left[\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\right] = \frac{1}{N}\sum_{k=0}^{N-1} \mathbb{E}\left[\tilde{\mathbf{x}}^*(f_k)\tilde{\mathbf{x}}^{\mathrm{T}}(f_k)\right] \tag{3.10}$$

for N measurement points.

Denote $\mathbf{G}(f_k) = \mathbb{E}\left[\tilde{\mathbf{x}}^*(f_k)\tilde{\mathbf{x}}^{\mathrm{T}}(f_k)\right]$, then $\mathbf{G}(f_k)$ gives the PSD matrix, which is a $d \times d$ Hermitian matrix. Eq. (3.10) can be simply expressed as

$$\sum_{n=0}^{N-1}\mathbf{R}(0) = \frac{1}{N}\sum_{k=0}^{N-1}\mathbf{G}(f_k) \tag{3.11}$$

3.2.3 Define the real-valued and one-sided spectral density matrix

In order to formulate a meaningful real-valued one-sided SD matrix, the DFT form of the Parseval's theorem as given by Eq. (3.10) will be used as our central equation.

Since in the DFT case, both the time function and the frequency domain function are periodic, we can choose the period as we want on both the time and the frequency axis. For the frequency axis, we will choose to follow normal practice having negative and positive frequencies, and where the maximum frequency is equal to the Nyquist frequency $f_{Ny} = 1/(2\Delta t)$, where Δt is the sampling time step. The period of the time function is then $T = N\Delta t$, and the frequency resolution is $\Delta f = 1/T$. This means that we have $N/2$ frequency steps from DC (Direct Current) to the Nyquist frequency. Thus we have $N/2+1$ frequency lines from DC to Nyquist, and then $N/2+1$ negative frequency lines, see Figure 1.(a), in which the classical discrete spectral density with 11 frequency lines from DC to Nyquist, and 9 negative frequency lines.

(a) The classical discrete spectral density   (b) The real-valued and one-sided spectral density
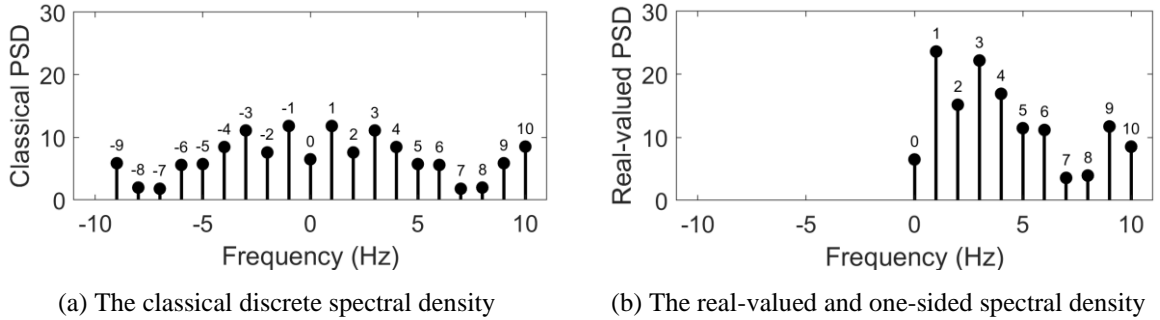
Figure 3.1. Demonstration of the discrete spectral density

With this choice of frequency axis, the discrete frequencies are given by

$$f_k = k\Delta f ; \quad k \in \left[-(N/2-1); N/2\right] \tag{3.12}$$

Performing the summation over the so defined frequency lines, the Parseval's theorem given by Eq. (3.10) is now expressed as

$$\sum_{n=0}^{N-1} \mathbf{R}(0) = \frac{1}{N} \sum_{k=-(N/2-1)}^{N/2} \mathbf{G}(f_k) \tag{3.13}$$

As we know that with real measurements, the DC matrix $\mathbf{G}(f_0)$ and the matrix $\mathbf{G}(f_{N/2})$ at Nyquist are always real. Note that the remaining matrices are complex, but fulfil

$$\mathbf{G}(-f_k) = \mathbf{G}^*(f_k) \tag{3.14}$$

we can now write Eq. (3.13) as a sum over the non-negative frequency lines

$$\sum_{n=0}^{N-1} \mathbf{R}(0) = \frac{1}{N} \left(\mathbf{G}(f_0) + \sum_{k=1}^{N/2-1} (\mathbf{G}(f_k) + \mathbf{G}^*(f_k)) + \mathbf{G}(f_{N/2}) \right) \tag{3.15}$$

And we see that all terms are now real-valued. This leads us to define the meaningful real-valued and one-sided spectral density

$$\mathbf{S}(f_k) = \begin{cases} \mathbf{G}(f_0); & k = 0 \\ \mathbf{G}(f_k) + \mathbf{G}^*(f_k); & k \in [1; N/2-1]; \\ \mathbf{G}(f_{N/2}); & k = N/2 \end{cases} \tag{3.16}$$

See Figure 3.1.(b), it shows the real-valued and one-sided spectral density only having 11 non-negative frequency lines. The result is that we can express the Parseval theorem summing over the non-negative frequency lines

$$\sum_{n=0}^{N-1} \mathbf{R}(0) = \frac{1}{N} \sum_{k=0}^{N/2} \mathbf{S}(f_k) \tag{3.17}$$

Every SD matrix $\mathbf{S}(f_k)$ can now be considered as the contribution to the covariance matrix from the frequency band $B_k = [f_k - \Delta f/2; f_k + \Delta f/2]$.

The defined real-valued one-sided SD matrix removes all the imaginary parts in the classical SD matrix. Intuitively, it may seem that this procedure stripped off some information from the original data. However, the Parseval's theorem tells us that the imaginary parts in the SD matrix do not contribute to the overall signal energy. Moreover, Eq. (3.17) gives the real distribution of the signal energy contained in the SD matrices along the frequency line. Apart from this, the imaginary part may contain some phase information. But for the mode shape estimation of a structure with non-changeable structural conditions and real modes, the imaginary parts of the SD matrix only give us noise. By removing it, we shall obtain a better mode shape estimation.

In order to eradicate our doubts, the subject of the next section will be focusing on verifying this real-valued SD via the Polyreference method based on a numerical model with different frequency spacing.

3.3 Modal parameters identification

3.3.1 Simulation case

Ultimately, we want to evaluate the performance between the classical complex SD matrix and real-valued SD matrix, then to make a conclusion whether it is legitimate to discard the imaginary part of the SD matrix in the OMA. The comparison items will be the identified modal parameters, i.e., natural frequencies, damping ratios, and mode shapes. The modal parameters are identified using the Polyreference technique, which we have clearly demonstrated in chapter 2.

Therefore, in this section, we consider a numerical model with five DOF's, where three modes are active, while the other two modes are inactive (i.e., by forcing the corresponding mode shapes to be zero). For the three active modes, two of them have fixed frequencies, while the other one is moving between them from the lower frequency to the higher frequency. In other words, the spacing between the two fixed frequencies and the moving frequency will be varied from closely spaced to reasonably separated modes.

Table 3.1 Basic settings for the simulation

Sample interval	$\Delta t = 0.323$
Data length for FFT	$\text{Nfft} = 512$
Number of segments.	$n_d = 32$
Total data points	$N = 512 \times 32$
Number of runs	$N_r = 100$
Nyquist frequency	$f_{Ny} = 1/2\Delta t = 1.545 \text{Hz}$
Noise level	1% of the maximum standard deviation of the 5 channels
Damping ratio	$\xi = 0.01$ for all five modes
Modal mass	$M = 1$ (kg)
First resonance frequency	$f_1 = 0.25 f_{Ny} = 0.386 \text{Hz}$
Third resonance frequency	$f_3 = 0.75 f_{Ny} = 1.159 \text{Hz}$
Second frequency	$f_2 \in [0.404, 1.129]$ Hz which is ranging from f_1 to f_3.
Noise modes	Randomly generated with frequencies between $0.25 f_1$ and f_{Ny}.

To be more specific, suppose the first frequency f_1 and the third frequency f_3 are fixed, while the second frequency f_2 will be moving along the frequency line from close to f_1 to f_3. For the two

inactive modes, the corresponding frequencies are randomly generated between $0.25 f_1$ and the Nyquist frequency f_{Ny}. The number of measurement channels is assumed to be five. Table 3.1 gives basic settings for the simulation.

By using the information given in Table 3.1, we design a simple vibration system, with the ith FRF $H_i(\omega)$ given by the following function,

$$H_i(\omega) = \frac{X_i(\omega)}{F_i(\omega)} = \frac{1}{M_i\left(\omega_i^2 - \omega^2 + \mathrm{i}2\xi_i\omega_i\omega\right)} , \quad i = 1, 2, \ldots, 5 \tag{3.18}$$

where $F_i(\omega)$ is the Fourier transform of the input response, while $X_i(\omega)$ is the corresponding Fourier transform of the output response. $\omega_i = 2\pi f_i$ is the ith frequency in radians. $H_i(\omega)$ is obtained by substituting the parameters shown in Table 3.1 into Eq. (3.18). Figure 3.3 shows the FRF amplitude plot and the corresponding phase plot of one of the simulations of the system. As we can see from the FRF plot, there are exactly five frequency components.

To simulate the stationary random output response of the system, a stationary random input (see Figure 3.4) is simulated first, then the randomly generated mode shapes are used to transform the input responses into modal forces. Since we have five modes, we can obtain five modal forces.

Note that the mode shape vectors or actual modes we used to generate the responses are generated in a random fashion, and between each vector, a maximum 20% Modal Assurance Criteria (MAC) value [110] limit is given. This is needed to secure a low linear correlation between the modes since, practically, normal modes should be statistically independent of each other. All the actual mode shape vectors are real. Figure 3.2 shows the MAC values of the generated mode shape vectors in one of the simulations.

Having obtained the modal forces, we take the Fourier transform of them, and then we can use the relation given by Eq. (3.18) to calculate the Fourier transform of modal coordinates $X_i(\omega)$ for each mode. Once obtained the $X_i(\omega)$ for each mode, we can take the inverse Fourier transform of them, then simply apply the inverse modal superposition technique [111] to $X_i(\omega)$, the output response of the system can be obtained.

It must be pointed out that, in order to reduce the mean bias, for every new position of f_2, the simulation will run $N_r = 100$ times. For each run, a new set of mode shape vectors and input and output responses will be generated. Besides, we will use the angle between the identified mode shape vector and the reference (actual) mode shape vector to assess the accuracy of the estimated mode shapes. Thus,

for every new value of f_2, since we have 100 sets of mode shape vectors, we will calculate the relative errors for each case. Having done this, we can average the relative errors then plot them against the frequency spacing as f_2 moving from f_1 to f_3. Figure 3.4 illustrates a sample input signal, while Figure 3.5 depicts a sample output put response from channel 2.
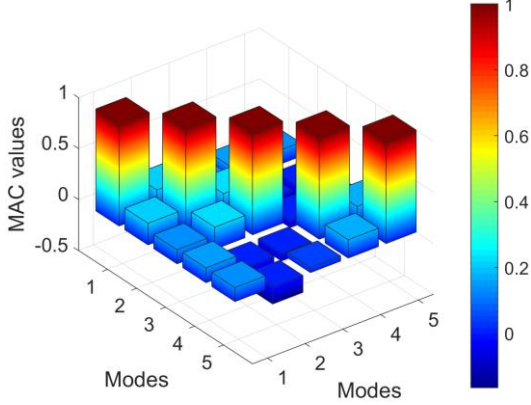
Figure 3.2 MAC values of the generated mode shapes
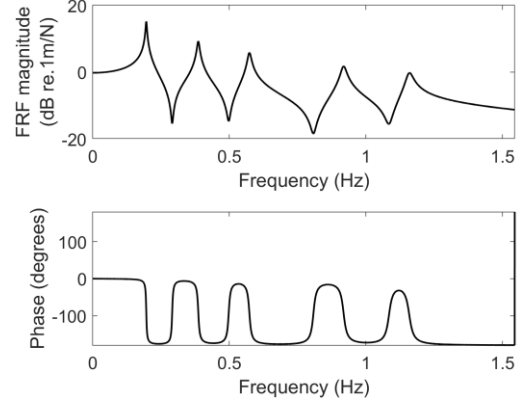
Figure 3.3 The Frequency Response Function (FRF) and the phase information of the system
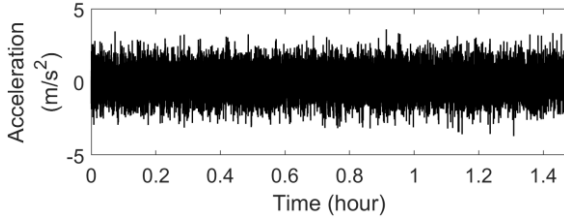
Figure 3.4 A sample input response
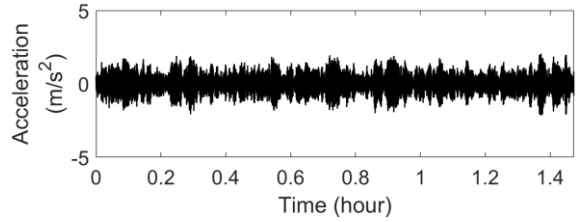
Figure 3.5 A sample output response of the system from channel 2

3.3.2 Identification results and discussion

The SD matrix of the responses is estimated from the simulated responses via Welch's method by using the Hanning window with 50% overlap. From here, the simulation will be split into two routes.

1. The first one keeps the classical SD matrix then takes them back to the time domain by using the inverse Fast Fourier transform (iFFT). What we obtain is the real-valued correlation function matrix.

2. The other one will follow the idea given by Eq. (3.16). Notice that the real-valued SD matrix we defined in Eq. (3.16) is one-sided, while the conventional iFFT requires the SD matrix to be double-sided. In principle, if we use the traditional method, we need to use the real part of the SD matrix for both the positive and negative frequency axis so as to satisfy both

requirements from the conventional iFFT and the real-valued SD matrix we defined. The reason being that the traditional iFFT in which the negative axis is put equal to the complex conjugate values on the positive axis. Just like the one-sided SD according to Eq. (3.16) yields the correct correlation matrix, also this iFFT gives the correct correlation matrix.

Another way is to keep the positive frequency component as it is given by Eq. (3.16), then fill up the negative frequency part with zeros. Having done this, we can apply the iFFT to this pseudo double-sided SD matrix to obtain the complex-valued matrix in the time domain. This procedure is unknown as the Hilbert transform [112]. Note that both approaches will produce a complex-valued matrix in the time domain because of the absence of the imaginary part of the SD matrix. The correlation matrix we wish to obtain is the real part of it.

The rest of the procedure is the same for the two cases. We apply the Polyreference method to the estimated correlation matrices obtained from the above procedure, but we take the transpose of all matrices so that each column represents an unbiased free response as mentioned in Brincker and Ventura [63].

Figure 3.6 shows the traditional singular value plot from one of the simulations based on the classical two-sided SD matrix, which is obtained by using the FDD (Frequency Domain Decomposition) method [63]. Particularly in Figure 3.6 (a), we plotted three types of singular values in one figure. The black line denotes the singular values obtained from the complex SD matrix, and the red dot line represents the singular values calculated from the real part of the SD matrix, while the blue line gives the total singular values (i.e., the sum of the singular value calculated at each frequency line) computed from the imaginary part of the SD matrix. It is not difficult to see that the red lines are almost identical to the black line, and the peaks shown in the blue line share the same frequency locations with the black and the red lines. Apart from this, even though the blue line is not as smooth as the black and the red line, but the overall trend of it is the power shifted version of the first singular values given by the black or the red line. For the case we considered in this study that the system is built upon the real mode shapes with low damping, the behaviour of the different singular values in Figure 3.6 (a) partly shows that the imaginary part of the SD matrix does not provide extra information to compare to its real counterpart.

Figure 3.6 (b) shows the recovered correlation response from the 2nd channel. Since the classical SD matrix is complex in the frequency domain, it has the real counterparts in the time domain. Therefore, we see a straight red line in Figure 3.6 (b).
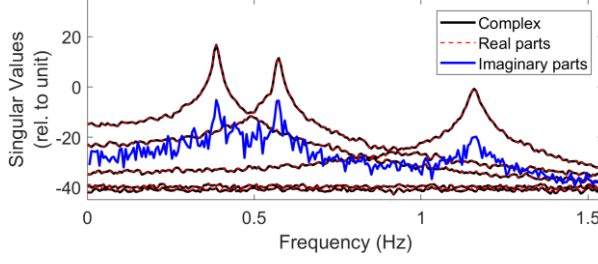
(a) Singular value plot based on classical SD matrix
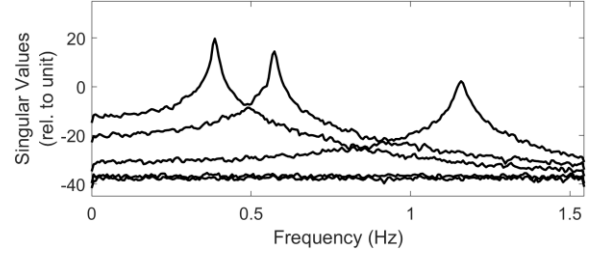
(a) Singular value plot based on the real-valued and one-sided SD matrix
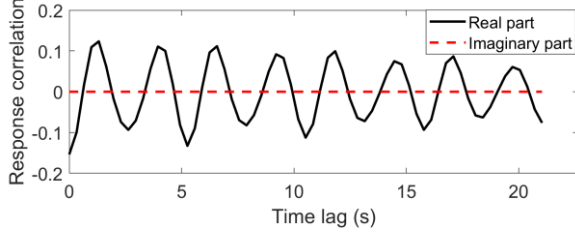
(b) Recovered correlation response from channel 2
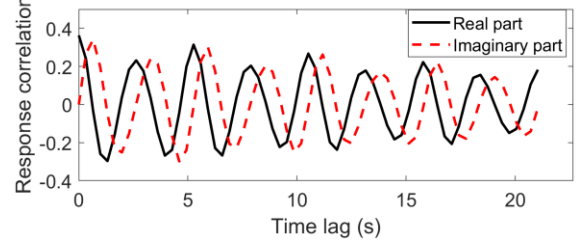
(b) Recovered correlation response from channel 2

Figure 3.6 Singular value plot and the Recovered correlation response obtained based on the classical two-sided SD matrix

Figure 3.7 Singular value plot and the Recovered correlation response obtained based on the real-valued and one-sided SD matrix

Figure 3.7 on the other hand, shows the singular value plot and the recovered correlation response from channel 2 based on the proposed real-valued one-sided SD matrix. Since the real-valued SD matrices are all real values in the frequency domain, then its counterparts in the time domain are then complex-valued. This is why we see an oscillating red line in Figure 3.7 (b), which is the imaginary part of the recovered correlation function.

The relative errors of the estimated frequencies between the reference frequencies for the three active modes are shown in Figure 3.8. Note that each point shown as a circle or an asterisk is an average of 100 runs. The abscissa of all plots shows the relative frequency spacing between f_2 and f_1, which is computed through the following fashion

$$r_f = \frac{f_2 - f_1}{f_1} \tag{3.19}$$

From Figure 3.8, we see that $r_f \in (0,2)$, which implies that $f_2 \in (f_1, 3f_1)$, and $f_3 = 3f_1$.

Since f_1 and f_3 are fixed, there is no apparent trend to tell from Figure 3.8 (a) and (c) for the first and the third mode. While for the second mode, the relative errors decrease with the increasing frequency spacing. Besides, the overall relative error for the first mode is larger than the third mode. We see such a phenomenon is because we assigned the same value to the damping ratio for all modes, within the

half power bandwidth, i.e., $\Delta f = 2 f \xi$, the higher the mode, the larger the resonance frequency, the wider the half-power bandwidth. As a result, within a certain measurement length, the frequency resolution is fixed; when the half-power bandwidth is wider, we will have more points fall into that bandwidth. Accordingly, in the time domain, there will be more cycles in that measurement length, which means more information, and more information means more accurate. This is why we obtain lower relative errors for the third mode, and when the second mode is moving toward it, the relative error of the second mode is getting smaller.

(1) Mode 1 frequency errors

(1) Mode 1 damping ratio errors

(2) Mode 2 frequency errors

(2) Mode 2 damping ratio errors

(3) Mode 3 frequency errors

(3) Mode 3 damping ratio errors
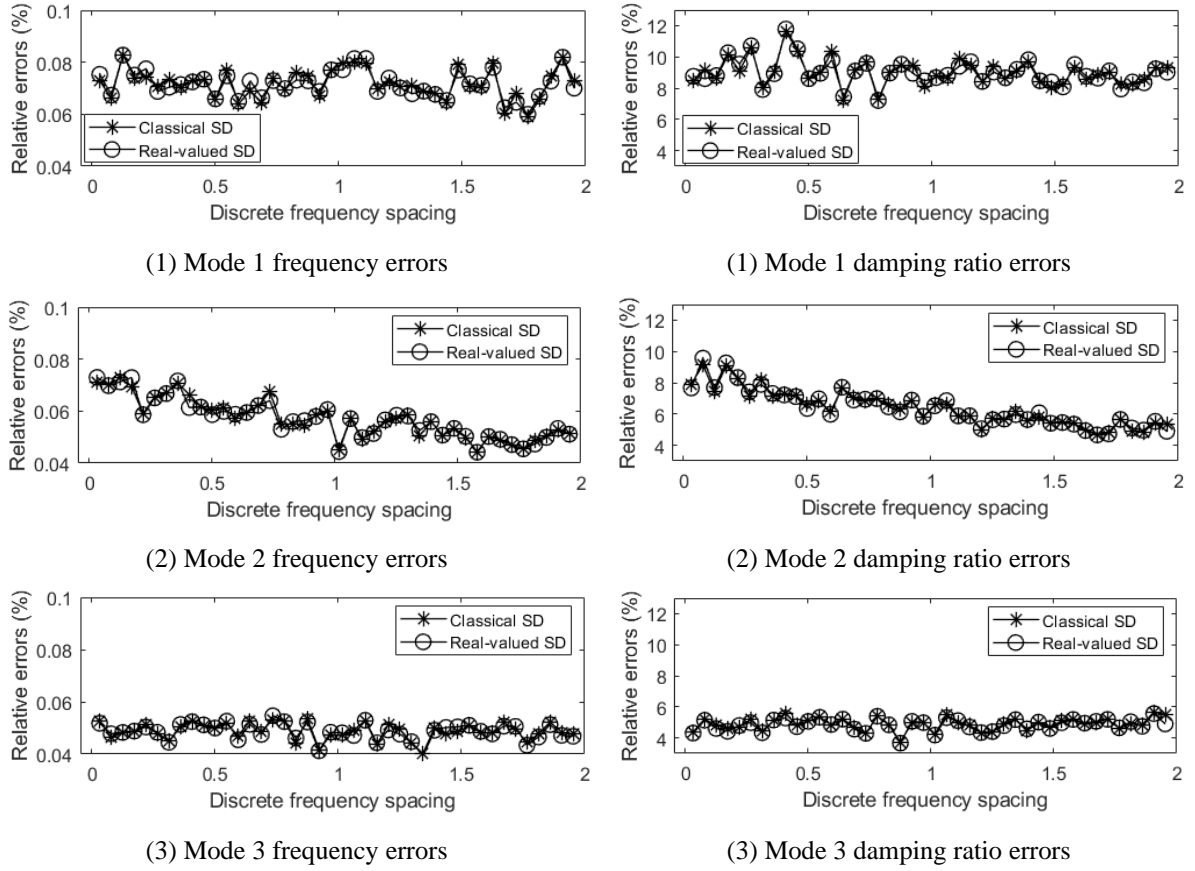
Figure 3.8 Relative error (%) of the estimated frequencies

Figure 3.9 Relative error (%) of the estimated damping ratios

Furthermore, if we compare the relative errors between the two different procedure for these three modes, we cannot really tell the differences between the identified results, for the relative errors we obtained based on the classical SD matrix are similar to the relative errors we obtained in terms of the real-valued SD matrix. Additionally, the errors for all three modes based on those two different procedures are all very low, and they are both less than 0.09%.

As for the estimated damping ratios, which are shown in Figure 3.9, we see the same phenomenon as given by the estimated frequencies, except that the relative errors are higher than the relative errors we obtained from the estimated frequencies.

(1) Mode 1 mode shape errors

(1) Mode 1 mode shape errors differences
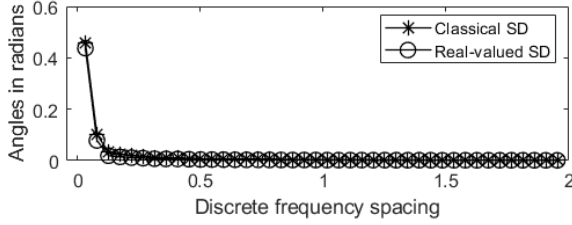
(2) Mode 2 mode shape errors

(2) Mode 2 mode shape errors differences
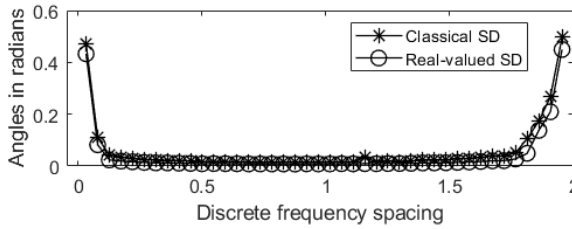
(3) Mode 3 mode shape errors
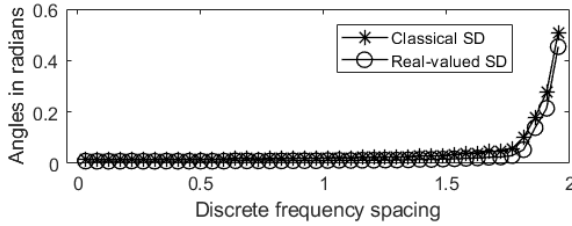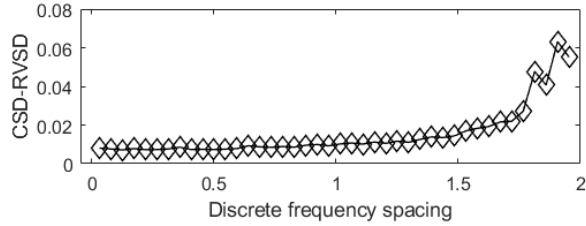
(3) Mode 3 mode shape errors differences

Figure 3.10 Averaged angles between the identified mode shape and the refence mode shape

Figure 3.11 The differences of averaged angles between the identified mode shape and the refence mode shape obtained from real-valued SD matrix and the classical SD matrix (Note: CSD—classic SD; RVSD—real-valued SD)

Figure 3.10 shows the angles between the identified mode shape and the reference mode shape. The angles are obtained from the MAC values. Since the MAC value gives the correlation between the normalised two vectors, which is the square of the cosine function of the angle between the two vectors, when the two vectors are close in direction (or in the opposite direction), the MAC value will be close to 1, accordingly, the angle between the two vectors is close to zero or π. However, since the MAC is a square correlation relation between two vectors, the angle we obtained based on the inverse of the square root of the MAC value will be between zero and $\pi/2 = 1.571$. For the analysis, the larger the angle, the poorer the estimate.

From Figure 3.10, we see that the first mode shows a decreasing trend in the angle value, while the third mode shows an increasing trend. As for the second mode, when it is moving from the first mode to the third mode, the angles first decrease, then increase. This behaviour of the angles shown in Figure 3.10 is consistent with our understanding of the close modes and the well-separated modes. It is well-known that when we get closer to the repeated pole case (when the frequency spacing approaches zero to either the first mode or the third mode), then the two closely spaced modes will interact by rotating the two mode shapes in their common subspace, see Brincker and Aenlle [113]. As a result, any linear combination of the two mode shapes are also the mode shapes for the structure for the case of repeated poles. In this case, the large angles, shown in Figure 3.10 when the two modes are close together, do not necessarily mean the estimated mode shape is a poor representation of the actual mode shapes. This, on the other hand, implies that the MAC is not a good criterion to be used to assess the case of close modes.

Regardless of the close mode problem, the angles obtained based on the real-valued SD matrix are smaller than the angles obtained based on the classical complex-values SD matrix. It is not very obvious in Figure 3.10. But in Figure 3.11, which gives the differences of averaged angles obtained from the real-valued SD matrix and the classical SD matrix, shows that the angles obtained based on the classical SD matrix are larger than angles obtained from the real-valued SD matrix, as their differences are all above zero. Besides, as we can see that the higher the modes, the larger the differences when the modes are well-separated.

The identification result in terms of the mode shapes shows the advantage of using the proposed real-valued SD matrix in identifying the mode shapes, while the natural frequencies and the damping ratios identification result says that the real-valued SD method is comparable with the classical SD method. Overall, it means that the defined real-valued SD matrix is rational, since the results in Figure. 3.8 and Figure. 3.9 tell us that the lack of the imaginary information of the SD matrix will not affect the identification of the natural frequencies and the damping ratios of the system.

3.3.3 Theoretical explanation of using the real spectral density matrix for identification

In the previous section, we showed numerically that the imaginary part of the SD matrix does not play a role in modal parameters estimation provided that the system has real modes. In this section, we provide a simple explanation based on the Periodogram of the vibration response for the legitimacy of this procedure.

The general solution of a second-order ordinary differential equation (ODE) with constant coefficient can be given by

$$\psi(x,t) = a(x)e^{st} + a^*(x)e^{-st} \tag{3.20}$$

where s and $-s$ represents the system poles. $a(x) = c(x)\phi(x)$, the constant coefficient $c(x)$ is the determined by the initial conditions, while $\phi(x)$ represents the eigenfunction of the system. Depending on the occurrence of the damping term in the ODE function, $\phi(x)$ and s can be complex-valued. Thus, more specifically, $a(x)$ can be given by a general form of

$$a(x) = \big(C(x) + iD(x)\big)\big(\phi_R(x) + i\phi_I(x)\big) \tag{3.21}$$

where $c(x) = C(x) + iD(x)$, $\phi(x) = \phi_R(x) + i\phi_I(x)$. Note that $C(x)$, $D(x)$, $\phi_R(x)$ and $\phi_I(x)$ are all real.

If we evaluate it on the imaginary line, i.e., $s = i\omega$, we get.

$$\psi(x,t) = a(x)e^{i\omega t} + a^*(x)e^{-i\omega t} \tag{3.22}$$

Suppose the vibration behaviour of a continuous system with d active modes can be characterised by a 2$^{\text{nd}}$ order ODE, then for each mode its vibration response can be given by Eq. (3.22). Then the overall response of the system is the sum of d equations of Eq. (3.22), that is

$$\psi_t(x) = \sum_{n=1}^{d} a_n(x)e^{i\omega_n t} + a_n^*(x)e^{-i\omega_n t} \tag{3.23}$$

Note that we used another notation $\psi_t(x)$ to represent the time discrete form of $\psi(x,t)$, where $t = 1, 2, \ldots$.

On the other hand, the finite Fourier series representation of $\{\psi_t(x)\}$ is given by

$$\psi_t(x) = \sum_{p=1}^{N/2} \big(A_p(x)\cos(\omega_p t) + B_p(x)\sin(\omega_p t)\big) \tag{3.24}$$

for $t = 1, 2, \ldots, N$, where $\omega_p = 2\pi p / N$, while $A_p(x)$ and $B_p(x)$ are the real-valued Fourier series coefficients, which are given by

$$A_p(x) = \frac{2}{N}\sum_{p=1}^{N/2}\psi_t(x)\cos(\omega_p t) \quad \text{and} \quad B_p(x) = \frac{2}{N}\sum_{p=1}^{N/2}\psi_t(x)\sin(\omega_p t) \tag{3.25}$$

Note that the mean value of $\psi_t(x)$ is assumed to be zero.

Now, rewrite Eq. (3.24) into the exponential form, we obtain

$$\psi_t(x) = \sum_{p=1}^{N/2} \left(\frac{A_p(x) - iB_p(x)}{2} e^{i\omega_p t} + \frac{A_p(x) + iB_p(x)}{2} e^{-i\omega_p t} \right) \tag{3.26}$$

In order to observe the right number of modes by using the Fourier analysis methods, we need to make sure $N/2 > d$. When $N/2 > d$, we will have $(N/2 - d)$ redundant modes or spurious modes. And if we extend d to $N/2$, we can relate $a(x)$ to $A(x)$ and $B(x)$, namely

$$a_p(x) = \frac{A_p(x) - iB_p(x)}{2} \quad \text{and} \quad a_p^*(x) = \frac{A_p(x) + iB_p(x)}{2} \tag{3.27}$$

The Periodogram of $\psi_t(x)$ at location x is given by

$$I_p(x) = \frac{1}{\pi N} \left(\sum_{t=1}^{N} \psi_t(x) e^{i\omega_p t} \right) \left(\sum_{t=1}^{N} \psi_t(x) e^{-i\omega_p t} \right) \tag{3.28}$$

which can be further reduced to a simpler form

$$I_p(x) = \frac{N}{4\pi} \left(A_p^2(x) + B_p^2(x) \right) \tag{3.29}$$

According to the relation between $a_p(x)$, and $A_p(x)$ and $B_p(x)$, we can write $I_p(x)$ in terms of $a_p(x)$, that is

$$I_p(x) = \frac{N}{4\pi} a_p^2(x) \tag{3.30}$$

Further we can obtain the cross-Periodogram between two locations x_1 and x_2, namely

$$I_p(x_1, x_2) = \frac{1}{\pi N} \left(\sum_{t=1}^{N} \psi_t(x_1) e^{i\omega_p t} \right) \left(\sum_{t=1}^{N} \psi_t(x_2) e^{-i\omega_p t} \right) \tag{3.31}$$

However, unlike $I_p(x)$, $I_p(x_1, x_2)$ is complex-valued, and its real and imaginary part is given by

$$\mathrm{Re}\left[I_p(x_1, x_2) \right] = \frac{N}{4\pi} \left(A_p(x_1) A_p(x_2) + B_p(x_1) B_p(x_2) \right)$$
$$\mathrm{Im}\left[I_p(x_1, x_2) \right] = \frac{N}{4\pi} \left(A_p(x_1) B_p(x_2) - A_p(x_1) B_p(x_2) \right) \tag{3.32}$$

Note that in the usual SD matrix, the $I_p(x)$ goes to the diagonal terms, while $I_p(x_1, x_2)$ forms the off-diagonal terms. If and only if $a_p(x)$ is real, $\mathrm{Im}\left[I_p(x_1, x_2) \right]$ can disappear, then the SD matrix

would be real. When the mode shapes are real or close to real, we can discard $\phi_I(x)$ in $\phi(x)$, then $a_p(x)$ is left with $\phi_R(x)(C(x)+iD(x))$. In this case, we have

$$A_p(x)=2\phi_R(x)C(x) \text{ and } B_p(x)=2\phi_R(x)D(x) \tag{3.33}$$

Substitute them into Eq. (3.29) and Eq. (3.32) we obtain

$$I_p(x)=\frac{N}{\pi}\phi_R^2(x)\left(C_p^2(x)+D_p^2(x)\right)2\phi_R(x)D(x)$$

$$\operatorname{Re}\left[I_p(x_1,x_2)\right]=\frac{N}{\pi}\phi_R^2(x)\left(C_p(x_1)C_p(x_2)+D_p(x_1)D_p(x_2)\right) \tag{3.34}$$

$$\operatorname{Im}\left[I_p(x_1,x_2)\right]=\frac{N}{\pi}\phi_R^2(x)\left(C_p(x_1)D_p(x_2)-C_p(x_1)D_p(x_2)\right)$$

From this equation, we notice that the imaginary part of the SD matrix should not be deemed as noises. It contains the same mode shape information as the real part but with different power (see Figure 3.6 (a)), which made it trivial for mode shape estimation provided that the mode shapes are real or close to real. As for the natural frequencies and damping ratios, the information is well stored in $c_p(x)$, where $c_p(x)=C_p(x)+iD_p(x)$, which is determined by the initial conditions.

3.4 Conclusion

In this chapter, a new concept, the real-valued one-sided SD matrix, which fulfils the Parseval's theorem, is verified by conducting a comparative simulation study with the classical complex-valued SD matrix. The simulation results clearly show that both approaches give the same level of accuracy for the measured frequencies and damping ratios. For the mode shapes, the error measured as the angles between the exact and the estimated mode shapes seems to be smaller than when we use the real-valued SD matrix. Upon the numerical analysis, we also provided a simple theoretical explanation for this new concept based on the Periodogram. Therefore, the results of our analysis support the initial hypothesis that we can remove the complex part of the SD matrix without losing quality in the identification process, provided that the mode shapes are real or close to real.

Besides, it is worth mentioning that the idea of the real-valued one-sided SD matrix is applied to most of the case studies throughout this thesis. We can do this is because the systems we concerned about have no or low damping force, then the mode shapes of such systems are real or close to real.

4. On the theory of the Frequency Domain Decomposition (FDD) identification technique and the close modes estimation via a density operator

4.1 Introduction

The popularity of operational modal analysis (OMA) has increased significantly over the last 10-20 years. The reason is that large structures are much easier to identify in this way since ambient excitation is used instead of artificial excitation, which is difficult to generate with sufficiently large magnitude at the low frequencies of interest. Further, simple and reliable identification techniques have been developed. In the last few years, three textbooks have been published in the area: Rainieri and Fabbrocino [97], Brincker and Ventura [63], and S. Au [114].

As one of the most popular identification techniques used in OMA, the Frequency Domain Decomposition (FDD) technique was first introduced in its simplest form by Brincker et al. [52], [115], and later as the enhanced FDD (EFDD) that also included damping estimation, Brincker et al. [53].

The technique is mainly popular due to its user-friendliness, where a large number of sensor channels, for instance, 10 degrees of freedom (DOF's) can be considered in one spectral density plot showing the singular values of the Power Spectral Density (PSD) matrix. See, for example, Figure 4.1, which shows a typical singular value plot of the stationary measurements. By observing the peaks in the first singular values, we can see three modes, and two of them are closely spaced in the frequency band from 3.5 Hz to 4 Hz.

It is a great simplification compared to the conventional way of considering individual PSD functions. For example, in the case of 10 DOF's we would have to consider 55 individual PSD functions so as to take all information into account. For such a reason, the FDD technique has also become widely used due to the user-friendly implementation in commercial software. The technique is protected by patent [116].

The basic routine of the FDD technique is the estimation of the mode shapes from the linear transform of the PSD matrix [52], [115]. And this linear transform is obtained by performing the Singular Value Decomposition (SVD) to the PSD matrix at each frequency line. Then the first singular vector, which

70

is in the column space of the PSD matrix, is chosen as the estimate of the mode shape at the consider peak. However, the two main explanations that exist for this procedure are not very well-founded.
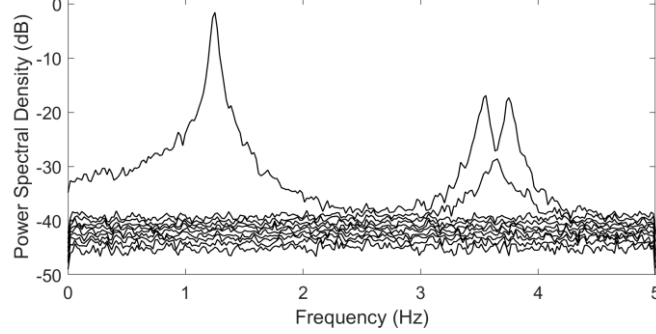
Figure 4.1. Singular value plot of the stationary responses gathered from 10 measurement locations

The original explanation, which is given in Brincker et al. [52], [115], considers the theoretical solution for the PSD matrix assuming white noise input. This explanation is based on the formula of the input and output PSD relationship, namely,

$$\mathbf{G}_{yy}(f) = \mathbf{H}_f \mathbf{G}_{xx}(f) \mathbf{H}_f^{\mathrm{H}} \tag{4.1}$$

where $\mathbf{G}_{xx}(f)$ and $\mathbf{G}_{yy}(f)$ are the input and output PSD matrices, respectively. While $\mathbf{H}_f$ denotes the Frequency Response Function (FRF) matrix. The superscript $*^{\mathrm{H}}$ denotes the Hermitian transpose. The fact is, the explanation based on Eq. (4.1) is not just convoluted, but also not actually related to the real application of the FDD.

Later a simpler explanation based only on the output PSD matrix was presented in Brincker and Ventura [63]. This time the output PSD matrix was obtained in terms of the Fourier transform of the correlation function matrix of the measurements. To be more specific, let $\mathbf{y}(t)$ and $\mathbf{q}(t)$ denote the output normal coordinates (or geometric coordinates) and corresponding modal coordinates (or generalised coordinates), respectively. By the mode superposition technique, we have $\mathbf{y}(t) = \mathbf{\Psi}\mathbf{q}(t)$. As a result, the correlation function matrix of $\mathbf{y}(t)$ is given by

$$\mathbf{R}_{yy}(\tau) = \mathbb{E}\left[\mathbf{y}(t)\mathbf{y}^{\mathrm{T}}(t+\tau)\right] = \mathbf{\Psi}\mathbb{E}\left[\mathbf{q}(t)\mathbf{q}^{\mathrm{T}}(t+\tau)\right]\mathbf{\Psi}^{\mathrm{T}} = \mathbf{\Psi}\mathbf{R}_{qq}(\tau)\mathbf{\Psi}^{\mathrm{T}} \tag{4.2}$$

where $\mathbf{\Psi}$ is the mode shape matrix. If the modal coordinates are assumed to be independent, then the PSD matrix of the modal coordinates is diagonal. In this case, when we decompose the PSD matrix $\mathbf{G}_{yy}(f)$ as follows by using the normal FDD estimation procedure, i.e., SVD technique

$$\mathbf{G}_{yy}(f) = \mathbf{V}_f \mathbf{\Sigma}_f \mathbf{V}_f^{\mathrm{H}} \tag{4.3}$$

71

the first singular vectors in $\mathbf{V}_f$ are deemed as the mode shape estimates, while the diagonal matrix $\mathbf{\Sigma}_f$ is the PSD matrix of the modal coordinates.

Again, this explanation is not very satisfactory. The reason is that it is not a necessary condition for obtaining good mode shape estimates that the modal coordinates are uncorrelated. It is easy to find examples where the modal coordinates are significantly correlated, but Eq. (4.3) can still be used to obtain good estimates of the involved mode shapes using the information in the column space of $\mathbf{V}_f$.

However, when two modes are very closely spaced, the information delivered by the FDD may not be accurate or enough for us to determine the modal parameters. For instance, when two modes fall into the limit of the spectral resolution, then we are unable to resolve those two modes based on the first singular value plot. As a result, we are unable to give a proper natural frequency estimation as well as the mode shape estimations for those two close modes. Therefore, in section 4.4 of this chapter, we will introduce a density operator (also called a density matrix) [117] instead of the PSD matrix $\mathbf{G}_{yy}(f)$ to identify the mode shapes. Since the density operator is derived based on the PSD matrix $\mathbf{G}_{yy}(f)$, the identification process of using the density operator is similar to the FDD that we still need to find the peak frequencies. However, those peak frequencies are different from the ones given by the FDD. We will show that the best mode shape estimates do not necessarily occur at the peak frequencies of the first singular value plot of the FDD.

To move from the PSD matrix to the density operator, we need to have a different interpretation of $\mathbf{G}_{yy}(f)$ and the FDD, as the two classical explanations for the FDD given by Eq. (4.1) and Eq. (4.3) are not good enough to support our new application and derivation, not to mention the rules for how to estimate the mode shapes from the column space of $\mathbf{V}_f$ are never well founded. Therefore, one of the aims of this chapter is to reinterpret the FDD technique, and upon which we will propose a new way to tackle the close modes modal parameters estimation problem based on the density operator.

Except for this introduction section, the work in this chapter has three essential sections. In section 4.2, we aim to establish our new theory for the FDD and a density operator, and it comprises two parts. In the first part, we define a PSD estimator based on the Periodogram; while in the second part, we first investigate the property of this PSD estimator and redefine the FDD in terms of the Principal Component Analysis (PCA), then introduce a density operator and derive it based on a PSD estimator.

In section 4.3, we explore the possibility of using the FDD to deal with a certain kind of nonstationary processes. Although the FDD is built upon the Fourier transform, it is naturally not suitable for dealing with the nonstationary processes in a wide sense, as it is unable to provide any time information of the data. However, if only the frequency components are needs, and that frequency is not a function of time, then we can use the FDD to identify the modal parameters from such measurements. The work in section

4.3 may be incompatible with the main discussion of this chapter. However, the discussion in this section can advocate the application of the FDD to the nonstationary response we would obtain from a bridge subject to a moving vehicle excitation since the frequency of the bridge does not change with time, as only the amplitude of the measurements changes.

While in section 4.4, the close mode analysis is discussed in depth. The model we used is inherited from chapter 3 with only a few changes to the parameters. Since the natural frequencies and damping ratio cannot be obtained by analysing a density operator alone, we design a two-stage method, namely, Enhanced Polyreference. By using it we can extract all the modal parameters (i.e., natural frequencies, damping ratios, and mode shapes) we can get in an OMA analysis. As its name implies, it is a combination of the Polyreference method. In the first stage, we analyse a density operator to get the mode shape estimation, then use it to decompose the normal coordinates into a set of modal coordinates; in the second stage, we apply the Polyreference method to those modal coordinates to extract the natural frequencies and damping ratios. In order to evaluate the performance of our proposed method, the same data is analysed by the Polyreference method and the data-driven Stochastic Subspace Identification (SSI-DATA) method. Details of those two methods are presented in chapter 2 (see section 2.4.1.2 and section 2.4.3.2). In our analysis, different cases regarding different MAC values of two close modes, damping ratios, and noise levels are investigated to reveal the underlying issues associated with the close modes identification problem.

4.2 Theoretical background

4.2.1 Part I — Spectral analysis of a random process

In the first part, our purpose is to define a valid PSD estimator based on the stationary theory of a time series. First, we will present the mathematical definition of a stochastic process and its stationarity properties in section 4.2.1.1. Then in the next three subsections, we will discuss two common approaches, i.e., correlogram and Periodogram, to obtain a PSD function. However, we chose the Periodogram approach, and reasons are revealed in section 4.2.2.4 , which in the second part of the theoretical background. Based on our preference, in section 4.2.1.5, several smooth methods are discussed to control the variance of a PSD function. While in section 4.2.1.6, the matrix notation of a PSD estimator is presented, in the meantime, the smoothing process is converted into a matrix form.

4.2.1.1 Definition of a stochastic process

A measurable space is an ordered pair $(E, \mathcal{U})$, where E is a set and $\mathcal{U}$ is a σ-algebra on E, while a measure space is an ordered triple $(E, \mathcal{U}, \mu)$, where μ is a measure on $(E, \mathcal{U})$ [118]. So, given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and a measurable space $(S, \mathcal{O})$, where Ω is a set of outcomes, $\mathcal{A}$ is a σ-algebra of events, $\mathbb{P}$ is the probability measure on the sample space $(\Omega, \mathcal{A})$, S is a metric space and $\mathcal{O}$ is its corresponding Borel σ-algebra, a random variable X is a mapping from a probability space to a metric space. In other words, as a random element of S, X is defined on its domain Ω and in its range S [119]. If $\omega \in \Omega$ denotes a generic outcome or a sample point, then $X(\omega)$ measures an observable quantity of that outcome [120].

For a random process, we need to define another measurable space $(D, \mathcal{T})$ with D denotes the index set and $\mathcal{T}$ be its corresponding Borel σ-algebra. Hence, a random process is a collection of random variables on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with state-space $(S, \mathcal{O})$ and index set D, which can be written as $\{X(t) : t \in D\}$ such that $X(t)$ takes values in S for each $t \in D$. Conventionally, the elements $t \in D$ has the meaning of time, and $D = \mathbb{N}$ for the discrete case, while $D = [0, \infty)$ for the continuous case. Under such interpretation, $X(t) \in S$ is the state of the random process at time $t \in D$, then $(D, \mathcal{T})$ is called the time space.

Apart from $\{X(t) : t \in D\}$, a random process can also be written as $\{X(t, \omega) : t \in D\}$, where $\omega \in \Omega$, or $X(t) : D \times \Omega \to S$, where $\times$ means taking the Cartesian product of two sets. All three expressions for a random process are very useful, and sometimes we will simply use $\{X(t)\}$ to denote a random process. Especially for the last two, which explicitly reflects that a random process is actually a function of two variables. Therefore, we will be using them in the following content interchangeably without making further explanation.

Note that a random variable is written in capital, e.g., $X(t)$ or X_t, while a particular realisation of the ensemble will be denoted by the lowercase, e.g., $x(t)$ or x_t. For time series, we call the infinite set of time series an ensemble, and the observed time series is just one example of the infinite set of time series that might have been observed.

Unlike nonstationary processes describe systems which are evolving with time, stationary processes generally arise from stable systems which have attained a steady-state mode of operation. The word

"steady-state" means that the system is in a state of statistical equilibrium, which implies that the statistical properties of the process do not change over time [121]. If $\{X(t)\}$ is such a process such that the joint distribution of $X(t_1),\ldots X(t_k)$ is the same as the joint distribution of $X(t_1 + \tau)$, $\ldots, X(t_k + \tau)$ for all $t_1, \ldots, t_k, \tau$, $k = 1, 2, \ldots$, where τ is the lag, then $\{X(t)\}$ is said to be strictly stationary [122].

However, such a definition is too restricted to be applied in practice. Therefore, in most cases, we only require a random process to be stationary up to order 2, which means that its mean is a constant independent of t and its autocovariance function depends only on the lag τ, i.e.,

$$\mathbb{E}\big[X(t)\big] = \mu_X < \infty \tag{4.4}$$

and

$$\mathrm{cov}\big[X(t), X(t+\tau)\big] = \mathbb{E}\big[\big(X(t) - \mu_X\big)\big(X(t+\tau) - \mu_X\big)\big] = R_{XX}(\tau) - \mu_X^2 \tag{4.5}$$

where $R_{XX}(\tau)$ is known as the autocorrelation function. When $\tau = 0$, $\mathrm{cov}\big[X(t), X(t)\big] = \sigma_X^2 < \infty$, where σ_X^2 denotes the variance of the random process, such that $\{X(t)\}$ is said to be second-order stationary or weakly stationary [122].

Generally, it is not difficult to see that weakly stationary does not imply strictly stationary, except for the Gaussian process. It is because, for a Gaussian process, the multivariate distribution is completely characterised by its first and second moments. However, strictly stationary does not imply weakly stationary as well. A typical example is an independent and identically distributed (i.i.d) Cauchy process. This process is certainly strictly stationary, but both its first and second moments do not exist, and no joint moments exist up to any order [121]. Note that in this chapter, without particular mentioning, when we say stationary, we mean weakly stationary.

Although processes arise in practice are all real-valued, sometimes it is convenient to define a complex variable. Suppose $\{X(t)\}$ is a complex-valued process of the form $X(t) = X_R(t) + \mathrm{i} X_I(t)$, where $\{X_R(t)\}$ and $\{X_I(t)\}$ are both real-valued processes. The subscription R and I suggests the real and imaginary components of a random variable.

Since the $X(t)$ is the combination of the two real random processes, the statistical properties of it are determined by the joint probability density function (PDF) of $X_R(t)$ and $X_I(t)$. Generally, $\{X(t)\}$

is said to be strictly stationary if both $\{X_R(t)\}$ and $\{X_I(t)\}$ are strictly stationary [121]. Whereas if $\{X(t)\}$ is weakly stationary or stationary up to order 2. Then the mean of $\{X(t)\}$, which is given by

$$\mathbb{E}\big[X(t)\big] = \mathbb{E}\big[X_R(t)\big] + i\mathbb{E}\big[X_I(t)\big] = \mu_X < \infty \tag{4.6}$$

and its autocovariance function

$$\text{cov}\big[X(t), X(t+\tau)\big] = \mathbb{E}\Big[\big(X^*(t) - \mu_X^*\big)\big(X(t+\tau) - \mu_X\big)\Big] = R_{XX}(\tau) - \mu_X^*\mu_X \tag{4.7}$$

depends only on the lag τ. Where * denotes the complex conjugate.

4.2.1.2 Power spectral density function of a stationary process

Now, assume we have a well-behaved deterministic function $f(t)$ with finite number of discontinuity and bounded variation, a sufficient condition for the existence of Fourier transform (or integral) of $f(t)$ is that it must be absolutely integrable over $(-\infty, \infty)$, i.e., $\int_{-\infty}^{\infty} |f(t)| dt < \infty$. On the other hand, consider a single realisation, $x(t)$, which is a well-behaved non-periodic function, of a continuous parameter stationary process $\{X(t)\}$. It is obvious that the very nature of the stationary process has prohibited $x(t)$ to possess a valid Fourier integral, as the process needs infinite energy to keep sustainable over $(-\infty, \infty)$.

In other words, the process will almost certainly not decay at infinity, and the absolutely integrable condition for the existence of the Fourier integral of a reasonably behaved function cannot be satisfied. Since $x(t)$ is aperiodic, we are unable to write $x(t)$ as a Fourier series as well.

Now it seems that we cannot examine a stationary process in the frequency domain, as we are unable to decompose the process on the Fourier basis formed by the sinusoidal functions. However, the problem can be nicely resolved by evaluating the power rather than the energy of a stationary random process, as it is often the case that the power (i.e., energy per unit time) contributed by the various frequency component will be finite [121]. Correspondingly, the power spectral density of a stationary random process will converge to a finite limit when $t \to \infty$.

To be more specific, for a particular realisation of a stationary random process defined over $(-\infty, \infty)$, we can define a windowed or truncated process in the interval $[-T, T]$ but is 0 outside this interval,

$$x_T(t) = \begin{cases} x(t) & t \in [-T, T] \\ 0 & \text{otherwise} \end{cases} \tag{4.8}$$

Clearly, this truncated process $x_T(t)$ is absolutely integrable over $(-\infty, \infty)$, and it has finite energy. Thus, we can express it as a Fourier integral. Then we can write out the average power of this truncated process over $[-T, T]$ [123], that is

$$P_{x_T} = \frac{1}{2T} \int_{-T}^{T} x_T^2(t) \, \mathrm{d}t = \frac{1}{2T} \int_{-\infty}^{\infty} x_T^2(t) \, \mathrm{d}t \tag{4.9}$$

where $\int_{-T}^{T} x_T^2(t) \, \mathrm{d}t$ gives the energy of the process in $[-T, T]$.

According to Parseval's theorem [123], P_{x_T} can be expressed by the integration of the energy spectral density, namely

$$P_{x_T} = \frac{1}{2T} \int_{-\infty}^{\infty} \left| \tilde{x}_T(f) \right|^2 \, \mathrm{d}f \tag{4.10}$$

where $\tilde{x}_T(f)$ denotes the Fourier transform of $x_T(t)$. When $T \to \infty$, we obtain that

$$P_x = \lim_{T \to \infty} \frac{1}{2T} \int_{-\infty}^{\infty} \left| \tilde{x}_T(f) \right|^2 \, \mathrm{d}f = \int_{-\infty}^{\infty} G_{xx}(f) \, \mathrm{d}f \tag{4.11}$$

where $G_{xx}(f) = \lim_{T \to \infty} \left(\left| \tilde{x}_T(f) \right|^2 / 2T \right)$ is known as the power spectral density (PSD) function or simply auto-spectrum of $x(t)$, and we require it to be finite for all f.

However, since $x(t)$ is a particular realisation of $\{X(t)\}$, $G_{xx}(f)$ is a random variable, so does P_x. Thus, we shall take the expectation value of $G_{xx}(f)$, which leads us to the finial expression for the PSD function for a random process with different realisations, namely [123]

$$G_{XX}(f) = \lim_{T \to \infty} \frac{1}{2T} \mathbb{E}\left[\left| \tilde{X}_T(f) \right|^2 \right] = \lim_{T \to \infty} \frac{1}{2T} \mathbb{E}\left[\tilde{X}_T^*(f) \tilde{X}_T(f) \right] \tag{4.12}$$

where $\tilde{X}_T(f)$ is the Fourier transform of a generic truncated random process $X_T(t)$. Note that for a single random process, $G_{XX}(f)$ is real and nonnegative. It is also an even function if random process is real. In other words, when $X_T(t)$ is real valued, we have $\tilde{X}_T(-f) = \tilde{X}_T^*(f)$.

4.2.1.3 The Wiener-Khintchine-Einstein theorem

Now that we have the expression for the PSD function of a random process, but to calculate $G_{XX}(f)$ directly from limitation expression given by Eq. (4.12) is formidable. Nevertheless, the Wiener-Khintchine-Einstein theorem provides us a solution for this problem.

Let $\{X(t)\}$ be a stationary process with zero mean, followed by our aforementioned discussion, its autocorrelation function is given by $R_{XX}(\tau) = \mathbb{E}\left[X(t)X(t+\tau)\right]$. The Wiener-Khintchine-Einstein theorem [123] says that if $R_{XX}(\tau)$ is absolutely integrable, then its possesses a Fourier transform, which is the PSD function $G_{XX}(f)$ we are looking for. In other words, the autocorrelation function and the spectral density are Fourier transforms of each other, namely,

$$R_{XX}(\tau) = \int_{-\infty}^{\infty} G_{XX}(f) e^{i2\pi f\tau} df \quad \text{and} \quad G_{XX}(f) = \int_{-\infty}^{\infty} R_{XX}(\tau) e^{-i2\pi f\tau} d\tau \qquad (4.13)$$

or

$$R_{XX}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G_{XX}(\omega) e^{i\omega\tau} d\omega \quad \text{and} \quad G_{XX}(\omega) = \int_{-\infty}^{\infty} R_{XX}(\tau) e^{-i\omega\tau} d\tau \qquad (4.14)$$

where Eq. (4.14) is expressed in terms of the radial frequency variable $\omega = 2\pi f$.

Similarly, we can obtain a Fourier transform pair between the cross-covariance function $R_{XY}(\tau)$ and the cross spectral density function (or cross-spectrum) $G_{XY}(f)$ of two real-valued random process $\{X(t)\}$ and $\{Y(t)\}$,

$$R_{XY}(\tau) = \int_{-\infty}^{\infty} G_{XY}(f) e^{i2\pi f\tau} df \quad \text{and} \quad G_{XY}(f) = \int_{-\infty}^{\infty} R_{XY}(\tau) e^{-i2\pi f\tau} d\tau \qquad (4.15)$$

Note that in general $R_{XY}(\tau) = R_{YX}(-\tau)$. Due to this property of the cross-covariance function, the cross-spectral density function is Hermitian symmetric, i.e., $G_{XY}(f) = G_{YX}(-f) = G_{XY}^{*}(-f)$ [123].

4.2.1.4 The time-averaged spectral density function

From the last subsection, we know that the spectral density function of a stationary random process is defined as the limitation of the expected average power. However, the limiting operation $T \to \infty$ can never be performed, as the record length T will always be finite in practice. Apart from this, an infinite ensemble is impossible to be obtained for real data. As a result, the expectation operation $\mathbb{E}[\cdot]$ will always be taken over only a finite number of ensemble elements. For this reason, it is almost impossible

to compute the PSD function via Eq. (4.12). Nevertheless, The Wiener-Khintchine-Einstein theorem provides us an alternative way to compute the PSD function by taking the Fourier transform of the correlation function.

To use the Wiener-Khintchine-Einstein theorem to obtain the PSD function of a random process, we first need to evaluate $R(\tau)$, which is also given by an ensemble average. Unfortunately, we do not normally have access to the ensemble in reality. In many cases, there are only one or a few sample records of data with limited length are available to us. So again, we are in trouble. However, for a stationary random process, it is possible to describe its properties by computing the time averages over specific sample functions in the ensemble [124]. In other words, we can invoke the ergodicity property of a stationary random process by using the values taken by a typical sample over time to represent the values taken across the ensemble [125].

It seems that the best way to obtain the PSD function for a stationary process is through analysis of the correlation function. However, as we mentioned above, in practice, both the ensemble and $T \rightarrow \infty$ are unattainable, and the measurements are often digitalised. In this case, we will have more options to approximate a PSD function either by using a correlation function-based technique (i.e., correlogram, based on the Wiener-Khintchine-Einstein theorem) or by using a direct method (i.e., Periodogram, based on Eq. (4.12)) [126]. In this work, we will adopt the latter method, as it can help us to define a linear operator we need to understand the FDD and the density matrix.

Now let us assume that $\{X(n)\}$ is a real-valued discrete-time random process with N observations, then the limiting operation in Eq. (4.12) can be omitted. Since the random process is discrete, the auto-spectrum for a finite sequence will be defined only at selected discrete frequencies. Thus, at any such frequencies f_k, where $k = 0, 1, \ldots, N-1$, the auto-spectrum is given by

$$\bar{G}_{XX}(f_k) = \frac{1}{N} \mathbb{E}\left[\tilde{X}^*(f_k) \tilde{X}(f_k) \right] \tag{4.16}$$

While the cross-spectrum of two finite-length process $\{X(n)\}$ and $\{Y(n)\}$ with N sample points, we have

$$\bar{G}_{XY}(f_k) = \frac{1}{N} \mathbb{E}\left[\tilde{X}^*(f_k) \tilde{Y}(f_k) \right] \tag{4.17}$$

where $\bar{G}_{XX}(f_k)$ is real-valued, whereas $\bar{G}_{XY}(f_k)$ Hermitian symmetric. $\tilde{X}(f_k)$ and $\tilde{Y}(f_k)$ are the Discrete Fourier Transform (DFT) of $X(n)$ and $Y(n)$, respectively, and they are given by

$$\tilde{X}(f_k) = \sum_{n=0}^{N-1} X(n) e^{-i2\pi f_k n} \quad \text{and} \quad \tilde{Y}(f_k) = \sum_{n=0}^{N-1} Y(n) e^{-i2\pi f_k n} \tag{4.18}$$

where the frequency $f_k = k/N$ are the Fourier frequencies. For the inverse DFT we have

$$X(n) = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{X}(f_k) e^{i2\pi f_k n} \quad \text{and} \quad Y(n) = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{Y}(f_k) e^{i2\pi f_k n} \tag{4.19}$$

With the PSD expression in Eq. (4.16), we can deem it as the time-averaged mean square value of a complex-valued random process $\{\tilde{X}(f_k)\}$. More precisely, $\overline{G}_{XX}(f_k)$ gives the time-averaged auto spectral density function when the double-frequency auto spectra is evaluate along the f axis [124]. Thus, to distinguish the PSD function for an unbounded random process, we will refer to the name "time-averaged" PSD function or spectrum rather than the PSD function and put a bar on top of $G_{XX}(f_k)$ to mean the "time-averaged". Accordingly, Eq. (4.17) gives the time-averaged cross-spectrum formed by two complex-valued processes $\{\tilde{X}(f_k)\}$ and $\{\tilde{Y}(f_k)\}$.

4.2.1.5 Nonparametric PSD estimator

Since we only have a limited number of records in reality, we must consider how to remove the expectation notation in Eq. (4.16) and Eq. (4.17). If we remove the expectation notation in Eq. (4.16) directly, then the time-averaged the auto-spectrum $\overline{G}_{XX}(f_k)$ gives the Periodogram of the finite-length process $\{X(n)\}$. Note that discussion in this section will be focusing on the auto-spectrum, for the cross-spectrum, but the reasoning applies to the cross-spectrum.

For one of the realisations of the random process, after removing the expectation notation, we have

$$\hat{G}_{XX}(f_k) = \frac{1}{N} \tilde{x}^*(f_k) \tilde{x}(f_k) \tag{4.20}$$

where a hat is put on top of $G_{XX}(f_k)$ to denote an estimator. Although as a spectral estimator the Periodogram is asymptotically unbiased, it is inconsistent with large variance, which made it a very poor approximation for the true PSD [126].

As we know that if the stationary process is Gaussian, then its Fourier transform is Gaussian as well by linearity. However, the probability distribution of Periodogram values is far from Gaussian. For a Gaussian process, the Periodogram estimates, defined at a set of frequencies, each of them follows a chi-squared distribution with two degrees of freedom, i.e., χ_2^2. Its variance does not tend to 0 as the data length increases. It is to say that, even if the length of the time series increases, the Periodogram

does not become smoother accordingly. As a result, there will be more pikes packed closer together, but the precision of each estimate does not change. This is because the Periodogram is an inconsistent spectral estimator that even if the length of the processed sample increases without bound, the Periodogram continuous fluctuate around the true PSD with a nonzero variance [126]. This means that the Periodogram is a very poor approximation.

However, the variance problem can be alleviated to some extent by smoothing of some kind but at the expense of increasing its bias and, hence, decreasing the average resolution. In order to cure the aforementioned difficulties of the basic Periodogram approach, the simplest way to do is to approximate the expectation in Eq. (4.16) with an average on finite records.

Suppose there are n_d different sub-records, each of length M, then a smoothed PSD estimator can be given by

$$\hat{G}_{XX}\left(f_k\right) = \frac{1}{Mn_d}\sum_{i=1}^{n_d}\tilde{x}_i^*\left(f_k\right)\tilde{x}_i\left(f_k\right) \tag{4.21}$$

In practice, if a measurement record is of length N, we can divide it into n_d segments, each of length M, then $n_d = N/M$. Hence, if there are n_d subsamples at frequency f_k, the resulting estimates will be distributed like $\chi^2_{2n_d}$, i.e., χ^2 with $2n_d$ degrees of freedom. As well as having smaller variance, the estimate will approximate a Gaussian distribution due to the fact that, as n_d increases, the $\chi^2_{2n_d}$ distribution tends to normality. For a reasonable approximation, around 50 such observations are required if there is no overlap [127]. It is worth mentioning that Eq. (4.21) is often known Bartlett method if a N points sequence is partitioned into n_d subsamples without overlapping between the successive segments. Normally, the Bartlett estimator can reduce the variance of the original Periodogram by the factor of n_d.

To further reduce the statistical variability of the estimated spectrum, we can overlap the segments and window each segment prior to computing the Periodogram. And this is the basic idea of the Welch method [126]. Let $x_i\left(n\right)$ be the ith data segment, $i = 1, \ldots, n_d$, then define

$$x_i\left(n\right) = x\left(\left(i-1\right)K + n\right), \quad n = 0, \ldots, M-1 \tag{4.22}$$

where $\left(i-1\right)K$ is the starting point for the ith sequence of observations. When $K = M$, we get the sample splitting used by the Bartlett method, which means there is no overlap. Typically, $K = M/2$ (50% overlap) is recommended by the Welch method.

As for the window, it is applied to each segment directly such that the DFT of the windowed sequence for each segment is given by

$$\tilde{x}_i(f_k) = \sum_{n=0}^{M-1} w_p(n) x_i(n) e^{-i2\pi f_k n} \tag{4.23}$$

where $\{w_p(n)\}$ denote the M points window sequence, the subscription p means "Periodogram". For a windowed Periodogram, it is called a temporal window or a taper. Substitute Eq. (4.23) into Eq. (4.21) and divided by the power of the temporal window, we obtain the Welch estimate of PSD, i.e., a windowed Periodogram.

$$\hat{G}_{XX}(f_k) = \frac{1}{M n_d P_w} \sum_{i=1}^{n_d} \tilde{x}_i^*(f_k) \tilde{x}_i(f_k) \tag{4.24}$$

where P_w is the power of the window, and it is given by

$$P_w = \frac{1}{M} \sum_{n=0}^{M-1} |w_p(n)|^2 \tag{4.25}$$

Compared to the Bartlett method, empirical evidence showed that the Welch method could offer lower variance, but the difference in the variances corresponding to the two methods is not significant [126].

Both the Bartlett method and the Welch method are modified or refined Periodogram-based methods. Apart from these two methods, the Daniell method [128], which averages Periodogram values locally around the frequency of interest, is also well-known. Although these methods come in different forms, they are all variations on the same theme and can all be related to the Blackman-Tukey method [129].

Unlike the refined Periodogram-based method, the Blackman-Tukey method is implemented by applying a lag window to truncate the estimated correlation sequence before taking the Fourier transform of it. It was inspired by the intuition that the poor statistical quality of the Periodogram PSD estimator was caused by the poor accuracy of the estimated correlation function in the Correlogram for extreme lags and the large number of correlation estimation errors that are cumulatively summed up in the Correlogram [126]. Let $\{w_c(n)\}$ denote the lag window sequence, where c means "Correlogram", then the windowed Periodogram and the windowed Correlogram have the same average behaviour if the temporal and lag window are related in the following way

$$w_c(n) = \frac{1}{M} \sum_{k=0}^{M-1} w_p(k) w_p^*(k-n) \tag{4.26}$$

The Fourier transform of $w_c(n)$ is normally known as the spectral window, which convolutes with the Periodogram to give the Blackman-Tukey spectral estimator. Although the Periodogram is nonnegative,

it should be pointed out that the Blackman-Tukey spectral estimator may take negative values if $\{w_c(n)\}$ is a rectangular window [126]. In order to obtain a nonnegative Blackman-Tukey spectral estimator, the lag window needs to be positive semidefinite, i.e., and its Fourier transform should be nonnegative.

Typically, the modified Periodogram-based methods reduce the variance of the estimated spectrum at the expense of some reduction in average resolution (hence, increasing bias). However, the unmodified Periodogram can be a satisfactory estimator for discrete spectra corresponding to sinusoidal signals, especially when no two spectral lines in the spectrum are separated by less than $1/N$ provided that the sequence has N observation points [126].

The reason is that the unmodified Periodogram is equivalent to a Blackman-Tukey windowed estimator with a rectangular window of maximum length equal to $2N+1$. And of all window functions, the rectangular window has the narrowest main lobe, which means that it can afford maximum spectral resolution and bring the least smearing. However, because of its large number of nonzero side lobes, it will bring the most significant leakage. Theoretically, the Bartlett estimate is similar to the Blackman-Tukey estimate with a rectangular lag window of length M, which means that compared to other modified Periodogram estimates, the Bartlett estimator is expected to have the least smearing but the most significant leakage [126]. So, one should note that it is always a good idea to consider spectra with different amounts of smoothing when estimating a continuous PSD, but excessive smoothing will blur the features we are looking for.

Although it is an unusual luxury to have many independent observations of the same random process in practice, we can have such luxury to generate as many independent observations as we want in a simulation. Therefore, in our first case study in section 4.3.2, we will adopt the core idea of the Bartlett method to generate a Periodogram estimator for a multivariate process by averaging the power spectrum at each frequency line because the components that need to be identified are sinusoids. While in our analysis of the close modes in section 4.4.2, the Welch method will be adopted to smooth the PSD from a stationary random process.

4.2.1.6 Matrix notation of the PSD estimator for a multivariate process

Now suppose we have d channel measurements of the real and zero-mean stationary random process, each of which has N sample points. If we divide our records into n_d segments, each of length M, then apply an appropriate window function and take the Fourier transform of them, we can obtain a three dimensional data matrix $\tilde{\mathbb{X}}$ with size $d \times n_d \times M$, namely,

$$\tilde{\mathbb{X}} = \begin{bmatrix} \tilde{\mathbf{X}}(f_0) & \tilde{\mathbf{X}}(f_1) & \cdots & \tilde{\mathbf{X}}(f_{M-1}) \end{bmatrix} \tag{4.27}$$

where each of $\tilde{\mathbf{X}}(f_k)$, $k = 0, 1, \ldots, M-1$, is a $d \times n_d$ complex-valued matrix, and it is given by

$$\tilde{\mathbf{X}}(f_k) = \begin{bmatrix} \tilde{\mathbf{y}}_1(f_k) & \cdots & \tilde{\mathbf{y}}_{n_d}(f_k) \end{bmatrix} \tag{4.28}$$

where $\tilde{\mathbf{y}}_i(f_k)$ are the $d \times 1$ vectors. Note that we used $\mathbf{y}$ instead of $\mathbf{x}$ to address the formulation of $\tilde{\mathbb{X}}$ as a three dimensional matrix.

According to the PSD estimator given by Eq. (4.24) with n_d subrecords at each frequency line, we can obtain the PSD estimator for a multivariate process as follows

$$\hat{\mathbf{G}}(f_k) = \frac{1}{Mn_d} \sum_{i=1}^{n_d} \tilde{\mathbf{y}}_i^*(f_k) \tilde{\mathbf{y}}_i^{\mathrm{T}}(f_k) = \frac{1}{Mn_d} \tilde{\mathbf{X}}^*(f_k) \tilde{\mathbf{X}}^{\mathrm{T}}(f_k), \quad k = 0, 1, \ldots, M-1 \tag{4.29}$$

where $\hat{\mathbf{G}}(f_k)$ is a $d \times d$ complex-valued matrix which incorporates both the auto-spectrum (diagonal, real-valued) and the cross-spectrum estimator (off-diagonal, complex-valued). Since the cross-spectrums are Hermitian symmetric, then $\hat{\mathbf{G}}(f_k)$ is a Hermitian matrix with $\hat{\mathbf{G}}^*(f_k) = \hat{\mathbf{G}}^{\mathrm{T}}(f_k)$ or $\hat{\mathbf{G}}(f_k) = \hat{\mathbf{G}}^{\mathrm{H}}(f_k)$, and it is non-negative definite, where the superscription $\cdot^{\mathrm{H}}$ denotes Hermitian conjugate or conjugate transpose, i.e., $\cdot^{*\mathrm{T}}$. Besides, it is not difficult to see that $\hat{\mathbf{G}}(f_k)$ is a M scaled complex-valued empirical or sample covariance matrix (or autocorrelation matrix when the mean is zero).

In practice, we use Eq. (4.29) to compute $\hat{\mathbf{G}}(f_k)$, but in our theoretical discussion we can also use the following more generic expression, namely

$$\bar{\mathbf{G}}(f_k) = \frac{1}{M} \mathbb{E}\left[\tilde{\mathbf{X}}^*(f_k) \tilde{\mathbf{X}}^{\mathrm{T}}(f_k) \right] \tag{4.30}$$

where $\bar{\mathbf{G}}(f_k)$ is a time-averaged spectrum, and it is also a M scaled covariance matrix that the M scaled empirical covariance matrix $\hat{\mathbf{G}}(f_k)$ wants to approximate.

Particularly, if we consider the real-valued spectral density matrix we introduced in chapter 3, $\tilde{\mathbf{X}}(f_k)$ for $k = 1, \ldots, M/2 - 1$, can be effectively given by

$$\tilde{\mathbf{X}}(f_k) = \begin{bmatrix} \tilde{\mathbf{y}}_1(f_k) & \tilde{\mathbf{y}}_1^*(f_k) & \cdots & \tilde{\mathbf{y}}_{n_d}(f_k) & \tilde{\mathbf{y}}_{n_d}^*(f_k) \end{bmatrix} \tag{4.31}$$

While the at f_0 and $f_{M/2}$, $\tilde{\mathbf{X}}(f_k)$ will still be given by Eq. (4.28). When the above definition of $\tilde{\mathbf{X}}(f_k)$ is substituted into Eq. (4.29), the PSD estimator $\hat{\mathbf{G}}(f_k)$ becomes a real-valued symmetric

matrix. In this case, if the mode shapes of the system are real-valued, by using the real-valued spectral density matrix, a set of real-valued mode shapes estimations will be generated from the FDD.

Note that the rank of $\hat{\mathbf{G}}(f_k)$ depends on the number of sensors d and the number of subsamples n_d. If the real-valued spectral density matrix is not considered, the rank of $\hat{\mathbf{G}}(f_k)$ is determined by n_d. In this case, in order to obtain a full rank of the $\hat{\mathbf{G}}(f_k)$, one should make sure that at least d numbers of averages is being used, i.e. $n_d \geq d$. However, it is just a necessary condition, as we will not be able to obtain a full rank of $\hat{\mathbf{G}}(f_k)$ if the actual modes of the system is less than d. In other words, if and only if $r \geq d$, where r denotes the actual mode the system contains, we can obtain a full rank of $\hat{\mathbf{G}}(f_k)$ by averaging it for $n_d \geq d$ times at f_k.

4.2.2 Part II — Decomposition of the PSD estimator

Now that we have obtained the PSD estimator for a multivariate process, our objective in this section is to explore possible ways to interpret the information contained in a PSD estimator. Under the conventional FDD setting, the Singular Value Decomposition (SVD) technique is used to analyse the PSD estimator $\hat{\mathbf{G}}(f_k)$. However, in this chapter, we will drop this idea temporarily and approach it by using the Spectral theorem. We can do this is because at each f_k, $\hat{\mathbf{G}}(f_k)$ is an M scaled empirical covariance matrix, and it is Hermitian. For such a matrix, it has some special properties, which we will explore in this section. Therefore, our first step is to understand the mathematics behind $\hat{\mathbf{G}}(f_k)$, then build the connection between $\hat{\mathbf{G}}(f_k)$ and the modal information of the system. Eventually, our analysis will lead us to define a density operator for the close modes analysis.

Our discussion will start with the definition of some linear operators. Then we will explain what a Hilbert space is and some of its properties. Upon which, we shall present the mathematical properties of a self-adjoint operator in a Hilbert space and introduce the Spectral theorem in the meantime. We will then redefine the FDD in terms of the Principal Component Analysis (PCA) in section 4.2.2.4. Finally, we will introduce a density operator in section 4.2.2.5 and derive it from a PSD estimator in section 4.2.2.5 III. The idea of pure and mixed states is explained in section 4.2.2.5 II. Particularly, the reason that we can use the FDD and our proposed method to estimate the mode shapes in terms of a PSD estimator is given in section 4.2.2.5 III.

4.2.2.1 Linear operators and subspace operation

Let a field $\mathbb{F}$ be either $\mathbb{R}$ or $\mathbb{C}$. Suppose V and W are the vectors spaces over $\mathbb{F}$. A linear map from V to W is a function $T:V \rightarrow W$ that (1) (additivity), $\forall f,g \in V$, $T(f+g)=T(f)+T(g)$; (2) (homogeneity) $\forall \alpha \in \mathbb{F}$, and $\forall f \in V$, $T(\alpha f)=\alpha T(f)$, where $\forall$ means for all [130].

If we define the set of all linear maps from V to W is denoted by $\mathcal{L}(V,W)$, and suppose $T \in \mathcal{L}(V,W)$, then the adjoint of T is the function $T^{\dagger}:W \rightarrow V$ such that $\langle Tf,g \rangle = \langle f,T^{\dagger}g \rangle$, $\forall f \in V$ and $\forall g \in W$ [130]. Where $\langle \cdot,\cdot \rangle$ gives the inner product, † denotes the conjugate transpose.

Now if a function $T:V \rightarrow V$, $\forall f \in V$, then such a linear map is called an operator. Accordingly, the set of all operators on V is denoted as $\mathcal{L}(V)$. Particularly, when the dimensional of V is finite and $T \in \mathcal{L}(V)$, then T is invertible.

For an operator $T \in \mathcal{L}(V)$, if $T = T^{\dagger}$, then it is called self-adjoint. Followed by the definition of an adjoint operator, we have $\langle Tf,g \rangle = \langle f,Tg \rangle$, $\forall f,g \in V$. If $\langle Tf,f \rangle \geq 0, \forall f \in V$, then $T \in \mathcal{L}(V)$ is called a positive semidefinite operator. When an operator commutes with its adjoint on an inner product space (introduced in section 4.2.2.2), namely $TT^{\dagger} = T^{\dagger}T$, then it is called normal, and every self-adjoint operator is normal [130]. Furthermore, if $TT^{\dagger} = T^{\dagger}T = I$, where I is an identify operator, then T is called unitary or isometry, and it preserved the inner product, i.e., $\langle Tf,Tg \rangle = \langle f,g \rangle$, $\forall f,g \in V$.

It is worth mentioning that the Fourier transform is a unitary operator on a square integral Hilbert space, with a set of orthonormal sinusoidal functions as its basis. The Parseval's theorem we mentioned in section 2.4.1.2 is derived based on this property.

Note that a self-adjoint operator is also known a Hermitian operator. The PSD estimator $\hat{\mathbf{G}}(f_k)$ is such an operator. Note that a real and symmetric matrix is a special case of a Hermitian matrix. Thus, either we use the real-valued PSD estimator, or the complex-valued one, the mathematical theorems we presented in this section applies to both.

Now suppose a vector space V has a direct sum (denoted by $\oplus$ or $+$) decomposition, i.e.,

$$V = U_1 \oplus \cdots \oplus U_n \tag{4.32}$$

where U_i are proper subspaces of V which satisfy the following three conditions: (1) (additive identity) $0 \in U$; (closed under addition) $f,g \in U \Rightarrow f+g \in U$; (3) (closed under scalar multiplication) $\alpha \in \mathbb{F}$ and $f \in U \Rightarrow \alpha f \in U$. Note that the direct sum gives the smallest subspace of V, which

means that for each $f_i \in U_i$, the only way to write $0 = f_1 + \cdots + f_n$ is by taking each f_i equal to 0. In this case, the set of all linear combinations of a list of vectors $f_1, \ldots, f_n$ in V is called the span of $f_1, \ldots, f_n$. Therefore, with the direct sum, to understand the behaviour of T, we only need to understand the behaviour of $T\big|_{U_i}$, where $T\big|_{U_i}$ denotes the restriction of T to the smaller domain U_i. [130].

Besides, when T is an operator on U, where U is a subspace of V, and T maps each U into itself, then a subspace U of V is called invariant under T if $f \in U_i$ implies $Tf \in U$ [130]. A good example is given by the eigenvalues and the eigenvectors. It says that if U is invariant under an operator $T \in \mathcal{L}(V)$, then followed by its definition, we have $Tf \in U$, and hence there is a scalar $\lambda \in \mathbb{F}$, such that $Tf = \lambda f$, where λ is called the eigenvalue of T. Conversely, if $Tf = \lambda f$ for some $\lambda \in \mathbb{F}$, then the span of f is a one-dimensional subspace of V invariant under T [130].

4.2.2.2 Hilbert spaces

A metric or distance is a function which assigns a real number $d(f, g)$ to every two elements $f, g \in V$. While a metric space is an order pair (V, d), where V is a nonempty set and d is a metric on V, i.e., $d : V \times V \to \mathbb{R}$, such that the following holds: (1) (non-negativeness) $d(f, g) \geq 0$, $\forall f, g \in V$; (2) (identification) $\forall f, g \in V$ we have $d(f, g) = 0 \Leftrightarrow f = g$; (3) (symmetry) $\forall f, g \in V$, $d(f, g) = d(g, f)$; (4) (triangular inequality) $\forall f, g, h \in V$, $d(f, h) \leq d(f, g) + d(g, h)$ [131].

In general, one could define lots of distances. For instance, given two vectors in $x = (x_1, \ldots, x_n)^{\mathrm{T}}$ and $y = (y_1, \ldots, y_n)^{\mathrm{T}}$ in $\mathbb{F}^n$, we can define $d_1(x, y) = \sum_{i=1}^{n} |x_i - y_i|$, $d_2(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$ or $d_\infty(x, y) = \max_{i \in \{1, \ldots, n\}} |x_i - y_i|$, where $n \in \mathbb{Z}^+$, $d : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{R}$, and $d_2(x, y)$ is well-known as the Euclidean distance. Note that the subscriptions were used to distinguish between different metric on a given space. Among all the distances, we are particularly interested in a subset of functions called norms.

A norm is a real-valued function defined on the vector space, a normed vector space or a normed space V on $\mathbb{F}$ is a function mapping V to the non-negative real numbers, namely, $\|\cdot\| : V \to \mathbb{R}$ such that for $f \in V$ has the following properties: (1) (positive definite) $\|f\| \geq 0$, $\forall f \in V$ and $\|f\| = 0$ if and only if $f = 0$; (2) (homogeneity) $\|\alpha f\| = |\alpha| \|f\|$, $\forall f \in V$ and $\forall \alpha \in \mathbb{F}$, where $|\cdot|$ denotes the absolute value; (3) (triangle inequality) $\|f + g\| \leq \|f\| + \|g\|$, $\forall f, g \in V$.

We can verify based on the properties of a metric space that a normed vector space is a metric space indeed. Thus, we can define $d : V \times V \to \mathbb{R}$ by $d(f, g) = \|f - g\|$, where $d(f, g) \geq 0$. So, followed by our previous example, we can denote them as $d_1(x, y) = \|x - y\|_1$, $d_2(x, y) = \|x - y\|_2$, and $d_\infty(x, y) = \|x - y\|_\infty$, for any given points in $x, y \in \mathbb{F}^n$.

Now suppose $(E, \mathcal{U}, \mu)$ is a measure space, given a finite number $1 \leq p < \infty$ (here we only want to consider the finite case, but in general $p = \infty$ is allowed), the notation $L^p(\mu)$ is defined to the set of $\mathcal{U}$-measurable function $f : E \to \mathbb{F}$, such that the p-norm of f is denoted by $\|f\|_p = < \infty$, where $\|f\|_p$ is defined by

$$\|f\|_p = \left(\int |f|^p \, d\mu \right)^{1/p} \tag{4.33}$$

When μ is a counting measure on $\mathbb{Z}^+$, $L^p(\mu)$ is often denoted by ℓ^p, then a sequence of scalars $x = (x_1, x_2, \ldots)$ in $\mathbb{F}$, Eq. (4.33) becomes

$$\|x\|_p = \left(\sum_{i=1}^{\infty} |x_i|^p \right)^{1/p} \tag{4.34}$$

And it is said p-summable if $\|x\|_p < \infty$. When $p = 1$ it is said "absolutely summable". When $p = 2$, we have a "square summable" sequence.

Now comes a critical question that if our space is complete or not. And the answer lies in the convergence of a Cauchy sequence.

Let $f_1, f_2, \ldots$ be a sequence defined on a metric space (V, d), it is said to be Cauchy sequence if for every $\varepsilon > 0$, there exists an integer $n \in \mathbb{Z}^+$, such that $i, j \geq n \implies d(f_i, f_j) < \varepsilon$. This definition says that the distance between f_i and f_j decreases to zero as i and j increase. Particularly, according to the triangular inequality property of a metric space, we can prove that every convergent sequence is a Cauchy sequence, and all Cauchy sequence in $\mathbb{F}$ converge. However, a Cauchy sequence need not be convergent, for instance, if f is irrational in $\mathbb{Q}$ [131].

For a metric space V, if every Cauchy sequence in V converges to some element of V, then V is called complete, and a complete normed space is called a Banach space in which every Cauchy sequence converges. As for the Hilbert space, it is defined to be an inner product space that is a Banach space with its norm determined by the inner product [118].

For vectors in $\mathbb{R}^n$ or $\mathbb{C}^n$, we know how to measure the angle between vectors by computing the dot product, while the inner product is a generalisation of the dot product, which can provides us with a generalisation of the notion of orthogonality to a much wider range of spaces than just $\mathbb{R}^n$ or $\mathbb{C}^n$ [131]. By definition, an inner product on a vector space V is real-valued function $\langle \cdot, \cdot \rangle$ on $V \times V$ such that satisfying the following properties: (1) (non-negativity) $\langle f, f \rangle \geq 0$, $\forall f \in V$; (2) (Uniqueness) $\langle f, f \rangle = 0$ if and only if $f = 0$; (3) (conjugate symmetry) $\langle f, g \rangle = \langle g, f \rangle^*$, $\forall f, g \in V$; (4) (linearity in the first entry) $\langle f + g, h \rangle = \langle f, h \rangle + \langle g, h \rangle$, $\forall f, g, h \in V$; (5) (homogeneity in the first entry) $\langle \alpha f, g \rangle = \alpha \langle f, g \rangle$, $\forall f, g \in V$ and $\forall \alpha \in \mathbb{F}$. Note that the linearity and the homogeneity can also be defined in the second entry. When $\langle f, g \rangle = \langle g, f \rangle = 0$, $\forall f, g \in V$, then we say f is orthogonal to g.

By definition, each inner product induces a norm, so every vector space that has an inner product is a normed space. So a norm associated with an inner product is denoted by $\|f\| = \sqrt{\langle f, f \rangle}$, for $f \in V$. However, if we compare the definitions of a norm and an inner product, it is not difficult to see that not all norms are induced from inner products, for example, out all ℓ^p spaces, only ℓ^2 is a Hilbert space [131]. Therefore, a Hilbert space, is a vector space whose norm is induced from an inner product and it is complete with respect to this norm, is a special type of Banach space. Particularly, in the finite dimensional case, the completeness always holds.

4.2.2.3 Spectral theorem

Defined upon the inner product spaces, the Hilbert spaces have some nice properties, for instance, every Hilbert space has an orthonormal basis. When associated with a compact and self-adjoint operator, important conclusions can be drawn to help us analyse the PSD estimator of our interest. And the Spectral theorem which gives us a fundamental decomposition for self-adjoint operators on a Hilbert space will be the conclusion we want to achieve in this subsection.

Every vector space has a basis. Usually the notation $\{e_i\}_{i \in \Gamma}$ is used to express the notion of a basis of a vector space V, and we call it a family in V. We can think of it as a function e from a set Γ to V, with the value of the function e at $i \in \Gamma$ denoted by e_i, where Γ can be given by $\{1, \ldots, n\}$ or any arbitrary set that might not be a finite set.

Usually, we want each element in the family $\{e_i\}_{i \in \Gamma}$ to be linearly independent, and it is linearly independent if there does not exist a finite nonempty set J of Γ and a family $\{a_j\}_{j \in J}$ in V such that

$\sum_{j\in J} a_j e_j = 0$, where $\mathbb{F}\backslash\{0\}$ means the complement of the set $\{0\}$. In other words, the only choice of $\{a_j\}_{j\in J}$ that makes $\sum_{j\in J} a_j e_j = 0$ is that all the elements in $\{a_j\}_{j\in J}$ equal zero. And a vector space is said to be finite-dimensional if there exists a finite set Γ and a family $\{e_i\}_{i\in\Gamma}$ in V such that the span of $\{e_i\}_{i\in\Gamma}$, which denoted as $\{e_i\}_{i\in\Gamma}$, is equal to V. For a family $\{e_i\}_{i\in\Gamma}$ to be the basis of V, it needs to be linearly independent and its $\text{span}\{e_i\}_{i\in\Gamma} = V$ [118]. In this work, we will only consider a finite-dimensional space.

A family $\{e_i\}_{i\in\Gamma}$ is called an orthonormal family if $\langle e_i, e_j \rangle = 1$ only if $i = j$, otherwise $\langle e_i, e_j \rangle = 0$ for $i \neq j$. In a Hilbert space, an orthonormal family $\{e_i\}_{i\in\Gamma}$ is called an orthonormal basis of V if $\text{span}\{e_i\}_{i\in\Gamma}^* = V$, and every Hilbert space has an orthonormal basis. A special case is when e_i is an element of ℓ^2, $i \in \mathbb{Z}^+$, all other components are zero except for the ith component which is 1, i.e., $e_i = (0,\ldots,0,1,0,\ldots)$, then we call e_i the ith standard basis vector. Accordingly, we call the family $\{e_i\}_{i\in\mathbb{Z}^+}$ the sequence of standard basis vector or standard basis [131].

An important topic in operator theory is the spectrum of an operator. Suppose T is a bounded operator on a finite-dimensional Banach space V, if $T - \lambda I$ is not injective (implies not invertible), then $\lambda \in \mathbb{F}$ is called an eigenvalue of T, then the nonzero vector $e \in V$ is called an eigenvector of T corresponding to an eigenvalue of $\lambda \in \mathbb{F}$ if $Te = \lambda e$. Note that a linear operator is called bounded if $\|T\| < \infty$, the Banach space of bounded linear operator from V to V is denoted as $\mathcal{B}(V)$. Besides, it is worth mentioning that when dealing with an finite-dimensional Banach space, the spectrums and eigenvalues are equal, but it may not be the case for the infinite-dimensional case [118]. But for this study, we will use them interchangeably since we only care about the finite case.

Now, let V be a finite-dimensional Hilbert space and $T \in \mathcal{B}(V)$, then by choosing a basis for V, we can identify T with an $n \times n$ matrix. When $T = T^\dagger$, then it is self-adjoint, then all eigenvalue of T are real, and eigenvectors of T corresponding distinct eigenvalues are orthogonal [131]. The nicest operator on V are those for which there is an orthonormal basis of V consisting of eigenvectors of T with respect to which the operator has a diagonal matrix [118]. And such an operator should not only be self-adjoint but also be compact. By compact, we mean if for every sequence $f_1, f_2, \ldots$ in a Hilbert space V, the sequence $Tf_1, Tf_2, \ldots$ has a convergent subsequence, then the operator T is said to be compact, and the collect of compact operators on V is denoted by $\mathcal{C}(V)$. Every compact operator on a Hilbert space is a bounded operator [118].

Now we are in place to introduce the Spectral theorem for self-adjoint compact operators. Let T be a self-adjoint compact operator on a Hilbert space V, then there is an orthonormal basis of T consisting of eigenvectors of T, and there is a countable set J, an orthonormal family $\{e_i\}_{i \in J}$ in V, and a family $\{\lambda_j\}_{j \in J}$ in $\mathbb{R} \setminus \{0\}$, for every $f \in V$ such that [118]

$$Tf = \sum_{j \in J} \lambda_j \langle f, e_j \rangle e_j \tag{4.35}$$

where

$$f = \sum_{j \in J} \langle f, e_j \rangle e_j \tag{4.36}$$

Here, the orthonormal family $\{e_i\}_{i \in J}$ corresponds to $\{\lambda_j\}_{j \in J}$ along with the 0 vector, is often known as the eigenspaces of T and usually denoted by $E(\lambda_j, T)$. Besides, from the Spectral theorem, we know that for a self-adjoint compact operator on a nonzero Hilbert space, it has at least one eigenvalue [118], and even with repeated eigenvalues, we are assured that we will have a complete set of orthonormal eigenvectors to span the space [132]. In other words, the sum of eigenspaces of T is a direct sum, i.e.,

$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_n, T) \tag{4.37}$$

and each $E(\lambda_j, T)$ is a one-dimensional subspace of V.

Note that when the V is a complex Hilbert space, and $T \in \mathcal{B}(V)$, then T is self-adjoint $\Leftrightarrow$ $\langle Tf, f \rangle \in \mathbb{R}$, $\forall f \in V$. Furthermore, on a complex Hilbert space, T is said to be positive or positive semidefinite $\Leftrightarrow \langle Tf, f \rangle \geq 0$ for every $f \in V$, and T is called positive definite $\Leftrightarrow \langle Tf, f \rangle > 0$ for every $f \neq 0$ [131]. When T is self-adjoint and positive, T not only has non-negative eigenvalues, but there also exists an operator $R \in \mathcal{B}(V)$ such that $T = R^* R$.

Besides, before moving to the next subsection, it is worth defining the orthogonal projection. Let U be a closed subspace on a Hilbert space V. Given $f \in V$, the unique vector $g \in U$ that is closest to f is called the orthogonal projection of f onto U [131]. Then the orthogonal projection of V onto U is the function $P_U : V \to V$ defined by $P_U f = g$ for $f \in V$ and $g \in U$. Note that for an orthogonal projection operator we have $P_U^2 = P_U$, and $P_U f = f$ if and only if $f \in U$.

If the orthogonal complement of U is denoted by $U^\perp$, where $U^\perp$ is a set of all vectors in V that are orthogonal to every vectors in U, i.e., $U^\perp = \left\{ h \in V : \langle g, h \rangle = 0, \forall g \in U \right\}$. Then $f - g = h$ is the orthogonal projection of f onto $U^\perp$, and h is orthogonal to g for every $g \in U$.

Now suppose $\{e_i\}_{i \in J}$ gives a orthonormal basis of U, then for every $f \in V$ given by Eq. (4.36), we have

$$P_U f = \sum_{j \in J} \langle f, e_j \rangle e_j \tag{4.38}$$

4.2.2.4 Frequency Domain Decomposition (FDD)

Mathematically speaking, the FDD technique, which is proposed to analyse $\hat{\mathbf{G}}(f_k)$, uses the Singular Value Decomposition (SVD) as the decomposition technique instead of the Spectral theorem (also known as the Spectral decomposition or Eigenvalue decomposition (EVD)), is actually unnecessary, especially for a positive semidefinite Hermitian matrix like $\hat{\mathbf{G}}(f_k)$. Therefore, in this subsection, we are going to unravel the reason for this.

For a Hermitian matrix, like the one we are interested in, we can decompose it into a diagonal matrix with the diagonal entries being the eigenvalues of the Hermitian matrix. Namely,

$$\hat{\mathbf{G}}(f_k) = \mathbf{U}_k \mathbf{\Lambda}_k \mathbf{U}_k^{\mathrm{H}} \tag{4.39}$$

where $\mathbf{U}_k = (\mathbf{u}_{k1}, \ldots, \mathbf{u}_{kd})$ is an $d \times d$ unitary matrix, while $\mathbf{\Lambda}_k = \mathrm{diag}\left([\lambda_{k1}, \ldots, \lambda_{kd}]^{\mathrm{T}} \right)$ is a $d \times d$ diagonal matrix with non-negative real eigenvalues. The columns of $\mathbf{U}_k$ form an orthonormal basis of the Hilbert space consisting of eigenvectors of $\hat{\mathbf{G}}(f_k)$.

Note that $\hat{\mathbf{G}}(f_k)$ is a special matrix, it is built in terms of a random process. So, $\hat{\mathbf{G}}(f_k)$ is a random matrix. When the original process is an i.i.d. Gaussian process, the Fourier transformed of the data matrix $\tilde{\mathbf{X}}(f_k)$ at each frequency line is filled with complex i.i.d. Gaussian entries. In this case, $\hat{\mathbf{G}}(f_k)$ is known as a Wishart matrix in the Random Matrix Theory (RMT), and Wishart matrices are positive semidefinite [133]. From our discussion, we know that, for a positive semidefinite matrix, all the eigenvalues must be non-negative [130]. Such a nice property actually applies to all covariance matrices [134].

Apart from this, the probability of sampling two eigenvalues to be very close to or very far apart from each other is very small, in other words, the space between to eigenvalues $0 < s = \lambda_j - \lambda_i < \infty$, where $i \neq j$ [133]. Because of this, we will only consider the case when the eigenvalues are all distinct.

When dealing with linear map from a vector space V to a vector space W, i.e., $T : V \rightarrow W$, we can consider the matrix of a linear map regarding a basis of V and a basis of W. Whereas for operators, since the mapping is happening in the same space, we almost always use only one basis and make it play both roles, which is what the Spectral theorem does. As for the SVD, it allows us to make good use of two different bases for the matrix of an operator [130].

Assume that $T \in \mathcal{B}(V)$ is an operator on a finite-dimensional Hilbert space V, then there exists a countable set J, orthonormal families $\{e_i\}_{i \in J}$ and $\{v_i\}_{i \in J}$ in V, and a family $\{s_j\}_{j \in J}$ of positive numbers, for every $f \in V$ such that [118]

$$Tf = \sum_{j \in J} s_j \langle f, e_j \rangle v_j \tag{4.40}$$

is called a SVD of the compact operator of T. And the positive numbers $\{s_j\}_{j \in J}$ are the so-called singular values, which are usually arranged in descending order. While $\{e_i\}_{i \in J}$ and $\{v_i\}_{i \in J}$ are often known as the right singular vectors and left singular vectors, respectively.

Now if we compose T with its adjoint, then we obtain the self-adjoint operator $T^\dagger T : V \rightarrow V$. Note that this operator is also compact since the product of a compact operator and a bounded operator is also compact [131]. By the Spectral theorem, we have

$$T^\dagger Tf = \sum_{j \in J} \lambda_j \langle f, e_j \rangle e_j \tag{4.41}$$

where $\{e_i\}_{i \in J}$ and $\{\lambda_j\}_{j \in J}$ are the eigenvectors and the eigenvalues with respect to $T^\dagger T$ in V, respectively. From Eq. (4.40), it implies $Te_j = s_j v_j$, and from Eq. (4.41), we know $T^\dagger Te_j = \lambda_j e_j$, then $T^\dagger Te_j = s_j^2 e_j$. Since $\lambda_j = \langle \lambda_j e_j, e_j \rangle = \langle T^\dagger Te_j, e_j \rangle = \langle Te_j, Te_j \rangle = \|Te_j\|^2 \geq 0$, it is not difficult to see that $s_j = \sqrt{\lambda_j}$, which mean the Spectral theorem and the SVD is interconnected. Usually, the orthonormal bases $\{e_i\}_{i \in J}$ and $\{v_i\}_{i \in J}$ in SVD are not uniquely determined by T, but the singular values are uniquely determined as positive square roots of the positive eigenvalues of $T^\dagger T$.

Although the Spectral theorem is powerful enough to analysis $\hat{\mathbf{G}}(f_k)$, and it is also interconnected to the SVD, the SVD does possess some numerical advantages when calculating an ill-conditioned matrix.

In this case, the condition number of a matrix is the square of the condition number of the data matrix, and the Spectral theorem may be less accurate than SVD. Besides, since the singular values are non-negative by definition, and they are automatically sorted in a descending order, which makes it more convenient to use in practice. Therefore, even if the Spectral theorem is applicable in the FDD, the SVD is usually favoured in the application.

Besides, it is worth pointing out that the reason that we prefer the Periodogram over the Correlogram to define a PSD estimator is because by using the Periodogram we can we treat $\hat{\mathbf{G}}(f_k)$ as a covariance matrix in the frequency domain. Whereas the Correlogram computed from the Fourier transform of a sample correlation function is unable to provide us such an extended understanding of a PSD estimator. When it is a covariance matrix, the decomposition of $\hat{\mathbf{G}}(f_k)$ can also be explained by using some concepts from statistics, and the net result will be equivalent. For example, one of the most popular statistical methods for analysing a multivariate process is the Karhunen–Loève transform [135][136], whose discrete version is often known as the Principal Component Analysis (PCA) [137]. In practice, the PCA is usually fulfilled by performing the Spectral theorem, SVD, or alternating least squares [138] (which is designed to better handle missing values) on a covariance matrix.

The idea of the PCA is, given an empirical covariance matrix and a normalised random vector on an inner product space, we want to maximise the empirical variance of the data points along the direction of that random vector. And the direction of an eigenvector, which corresponds to the largest eigenvalue of that empirical covariance matrix, happens to carry the most empirical variance. To be more specific, suppose we are given $\hat{\mathbf{G}}(f_k)$, a M scaled $d \times d$ empirical covariance matrix, and a $d \times 1$ random vector $\mathbf{v}$, which is normalised $\langle \mathbf{v}, \mathbf{v} \rangle = \mathbf{v}^{\mathrm{H}} \mathbf{v} = 1$, then when we pre multiply $\mathbf{v}^{\mathrm{H}}$ and post multiply $\mathbf{v}$ to $\hat{\mathbf{G}}(f_k)$, this actually gives us what the scaled sample variance of the data points $\tilde{\mathbf{y}}_i(f_k)$ along the direction $\mathbf{v}$ is, i.e.,

$$
\begin{aligned}
\mathbf{v}^{\mathrm{H}} \hat{\mathbf{G}}(f_k) \mathbf{v} &= \frac{1}{Mn_d} \sum_{i=1}^{n_d} \mathbf{v}^{\mathrm{H}} \tilde{\mathbf{y}}_i^*(f_k) \tilde{\mathbf{y}}_i^{\mathrm{T}}(f_k) \mathbf{v} \\
&= \frac{1}{Mn_d} \sum_{i=1}^{n_d} \left(\mathbf{v}^{\mathrm{T}} \tilde{\mathbf{y}}_i(f_k) \right)^* \left(\mathbf{v}^{\mathrm{T}} \tilde{\mathbf{y}}_i(f_k) \right)^{\mathrm{T}} = \frac{1}{Mn_d} \sum_{i=1}^{n_d} \left| \mathbf{v}^{\mathrm{T}} \tilde{\mathbf{y}}_i(f_k) \right|^2
\end{aligned}
\tag{4.42}
$$

Thus, $\mathbf{v}^{\mathrm{H}} \hat{\mathbf{G}}(f_k) \mathbf{v}$ is the M scaled empirical variance of $\mathbf{v}^{\mathrm{T}} \tilde{\mathbf{y}}_1(f_k)$, ..., $\mathbf{v}^{\mathrm{T}} \tilde{\mathbf{y}}_{n_d}(f_k)$. Here, the inner product $\mathbf{v}^{\mathrm{T}} \tilde{\mathbf{y}}_i(f_k)$ is actually the coordinates of $\tilde{\mathbf{y}}_i(f_k)$ when projected on $\mathbf{v}$. So, if $\mathbf{v}^{\mathrm{H}} \hat{\mathbf{G}}(f_k) \mathbf{v}$ is large, it means when we look at those data points as projected onto the axis generated by $\mathbf{v}$ will have a lot of variances. The larger the variance, the further away they are from each other on average. In summary, the idea of PCA is to try to identify those directions along which we have a lot of variances.

In other words, we want to maximise the empirical variance $\mathbf{v}^{\mathrm{H}}\hat{\mathbf{G}}(f_k)\mathbf{v}$ of the data points projected on to the direction $\mathbf{v}$ when $\mathbf{v}$ is of norm 1, which is actually a non-convex problem [139]. However, we can still solve it by turning it into a convex problem and attain the maximum at the boundary. Actually, the PCA is one of the fanciest non-convex problems we can solve efficiently by using the Spectral theorem or SVD. Therefore, if we rewrite Eq. (4.39) as follows

$$\mathbf{U}_k^{\mathrm{H}}\hat{\mathbf{G}}(f_k)\mathbf{U}_k = \frac{1}{M}\mathbf{U}_k^{\mathrm{H}}\left(\frac{1}{n_d}\tilde{\mathbf{X}}^*(f_k)\tilde{\mathbf{X}}^{\mathrm{T}}(f_k)\right)\mathbf{U}_k = \mathbf{\Lambda}_k \tag{4.43}$$

Then the eigenvalues in $\mathbf{\Lambda}_k$ actually give the M scaled empirical variance of the data points along the eigenvectors. And the sum of the eigenvalues in $\mathbf{\Lambda}_k$ gives the M scaled total empirical variance of the data points at f_k, i.e., $\mathrm{tr}(\mathbf{\Lambda}_k) = \lambda_{k1} + \cdots + \lambda_{kd}$, where $\mathrm{tr}(\mathbf{\Lambda}_k)$ denotes of the trace of $\mathbf{\Lambda}_k$.

Note that those eigenvectors or singular vectors which maximise the data variance is often called principal components, whereas the eigenvalues and the singular values are known as the latent factors or the principal component variance. In the following content, we will use these nomenclatures interchangeably without further clarification. However, to be consistent, we will stick to the Spectral theorem in theoretical explanation but use the SVD in practice.

Therefore, statistically speaking, the FDD is the variance-based method. By using the PCA, we can measure the spread between the data points, and we want them to be as spread as possible in the direction in which we project. The wider the spread, the larger the variance, then the more information we have in a direction. Usually, we are interested in two directions that correspond to the first two largest sample variances when using the FDD. But in most cases, we plot all the singular values against the frequency line in one plot.

For a random process with multimodal distribution, when all the modes are well-separated, we can observe those modes from the first singular value plot without any difficulties. But when two modes are very close together that they cannot be separated in frequency less than the frequency resolution f_s/Nfft, where f_s is the sampling frequency and Nfft denotes the number of Fast Fourier Transform (FFT) points used in a calculation, in this case, if we only look at the first singular values alone, we will not be able to tell if there is a close mode. However, we can detect the presence of the close modes by analysing the second singular value plot.

Generally, when two modes are very close to each other around f_k, we usually see that the second singular value takes a large proportion in the total variance around f_k, and the closer the two modes, the larger the second singular value is around f_k, vice versa. Thus, by analysing the behaviour of the second singular values, we can have an idea about the presence of the close modes, even if the two close

modes are inseparable in the first singular value plot. Since the natural frequency is estimated by the location of the peaks in the first singular value plot in FDD, then FDD is not a good tool be used to identify the closed modes frequencies and the corresponding mode shapes, as the mode shape identification depends on the frequency identification. However, it is a good technique to be used to detect the close modes by analysing the second singular values.

Given a covariance matrix, our goal is to measure the spread or variability between the data points. The conventional FDD method says that we can use the variance to quantify the statistic dispersion or uncertainty. However, according to our discussion that the such a variance-based method is incapable of identifying the close modes mode shapes if the frequencies of the close modes are inseparable. Now, we want to consider that whether or not we can identify the close modes mode shapes if we use another measure to quantify the statistical dispersion instead of the variance. Fortunately, a good candidate is the entropy. It makes intuitive sense that when dealing with a random process with unimodal distribution, increasing the variance means the increasing of the uncertainty. However, for a random process with multimodal distribution, the distribution of possible outcomes spreads around two or more peaks. In this case, the variance of the distribution will be too poor to capture the overall spread [140]. In other words, the variance is not a good measure of the uncertainty in the case of multimodal distribution. Therefore, in the next section, we will introduce a density operator, which will lead us to compute the Von Neumann Entropy (VNE) or the purity of a density operator. One will see in section 4.4.2.3 that our new measure can give contrasting results to the variance when two modes are very close to each other that the FDD cannot separate them.

4.2.2.5 Density operators

With the understanding of the decomposition of a Hermitian operator on a finite-dimensional Hilbert space, now we can talk about the density operator (also known a density matrix) in Quantum Mechanics (QM). Our discussion will start with the Schrödinger equation and a stationary state because we want to establish a connection between as a stationary random process and the stationary state of a physical system. After that, we can define a density operator, which is a more capable description of the states of a physical system, as it can encode all the accessible information about the system. Finally, with the definition of a density operator, we can talk about how to obtain it via a PSD estimator. And in this final subsection, we will also explain why we can use the FDD to estimate the mode shape.

However, before our discussion, we want to clarify the potential change of the notations. In QM, the Dirac notation [117], which is to denote vectors and convectors, is used ubiquitously. Thus, to be consistent, we will use Dirac notation when it is appropriate. By definition, vectors are denoted by kets, i.e., $\left| f \right\rangle$ for $f \in V$, then for each ket we can associate it to a covector in the dual space called bra, i.e., $\left\langle g \right|$ for $g \in V'$. Here, the dual space of V, denoted by V', is a vector space of all linear functionals

on V, i.e., $V' \in \mathcal{L}(V, \mathbb{C})$. When V has finite dimension, so does V', and their dimensions are the same. With this new notation, the inner product of two objects is denoted as $\langle f | g \rangle$, while the outer product $f \otimes g$, is denoted by $|f\rangle\langle g|$.

I.   Schrödinger equation and a stationary state

In structural dynamics, the response of a vibrating system $y(x,t)$ can be expressed it in terms of some modal coordinates [141], i.e.,

$$y(x,t) = \sum_{i=1}^{\infty} q_i(t) \psi_i(x) \tag{4.44}$$

where $q_i(t)$ is the ith modal coordinates, and $\psi_i(x)$ is the corresponding mode shape.

In Classical Mechanics (CM), the configuration or a state of a system is specified by a point in the space of the coordinates and the momenta in a fully deterministic way. However, governed by the position-momentum uncertainty principle, it is impossible to know both the position and momentum of a particle, such as an electron, at every point along the trajectory. Therefore, we have to choose another variable to specify the state of a quantum system. And this variable is the wavefunction denoted as $\psi(x)$. When we square it, namely $|\psi(x)|^2$, it gives the probability or probability density that an object in the state $\psi(x)$ to be found at position x. This definition implies that a wavefunction must be normalised to satisfy the probability principle that the total probability must be equal to 1, i.e., $\int_{-\infty}^{\infty} |\psi(x)|^2 \, dx = 1$.

When a wavefunction is evolving in time, it becomes a function of space x and time t, i.e., $\psi(x,t)$, and it is governed by the Schrödinger equation. Note that this notation gives us a scalar wave and $\psi(x,t) \in \mathbb{Q}$ is complex-valued function. For a quantum object, the configuration or state of it is completely specified by a wavefunction.

For a free particle with wavenumber k and frequency ω, its one-dimensional wavefunction is given by $\psi(x,t) = e^{i(kx-\omega t)}$, which is also known a de Broglie wave. The wavenumber k and the frequency ω together determine the momentum p and the energy E of a particle through the relations $p = \hbar k$, $E = \hbar \omega$ and $E = p^2/2m$, where m is the mass of the particle, $\hbar$ is Planck's reduced constant. Here, e^{ikx} is often known as a plan wave, while $e^{-i\omega t}$ is the global phase factor. For this free particle, the Schrödinger equation is given by

$$i\hbar \frac{\partial}{\partial t}\psi(x,t) = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\psi(x,t) \tag{4.45}$$

When a particle moving in some external time-independent potential $V(x)$, the Schrödinger equation is modified to

$$i\hbar \frac{\partial}{\partial t}\psi(x,t) = \left(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}+V(x)\right)\psi(x,t) \tag{4.46}$$

Although we are not dealing with a quantum system here, what we have at hand is a collection of the system responses $\{y(x,t)\}$. The same as the wavefunction $\psi(x,t)$, the vibration response of a classical system $y(x,t)$ is also a function of space and time. This implies that analysis conducted in QM regarding the wavefunctions should be applicable in modal analysis. Particularly, the idea of the stationarity of a random process is equivalent to the definition of a stationary state of a quantum system, as they both have the property of time independence. So, a stationary state in QM really just means the probabilities of outcomes of a measurement of any property of the system remains the same regardless of when we measure it. Thus, because of the aforementioned reasons, we will discuss a wavefunction $\psi(x,t)$ first before we come back to $y(x,t)$.

A stationary state is related to a Hamiltonian operator $\hat{H}$, which is also known as the energy operator $\hat{E}$, and it is given by the term in the brackets of Eq. (4.46) has the units of energy. When a quantum system is in a stationary state, we can write the wavefunction as the product of spatial and temporal terms, namely $\psi(x,t) = q(t)\phi(x)$, and called it a full stationary state. Now if we plug this expression of a wavefunction into Eq. (4.46) and replace the complicated term in the brackets of Eq. (4.46) with a Hamiltonian operator $\hat{H}$, we obtain

$$i\hbar \frac{dq(t)}{dt}\phi(x) = q(t)\hat{H}\phi(x) \tag{4.47}$$

Since the Hamiltonian is independent of t, $q(t)$ is moved across $\hat{H}$. If we group the time-independent term on one side of the equation and the time-dependent term on the other side of the equation, we get

$$i\hbar \frac{1}{q(t)}\frac{dq(t)}{dt} = \frac{1}{\phi(x)}\hat{H}\phi(x) \tag{4.48}$$

Since the equality has to be satisfied for all x and t in Eq. (4.48), the only possibility is that this equation equals to a constant E with units of energy. Thus, by equalling both sides of Eq. (4.48), we

can obtain a time-dependent solution $q(t) = e^{-iEt/\hbar}$ from the left hand side, and an eigenvalue equation for a Hamiltonian operator from the right hand side, namely

$$\hat{H}\phi(x) = E\phi(x) \tag{4.49}$$

where the eigenvalue of $\hat{H}$ is given by E, accordingly, $\phi(x)$ denotes an eigenstate of $\hat{H}$, and $\int_{-\infty}^{\infty} |\phi(x)|^2 \, dx = 1$. By contrast with Eq. (4.46), Eq. (4.49) is often known as the time-independent Schrödinger equation. Since a Hamiltonian is Hermitian and positive, the eigenvalues of it must be real and non-negative, i.e., $E = E^*$.

Usually, a Hermitian operator in QM is associated with an observable. The definition of observables is related to the measurements of a quantum system. Roughly speaking, any measurable properties of a physical system, such as position, momentum, energy, etc., are known as an observable. Particularly in QM, as an operator, an observable carries all the values that the corresponding physical quantity could have through its eigenvalues. When we conduct a measurement for a particular physical property of a quantum system, say energy, the outcome can be any of the possible values of E, where E comes from Eq. (4.49). The procedure of associating a Hermitian operator with the observable properties of a quantum system will further discussed in the next subsection. But now, let us finish our discussion of a stationary state.

Since for a stationary state its wavefunction is given by $\psi(x,t) = q(t)\phi(x)$, we can multiply $q(t)$ on both sides of Eq. (4.49), and obtain $\hat{H}\psi(x,t) = E\psi(x,t)$. It means that the full stationary state is also an eigenstate for the Hamiltonian operator $\hat{H}$, and it must satisfy the normalisation condition. In other words, the total probability must be conserved, namely

$$\int_{-\infty}^{\infty} \psi^*(x,t)\psi(x,t)\,dx = \int_{-\infty}^{\infty} e^{-iE^*t/\hbar}e^{-iEt/\hbar}\phi^*(x)\phi(x)\,dx = \int_{-\infty}^{\infty} |\phi(x)|^2 \, dx = 1 \tag{4.50}$$

Based on this normalisation condition and the statistical property of $|\phi(x)|^2$ (probability density), we can obtain an expected value of a Hamiltonian H, which is denoted by $\langle H \rangle$, such that

$$\langle H \rangle = \int_{-\infty}^{\infty} \psi^*(x,t)\hat{H}\psi(x,t)\,dx = \int_{-\infty}^{\infty} \phi^*(x)\hat{H}\phi(x)\,dx \tag{4.51}$$

The above equation says that although $\psi(x,t)$ differs from $\psi(x)$ by a global phase factor $e^{-iEt/\hbar}$ of constant magnitude, the statistical measurement predicted for these two states are the same. This gives the exact reason that the state $\psi(x,t)$ given by $e^{-iEt/\hbar}\phi(x)$ is called stationary.

Since the statistical properties of a stationary state do not change with time, it means that if a state has all the information about the system at some time t, then the state of the system at time $t + dt$ depends only on the state at time t and on a multiplicative factor $U(t, t+dt)$, namely, [142]

$$\psi(x, t+dt) = U(t, t+dt)\psi(x, t) \tag{4.52}$$

From Eq. (4.50) we know that $\psi(x, t)$ is normalised to unity, when $\psi(x, t+dt)$ is also normalised, $U(t, t+dt)$ is given by $e^{-iEdt/\hbar}$, and hence $|U(t, t+dt)|^2 = 1$, which means $U(t, t+dt)$ is unitary, and we call it the time evolution operator, and simply denoted it by U.

Similarly, when we apply the Fourier transform to a full stationary state $\psi(x, t) = q(t)\phi(x)$ with respect to t, the operation only acts on $q(t)$. Apart from this, since the Fourier transform is unitary, as we mentioned in section 4.2.2.1, then the Fourier transform of $\psi(x, t)$ with respect to t, denoted by $\tilde{\psi}(x, f)$, should preserve all the information of the system in $\psi(x, t)$.

Besides, since the wavefunction $\psi(x, t)$ is a solution of the Schrödinger equation, while the Schrödinger equation is linear, which means that the Schrödinger equation admits more general solutions than a single wavefunction. In other words, any superposition of the wavefunction $\psi(x, t)$ will be a solution to the Schrödinger equation. Thus, given a system in a superposition, its general time-dependent wavefunction could be given by

$$\Psi(x, t) = \sum_i \alpha_i \psi_i(x, t) \tag{4.53}$$

where α_i are the arbitrary complex coefficients satisfying normalisation condition, i.e., $\sum_i |\alpha_i|^2 = 1$.

While $\psi_i(x, t)$ give the normal modes of the system. Apply the Fourier transform to Eq. (4.53), we get

$$\tilde{\Psi}(x, f) = \sum_i \alpha_i \tilde{\psi}_i(x, f) \tag{4.54}$$

where $\tilde{\Psi}(x, f)$ is the wavefunction we want to work on. Therefore, in the next subsection, we will first derive the density operator in a general form, then specify it in our case.

II. Density operators for pure and mixed states

In this subsection, we are going to wrap all the discussions we mentioned in the previous sections and introduce a density operator and the associated purity condition and Von Neumann entropy. Note that the Dirac notations will be used instead of the normal representation we used in our previous discussion.

In the last subsection, we have encountered the concept of observables, and we discussed the energy of a physical system in particular. We said that every observable is associated with a Hermitian operator. The truth is, it is quite possible for us to construct all manner of Hermitian operators to be associated with any given physical system without having a readily identifiable physical feature (such as position, momentum, energy, etc.) of the system, as it might be the case that it is not clear how much such observables could be measured. In this case, we shall consider a general observable A for a system with a discrete set of values $\{a_i\}_{i \in \mathbb{Z}^+}$, and using $\hat{A}$ to represent its associated Hermitian operator. Then $\hat{A}$ should have its eigenvalues given by $\{a_i\}_{i \in \mathbb{Z}^+}$, and for each a_i, the corresponding eigenfunction is denoted by $|\phi_i\rangle$, such that

$$\hat{A}|\phi_i\rangle = a_i|\phi_i\rangle \tag{4.55}$$

where $\langle\phi_i|\phi_j\rangle = \delta_{ij}$, if and only if $i = j$, $\delta_{ij} = 1$, otherwise $\delta_{ij} = 0$. Then a set of eigenstates $\{|\phi_i\rangle\}_{i \in \mathbb{Z}^+}$ form a complete, orthonormal set such that any state of the system $|\psi\rangle$, it can be expressed as a linear combination of the eigenstates

$$|\psi\rangle = \sum_i c_i|\phi_i\rangle = \sum_i |\phi_i\rangle\langle\phi_i|\psi\rangle \tag{4.56}$$

where c_i is given by $\langle i|\psi\rangle$ provided that $|\psi\rangle$ is normalised, i.e, $\langle\psi|\psi\rangle = 1$, and $|\langle\phi_i|\psi\rangle|^2$ gives the probability of obtaining the result a_i on measuring $\hat{A}$. Note that the same as the superposition we mentioned in the last subsection, for any state vector $|\psi\rangle$, Eq. (4.56) also gives a superposition, which is the superposition of states $|\phi_i\rangle$ with amplitude $\langle\phi_i|\psi\rangle$ for the state $|\phi_i\rangle$. Usually, $\langle\phi_i|\psi\rangle$ is called the probability amplitude for the eigenstate $\langle\phi_i|$, which is the inner product of between the state vectors. Besides, the state vector $|\psi\rangle$ can be a time-dependent solution of the Schrödinger equation. In this case, when the state is stationary, Eq. (4.56) can be expressed as

$$|\psi(t)\rangle = \sum_i c_i(t)|\phi_i\rangle \tag{4.57}$$

where $c_i(t)$ are the time-dependent coefficients. In stochastic process analysis, Eq. (4.57) actually gives the Karhunen–Loève (KL) expansion [135][136]. In this case, $c_i(t)$ are known as the zero mean uncorrelated coefficients.

Associated with any isolated physical system is a complex Hilbert space, which is called a state-space, and it is formed by a collection of all the possible state vectors. According to our discussion in section 4.2.2.2 and 4.2.2.3, we know that for a finite-dimensional Hilbert space, the eigenvectors of a self-adjoint compact operator forms a complete set of basis states, and the dimension of the state-space is given by the number of basis states. It means that any state vectors in the state-space can be expressed by a linear combination of the basis vectors. In other words, there is zero probability that for a state $|\psi\rangle$ of the system for which $\langle \phi_i | \psi \rangle = 0$ for every state. Therefore, the completeness of a Hilbert space is very important, which guarantees the system state is measurable. However, when such a condition is violated, for instance, when the space is infinite-dimensional, it is possible for a system to be in a state which cannot be represented as a linear combination of the eigenstates. As we mentioned before, we will not consider this case in this work.

Followed by our previous discussion, we can obtain the expectation value for the observable A, which is given by the following expression

$$\langle A \rangle_\psi = \sum_i a_i \langle \psi | \phi_i \rangle \langle \phi_i | \psi \rangle = \sum_i \left| \langle \phi_i | \psi \rangle \right|^2 a_i = \sum_i \langle \psi | \hat{A} | \phi_i \rangle \langle \phi_i | \psi \rangle = \langle \psi | \hat{A} | \psi \rangle \qquad (4.58)$$

where $\left| \langle \phi_i | \psi \rangle \right|^2$ gives the probability to measure the eigenvalue a_i.

Although the complete information about a system is contained in its wavefunctions or state vectors, they can only describe pure states of a system. Such idealised descriptions cannot characterise the statistical mixtures of the states. To resolve this problem, John Von Neumann [143] introduced the density operator formalism to count both pure and mixed states situations of a physical system. Based on Von Neumann's postulate, we can make use of the projection operators to describe the collapse of a wavefunction upon measuring an observable A for a system in state $|\psi\rangle$ to an eigenstate $|\phi_i\rangle$ with a result a_i, which gives a basis of the methods of density operators for pure and/or mixed states. Although the collapse of a wavefunction only applies to a quantum system, we can still use the idea of density matrix to do analysis in Classical Mechanics.

First, let us consider an observable A in the pure state $|\psi\rangle$, and $\langle \psi | \psi \rangle = 1$. The density matrix ρ for this pure state is defined as

$$\rho_\psi = |\psi\rangle \langle \psi| \qquad (4.59)$$

which is an outer product of the wavefunction and its conjugate. One can easily validate that ρ_{ψ} is a projection operator, i.e., $\rho_{\psi}^2 = \rho_{\psi}$, and it is self-adjoint and positive with its trace equals 1, namely,

$$\text{tr}\left(\rho_{\psi}\right) = \sum_i \langle i | \rho_{\psi} | i \rangle = 1 \tag{4.60}$$

where $|i\rangle$ is the standard basis vector with 1 in the ith entry zero everywhere. Since $\rho_{\psi}^2 = \rho_{\psi}$, then $\text{tr}\left(\rho_{\psi}^2\right) = 1$.

Since ρ_{ψ} is a projection operator, when we apply it to $\hat{A}$, we will project $\hat{A}$ onto to the space spanned by $|\psi\rangle$, i.e., $\rho_{\psi}\hat{A} = |\psi\rangle\langle\psi|\hat{A}$. While by taking the trace of $\rho_{\psi}\hat{A}$, we can get the expectation value of A, that is

$$\langle A \rangle = \text{tr}\left(\rho_{\psi}A\right) = \sum_i \langle i | \psi \rangle \langle \psi | A | i \rangle = \sum_i \langle \psi | A | i \rangle \langle i | \psi \rangle = \langle \psi | A | \psi \rangle \tag{4.61}$$

On the other hand, since $|\psi\rangle$ may in a linear superposition of several eigenstates, substitute Eq. (4.56) into Eq. (4.59), we can obtain

$$\rho_{\psi} = |\psi\rangle\langle\psi| = \sum_{i,j} c_i c_j^* |\phi_i\rangle\langle\phi_j| = \sum_{i,j} \rho_{ij} |\phi_i\rangle\langle\phi_j| \tag{4.62}$$

where ρ_{ij} are the density matrix elements, which can also be given by $\rho_{ij} = \langle i | \rho_{\psi} | j \rangle$.

When $i = j$, we can denote $P_i = |\phi_i\rangle\langle\phi_i|$, which gives our usual rank one projection operator we discussion in section 4.2.2.3. When acting on any vector $|\psi\rangle$, it projects $|\psi\rangle$ onto a one-dimensional subspace of the state-space spanned by $|\phi_i\rangle$. Similarly, by applying the projection operator $P_i + P_j$, where $i \neq j$, to any vector $|\psi\rangle$, it will give us a vector in the subspace spanned by $|\phi_i\rangle$ and $|\phi_j\rangle$. And if we combine the projection operators of P_i, we will obtain a projection operator projects onto the state-space itself. Therefore, when $i = j$, Eq. (4.62) becomes

$$\rho_{\psi} = \sum_i \rho_{ii} |\phi_i\rangle\langle\phi_i| \tag{4.63}$$

where $\rho_{ii} \geq 0$ gives the probability of observing the eigenstate $|\phi_i\rangle$, and $\sum_i \rho_{ii} = 1$. Eq. (4.63) actually gives us the decomposition of a density operator based on the eigenstates of the state-space. In this case, ρ_{ii} is the eigenvalue of the ρ_{ψ}.

Statistically, given an ensemble of objects in states $\{|\psi_i\rangle\}_{i\in\mathbb{Z}^+}$, if all the objects are in the same state, then the ensemble is represented by a pure state, then we can use either a wavefunction or the density matrix presented in Eq. (4.59) to describe the system. However, if not all of the objects in the ensemble are in the same state, then we can only use the density matrix to encode all the possibilities. Therefore, similar to Eq. (4.63), the density matrix for a mixed state is given by the weighted sum of the pure states

$$\rho_{\text{mix}} = \sum_i p_i |\psi_i\rangle\langle\psi_i| \tag{4.64}$$

where p_i gives the probability the system is in state $|\psi_i\rangle$, and $\sum_i p_i = 1$. While $|\psi_i\rangle\langle\psi_i|$ gives the density matrix for a pure state.

Given the definition given by Eq. (4.64), we can verify that the trace of ρ_{mix}^2 will be less than 1 when the system is in a mixed state. To show this, we first calculate ρ_{mix}^2, i.e.,

$$\rho_{\text{mix}}^2 = \sum_{i,j} p_i p_j |\psi_i\rangle\langle\psi_i|\psi_j\rangle\langle\psi_j| = \sum_i p_i^2 |\psi_i\rangle\langle\psi_i| \tag{4.65}$$

Particularly when $\{|\psi_i\rangle\}_{i\in\mathbb{Z}^+}$ are the eigenstates, then $\{p_i^2\}_{i\in\mathbb{Z}^+}$ gives the eigenvalues of ρ_{mix}^2,

As we can see that unlike the density matrix of a pure state, $\rho_{\text{mix}}^2 \neq \rho_{\text{mix}}$, the trace of ρ_{mix}^2 which is given by

$$
\begin{aligned}
\text{tr}\left(\rho_{\text{mix}}^2\right) &= \sum_n \langle n| \sum_{i,j} p_i p_j |\psi_i\rangle\langle\psi_i|\psi_j\rangle\langle\psi_j|n\rangle \\
&= \sum_{i,j} p_i p_j \langle\psi_i|\psi_j\rangle\langle\psi_j| \sum_n |n\rangle\langle n|\psi_i\rangle \\
&= \sum_{i,j} p_i p_j \left|\langle\psi_i|\psi_j\rangle\right|^2 = \sum_i p_i^2
\end{aligned} \tag{4.66}
$$

By definition, $\sum_i p_i = 1$, thus, $\text{tr}\left(\rho_{\text{mix}}^2\right) = \sum_i p_i^2 \leq 1$, the equality obtains if and only if the system is in pure states, otherwise the $\text{tr}\left(\rho_{\text{mix}}^2\right)$ is strictly smaller than 1 for mixed states. Therefore, the trace of ρ_{mix}^2 is a good measure for the mixedness or purity condition of a density matrix, and we call $\text{tr}\left(\rho_{\text{mix}}^2\right)$ the purity for short [144]. While for a maximally mixed state, when the system has dimension d, i.e, d possible states, we have $\text{tr}\left(\rho_{\text{mix}}^2\right) = 1/d > 0$ [142], and hence we have the following criteria,

$$0 < \frac{1}{d} \leq \text{tr}\left(\rho_{\text{mix}}^2\right) \leq 1 \tag{4.67}$$

With a density operator, we can use another measure, i.e., Von Neumann entropy [142], to quantify the departure of the system from a pure state. The Von Neumann entropy is an extension of Shannon entropy [145] to quantum states. Unlike the variance, the entropy is good for a process with a multimodal distribution. When a system is in a mixed state, by computing the Von Neumann entropy, we will know the degree of mixing of the state describing a given finite system. If we denote the density matrix of a system simply by ρ, the Von Neumann entropy $S(\rho)$ is given by,

$$S(\rho) = -\mathrm{tr}(\rho \log_2 \rho) \tag{4.68}$$

To compute $S(\rho)$, sometimes it is convenient to apply the eigenvalue decomposition to ρ first. Hence, if $\{\bar{\lambda}_i\}_{i \in \mathbb{Z}^+}$ are the eigenvalues of ρ, then $S(\rho)$ can be re-expressed as

$$S(\rho) = -\sum_i \bar{\lambda}_i \log_2 \bar{\lambda}_i \tag{4.69}$$

Note that if $\bar{\lambda}_i = 0$, then $\bar{\lambda}_i \log \bar{\lambda}_i \equiv 0$, as an event of probability zero should not be contributed to the entropy. When the system is in a pure state, $S(\rho)$ is equal to 0. If the system is in a maximally mixed state, then $S(\rho)$ will reach its maximum too. Therefore, the purity condition given by $\mathrm{tr}(\rho_{\mathrm{mix}}^2)$ is equivalent to $S(\rho)$, they deliver the same information about the system. In our analysis, we will use $\mathrm{tr}(\rho_{\mathrm{mix}}^2)$ to do the identification, but plot $S(\rho)$ in the meantime.

III.   Constructing a density operator from the PSD estimator

Suppose we have a $d < \infty$ dimensional Hilbert space, which has a complete set of basis vectors to span the space. To use the idea of a density matrix, we first need to think of our target mode shapes as the state vectors in this state-space, and our conventional modal space, which is spanned by the mode shape vectors, is a subspace of this d dimensional inner product space. The general time-dependent state vector of a stationary state living in the modal space, following the usual mode superposition idea in structural dynamics, is simply a linear superposition of the mode shape vectors times the corresponding time-dependent factor. Therefore, when we take the Fourier transform of the state vector with respect to time t, it preserves all the information of a state. Then by using our usual notation, any state vector $\tilde{\mathbf{y}}(f_k)$ can be written as a linear combination of the mode shape vectors, just like the superposition equation given by Eq. (4.54) or the Fourier transform of Eq. (4.44), namely

$$\tilde{\mathbf{y}}(f_k) = \sum_{j=1}^{r} \boldsymbol{\psi}_j \tilde{q}_j(f_k) = \boldsymbol{\Psi} \tilde{\mathbf{q}}(f_k) \tag{4.70}$$

where $r \le d$ is the number of modes, $\tilde{q}_j$ denotes the Fourier transform of the j th generalise coordinates, $\tilde{\mathbf{q}}(f_k)$ is a $r \times 1$ vector. While $\boldsymbol{\psi}_j$ denotes the j th mode shape vector with size $d \times 1$, and $\boldsymbol{\Psi}$ is a $d \times r$ mode shape matrix.

On the other hand, according to our discussion of a PSD operator $\hat{\mathbf{G}}(f_k)$, the eigenvectors of $\hat{\mathbf{G}}(f_k)$ forms a complete set of basis vectors of the measurement state-space. Therefore, any state $\tilde{\mathbf{y}}(f_k)$ in the state-space of $\hat{\mathbf{G}}(f_k)$ at each frequency location can be expanded in terms of the eigenvectors $\mathbf{u}_{kj}$ of $\hat{\mathbf{G}}(f_k)$ as

$$\tilde{\mathbf{y}}(f_k) = \sum_{j=1}^{d} \mathbf{u}_{kj} z_j(f_k) = \mathbf{U}_k \mathbf{z}(f_k) \tag{4.71}$$

where z_j is the zero-mean uncorrelated coefficient, and $\mathbf{z}(f_k)$ is a $d \times 1$ vector. When normalised, $\left| z_j(f_k) \right|^2$ gives the probability of finding $\tilde{\mathbf{y}}(f_k)$ is the eigenstate $\mathbf{u}_{kj}$. Note that Eq. (4.71) can also be obtained by applying the KL expansion [136] to the random process $\tilde{\mathbf{y}}(f_k)$ in the frequency domain.

Likewise, we can express our mode shape vector $\boldsymbol{\psi}_i$ in terms of the eigenvectors of $\hat{\mathbf{G}}(f_k)$, namely

$$\boldsymbol{\psi}_i = \sum_{j=1}^{d} \mathbf{u}_{kj} \beta_j(f_k) = \mathbf{U}_k \boldsymbol{\beta}(f_k) \tag{4.72}$$

where β_j is the constant coefficient, and $\boldsymbol{\beta}(f_k)$ is a $d \times 1$ vector. After normalisation, $\left| \beta_j(f_k) \right|^2$ gives the probability of finding $\boldsymbol{\psi}_i$ in the eigenstate $\mathbf{u}_{kj}$.

The matrix form of Eq. (4.72) is given by

$$\boldsymbol{\Psi} = \mathbf{U}_k \mathbf{Z}_k \tag{4.73}$$

When $\boldsymbol{\Psi}$ is of size $d \times r$, then $\mathbf{Z}_k$ is $d \times r$ matrix, and we can easily obtain $\mathbf{Z}_k$ via the following equation by using the unitary property of $\mathbf{U}_k$,

$$\mathbf{U}_k^{\mathrm{H}} \boldsymbol{\Psi} = \mathbf{Z}_k \tag{4.74}$$

Similarly, we can also inverse the process to show how much of each mode shape vectors contribute to the first eigenvector, which would be a useful indicator to be used alongside the FDD, as it uses the first singular vector to estimate the mode shape, i.e.,

$$\mathbf{u}_{1k} = \sum_{i=1}^{r} \mathbf{\psi}_i c_i \left(f_k \right) = \mathbf{\Psi} \mathbf{c} \left(f_k \right) \tag{4.75}$$

where $\mathbf{u}_{1k}$ is the first singular vector at frequency f_k, when normalised $\left| c_i \left(f_k \right) \right|^2$ are the probability contributed by each mode shape vector, and we name it as contribution scores. $\mathbf{c} \left(f_k \right)$ is a $r \times 1$ coefficient vector, which is computed as follows,

$$\mathbf{c} \left(f_k \right) = \left(\mathbf{\Psi}^{\mathrm{H}} \mathbf{\Psi} \right)^{-1} \mathbf{\Psi}^{\mathrm{H}} \mathbf{u}_1 \left(f_k \right) \tag{4.76}$$

If a mode is dominant, then the corresponding coefficient should be close to 1.

Therefore, both $\left| \beta_j \left(f_k \right) \right|^2$ and $\left| c_i \left(f_k \right) \right|^2$ are useful indicators of the closeness of the mode shape vectors and the eigenvectors, when we plot them against the frequency line. Especially for the contribution scores, it will give us a trajectory of how a certain mode develops with the increase of the frequency, and it is quite useful when we use the first singular vectors to estimate the mode shapes.

In particular, the larger the value of $\left| \beta_1 \left(f_k \right) \right|^2$, the closer the eigenvector $\mathbf{u}_{k1}$ is to the $\mathbf{\psi}_i$. This explains why we can use the first singular vectors to estimate the mode shapes regardless the mode shape vectors are orthogonal to each other or not. Similarly, if the ith $\left| c_i \left(f_k \right) \right|^2$ is larger than the others, then $\mathbf{u}_{k1}$ can be explained by ith mode shape vector $\mathbf{\psi}_i$.

Besides, according to our discussion of the PCA, we know that the first singular vector represents the direction that explains most of the variance or information in the data point. For each matrix, all of the singular vectors are orthogonal to each other. Thus, for a random state $\tilde{\mathbf{y}} \left(f_k \right)$ and the mode shape $\mathbf{\psi}_i$, if $\left| z_1 \left(f_k \right) \right|^2$ is close to one, in the meantime, $\left| \beta_1 \left(f_k \right) \right|^2$ is close to 1, then $\tilde{\mathbf{y}} \left(f_k \right)$ can be explained by $\mathbf{\psi}_i$, as the eigenvector $\mathbf{u}_{k1}$ can explain both $\tilde{\mathbf{y}} \left(f_k \right)$ and $\mathbf{\psi}_i$. When both $\left| z_1 \left(f_k \right) \right|^2$ and $\left| z_2 \left(f_k \right) \right|^2$ are large, then $\tilde{\mathbf{y}} \left(f_k \right)$ is explained by two modes, meaning that there is might be close modes in presence.

In the case when two modes are very close, or even for the case of repeated poles, if $\left| \beta_1 \left(f_k \right) \right|^2$ is close to 1, then we can use the first singular vector to explain one mode, while using the second singular vector to explain the other mode provided that those two modes are almost orthogonal two each other.

Besides, by observing the plot of $\left| \beta_j \left(f_k \right) \right|^2$ or $\left| c_i \left(f_k \right) \right|^2$ against frequency, we can also get some insight about close modes. However, they are only useful indicators for close modes detection when we know the reference mode shapes. To see them both in action, we will use the contribution scores in the

first case study, as the modes are not too close such that we still use the first singular vectors to estimate the mode shape; while in the second case study, we use the probability indicator $\left|\beta_j\left(f_k\right)\right|^2$ because the second singular vectors will play a part in one of the cases.

With the above discussion of the mode shape vectors and the eigenvector, it will be quite straightforward to form a density operator in terms of our PSD estimator $\hat{\mathbf{G}}\left(f_k\right)$. The only trick is that at each frequency line, we see the system is in a mixed state. The underlying mode shape vectors are the state vectors of the system, each of which is a linear superposition of the eigenstates, see Eq. (4.72). Thus, if we have r modes, then at each frequency line, a density operator of a mixed state can be given by a linear combination of the pure states formed by the outer product ($\otimes$) of each mode shape vector, namely,

$$\rho_k = \sum_i^r p_{ki}\left(\boldsymbol{\psi}_i \otimes \boldsymbol{\psi}_i\right) = \sum_i^r p_{ki}\boldsymbol{\psi}_i^*\boldsymbol{\psi}_i^{\mathrm{T}} \tag{4.77}$$

where $\left\langle \boldsymbol{\psi}_i, \boldsymbol{\psi}_i \right\rangle = 1$. Note that by normalisation, the orthogonality property of the mode shape vectors remains unchanged.

Since at different frequency line, the contribution of each mode to a random state $\tilde{\mathbf{y}}\left(f_k\right)$ is different, then ρ_k will be different at different frequency locations.

It is clear that we cannot obtain a density matrix via Eq. (4.77) because we do not know about the mode shape vectors. The only thing we have is a PSD estimator. Fortunately, we are able to construct a density matrix out of our PSD estimator $\hat{\mathbf{G}}\left(f_k\right)$. But before we convert $\hat{\mathbf{G}}\left(f_k\right)$ into a density operator, we shall rewrite Eq. (4.77) into the following format

$$\rho_k = \sum_i^r \left(\sqrt{p_{ki}}\boldsymbol{\psi}_i \otimes \sqrt{p_{ki}}\boldsymbol{\psi}_i\right) = \boldsymbol{\Psi}_p^*\boldsymbol{\Psi}_p^{\mathrm{T}} \tag{4.78}$$

where $\boldsymbol{\Psi}_p$ is the probability scaled mode shape matrix.

Now if we revisit Eq. (4.29), and use Eq. (4.70) to replace $\tilde{\mathbf{y}}\left(f_k\right)$, we get

$$\hat{\mathbf{G}}\left(f_k\right) = \frac{1}{Mn_d}\sum_{i=1}^{n_d}\left(\tilde{\mathbf{y}}\left(f_k\right) \otimes \tilde{\mathbf{y}}\left(f_k\right)\right) = \frac{1}{Mn_d}\sum_{i=1}^{n_d}\left(\tilde{\mathbf{q}}_i\left(f_k\right)\boldsymbol{\Psi} \otimes \tilde{\mathbf{q}}_i\left(f_k\right)\boldsymbol{\Psi}\right) = \boldsymbol{\Psi}_q^*\boldsymbol{\Psi}_q^{\mathrm{T}} \tag{4.79}$$

where $\boldsymbol{\Psi}_q$ is the generalised coordinates scaled mode shape matrix.

Compare Eq. (4.79) to Eq. (4.78), we see that the PSD estimator $\hat{\mathbf{G}}(f_k)$ differs from the density matrix ρ_k only by a scaling factor of constant magnitude. Therefore, if we normalise $\hat{\mathbf{G}}(f_k)$ by its trace, then it should lead us to ρ_k, namely

$$\rho_k = \hat{\mathbf{G}}(f_k) \Big/ \mathrm{tr}\big(\hat{\mathbf{G}}(f_k)\big) \tag{4.80}$$

It is worth pointing out that by using the FDD, we see $\hat{\mathbf{G}}(f_k)$ as a Hermitian operator associated with a general observable without a particular physical meaning apart from the average power or variance, which quantifies the vibrational status of a system. While a density matrix measures the mixture of the states of a system. When we use them together, we can make the best use of $\hat{\mathbf{G}}(f_k)$. However, in our close mode analysis, due to the intrinsic restriction of the FDD, i.e., spectral resolution problem and mode shape measuring procedure, we will mainly use the density operator to estimate the mode shapes.

Besides, since the density matrix ρ_k is a scaled version of the PSD estimator $\hat{\mathbf{G}}(f_k)$, and the mode shapes are required to be normalised, then there will be no impact on the estimated mode shapes based on the singular vectors, if we use the density matrix ρ_k to estimate the mode shapes. In other words, when we apply either the Spectral decomposition or SVD to ρ_k, namely

$$\rho_k = \mathbf{U}_k \mathbf{\Lambda}_\rho(f_k) \mathbf{U}_k^{\mathrm{H}} \tag{4.81}$$

only the eigenvalue matrix $\mathbf{\Lambda}_\rho(f_k)$ is different from the one we obtained in Eq. (4.39), i..e, $\mathbf{\Lambda}_\rho(f_k) \neq \mathbf{\Lambda}_k$. But it is also a $d \times d$ diagonal matrix, and $\mathbf{\Lambda}_\rho(f_k) = \mathrm{diag}\left(\begin{bmatrix} \bar{\lambda}_{k1}, \ldots, \bar{\lambda}_{kd} \end{bmatrix}^{\mathrm{T}}\right)$ with non-negative real eigenvalues.

Since $\mathrm{tr}\big(\hat{\mathbf{G}}(f_k)\big) = \mathrm{tr}(\mathbf{\Lambda}_k) = \sum_i^d \lambda_{ki}$, then ρ_k can also be obtained from the Spectral decomposition of $\hat{\mathbf{G}}(f_k)$, that is

$$\rho_k = \mathbf{U}_k \big(\mathbf{\Lambda}_k \big/ \mathrm{tr}(\mathbf{\Lambda}_k)\big) \mathbf{U}_k^{\mathrm{H}} \tag{4.82}$$

This equation gives the relation between the FDD and the density operator decomposition. And it is obvious that $\mathrm{tr}\big(\mathbf{\Lambda}_\rho(f_k)\big) = 1$.

4.3 Application to the nonstationary random process with a case study

In section 4.2, we elaborated on the theoretical background of the techniques we will be using in this chapter. Upon redefining the FDD, we introduced a density operator, which was designated to tackle the mode shapes estimation for two close modes. Follow the order of our theoretical discussion of the two methods, we will consider the FDD in this section and the close modes analysis by using a density operator in section 4.4.

The application of the FDD to the stationary signals is the main argument for the technique ever since it has been proposed. Therefore, in this section, we want to show that the FDD method can be applied to a certain category of the nonstationary random process with a simulation case. Note that this does not imply that the FDD is a technique designed to be used to deal with the nonstationary random process. On the contrary, the use of the Fourier basis has naturally put it in the stationary data analysis technique category. The reason it can be used to deal with some nonstationary processes lies in the behaviour of the spectrum rather than the explanation we gave to the FDD in section 4.2.2.4. However, it is unlikely to explain in-depth the spectral analysis with respect to the nonstationary random process in this chapter, which is also not our purpose. Instead, we will demonstrate the properties of the FDD and its capability to be used to deal with the nonstationary random process with a designed random sequence.

4.3.1 Overview of the nonstationary process analysis

In general, the nonstationary behaviour of a time series is widely seen in nature, e.g., stock market data economics or speech signals. When dealing with nonstationary processes, one of the difficulties is how to set up an adequate asymptotic theory to solve them [146]. As the basis for a general asymptotic theory, the assumption of stationarity guarantees that the increase of the sample size leads to more information of the same kind [146]. Clearly, this is not the case for the nonstationary processes. Information on the behaviour of the process at the beginning of the time interval cannot be inferred from the future state of the process.

On the other hand, the standard Fourier analysis is crippled by the time-dependent spectrum feature of the nonstationary process, as it only preserves the frequency resolution but no time resolution. In order to overcome this problem, many solutions have been developed in the past decades which can represent a signal in the time and frequency domain simultaneously.

Among all the possible time-frequency analysis tools, the Short Time Fourier Transform (STFT) (spectrogram) is a standard method for the analysis of signals whose spectral content is varying. However, the inherent trade-off (also known as the uncertainty principle) between the time and

frequency resolution has made the STFT method, which uses the fixed window length, problematic, especially when the spectral content is changing so rapidly that finding an appropriate short-time window is indefinite since there may not be any time interval for which the signal is more or less stationary [147].

To counteract the shortcomings of the STFT, many variations of it have been proposed over the decades. For instance, since the boundary of the uncertainty principle (best simultaneous resolution of both) is reached with a Gaussian window, a Gaussian window function was introduced to the Fourier transform of the signal, which is known as the Gabor transform [148]. Later, in order to have a fundamental analysis and a clarification of the physical and mathematical idea to understand the time-varying spectrum, various time-dependent spectra definitions for nonstationary processes have been proposed, such as Page's instantaneous power spectra [149], the Wigner-Ville distribution [150], evolutionary spectral [151] and so on.

As an alternative development to improve the spectrogram, the wavelet transform or wavelet analysis [152], [153], [154] has received much attention in recent years. By using a fully scalable modulated window, the signal cutting problem associated with the STFT is resolved. As a result, it can give good time resolution for high-frequency events as well as good frequency resolution for low-frequency events.

Apart from these traditional time-frequency analysis methods, techniques such as the Hilbert-Huang transform, which combines the Empirical Mode Decomposition (EMD) and the Hilbert spectral analysis is designed specifically for analysing data from nonlinear and nonstationary processes [155], or Empirical Orthogonal Function (EOF) expansion (another name of PCA used in meteorology and oceanography) [156], [157] are also widely used to deal with the nonstationary processes. Apart from the aforementioned examples, some special techniques, such as taking differences or considering the data on small time intervals, have been applied to make use of the stationary techniques, and this is known as the ARIMA (Autoregressive Integrated Moving Average) modal time series forecasting [102]. Other methods, such as the Artificial Neural Network (ANN) [158] or the Bayesian approach [159] are also on some occasions.

In general, a process is only stationary or not, and it can either be globally nonstationary or locally nonstationary. A typical example of the global nonstationary random process is a random walk. The process and the parameters of such a global nonstationary process are fixed some time ago (usually infinity), and then it evolves based on a fixed rule. When the parameters of the ARMA (Autoregressive Moving Average) models themselves evolve over time, the random process will show some local nonstationary phenomena. For this type of nonstationary process, the estimation of the parameters is very difficult, especially when the parameters change in an unstructured way.

Technically speaking, time series can be nonstationary in many ways. The aforementioned two types of nonstationary processes imply two ways to produce them. Both of them will be used to generate the

nonstationary process we will be using to test the FDD in the next subsection. Since the FDD is a spectral density analysis technique, which is defined for stationary processes, the application of it to the nonstationary processes is quite limited. Thus, the type of random process must be carefully selected. According to our analysis, it can be used to analyse a certain type of locally stationary processes [160]. The simulation in the next section will demonstrate this statement.

4.3.2 Case study I – nonstationary process analysis

In this section, we will consider a certain type of nonstationary processes that the FDD can be applied to. This type of nonstationary process, if we denote it as $\{q_i(n)\}_{i\in\mathbb{Z}^+}$, needs to have the following property. First, $\{q_i(n)\}_{i\in\mathbb{Z}^+}$ must possess what we can loosely describe as an "oscillatory form", and we can characterise this property by saying that the Fourier transform of such a function will be concentrated around a particular point f_1 (or around $\pm f_1$ in the real case). Particularly, f_1 should not be a function of time. Thus, if we have a non-periodic process $\{q_i(n)\}_{i\in\mathbb{Z}^+}$, whose Fourier transform has an absolute maximum at the point f_1 we may define f_1 as the frequency of this function, the argument being that locally $\{q_i(n)\}_{i\in\mathbb{Z}^+}$ behaves like a sine wave with (conventional) frequency f_1, modulated by a "smoothly varying" amplitude. A good example of Eq. (4.83) is the free vibration responses of signal mode, which can be interpreted as consisting of just two frequency components ($\pm f_1$) with time-varying amplitude.

Now if we translate the aforementioned description of $\{q_i(n)\}_{i\in\mathbb{Z}^+}$ into equation, we can obtain the following expression

$$q_i(n) = a_i(n)u_i(n) \tag{4.83}$$

where $\{u_i(n)\}_{i\in\mathbb{Z}^+}$ is a covariance stationary random process with zero mean and finite variance, which consists at least one oscillatory component. Let f_1 being the lowest centre frequency. While $a_i(n)$ represents the amplitude modulated signal with frequency f_0, which can be deterministic or random.

Thus, in order to observe the centre frequencies carried by $\{q_i(n)\}_{i\in\mathbb{Z}^+}$, the amplitude of the process must varies slowly compared to the lowest frequency of $u_i(n)$. In other words, when the modulating frequency f_0 is far smaller than the centre frequency f_1, the two stationary spectral components at $f_1 \pm f_0$ collapse into a single stationary auto-spectrum [124]. Note that the suffix i in Eq. (4.83)

denotes the number of modes. In this study case, we will only consider three modes, thus, $i = 1, 2, 3$. As for $q_i(n)$, it represents the modal coordinates.

In this case, the stationary autospectrum density $G_{uu}(f)$ can be estimated by the time-averaged autospectral density $\bar{G}_{qq}(f)$ over the nonstationary record $q(n)$. When $\bar{G}_{qq}(f)$ is normalised (i.e., $\int \bar{G}_{qq}(f) df = 1$) [124], we have $\bar{G}_{qq}(f) \approx G_{uu}(f)$, where $\bar{G}_{qq}(f)$ is given by Eq. (4.16). Additionally, if the rate of change of $a_i(n)$ increases with time, the bandwidth of $\bar{G}_{qq}(f)$ will increase relative to $G_{uu}(f)$. However, as long as the difference between f_1 and f_0 is small and it is within the resolution limit, the centre frequency can still be estimated to a reasonable accuracy.

In this study, if we assume $a_i(n)$ is a random process generated by an Autoregressive (AR) process with a fixed parameter. According to our analysis, $a_i(n)$ needs to vary very slow compare to the lowest frequency of $u_i(n)$ in order for the FDD to work. A good candidate is the AR(1) process, which is also known as the random walk, and it varies very slow globally. Therefore, $a_i(n)$ is given by

$$a_i(n) = a_i(n-1) + \varepsilon_i(n) \tag{4.84}$$

where $\{\varepsilon_i(n)\}_{i \in \mathbb{Z}^+}$ is a Gaussian white noise process, i.e., $\mathbb{E}[\varepsilon_i(n)] = 0$, $\mathbb{E}[\varepsilon_i^2(n)] = \sigma_i^2$, $\mathbb{E}[\varepsilon_i(n)\varepsilon_i(m)] = 0$, $\forall n$, $\forall m \neq n$.

On the other hand, according to the Wold's decomposition theorem, any zero-mean covariance stationary time series $\{X_t\}$ can be decomposed as [161],

$$X_t = \eta_t + Z_t \tag{4.85}$$

where $\{\eta_t\}$ is the a linearly deterministic process which is perfectly predictable without a time trend, such as the linear combination of the sinusoidal with constant coefficients. While $\{Z_t\}$ is a MA(∞) (Moving average) process with square summable moving average terms, i.e,

$$Z_t = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} \tag{4.86}$$

where $\theta_0 = 1$, $\sum_{j=0}^{\infty} \theta_j^2 < \infty$, and $\{\varepsilon_t\}$ is Gaussian white noise. Note that $\{\eta_t\}$ is uncorrelated with $\{\varepsilon_t\}$, namely, $\mathbb{E}[\eta_t \varepsilon_s] = 0$, $\forall t, s$.

Therefore, we shall construct a time series $\left\{u_i(n)\right\}_{i \in \mathbb{Z}^+}$ in the following way

$$u_i(n) = b_{i1} \sin\left(2\pi f_i(n-n_0) - \varphi\right) + b_{i2}\mu(n) + z_i(n) \tag{4.87}$$

where

$$\mu(n) = \begin{cases} \cos(2\pi f_\alpha n) & n < n_s \\ 0 & n \geq n_s \end{cases} \tag{4.88}$$

and f_α is assumed to be 20 (Hz). This frequency component only occurs when $n < n_s$. Here, we assume $n_s = 1$ (s). While b_{i1} and b_{i2} are constant coefficients, which will be generated randomly (follow the normal distribution) for each mode in this study. f_r are the frequencies we wish to identify by using the FDD. In this simulation, we assume $f_1 = 1$ (Hz), $f_2 = 2$ (Hz) $f_3 = 18$ (Hz). φ represents the random phase shift. n_0 denotes the random time shift. $\left\{z_i(n)\right\}$ is the Gaussian white noise, which is a covariance stationary random process. Note that $\mu(n)$ is shared by all three modes, and we introduce $\mu(n)$ in Eq. (4.87) is because we want to show the behaviour of the second or maybe higher singular values in the conventional FDD singular value plot when there is a correlated term within some modes,

It is worth mentioning that another reason we can build $u_i(n)$ in such a way, i.e., a combination of the sine and cosine waves and the white noise, is because sine and cosine waves are themselves "stationary", and it is natural that they should form the basic elements used in building up models of stationary processes.

The sequence is sampled in 2 seconds at interval 0.001, which means the sampling frequency f_s is 1000 (Hz). As a result, the Nyquist frequency is equal to 500 Hz. While the mode shapes we will be using in this section will be generated by the following function

$$\psi_i(j) = \sin^2\left(\frac{j\pi}{d+1}\right)\sin\left(\frac{ij\pi}{d+1}\right), \quad j = 1,\ldots,d \tag{4.89}$$

Note that, here we assume there are 10 measurement locations, i.e., $d = 10$, with only three active modes, and the f_a component exist for all three active modes. Now according to the mode superposition technique, the j th measurement is given by

$$y_j(n) = \sum_{i=1}^{3} q_i(n)\psi_i(j), \quad j = 1,2,\ldots,d \tag{4.90}$$

Note that all the frequency components concerned in this model do not have any trend against time. In other words, although they can occur at different time instants, they remain constant over a period of time.

4.3.3 Simulation and results

As we mentioned in the previous section, this study will only take the ensemble averaging of the Periodogram into consideration. Since we are doing simulation, multiple independent observations can be easily achieved. To make such an averaging feasible, the mode shapes and frequencies for each mode to be identified should be the same for all numerical experiments. Additionally, the mode shapes we calculated based on Eq. (4.89) are all real-valued. Thus, we can use the real-valued spectral density matrix approach to compute the spectral density matrix we introduced in chapter 3.

The simulation will run for $N_r = 1, 10, 50, 100$ times. Figure 4.2 gives the simulated responses for N_r =1, as a demonstration. Figure 4.2 (a) illustrates the random walk response within 2 seconds. Figure 4.2 (b) presents the stationary process generated according to Eq. (4.87). Figure 2 (c) gives the generalised coordinates of the second mode given by Eq. (4.83). Figure 4.2 (d) depicts the measured response from channel 5. It is obvious that both the generalised coordinates and the measured responses are nonstationary.

Figure 4.3 illustrates the time-frequency plot of the measurements from channel 5, and it is generated by using the shot time Fourier transform (STFT). 256 point of the Kaiser window was used with a shape factor of 5, and the Fast Fourier Transform (FFT) length is 512 with 220 points overlapping. It is clearly shown from Figure 4.3 (a) that the frequency component f_a only occurs within a certain time, i.e, $n < 1$ (s), and it is observable in Figure 4.3 (a) only when it receives enough energy. While from Figure 4.3 (b), we can see that the amplitude (power) of the signal is changing all the time. Whereas for stationary signals, we shall have a constant power over all time.

As for the calculation of the spectral density matrix, the FFT is applied to all 10 channels' full-length record (i.e., 2 seconds) with a rectangular window first, and then we can calculate the real-valued spectral density matrix as we introduced in chapter 3. Since the duration is only 2 seconds for all the records, the frequency components less than 0.5 (Hz) are unresolvable. Note that the signals generated from all channels are all normalised before the spectral density matrix is formulated. The total energy from each finite observation is used to normalise the multivariate processes. Besides, as we mentioned is the earlier section, the different windowing effects will not be considered for the FDD analysis.

(a) Random walk from $a_2(n)$

(c) Generalised coordinates from $q_2(n)$

(b) Stationary process from $u_2(n)$

(d) Measurememnts from $y_5(n)$
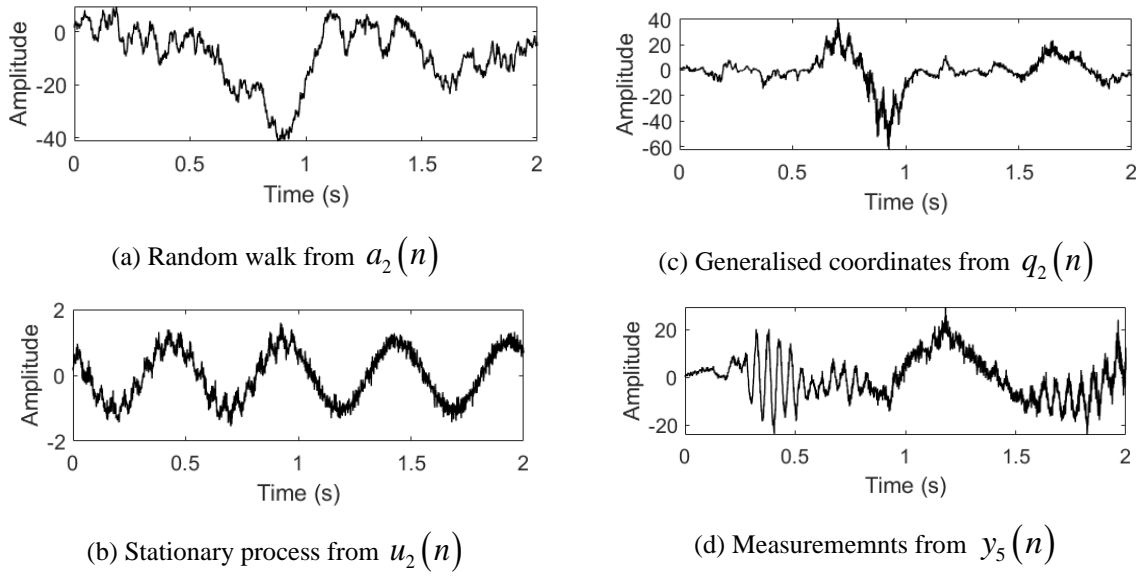
Figure 4.2. Demonstration of the simulated responses

(a) Plan view of the time-frequency plot

(b) Front view of the time-frequency plot
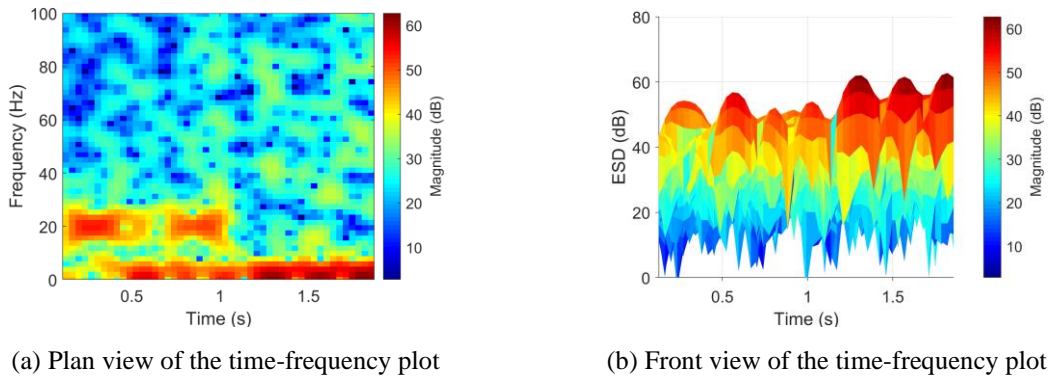
Figure 4.3. STFT of the measurements from $y_5(n)$

Figure 4.4 shows conventional singular value plots with different averaging times. From this figure, we can observe that as the number of averaging times increases, the singular value plots getting smoother and smoother. According to the FDD, to estimate the mode shapes, we first need to identify the location of the peak frequencies. In the case of $N_r = 1$, we can see that the singular value plot is quite zigzag, and it is difficult to observe all three modes we wish to identify. However, there is no direct correspondence between a good frequency estimation and a good mode shape estimation from the FDD. From the corresponding contribution score plot given in Figure 4.5 (a) of case $N_r = 1$, we can see that mode 1 and mode 2 govern a small range of the frequency band, respectively, as the contribution scores for these two modes are close to 1. It means that we can use any of the singular vectors from those frequency bands to estimate the mode shape of the two modes. This explains why we can obtain a good

estimation of the mode shape for the first mode while getting a poor frequency estimation based on the peak frequency of the FDD.

(a) $N_r = 1$. (b) $N_r = 10$. (c) $N_r = 50$. (d) $N_r = 100$.

Estimated frequency $f_1 = 0.5$, $f_2 = 2$, $f_3 = 18$ (Hz)

Estimated frequency $f_1 = 1$, $f_2 = 2$, $f_3 = 18$ (Hz)

Estimated frequency $f_1 = 1$, $f_2 = 2$, $f_3 = 18$ (Hz)

Estimated frequency $f_1 = 1$, $f_2 = 2$, $f_3 = 18$ (Hz)
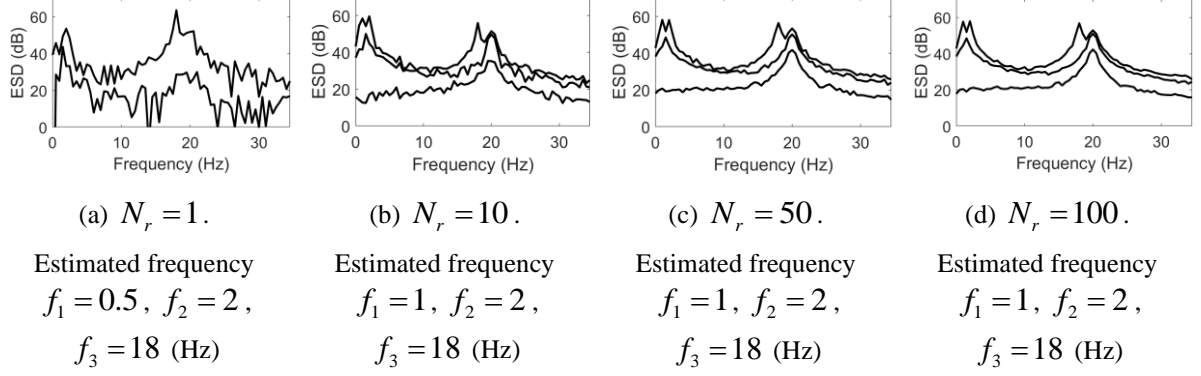
Figure 4.4. Singular value plot for 4 different number of example runs. ESD represents Energy Spectral Density. The estimated frequencies are shown below the figure for each case.

(a) $N_r = 1$ (b) $N_r = 10$ (c) $N_r = 50$ (d) $N_r = 100$
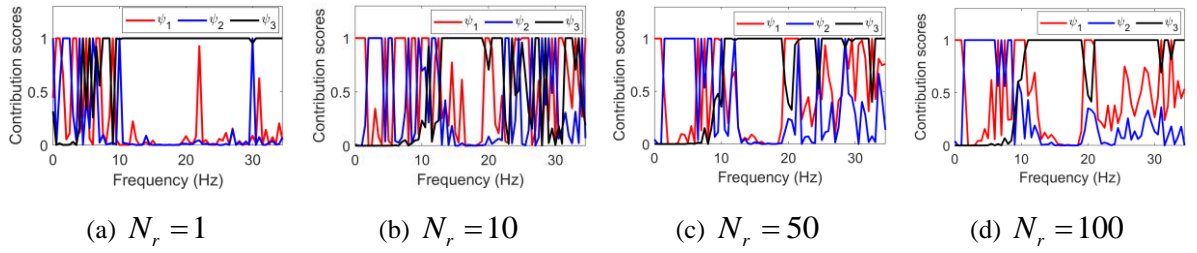
Figure 4.5. Contribution scores plot for 4 different number of example runs

(a) $N_r = 1$ (b) $N_r = 10$ (c) $N_r = 50$ (d) $N_r = 100$
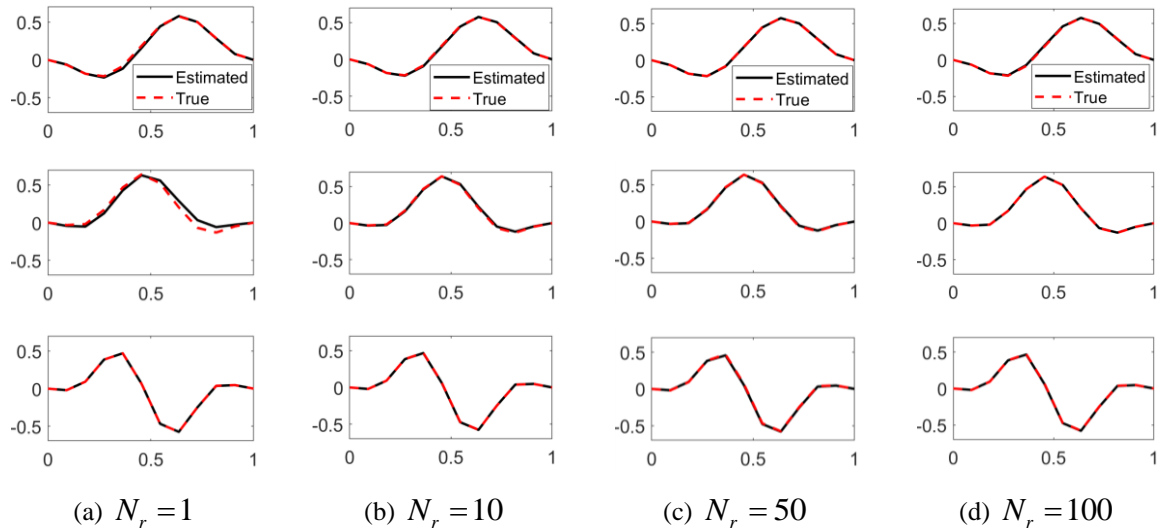
Figure 4.6. Estimated mode shapes for 4 different number of example runs. All the mode shapre are normalised with length euqal to 1

When the averaging times increased to 10, we can see very clearly that there are three peaks in the frequency band [0, 19] (Hz), which are corresponding to the three modes we are looking for. In the

vicinity of a resonant frequency, a certain mode will be governing, then the contribution score of that mode is close to 1. In the meantime, compared to Figure 4.6 (a), we get a better estimation of the mode shapes in Figure 4.6 (b).

As the averaging times keep growing, we see a smoother and smoother singular value plot in Figure 4.4 (c) and (d). Likewise, the dominant frequency band for each mode is getting wider and wider in the contribution score plot in Figure 4.5 (c) and (d). As for the estimated mode shapes, they are getting closer and closer to the true values. It is true from our simulation case that the more averages, the better. However, we as shall observe from Figure 4.4 (c) and (d), and Figure 4.6 (c) and (d), after a certain number of averages, the improvement of the identification accuracy is negligible.

Apart from the aforementioned phenomenon that we can observe from the plots given by Figure 4.4, Figure 4.5, and Figure 4.6, it must be pointed out that we did not deliberately remove the higher order singular values in Figure 4.4. It is only because the rest of the singular values are too small to be shown in the scale of Figure 4.4. Particularly, in Figure 4.4 (a), since we are using the real-valued spectral density matrix approach, this procedure actually gives a one-time average. Thus, we can see two prominent singular value plots in Figure 4.4 (a) rather than one. On the other hand, even if we have 10 channels, and we conducted more than 10 averages, we still obtain 3 prominent singular value plots rather than 10 in Figure 4.4 (b), (c) and (d). It is because we only have three modes, and the large number of averaging does not guarantee a full-rank spectral density matrix. Therefore, in our study case, we shall obtain three prominent singular values when $N_r \geq 2$.

Another phenomenon we must address is the frequency component f_a, which is the correlation term among all three modes. In our previous understanding about the second singular values, the behaviour of which, we treat it as an indication for the close modes. However, when more than one modal coordinate is correlated, then there will be more than one direction in the subspace can explain the data at f_a. As a result, we shall see large higher order singular values at f_a, which has a similar trend as the first one. Apart from this, we can see that even if the third mode $f_3 = 18$ (Hz) is very close to the additional frequency component $f_a = 20$ (Hz), the peak of the second singular value does not occur at the dip point between the two close frequencies. Furthermore, we may conclude that the mode shapes only correspond to the uncorrelated modes given by the first term in $u_i(n)$, and this can be easily verified by observing the contribution scores plot in Figure 4.5. Thus, the conventional understanding of the behaviour of the second singular values only valid for uncorrelated frequency components.

Now that we have experimentally shown the possibility of using the FDD to analysis nonstationary data, we can safely apply the FDD to a similar scenario in chapter 6 to extract the mode shapes from responses of a bridge subject to a moving vehicle excitation under the assumption that the impact of the vehicle

is insignificant and the bridge frequencies do not change with time, only the amplitude of the responses diminish over time. In the next section, we will move on to the other topic of this chapter, namely close modes analysis. Instead of working on nonstationary data, to keep our discussion as simple as possible, the stationary data is used.

4.4 Close modes analysis

4.4.1 Introduction of the close modes analysis

Due to structural symmetries or large-scale repetitive structures with little damping, close modes widely exist in our engineering structures, such as those large flexible aerospace structures or cables on a cable-supported bridge. Accurate determination of the modal parameters of such modes is very important in response and stability calculations [162]. However, compared to the well-separated modes, close modes can bring a lot of uncertainties to our identification results. To understand such uncertainties and to produce better identification result, the close modes problem has been approached by researchers from different perspectives. Nevertheless, we can classify them into three groups.

In the first group, researchers like Brincker and Lopez-Anelle [113] and Au et al. [163]–[166] were trying to understand the uncertainty problem of the close modes. Brincker et al. [167] showed that in the case of a set of two closely spaced eigenvalues, the mode shapes become very sensitive to small changes in the system. Then later, Brincker and Lopez-Anelle [113] investigated the intrinsic uncertainty of an undamped system with 3 DOFs in terms of the perturbation theory [113]. The system they considered is a simple mass sprung system with a set of almost orthogonal mode shapes. They pointed out that in the case of repeated eigenvalues, the undefined individual mode shapes are carried over to the case of closely spaced eigenvalues as an increased sensitivity of the mode shapes when the eigenvalues approach each other. They also showed that the main perturbation of the mode shapes is a rotation in the subspace defined by the initial eigenvectors.

If Brincker and Lopez-Anelle [113] explored the uncertainty problem of close modes via forward thinking, then Au et al. [163]–[166] approached it from the identification point of view. Zhu et al. [163] studied the identification uncertainty of closely-spaced modes based on a multi-mode model with Fast Fourier transform data on the same frequency band. They found that with the increase of the signal-to-noise ratio, the identification uncertainty under different MAC values converges to the uncertainty law of well-separated modes. Later, Au and Brownjohn [164] investigated analytically the identification uncertainty of close modes, especially the mode shapes, under asymptotic conditions of long data and high signal-to-noise ratio based on a Bayesian approach. By working on the eigenvalue properties of a Fisher information matrix, they revealed that the mode shape uncertainty occurs in two characteristic

types of mutually uncorrelated principal directions, one perpendicular and one within the mode shape subspace spanned by the mode shapes. Whereafter Au et al. [165], [166] developed the 'uncertainty law' for close modes, which dictates the achievable precision of modal properties regardless of the identification algorithm used up to modelling assumptions and the use of probability. In their work, they revealed a fundamental definition that quantifies the closeness modes and demystifies the roles of various governing factors.

In the second group, researchers tend to improve the estimation results via different methods. Depending on whether or not the measurement is intact, we can further divide those methods into invasive and non-invasive methods. In the invasive category, the fictitious modification is often applied to the test data, and such method is widely seen in modal testing. Usually, the receptances of a symmetric structure are processed using fictitious mass or stiffness modifications. With the modified measured receptance data, the two close mods involved will become somehow separated, and hence their associated modal parameters can be analysed accurately by using the conventional modal analysis techniques. Both Lin and Lim [162] and Mottershead et al. [168] have validated such an approach.

While in the non-invasive category, researchers choose to search for or develop new techniques to extract the modal information of the close modes as accurate as possible. The work from Brincker and Lopez-Anelle [113] and Au et al. [163]–[166] gave us significant insight into how the modal parameters behave when two modes are getting closer to each other. However, in order to quantify to what extent an identification method can accurately identify the close modes modal parameters is a question, as it is usually the case that for the case of two modes which are close but not very close, many of the existing identification methods can be quite effective.

Wu et al. [169] used the following mode separation parameter

$$\gamma_j = \frac{2\left(f_j - f_i\right)}{f_j + f_i} \tag{4.91}$$

proposed by Lin and Lim [162] to compare his work with many others' results, and this parameter is also used by Chen [170]. Here, i, j represent the index of the modes, and we assume $j = 2$, $i = 1$.

However, Eq. (4.91) is too simply to truly reflect the closeness of two modes, as it only takes the difference in frequency into account but ignored the effect of system damping and potentially the spectral resolution. It is well known that damping is also an important influence factor in identification. For modes with the same damping ratio, according to half-power bandwidth, i.e., $\Delta f = 2 f \xi$, the higher the frequencies, the wider the shape of modes will spread over the frequency band. Therefore, in order to take the damping ratio into consideration, Srikantha Phani and Woodhouse [171] defined a modal overlap factor, namely

$$\gamma_j = \frac{f_j \xi_j}{\left(f_j - f_i\right)} \tag{4.92}$$

and this factor is further adopted by Rainieri and Fabbrocino [172] and [173].

Apart from the aforementioned closeness index, Zhu et al. [163] proposed a proximity index, which is essentially Eq. (4.91) divided by the damping ratio of the two modes, where they assumed the two modes has the same damping ratio. Later, Au et al. [165] suggested a new formula to encapsulate differences in frequencies and damping ratios in an overall sense, i.e.,

$$\gamma_j = \sqrt{\gamma_f^2(j) + \gamma_\xi^2(j)} \tag{4.93}$$

where γ_f and γ_ξ are the frequency and damping disparity parameters, respectively, and they have the following expressions

$$\gamma_f(j) = \frac{f_j \xi_j}{f_i \xi_i} - 1 \quad \text{and} \quad \gamma_\xi(j) = \frac{f_j - f_i}{f_i \xi_i} \tag{4.94}$$

Note that the term 'disparity' is used here to describe the γ_f and γ_ξ, as they are not simple difference of modal properties. And a disparity of 0.5 is considered to be very close, 1 is close, 5 is separated, and 10 is well-separated [165].

As for the spectral resolution, it is a data measuring and processing issue not a system problem. It is beneficial to know it when comparing two methods in the frequency domain. However, when only the time domain methods were used to do the identification, the spectral resolution may not be provided by the authors. Therefore, in this work, we adopt Eq. (4.93) as the closeness index to quantify the goodness of our proposed method.

In Table 4.1, we listed some of the methods used by different authors to tackle the close modes identification problem, and quantify the closeness of two modes in their experimental model in terms of Eq. (4.93). Note that we only selected the data which have a control group in each reference paper. In other words, the frequencies and damping ratios in Table 4.1 are the exact or theoretical values for their experiments and by using their proposed methods, they claimed that they can successfully identify the close modes modal parameters. The reason that we did not use their real measurement results is that some of the identification methods may give consistent biased estimates, especially for the damping ratios. We will show this phenomenon in our study. Therefore, it is undesirable to do the comparison based on their estimated results.

It is worth mentioning that Table 4.1 only gives an incomplete statistic about the work people are doing in the field to tackle the close modes problem, and each method has its advantages and disadvantages.

If one wants to know more about each method listed in Table 4.1 and how accurate they are, one should refer to each paper.

Table 4.1 Basic settings for the simulation (variant parameters)

No.	References	Identification methods	Natural frequency (Hz)		Damping ratio (%)		Eq. (4.93)
			f_1	f_2	ξ_1	ξ_2	
1	Yan and Miyamoto [174]	Wavelet-based method	1	1.06	1	1	3.344
2	Tan et al. [175]	Continuous wavelet transform (CWT) and patter search	1	1.1	3	2	6.000
3	Bao et al. [176]	Improved Hilbert–Huang transform (HHT)	1.09	1.15	1	1	5.505
4	Li et al. [177]	Improved HHT with the application of blind source separation (BSS), singular spectrum analysis and automatic moving-window	0.3425	0.3713	2	2	4.205
5	Kim and Chen [178]	Wavelet packet decomposition and complex envelope displacement analysis	1.180	1.304	1.11	1.23	9.470
6	Wang and Chen [179]	Analytical mode decomposition theorem and random decrement technique	0.175	0.194	3	3	3.621
			0.511	0.572	3	3	3.981
7	Wang et al [180]		1.180	1.304	1.11	1.23	9.470
8	McNeill and Zimmerman [181]	Blind modal identification by adapting blind source separation techniques, whitening and joint approximation diagonalisation	0.342	0.371	2	2	4.241
9	Yang and Nagarajaiah [182]	Novel blind source separation technique learning rule complexity pursuit	0.3425	0.3713	0.2324	0.2143	36.182
10	Yang and Nagarajaiah [183]	Novel blind source separation technique termed spare component analysis	0.3425	0.3713	2	2	4.205
			0.3425	0.3713	3.02	2.79	2.784
11	Guo and Kareem [184]	Time-frequency blind source separation	0.2	0.205	1	1.5	2.557
12	Brincker et al. [52]	Frequency Domain Decomposition (FDD)	55.055	55.121	0.72	0.72	0.167

From Table 4.1, we can see that methods, such as wavelet transform, HHT, and BSS, which are popular ones to be used to identify modal parameters from the nonstationary responses of a system, are also

very popular in the community to be used to identify close modes. However, according to the closeness index we adopted in this work, we can see that, except for the last one from Table 4.1, all cases should not be classified as very close. However, it is still inappropriate to conclude the goodness of the listed methods based on a single closeness index without considering the different systems or data each author used in their research. To compare the performance of each method in dealing with close modes identification, we should apply those methods to the same data set under the same condition.

Besides, although all methods mentioned in Table 4.1 belong to the class of non-parametric methods, FDD is built to deal with stationary data. It is not to say that the other three are more powerful than the FDD, as each method has its own value. Actually, it is more important for us to adopt an appropriate method to deal with different situations. For example, if we are dealing with stationary measurements, we will not get extra benefits from using the wavelet transform. On the contrary, we may over complicate the problem by investing lots of energy in finding an appropriate mother wavelet to decompose the signal onto different bases, in the meantime, sacrifice some frequency precision to get trivial temporal information. Apart from this, when implementing the discrete wavelet transform, one has to overcome drawbacks, such as shift sensitivity, poor directionality, and lack of phase information problem [185], that come along with this technique.

In addition to the above discussion, very few references listed in Table 4.1 have any discussion about the mode shape identification. Although Wang et al. [180], Yang and Nagarajaiah [183], Guo and Kareem [184], and Brincker et al. [52], etc. may present their mode shapes identification results in terms of the actual mode shape values or the Modal Assurance Criterion (MAC) values [110], [186], none of them have addressed any mode shape estimation problems associated with close modes. Particularly, the mode shapes, which are given by Wang et al. [180] and Brincker et al. [52], are almost orthogonal.

Except for the non-parametric methods listed above, it was reported that parametric methods such as the Natural Excitation Technique Eigensystem Realization Algorithm (NExT-ERA) [173], Stochastic Subspace Identification (SSI) algorithms [169] could also give satisfactory results. Unlike those non-parametric methods, all parametric system identification techniques require at least one user-defined integer, i.e., the model order, which equals the number of eigenvalues present in the model, and hence twice the number of eigenfrequencies in theory [187]. Other methods, such as the one based on the Bayesian statistics proposed by Au [188], [189] was also applied to the case of close modes [163], [164].

As for the third group, we identify it to be the assessment problem for the estimated mode shapes of the close modes. Unlike the natural frequencies and damping ratios, the indefiniteness of the mode shapes of two modes when they are extremely close to each other adds an extra level of difficulty for us to quantify the accuracy of our estimates. It was argued by D'Ambrogio and Fregolent [190] and Brincker and Lopez-Anelle [113] that the classical MAC should not be applied to closely spaced mode cases because of the increased sensitivity of the mode shapes, but instead a generalised measure correlating

the subspace should be used. However, in this work, the MAC will still be used, as we want to test how robust our proposed method can identify the individual mode shape vectors for the close modes case compared to other promising methods, such as Polyreference [63] and SSI-DATA [58].

In the next section, we will first present the basic settings for the numerical model we adopted in this work, then give a brief explanation for the identification methods we want to assess in section 4.2.2.2. Having done this, we will elaborate on two sample runs so as to explain how to use the proposed method to do mode shape identification for two close modes in section 4.4.2.3. After this, we shall show our analysis results in terms of 8 difference cases in section 4.4.2.4. Finally, we will give a further discussion about the close mode analysis regarding the case of repeated poles.

4.4.2 Case study II – close modes analysis

4.4.2.1 Numerical experiment setup

The numerical model in our close mode analysis is similar to the one we used in chapter 3, where we have a dynamic system that generates stationary acceleration responses. However, instead of having 5 DOF's, we consider a system with 10 DOF's, where three modes are active, while the rest of them are inactive (i.e., by forcing the corresponding mode shapes to be zero). For the three active modes, two of them have fixed frequencies, while the other one is moving between them from the lower frequency to the higher frequency.

To be more specific, suppose f_1, f_2, f_3 are the frequencies correspond to the three active modes, and $f_1 \leq f_2 \leq f_3$. We assume the first frequency f_1 and the third frequency f_3 are fixed, while the second frequency f_2 moves along the frequency line in the direction from f_1 to f_3. For the 7 inactive modes, the corresponding frequencies are randomly generated between $0.25 f_1$ and the Nyquist frequency f_{Ny}. The number of measurement channels is assumed to be 10.

Table 4.2 Basic settings for the simulation (fixed parameters)

Sample interval	$\Delta t = 0.2$
Data length for FFT	$\text{Nfft} = 1024$
Number of segments.	$n_d = 32$
Total data points	$N = 1024 \times 32$
Number of runs	$N_r = 100$
Nyquist frequency	$f_{N_y} = 1/2\Delta t = 2.5$
Modal mass	1 (kg)
First resonance frequency	$f_1 = 0.25 f_{N_y} = 0.625\,\text{Hz}$
Third resonance frequency	$f_3 = 0.75 f_{N_y} = 1.875\,\text{Hz}$
Second frequency	$f_2 \in [0.625, 0.6445]$ Hz which is ranging from f_1 to f_3.
Noise modes	Randomly generated with frequencies between $0.25 f_1$ and f_{Ny}.

Table 4.3 Basic settings for the simulation (variant parameters)

MAC	MAC_1	$\text{MAC\_1}(\psi_1, \psi_2) = 0.73$, $\text{MAC\_1}(\psi_1, \psi_3) = 0$, $\text{MAC\_1}(\psi_2, \psi_3) = 0$
	MAC_2	$\text{MAC\_2}(\psi_1, \psi_2) = 0$, $\text{MAC\_2}(\psi_1, \psi_3) = 0$, $\text{MAC\_2}(\psi_2, \psi_3) = 0$
Damping ratio	Damp_1	$\xi = 0.1\%$ for all 10 modes
	Damp_2	$\xi = 1\%$ for all 10 modes
Noise level	Noise_1	1% of the maximum standard deviation of the 10 channels
	Noise_2	10% of the maximum standard deviation of the 10 channels

Table 4.2 and Table 4.3 give basic settings for the simulation. In terms of the correlation relationship between the two close modes, we divide the possible combinations of the variant parameters in Table 4.3 into two categories, namely, MAC_1 and MAC_2. In each category, we consider 4 cases. All the possible combinations are listed in Table 4.4. In the case of MAC_1, the mode shapes of the two close modes are highly correlated to each other, but they are orthogonal to the third mode. Whereas in the case of MAC_2, the mode shapes of all three modes are orthogonal to each other.

Table 4.4 Cases to be considered

MAC_1	MAC_2
(a1) (MAC_1, damping_1, noise_1)	(b1) (MAC_2, damping_1, noise_1)
(a2) (MAC_1, damping_2, noise_1)	(b2) (MAC_2, damping_2, noise_1)
(a3) (MAC_1, damping_1, noise_2)	(b3) (MAC_2, damping_1, noise_2)
(a4) (MAC_1, damping_2, noise_2)	(b4) (MAC_2, damping_2, noise_2)

4.4.2.2 Identification methods

We will consider essentially three methods in our case study, Enhanced Polyreference, Polyreference, and SSI-DATA. In section 2.4.1.2 and section 2.4.3.2, we have presented the theoretical background for Polyreference and SSI-DATA, respectively.

As for the Enhanced Polyreference, it is just an easy name we give it to the two-stage identification procedure to identify the close modes modal parameters. It is well known that by using FDD alone, we are unable to identify the damping ratios. The density operator we proposed in this work also suffers from this problem. In other words, we need to take a further step to obtain the damping ratios. Therefore, we designed a two-stage method, which we call Enhanced Polyreference, to help us to extract the natural frequencies, damping ratios, and mode shapes from our measurements. As its name suggests, the further step we choose to identify the damping ratios and natural frequencies is the Polyreference method. Thus, the Enhanced Polyreference is a two-stage method. In the first stage, we use the proposed density operator to identify the mode shapes, then use the identified mode shape to decouple the system into a series of generalised coordinates. In the second stage, the Polyreference is applied to the generalised coordinates to extract the natural frequencies and damping ratios. Note that one is free to choose any other methods in the second stage to identify the natural frequencies and damping ratios.

Besides, the word 'Enhanced' also implies that this two-stage method is better than the conventional Polyreference method. In the conventional Polyreference method, identifying the physical poles from a set of eigenvalues has always been a challenge in the absence of reference information. Such a problem is quite common in any time domain methods relying on the eigenvalue decomposition of a companion matrix of a system. In other words, the SSI-DATA will be suffering from this problem as well. However, by using the proposed Enhanced Polyreference, we can effectively avoid such a problem since the number of poles can be determined in the first stage in the frequency domain, and in the meantime, we can obtain an estimation of the mode shapes of a system.

Since the singular value plot of a PSD estimator against the frequencies provides us a convenient way to detect the presence of the close modes, we could use the FDD as an auxiliary technique to detect the

close modes, then use the density operator to estimate the mode shapes. Once we obtained the mode shapes for all possible modes, we can use them to decouple the original measurements or geometric coordinates into a set of generalised coordinates. Unlike the original geometric coordinates, which are the superposition of the generalised coordinates, each generalised coordinate is a vibrating response for each mode. Then when we apply the conventional Polyreference method to the generalised coordinates, there will be no superfluous eigenvalues left, which largely simplifies the problem. However, it is worth pointing out that although it is important to obtain a good mode shape estimation in the first stage, unfortunately, it does not guarantee good estimations for the natural frequencies and damping ratios in the second stage. Results in section 4.4.2.4 clearly demonstrate this phenomenon.

As we just mentioned, for the Polyreference and the SSI-DATA, we need to select the physical poles from a bunch of eigenvalues that are contaminated by the spurious poles. However, there is no easy way to do this. One of the most widely used technique is to use a stabilisation diagram [58], which is a model order vs. eigenfrequency diagram. The stabilisation diagram is constructed based on the fact that with the increase of the model order, the physical modes of the system appear at nearly the same eigenfrequency, whereas the spurious modes do not [191]. Based on this principle, a set of parametric models can be identified for a wide range of model orders. Typically, model orders are larger than the number of modes in the considered frequency band. When we plot the modes of all these models in a stabilisation diagram, the physical modes should then show up as vertical lines in this diagram, while the spurious modes tend to show scattering values as frequency or damping values shift all around.

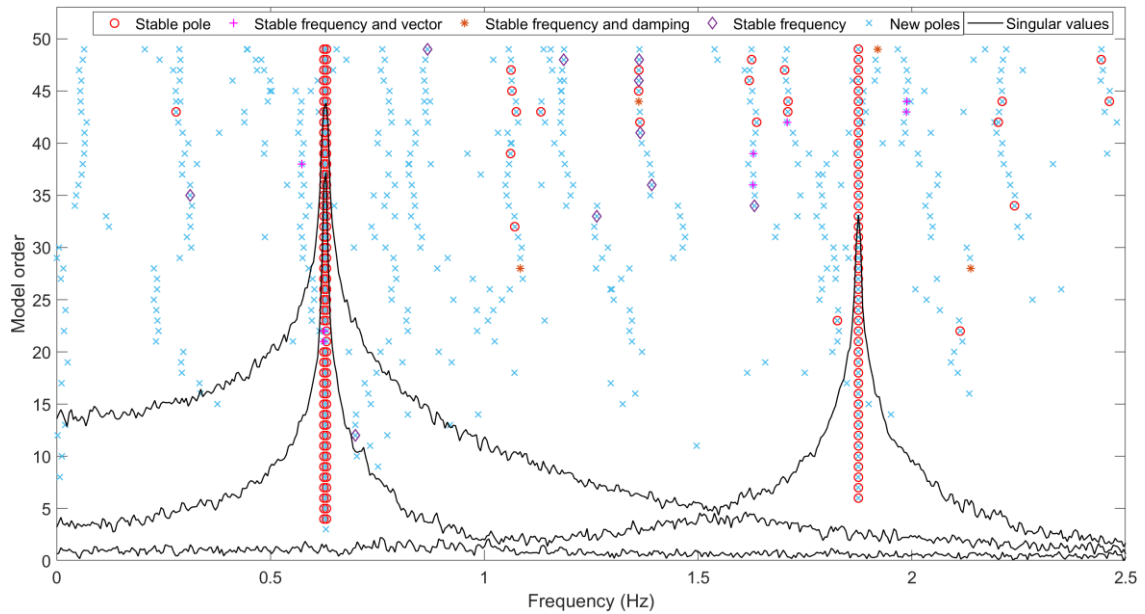
Figure 4.7 Stabilisation diagram

Figure 4.7 shows a typical stabilisation diagram generated based on SSI-DATA from one of the (a1) (MAC_1, damping_1, noise_1) cases. The model order ranges from 3 to 50. Conventionally, we define the different poles based on some stability criteria for each of the modal parameters of interest. In this

work, the criteria are 1% for frequency stability, 5% for damping ratios, and 2% for eigenvectors (MAC) [191]. Based on these criteria, different types of poles are defined as follows:

- Stable pole (frequency, damping, and vector): poles with frequency, damping, and eigenvector deviations within the criteria.
- Stable frequency and vector: poles with frequency and eigenvector deviations within the criteria.
- Stable frequency and damping: poles with frequency and damping deviations within the criteria.
- Stable frequency: poles with only frequency deviations within the criteria.

From Figure 4.7, it is not difficult to see that the SSI-DATA successfully identified two close modes and one well-separated mode.

Although the stabilisation diagram has become a standard method in modal analysis, the selection of physical modes as columns in the diagram is often not straightforward. In other words, the results may depend on the judgement of the analyst, and possible additional validation criteria may be needed [192]. In order to make the interpretation of the stabilisation diagram more effective, researchers in the past two decades have been trying to apply all kinds of clustering methods to the identification process as automated as possible. In the work of Reynders et al. [187], the authors gave an overview of many clustering methods people used to interpret the stabilisation diagram. In this work, we adopted the one proposed by Magalhães et al. [193], which is a simple hierarchical clustering procedure.

In the first step, the similarity between all the pairs of stable estimated modes is calculated based on the following distance measure,

$$d_{ij} = d\left(f_i, f_j\right) + 1 - \text{MAC}\left(\varphi_i, \varphi_j\right) \tag{4.95}$$

where

$$d\left(f_i, f_j\right) = \frac{\left|f_j - f_i\right|}{\max\left(\left|f\right|_i, \left|f_j\right|\right)} \quad \text{and} \quad \text{MAC}\left(\varphi_i, \varphi_j\right) = \frac{\left|\varphi_i^* \varphi_j^{\text{T}}\right|^2}{\left\|\varphi_i\right\|_2^2 \left\|\varphi_j\right\|_2^2} \tag{4.96}$$

f_i and φ_i represent the eigenfrequencies and eigenvectors, respectively, are the estimates of the natural frequencies and mode shapes. If d_{ij} is small, it means the i th and the j th estimates present similar natural frequencies and mode shapes. Hence, it is likely that they represent the same physical mode, then they should be included in the same cluster.

In the second step, the single linkage method is used to compute the distance between clusters. With this option, the distance between two clusters is defined to be the shortest distance between any point in one cluster to any point in the other [193]. In the third step, a limit distance of 0.02 is used to select the Hierarchical tree cut level, i.e., define the number of resulting clusters [193].

Although the clustering methods proposed by Magalhães et al. [193] can generate the identified the modal parameters automatically, the analysis will not always be successful. It means that there will still be spurious modes left in the final results. Besides, it contains parameters that need to be specified or tuned by the user, such as the cut off distance. Apart from this, different linkage methods can also give different results as well, e.g., when the two modes are close together, some linkage methods may not be able to separate the two clusters. Thus, in order to produce a stable result to compare with our proposed method, the estimated results were further compared to the exact values to select the most closed ones. The criteria for this last stage selection is simply based on the MAC. The estimation results obtained via the Polyreference and the SSI-DATA were fed to the same selection program.

4.4.2.3 Mode shapes selection criteria

In the section, we introduce the mode shape selection criteria in terms of the correlation relationship between two modes according to the purity or the Von Neumann entropy plot. We will separate our discussion into two main cases, and one is MAC_1, the other is MAC_2. For these two cases, we will use the lower damping damp_1 and lower noise level noise_1 as examples to demonstrate the selection criteria. Figure 4.8 compares the conventional singular value plot obtained from FDD (Eq. (4.39)) with the entropy plot obtained based on Eq. (4.69) and the purity plot obtained in terms of Eq. (4.66). The first column depicts the full view of each plot, while the second column shows a close shot around the two close modes.
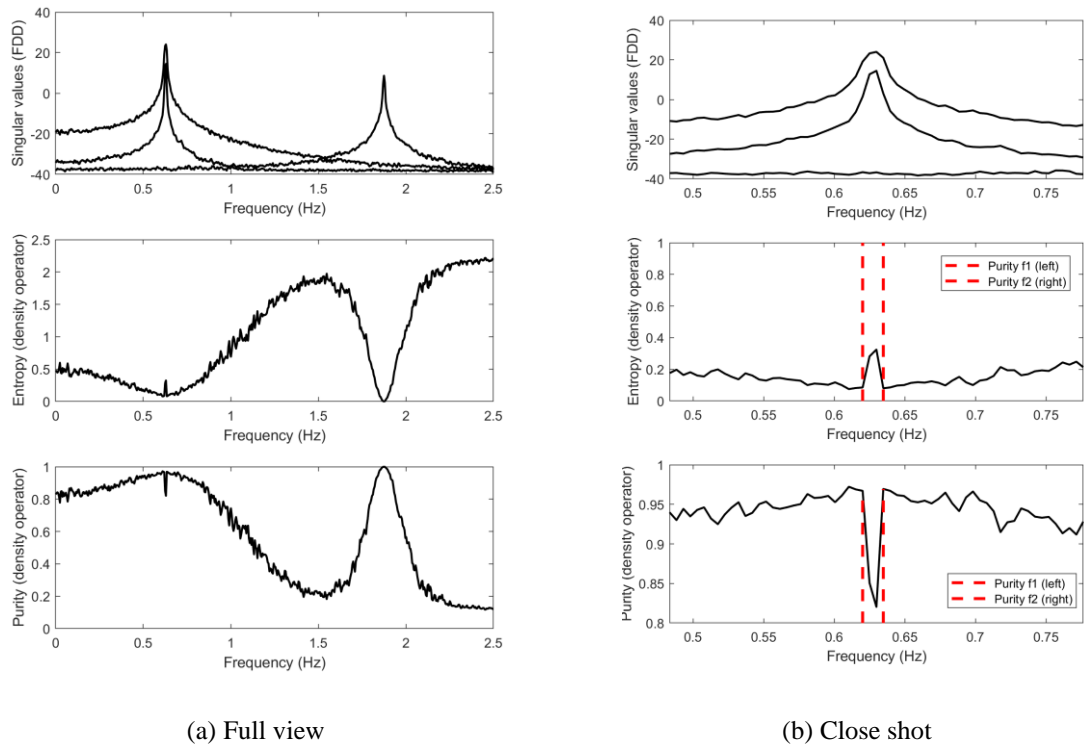

(a) Full view   (b) Close shot

Figure 4.8 Singular value plot (top), Von Neumann entropy plot (middle) and Purity plot (bottom) of (a1) (MAC_1, damping_1, noise_1)

The singular values plot produced by the FDD given by the first row of Figure 4.8 clearly shows that there are two close modes since the second singular values are very large in the vicinity of the peak value of the first singular value plot. However, in the close shot of the singular value plot, we can see that there is only one peak in the first singular value plot. Clearly, if we still want to use the FDD to estimate the mode shapes for the two close modes, we will be in trouble.

Figure 4.9 gives the hindsight information of how the first three singular vectors contribute to the three mode shape vectors we are concerned of based on Eq. (4.72). Note that the acquisition of Figure 4.9 is depended on the prior information of the mode shape vectors, without it, it is unattainable. From Figure 4.9 we see that when there is a pair of close modes, ($\boldsymbol{\psi}_1 \_ \mathbf{u}_{k1}$, $\boldsymbol{\psi}_1 \_ \mathbf{u}_{k2}$) and ($\boldsymbol{\psi}_2 \_ \mathbf{u}_{k1}$, $\boldsymbol{\psi}_2 \_ \mathbf{u}_{k2}$) show some reversed pattern, where $\boldsymbol{\psi}_i \_ \mathbf{u}_{kj}$ represents the probability to find the i th mode shape in the direction of the j th singular vector at frequency f_k, which is the same as $\left| \beta_j \left(f_k \right) \right|^2$ in Eq. (4.72). We can also see that $\boldsymbol{\psi}_1 \_ \mathbf{u}_{k1}$ is large, $\boldsymbol{\psi}_2 \_ \mathbf{u}_{k1}$ is large as well, meaning that $\mathbf{u}_{k1}$ is almost equally important to $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$. Near the peak frequency of the FDD where close modes occur, $\boldsymbol{\psi}_1 \_ \mathbf{u}_{k1}$ and $\boldsymbol{\psi}_2 \_ \mathbf{u}_{k1}$ reach their maximum at different frequency lines, which are marked by the two red dashed vertical lines. These two vertical lines are the same as the vertical lines shown in Figure 4.8, and they can be easily obtained by observing the Von Neumann entropy plot or the purity plot.
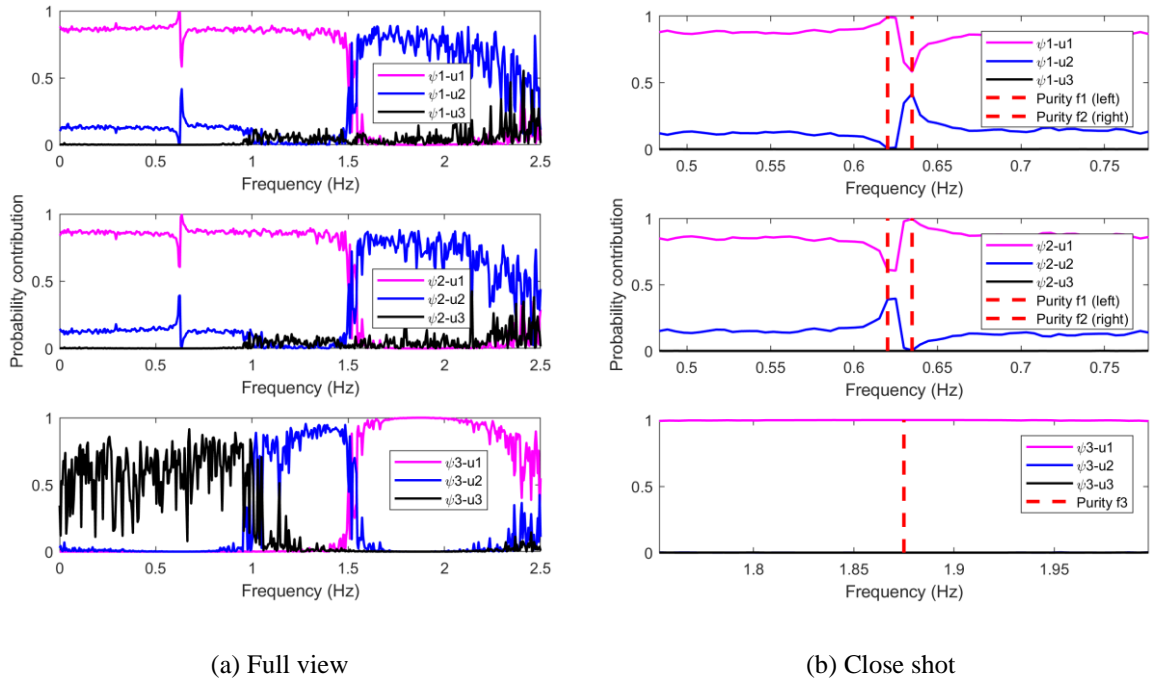

(a) Full view   (b) Close shot

Figure 4.9 Probability of the first three singular vectors contributed to mode 1 (top), mode two (middle) and mode 3 (bottom) of (a1) (MAC_1, damping_1, noise_1)

In the Von Neumann entropy plot, we can see a small peak between the two vertical lines, while in the purity plot, we see a dip. As we mentioned in section 4.2.2.5, the purity and the Von Neumann entropy behaves in an opposite way. When the system is in a pure state, the entropy will approach zero, while the purity approaches 1. In the vicinity of the 3rd mode, there is no other mode close to it. The behaviour of the entropy and the purity is clearly depicted. However, such a phenomenon is not prominent around the two close modes. On the contrary, in a wide range of frequency band, the entropy line is close to zero, while the purity line is close to one. The main reason for this is that the mode shapes of the two close modes are highly correlated, and they behave in a way almost like a single mode. Nevertheless, we can still see a small change in the entropy and the purity around the peak frequency of the two close modes. Note that without specification, we refer the peak frequency to the peak in the first singular value plot from the FDD. When we need to define other peak frequencies, we will specify them accordingly.

It seems that with the increase of the frequency, the system gradually approaches a pure state at the left vertical line, then somehow, due to the presence of close modes, it left the pure state and arrives at a mixed state; after that the system decorrelates again at the second vertical line. While are other frequencies, for instance, around 1.5 (Hz), the system has large entropy and small purity, which means that system is in a mixed state. Unlike in the frequency band where the two correlated modes are mixed, the mixture around 1.5 (Hz) is mainly caused by two orthogonal modes, namely, mode 2 and mode 3. One can tell from the middle and the bottom in Figure 4.9 that around 1.5 (Hz), the system is mainly governed by mode 2 and mode 3, as $\mathbf{u}_{k1}$ has a dramatic change from $\mathbf{\psi}_2$ to $\mathbf{\psi}_3$, see the behaviour of $\mathbf{\psi}_2 \_ \mathbf{u}_{k1}$ is large, $\mathbf{\psi}_3 \_ \mathbf{u}_{k1}$ around 1.5 (Hz).

Therefore, to estimate the mode shapes for each mode, we should use the singular vectors that correspond to the frequency locations of those two vertical lines. Besides, since the entropy plot and the purity plot are equivalent, we can use either of them to do the identification. In this work, the purity is preferred because it is easier to find peaks than dips. And the location of those two vertical lines for the two close modes is identified as the highest purity values occur around the dip.

As a comparison, Figure 4.10 and Figure 4.11 depict the purity plot for the cases of (a2) (MAC_1, damping_2, noise_1) and (a3) (MAC_1, damping_1, noise_2) with the same system parameters as Figure 4.8. It is not difficult to see that when we increase the damping ratio, the dip between the two vertical lines disappeared, whereas when we increase the noise level, the dip remains. This phenomenon shows that the proposed method is noise insensitive. In the meantime, unlike the case of (a1) (MAC_1, damping_1, noise_1) and (a2) (MAC_1, damping_2, noise_1), the purity plot of (a3) (MAC_1, damping_1, noise_2) has very sharp peaks around the resonance frequencies. It implies that the noise level has an impact on the purity of a system along the frequency lines. And in the presence of a high noise level, the pure state of a system can only exists in a very short frequency band.
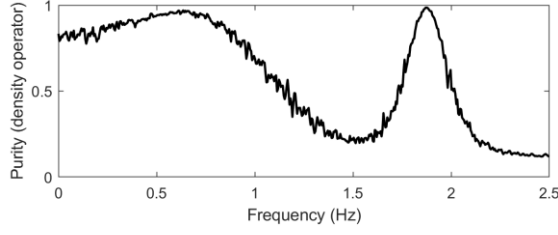
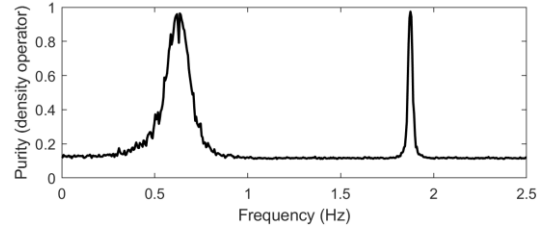Figure 4.10 Purity plot of (a2) (MAC_1, damping_2, noise_1)

Figure 4.11 Purity plot of (a3) (MAC_1, damping_1, noise_2)

As for the case when the mode shapes of two close modes are orthogonal to each other, the singular value plot obtained from the FDD shown in Figure 4.12 gives a similar result as its counterparts in Figure 4.8 that there is only one peak in the first singular value plot where there are two close modes. However, when we compare the two full views of the singular value plot, we see that in Figure 4.8, the second singular values are smaller than the second singular values in Figure 4.12. In other words, the spacing between the first and the second singular values behaves differently for these two cases, namely MAC_1 and MAC_2. Accordingly, the Von Neumann entropy plot and the purity plot of these two cases are different from each other. In the case of MAC_2, the entropy plot shows dips around the peak frequency, while the purity plot shows large peaks around the peak frequency. Nevertheless, when the system approaches a pure state, its Von Neumann entropy will be close to zero, while its purity will be close to 1. Therefore, the criteria for finding the right frequency locations to estimate the mode shapes is essentially the same as we mentioned before for the MAC_1 case, which is the place where the system approaches a pure state with its purity close to 1. In the close view of Figure 4.12, we have clearly marked those locations for two close modes with red dashed vertical lines.

As for the probability contribution of the singular vectors to each mode shape, from Figure 4.13 we see that in the lower frequency band, the first two singular vectors are constantly oscillating to be in line with either ψ_1 or ψ_2. This phenomenon only occurs when the mode shapes of two modes are almost orthogonal. Between the two vertical lines, we can see from the close view of Figure 4.13 that there is an exchange of the probabilities of the first two singular vectors. As we know, the location of the vertical line corresponds to the location of the peak in the purity plot. Thus, the left vertical line indicates that the $\mathbf{u}_{k1}$ can fully explain ψ_1 because the probability $\psi_1 \_ \mathbf{u}_{k1}$ is close to 1. In the meantime, at the same frequency location, $\psi_2 \_ \mathbf{u}_{k2}$ is close to 1, meaning that we can use the first singular vector at the left vertical line to estimate the first mode while using the second singular vector at the same frequency location to estimate the second mode. This is the biggest contrast of the behaviours of the singular vectors between the two mode shape cases, i.e., MAC_1 and MAC_2.
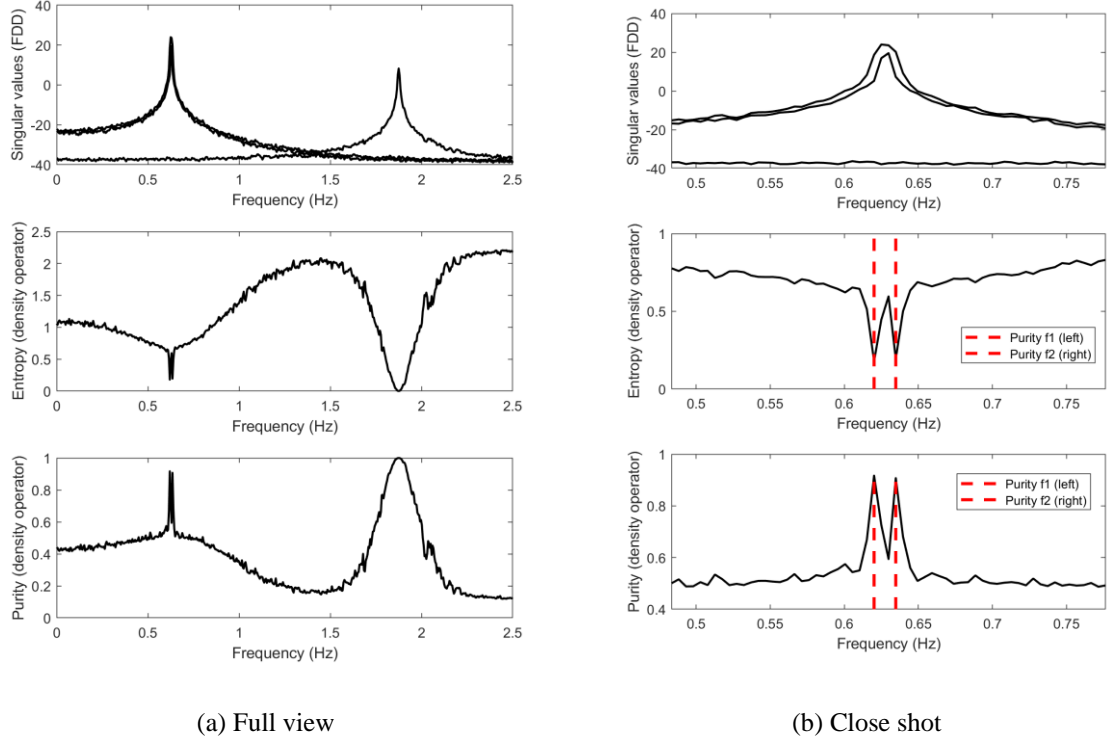
(a) Full view (b) Close shot

Figure 4.12 Singular value plot (top), Von Neumann entropy plot (middle) and Purity plot (bottom) of (b1)
(MAC_2, damping_1, noise_1)

In the case of MAC_1, if we use the first two singular vectors at the same frequency location, for instance, at one of the vertical lines, to estimate the mode shapes of two close modes, we will then scarify some accuracy for one of the modes. While in the case of MAC_2, we may use the first two singular vectors at the same location to estimate the mode shapes of two close modes without losing any precision for each mode. However, when two modes are extremely close, or they are repeated poles, it is difficult to predict the correspondence between the individual mode shapes and the singular vectors. In this case, we will not be able to identify any peaks around the peak frequency in the purity plot as well. Therefore, depending on different mode shape cases, we use different identification criteria to identify the mode shapes when we failed to identify the two vertical lines, or there is no prominent peak in the purity plot for the two close modes.

For MAC_1, since the system usually approaches the mixed states around the peak frequency, we then select the two singular vectors corresponding to the frequencies on two sides of the peak frequency. In other words, if f_k is the peak frequency, Δf is the frequency resolution, then the two locations is given by $f_k - a\Delta f$ and $f_k + a\Delta f$, where α is a constant, here we assume it equals 1.

While for MAC_2, we consider three cases. Before discussing those three cases, first, we recognise that there might be a situation that there is only one prominent peak (i.e., two peaks merged into one) in the

purity plot for the two close modes. In this case, the rule of thumb is to compare the location of this peak in the purity plot with the peak in the first singular value plot.

To be more specific, suppose f_k is the peak frequency and f_p is the purity peak frequency, if $f_k < f_p$, we use the first singular vector at f_p to approximate the second mode shape vector and the second singular vector to approximate the first mode shape vector. For the case when $f_k > f_p$, the selection rule of the singular vectors at f_p is reversed. When $f_k = f_p$ or there is no prominent peak in the purity plot, we just use the first two singular vectors at the peak frequency to estimate the two mode shapes, but with the risk to losing some precision for each mode shape. It is possible that both the first and the second singular vectors are imperfect estimations for the mode shapes, or the correspondence between the singular vectors and mode shape vectors are completely swapped. As a result, the estimation for the mode shapes when two modes are extremely close or identical will not be as good as the case when we can successfully identify the two peaks for the close modes in the purity plot.

Note that the different appearance between the entropy or purity plot of MAC_1 and MAC_2 can be used to determine if the two close modes are orthogonal or correlated to each other.
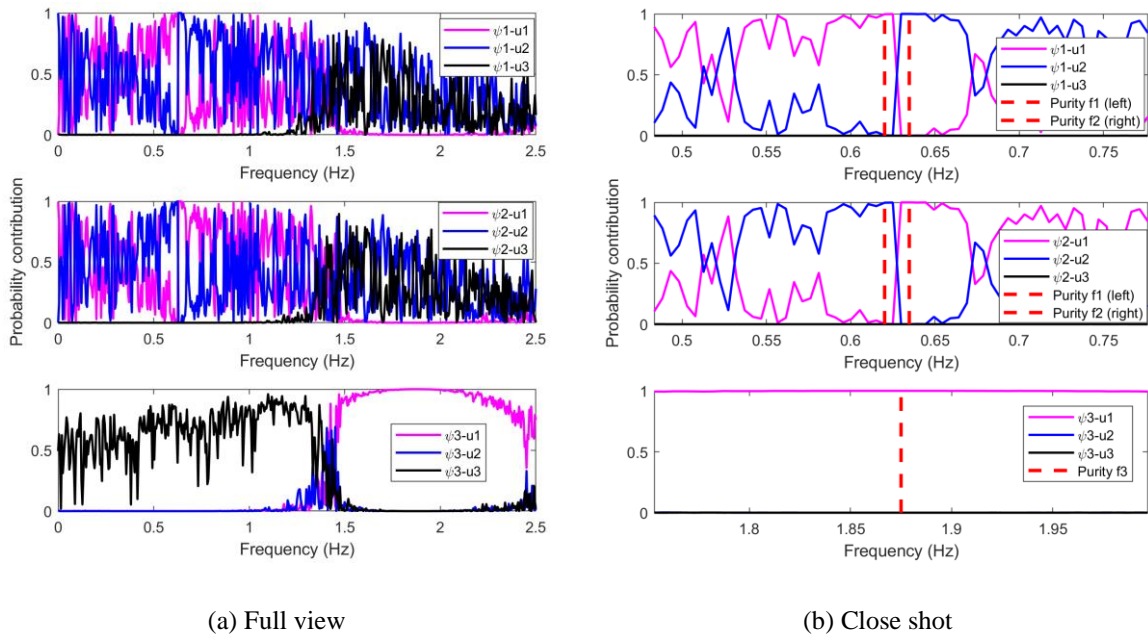

(a) Full view   (b) Close shot

Figure 4.13 Probability of the first three singular vectors contributed to mode 1 (top), mode two (middle) and mode 3 (bottom) of (a1) (MAC_2, damping_1, noise_1)

Although there are some differences between the case of MAC_1 and MAC_2, the damping ratios and the noise level have the same impact on them. As a comparison, the purity plot of the case (b2) (MAC_2, damping_2, noise_1) and case (b3) (MAC_2, damping_1, noise_2) are presented in Figure 4.14 and

Figure 4.15, respectively. The same as we discussed before in the cases of MAC_1, the purity of a system is damping sensitive but noise insensitive. The frequency band for a pure state decreases with the increase of the noise level.
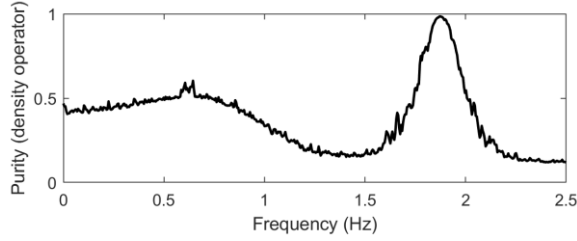

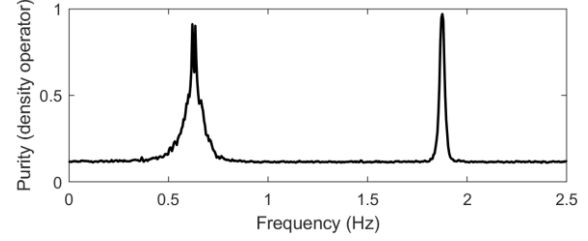Figure 4.14 Purity plot of (b2) (MAC_2, damping_2, noise_1)

Figure 4.15 Purity plot of (b3) (MAC_2, damping_1, noise_2)
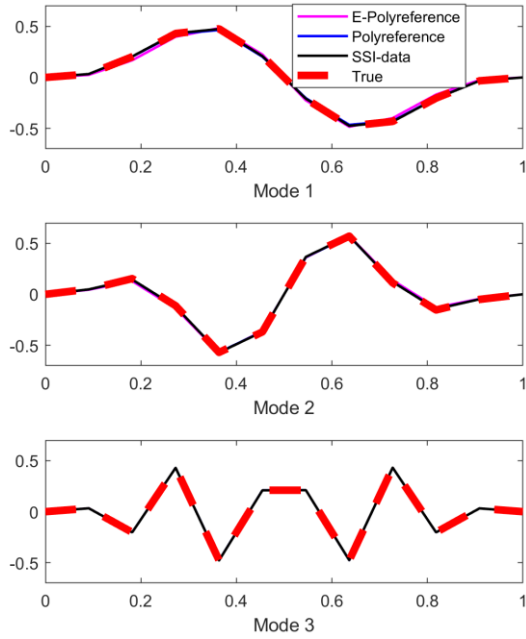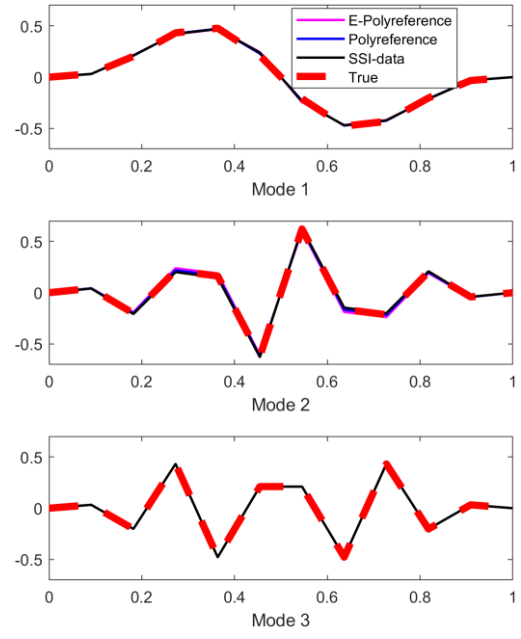
(a1) (MAC_1, damping_1, noise_1)　　　　(b1) (MAC_2, damping_1, noise_1)

Figure 4.16 Identified mode shapes of (a1) (MAC_1, damping_1, noise_1) and (b1) (MAC_2, damping_1, noise_1). Mode 1 (top), mode 2 (middle), mode 3 (bottom)

In order to show how good the results are by using the proposed method, the estimated mode shapes are presented in Figure 4.16. It is clearly shown from Figure 4.16 that when $f_1 = 0.625$ (Hz), $f_2 = 0.6309$ (Hz), $\xi_1 = \xi_2 = 0.1$ % , and the disparity $\gamma = 9.4$, all three methods can successfully identify the mode shapes while the FDD cannot because we are unable to specify the peak frequencies

for the two close modes based on the first singular value plot. Besides, the large disparity, which is given by this example, also verified our statement in section 4.4.1 that it is inappropriate and meaningless to compare the goodness of each technique used by different authors to tackle close modes identification problems based on the same closeness index.

Table 4.5 Identified modal parameters of (a1) (MAC_1, damping_1, noise_1) for one example run

No.	Natural frequencies (Hz)				Damping ratios (%)			
	True	M1	M2	M3	True	M1	M2	M3
1	0.625	0.6251	0.625	0.6251	0.1	0.13	0.19	0.08
2	0.6309	0.6309	0.6309	0.6309	0.1	0.12	0.19	0.08
3	1.875	1.8752	1.8752	1.8752	0.1	0.12	0.13	0.1

(Note: M1 – Enhanced Polyreference; M2 – Polyrference; M3 – SSI-DATA.)

As for the natural frequencies and damping ratios, the identified results are presented in Table 4.5 and Table 4.6 for the two cases we considered here. It is interesting to note that the proposed two-stage method (I.E., Enhanced Polyreference) gives better damping ratio estimates than the conventional Polyreference method. Although the natural frequencies estimated from the three methods are comparable, the SSI-DATA does produce better damping ratio estimates than the other two methods. This pattern will be shown visually in section 4.4.2.4 with a result of 100 times averaged multiple runs.

Table 4.6 Identified modal parameters of (b1) (MAC_2, damping_1, noise_1) for one example run

No.	Natural frequencies (Hz)				Damping ratios (%)			
	True	M1	M2	M3	True	M1	M2	M3
1	0.625	0.6254	0.6253	0.6253	0.1	0.16	0.23	0.12
2	0.6309	0.6306	0.6306	0.6307	0.1	0.13	0.2	0.08
3	1.875	1.8749	1.8749	1.875	0.1	0.13	0.14	0.11

(Note: M1 – Enhanced Polyreference; M2 – Polyrference; M3 – SSI-DATA.)

4.4.2.4 Estimated modal parameters

In the last subsection, we showed that in the vicinity of the peak frequency of two close modes, the system is usually in a mixed state. In order to estimate the individual mode shapes properly, we have to select the singular vector slightly away from the perturbed region on the frequency line. The Von Neumann entropy plot and the purity plot both provided us a way to find the best locations to estimate the mode shape vectors. To show the difference between the FDD peak frequencies and purity peak frequencies, we plotted the averaged frequencies of these two kinds of peak frequencies for the 8 cases in terms of Table 4.4 in Figure 4.17. Each point in Figure 4.17 is an average of 100 runs.

(a1) (MAC_1, damping_1, noise_1)
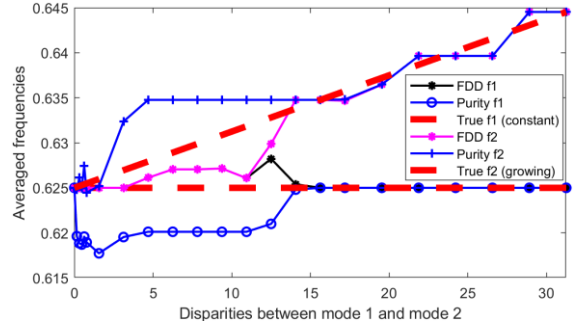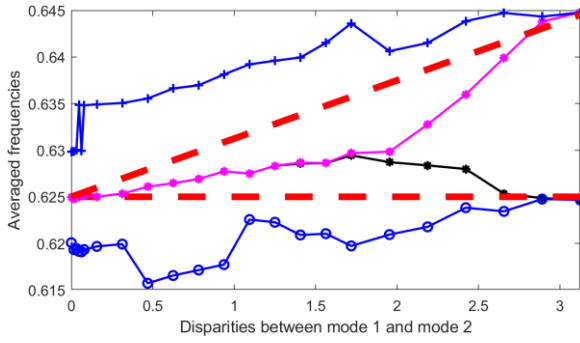
(b1) (MAC_2, damping_1, noise_1)

(a2) ((MAC_1, damping_2, noise_1)

(b2) (MAC_2, damping_2, noise_1)

(a3) (MAC_1, damping_1, noise_2)
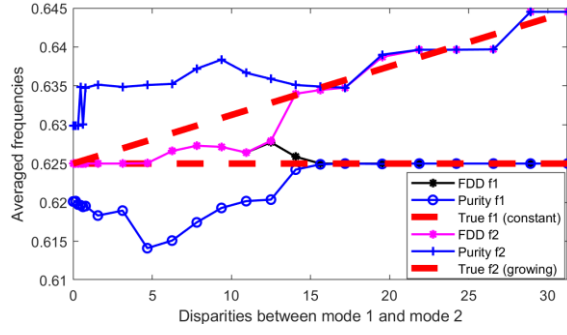
(b3) (MAC_2, damping_1, noise_2)
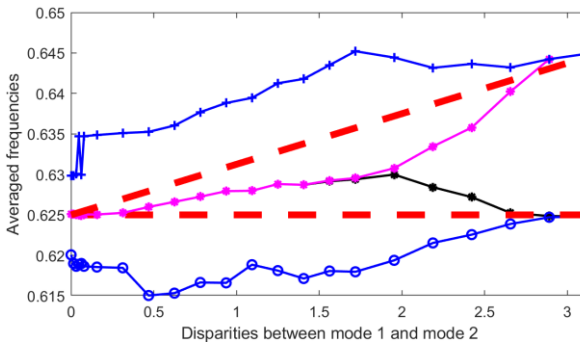
(a4) (MAC_1, damping_2, noise_2)
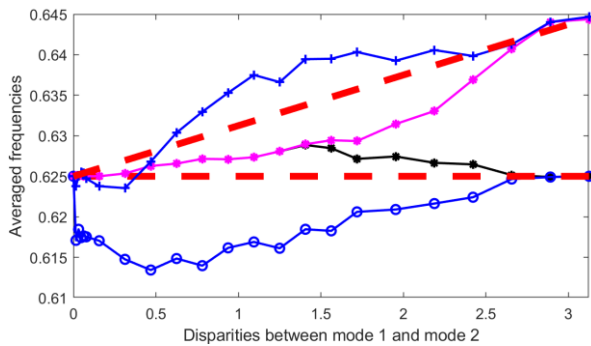
(b4) (MAC_2, damping_2, noise_2)

Figure 4.17 Trajectories of the averaged peak frequencies of the FDD and the purity for the two close modes of the 8 cases. All the frequencies are in Hz.

From Figure 4.17, we see that with the increasing disparity between the two close modes, the two kinds of peak frequencies converge for both modes eventually. And the convergence rate is only affected by the damping ratio. The larger the damping ratio, the slower the convergence. After the FDD peak frequencies and the purity peak frequencies converged, we can use either of them to identify the mode shapes and the natural frequencies.

Besides, for all 8 cases in Figure 4.17, the FDD peak frequency line start at f_1 for both modes with a single peak, then it gradually moves towards f_2. As the disparity between the two modes increases, the FDD peak frequency line splits into two frequency lines before each line converges to the corresponding true value lines. In general, before the FDD peak frequency line splits, we cannot use the FDD to identify the modal parameters of the two close modes, as it is unable to distinguish them from the peak frequency. This explains that for the cases with lower damping ratios in the disparity range from 3 to 15, the Enhanced Polyreference method can successfully identify the natural frequencies, damping ratios, and mode shapes to a very high accuracy while the conventional FDD fails. While in the case of higher damping ratios, for the FDD to be successful, the disparity between the first two modes needs to be larger than 2.8.

Furthermore, it must be pointed out that although the FDD peak frequency line of the two close modes separated from each other at a specific frequency, meaning that we can observe two peaks in the first singular value plot. Unfortunately, those two peak frequencies are far from the corresponding true values. Hence, it is still not a good idea to use the FDD to identify the natural frequencies, as the estimates will be biased. For example, the FDD peak frequencies line splits at a disparity of 1.7, but the two split lines are still in between the two true frequencies lines.

Apart from the above discussion, it is worth mentioning that since we used different criteria to estimate the mode shapes from the purity plot in terms of the correlations relationship between the mode shapes of the two close modes, the starting points of the frequencies found by evaluating the purity plot are different for the cases of MAC_1 and MAC_2.

In the previous subsection, we compared the modal parameters identified from three methods, i.e., Enhanced Polyreference, Polyreference, and SSI-DATA, in terms of a single numerical experiment. In this subsection, we want to show the performance of the proposed method with multiple runs. The estimated natural frequencies, damping ratios, and mode shapes (averaged MAC values) are presented in Figure 4.18, Figure 4.19, and Figure 4.20, respectively. All the results are plotted against the disparities between the first two close modes. For some outliers, we removed it based on the Interquartile Range (IQR) method. If the 25th and the 75th percentiles are denoted by Q1 and Q3, respectively, then IQR is equal to Q3 – Q1. By using the IQR method, we will only keep anything between (Q1 – 1.5*IQR) and (Q3 + 1.5*IQR) and discard the rest. In other words, each point in the figure is an average of 100 runs, but only with those that satisfy the IQR criteria. As for the mode shapes,

for the 3 active modes, they are fixed for different disparity values for all cases of MAC_1 and MAC_2, and they were shown in Figure 4.16.
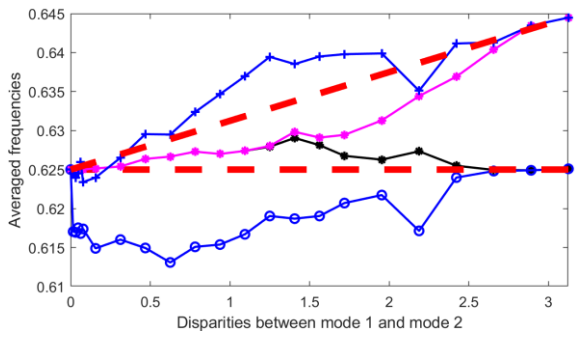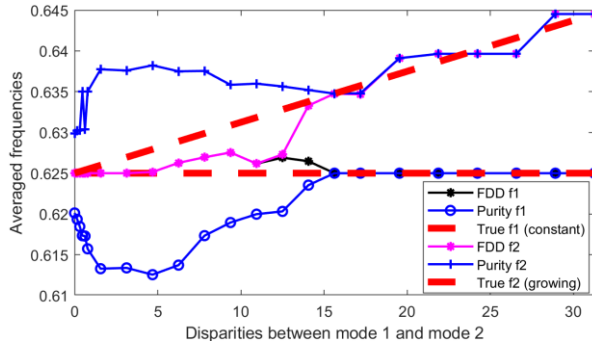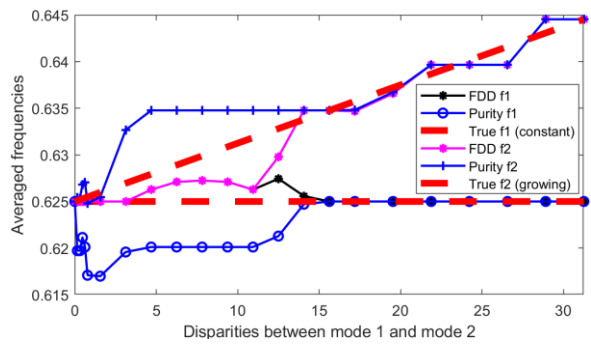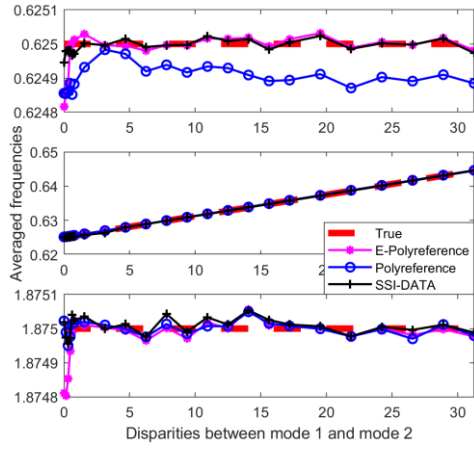

(a1) (MAC_1, damping_1, noise_1)   (b1) (MAC_2, damping_1, noise_1)
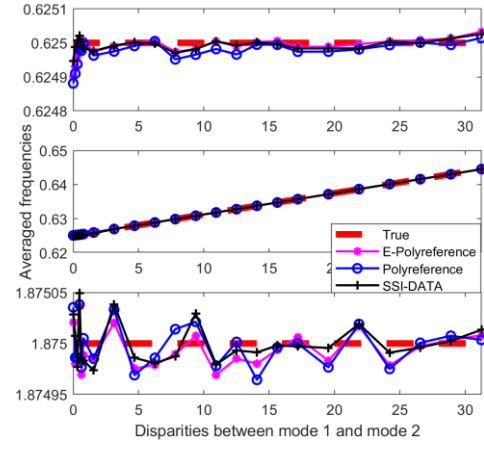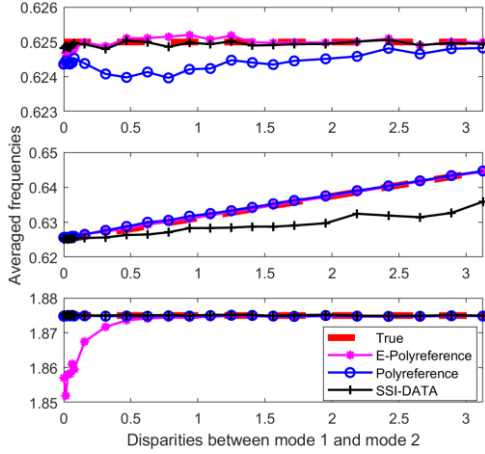
(a2) (MAC_1, damping_2, noise_1)   (b2) (MAC_2, damping_2, noise_1)

(a3) (MAC_1, damping_1, noise_2)   (b3) (MAC_2, damping_1, noise_2)

(a4) (MAC_1, damping_2, noise_2)   (b4) (MAC_2, damping_2, noise_2)

Figure 4.18 Identified averaged natural frequencies of 3 active modes. Mode 1 (top), mode 2 (middle), mode 3 (bottom). All the frequencies are in Hz

From Figure 4.18 we see that all three methods give better natural frequency estimates in the case of MAC_2 regardless of the value of damping ratios and noise. Although the Enhanced polyreference does not perform very well in the cases of MAC_1 when two modes are extremely close to each other, one can see from (a1) and (a2) or (a3) and (a4) that it converges to the true as very fast, and the lower the damping, the faster it converges. Besides, it is clear from all four cases of MAC_1 that the two close modes have an impact on the well-separated mode, when the two close modes have correlated mode shapes. And it is mysterious that SSI-DATA gives poor frequency estimates of mode 2 in the case of (a2) (MAC_1, damping_2, noise_1), and Polyreference gives consistently poorer frequency estimates in mode 1 in all 4 cases of MAC_1 and 2 MAC_2 cases of the 3rd mode with higher damping level.

(a1) (MAC_1, damping_1, noise_1)   (b1) (MAC_2, damping_1, noise_1)
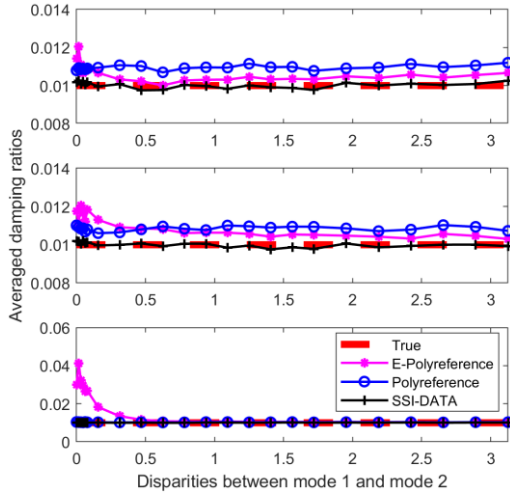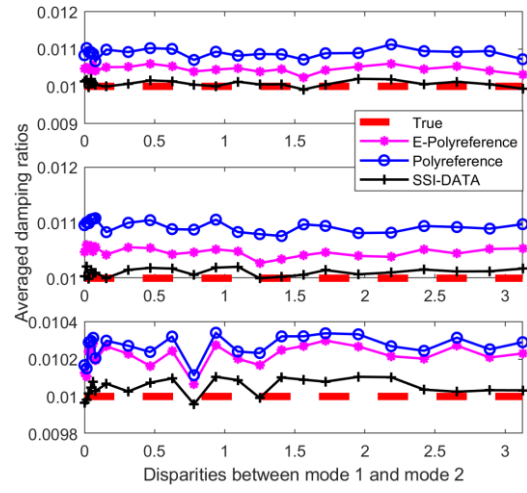
(a2) (MAC_1, damping_2, noise_1)   (b2) (MAC_2, damping_2, noise_1)
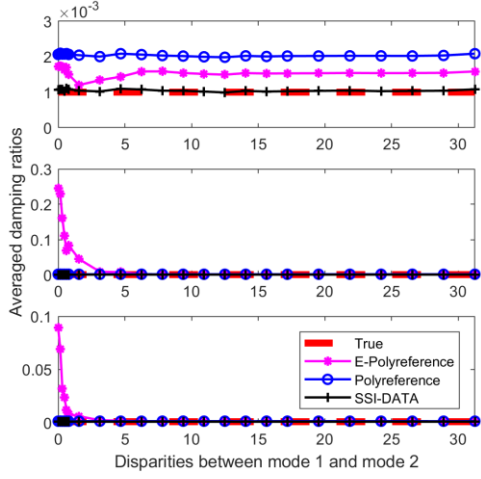
(a3) (MAC_1, damping_1, noise_2)   (b3) (MAC_2, damping_1, noise_2)
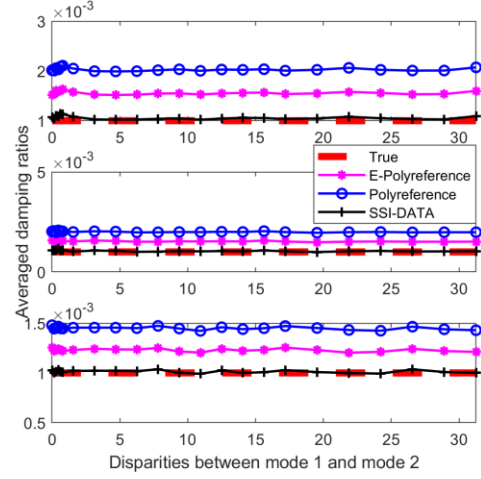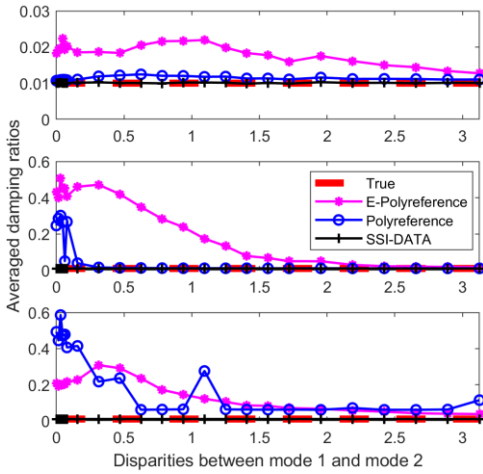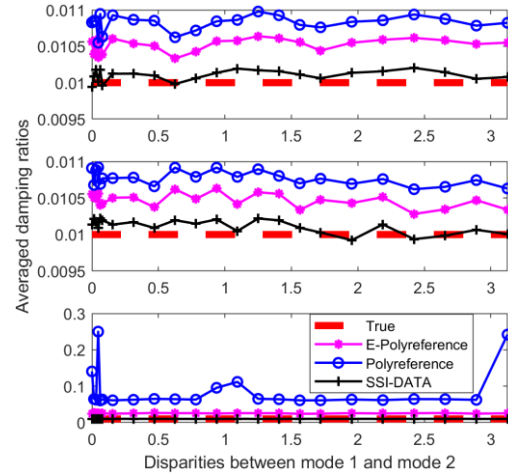
(a4) (MAC_1, damping_2, noise_2)   (b4) (MAC_2, damping_2, noise_2)

Figure 4.19 Identified averaged damping ratios of 3 active modes. Mode 1 (top), mode 2 (middle), mode 3 (bottom).

As for the estimated damping ratios shown in Figure 4.19, we see that all cases indicate that SSI-DATA consistently gives the best damping ratio estimation. While for the Enhanced Polyreference and Polyreference, except for the last 3 cases in the MAC_1 category, the Enhanced Polyreference gives consistently better damping ratio estimates than Polyreference in all modes. However, in the case of (a4) (MAC_1, damping_2, noise_2), the damping ratios estimated from Enhanced Polyreference show a large deviation from the true values. Besides, when compare the frequency estimates and the damping ratio estimates in the case of (a2) (MAC_1, damping_2, noise_1), it is interesting to note that although SS-DATA gives poor frequency estimates, the estimated damping ratios are very close to the true values.
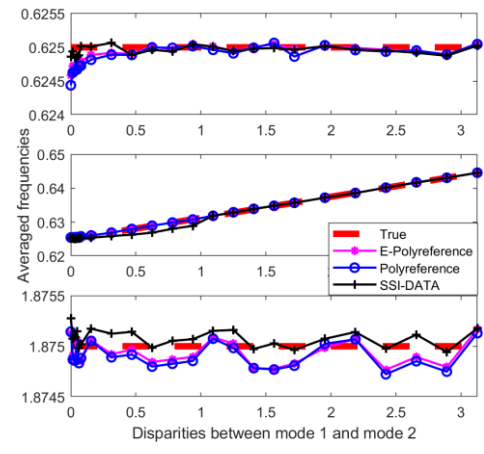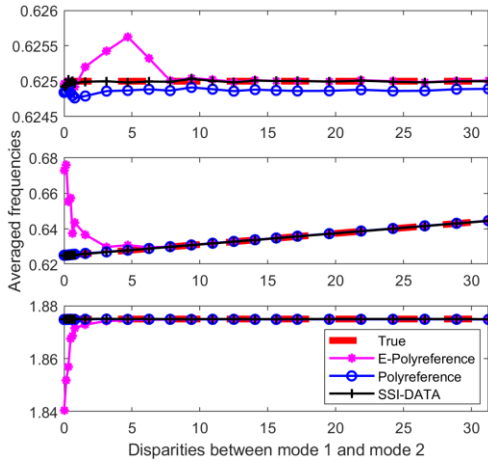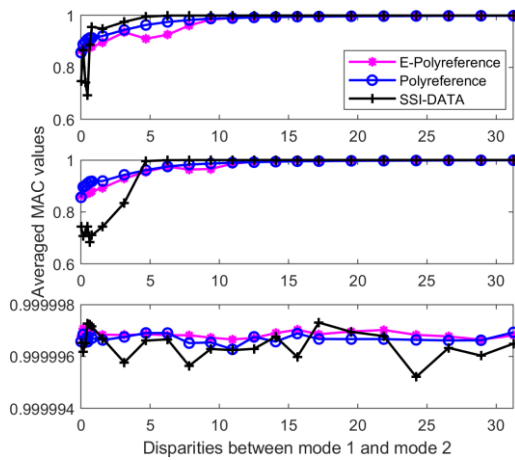

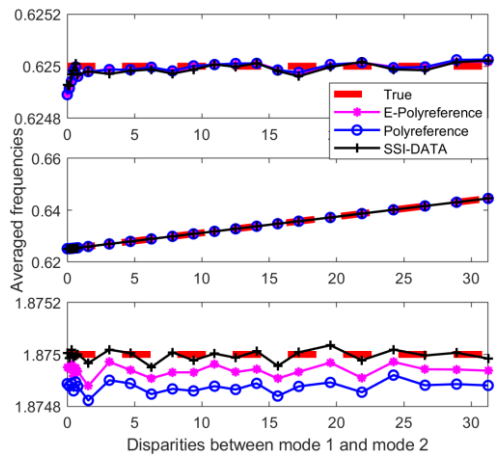(a1) (MAC_1, damping_1, noise_1)   (b1) (MAC_2, damping_1, noise_1)

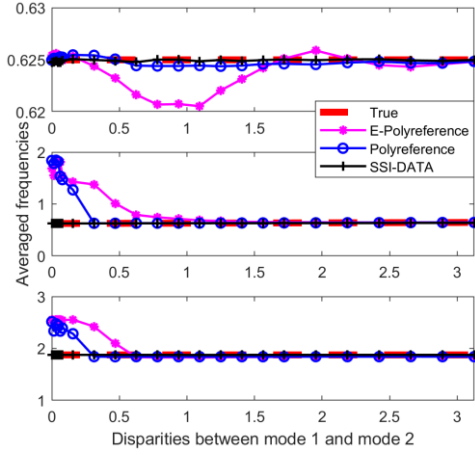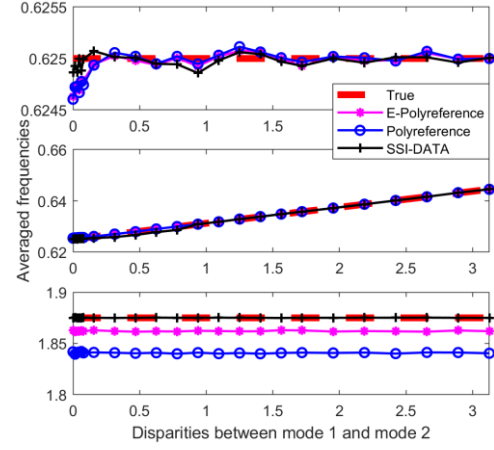(a2) (MAC_1, damping_2, noise_1)   (b2) (MAC_2, damping_2, noise_1)

(a3) (MAC_1, damping_1, noise_2)　　　　(b3) (MAC_2, damping_1, noise_2)
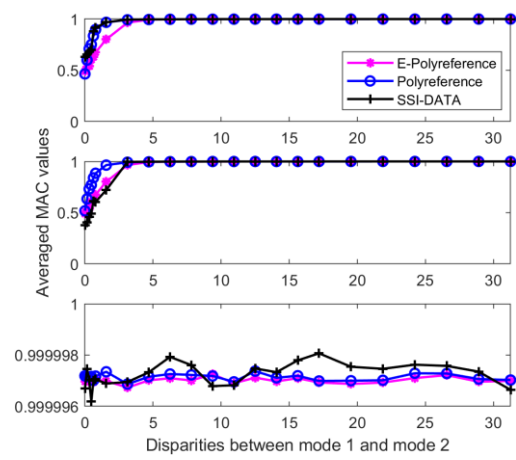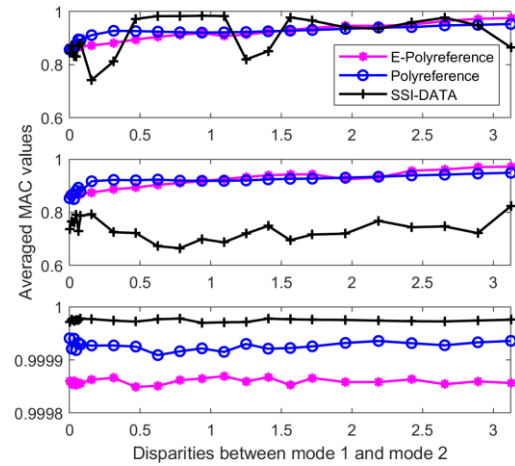
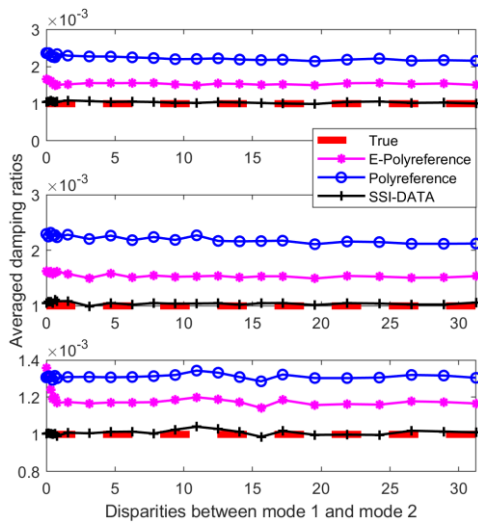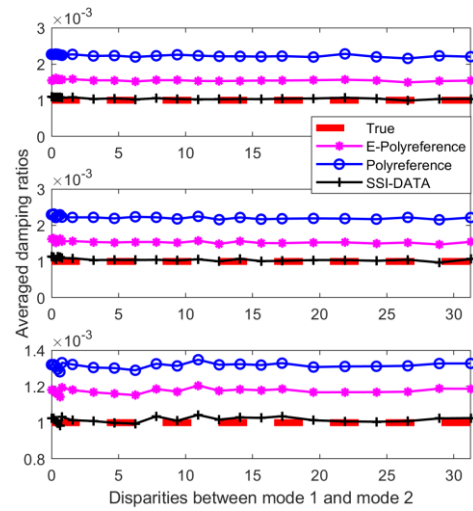(a4) (MAC_1, damping_2, noise_2)　　　　(b4) (MAC_2, damping_2, noise_2)
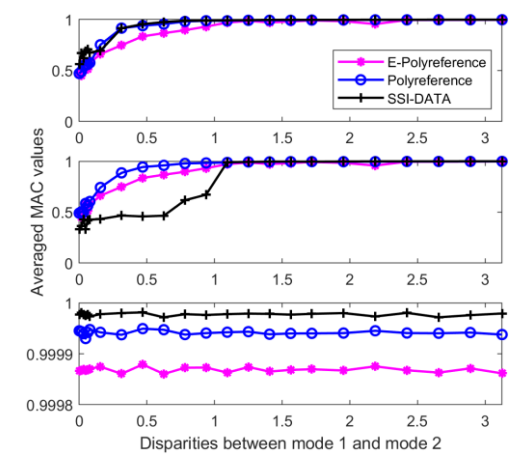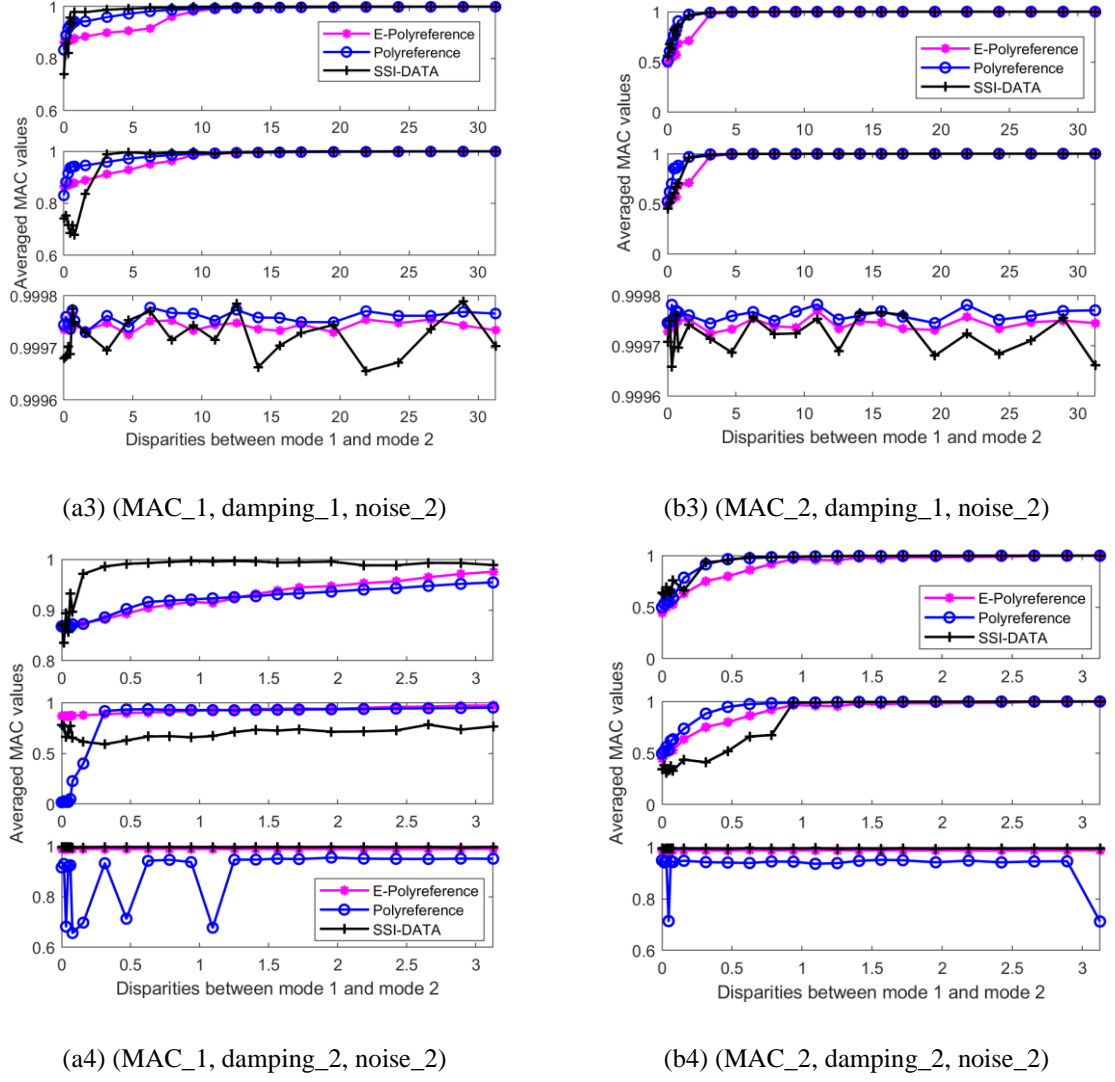
Figure 4.20 Identified averaged MAC values of 3 active modes. Mode 1 (top), mode 2 (middle), mode 3 (bottom).

The averaged MAC results between the estimated mode shapes and the reference mode shapes illustrated in Figure 4.20 shows that the Enhanced Polyreference gives consistently good estimates. However, all three methods struggled to identify the mode shapes for the two close modes when the disparity between them is extremely small in the cases of MAC_2. While in the cases of MAC_1, the MAC values of the mode shapes estimated via the Enhanced Polyreference are all larger than 0.8, even for the repeated pole case.

Additionally, since different correlation condition between two mode shapes of the close modes can result in different purity plot, we can expect that the mode shape estimation results for the two cases MAC_1 and MAC_2 would be different. And it obvious in Figure 4.20 that the cases of MAC_2 give relatively better mode shape estimation results than the cases of MAC_1, as the averaged MAC values converge to 1 faster with the increase of the disparity between the two close modes. The same pattern is reflected in the natural frequencies and damping ratio estimations as well. And this phenomenon is

more prominent in the cases of (a4) (MAC_1, damping_2, noise_2) and (b4) (MAC_2, damping_2, noise_2) when the system has a large damping ratio and the corresponding vibration measurements have a high noise level.

As for the SSI-DATA, by comparing (a2) (MAC_1, damping_2, noise_1) and (b2) (MAC_2, damping_2, noise_1), (a4) (MAC_1, damping_2, noise_2) and (b4) (MAC_2, damping_2, noise_2), we may conclude that the mode shape estimation accuracy of this method is sensitive to the correlation relationship between the two close modes. And the effect of the correlation between the close modes seems to have a bigger impact on the second mode than the first mode.

Besides, compared to the other two methods, the Polyreference seems to perform very poorly when the system has both large damping and large noise in the measurements. Especially in the case of (a4) (MAC_1, damping_2, noise_2), when the first two modes are extremely close to each other, it almost failed to identify the mode shapes for the second mode. While for the third mode, which is well-separated from the first two modes, the Polyreference struggled to identify the mode shape when the first two modes are close to each other. When we compare (a4) (MAC_1, damping_2, noise_2) with (b4) (MAC_2, damping_2, noise_2), it is not difficult to see that the Polyreference still gives poorer estimates for the well-separated mode.

Regarding the common understanding about the mode shape estimates via the decomposition of a PSD estimator that only if the two mode shapes are orthogonal to each other, the first singular vectors are unbiased estimates of the mode shapes. According to our theoretical explanation given in section 4.2.2.5 III and the estimation results illustrated in Figure 4.20, we can conclude that such understanding is misplaced. Because regardless of the orthogonality of the mode shapes, when we use the FDD or by analysing a density operator to estimate the mode shapes using the first singular vectors, the first singular vectors in the frequency domain are rotating in the state-space. At different frequency lines, they point in different directions. However, those directions are not the ones that best explain the mode shapes vectors but the ones that best explain a random state vector $\tilde{\mathbf{y}}\left(f_k\right)$. Only if $\tilde{\mathbf{y}}\left(f_k\right)$ is pointing to the same direction as one of the mode shapes, the corresponding singular vector can be deemed as a good estimate of that mode shape.

One should bear in mind that the first singular vector does not discriminate the mode shapes in virtue of the correlation relationship among them. All the mode shapes are just a vector in the state-space. Therefore, there should not be any bias in the estimated mode shapes obtained based on the first singular vectors via the FDD or the proposed method, provided that the state vector $\tilde{\mathbf{y}}\left(f_k\right)$ is in the same direction of the corresponding mode shape. However, bias does exist when we use the first two singular vectors to estimate two nonorthogonal mode shapes.

Now that we have tested the performance of our proposed method in estimating natural frequencies, damping ratios, and individual mode shape vectors based on 8 different cases with the increasing disparity between. In comparison with Polyreference, the Enhanced Polyreference has the ability to improve the estimation accuracy of the natural frequencies and damping ratio when two modes are slightly separated in the sense of the system purity, i.e., we can identify two peaks in the purity plot for the two close modes. As for the mode shapes, the estimation based on the proposed method for all 8 cases is robust compared to the other two methods. Therefore, even if the proposed method cannot produce the best damping ratio estimation compared to SSI-DATA when two modes are extremely close or identical, the overall performance is good and stable, and it is noise insensitive for the mode shape estimation. Apart from this, unlike the Polyreference and SSI-DATA, we do not need to worry about the occurrence of the spurious modes when using the proposed method, and it is less dependent on the human-defined parameters. Except for the spectral resolution, which needs to be specified, we can easily automate the whole process without too much trouble.

It is true that no method can perfectly adapt to all situations and give the best estimation results. There will always be some kind of compromise. But overall, the proposed method is user friendly and easy to implement, and it is capable of dealing with close modes. In the next section, we will have a further discussion of the close modes problem with respect to the points we may have missed so far.

4.4.3 Further discussion and future work

For close mode analysis, there are still lots of unsolved puzzles around this topic which worth exploring. For example, we have noticed that if there is a pair of close modes in the system, it could have some sort of impact on the estimation of the well-separated modes depending on the identification methods were used as well as the modal parameters are concerned. Such a phenomenon needs to be further investigated. Also, this work only studied two closely spaced modes the case of the cluster modes was untouched. As an important case, the cluster modes case will be investigated in future work.

Apart from this, although we included the repeated pole case in this study, this problem is rather complicated than others. It is true that in the forward analysis, we can obtain a complete set of eigenvectors from the characteristic equation of motion to form the modal space provided that both stiffness and mass matrix is positive semidefinite. But such eigenvectors are only one representative of many potential mode shape vectors in the modal space that can explain the close modes. In other words, the mode shape vectors we obtained from any FE model are not unique, and different computing setup could give us different eigenvectors.

This can be further demonstrated by considering the eigenvalue problem of a 3×3 positive definite symmetric matrix $\mathbf{A}$, i.e., $\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$, $\mathbf{x} = \left(x_1, x_2, x_3 \right)^{\mathrm{T}}$. When we solve for each eigenvector, we need

to solve $(\mathbf{A} - \lambda_k \mathbf{I})\mathbf{x} = 0$ for each eigenvalue we obtained, here $k = 1, 2, 3$. If all three eigenvalues of this matrix are distinct, for each λ_k, there will be two independent linear equation expressed by the components of vector $\mathbf{x}$, i.e.,

$$\begin{cases} a_1 x_1 + b_1 x_2 + c_1 x_3 = 0 \\ a_2 x_1 + b_2 x_2 + c_2 x_3 = 0 \end{cases} \tag{4.97}$$

where a, b, c are the constant coefficients derived from $(\mathbf{A} - \lambda_k \mathbf{I})$. Each equation in Eq. (4.97) represents an equation of a plane which passes through the origin in the 3-dimensional Euclidean space. Solving Eq. (4.97) is equivalent to find the intersection between two planes, which is a line, and it passes through the origin. And it is realised by giving one of the components x a convenient value (say 1), and solving for the other x's. This process implies that all the vectors connecting the origin and the points on this intersection line is a valid representation of the eigenvector corresponding to λ_k. Usually, we would like to normalise the eigenvector to length 1.

In the case of repeated poles, two of λ s will be identical. Assume λ_1 is a double root of the characteristic equation, it will be a complete eigenvalue when $(\mathbf{A} - \lambda_k \mathbf{I})\mathbf{x} = 0$ has two independent solutions. This will happen when $(\mathbf{A} - \lambda_k \mathbf{I})\mathbf{x} = 0$ has essentially only one equation, and the other two equations are constant multiples of it or identically zero. This means that we can only obtain one equation in Eq. (4.97), and it is a plane equation. It implies that all the vectors connecting the origin and the points on this plane is a valid representation of the eigenvector corresponding to λ_1. Unlike the previous case that we can use a unit vector to span the intersection line of two planes, we need two vectors to span a plane in the case of repeated eigenvalues. Because of this, it becomes problematic when the conventional MAC is used to assess the correlation between two modal vectors. To resolve this problem, D'Ambrogio and Fregolent [190] proposed to evaluate the correlation between a modal vector and a subspace spanned by two modal vectors, while Brincker and Lopez-Anelle [113] suggested using the angle between two subspaces. However, both of their discussion was restricted to the subspace of order 2. Therefore, whether or not the proposed solutions can be extended to the case with more than 2 repeated eigenvalues needs to be further investigated.

It must be pointed out that we adopted neither of the proposed methods given by D'Ambrogio and Fregolent [190] and Brincker and Lopez-Anelle [113] to assess the accuracy of our estimates for the close modes. On the contrary, we applied MAC to all our cases. It is because we want to test how well our proposed method, in the meantime compared to other promising methods, can estimate the individual mode shape vectors in a frequency band that the conventional FDD fails. In this case, low MAC values when does indicate bad mode shape estimates.

Besides, when we use two fixed mode shape vectors to simulate the system responses, we have automatically assumed that the system under the current measurement is oscillating in the direction governed by those two predefined mode shape vectors. According to our theoretical explanation of the FDD in terms of PCA, which shall also apply to our density operator, that the first and the second singular vectors point to the directions that explain most of the information of the system. Therefore, in the case when two predefined mode shapes are orthogonal, we can use the first two singular vectors to estimate the mode shapes. However, this may only work well when two modes are very close but not identical. Also, it is quite challenging for us to locate a proper frequency location that could give us the best estimates when there is no prominent peak in the purity plot. Moreover, there is no direct correspondence between the individual mode shape vectors and the first two singular vectors when the system is given a white noise excitation at all locations. More precisely, the first two singular vectors alternate between the two mode shapes. Therefore, for a damped system with close modes under the ambient excitation, when two modes are very close but not identical, their modes become indefinite. Hence, the relation between the estimated mode shapes and the reference mode shapes should be more appropriately described by statistical laws, and this is something that need further investigation.

Usually, when we use the conventional FDD to estimate the mode shapes, our estimates are complex-valued. In this case, the common practice is to keep the real part and discard the imaginary part. However, it is unwise to do so when we have two orthogonal close modes. Since the mode shapes are rotating in the subspace, the estimated mode shapes based on the singular vectors may have a small real part and large imaginary part. And it is likely that the imaginary part of a singular vector gives better estimates, even if the reference mode shape vector is real. Therefore, caution is needed when dealing with mode shape estimation for close modes. In this work, when calculating the MAC values, we used the complex-valued singular vectors. Although one should mind this identification issue, it is not a problem for well-separated modes. So, the regular practice is still valid.

Apart from the above discussion, it is worth showing the difference between the singular value plot of a density operator and the FDD. In the last subsection of section 4.2.2.5, we have discussed the connection between the FDD and the decomposition of a density operator, see Eq. (4.82). We are now quite familiar with the FDD, and its corresponding singular value plot has been shown throughout this chapter, see for examples in section 4.2.2.3. The same as the FDD, we can also plot the singular values of a density operator against the frequency line. In Figure 4.21, the first three singular values are plotted for the same cases of (a1) (MAC_1, damping_1, noise_1) and (b1) (MAC_2, damping_1, noise_1) that we have discussed in section 4.2.2.3. It is interesting to note that the first singular value plot of a density operator has the same trend as the purity plot of the density operator, while the second singular values behave in the same way as the Von Neumann entropy plot. Such correspondence between the first two singular values of a density operator and its purity and the entropy plots need to be investigated further.

<div align="center">(a1) (MAC_1, damping_1, noise_1)   (b1) (MAC_2, damping_1, noise_1)</div>
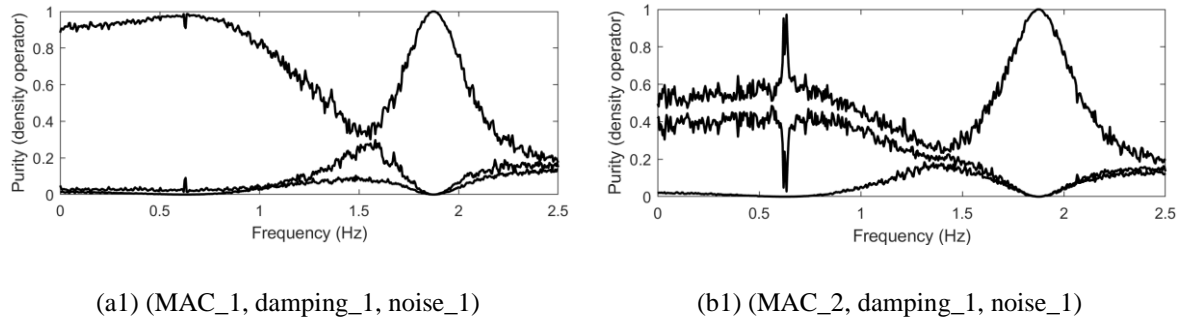
<div align="center">Figure 4.21 Singular values plot of a density operator against the frequency line. Only the first three singular values are considered.</div>

Finally, we want to point out that there is more than one way to obtain a plot that is similar to the purity plot we introduced in this chapter. For example, we can plot $\bar{\lambda}_{k1} \log_2 \bar{\lambda}_{k1}$ or $\bar{\lambda}_{k2} \log_2 \bar{\lambda}_{k2}$ against the frequency line, where $\bar{\lambda}_{k1}$ and $\bar{\lambda}_{k2}$ are the first and second singular value of a density operator at f_k; or we can plot $\log_2 \left(\lambda_{k1} / \lambda_{k2} \right)$ against the frequency line, where λ_{k1} and λ_{k2} are the first and second singular value of a PSD estimator at f_k. One can verify that the curves of those functions are similar to the purity plot. Therefore, we have multiple options to do the same analysis. However, compared to the Von Neumann entropy and the purity, which are well justified by the statistical theory and quantum mechanics, the other indicators bear less mathematical and physical foundation. Hence, we establish our theory based on the Von Neumann entropy and the purity of a density operator.

4.5. Conclusion

This chapter comprises 3 essential sections, which are given in section 4.2, section 4.3, and section 4.4, respectively. In section 4.2, the theoretical background for the FDD and a density operator is presented. The theory is given in two parts. In the first part, a PSD estimator is defined based on the Periodogram. While in the second part, the property of that PSD estimator is discussed. In the meantime, the FDD is redefined in terms of the PCA, and a density operator, which comes from quantum mechanics, is introduced and derived based on the PSD estimator to tackle the close modes estimation problem.

In section 4.3, instead of applying the FDD to the stationary data, we tested its ability to deal with some nonstationary random processes with some simulated data. The discussion in this section served as a preparation for the modal analysis in chapter 6, as we want to use the FDD to identify the mode shapes of a bridge by using the output only responses of it in the first stage of our Vehicle Bridge Interaction (VBI) analysis. Since in our VBI model, the impact of the vehicle is insignificant, and the bridge's modal parameters do not change over time, only the amplitude of the bridge responses diminish with time, and we can confidently apply the FDD to the bridge's responses to extract the mode shapes.

While in section 4.4, we gave an in-depth discussion over the close modes problem based on our proposed method. The model we used here is similar to the one we used in chapter 3, with only a few changes to the parameters. Since the natural frequencies and damping ratio cannot be obtained by analysing a density operator alone, we designed a two-stage method, namely, Enhanced Polyreference. By using it we can extract all the modal parameters (i.e., natural frequencies, damping ratios, and mode shapes) we can get in an OMA analysis. As its name implies, it is a combination of the Polyreference method. In the first stage, we analysed a density operator to get the mode shape estimation, then used it to decompose the normal coordinates into a set of modal coordinates; in the second stage, we applied the Polyreference method to those modal coordinates to extract the natural frequencies and damping ratios. However, it must be pointed out that there is no particular reason to use the Polyreference in the second stage, and any pole analysing method fulfil the job. In order to evaluate the performance of our proposed method, the same data is also analysed by the Polyreference method and the data-driven Stochastic Subspace Identification (SSI-DATA) method.

In our analysis, different cases regarding different MAC values of two close modes, damping ratios, and noise levels are investigated to reveal the underlying issues associated with the close modes identification problem. The results show that the Enhanced Polyreference improves the estimation accuracy of the natural frequencies and the damping ratios in comparison with the Polyreference, while the SSI-DATA can provide the best damping ratios estimation. When it comes to the mode shape identification, Enhanced Polyreference uses a density operator in the first stage to estimate the mode shapes gives the most robust performance than the comparison methods. Although the mode shape estimation procedure proposed in this work does not work perfectly when two modes are extremely close or identical, when the disparity between the two close modes are slightly larger, it gives comparable identification precision of the mode shapes estimates when compared to Polyreference and SS-DATA. Most importantly, the proposed method can work in a frequency band where the FDD fails.

Besides, it is worth mentioning that the use of the Periodogram to define the PSD estimator is an essential step for us to use the PCA to explain the FDD, as it offers us a simple way to define a series covariance matrices in the frequency domain at each frequency line. With the idea of PCA, we can deem the FDD as a variance-based method. On the other hand, we can view a PSD estimator from another perspective. When transforming it into a density operator, we can use the Von Neumann entropy or the purity of a density operator to assess the state conditions of a system. If we plot the Von Neumann entropy or the purity of each density operator at each frequency line, the resulting curve can give us some insight into the system, whether it is in a pure state or a mixed state. When a system is in a pure state, the entropy approaches 0; accordingly, the purity approaches 1, whereas in a mixed state, the entropy will be large, and purity will be less than 1. Our analysis showed that the best mode shape estimates are not necessarily given by the peak frequency of the FDD. Rather, when two modes are very close to each other in the case that the first singular value plot of the FDD is unable to resolve it,

the best mode shape estimates are corresponding to the frequency locations where a system approaches a pure state. This is a significant result, as it reveals the protentional bias in using the FDD to estimate the mode shapes and the natural frequencies, see section 4.4.2.4.

As we mentioned in section 4.2.2.4, the FDD is very good at detecting the close modes, but it is unable to resolve the mode shape and natural frequency identification problem when two modes are inseparable in the first singular value plot. Although good estimation results of the mode shape can be obtained by analysing a density operator when two modes are very close but not extremely close or identical, the dependence on the peak picking of the purity plot will cripple it when a system has large damping ratios. Nevertheless, the proposed method is noise insensitive, and for well-separated modes, it can also be used to estimated natural frequencies, as the peak frequencies in a purity plot converge to their FDD counterparts. Apart from this, according to our discussion, different correlation relations between the two close modes will generally give a different view of the entropy or the purity plot. Such information can be used to make reference to the orthogonality of mode shapes. Besides, when used in combination with the Polyreference method, the physical pole selection problem associated with the original method is disappeared, which largely simplifies the identification procedure.

Furthermore, since the simply supported beam has well-separated modes, there is no need to use the Enhanced Polyreference method to estimate the mode shapes and natural frequency. Therefore, in the subsequent VBI analysis we will use the FDD to do the first stage identification.

5. A method for input-output system identification for moving force excitation

5.1 Introduction

As mentioned in chapter 1 that in order to develop a methodology to identify bridge modal parameters from vehicle induced vibration, we face two challenges, i.e., the correlation and the nonstationary nature of the measurements which originated from the spatial and temporal variation of the moving load from a passing vehicle. Overcome these two challenges is vital to estimate the modal parameters of interest.

In the last chapter, we reformulated the theoretical background of the FDD to adapt to our nonstationary case. With such an extension, now we can confidently apply our two-stage strategy to extract the structural modal parameters from the system. In other words, we can use the FDD to estimate the mode shapes of the bridge by using the measurements solely from the bridge in the first stage, then use those identified mode shapes to decompose the system via the mode superposition technique. In which case, both the input signal from the vehicle and the output signal from the bridge will be decoupled at the same time. Then a complex SIMO system is being effectively decomposed into a series of SISO systems represented by the Frequency Response Functions (FRFs), a subspace identification procedure can be conducted to extract the modal parameters from the system in this fashion.

In order to explain the proposed method, this chapter will first start with the discussion of the decorrelation of the input-output measurements. Then we will move on to present our two-stage system identification procedure. Two methods will be presented: one is the straightforward way through the optimisation of the FRFs, the other one will be the simple inverse method based on linear regression. Also, we will address the following question: should we use the estimated complex mode shape to decompose the system or its real counterpart?

5.2 Decomposition of the bridge system

To decorrelated the input and output of the measurement to identify the bridge modal parameters, we can think of transforming the correlated input-output measurements of the VBI system into independent subsystems, which are characterised by the modal coordinates. This decorrelation procedure, which is associated with the decomposition of the bridge system and its external moving force, can be achieved by applying the well-founded modal superposition technique to the equation of motion of the bridge

system. And to apply this technique, we need to have a clear understanding of the properties of the mode shapes associated with an equation of motion described by a special 2nd order ODE. Thus, in this section, we start our discussion with a system described by this special 2nd order ODE, then we can derive the important properties of the modes shapes associated with this system equation.

5.2.1 Characteristics equation of motion

Based on Newton's law of motion, excluding the damping, the equation of motion for a n DOF system is given by

$$\mathbf{M\ddot{y}}(t) = \mathbf{f}(t) - \mathbf{Ky}(t) \tag{5.1}$$

where $\mathbf{f}(t) \in \mathbb{R}^n$ is the external force vector, and $\mathbf{y}(t) \in \mathbb{R}^n$ is the state vector. $\mathbf{M}$ and $\mathbf{K}$ represent the mass and stiffness matrices, respectively, and they both are $n \times n$ matrix. For simplicity, in the following analysis we will drop the parenthesis with time.

For no external forcing, the system equation of motion can reduce to

$$\mathbf{M\ddot{y}} + \mathbf{Ky} = \mathbf{0} \tag{5.2}$$

which is a second order differential equation, and it has a general solution which is the combination of pure exponential solutions. To make it clearer, we shall rewrite the equation of the motion as

$$\ddot{\mathbf{y}} = -\mathbf{M}^{-1}\mathbf{Ky} \tag{5.3}$$

And if we define

$$\mathbf{A} = \mathbf{M}^{-1}\mathbf{K} \tag{5.4}$$

Then

$$\ddot{\mathbf{y}} = -\mathbf{Ay} \tag{5.5}$$

Now recall that in calculus, with two initial conditions, typically, the displacement $\mathbf{y}(0)$ and the velocity $\dot{\mathbf{y}}(0)$, we assume that all solutions to the differential equation will have an exponential form, namely

$$\mathbf{y} = e^{st}\boldsymbol{\phi} \tag{5.6}$$

where $\boldsymbol{\phi} \in \mathbb{R}^n$ is the amplitude.

This is a pure solution of the linear differential equation, the other solutions will be mixtures of these pure solutions, and the mixtures are adjusted to fit the initial conditions. The pure solutions are the mode shapes of the system.

Now substituting this solution into Eq. (5.2), and define that $s = i\omega$ gives

$$\mathbf{A}\boldsymbol{\phi} = \omega^2 \boldsymbol{\phi} \tag{5.7}$$

The vector $\boldsymbol{\phi}$ must be an eigenvector of $\mathbf{A}$, the corresponding eigenvalue is now $\lambda = \omega^2$, this equation is the characteristic equation. Now we have reduced the ODE problem to an eigenvalue problem.

As we know from structural mechanics, the stiffness matrix $\mathbf{K}$ and mass matrix $\mathbf{M}$ are both symmetric positive definite matrices. However, the eigenvalues and eigenvectors of the characteristic equation depend on matrix $\mathbf{A}$. Unfortunately, matrix $\mathbf{A}$ is not generally symmetric. It means the eigenvectors of the characteristic equation do not have to be orthogonal to each other directly. And to calculate the eigenvalues and eigenvectors from an asymmetric matrix with current basis (defined by $\mathbf{A}$) is very costly. For this reason, we often try to find another basis in the same vector space, which not only preserves the system energy but also can largely simplify the procedure to find the eigenvalues and eigenvectors. Basically, to change the base, the conventional way is to manipulate the mass matrix $\mathbf{M}$, which can be done by using Cholesky decomposition [194] [195].

5.2.2 M-orthogonality of the mode shapes

Since the mass matrix $\mathbf{M}$ is positive definite, therefore we can always split it into $\mathbf{D}^{\mathrm{T}}\mathbf{D}$ provided that $\mathbf{D}$ has independent columns. This can be achieved by using Cholesky decomposition and $\mathbf{D}^{\mathrm{T}}$ will be a lower triangular matrix. Now, if we substitute $\mathbf{M} = \mathbf{D}^{\mathrm{T}}\mathbf{D}$ into the characteristic equation, we will have

$$\mathbf{K}\boldsymbol{\phi} = \lambda \mathbf{D}^{\mathrm{T}}\mathbf{D}\boldsymbol{\phi}\mathbf{K} \tag{5.8}$$

As we know matrix $\mathbf{D}$ has independent columns, so it is non-singular, its inverse exists. In the meantime, if we assume a new vector $\mathbf{v}$ is related to $\boldsymbol{\phi}$ by the nonsingular matrix $\mathbf{D}$, $\mathbf{v} = 1/\sqrt{M}\,\mathbf{D}\boldsymbol{\phi}$, where $1/\sqrt{M}$ is a scaling coefficient, then plug this relation into this equation, we will have

$$\mathbf{K}\mathbf{D}^{-1}\mathbf{v} = \lambda \mathbf{D}^{\mathrm{T}}\mathbf{v} \tag{5.9}$$

And if we write $\boldsymbol{\Gamma}$ for $\mathbf{D}^{-1}$, and multiply through by $\left(\mathbf{D}^{\mathrm{T}}\right)^{-1} = \boldsymbol{\Gamma}^{\mathrm{T}}$, this becomes a standard eigenvalue problem for a single symmetric matrix $\boldsymbol{\Gamma}^{\mathrm{T}}\mathbf{K}\boldsymbol{\Gamma}$. Then Eq. (5.9) becomes

$$\boldsymbol{\Gamma}^{\mathrm{T}}\mathbf{K}\boldsymbol{\Gamma}\mathbf{v} = \lambda \mathbf{v} \tag{5.10}$$

The transformation from

$$\mathbf{A}\boldsymbol{\phi} = \lambda\boldsymbol{\phi} \text{ to } \boldsymbol{\Gamma}^{\mathrm{T}}\mathbf{K}\boldsymbol{\Gamma}\mathbf{v} = \lambda\mathbf{v} \tag{5.11}$$

More precisely, from

$$\mathbf{A} \text{ to } \boldsymbol{\Gamma}^{\mathrm{T}}\mathbf{K}\boldsymbol{\Gamma} \tag{5.12}$$

For the eigenvectors, the symmetric property of matrix $\boldsymbol{\Gamma}^{\mathrm{T}}\mathbf{K}\boldsymbol{\Gamma}$ forces its eigenvectors $\mathbf{y}$ to be orthogonal, i.e.,

$$\mathbf{v}_i^{\mathrm{T}}\mathbf{v}_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{5.13}$$

This will lead to the so-called M-orthogonal (where M is the acronym of mass) of the eigenvectors or orthogonality of the eigenvectors with respect to the mass matrix of $\mathbf{K}\boldsymbol{\phi} = \lambda\mathbf{M}\boldsymbol{\phi}$,

$$\mathbf{v}_i^{\mathrm{T}}\mathbf{v}_j = \frac{1}{\sqrt{M_i M_j}}\boldsymbol{\phi}_i^{\mathrm{T}}\mathbf{D}_i^{\mathrm{T}}\mathbf{D}_j\boldsymbol{\phi}_j = \frac{1}{\sqrt{M_i M_j}}\boldsymbol{\phi}_i^{\mathrm{T}}\mathbf{M}\boldsymbol{\phi}_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{5.14}$$

Particularly when $i = j$, we have

$$\boldsymbol{\phi}_i^{\mathrm{T}}\mathbf{M}\boldsymbol{\phi}_i = M_i \tag{5.15}$$

which is the so-called the i th modal mass.

From Eq. (5.14), we can conclude that the eigenvectors $\boldsymbol{\phi}$ of characteristic equation Eq. (5.2) are independent of each other, but the orthogonal property cannot be directly derived from the matrix $\mathbf{A}$ for the general case. If matrix $\mathbf{A}$ happens to be a symmetric matrix, then the eigenvectors of $\mathbf{A}$ will be orthogonal to each other directly, otherwise, the eigenvectors will satisfy the M-orthogonality, which we just defined.

Now if we have r $\boldsymbol{\phi}_i$ s and r $\mathbf{v}_i$ s, then we can group them into matrix $\boldsymbol{\Phi}$ and matrix $\mathbf{V}$, respectively, then

$$\mathbf{K}\boldsymbol{\Phi} = \boldsymbol{\Lambda}\mathbf{M}\boldsymbol{\Phi} \tag{5.16}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix contains all the eigenvalues.

With the matrix notation, Eq. (5.15) is given by

$$\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{M}\boldsymbol{\Phi} = \mathbf{m} \tag{5.17}$$

where $\mathbf{m} = \mathrm{diag}(M_1,\ldots,M_r)$ is the modal mass matrix.

So if we pre-multiply $\boldsymbol{\Phi}^{\mathrm{T}}$ to Eq. (5.16), we will have

$$\mathbf{\Phi}^{\mathrm{T}}\mathbf{K}\mathbf{\Phi} = \mathbf{\Lambda}\mathbf{\Phi}^{\mathrm{T}}\mathbf{M}\mathbf{\Phi} = \mathbf{\Lambda}\mathbf{m} \tag{5.18}$$

If we define $\mathbf{k} = \mathbf{\Lambda}\mathbf{m} = \mathrm{diag}\left(K_1, \ldots, K_r\right)$, then it gives the modal stiffness matrix, where K_i is the i th modal stiffness. Since $\lambda = \omega^2$, then

$$\omega_i^2 = \frac{K_i}{M_i} \tag{5.19}$$

On the other hand, let us recall the mode superposition technique we discussed in chapter 4. As a quick review, we suppose the displacement vector of the structure is denoted by $\mathbf{u}$, according to the mode superposition technique, $\mathbf{u}$ can be expressed in terms of a set of mode shapes $\mathbf{\Psi}$ of amplitudes $\mathbf{q}$,

$$\mathbf{u} = \mathbf{\Psi}\mathbf{q} \tag{5.20}$$

where $\mathbf{u} \in \mathbb{R}^d$ denote a sample vector of d sensors (or the number of the independent points that define the selected displacement patterns), which is known as the geometric coordinates. If the dynamic behaviour of the system can be represented by only r modes, then $\mathbf{q} \in \mathbb{R}^r$ denotes a sample vector of r modal coordinates. Accordingly, $\mathbf{\Psi}$ is a $d \times r$ mode shape matrix.

Note that the modal superposition technique given by Eq. (5.20) can also apply to the state vector $\mathbf{y}$ as well. With the mass normalised mode shape matrix, $\mathbf{y}$ is expressed as

$$\mathbf{y} = \mathbf{\Phi}\mathbf{q} \tag{5.21}$$

Normally, Eq. (5.20) and Eq. (5.21) is connected with an observation matrix $\mathbf{C}_d$, which is

$$\mathbf{u} = \mathbf{C}_d\mathbf{y} \tag{5.22}$$

Thus,

$$\mathbf{\Psi} = \mathbf{C}_d\mathbf{\Phi} \tag{5.23}$$

Furthermore, when $\mathbf{\Psi}$ is rank deficient but has full column rank, i.e., $\mathbf{\Psi}$ is a $d \times r$ matrix with r ($< d$) independent columns, we can establish the relation between $\mathbf{u}$ and $\mathbf{q}$ through a least square sense. For example, if we want to find $\mathbf{q}$ based on $\mathbf{\Psi}$ and $\mathbf{u}$, we can project the vector $\mathbf{u}$ in the d dimensional space onto the subspace determined by the r columns in $\mathbf{\Psi}$, then find the projection of $\mathbf{u}$ which is the closest to itself, that is

$$\hat{\mathbf{u}} = \mathbf{\Psi}\left(\mathbf{\Psi}^{\mathrm{T}}\mathbf{\Psi}\right)^{-1}\mathbf{\Psi}^{\mathrm{T}}\mathbf{u} \tag{5.24}$$

Where the estimated modal coordinates are given by

$$\hat{\mathbf{q}} = \left(\mathbf{\Psi}^{\mathrm{T}} \mathbf{\Psi} \right)^{-1} \mathbf{\Psi}^{\mathrm{T}} \mathbf{u} \tag{5.25}$$

So, in this subspace, Eq. (5.20) also stands, but it is relation between the estimated terms.

$$\hat{\mathbf{u}} = \mathbf{\Psi} \hat{\mathbf{q}} \tag{5.26}$$

Note that

$$\mathbf{\Psi}^{\dagger} = \left(\mathbf{\Psi}^{\mathrm{T}} \mathbf{\Psi} \right)^{-1} \mathbf{\Psi}^{\mathrm{T}} \tag{5.27}$$

is the pseudo inverse of $\mathbf{\Psi}$.

5.2.3 System damping

The M-orthogonal property of the mode shapes we discussed in section 5.2. plays a vital role in structural vibration analysis. However, this property is not applicable to all damping systems. Unfortunately, all real structures are damped to some extent, and in most cases, the actual damping mechanism of a system is unknown. To simulate the energy dissipating phenomenon, we have to model the damping force. Practically, there are two damping models we normally encounter in modal analysis, namely, viscous damping and hysteretic damping [73], [141].

Table 5.1 Summary of the system damping models

	Viscous damping	Hysteretic damping
Equation of motion	$\mathbf{M}\ddot{\mathbf{y}} + \mathbf{C}\dot{\mathbf{y}} + \mathbf{K}\mathbf{y} = 0$	$\mathbf{M}\ddot{\mathbf{y}} + \left(\mathbf{K} + \mathrm{i}\mathbf{D} \right)\mathbf{y} = 0$
Proportional damping	$\mathbf{C} = \alpha_0 \mathbf{M} + \alpha_1 \mathbf{K}$	$\mathbf{D} = \alpha_0 \mathbf{M} + \alpha_1 \mathbf{K}$
General damping	$\mathbf{C}$	$\mathbf{D}$

In table 5.1, a summary of the commonly used damping models is presented. Note that the proportional damping matrices $\mathbf{C}$ and $\mathbf{D}$ in the table are given in the Rayleigh damping form, where α_0 and α_1 are proportionality factors.

From the equation of motion for those two damping models, we can see that the hysteretic damping case can be treated as an eigenvalue problem if one endeavours to find the damped vibration modes of the system regardless of the form of the damping matrix. The orthogonality property of the damped mode shapes is directly valid for this case since its equation of motion has the same format as Eq. (5.2), one only needs to replace $\mathbf{K}$ to $\mathbf{K} + \mathrm{i}\mathbf{D}$. However, the eigenvector obtained for this case will be complex. Also, note that the concept of damped mode shapes should not be confused with the mode

shapes or normal modes we defined from the undamped characteristic equation of motion, as the damped mode shapes can be complex rather than real due to the occurrence of the damping.

For the viscous damping model, it is a general second order differential problem, which is more difficult than the eigenvalue problem if we use the general form of the damping element. However, when the damping matrix is proportional to the mass and stiffness matrices, it is theoretically easier for analysis. Because the proportional damping possesses a particular advantage that the modes of a structure with the proportional damping are almost identical to those of the undamped version of the model [73]. Therefore, it is common practice to analyse the undamped system in full then make a correction for the presence of the damping before we derive the modal properties of a proportionally damped system. However, this procedure is only valid in the case of this special type of distribution of damping, which may not apply to actual structures in modal tests [73]. Apart from this, the orthogonality property of the undamped mode shapes can be used to uncouple the damping forces as well, which is why the viscous damping model is widely used in engineering.

To be more specific, when the damping matrix is a linear combination of the mass and stiffness matrices, the M-orthogonality of the mode shapes of the undamped system can also diagonalise the damping matrix. In other words, the normal coordinates transformation serves to uncouple the damping forces in the same way that it uncouples the inertia and elastic forces.

Hence the Rayleigh form of the damping matrix will permit uncoupling of the equations of motion. In general, then, the orthogonal damping matrix may be of the form [141]

$$\mathbf{C} = \mathbf{M} \sum_{j=0}^{\infty} \alpha_j \left(\mathbf{M}^{-1} \mathbf{K} \right)^j \tag{5.28}$$

in which as many terms may be included as desired. Proportional damping is obviously contained in Eq. (5.28).

For a non-rotating linear structure, the damped mode shapes can be complex only if the damping is distributed in a non-proportional way [73], and this situation is widely seen in real structures. It is because the majority damping in the real structure is normally concentrated at the joints between components of a structural assembly, which does not have a proportional distribution, whilst the internal (hysteresis) damping of most structural elements is distributed essentially proportional to the stiffness distribution [73]. Therefore, this bias ingredient of non-proportionality for complex modes is likely to exist in most structures [73]. Besides, it must be pointed out that non-proportionality is only a necessary condition for the presence of the complex modes, it is not sufficient, at least not if the degree of complexity is to be other than trivial [73].

5.2.4 Complexity of the mode shapes

In the last section, we had a brief introduction about the system damping and its impact on the mode shapes. In this section, we want to elaborate on the definition of the mode shapes and give a further discussion about the complexity of the mode shapes. It is worth mentioning that the definition of the mode shapes varies from field to field. Even in engineering, people have a different understanding of what mode shapes are.

For example, in structural dynamics, the mode shapes, also known as normal modes, are defined according to the characteristic equation Eq. (5.2), which will always be real. While researchers working particularly in modal analysis, they see normal modes as the characteristic modes of the structure in its actual damped state [73], which could either be real or complex, and the imaginary part of the mode shape contains some important phase information. Based on such definition, the complex modes are commonly seen in practice. Accordingly, researchers in the modal analysis would argue that the estimated mode shapes are complex is normally due to the following reasons, such as system damping (not all system damping result in complex mode shapes, see section 5.2.3), poor measurement or analysis method or the occurrence of the repeated poles.

For the mode shape complicity caused by the analysis method, a good example is the estimated mode shapes obtained based on the classical FDD technique. The classical FDD always gives complex mode shape estimates even if the structure has real modes.

Clearly, these two definitions about the mode shapes are not consistent. Apart from this, it is believed that the real mode has the appearance of a standing wave, while the complex mode is better described as exhibiting of a travelling wave [73]. Such perception about the real mode and the complex mode is also misleading. The appearance of the standing wave or the travelling wave is our observation regarding the deflection shape of the structure, not the mode shape itself. The deflection shape is usually the linear combination of different mode shapes. Besides, if each mode is seeing as a standing wave, then the linear combination of two or more standing waves will produce a travelling wave.

It is true that when the damping force (non-proportional) is introduced into Eq. (5.1), the calculated mode shapes will be complex. For each mode, we need to use two eigenvectors (they are complex conjugates) to describe it in order to generate the real displacement. This is similar to the case of repeated poles that one frequency is corresponding to the linear combination of two eigenvectors. As a result, we constantly observe complex modes for an symmetric structure even in the absence of damping [73].

Suppose a general expression of the structural response can be given by

$$y(x,t) = Be^{i(\kappa x + \omega t)} \tag{5.29}$$

for some frequency ω and wave number κ. $y(x,t)$ can be seen as the continuous form of the state vector $\mathbf{y}(t)$ described in two-dimensions, and x is the dimension along the DOFs of the structure.

Define

$$w(x) = Be^{i\kappa x} \tag{5.30}$$

where B is the amplitude, and it can be real or complex. Then $w(x) = y(x,0)$ gives the initial state of vibration, which we can see it as mode shapes.

Since there is no complex value in the real world, the observation of $y(x,t)$ is given by

$$y(x,t) = w(x)e^{i\omega t} + w(x)^* e^{-i\omega t} = Be^{i(\kappa x + \omega t)} + B^* e^{-i(\kappa x + \omega t)} \tag{5.31}$$

where B^* is the complex conjugate of B.

If $B = \beta_1 + i\beta_2$, expand the above equation, we obtain

$$y(+x,t) = 2\sqrt{\beta_1^2 + \beta_2^2}\cos(\omega t + \kappa x + \theta) \tag{5.32}$$

where $\theta = \arctan(\beta_2/\beta_1) \in [0, \pi/2]$. Eq. (5.32) represents a leftward-moving sinusoidal wave. Here, we added a plus sign in front x to demonstrate the direction of the travelling wave. From Eq. (5.31) and Eq. (5.32), we learned that if $w(x)$ is complex, the net result of $y(x,t)$ is a travelling wave, which is what we observe in real life.

Similar to Eq. (5.31), the observation of $w(x)$ in real life is given by is

$$w(x) = Be^{i\kappa x} + B^* e^{-i\kappa x} \tag{5.33}$$

Expand Eq. (5.33), we get

$$w(x) = 2\beta_1 \cos(\kappa x) - 2\beta_2 \sin(\kappa x) \tag{5.34}$$

Or we can express it as

$$w(x) = 2\sqrt{\beta_1^2 + \beta_2^2}\cos(\kappa x + \theta) \tag{5.35}$$

which means the complex expression of $w(x)$ given by Eq. (5.30) is corresponding to a real shape expressed as Eq. (5.35).

Since the $w(x)$ given by Eq. (5.34) is two times the real part of $w(x)$ given by Eq. (5.30). Thus, it does not matter if we substitute the real part of Eq. (5.30) or the real expression of $w(x)$ given by Eq. (5.35) into Eq. (5.31). Suppose we substitute the real express of $w(x)$ given by Eq. (5.35) into Eq. (5.31), it simply becomes

$$y(x,t) = 4\sqrt{\beta_1^2 + \beta_2^2}\,\cos(\omega t)\cos(\kappa x + \theta) \tag{5.36}$$

which represents standing wave in real life, with $\left|4\sqrt{\beta_1^2 + \beta_2^2}\,\cos(\omega t)\right|$ being the amplitude. This corresponds to the case that if we have real mode shapes, the net result of our observation is a standing wave.

Now consider the small-amplitude free vibration of a uniform string of unit length, the string lies at rest along the interval $0 \le x \le 1$, prior $t = 0$, and it has two fixed ends, and $w(0) = w(1) = 0$. With such a hard boundary condition, based on Eq. (5.32) the rightward-moving wave can be given by

$$y(-x,t) = 2\sqrt{\beta_1^2 + \beta_2^2}\,\cos(\omega t - \kappa x + \theta) \tag{5.37}$$

The total wave is therefore,

$$y(x,t) = y(+x,t) + y(-x,t) = -4\sqrt{\beta_1^2 + \beta_2^2}\,\sin(\omega t + \theta)\sin(\kappa x) \tag{5.38}$$

which represents a standing wave, with $\left|4\sqrt{\beta_1^2 + \beta_2^2}\,\sin(\omega t + \theta)\right|$ being the amplitude. This is the standing wave generated by a so-called complex mode. Similarly, based on Eq. (5.36), which is derived according to a real mode, we can obtain a standing wave as well,

$$y(x,t) = 8\sqrt{\beta_1^2 + \beta_2^2}\,\cos(\omega t)\cos(\theta)\cos(\kappa x) \tag{5.39}$$

with $\left|8\sqrt{\beta_1^2 + \beta_2^2}\,\cos(\omega t)\cos(\theta)\right|$ being the amplitude.

Particularly for this special case that $w(0) = w(1) = 0$, if we substitute this boundary condition into Eq. (5.34), we have $\beta_1 = 0$ and $\theta = \pi / 2$, B is pure imaginary, then Eq. (5.38) becomes

$$y(x,t) = -4\beta_2\cos(\omega t)\sin(\kappa x) \tag{5.40}$$

On the other hand, if we substitute $\beta_1 = 0$ and $\theta = \pi / 2$ into Eq. (5.36), we obtained

$$y(x,t) = 4\beta_2\cos(\omega t)\sin(\kappa x) \tag{5.41}$$

which is equivalent to Eq. (5.40).

According to our analysis, even if the structure has complex mode shape, the structure can still vibrate like a standing wave, as long as the combination of the leftward-moving wave $y(+x,t)$ and reflection wave $y(-x,t)$ produces a standing wave. Therefore, it is not justified to use the observed standing wave and the travelling wave to refer to the real or complex mode shapes without mentioning the boundary conditions. However, if there is some energy loss at the reflection point or along the string, which alters the wavenumber, phase, or the amplitude of the leftward-moving wave and the reflection wave, the combination of them is unlikely to generate a standing wave, what we see will still be a travelling wave.

It should be pointed out that the contribution of the force is not included here. It is true that the forced excitation response will constantly show relative phase differences between the structural responses of different locations, which, as a result, indicate a complex mode [73]. However, when talking about the normal modes, we normally refer to each mode shape with a certain natural frequency. Under forced vibration, when the excitation frequency of the structure differs a lot from its natural frequencies, the vibration shape will not be able to reveal the mode shape of the system. Even if the operating deflection shape reflects the shape of the nearby mode when an excitation frequency is close to one of the natural frequencies of the system, due to the contribution of all other modes (albeit small), such force-induced deflection shape will not be identical to the shape of the nearby mode [73], and it is usually referred to the forced vibration mode, or an operating deflection shape (ODS) [73]. Here, we shall separate this forced vibration mode from our discussion of the normal modes of the system.

To sum up, we separate the complexity of the mode shapes into two categories: theoretical complexity and estimation complexity or observation complexity. We recognise the complexity of the mode shapes generated by the introduction of the system damping as the theoretic complexity. As for the estimation complexity, we admit it as the contribution of the system damping, repeated poles, bad measurements, analysis method, etc. It means that the estimation complexity of the mode shapes for the real structure is the combination of all sources. Whether we should discard the imaginary part of the estimated mode shapes depends on the damping distribution of the system and how significant the damping force is.

Additionally, when we say the structure has real modes, we normally refer this to our observation of the standing wave, which is a misleading concept. Nevertheless, when a structure, such as a long-span bridge, has hard boundaries and its damping is very low, we can still observe the standing wave when it vibrates with a dominant mode. Therefore, we usually deem such structure has real modes.

5.2.5 Decoupling the damped system

In this section, we consider a linear damped system with the proportional damping element. The system equation of motion is presented as follows with the introduction of the damping force to Eq. (5.1), that is

$$\mathbf{M\ddot{y}} + \mathbf{C\dot{y}} + \mathbf{Ky} = \mathbf{f} \tag{5.42}$$

In terms of our aforementioned discussion of the system damping, we know that for a viscous damping model, when the damping matrix is proportional to the mass and stiffness matrix, it can be decomposed in the same way as the mass and stiffness matrix. Therefore, we can utilise the M-orthogonal property of the mode shapes we obtained from the undamped system given by Eq. (5.2) to decouple the system expressed as Eq. (5.42).

Now first substituting Eq. (5.21) into this equation using the mass normalised mode shapes, we obtain

$$\mathbf{M\Phi\ddot{q}} + \mathbf{C\Phi\dot{q}} + \mathbf{K\Phi q} = \mathbf{f} \tag{5.43}$$

Then if we pre-multiply both sides by $\mathbf{\Phi}^{\mathrm{T}}$, we arrive at

$$\mathbf{\Phi}^{\mathrm{T}}\mathbf{M\Phi\ddot{q}} + \mathbf{\Phi}^{\mathrm{T}}\mathbf{C\Phi\dot{q}} + \mathbf{\Phi}^{\mathrm{T}}\mathbf{K\Phi q} = \mathbf{\Phi}^{\mathrm{T}}\mathbf{f} \tag{5.44}$$

where $\mathbf{P} = \mathbf{\Phi}^{\mathrm{T}}\mathbf{f}$ is the so-called modal force, and Eq. (5.44) denotes the decoupled system equation of motion, which can be further expressed as

$$M_i\ddot{q}_i + C_i\dot{q}_i + K_iq_i = P_i, \quad i = 1,\ldots,r \tag{5.45}$$

where C_i is the i th modal damping coefficient, which can be further given by

$$C_i = 2M_i\xi_i\omega_i \tag{5.46}$$

where ξ_i is i the damping ratio associated with the mode shape $\mathbf{\phi}_i$.

Therefore, from Eq. (5.19) and Eq. (5.46), Eq. (5.45) can be simplified as

$$\ddot{q}_i + 2\xi_i\omega_i\dot{q}_i + \omega_i^2q_i = \frac{P_i}{M_i} \tag{5.47}$$

It should be pointed out that Eq. (5.42) is the general equation of motion for a second order linear system, and it imposes no restriction on the form of the external forces, which means it also applies to the moving force. The mode shape matrix can be used to transform the geometric forces into the modal forces corresponding to each mode.

5.3 Two-stage system identification method

Having discussed the mode shape properties in the previous sections, now we can present the two-stage system identification strategy. As previously mentioned, the mode shape matrix serves as a medium to transform the geometric coordinates to modal coordinates and external forces to modal forces. When diagonalization of the damping matrix is achievable also, a damped bridge system can be decomposed into a series of uncoupled SISO systems using the mode shape matrix in the case of a moving force.
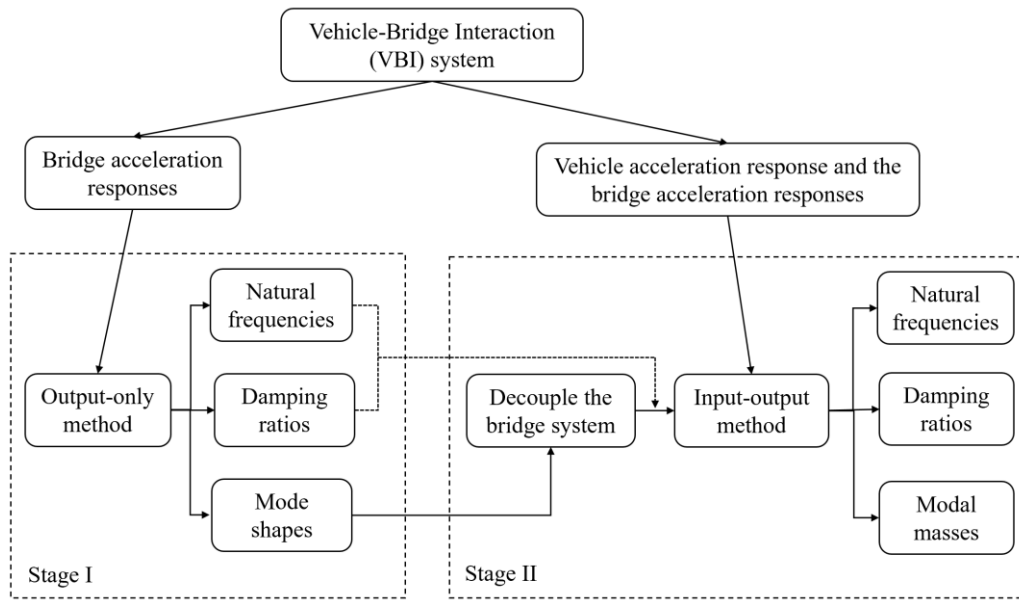
Figure 5.1. The flow chart of the proposed two-stage identification method

Therefore, in the first stage, we can use an appropriate output-only method, such as the FDD technique, to identify the mode shapes of the bridge. Then use the estimated bridge mode shapes to decouple the bridge system. The modal parameters, i.e., natural frequencies, damping ratios, and the modal masses, can then be identified in the second stage. Since the first stage techniques have been discussed extensively in the previous chapter, this section mainly focuses on presenting the second stage strategy, and an optimisation procedure and a least square method are discussed. The diagram shown in Figure 5.1 gives the flow chart of the two-stage process. Note that the identified natural frequencies and damping ratios in the first stage are not necessary inputs for the second stage identification. If we use an optimisation method in the second stage, they can be used as initial guesses. However, if a least square method is adopted, then those estimates are redundant.

5.3.1 Modal force of the moving load

Suppose the tires of a vehicle are connected to the bridge surface all the time when traversing a bridge, then the decoupled equation of motion for the bridge can be easily derived using the continuous form of Eq. (5.47). Figure 5.2 shows an example of a moving load traversing a bridge with length L.
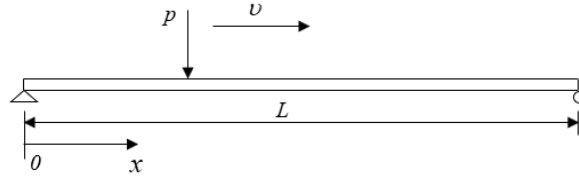
Figure 5.2. Simply supported Euler-Bernoulli beam model with moving point load

Now define $p(x,t)$ as the distributed vertical loading on a bridge, where x denotes the location of the vertical loading. The continuous form of the decoupled bridge system is given by

$$\ddot{q}_i(t) + 2\xi_i\omega_i\dot{q}_i(t) + \omega_i^2 q_i(t) = \frac{1}{M_i}\int_0^L \phi_i(x)p(x,t)\,\mathrm{d}x \qquad (5.48)$$

where L is the bridge length. $\phi_i(x)$ is the i th continuous mode shape.

In the case of a moving concentrated load with a given time varying intensity $f(t)$ and speed υ, the load $p(x,t)$ can be expressed as [196]

$$p(x,t) = \begin{cases} \delta(x-\upsilon t)f(t) & 0 \le t \le t_d \\ 0 & t > t_d \end{cases} \qquad (5.49)$$

where $t_d = L/\upsilon$ is the time required for the load to cross the bridge of length L.

With Eq. (5.49), the right hand side of the Eq. (5.48) is given by

$$f_i = \begin{cases} \dfrac{1}{M_i}\int_0^L \phi_i(x)\delta(x-\upsilon t)f(t)\,\mathrm{d}x = \dfrac{1}{M_i}\phi_i(\upsilon t)f(t) & 0 \le t \le t_d \\ 0 & t > t_d \end{cases} \qquad (5.50)$$

5.3.2 Stage I — estimation of the bridge mode shapes

In this section, we shall answer two questions. One is whether or not we can use the mode shapes we estimated from the bridge responses under moving force excitation to decouple the bridge system. The other one is whether we should use the real or the complex mode shapes to decouple the system, as the mode shape estimates can be complex-valued.

5.3.2.1 Legitimacy of the usage of the estimated mode shapes

In our VBI bridge modal parameters analysis, we shall separate the bridge state into two-stages: vehicle moves across on the bridge, and vehicle leaves the bridge. When we use the output-only identification technique such as the FDD to estimate the bridge mode shapes based on the bridge output responses only when the vehicle is moving across the bridge, the mode shapes we identified is actually the mode shape of the whole VBI system (i.e., the combination of the vehicle subsystem and the bridge subsystem). When the vehicle is moving across the bridge, the location of the vehicle is changing all the time. As a result, the configuration of the whole VBI system is changing all the time. In other words, the mass and stiffness matrices of this VBI system are not constant anymore.

Apparently, the mode shapes of the whole VBI system are different from the bridge's actual mode shapes. However, as both the size and weight of the vehicle are usually smaller than the bridge, the bridge's vibration will still dominate the vibration of the whole system. Apart from this, when the excitation frequency is close to being one of the system's natural frequencies, the operating deflection shape will reflect the shape of the nearby mode but will not be identical to it because of the contributions, albeit small, of all the other modes [73]. Therefore, we can potentially use the estimated mode shapes of the whole VBI system to represent the bridge's estimated mode shapes, provided that the estimated mode shapes of the whole VBI system and the bridge mode shapes are highly correlated.

After the vehicle exits the bridge, the bridge subsystem enters a free vibration stage. It is beneficial to add the free vibration part into the VBI analysis of the bridge subsystem. This is because the longer the time history, the smaller the PSD bias we have if we use the FDD to estimate the mode shapes. Therefore, the whole-time history of the bridge system is used in our analysis, and it is used in both stages of the proposed method.

Besides, it is worth mentioning that it is commonly believed that the VBI effect can offer an additional damping effect (i.e., an apparent increase in the damping ratio) and additional mass effect (i.e., an apparent decrease in natural frequency) on the railway bridge response [197], and it is improper to apply the FDD on the vibration responses excited by spatially correlated excitation [14], as it might result in falsely identified structural dynamic properties. Nonetheless, this is not a problem for this study for the following reasons.

Although the estimated bridge properties by using the FDD may result in erroneous results under the vehicle excitation, the actual context studied in [14] is under the traffic excitation, which is similar to the case of a passing train scenario. The actual mass ratio between the vehicle and the bridge is small, and the estimation is based on the output-only time histories of the bridge under forced vibration. As a result, the estimated resonance frequencies of the bridge based on an output-only identification method can be effectively alerted. Whereas the proposed method in this study only concerns a single vehicle, the mass ratio between the vehicle and the bridge is significantly smaller than the traffic excitation case.

Apart from this, not only the VBI interaction responses of the bridge but also the free vibration of the bridge responses is included in the identification procedure. This is essential because the inclusion of the free decay part of bridge responses will introduce more energy to the bridge spectral density at each frequency point, including the resonance frequencies. When the bridge properties do not change with time, more data means more information and better estimation. It is true that when we use the forced vibration responses only of the bridge to estimate the bridge properties, we will end up with biased results. But when the bridge vibration is dominated, and the free decay vibration of the bridge is included in the analysis, the effect of the vehicle frequency that contributes to the bridge spectral density is limited. It means that a frequency domain technique such as the FDD should be able to extract the information we need from the bridge forced and free vibrations. In chapter 6, we will present our results of the FDD application in terms of two VBI systems.

5.3.2.2 Complexity problem of the estimated mode shapes

As for the complexity problem of the estimated mode shapes, from our previous discussion about the mode shape, we know that the observed complex modes are common in practice, and its complexity is depending on the damping force, repeated poles, measurements, or analysis method. Among the sources of the complexity, we wish to preserve the one contributed by the system damping. However, it is not possible for us to separate it from the influence of other sources, such as the ones come from bad measurements and analysis method. Note that the complexity of the estimated mode shapes for the repeated pole case is just a result of our observation. It does not necessarily mean the system has complex modes.

Now this complexity property of the estimated mode shapes poses a problem for the two-stage system identification method we proposed in this thesis because the second stage identification of the proposed method depends on the goodness of the estimated mode shapes. Therefore, whether to use the real or the complex mode shapes to decouple the system is an important question to answer for this work.

For a structure with low damping, its theoretical mode shapes can be real or close to real. In this case, most of the complexity of the estimated mode shape would come from bad measurements and analysis methods and such contribution might be more significant than the one contributed by the damping force.

In this case, keeping the imaginary part of the estimated complex mode shape will introduce extra errors for further analysis. On the other hand, the complex mode shape always corresponds to a real shape (see Eq. (5.35)), which we can perceive in real life, and that real shape is twice the amplitude of the real part of the complex mode shape. When a structure has close to a real mode shape with low damping, it is more appropriate to use the real part of the estimated mode shape or its corresponding real shape to represent the mode shapes of the structure.

Besides, it must be pointed out that the essence of the modal superposition technique is to use a set of base functions to decompose the measurements into a set of independent modal coordinates. As long as the basis functions are independent of each other, we can always find a rotated version of them to make it satisfy the M-orthogonal property of the mode shapes we talked about in the last section. Since in the second stage of our method, all we need is to use the M-orthogonal property of the mode shapes to decouple the bridge system, as long as we have independent bases, the second stage analysis is valid.

Additionally, the real-valued one-sided spectral density matrix we presented in chapter 3, if used in conjunction with the FDD, the mode shapes we estimated will be real immediately. Since the simulation case used in the next chapter is a simply supported beam, and it is assumed to have proportional viscous damping, the structure has real mode shapes theoretically. Therefore, we will use the estimated real mode shapes to decouple the system, and the real-valued one-sided spectral density matrix is adopted with the application of the FDD in this study.

Furthermore, it is also worth mentioning that, as our purpose is to identify the modal parameters of the bridge, what we decoupled is the bridge system. The vehicle system only provides the external force. Thus, we are not talking about using the bridge mode shapes to decouple the vehicle system here, but we are talking about using the bridge mode shapes to decouple the measured response of the bridge system.

5.3.3 Stage II — construction of the Frequency Response Functions (FRFs) for the moving force excited system

Eq. (5.48) gives the general form of the decoupled equation of the motion of the beam system subjected to a moving force. When the force is on the bridge, the structure undergoes forced vibration. After the force leaves the bridge, it will experience free vibration until the vibration dies out because of the damping.

If we assume the bridge is at rest before the moving force enters the bridge, and we measure its vibration from the time the moving force enters the bridge until it returns to its static state, we can still use Eq. (5.48) to described the vibration of the structure for each mode, due to the additive property of a linear system. In other words, if Eq. (5.48) represents the moving force induced vibration equation of motion

of the bridge and the free vibration of the bridge after the vehicle leaves the bridge, which is given by the following equation

$$\ddot{q}_i^f(t) + 2\xi_i\omega_i\dot{q}_i^f(t) + \omega_i^2 q_i^f(t) = 0 \tag{5.51}$$

can be added to Eq. (5.48).

Therefore, we can express $q_i(t)$ as

$$q_i(t) = q_i^v(t) + q_i^f(t) = \begin{cases} q_i^v(t) & 0 \le t \le t_d \\ q_i^f(t) & t > t_d \end{cases} \tag{5.52}$$

Now if we take the Fourier transform of both sides of Eq. (5.48),

$$\int_0^\infty \ddot{q}_i(t)e^{-i\omega t}\mathrm{d}t + 2\xi_i\omega_i\int_0^\infty \dot{q}_i(t)e^{-i\omega t}\mathrm{d}t + \omega_i^2\int_0^\infty q_i(t)e^{-i\omega t}\mathrm{d}t$$

$$= \frac{1}{M_i}\int_0^{t_d}\int_0^L \phi_i(x)p(x,t)e^{-i\omega t}\mathrm{d}x\mathrm{d}t \tag{5.53}$$

Denote

$$\tilde{q}_i(\omega) = \frac{1}{2\pi}\int_0^\infty q_i(t)e^{-i\omega t}\mathrm{d}t \tag{5.54}$$

and

$$\tilde{P}_i(\omega) = \frac{1}{2\pi}\int_0^{t_d}\int_0^L \phi_i(x)p(x,t)e^{-i\omega t}\mathrm{d}x\mathrm{d}t \tag{5.55}$$

In the meantime, observe that,

$$\tilde{\dot{q}}_n(\omega) = \frac{1}{2\pi}\int_{-\infty}^\infty \dot{q}_n(t)e^{-i\omega t}\mathrm{d}t = i\omega\tilde{q}_n(\omega)$$

$$\tilde{\ddot{q}}_n(\omega) = \frac{1}{2\pi}\int_{-\infty}^\infty \ddot{q}_n(t)e^{-i\omega t}\mathrm{d}t = -\omega^2\tilde{q}_n(\omega) \tag{5.56}$$

Substitute Eq. (5.54) to Eq. (5.56) into Eq. (5.53), we obtain

$$-\omega^2\tilde{q}_i(\omega) + i2\xi_i\omega_i\omega\tilde{q}_i(\omega) + \omega_i^2\tilde{q}_i(\omega) = \frac{\tilde{P}_i(\omega)}{M_i} \tag{5.57}$$

Now rearrange Eq. (5.57), we shall have the Frequency Response Function for our moving force excitation case

$$H_n(\omega) = \frac{\tilde{q}_n(\omega)}{\tilde{P}_n(\omega)} = \frac{1}{M_n\left(\omega_n^2 - \omega^2 + i2\xi_n\omega_n\omega\right)} \tag{5.58}$$

which has the same format as the normal FRF, the only difference is given by the generalised moving force. This equation is significant. It is because the once highly correlated MIMO system has been transformed into a series of SISO systems, and the FRF we are concerned with here is the simple direct point FRF, no direct transfer or cross transfer FRFs are associated. Thus, we can use some very simple optimisation or regression methods to identify the parameters from the above equation.

Besides, it is worth mentioning that by using (5.58) to estimate the natural frequencies of the system shall result in better estimations if we compare the results with the estimates we obtain from any output-only method. This is because, in the construction of the FRF, the effect of the input carried through the output-only method is eliminated.

5.3.3.1 Optimisation of the Accelerance

In practice, the dynamic response we normally measure on structures is the acceleration. Therefore, for practical use, we shall use $\tilde{\ddot{q}}_i(\omega)$ rather than $\tilde{q}_i(\omega)$ in Eq. (5.58). Then we have

$$I_i(\omega) = \frac{\tilde{\ddot{q}}_i(\omega)}{\tilde{P}_i(\omega)} = \frac{-\omega^2 \tilde{q}_i(\omega)}{\tilde{P}_i(\omega)} = -\omega^2 H_i(\omega) = \frac{-\omega^2}{M_i\left(\omega_i^2 - \omega^2 + \mathrm{i}2\xi_i\omega_i\omega\right)} \tag{5.59}$$

which is called the Accelerance.

In order to identify the modal parameters ω_i, ξ_i and M_i in Eq. (5.59), we can choose an appropriate optimization method or a hybrid optimization strategy [198] to realise it.

In numerical optimization, the aim is to minimise an objective function, depends on the parameters of the system, the general mathematical formulation in this case can be written as

$$\min_{\eta_i \in \Omega} J\left(\eta_i; \omega\right) \tag{5.60}$$

where $J\left(\eta_i; \omega\right)$ is the objective function, and $\eta_i = \left(\omega_i, \xi_i, M_i\right)$ represents the three-dimensional parameter vector. It is chosen so that the objective function $J\left(\eta_i; \omega\right)$ best agrees with the observation in some sense. Ω is the feasible set, in unconstrained optimization, it equals $\mathbb{R}^m$; in constrained optimization scenario, it is a subset of $\mathbb{R}^m$, and parameters will be subject to some constraints in the meantime [199].

Now with the expression of the Accelerance given by Eq. (5.59) we consider the following objective functions,

$$J\left(\eta_i; \omega\right) = \frac{1}{2}\sum_{j=1}^{m} R_j^2\left(\eta_i\right) \tag{5.61}$$

where m is the number of points, and $R_j\left(\eta_i\right)$ can have the following three forms

$$R_{1j}(\eta_i) = I_i(\omega_j) - I_{ij} = \text{Re}\Big[I_i(\omega_j)\Big] - \text{Re}\Big[I_{ij}\Big] + \text{i}\Big(\text{Im}\Big[I_i(\omega_j)\Big] - \text{Im}\Big[I_{ij}\Big]\Big) \qquad (5.62)$$

and

$$R_{2j}(\eta_i) = \Big|I_i(\omega_j) - I_{ij}\Big| = \sqrt{\Big(\text{Re}\Big[I_i(\omega_j)\Big] - \text{Re}\Big[I_{ij}\Big]\Big)^2 + \Big(\text{Im}\Big[I_i(\omega_j)\Big] - \text{Im}\Big[I_{ij}\Big]\Big)^2} \qquad (5.63)$$

and

$$R_{3j}(\eta_i) = \Big|I_i(\omega_j)\Big| - \Big|I_{ij}\Big| = \sqrt{\left(\sqrt{\text{Re}\Big[I_i(\omega_j)\Big]^2 + \text{Im}\Big[I_i(\omega_j)\Big]^2} - \sqrt{\text{Re}\Big[I_{ij}\Big]^2 + \text{Im}\Big[I_{ij}\Big]^2}\right)^2} \qquad (5.64)$$

where $I_i(\omega_j)$ and I_{ij} represent the parametrized model and the observed Accelerence, respectively. m is the number of residuals. In this case, it is also the number of selected frequency lines from the transfer function, and the range of which is chosen here as $[0.9, 1.1]$ times the natural frequency of the mode. Note that the sum of squares on the right side of the equation is divided by 2, which has no actual effect on the optimal solution. It's introduced only for the convenience of the differentiation so that the square exponent will be cancelled out. More generally, Eq. (5.61) can be normalised by the number of summands m [56], then the objective function becomes the average error per data point, or in a sense, the mean square error between the predicted model and the measured sample divided by 2. Then the sum of squared errors from different datasets with differing sizes are comparable since the error has been averaged over its sample length.

Among the three representations of $R_j(\eta_i)$, $R_{2j}(\eta_i)$ and $R_{3j}(\eta_i)$ are continuous functions. Thus, we can treat them as a least-square problem. Methods such as the Gauss-Newton line search method and the Levenberg-Marquardt trust-region method can be applied. As for $R_{1j}(\eta_i)$, it is a direct measure of the discrepancy between the model and the observations. Here, it represents the difference between the observed and predicted FRFs. Since the least square methods need to take the partial derivative of $R_{1j}(\eta_i)$ directly to form a Jacobian matrix, which, in this case, is a non-smooth function ($R_{1j}(\eta_i)$ is a complex function) in $\mathbb{R}^m$. As a result, the direct application of least square methods for the objective function with $R_{1j}(\eta_i)$ is not available. Nevertheless, the objective function itself is continuous – its derivative information can still be calculated through either analytical or numerical methods (finite differencing) [199].

In principle, one can choose other forms of the objective function other than Eq. (5.61) to optimize. For example, the conventional Circle-Fitting method [64], which is trying to find the FRF circle centre position and the radius, uses the mass normalised mode shape, but the mass information is incorporated

in the modal constant, which is not a direct variable in the objective function. In other words, the modal mass is not being optimized, and without the mode shape information, it's unable to be extracted neither.

In order to formulate the optimisation strategy, we shall notice that the problem is dealt with herein has natural constraints on the variables. Specifically, the natural frequencies, damping ratios, and modal masses have positive values. However, it may be safe to disregard these constraints as they may not affect the solution and do not interfere with algorithms [66]. Therefore, in theory, both unconstrained optimization and constrained optimization techniques are applicable.

Generally speaking, the global solution for an optimization problem is difficult to find, but the situation may be improved when constraints are added, as the feasible set might exclude many of the local minima, and it may be comparatively easy to pick the global minimum from those that remain [66]. Nonetheless, there are some issues that need to be addressed here, the most important one being the effect of constraints imposed on the variables. Even though constrained optimization methods have some advantages to trace the global minimum as previously mentioned, constraints can also make things more difficult. In other words, in some circumstances, the problem has a unique solution without constraints, but with constraints, the algorithm could illegalise certain points that might otherwise be the global optimum or find several feasible points near the neighbourhood of the local minimiser [199]. This means the algorithm could converge at different local solutions for each run even if the same sequence is used. In some cases, the different constraints added to the function could result in different results.

For this reason, a hybrid optimization strategy is adopted. Specifically, the search for a minimiser begins with two unconstrained optimization methods, followed by a constrained optimization method with a "hot start" obtained from the previous step. This unconventional procedure is used to avoid the unpleasant termination of the mostly preferred unconstrained optimization program, which might stop at a local minimum, which is far from a global minimum. In the meantime, it can preserve both the advantages of the constrained and the unconstrained optimization methods to ensure a solution to Eq. (5.61) will always be found in the reasonable range of the parameters.

The first unconstrained optimization algorithm used here is a first-order method or direct search method, which does not require derivative information and approximates the objective without recourse to Taylor expansions. In other words, such methods rely exclusively on values of the objective function and compare each trial solution with the best previous solution. Basically, there are three methods that can be drawn from this genre, pattern search methods, simplex methods (not the simplex method for the linear programming), and methods with adaptive sets of search directions [198], [200]. These three basic direct search methods are readily available in the MATLAB toolbox. Particularly for the non-smooth problems, techniques like the simplex method can be superior to other approaches. In this research, the simplex algorithm of Nelder and Mead [201] is used.

Even though gradient-free methods can find many good solutions and can be used to deal with nondifferentiable problems, they are not necessarily guaranteed to find the true global optimal solutions. In other words, as one of the disadvantages mentioned before, the direct search methods used here can sometimes stop at places that are far from the global minimum, which will give a defective "hot start" for the next step. On the other hand, with the gradient information, the gradient-based optimizers are efficient at finding local minima for nonlinear problems. Hence, another method based on the trust-region technique, namely Levenberg-Marquardt method [199], is employed at the same time to ensure the next step can enjoy superior local convergence properties.

To be more specific, both $R_{1j}(\eta_i)$ and $R_{2j}(\eta_i)$ will be used in the first step to generate the "hot start". For the objective function using $R_{1j}(\eta_i)$, the unconstrained derivative-free optimisation method is used. While for the objective function based on $R_{2j}(\eta_i)$, the unconstrained gradient-based optimisation method is applied. Then $R_{3j}(\eta_i)$ is used in the second step to form a constrained optimisation problem. As for the constraints, the natural frequencies are bounded between zero and the Nyquist frequency, while the constraints of the damping ratios and modal masses, in this case, are released up to the software precision.

Optimization algorithms are iterative. They need an initial guess of the parameters and generate a sequence of the improved estimates until they terminate. In many cases, good initial guesses result in better optimisation results. For the first two algorithms, the starting points are selected based on the previously identified information (natural frequencies and damping ratios) and general knowledge of the parameters (modal masses). Since two algorithms are being used in the first step, two groups of potential solutions are available when both of them successfully terminate at a solution. There will be eight combinations of the results. To select the best starting point for the next step, all eight combinations are substituted into Eq. (5.61) with $R_{3j}(\eta_i)$ to see which combination gives the least value. By using this hybrid method, the whole optimisation program will be less sensitive to small changes to the initial conditions.

5.3.3.2 Regression of the Apparent mass

Instead of optimising the Accelerance directly, which needs an initial estimate, we can evaluate its inverse, which is called the Apparent mass, by using the regression method, which does not require an initial guess.

The Apparent mass is defined as

$$A_i(\omega) = I_i(\omega)^{-1} = M_i \left(1 - \frac{\omega_i^2}{\omega^2} - \mathrm{i}\frac{2\xi_i\omega_i}{\omega} \right) \tag{5.65}$$

Since $K_i = M_i\omega_i^2$, we can rewrite the above equation as

$$A_i(\omega) = M_i - K_i \frac{1}{\omega^2} - \mathrm{i}2\xi_i \sqrt{K_i M_i} \frac{1}{\omega} \tag{5.66}$$

Now we can separate the real and imaginary parts of the Apparent Mass,

$$\mathrm{Re}\left[A_i(\omega) \right] = M_i - K_i \frac{1}{\omega^2}$$
$$\mathrm{Im}\left[A_i(\omega) \right] = -2\xi_i\omega_i M_i \frac{1}{\omega} \tag{5.67}$$

Further, we can get

$$\mathrm{Re}\left[A_i(\omega) \right] = \begin{pmatrix} 1 & \dfrac{1}{\omega^2} \end{pmatrix} \begin{pmatrix} M_i \\ -K_i \end{pmatrix}$$
$$\mathrm{Im}\left[A_i(\omega) \right] = \begin{pmatrix} \dfrac{1}{\omega} \end{pmatrix} \begin{pmatrix} -2\xi_i \sqrt{K_i M_i} \end{pmatrix} \tag{5.68}$$

For now, if we have the measured $A_n(\omega)$ which is complex, we can always separate its real and imaginary part to do the analysis independently. At each frequency value there will be a corresponding measured $A_n(\omega)$ complex value. If we evaluate Eq. (5.68) at multiple frequency locations, we can obtain the matrix form of Eq. (5.68) as follows

$$\mathrm{Re}\left[\mathbf{A}_i(\omega) \right] = \mathbf{D} \begin{pmatrix} M_i \\ -K_i \end{pmatrix} = \mathbf{D}\mathbf{X}$$
$$\mathrm{Im}\left[\mathbf{A}_i(\omega) \right] = \mathbf{v} \begin{pmatrix} -2\xi_i \sqrt{K_i M_i} \end{pmatrix} \tag{5.69}$$

where matrix $\mathbf{D}$ is a tall triangular matrix with full column rank, so does the vector $\mathbf{v}$. $\mathbf{A}_i(\omega)$ is a column vector of the Apparent mass.

Therefore, to calculate the modal properties and the damping ratios, the pseudo-inverse is needed for the real value equation, but for the imaginary value equation, it can simply be done by vector projection, namely,

$$\begin{pmatrix} X_{11} \\ X_{21} \end{pmatrix} = \mathbf{X} = \mathbf{D}^\dagger \mathrm{Re}\left[\mathbf{A}_i(\omega) \right] = \left(\mathbf{D}^\mathrm{T}\mathbf{D} \right)^{-1} \mathbf{D}^\mathrm{T} \mathrm{Re}\left[\mathbf{A}_i(\omega) \right]$$
$$-2\xi_i \sqrt{K_i M_i} = \frac{\mathbf{v}^\mathrm{T} \mathrm{Im}\left[\mathbf{A}_i(\omega) \right]}{\mathbf{v}^\mathrm{T}\mathbf{v}} = z \tag{5.70}$$

173

Since

$$\begin{pmatrix} M_i \\ -K_i \end{pmatrix} = \begin{pmatrix} X_{11} \\ X_{21} \end{pmatrix} = \mathbf{X} \tag{5.71}$$

The concerned modal parameters will be given by

$$
\begin{aligned}
K_i &= -X_{21} \\
M_i &= X_{11} \\
\omega_i &= \sqrt{\frac{K_i}{M_i}} = \sqrt{\frac{-X_{21}}{X_{11}}} \\
\xi_i &= \frac{-z}{2\sqrt{K_i M_i}} = \frac{-z}{2\sqrt{-X_{21} X_{11}}}
\end{aligned}
\tag{5.72}
$$

In practice, if we have multiple measurements, we can introduce a weighting matrix to each line of Eq. (5.70), which will give the weighted least square solution for the modal parameters

$$
\begin{pmatrix} X_{11} \\ X_{21} \end{pmatrix} = \mathbf{X} = \mathbf{D}^\dagger \operatorname{Re}\left[A_n(\omega)\right] = \left(\mathbf{D}^\mathrm{T} \mathbf{W}_R^\mathrm{T} \mathbf{W}_R \mathbf{D}\right)^{-1} \mathbf{D}^\mathrm{T} \mathbf{W}_R^\mathrm{T} \mathbf{W}_R \operatorname{Re}\left[A_n(\omega)\right]
$$
$$
-2\xi_n \sqrt{K_n M_n} = \frac{\mathbf{v}^\mathrm{T} \mathbf{W}_I^\mathrm{T} \mathbf{W}_I \operatorname{Im}\left[A_n(\omega)\right]}{\mathbf{v}^\mathrm{T} \mathbf{W}_I^\mathrm{T} \mathbf{W}_I \mathbf{v}} = z
\tag{5.73}
$$

The entries in the weight matrices $\mathbf{W}_R$ and $\mathbf{W}_I$ are defined in the following way that if the errors in the measured Apparent mass at different frequencies are independent of each other, and their variances are σ_{Rj}^2 for the real part and σ_{Ij}^2 for the imaginary part at the j th entry, then the right weights are $w_{Rj} = 1/\sigma_{Rj}$ and $w_{Ij} = 1/\sigma_{Ij}$, where the subscriptions R and I denote real and imaginary, respectively. Thus, a more accurate measurement, which means a smaller variance, gets a heavier weight. With the independence assumption, both $\mathbf{W}_R$ and $\mathbf{W}_I$ are diagonal matrices with diagonal terms equal w_{Rj} and w_{Ij}, respectively, and all the off-diagonal entries equal to zero. The case where the observations at different frequencies are not independent is not discussed here.

5.4 Conclusion

Starting with a discussion of the decomposition of the bridge system, this chapter presents a two-stage input-output method to solve the vehicle induced bridge modal parameters identification problem. We first discussed some properties of the mode shapes of a system defined by the second order differential equation and clarified some misunderstandings about the mode shapes in the structural dynamics and general modal testing communities.

Having done this, we answered the question that if we should use the estimated complex mode shape or its real counterpart to decompose the system. Since we only need a set of independent basis functions, we decided to use the real shapes, which can be easily estimated by using use the FDD technique in terms of the real-valued one-sided SD matrix proposed in chapter 2.

Later, the second stage of the proposed method is presented with an optimisation procedure based on the Accelerance and a regression approach based on the Apparent mass. Despite the Accelerance and the Apparent mass are formulated in two different ways, and the optimisation method needs an initial guess while the regression method does not need it, these two have a reciprocal relationship, and they are equivalent in estimating the modal parameters. However, regardless of the theoretical and the numerical differences between the two methods, the Apparent mass, when evaluates the real and imaginary part of it separately, is less meaningful than the Accelerence. Therefore, the verification in the next chapter will adopt the optimisation procedure based on the Accelerance.

6. Verification of the proposed input-output system identification method

6.1 Introduction

The VBI problem is a coupled linear dynamic problem, which is intrinsically convoluted. In order to identify the vehicle induced bridge parameters identification problem, we proposed a two-stage input-output system identification methodology in chapter 5. Now, in this chapter, this two-stage identification method is verified by using two types of moving force, i.e., a moving point load and a quarter car, traversing a simply supported beam, respectively. These two types of moving force models correspond to two VBI systems, which have some different properties. For instance, the road roughness effect cannot be incorporated in the point load system. To make our discussion more generic, both systems will be discussed on the nondimensional basis.

The discussion of this chapter is presented in the following fashion. We first give a brief introduction about the VBI systems, then focus on deriving the equation of the motions for the two proposed two VBI systems. Having done this, we model these two VBI systems in MATLAB, then use these numerical models to verify our proposed two-stage method. At the end of this chapter, a discussion about the identification impact factors is presented.

6.2 Vehicle-bridge interaction system

6.2.1 Car models and bridge models

There are three car models that have been intensively investigated in literature, i.e., moving load, moving mass, and moving sprung mass models [31]. Among the three vehicle models, the moving load model is the simplest one, which has been adopted by researchers in studying traffic-induced bridge vibration [14], [49]. However, by using this model, the interaction between the two subsystems is intractable. Therefore, this model is only good for computing the response of the bridge system. When the inertia of the vehicle cannot be neglected, the moving load model is replaced by a moving mass model . However, the bouncing action of the moving mass relative to the bridge is not captured by this model, which is the well-known drawback of it. Since the impact of bouncing action of the vehicle is expected to be significant, especially in the presence of pavement roughness or for vehicles moving at

high speeds [31], the sprung mass model is proposed to incorporate elastic and damping effects of the suspension systems. For instance, a quarter car model, see Figure 6.3.

Note that, although the use of a more sophisticated vehicle model can make the simulation more realistic, it may result in some computation problems, such as divergence or slow convergence problem when searching for a large number of contact forces in the process of iteration. Whereas by using the simplified models, the essential dynamic characteristics of the bridge can be captured with a sufficient degree of accuracy [31]. As a preliminary study of the feasibility of the proposed input-output system identification method, and our focus is on the system identification of the bridge modal parameters. Therefore, we consider the moving load and the quarter car model in this thesis.

In general VBI analysis, especially for the moving load identification, we usually need to establish the relationship between the moving force and the bridge response, which means that a known and well-defined bridge model is required [25]. Typically, there are two kinds of approaches widely adopted in the VBI analysis, i.e., analytical model based approach and finite element model based approach [23]. Although there is no restriction on the type of structures considered for the VBI problems, a simpler bridge model requires less preparation and computation efforts [31]. Apart from this, compared to the analytical approaches, the finite element approach is more flexible for general application, and it is more convenient for modelling some complex bridge structures. To simplify the problem, the simply supported Euler-Bernoulli beam is considered in this study. In this case, we can formulate the equation of motion for the VBI system analytically, and the 4th order Runge-Kutta method [202] is used to simulate the VBI system.

6.2.2 Road roughness

In VBI analysis, the role of road roughness cannot be ignored. Depended primarily on the workmanship involved in the construction of pavement or rail tracks and on how they are maintained [31], as it has been reported that the impact response of bridges can be significantly affected by the road unevenness [203]. Such impact, from the viewpoint of the structural dynamics, is caused by the frequencies implied by the surface roughness or rail irregularities [31]. And such spatial frequencies will be introduced to the bridge responses via the moving vehicle, see for instance Eq. (6.7), where the road roughness is incorporated in the vehicle's equation of motion. When a vehicle is moving across a bridge, the bridge will be excited by the contact force, see Eq. (6.5) and Eq. (6.6). Due to the sprung-mass configuration of the vehicle, the road roughness is affecting both the bridge and the vehicle systems. As a result, resonance may occur on the bridge and traversing vehicles if the vehicle frequencies equal to or close to any of the frequencies implied by the road unevenness [31].

To account for its random nature, the road profile can be modelled with a periodic modulated random process [204]. According to the ISO-8608 [205] specification, the road surface roughness can be simulated based on the displacement PSD given by the following formula

$$G_d(n) = G_d(n_0)\left(\frac{n}{n_0}\right)^{-w} \tag{6.1}$$

Where $n_0 = 0.1$ cycle/m is the reference spatial frequency, and w is the exponent of the fitted PSD, usually assumed it takes the value 2. As for n, which denotes the spatial frequency per meter, is usually determined by the vehicle speed and time frequency of the vehicle. The general expression for this relationship is given by ISO-8608 [205],

$$n = \frac{f_v}{\upsilon} \tag{6.2}$$

where f_v (Hz) is the time frequency of the vehicle, υ denotes the vehicle speed. In this study, the range of spatial frequency is taken as 0.01 – 100 cycle/m.

Eq. (6.1) gives an estimate on the degree of roughness of a road from the value of $G_d(n_0)$. In ISO-8608, the road surface is divided into eight classes, see Table 6.1, which gives the guidance of the functional value $G_d(n_0)$ in terms of each roughness class.

Table 6.1 Road classification

| Road class	Degree of roughness $G_d(n_0) \times 10^{-6} \, m^3$ ($n_0 = 0.1$ cycle/m)			
	Lower limit	Geometric mean	Upper limit	Adopted value
A	—	16	32	0.001
B	32	64	128	8
C	128	256	512	16
D	512	1024	2048	32
E	2048	4094	8192	64
F	8192	16384	32768	128
G	32768	65536	131072	256
H	131072	262144	—	215

Now by applying the inverse Fourier transform on $G_d(n)$, we can calculate the road surface roughness function $r(x)$, which can be simply given by

$$r(x) = \sum_{i=1}^{N_r} \sqrt{4G_d(n_i)\Delta n} \cos(2\pi n_i x + \theta_i) \tag{6.3}$$

where N_r is the number of data points, and Δn is the sampling interval of the spatial frequency, and it is taken as 0.04 cycle/m in this study [206]. Note that, if we the geometric mean value of $G_d(n)$ in the above equation, the amplitude of roughness, i.e., $\sqrt{4G_d(n_i)\Delta n}$, will be too large to be compatible with the road roughness from the field [206]. Therefore, we take the square root of the geometric mean of the functional value $G_d(n)$ in Table 6.1. for the simulation in this study. As for θ_i, it represents a set of independent random phase angles uniformly distributed between 0 and 2π. As an example, three different road roughness levels are depicted in Figure 6.1 (a) – (c).
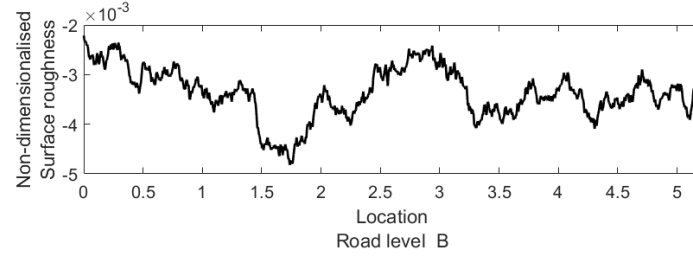
(a) Nondimensionalised road roughness level B
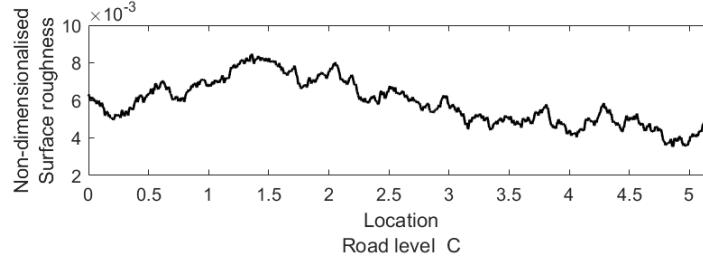
(b) Nondimensionalised road roughness level C
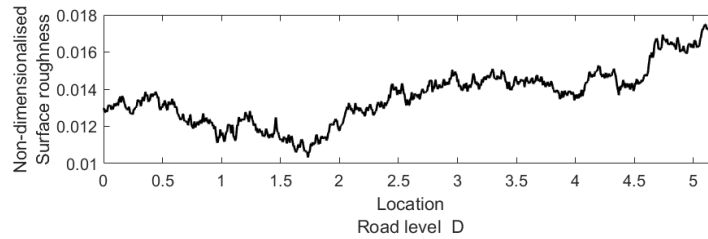

(c) Nondimensionalised road roughness level D

Figure 6.1 Nondimensionalised road roughness

6.3 VBI systems equation of motion

In this section, our main task is to derive the equation of motion for the two VBI systems (see Figure 6.2) to verify the proposed two-stage identification method. Figure 6.2 (a) shows a point load moving across a simply supported beam, while Figure 6.2 (b) depicts a quarter car traversing a simply supported beam. Where $p(x,t)$ is the contact force between the two subsystems (vehicle and the bridge); L is the length of the beam; x denotes the location on the bridge in terms of the left entrance; υ represents the speed of the vehicle. To generalise the problem, both of the considered systems are nondimensionalised accordingly.
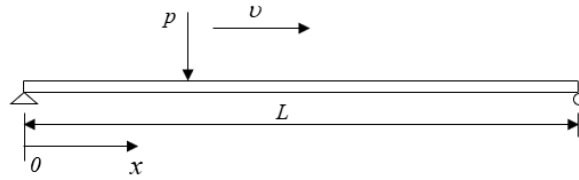
(a) Simply supported Euler-Bernoulli beam model with moving point load
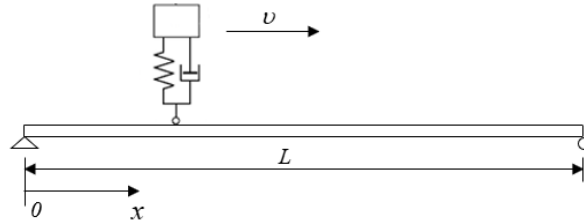
(b) Simply supported Euler-Bernoulli beam model with quarter car model

Figure 6.2. Moving forces (a moving load and a quarter car) traversing a simply supported beam

6.3.1 Subsystems

6.3.1.1 System Equations—Bridge subsystem

From last chapter we know that the vibration equation of the motion for a decoupled system subjected to a moving force is given by Eq. (5.48)

$$\ddot{q}_i(t) + 2\xi_i\omega_i\dot{q}_i(t) + \omega_i^2 q_i(t) = \frac{1}{M_i}\int_0^L \phi_i(x)\,p(x,t)\,\mathrm{d}x \tag{6.4}$$

where $q_i(t)$ is the ith modal coordinates and $\phi_i(x)$ is the ith continuous mode shape. ω_i and ξ_i represent the ith natural frequency and damping ratio, respectively. $p(x,t)$ is the distributed vertical loading.

In the case of a moving concentrated load with a given time varying intensity $f(t)$ and speed υ, the load $p(x,t)$ can be expressed as Eq. (5.49) [196]

$$p(x,t) = \begin{cases} \delta(x - \upsilon t)f(t) & 0 \le t \le t_d \\ 0 & t > t_d \end{cases}$$

(6.5)

where $t_d = L/\upsilon$ is the time required for the load to cross the bridge of length L.

It is worth pointing out that although the energy loss of a structure is more accurately modelled by the hysteretic damping [141], viscous damping is commonly used for the bridge structure. Not only because the viscous damping model can linearise the equation of motion and facilitate it is solving [207], but also because it is deemed to be sufficient to reproduce the bridge responses accurately [208]. Therefore, we followed the convention by adopting the viscous damping in Eq. (6.4).

6.3.1.2 System Equations — Vehicle subsystem

When the vehicle is modelled as a point load, $p(x,t)$ is given by Eq. (6.5). To be more specific, we can define $f(t)$ as

$$f(t) = -M_v\left(g + \ddot{z}_1(t)\right)$$

(6.6)

where M_v is the mass of the vehicle, $\ddot{z}_1(t)$ represents the vehicle body acceleration, measured according to the equilibrium position, while $g = 9.8\,m/s^2$ is the acceleration due to gravity.

Note that Eq. (6.6) is defined for the point load model; it also applies to the quarter car model (see Figure 6.3). It is because the active force between the vehicle and bridge is the contact force; no matter what kind of car models we use for each vehicle axis, such a contact force can also be modelled as a point load. However, the use of the quarter car model is more advanced than the moving load case we just discussed since it incorporates the mass-sprung effect of the vehicle into the analysis.
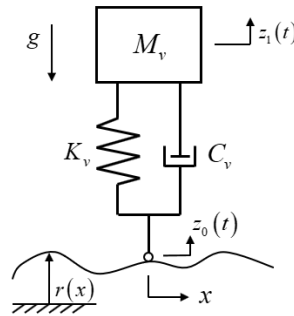

Figure 6.3 A typical quarter car model

For a quarter car model, given by Figure 6.3, the vehicle subsystem vibration can be depicted by the following equation of motion,

$$
\begin{cases}
M_v \ddot{z}_1(t) + C_v\left(\dot{z}_1(t) - \dot{z}_0(t)\right) + K_v\left(z_1(t) - z_0(t)\right) = 0 \\
z_0(t) = \displaystyle\sum_{i=1}^{N} \phi_i(\upsilon t) q_i(t) + r(\upsilon t)
\end{cases}
\tag{6.7}
$$

where $z_1(t)$ is the vehicle body displacement measured at the equilibrium position; $z_0(t)$ is the contact point displacement, while $r(\upsilon t)$ is the road roughness, which is given by Eq. (6.3). N determines the number of modes for the bridge system. As for C_v and K_v, they represent the vehicle damping coefficient and spring constant, respectively.

Note that the quarter car model we adopted in this study, as given by Figure 6.3, has included the mass of the wheel to the vehicle body (suspended) mass M_v. Likewise, the sprung and damping effect of the wheel are also included in K_v and C_v, respectively.

If we denote the vehicle frequency and damping ratio as ω_v and ξ_v, respectively, then

$$
C_v = 2\xi_v \omega_v M_v \ \text{and} \ K_v = M_v \omega_v^2
\tag{6.8}
$$

6.3.2 Point load system

6.3.2.1 Nondimensionalised equation of motion for the point load system

For point moving load, the modal force P_i for the ith mode is

$$
P_i = \int_0^L \phi_i(x) p(x,t)\,\mathrm{d}x = \phi_i(\upsilon t) f(t)
\tag{6.9}
$$

When the load is moving on the bridge, by substituting the above expression into Eq. (6.4), we can obtain the VBI system for this moving load case

$$
\ddot{q}_i(t) + 2\xi_i \omega_i \dot{q}_i(t) + \omega_i^2 q_i(t) = \frac{1}{M_i}\phi_i(\upsilon t) f(t)
\tag{6.10}
$$

Again, substitute Eq. (6.6) into the above equation, we will reach

$$
\ddot{q}_i(t) + 2\xi_i \omega_i \dot{q}_i(t) + \omega_i^2 q_i(t) = -\frac{M_v}{M_i}\phi_i(\upsilon t)\left(g + \ddot{z}_1(t)\right)
\tag{6.11}
$$

To nondimensionalise this system equation, we observe that we have the following eight parameters, i.e., M_v, M_i, ω_i, ω_1, ξ_i, L, υ, g, and three dimensions, i.e., mass, length, and time, where ω_1 is the fundamental frequency of the bridge, which is used to nondimensionalise the bridge system. Therefore, we will obtain five nondimensionalised groups, exclude the damping ratio, which is already nondimensional, the other four of them can be defined in the following fashion,

$$\mu_i = \frac{M_v}{M_i}, \quad \kappa_i = \frac{\omega_i}{\omega_1}, \quad \beta = \frac{g}{\omega_1^2 L}, \quad \alpha = \frac{\upsilon}{\omega_1 L} \tag{6.12}$$

There are the four nondimensional parameters of system given by Eq. (6.11). Further, we have

$$\hat{q} = \frac{q}{L}, \quad \hat{z} = \frac{z}{L}, \quad \hat{x} = \frac{x}{L} = \frac{\upsilon t}{L} =, \quad \tau = \omega_1 t \tag{6.13}$$

With $\alpha = \upsilon / \omega_1 L$ and $\tau = \omega_1 t$, we can define

$$\hat{x} = \frac{x}{L} = \frac{\upsilon t}{L} = \frac{\upsilon}{\omega_1 L}\tau = \alpha\tau \tag{6.14}$$

We will then have the following relations

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} &= \omega_1 \frac{\mathrm{d}}{\mathrm{d}\tau} \\
\frac{\dot{\hat{q}}_i}{\omega_1} &= \frac{\mathrm{d}\hat{q}_i}{\mathrm{d}t} \cdot \frac{1}{\omega_1} = \frac{\mathrm{d}\hat{q}_i}{\mathrm{d}\tau}\omega_1 \cdot \frac{1}{\omega_1} = \frac{\mathrm{d}\hat{q}_i}{\mathrm{d}\tau} = \hat{q}_i' \\
\frac{\ddot{\hat{q}}_i}{\omega_1^2} &= \frac{\mathrm{d}^2\hat{q}_i}{\mathrm{d}t^2} \cdot \frac{1}{\omega_1^2} = \frac{\mathrm{d}^2\hat{q}_i}{\mathrm{d}\tau^2}\omega_1^2 \cdot \frac{1}{\omega_1^2} = \frac{\mathrm{d}^2\hat{q}_i}{\mathrm{d}\tau^2} = \hat{q}_i'' \\
\frac{\ddot{\hat{z}}_1}{\omega_1^2} &= \frac{\mathrm{d}^2\hat{z}_1}{\mathrm{d}t^2} \cdot \frac{1}{\omega_1^2} = \frac{\mathrm{d}^2\hat{z}_1}{\mathrm{d}\tau^2}\omega_1^2 \cdot \frac{1}{\omega_1^2} = \frac{\mathrm{d}^2\hat{z}_1}{\mathrm{d}\tau^2} = \hat{z}''
\end{aligned} \tag{6.15}
$$

On the other hand, for simply supported beam, we have

$$\omega_i = i^2\pi^2\sqrt{\frac{EI}{\bar{m}L^4}} = i^2\omega_1 \quad \phi_i(\upsilon t) = \sin\left(\frac{i\pi\upsilon}{L}t\right) \tag{6.16}$$

where E is the Young's modulus, I is the moment of inertia of the beam cross section, EI is the bending stiffness, $\bar{m}$ is the mass per unit length. For simplicity, we assume $\bar{m}$ and EI are constant.

By nondimensionalise Eq. (6.16)

$$\kappa_i = i^2 \quad \phi_i(\upsilon t) = \sin\left(\frac{i\pi\upsilon}{\omega_1 L}\tau\right) = \sin(i\pi\alpha\tau) \tag{6.17}$$

Substitute Eq. (6.12) to Eq. (6.17) into Eq. (6.11), we shall obtain the nondimensionalised form of the equation of motion for the VBI system.

$$\hat{q}_n''(\tau) + 2\xi_i \kappa_i \hat{q}_i'(\tau) + \kappa_i^2 \hat{q}_i(\tau) = \mu_i \sin(i\pi\alpha\tau)\left(-\beta - \hat{z}_1''(\tau)\right) \tag{6.18}$$

6.3.2.2 Equations for the simulation

To use MATLAB to generate the response, we shall write Eq. (6.18) for multiple modes into matrix form,

$$[M]\{\hat{q}''\} + [C]\{\hat{q}'\} + [K]\{\hat{q}\} = \{F\} \tag{6.19}$$

where

$$[M] = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \quad [C] = \begin{bmatrix} 2\xi_1\kappa_1 & 0 & \cdots & 0 \\ 0 & 2\xi_2\kappa_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 2\xi_N\kappa_N \end{bmatrix} \quad [K] = \begin{bmatrix} \kappa_1^2 & 0 & \cdots & 0 \\ 0 & \kappa_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \kappa_N^2 \end{bmatrix}$$

$$\{\hat{q}''\} = \{\hat{q}_1'', \quad \hat{q}_2'', \quad \cdots \quad \hat{q}_N''\}^T$$

$$\{\hat{q}'\} = \{\hat{q}_1', \quad \hat{q}_2', \quad \cdots \quad \hat{q}_N'\}^T$$

$$\{\hat{q}\} = \{\hat{q}_1, \quad \hat{q}_2, \quad \cdots \quad \hat{q}_N\}^T$$

$$\{F\} = -(\beta + \hat{z}'')\{\mu_1\phi_1, \quad \mu_2\phi_2, \quad \cdots \quad \mu_N\phi_N\}^T$$

$$\tag{6.20}$$

6.3.3 Quarter car system

Having discussed the moving load system, in this section, we move on to investigate the system equations with a moving quarter car model, see Figure 6.2 (b).

6.3.3.1 Nondimensionalised equation of motion for the quarter car system

Since the contact force between the quarter car and the bridge has the same format of the interaction force between the moving load and the bridge. Thus, from Eq. (6.7) and Eq. (6.11) the interaction VBI system equation of motion for a quarter car model is given by

$$
\begin{cases}
\ddot{q}_i(t) + 2\xi_i\omega_i\dot{q}_i(t) + \omega_i^2 q_i(t) = -\dfrac{M_v}{M_i}\phi_i(\upsilon t)\big(g + \ddot{z}_1(t)\big) \\[2mm]
M_v\ddot{z}_1(t) + 2\xi_v\omega_v M_v\big(\dot{z}_1(t) - \dot{z}_0(t)\big) + M_v\omega_v^2\big(z_1(t) - z_0(t)\big) = 0 \\[2mm]
z_0(t) = \displaystyle\sum_{i=1}^{N}\phi_i(\upsilon t)q_i(t) + r(\upsilon t)
\end{cases}
\tag{6.21}
$$

Please be aware of the notation n (spatial frequency) in Eq. (6.3) which has a different meaning from the n (number of modes) in Eq. (6.21) and the rest.

To nondimensionalise the system equation given by Eq. (6.21), we observe that we have the following 10 parameters, i.e., M_v, ξ_v, ω_v, M_i, ω_i, , ξ_i, L, υ, g, and three dimensions, i.e., mass, meter, and time. Therefore, we will obtain seven nondimensionalised groups, exclude the two damping ratios which are already nondimensional, the rest five of them will be given by

$$
\mu_i = \frac{M_v}{M_i}, \quad \kappa_i = \frac{\omega_i}{\omega_v}, \quad \beta = \frac{g}{\omega_v^2 L}, \quad \alpha = \frac{\upsilon}{\omega_v L}, \quad \widehat{r} = \frac{r}{L}
\tag{6.22}
$$

Further we have

$$
\widehat{q} = \frac{q}{L}, \quad \widehat{z} = \frac{z}{L}, \quad \tau = \omega_v t, \quad \widehat{x} = \frac{x}{L} = \alpha\tau
\tag{6.23}
$$

We will then have the following relations

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t} = \omega_v\frac{\mathrm{d}}{\mathrm{d}\tau} \\[2mm]
&\frac{\dot{q}_i}{\omega_v} = \frac{\mathrm{d}\widehat{q}_i}{\mathrm{d}t}\cdot\frac{1}{\omega_v} = \frac{\mathrm{d}\widehat{q}_i}{\mathrm{d}\tau}\omega_v\cdot\frac{1}{\omega_v} = \frac{\mathrm{d}\widehat{q}_i}{\mathrm{d}\tau} = \widehat{q}_i' \\[2mm]
&\frac{\ddot{q}_i}{\omega_v^2} = \frac{\mathrm{d}^2\widehat{q}_i}{\mathrm{d}t^2}\cdot\frac{1}{\omega_v^2} = \frac{\mathrm{d}^2\widehat{q}_i}{\mathrm{d}\tau^2}\omega_v^2\cdot\frac{1}{\omega_v^2} = \frac{\mathrm{d}^2\widehat{q}_i}{\mathrm{d}\tau^2} = \widehat{q}_i'' \\[2mm]
&\frac{\dot{z}_1}{\omega_v} = \frac{\mathrm{d}\widehat{z}_1}{\mathrm{d}t}\cdot\frac{1}{\omega_v} = \frac{\mathrm{d}\widehat{z}_1}{\mathrm{d}\tau}\omega_v\cdot\frac{1}{\omega_v} = \frac{\mathrm{d}\widehat{z}_1}{\mathrm{d}\tau} = \widehat{z}_1' \\[2mm]
&\frac{\ddot{z}_1}{\omega_v^2} = \frac{\mathrm{d}^2\widehat{z}_1}{\mathrm{d}t^2}\cdot\frac{1}{\omega_v^2} = \frac{\mathrm{d}^2\widehat{z}_1}{\mathrm{d}\tau^2}\omega_v^2\cdot\frac{1}{\omega_v^2} = \frac{\mathrm{d}^2\widehat{z}_1}{\mathrm{d}\tau^2} = \widehat{z}_1'' \\[2mm]
&\frac{\dot{\phi}_i(\upsilon t)}{\omega_v L} = \frac{\mathrm{d}\phi_i(\upsilon t/L)}{\mathrm{d}t}\cdot\frac{1}{\omega_v} = \frac{\mathrm{d}\phi_i(\alpha\tau)}{\mathrm{d}\tau}\omega_v\cdot\frac{1}{\omega_v} = \phi_i'(\alpha\tau) \\[2mm]
&\frac{\dot{r}(\upsilon t)}{\omega_v L} = \frac{\mathrm{d}\widehat{r}(\upsilon t/L)}{\mathrm{d}t}\cdot\frac{1}{\omega_v} = \frac{\mathrm{d}\widehat{r}(\alpha\tau)}{\mathrm{d}\tau}\omega_v\cdot\frac{1}{\omega_v} = \widehat{r}'(\alpha\tau)
\end{aligned}
\tag{6.24}
$$

Substitute Eq. (6.22) to Eq. (6.24) into Eq. (6.21), we shall obtain the nondimensionalised form of the equation of motion for the VBI system.

For the bridge subsystem, the equation of motion is given by Eq. (6.18),

$$\hat{q}_i''(\tau) + 2\xi_i \kappa_i \hat{q}_i'(\tau) + \kappa_i^2 \hat{q}_i(\tau) = -\mu_i \phi_i(\alpha\tau)\left(\beta + \hat{z}_1''(\tau)\right) \tag{6.25}$$

For the vehicle subsystem

$$\hat{z}_1''(\tau) + 2\xi_v \hat{z}_1'(\tau) + \hat{z}_1(\tau) = 2\xi_v \hat{z}_0'(\tau) + \hat{z}_0(\tau) \tag{6.26}$$

Since

$$\hat{z}_0(\tau) = \sum_{i=1}^{N} \phi_i(\alpha\tau)\hat{q}_i(\tau) + \hat{r}(\alpha\tau) \tag{6.27}$$

then

$$\hat{z}_0'(\tau) = \sum_{i=1}^{N} \phi_i'(\alpha\tau)\hat{q}_i(\tau) + \sum_{i=1}^{N} \phi_i(\alpha\tau)\hat{q}_i'(\tau) + \hat{r}'(\alpha\tau) \tag{6.28}$$

Now substitute Eq. (6.27) and Eq. (6.28) into the last equation of Eq. (6.26) and rearrange the equation, together with Eq. (6.25) we have

$$\begin{cases} \hat{q}_i''(\tau) + \mu_i \phi_i(\alpha\tau)\hat{z}_1''(\tau) + 2\xi_i \kappa_i \hat{q}_i'(\tau) + \kappa_i^2 \hat{q}_i(\tau) = -\beta\mu_i \phi_i(\alpha\tau) \\ \hat{z}_1''(\tau) - 2\xi_v \sum_{i=1}^{N} \phi_i(\alpha\tau)\hat{q}_i'(\tau) + 2\xi_v \hat{z}_1'(\tau) - \sum_{i=1}^{N}\left[2\xi_v \phi_i'(\alpha\tau) + \phi_i(\alpha\tau)\right]\hat{q}_i(\tau) + \hat{z}_1(\tau) \\ = 2\xi_v \hat{r}'(\alpha\tau) + \hat{r}(\alpha\tau) \end{cases} \tag{6.29}$$

For simply supported beam, from Eq. (6.16)

$$\kappa_i = \frac{\omega_1}{\omega_v} i^2, \quad \phi_i(\upsilon t) = \sin(i\pi\alpha\tau), \quad \phi_i'(\alpha\tau) = i\pi\alpha\cos(i\pi\alpha\tau) \tag{6.30}$$

Substitute the relations into Eq. (6.29),

$$\begin{cases} \hat{q}_i''(\tau) + \mu_i \sin(i\pi\alpha\tau)\hat{z}_1''(\tau) + 2\xi_i \kappa_i \hat{q}_i'(\tau) + \kappa_i^2 \hat{q}_i(\tau) = -\beta\mu_i \sin(i\pi\alpha\tau) \\ \hat{z}_1''(\tau) - 2\xi_v \sum_{i=1}^{N} \sin(i\pi\alpha\tau)\hat{q}_i'(\tau) + 2\xi_v \hat{z}_1'(\tau) - \\ \sum_{i=1}^{N}\left[2i\pi\alpha\xi_v \cos(\pi\alpha\tau) + \sin(i\pi\alpha\tau)\right]\hat{q}_i(\tau) + \hat{z}_1(\tau) = 2\xi_v \hat{r}'(\alpha\tau) + \hat{r}(\alpha\tau) \end{cases} \tag{6.31}$$

Define

$$b_i = \mu_i \sin(i\pi\alpha\tau), \quad d_i = 2\xi_v \sin(i\pi\alpha\tau), \quad g_i = 2i\pi\alpha\xi_v \cos(i\pi\alpha\tau) + \sin(i\pi\alpha\tau) \tag{6.32}$$

Then Eq. (6.31) becomes

$$\begin{cases} \widehat{q}_i''(\tau)+b_i\widehat{z}_1''(\tau)+2\xi_i\kappa_i\widehat{q}_i'(\tau)+\kappa_i^2\widehat{q}_i(\tau)=-\beta b_i \\ \widehat{z}_1''(\tau)-\sum_{i=1}^{N}d_i\widehat{q}_i'(\tau)+2\xi_v\widehat{z}_1'(\tau)-\sum_{i=1}^{N}g_i\widehat{q}_i(\tau)+\widehat{z}_1(\tau)=2\xi_v\widehat{r}'(\alpha\tau)+\widehat{r}(\alpha\tau) \end{cases} \tag{6.33}$$

6.3.3.2 Equations for the simulation

Again, to use MATLAB to generate the response, we write Eq. (6.33) into matrix form,

$$[M]\{\widehat{q}''\}+[C]\{\widehat{q}'\}+[K]\{\widehat{q}\}=\{F\} \tag{6.34}$$

where

$$[M]=\begin{bmatrix} 1 & 0 & \cdots & 0 & b_1 \\ 0 & 1 & \cdots & 0 & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & b_n \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}, \quad [C]=\begin{bmatrix} 2\xi_1\kappa_1 & 0 & \cdots & 0 & b_1 \\ 0 & 2\xi_2\kappa_2 & \cdots & 0 & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 2\xi_N\kappa_N & b_N \\ d_1 & d_2 & \cdots & d_N & 2\xi_v \end{bmatrix}$$

$$[K]=\begin{bmatrix} \kappa_1^2 & 0 & \cdots & 0 & 0 \\ 0 & \kappa_2^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \kappa_N^2 & 0 \\ g_1 & g_2 & \cdots & g_N & 1 \end{bmatrix}$$

$$\{\widehat{q}''\}=\{\widehat{q}_1'', \quad \widehat{q}_2'', \quad \cdots \quad \widehat{q}_N'', \quad \widehat{z}_1''\}^{\mathrm{T}}$$

$$\{\widehat{q}'\}=\{\widehat{q}_1', \quad \widehat{q}_2', \quad \cdots \quad \widehat{q}_N', \quad \widehat{z}_1'\}^{\mathrm{T}}$$

$$\{\widehat{q}\}=\{\widehat{q}_1, \quad \widehat{q}_2, \quad \cdots \quad \widehat{q}_N, \quad \widehat{z}_1(\tau)\}^{\mathrm{T}}$$

$$\{F\}=\{-\beta b_1, \quad -\beta b_2, \quad \cdots \quad -\beta b_N, \quad 2\xi_v\widehat{r}'(\alpha\tau)+\widehat{r}(\alpha\tau)\}^{\mathrm{T}} \tag{6.35}$$

6.3.4 FRF of the nondimensionalised bridge subsystem

For the considered two VBI systems, the equation of motion for the bridge subsystem is the same.

Thus, regardless which car model we use, we can take the Fourier transform of Eq. (6.25),

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\widehat{q}_i''(\tau)e^{-i\kappa\tau}\mathrm{d}\tau+2\xi_i\kappa_i\frac{1}{2\pi}\int_{-\infty}^{\infty}\widehat{q}_i'(\tau)e^{-i\kappa\tau}\mathrm{d}\tau+\kappa_i^2\frac{1}{2\pi}\int_{-\infty}^{\infty}\widehat{q}_i(\tau)e^{-i\kappa\tau}\mathrm{d}\tau$$
$$=\mu_i\frac{1}{2\pi}\int_{-\infty}^{\infty}\phi_i(\alpha\tau)(-\beta-\widehat{z}_1''(\tau))e^{-i\kappa\tau}\mathrm{d}\tau \tag{6.36}$$

This equation can be further written as

$$\tilde{\ddot{q}}_i''(\kappa) + 2\xi_i\kappa_i\tilde{\ddot{q}}_i'(\kappa) + \kappa_i^2\tilde{\ddot{q}}_i(\kappa) = \mu_i\tilde{\ddot{P}}_i(\kappa) \tag{6.37}$$

where

$$\tilde{\ddot{P}}_i(\kappa) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\phi_i(\alpha\tau)\left(-\beta - \ddot{z}_1''(\tau)\right)e^{-i\kappa\tau}\mathrm{d}\tau \tag{6.38}$$

is the Fourier transform of the nondimensionalised genialised force $\widehat{P}_i$.

Since

$$\tilde{\ddot{q}}_i'(\kappa) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\ddot{q}_i'(\kappa)e^{-i\kappa\tau}\mathrm{d}\tau$$

$$\tilde{\ddot{q}}_i''(\kappa) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\ddot{q}_i''(\kappa)e^{-i\kappa\tau}\mathrm{d}\tau \tag{6.39}$$

Use $\tilde{\ddot{q}}_i''(\kappa)$ to represent $\tilde{\ddot{q}}_i'(\kappa)$ and $\tilde{\ddot{q}}_i(\kappa)$, we have

$$\tilde{\ddot{q}}_i'(\kappa) = -i\frac{1}{\kappa}\tilde{\ddot{q}}_i''(\kappa)$$

$$\tilde{\ddot{q}}_i(\kappa) = -\frac{1}{\kappa^2}\tilde{\ddot{q}}_i''(\kappa) \tag{6.40}$$

Substitute Eq. (6.40) into Eq. (6.37) and rearrange the equation, we can obtain the Accelerence of the moving load system,

$$I_i(\kappa) = \frac{\tilde{\ddot{q}}_i''(\kappa)}{\tilde{\ddot{P}}_i(\kappa)} = \frac{\mu_i\kappa^2}{\kappa^2 - \kappa_i^2 - i2\xi_i\kappa_i\kappa} \tag{6.41}$$

Further we have

$$\mathrm{Re}\left[I_i(\kappa)\right] = \frac{\mu_i\kappa^2\left(\kappa^2 - \kappa_i^2\right)}{\left(\kappa^2 - \kappa_i^2\right)^2 + \left(2\xi_i\kappa_i\kappa\right)^2}$$

$$\mathrm{Im}\left[I_i(\kappa)\right] = \frac{\mu_i\kappa^2\left(2\xi_i\kappa_i\kappa\right)}{\left(\kappa^2 - \kappa_i^2\right)^2 + \left(2\xi_i\kappa_i\kappa\right)^2} \tag{6.42}$$

When $\kappa = 0$,

$$\mathrm{Re}\left[I_i(\kappa_i)\right] = 0$$

$$\mathrm{Im}\left[I_i(\kappa_i)\right] = \frac{\mu_i}{2\xi_i} \tag{6.43}$$

6.4 Input-output system identification and result

Up to now, we have discussed two types of VBI systems, the moving load case without considering the mass-sprung effect of the vehicle, and the quarter car model, which considers the mass-sprung effect of the vehicle. In this section, we will verify the proposed input-output method by using these two systems.

The identification results will be presented in the following manner as shown in the Figure. 6.4. The two systems are discussed in terms of noise-free and noise-corrupted scenarios. Particularly for the quarter car system, the effect of the road roughness is addressed. However, as we will point out later that the spectral density of the input signal of the quarter car is similar to the moving load case, the noise-contaminated case for the quarter car system without road roughness is not covered.
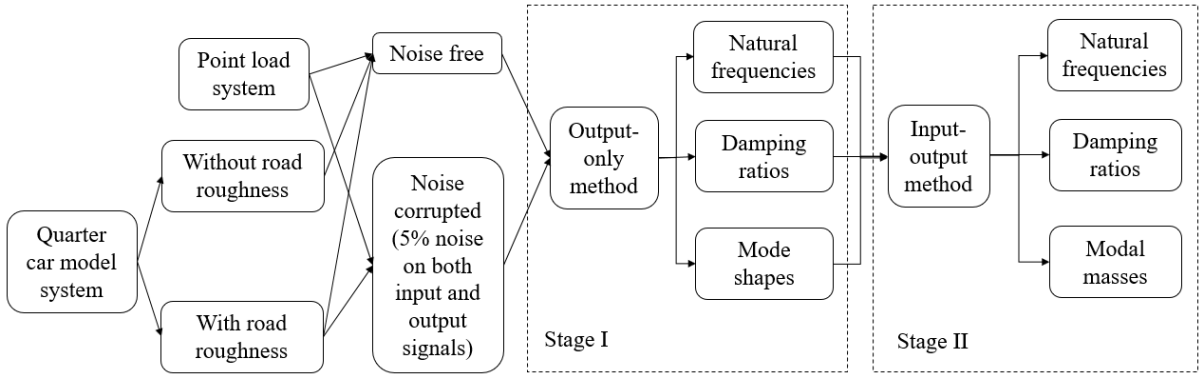

Figure 6.4 Identification process for the bridge subsystem

It is worth mentioning that, as an output-only method, by using the FDD alone, we cannot identify the damping ratios and the modal masses. However, we can use the identified mode shapes to decompose the measured responses, then evaluate the free vibrations of each mode. Thereby we can obtain the estimated damping ratios. As for the modal masses, without the input information, we are unable to extracted from the bridge measurements alone. Therefore, the output-only method we will use in this section is the combination of the FDD analysis plus the free decay analysis, the purpose of this is the same as the so-called Enhanced FDD method (EFDD) [90], which is the extension of the conventional FDD used for the identification of the damping ratios. However, the actual procedure is different, as the SDOF bell function is not evaluated.

For simplicity, the resonance frequencies obtained from FDD in this study are fed into the free decay analysis to extract the corresponding damping ratios. Note that even though we can obtain the natural frequencies from determining the zero-crossing time of the free vibration, the free decay analysis in this study is only used to identify the damping ratios, and this procedure is performed upon the generalised free vibration response of the bridge.

To obtain the damping ratios, the Hilbert transform [124] is used to find the envelope of the vibration response first. Specifically, let $x(t)$ denotes the original real signal, the Hilbert transform is defined to be the convolution between $x(t)$ and the Cauchy kernel $h_C(t) = 1/\pi t$. Since the Cauchy kernel is not defined at $t=0$, the integral defining the convolution may not converge. Therefore, the Hilbert transform is usually defined as the Cauchy principal value (denoted by p.v.) [112] of the convolution integral,

$$y(t) = \mathcal{H}\big[x(t)\big] = x(t) * h_C(t) = \frac{1}{\pi} \text{p.v.} \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau \tag{6.44}$$

Particularly if $x(t)$ is sinusoidal, for instance, $x(t) = \cos(\omega t)$, the Hilbert transform will impart a $-\pi/2$ phase shift to $x(t)$, then $y(t) = \cos(\omega t - \pi/2)$. As a result, if a reasonably behaved function can have a Fourier decomposition, then by taking the Hilbert of the original function, the transformed function will be shifted as a whole.

Thus, together with the original signal $x(t)$, we can obtain a complex helical sequence (analytic signal) $x_a(t) = x(t) + iy(t)$, where the real part is the original data while the imaginary part contains the 90-degree shift version of the original data. The decaying amplitude of the free vibration is then found by evaluating the absolute value of an analytic signal. Having done this, the amplitude data are used to fit an exponential function, namely $Ae^{-\kappa_k \xi_k \tau}$, where A is the initial amplitude, τ is the nondimensionalised time, while κ_k and ξ_k are the k th natural frequency ratio and damping ratio, respectively. Then the damping ratios can be found from the fitted exponential decay $\kappa_k \xi_k$ accordingly.

The process for finding the damping ratios is illustrated in Figure 6.5. The solid line depicts the original signal, the dotted line represents the shifted signal, while the circles are the selected points (the crossing points of the difference between the real and imaginary part of the analytic signal) on the envelope to demonstrate the decaying amplitude data to use to fit the exponential function. The data of this figure is from the noise-free moving load system discussed in the next section.
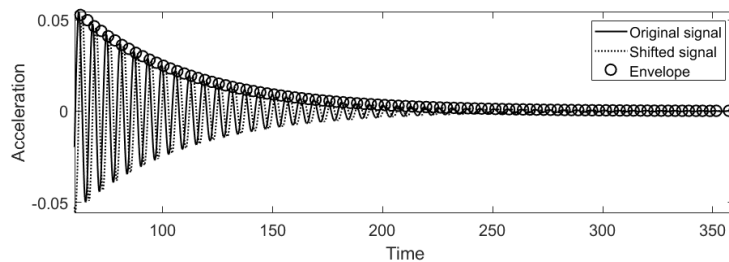

Figure 6.5. Generalised free vibration response of the 1st mode of the bridge

Apart from the mode shapes identified from the first stage, the natural frequencies and damping ratios identified from this stage serve two purposes. One purpose is to be used as a comparison with the parameters identified from the second stage. While the other one is to use them as the initial guess to the optimisation procedure we proposed in chapter 5 for the second-stage identification.

6.4.1 Evaluation of the proposed input-output system method based on the point load system

In this section, the proposed system identification approach will be demonstrated by using a moving load $p(x,t)$, traversing a simply supported Euler-Bernoulli beam, as shown in Figure. 6.2 (a). The beam is of length L and is assumed to be uniform, with bending stiffness EI and mass per unit length $\bar{m}$. The system is simulated based on Eq. (6.19) and Eq. (6.20) by using the ode45 function in MATLAB. To present the simulation and the results, a brief description of the simulation about the moving load system is given first. Then the identification results are presented according to Figure 6.4. The noise-free case is demonstrated first to verify the proposed method, and then the noise-corrupted case is presented in the same manner.

6.4.1.1 Generation of simulated data for a point load system

Before nondimensionalising, the simulation of this study is based on a 40 m long simply supported beam with measurement locations at 4m spacing. The bending stiffness and the mass per unit length of the bridge is set as $EI = 1.26\text{e}+11 (Nm^2)$ and $\bar{m} = 1.2\text{e}+04$ (kg/m), respectively. Note that these bridge parameters are primarily drawn from the reference [196]. The speed of the moving concentrated load is set as 48 (km/h), the lowest national speed limit in the UK. The natural frequencies are computed from Eq. (6.16), but the damping ratios are randomly assigned, and only the first five modes are considered, see Table 6.2. The first natural frequency is about 3.181 (Hz), and the corresponding damping ratio is 0.02, which are reasonable assignments based on the Institution of Civil Engineering (ICE) design and practice guides [209].

Table 6.2. Actual modal parameters for the bridge

Mode No. k	Actual natural frequency f_k (Hz)	Actual damping ratios ξ_k	Actual modal masses M_k
1	3.181	0.020	240000
2	12.725	0.020	240000
3	28.631	0.020	240000
4	50.900	0.020	240000
5	79.531	0.020	240000

As for the moving point load, it is defined based on Eq. (6.6). The mass of the vehicle M_v is assumed to be 4.63e+04 (kg), while the vertical acceleration of the vehicle, which simulates the dynamic effect of vehicle motion due to road roughness, is characterised by "coloured noise". Figure.6.6 (a) shows the simulated moving load acceleration $\hat{z}''(\tau)$ when the moving load is crossing the bridge. Note that the input signal is constructed from a normally distributed white noise with zero mean and $5\% M_v g$ standard deviation, then scaled by the corresponding frequency in the frequency domain, which is similar to the pink noise whose PSD is inversely proportional to the frequency of the signal.

To nondimensionalise this moving load system, we will first nondimensionalise the dimensional parameters according to Eq. (6.12), which is presented in Table 6.3, then nondimensionalise the time-dependent variables based on Eq. (6.13).

Table 6.3. Nondimensionalised actual parameters for the bridge

| Actual frequency ratio κ_k	Actual damping ratio ξ_k	Actual mass ratio μ_k	Gravitational Acceleration β	Nondimensionalised vehicle speed α
k^2 ($k=1,2,3,4,5$)	0.020 (same for all modes)	0.1929 (same for all modes)	6.13e+04	0.0167

In consideration of the spectral leakage problem in the subsequent Fourier analysis, the total measurement time length should be as long as possible to let the bridge response dies out. In terms of the nondimensionalised vehicle speed given by Table 6.3, the total measurement time length is set as 360, which according to Eq. (6.13), is equivalent to 18.0106s.

Figure.6.6 (b) illustrates the acceleration response of the bridge at the middle of the span due to the moving load. It clearly shows that within this period, the amplitude of the response attenuates to zero. As for the red vertical lines shown in Figure.6.6 (b), they depict the time that the vehicle enters and leaves the bridge, respectively.

Note that the same level of noise was added to the input signal (5% noise) and the acceleration responses (5% noise) measured from the bridge. In Figure 6.6, the solid black line depicts the noise-free case, while the red dot line demonstrates the noise-contaminated scenario. Besides, as one can observe from Figure 6.6 that both the input and the output signals are nonstationary.
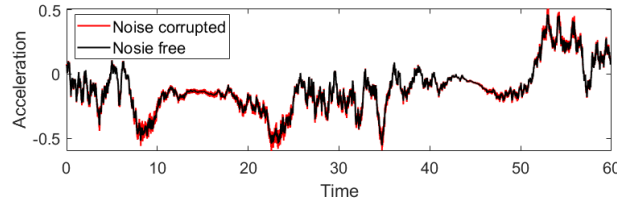
(a) Time history of the moving load acceleration when moving across the bridge (5% noise-corrupted)
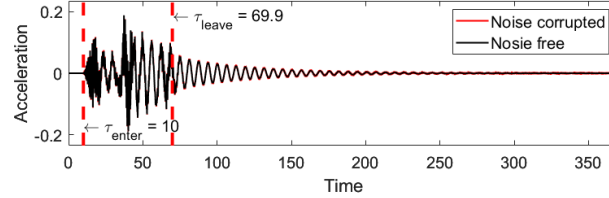
(b) Time history of the bridge acceleration (obtained at the middle of the bridge) (5% noise-corrupted)

Figure 6.6. Time history of the input (moving loading) and output (bridge) acceleration responses

As for the mode shape of the structure, for the simply supported beam, the mode shapes are sinusoids, they can be simulated based on Eq. (6.17). After being nondimensionalised, the original 40m long bridge will have unit length. Accordingly, the mode shapes of the bridge are confined within one, see Figure 6.7 (c).

6.4.1.2 Identification results

In this section, the noise-free identification results based on the moving load will be presented first, followed by the noise-contaminated case. In terms of our proposed two-stage method, the output-only analysis will be conducted first. In the first stage, the natural frequency ratios, mode shapes, and damping ratios are identified from the nondimensionalised bridge acceleration response only by using the FDD associated with the real-valued spectral density matrix.

I.   Noise-free

(a) Noise-free 1st and 2nd singular value plot (log-log scale)

(b) Actual and identified mode shape plot. From left to right, mode 1 to mode 5.
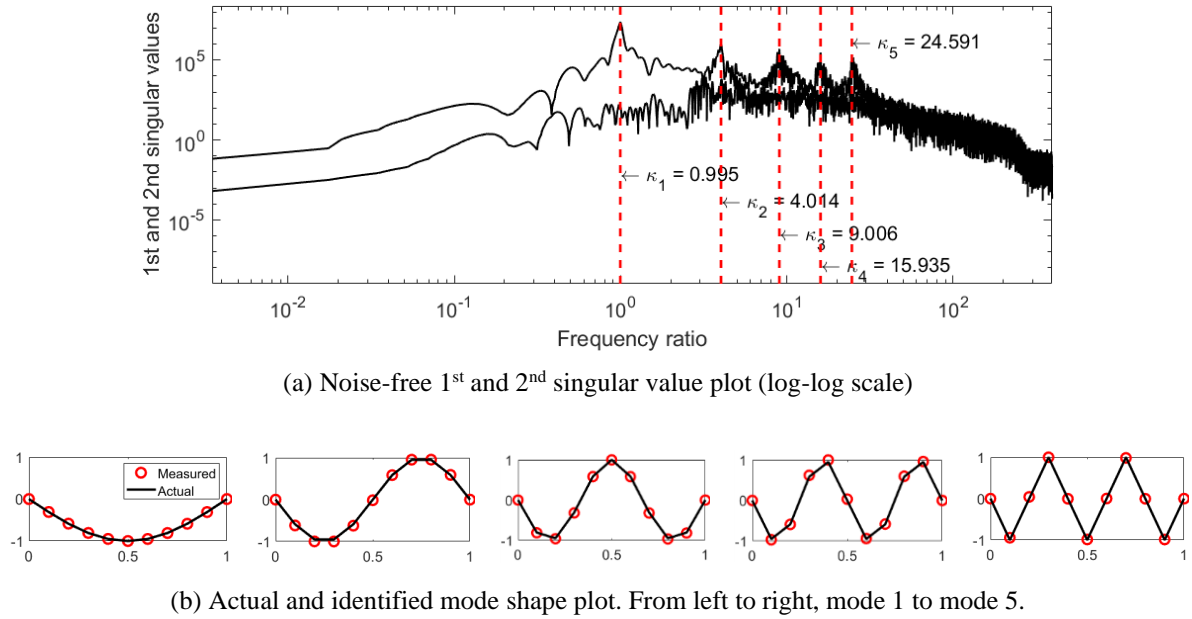
Figure 6.7 Noise-free output-only identification of the frequency ratios and mode shape identification based on the FDD according to the moving load system

Figure 6.7 (a) shows the traditional FDD singular value plot, while Figure 6.7 (b) is the comparison between the actual and the estimated mode shapes. From Figure 6.7 (a), we can see that there are five dominant peaks in the 1st singular value plot. Compare the identified peak values shown on the plot with the actual values given in Table 6.3, it is obvious that they are very close. Further, by looking at the actual and the identified mode shape comparison plot in Figure 6.7 (b), it is clear that at the dominant frequency ratios, the system is effectively governed by the corresponding modes. This means that the identified mode shapes for the VBI system can be used to represent the mode shapes of the bridge subsystem. Then we can use the estimated mode shapes to decouple the bridge system.

From Figure 6.8, we observe that the optimised line fit perfectly with the measured data for all five modes, and they are in line with the theoretical lines (calculated based on the actual modal parameter values), even for the higher modes. Note that the values are shown in Figure 6.7 and Figure 6.8 are all identified modal parameter values.

It is worth mentioning that, in the first stage, to identify the frequencies from the singular value plot, we can use the peak-picking method, namely, take the frequency value where the peak occurs. While in the section stage, the optimisation method we discussed in chapter 5 is used to extract the modal parameters in terms of Eq. (6.41). Whereas in the second stage, a frequency range needs to be defined for each mode. In this case, we can use the estimated frequencies obtained in the first stage to define the frequency ranges needed in the second stage. These frequencies are used as the reference frequencies,

and the boundary of the frequency range for each mode is set empirically with 10% below and 10% higher than the reference frequency.
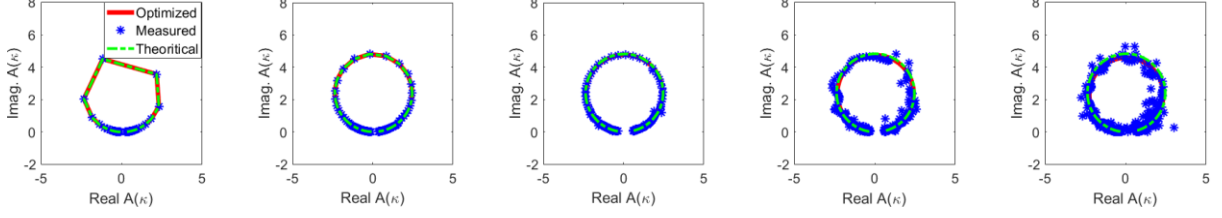
(a) Noise-free Nyquist Accelerance FRF plots for the SDOF systems
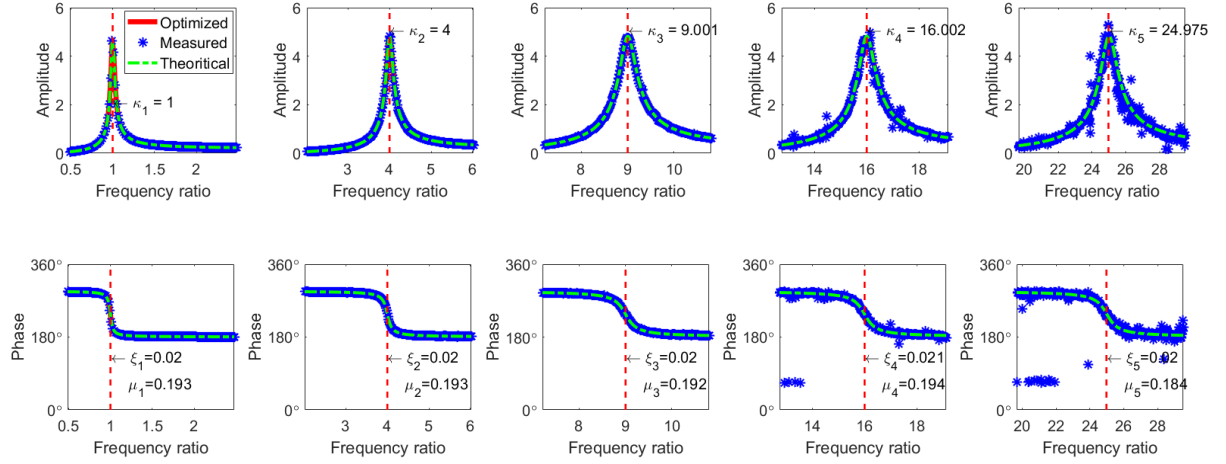
(b) Noise-free Accelerance FRF plots for the SDOF systems (linear scale) (1st row: amplitude ratio between the input and output responses; 2nd row: phase angle between the input and output responses)

Figure 6.8. Noise-free input-output system identification Accelerance FRF plots. From left to right, mode 1 to mode 5.

It is well known that for viscous damping, it is the Mobility (ratio between the response velocity and the input force) traces out an exact circle, while others, such as the Receptance (ratio between the response displacement and input force) and the Accelerance (ratio between the response acceleration and input force) only approximate to this shape [73]. From Figure 6.8 (a), we can see that except for the first mode, the Nyquist plot of the rest modes all close to a circle.

The estimated modal parameters together with their actual values and the relative errors (shown in the bracket below the estimates) from both the first stage and the second stage are grouped together in Table 6.4. Note that the values given in Table 6.4 are from a sample case, and a small change to the random input can result in a different value. Therefore, by analysing each value in the table is useless and unnecessary. Nevertheless, some overall phenomena can still be observed from the results given in Table 6.4.

According to Table 6.4, we can see that the accuracy of the identified frequency ratios is improved in the second stage, as the input-output method generates smaller relative errors. While for the damping ratios, the accuracy of the results obtained from the free decay analysis is comparable with the ones obtained from the optimisation procedure. As for the modal mass ratios, the optimisation procedure successfully identified them, with the highest relative errors being -4.642%.

Table 6.4 Noise-free dentification result based on the moving load system

|	Actual	1	4	9	16	25
κ_k	Output-only	0.9948 (-0.517)	4.0142 (0.356)	9.0058 (0.065)	15.9347 (-0.408)	24.5915 (-1.634)
	Input-output	1 (0.001)	4.0001 (0.002)	9.0006 (0.006)	16.0020 (0.013)	24.9748 (-0.101)
ξ_k	Actual	0.02	0.02	0.02	0.02	0.02
	Output-only	0.0201 (0.515)	0.0200 (-0.030)	0.0201 (0.348)	0.0202 (1.047)	0.0205 (2.663)
	Input-output	0.0200 (-0.014)	0.0201 (0.273)	0.0201 (0.352)	0.0205 (2.515)	0.0201 (0.385)
μ_k	Actual	0.1929	0.1929	0.1929	0.1929	0.1929
	Input-output	0.1931 (0.070)	0.1932 (0.154)	0.1918 (-0.584)	0.1940 (0.581)	0.1840 (-4.642)

(Note: κ_k — Frequency ratios; ξ_k — Damping ratios; μ_k — Modal mass ratios; values inside the brackets represent the relative errors (%)).

Besides, we can observe that for three modal parameters of interest, the estimated results obtained from the optimisation procedure has an overall increased relative error in terms of the number of the modes. In contrast, the results estimated from the output-only method does not have a clear trend. Overall, the proposed two-stage method successfully identified the bridge modal parameters by using both the input and the output signals, as the largest relative error among all three estimates is less than 6%.
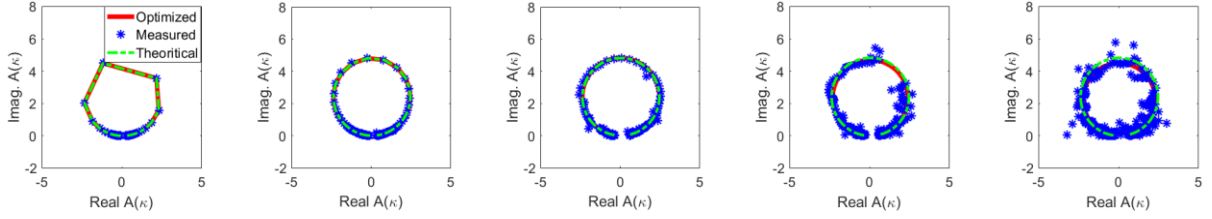
II.   Noise-corrupted
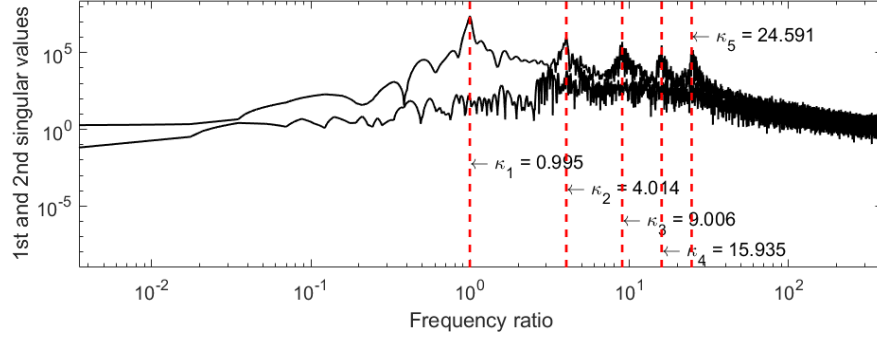
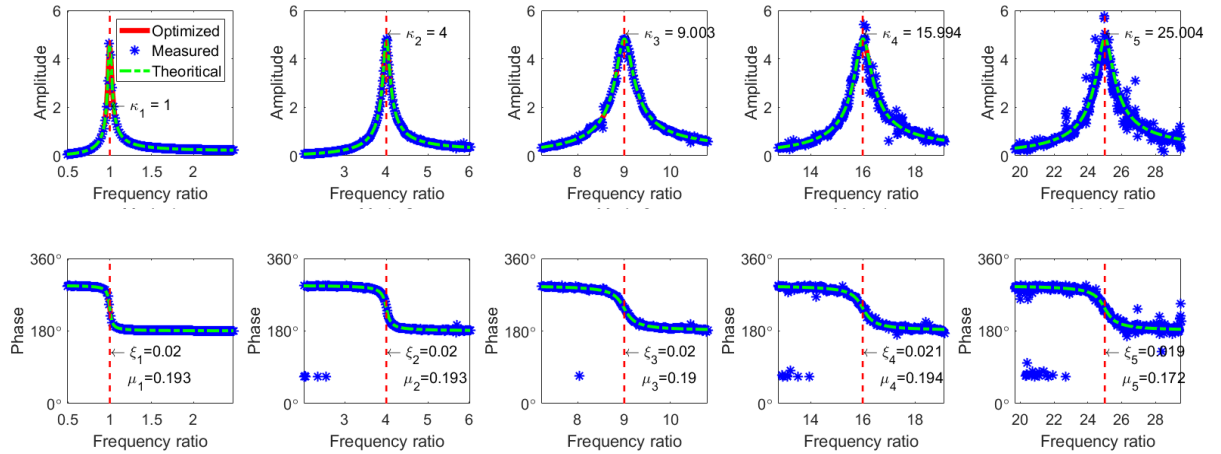(a) Noise-corrupted 1st and 2nd singular value plot (log-log scale)
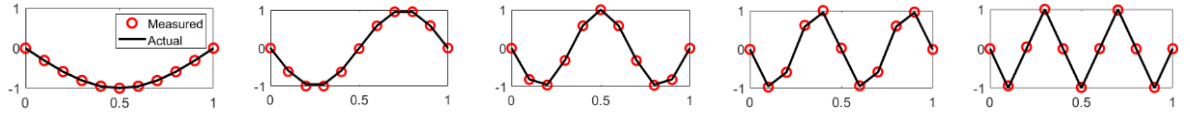
(b) Actual and identified mode shape plot. From left to right, mode 1 to mode 5.

Figure 6.9. Noise-corrupted frequency ratios and mode shape identification based on the FDD

(a) Noise-corrupted Nyquist Acceleration FRF plots for the SDOF systems

(b) Noise-corrupted Acceleration FRF plots for the SDOF systems (linear scale) (1st row: amplitude ratio between the input and output responses; 2nd row: phase angle between the input and output responses)

Figure 6.10. Noise-corrupted input-output system identification Acceleration FRF plots. From left to right, mode 1 to mode 5.

Table 6.5 Noise-corrupted dentification result based on the moving load system

	Actual	1	4	9	16	25
κ_k	Output-only	0.9948 (-0.517)	4.0142 (0.356)	9.0058 (0.065)	15.9347 (-0.408)	24.5915 (-1.634)
	Input-output	1 (0.004)	4.0001 (0.003)	9.0029 (0.033)	15.9945 (-0.035)	25.0044 (0.018)
ξ_k	Actual	0.02	0.02	0.02	0.02	0.02
	Output-only	0.0200 (-0.087)	0.0197 (-1.518)	0.0189 (-5.263)	0.0191 (-4.521)	0.0187 (-6.743)
	Input-output	0.0200 (0.188)	0.0200 (0.025)	0.0197 (-1.501)	0.0206 (3.103)	0.0192 (-4.226)
μ_k	Actual	0.1929	0.1929	0.1929	0.1929	0.1929
	Input-output	0.1932 (0.135)	0.1929 (-0.021)	0.1904 (-1.323)	0.1938 (0.435)	0.1723 (-10.695)

(Note: κ_k — Frequency ratios; ξ_k — Damping ratios; μ_k — Modal mass ratios; values inside the brackets represent the relative errors (%)).

From Table 6.5, we see that after adding the noise, we can see that there is an increase in the maximum relative error (-10.695%), which is coming from the 5th modal mass ratio. However, the proposed two-stage method can still identify the modal parameters to a certain level of accuracy. Compared to the output-only result, the input-output results of the frequency ratios and the damping ratios are generally better, as they still maintain lower relative errors.

6.4.2 Evaluation of the proposed input-output identification method based on the quarter car system

In this section, the proposed system identification approach will be demonstrated by using a quarter car model with a total vehicle mass M_v traversing a simply supported Euler-Bernoulli beam, as shown in Fig. 6.2 (b). The properties of the beam are the same as the beam we used in the moving load system. Likewise, the way to demonstrate the simulation and the identified results are the same as the previously discussed moving load system.

6.4.2.1 Generation of simulated data for a quarter car model system

The simulation of the quarter car model traversing a simply supported beam is similar to the moving load system we have discussed before. The weight of the moving load is assumed to be equal to the weight of the quarter car, i.e., $M_v = 4.63e+04$ (kg). However, since a spring and a damper are introduced into the vehicle system, the suspension stiffness K_v and the damping coefficient C_v need

to be defined. The value of these two coefficients are obtained according to reference [210] that K_v and C_v equal to 4.87e+06 (N/m) and 3.14e+05 (Ns/m), respectively. Accordingly, the frequency of the sprung mass system $\omega_v = \sqrt{K_v / M_v} = 10.2559$ (rad/s). $\xi_v = C_v / 2\omega_v M_v = 0.3306$.

For the quarter car system, it is simulated based on Eq. (6.34) and Eq. (6.35) by using the ode45 function in MATLAB. In the meantime, the road roughness is considered. In section 6.2.2, we had a brief introduction to the road roughness. In this section, we simulate the road roughness based on Eq. (6.3), and road class D (see Figure 6.1 (c)) is selected to represent a relatively bad road surface.

To see the effect of the road roughness on the identification results, our discussion in the next section is started with a road roughness-free case. Note that the noise-contaminated case for the quarter car system without road roughness will not be discussed. It is because the input spectral density (SD) of the quarter car is similar to the simulated moving load case.

Table 6.6. Nondimensionalised actual parameters for VBI system

| Actual frequency ratio κ_k	Actual bridge damping ratio ξ_k	Actual vehicle damping ratio ξ_v	Actual mass ratio μ_k	Gravitational Acceleration β	Nondimensionalised vehicle speed α
$k^2 \kappa_1$ ($\kappa_1 = 1.949$, $k = 1,2,3,4,5$)	0.020 (same for all modes)	0.3306	0.1929 (same for all modes)	0.0023	0.0325

The same as before, to nondimensionalise this quarter car model VBI system, we first nondimensionalise the dimensional parameters according to Eq. (6.22), which is presented in Table 6.6, then nondimensionalise the time-dependent variables based on Eq (6.23). And in consideration of the spectral leakage problem in the subsequent Fourier analysis, the total measurement time length should be as long as possible to let the bridge response dies out. In terms of the nondimensionalised vehicle speed given by Table 6.6, the total measurement time length is set as 150, which according to Eq. (6.23), is equivalent to 14.6257s.
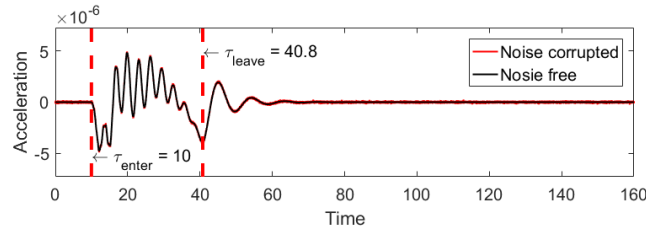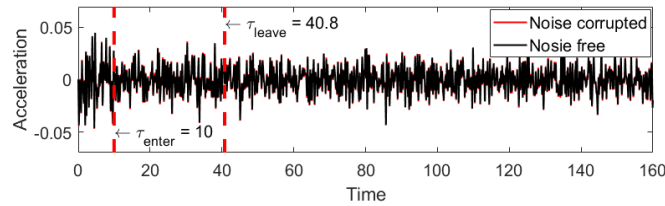
(a) Time history of the moving quarter car suspended mass acceleration when moving across the bridge (5% noise-corrupted without road roughness)
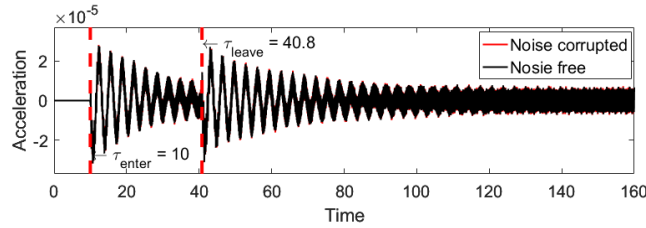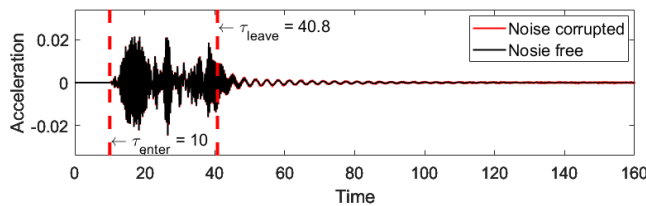

(b) Time history of the bridge acceleration (obtained at the middle of the bridge) (5% noise-corrupted without road roughness)

Figure 6.11 Time history of the input (moving quarter car suspended mass) and output (bridge) acceleration responses (without road roughness)

(a) Time history of the moving quarter car suspended mass acceleration when moving across the bridge (5% noise-corrupted)

(b) Time history of the bridge acceleration (obtained at the middle of the bridge) (5% noise-corrupted)

Figure 6.12 Time history of the input (moving quarter car suspended mass) and output (bridge) acceleration responses (with road roughness)

Figures 6.11 and Figures 6.12 depict the input and the output acceleration responses of the quarter car system. Figure.6.11 (a) and Figure 6.12 (a) show the simulated moving quarter car suspended mass acceleration $\hat{z}_1''(\tau)$ when the vehicle is crossing the bridge. In Figure. 6.11 (b) and Figure 6.12 (b), the acceleration response of the simply supported beam at the middle of the beam due to the moving quarter car clearly shows that within this period, the amplitude of the response attenuates to zero. As for the red

200

vertical lines shown in both Figure. 6.11 and Figure 6.12, they represent the time that the vehicle enters and leaves the bridge, respectively.

Compare Figure.6.11 (a) and Figure 6.12 (a), we can see that, without and with the effect of road roughness, the behaviour of the quarter car suspended mass acceleration $\hat{z}_1''(\tau)$ is different. This is because we assume the initial values of the vehicle, i.e., displacement and velocity, are equal to $\hat{r}(0)$ and $\hat{r}'(0)$. Thus, when there is no road roughness, Figure.6.11 (a), the contact point load acceleration between the two subsystems equals zero before the vehicle enters the bridge. In other words, the vehicle starts to move at $\tau = 10$.

Besides, the same level of noise was added to the input signal (5% noise) and the acceleration responses (5% noise) measured from the bridge. As one can observe from Figure 6.11 (a) and Figure 6.12 (a), the introduction of the road roughness transformed the nonstationary suspended mass acceleration signal into a stationary signal. However, in either case, the acceleration signal of the bridge is still nonstationary.

6.4.2.2 Identification results and discussion

In this section, we will present our simulation results for different cases first, namely, noise-free (without and with road roughness) and noise-contaminated cases, then discuss the results afterward.

### I. Identification results

Noise-free without road roughness
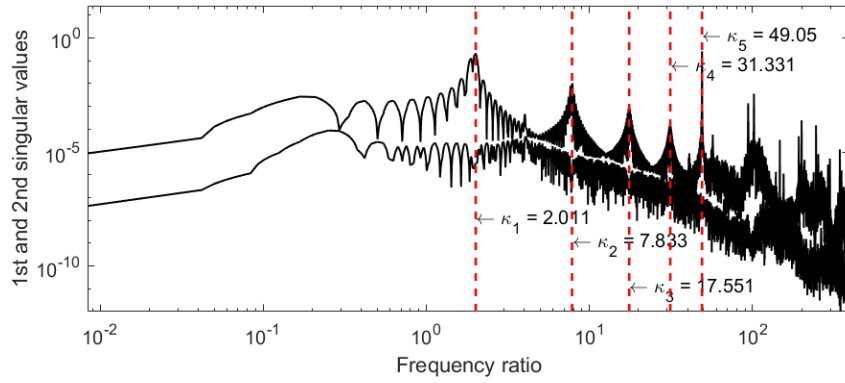
(a) Noise-free 1st and 2nd singular value plot (log-log scale)
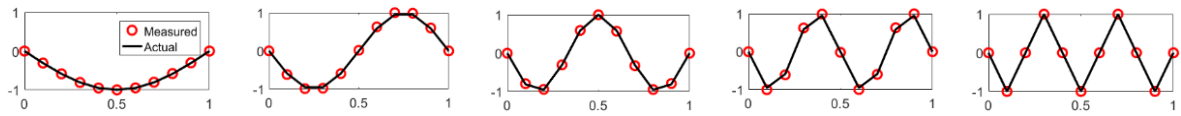
(b) Actual and identified mode shape plot. From left to right, mode 1 to mode 5.

Figure 6.13 Noise-free output-only identification of the frequency ratios and mode shape identification based on the FDD according to the quarter car system without roughness
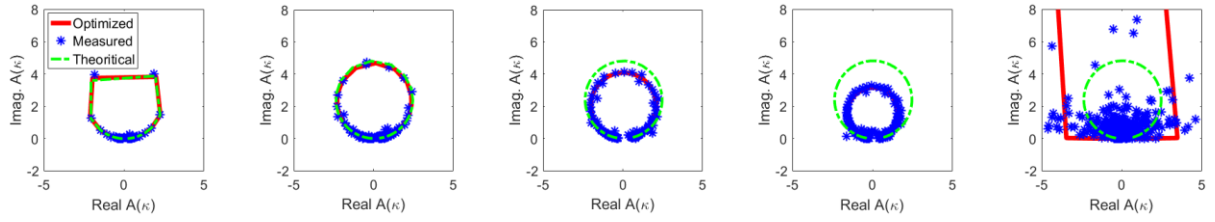
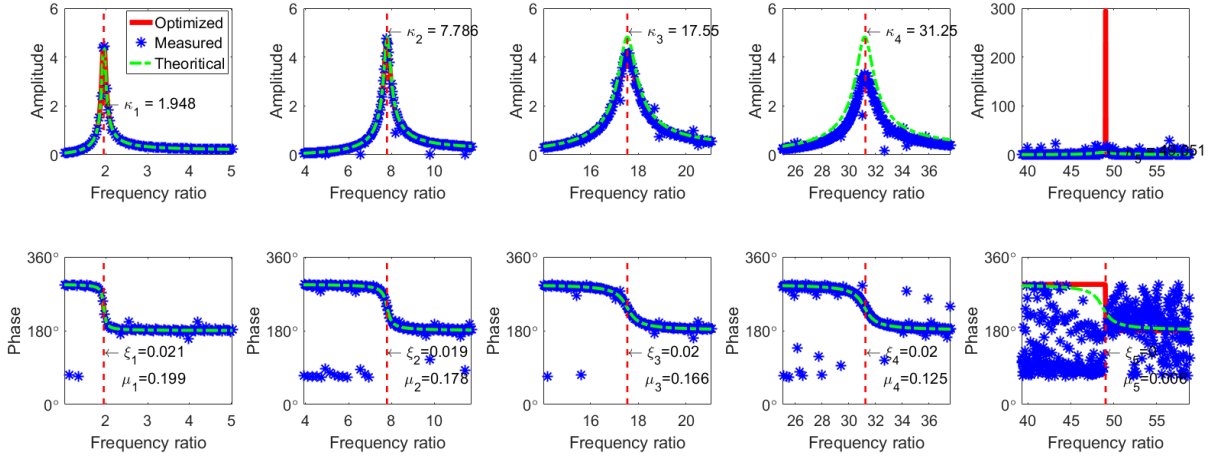(a) Noise-free Nyquist Accelerance FRF plots for the SDOF systems

(b) Noise-free Accelerance FRF plots for the SDOF systems (linear scale) (1st row: amplitude ratio between the input and output responses; 2nd row: phase angle between the input and output responses)

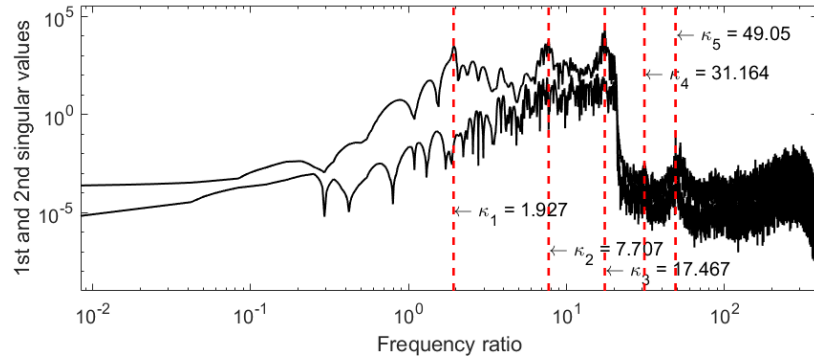Figure 6.14. Noise-free input-output system identification Accelerance FRF plots (without road roughness). From left to right, mode 1 to mode 5

Table 6.7 Noise-free dentification result based on the quarter car system with road roughness

κ_k	Actual	1.9489	7.7958	17.5405	31.1832	48.7237
	Output-only	2.0106 (3.162)	7.8329 (0.476)	17.5507 (0.058)	31.3315 (0.476)	49.0497 (0.669)
	Input-output	1.9476 (-0.068)	7.7864 (-0.120)	17.5503 (0.056)	31.2497 (0.213)	49.0508 (0.671)
ξ_k	Actual	0.02	0.02	0.02	0.02	0.02
	Output-only	0.0180 (-9.966)	0.0191 (-4.644)	0.0125 (-37.268)	0.0128 (-35.801)	0.0001 (-99.481)
	Input-output	0.0206 (2.800)	0.0190 (-5.182)	0.0202 (0.943)	0.0197 (-1.420)	0.0000 (-99.950)
μ_k	Actual	0.1929	0.1929	0.1929	0.1929	0.1929
	Input-output	0.1986 (2.949)	0.1780 (-7.727)	0.1655 (-14.202)	0.1246 (-35.398)	0.0059 (-96.933)

(Note: κ_k — Frequency ratios; ξ_k — Damping ratios; μ_k — Modal mass ratios; values inside the brackets represent the relative errors (%)).

Noise-free with road roughness

(a) Noise-free 1st and 2nd singular value plot (log-log scale)
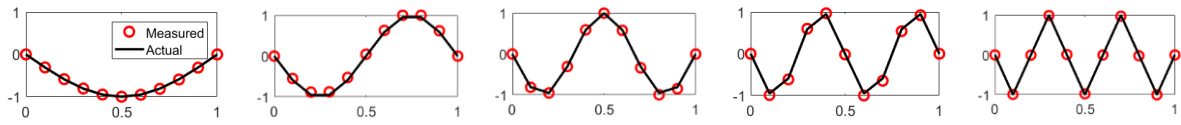
(b) Actual and identified mode shape plot. From left to right, mode 1 to mode 5.

Figure 6.15 Noise-free output-only identification of the frequency ratios and mode shape identification based on the FDD according to the quarter car system (with road roughness)
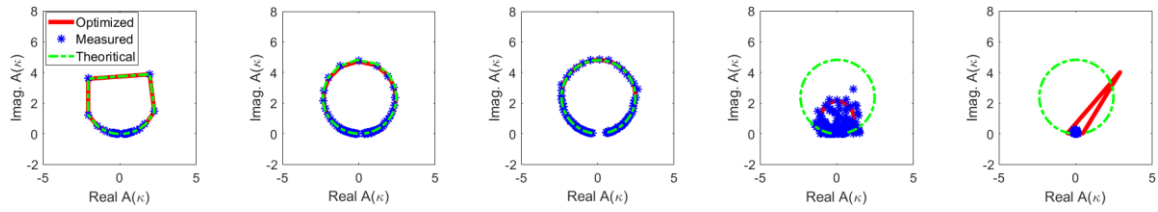
(a) Noise-free Nyquist Accelerance FRF plots for the SDOF systems
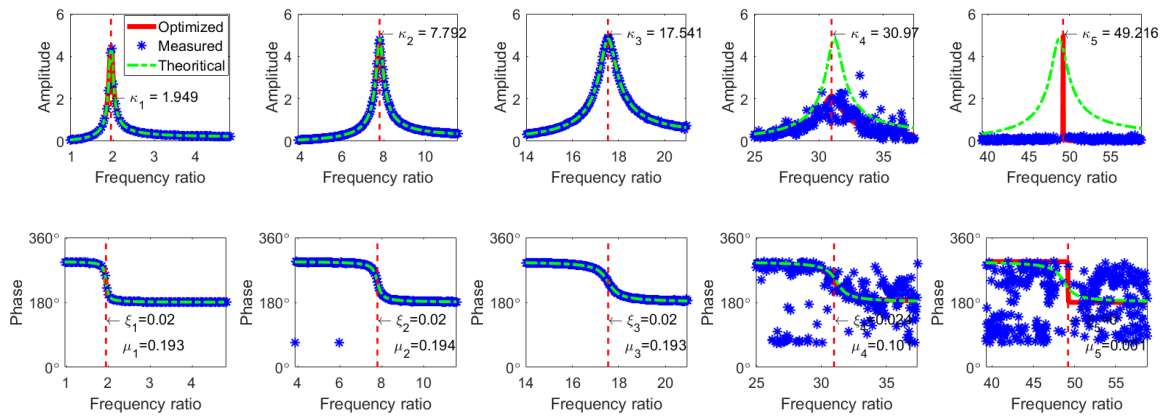
(b) Noise-free Accelerance FRF plots for the SDOF systems (linear scale) (1st row: amplitude ratio between the input and output responses; 2nd row: phase angle between the input and output responses)

Figure 6.16. Noise-free input-output system identification Accelerance FRF plots (with road roughness). From left to right, mode 1 to mode 5

Table 6.8 Noise-free dentification result based on the quarter car system with road roughness

κ_k	Actual	1.9489	7.7958	17.5405	31.1832	48.7237
	Output-only	1.9268 (-1.136)	7.7072 (-1.136)	17.4669 (-0.420)	31.1639 (-0.062)	49.0497 (0.669)
	Input-output	1.9490 (0.002)	7.7925 (-0.043)	17.5408 (0.001)	30.9698 (-0.684)	49.2158 (1.010)
ξ_k	Actual	0.02	0.02	0.02	0.02	0.02
	Output-only	0.0203 (1.311)	0.0203 (1.513)	0.0203 (1.376)	0.0090 (-55.047)	0.0071 (-64.485)
	Input-output	0.0200 (0.147)	0.0205 (2.396)	0.0200 (-0.068)	0.0236 (18.009)	0.0001 (-99.650)
μ_k	Actual	0.1929	0.1929	0.1929	0.1929	0.1929
	Input-output	0.1927 (-0.114)	0.1939 (0.4970)	0.1934 (0.2330)	0.1006 (-47.8710)	0.0009 (-99.557)

(Note: κ_k — Frequency ratios; ξ_k — Damping ratios; μ_k — Modal mass ratios; values inside the brackets represent the relative errors (%)).

Noise-corrupted with road roughness

(a) Noise-corrupted 1st and 2nd singular value plot (log-log scale)
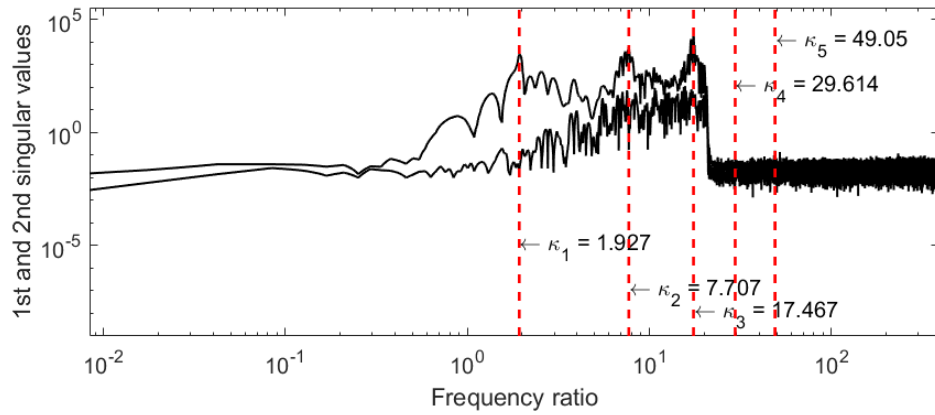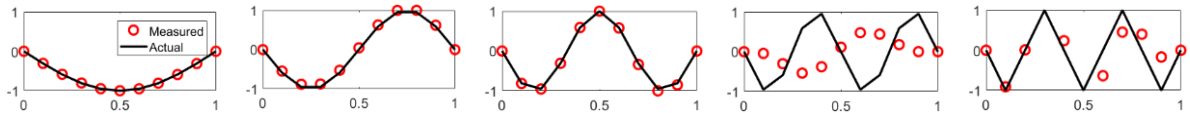

(b) Actual and identified mode shape plot. From left to right, mode 1 to mode 5.

Figure 6.17 Noise-corrupted output-only identification of the frequency ratios and mode shape identification based on the FDD according to the quarter car system (with road roughness)
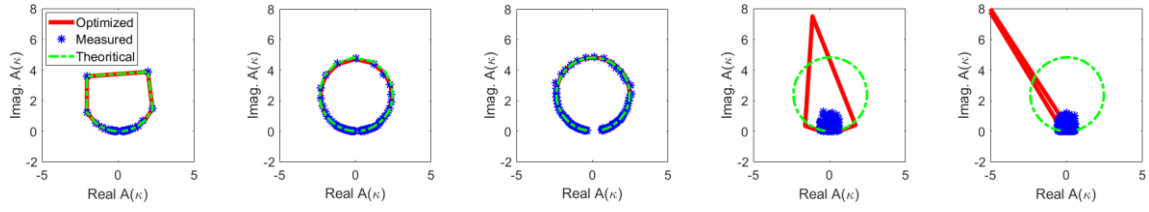
(a) Noise-corrupted Nyquist Accelerance FRF plots for the SDOF systems
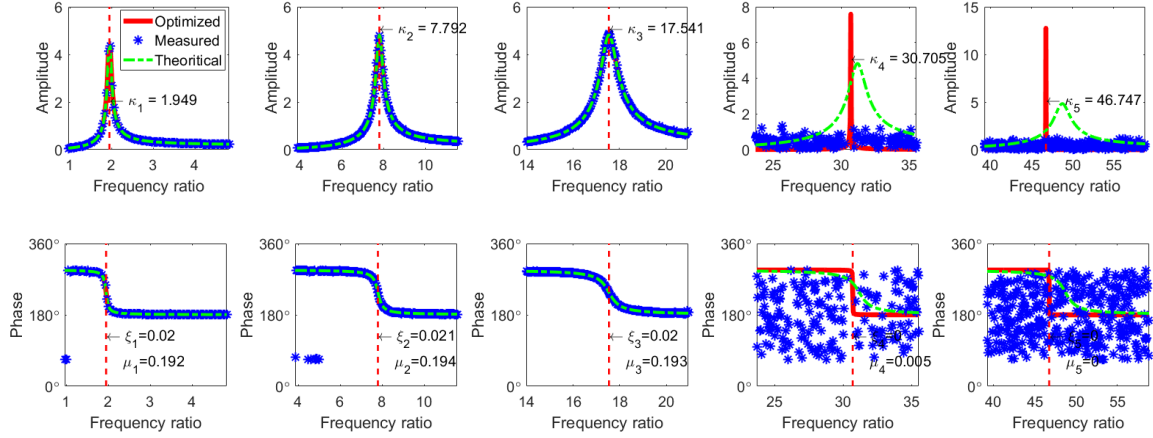
(b) Noise-corrupted Accelerance FRF plots for the SDOF systems (linear scale) (1st row: amplitude ratio between the input and output responses; 2nd row: phase angle between the input and output responses)

Figure 6.18. Noise-corrupted input-output system identification Accelerance FRF plots (with road roughness). From left to right, mode 1 and mode 5

Table 6.9 Noise-corrupted dentification result based on the quarter car system

κ_k	Actual	1.9489	7.7958	17.5405	31.1832	48.7237
	Output-only	1.9268 (-1.136)	7.7072 (-1.136)	17.4669 (-0.420)	29.6141 (-5.032)	49.0497 (0.669)
	Input-output	1.9490 (0.004)	7.7921 (-0.048)	17.5411 (0.003)	30.7052 (-1.533)	46.7473 (-4.056)
ξ_k	Actual	0.02	0.02	0.02	0.02	0.02
	Output-only	0.0192 (-4.033)	0.0090 (-54.866)	0.0201 (0.729)	0.0001 (-99.677)	0.0000 (-99.865)
	Input-output	0.0200 (-0.014)	0.0205 (2.729)	0.0200 (0.202)	0.0003 (-98.455)	0.0000 (-99.928)
ξ_k	Actual	0.1929	0.1929	0.1929	0.1929	0.1929
	Input-output	0.1925 (-0.236)	0.1941 (0.627)	0.1934 (0.254)	0.0047 (-97.542)	0.0004 (-99.776)

(Note: κ_k — Frequency ratios; ξ_k — Damping ratios; ξ_k — Modal mass ratios; values inside the brackets represent the relative errors (%)).

II.   Identification results analysis

We have a lot of similar figures and tables presented in the last subsection to demonstrate the ability of the proposed method for the quarter car case. However, we are unable to, and it is also unnecessary to discuss all the figures and values listed in each table because each plot and table are just one realisation of our numerical experiment. Instead, we will concentrate on explaining the most obvious phenomena one can observe from those figures and tables. Here, we identify the following two phenomena.

1. If we cannot successfully identify the mode shapes in the first place, then we are unable to identify the modal parameters in the second stage. We can see this from noise-corrupted with road roughness case. However, even if we can successfully identify all the mode shapes from the first stage identification, we could still fail to extract the rest of the modal parameter information from the second stage identification. This phenomenon is visible from the noise-free with road roughness case.

2. The accuracy of the identification results deteriorates with the increase of the mode number. Especially for the damping ratio and modal mass of the last two modes, we almost failed to identify them. This phenomenon can be observed from all three cases, i.e., noise-free without road roughness, noise-free with road roughness, and noise-corrupted with road roughness.

To explain these two phenomena, first, we shall observe the impact factors for the VBI system identification via the proposed method. Typically, noise level, mass ratios, vehicle frequency, vehicle speed, the maximum value of the spatial frequency, and the amplitude of the road roughness are all the potential factors that can influence the identification result. And for this proposed two-stage method, the accuracy of the estimated mode shapes from the first stage should also be counted as an influencing factor because we used the mode shape decomposed responses to estimate the rest of the modal parameters in the second stage of our identification.

Effect of the estimated mode shapes

It is obvious that if the mode shapes are inaccurate, then the decomposed responses will be distorted. As a result, we cannot obtain the correct modal information through the proposed method. This is verified in the noise-corrupted road roughness case. Figure 6.17 (b) shows that we successfully identified the first three mode shapes, with the 4th and the 5th being very bad. As a result, we successfully identified the frequency ratios, damping ratios, and modal mass ratios in the second stage with good accuracy for the first three modes. From Figure 6.18, we can see that the optimised curves of the Accelerance FRF for the first three modes fit the measured data perfectly, which are aligned with their theoretical counterparts, but not for the 4th and the 5th modes. The estimation results given in Table 6.9 verifies the plots shown in Figure 6.18.

However, a good mode shapes estimation in the first stage does not guarantee a good second stage identification, and this is a problem mainly concerned with the bridge response itself rather than the technique being used. An example is given by the noise-free with road roughness case. From Figure 6.15 (b), we see that all the mode shapes are properly estimated. However, when we look at the values listed in Table 6.8, we know we only successfully identified the modal parameters for the first three modes. And only for the first three modes do the optimised curves of the Accelerance FRF are aligned with their theoretical counterparts. The reason for this is that the last two modes received less energy from the vehicle excitation than the first three modes. We can see from Figure 6.15 (a) that there is an abrupt fall in the singular value plot. Compared to the first three modes, which are located on the plateau before the drop, the 4th and the 5th modes are seated on the right-hand side of the drop with narrow peaks and lower energy.

Since we can observe all the peaks on the singular value plot, the estimated mode shapes obtained from the corresponding first singular vectors can be good estimates of the actual mode shapes. However, due to the low energy the 4th and the 5th modes received from the excitation and the narrow active frequency ranges, we are unable to trace out a circle for the Accelerance around the corresponding resonance frequency. As a result, even with a good mode shape estimation from the first stage, we still failed to identify the modal parameters in the second stage.

Nondimensionalised spatial frequency

It seems that the abrupt change in the singular value plot of the bridge output response is the tipping point to separate a good estimation and a bad one. Thus, this tipping point is worthy of further investigation. Interestingly, we can calculate the location of this tipping point based on vehicle frequency, vehicle speed, and maximum value of the spatial frequency. Since the location of the tipping point is so important, we shall give it an official name, namely, the nondimensionalised spatial frequency, and it is given by the following equation

$$\varpi = \frac{2\pi n \upsilon}{\omega_v} \tag{6.45}$$

where n is the spatial frequency, υ is the vehicle speed, ω_v is the vehicle's natural frequency. Actually, ϖ is not unfamiliar to us, we showed that the spatial frequency is related to the vehicle frequency and vehicle speed through Eq. (6.2) in section 6.2.2. Hence, we named it the nondimensionalised spatial frequency in analogy to the spatial frequency n. The range of the plateau we saw in Figure 6.17 (a) is determined by this parameter, and it is exactly the frequency ratio that corresponds to the critical point or the dropping point of the plateau.

It is worth pointing out that since ϖ is a function of three parameters, and those three parameters are related to the vehicle and the road roughness. Therefore, we will discuss their effect separately in the following context.

(a) Effect of the input spectral density with road roughness

According to our identification results for the quarter car system with road roughness, the modes which are located on the left-hand side of the critical point of the plateau can definitely be identified by the proposed two-stage method. But for those located on the right-hand side of the critical point of the plateau, the proposed method is unable to generate a satisfactory result. Therefore, we can make an inference about the estimation results just by looking at the singular value plot of the output response of the bridge.

From the definition of the nondimensionalised spatial frequency, we know that it a function of three parameters the vehicle speed and the vehicle's natural frequency. So, it makes sense that we plot the generalised input spectral density (SD), the generalised output (modal coordinates) SD, and the Accelerance FRF together in the same figure for each mode in the following Figure 6.19.
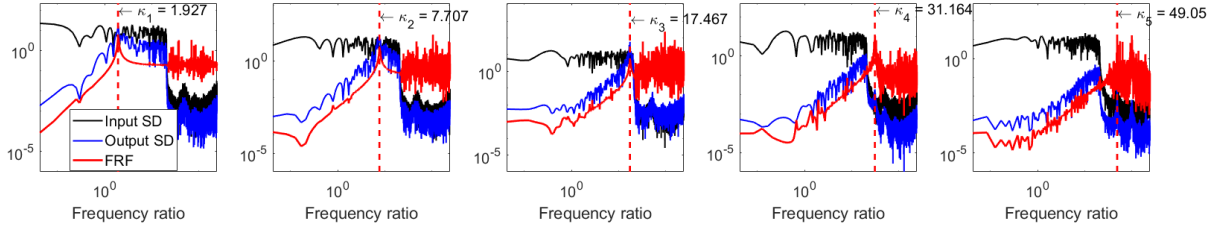
(a) Noise-free generalised input SD, generalised output SD and the Accelerance FRF plots for the quarter car system with road roughness
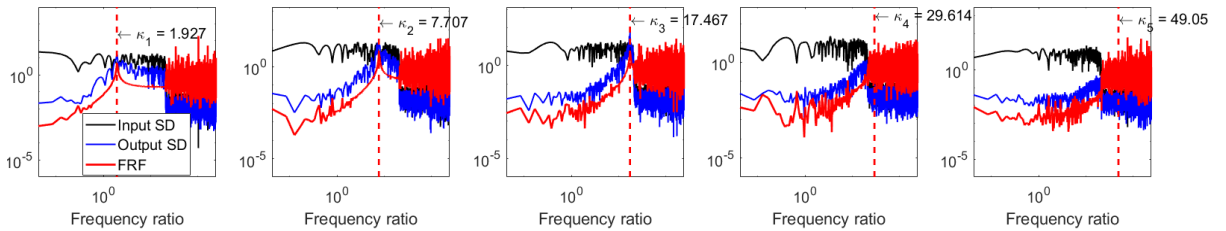
(b) Noise-corrupted generalised input SD, generalised output SD and the Accelerance FRF plots for the quarter car system with road roughness

Figure 6.19 Quarter car system generalised input SD, generalised output SD and the Accelerance FRF plots with road roughness. From left to right, mode 1 and mode 5

From Figure 6.19, we can observe three things. First, the input SD for each mode has a plateau-like the output SD as well. Second, the spectral density of the generalised input does not have a prominent peak for each mode. Third, if the output SD has a similar profile as the input SD along the frequency line,

the closer their trends are, the smoother the FRF is, which as a result, will give us a better identification result.

(b) Effect of the input spectral density without road roughness

As for the moving load system and the quarter car system without road roughness, we plotted the generalised input SD, generalised output SD, and the Accelerance FRF plots in Figure 6.20.
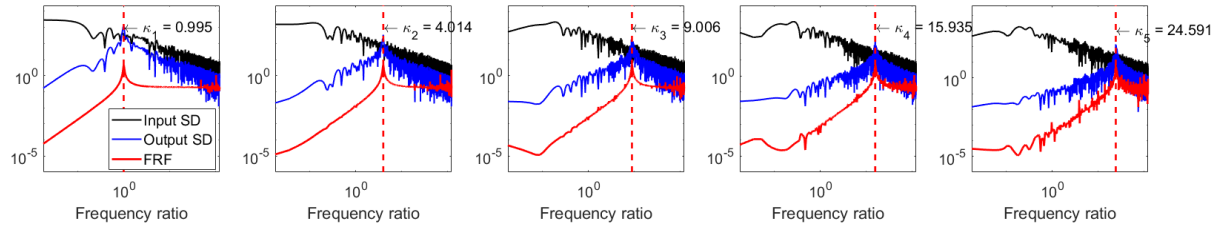
(a) Noise-free generalised input SD, generalised output SD and the Accelerance FRF plots for the moving load system
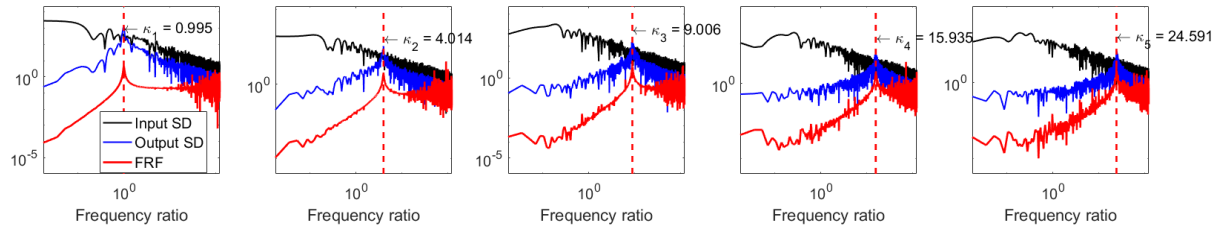
(b) Noise-corrupted generalised input SD, generalised output SD and the Accelerance FRF plots for the moving load system
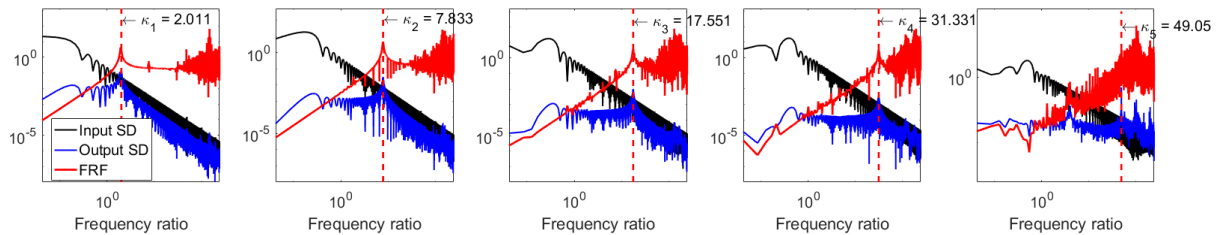
(c) Noise-free generalised input SD, generalised output SD and the Accelerance FRF plots for the quarter car system without road roughness

Figure 6.20 The generalised input SD, generalised output SD and the Accelerance FRF plots. From left to right, mode 1 and mode 5

Figure 6.20 shows that the generalised input SD for each mode for the moving load system has a similar trend as the generalised input SD for the quarter car system without road roughness. They all have an almost linear monotonic decreasing trend. The simulated pink noise (i.e., moving load) has higher power and decays slower than the generalised input SD for quarter car system without road roughness. If we compare Figure 6.20 (c) with Figure 6.19 (a), we see that the addition of the roughness has

effectively changed the generalised input SD. Besides, from Figure 6.20, it is not difficult to derive that even if the input SD is not flat, in light of our previous output-only analysis, as long as the input SD does not smear the Accelerance FRF of the system at the resonance frequency, we can still observe the peak in the singular value plot. If this condition is satisfied, this means that the FDD will be able to identify the modal shapes and natural frequencies of the system, provided that we have the input in the frequency range around the relevant natural frequency.

However, it does not guarantee good estimates of the modal parameters we identified in the second stage. If and only if the Accelerance FRF of the system is smooth around the resonate frequencies, we are able to get some good estimates, and this explains why the last two modes in Figure 6.14 give bad estimates. While for the bad estimates of the last two modes illustrated in Figure 6.16, the reasons are similar but with some extra features that the peaks of the last two modes are positioned at the right hand of the critical point, where the input SD and the output SD are both very small. As a result, the Accelerance FRF of the system is not smooth.

(c) Effect of the road roughness

Now recall the identification results in the previous sections. For the moving load system, we can identify the bridge modal parameters via the two-stage method to a certain accuracy. But for the quarter car system, we were struggling to identify the last two modes, especially the 5th modes. From Figure 6.20 (c), we can see that the Accelerance FRF of the 5th mode is very noisy. This implies that the smoother the Accelerance FRF, the better the estimation is. In other words, if the input and the output are both significantly greater than the respective noise levels that we can get a good estimate of the FRF, which is smooth.

Therefore, the effective frequency range we need to define for the second stage identification can be determined from the Accelerance FRF plot. All we need is to select a frequency range that encompasses the resonance frequency within the smooth region of the Accelerance FRF plot. When the vehicle speed and frequency are relatively fixed, the only influence factor of this frequency range is the spatial frequency, which is an important factor used in section 6.2.2 to determine the road roughness. To further demonstrate the effect of the road roughness, we plot the measured input SD in geometric coordinate for the quarter car system in the following Figure 6.21.

In this study, we treat the moving vehicle as an external input to the bridge system. However, we can also see the bridge as the external input to the vehicle system. It means that we could observe the bridge frequency component from the vehicle acceleration SD plot as well. Thus, the SD plot of the vehicle acceleration response will be the mix of bridge frequencies and the vehicle frequency. Since the bridge has two frequencies, i.e., spatial frequency and vibration frequency, we may see three kinds of frequencies in the vehicle acceleration SD plot. Usually, the spatial frequency of the bridge is very small (for a unit length beam, the first frequency is 0.5), and it occurs in a low frequency range. In Figure

6.21 (a), we can see that there is a hump around 0.5, which should be the contribution of the spatial frequency. While around 2, there is a prominent peak, which should be contributed by the frequency ratio $\kappa_1 = 1.949$. As for the vehicle's nondimensionalized frequency, it is equal to 1, and we can observe a small peak in Figure 6.21 (a) around 1.
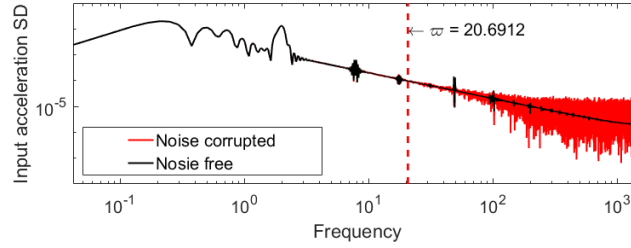
(a) Input acceleration SD plot for the for the quarter car system without road roughness
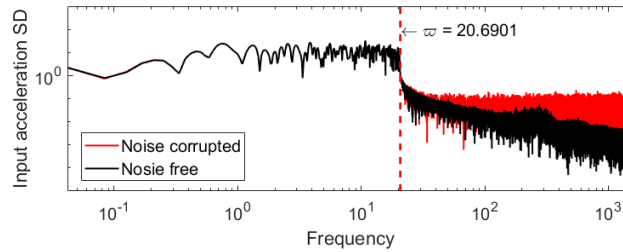
(b) Input acceleration SD plot for the for the quarter car system with road roughness

Figure 6.21 Input SD in geometric coordinate for the quarter car system

Now if we compare Figure 6.21 (a) with Figure 6.21 (b) we can see that the power of the road displacement PSD has an impact on the input SD, which means the amplitude of the road roughness can affect the identification results. Besides, the plateau in the input SD plot shown in Figure 6.21 (b) is consistent with its time domain counterpart illustrated in Figure 6.12 (b), that the road roughness transformed the nonstationary input signal to stationary signal when the vehicle is moving across the bridge. However, this transformation totally depends on the power of the road roughness compare to the vehicle's power.

Note that the nondimensionalised spatial frequency ϖ in Figure 6.21 (c) is obtained when $n = 100$ (cycle/m), $\upsilon = 13.333$ (m/s), and $\omega_v = 10.256$ (rad/s). As we see from Figure 6.19 that the first three modes are less than ϖ, therefore we can identify the first three modes very accurately.

(d) Variation of the nondimensionalised spatial frequency

Since the accuracy of the identification result largely depends on ϖ, it is worth showing how this parameter is affecting the identification. Without further explanation, we shall demonstrate the effect of this parameter with the following three cases, see Table 6.10. The identification results are presented

in Figure 6.22. For all three cases, the simulations are based on the 5% noise-corrupted quarter car system. Only the parameters shown in Table 6.10 are changed each time, while the rest of the system parameters remain the same with the quarter car system we discussed in the last section.

Here, the label, i.e., (a), (b), (c), is associated with the figure number in Figure 6.22. Under each label, it has two figures. The first one shows the generalised input SD, generalised output SD and the Accelerance FRF plots. In each plot, the estimated frequency ratios from the first stage identification are shown in the figure. The second figure gives Accelerance FRF plots for each mode. The identified frequency ratios, damping ratios, and the modal mass ratios from the second stage identification are all marked appropriate on the plot.

Note that among the three parameters, i.e., spatial frequency, vehicle speed, and vehicle natural frequency, only the vehicle natural frequency is related to the frequency ratios κ_k. According to the relationship defined between the frequency ratio and the vehicle's natural frequency, i.e., $\kappa_k = \omega_1 k^2 / \omega_v = k^2 \kappa_1$, where $\omega_1 = 2\pi f_1$, and f_1 is given by Table 6.2. Thus, if ω_v did not change, then the frequency ratios κ_k will stay the same as the previous examples. Now in case (a), we only increased the spatial frequency from 100 to 200. As a result, only the critical point of the plateau is being moved rightward, the rest of the system parameters stay the same. Based on our discussion, since the nondimensionalised spatial frequency ϖ in this case is larger than the frequency ratio of the 4th mode, i.e., $41.380 > 31.183$ (see Table 6.9), then we can identify the 4th mode modal parameters successfully.

Similarly, in case (b), when we change the vehicle speed, we only moved the tipping point of the plateau rightward, the rest of the system parameters stay the same. Since the new nondimensionalised spatial frequency is 55.608, which is larger than the 5th frequency ratio 48.7237 (see Table 6.9), we can identify all five modes properly.

Table 6.10 Nondimensionalised spatial frequency calculation breakdown

Figure label	Spatial frequency n (cycle/m)	Vehicle speed υ (m/s)	Vehicle natural frequency ω_v (rad/s)	Nondimensionalized spatial frequency $\varpi = 2\pi n \upsilon / \omega_v$	Modes that can be identified
(a)	200	13.333	10.256	41.380	1, 2, 3, 4
(b)	100	35.833	10.256	55.608	1, 2, 3, 4, 5
(c)	100	13.333	20.512	10.346	1, 2, 3
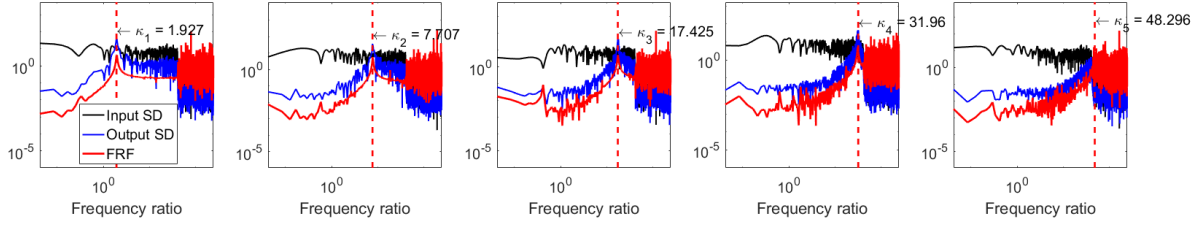
(a1) ($n = 200$ (cycle/m)) Noise-corrupted generalised input SD, generalised output SD and the Acceleance

FRF plots for the quarter car system with road roughness
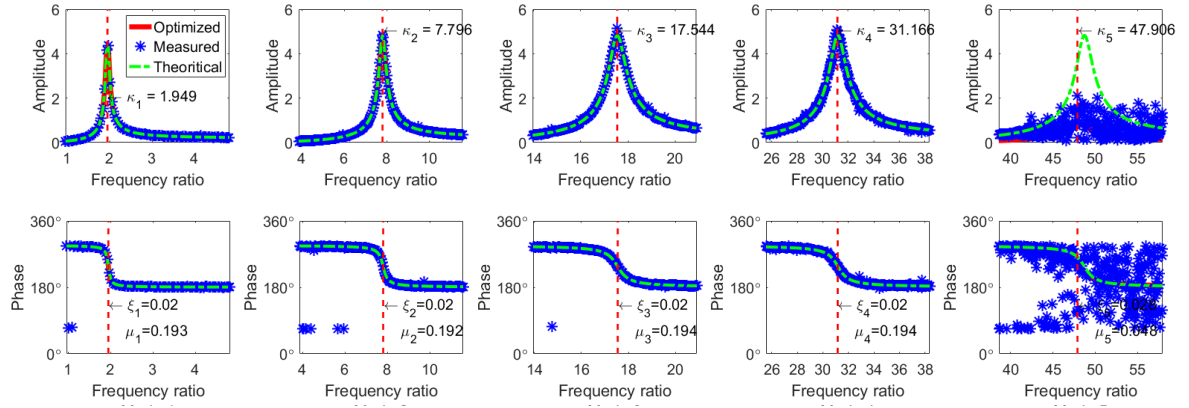
(a2) ($n = 200$ (cycle/m)) Noise-corrupted Acceleance FRF plots for the SDOF systems (linear scale) (1st row:

amplitude ratio between the input and output responses; 2nd row: phase angle between the input and output

responses). From left to rigth, mode 1 to mode 5.
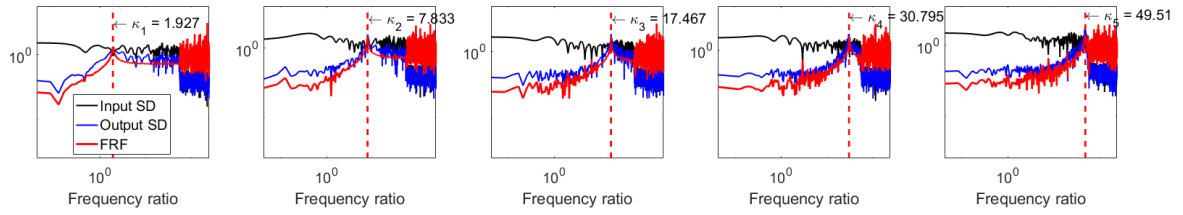
(b1) ($\upsilon = 35.8333$ (m/s)) Noise-corrupted generalised input SD, generalised output SD and the Acceleance

FRF plots for the quarter car system with road roughness
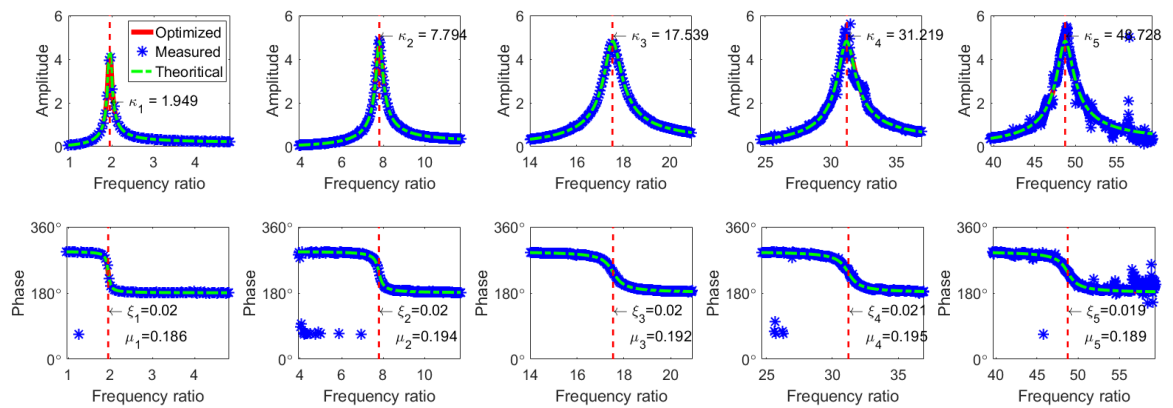
(b2) ($\upsilon = 35.8333$ (m/s)) Noise-corrupted Acceleance FRF plots for the SDOF systems (linear scale) (1st row:

amplitude ratio between the input and output responses; 2nd row: phase angle between the input and output

responses)
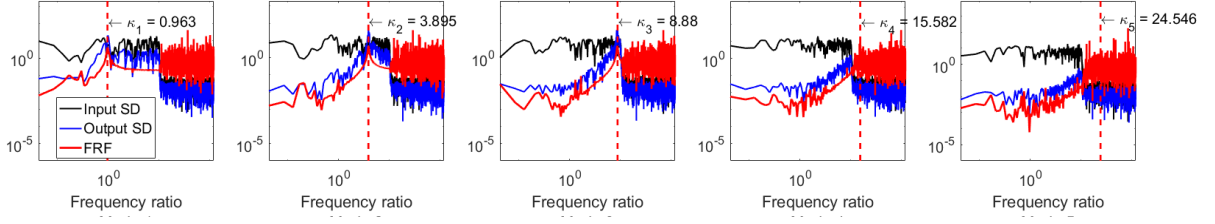
(c1) ($\omega_v = 10.3455$ (rad/s)) Noise-corrupted generalised input SD, generalised output SD and the Accelerance

FRF plots for the quarter car system with road roughness
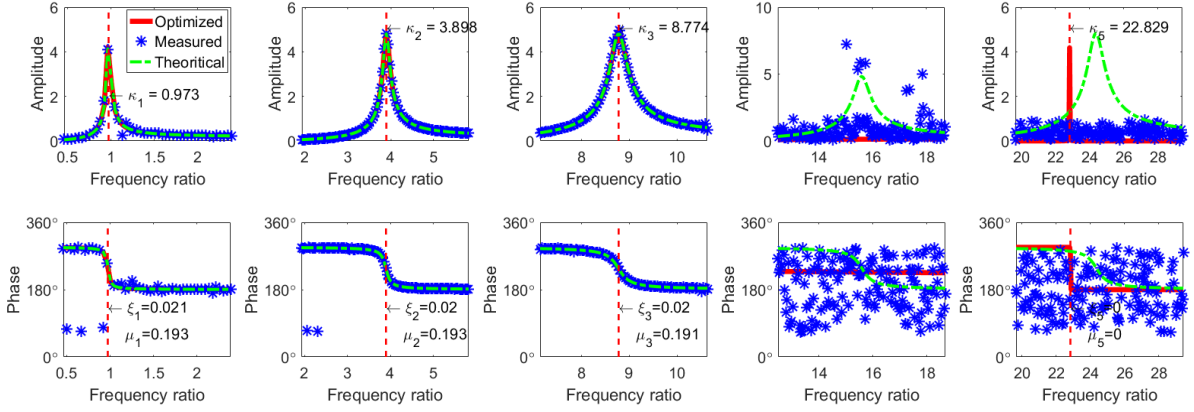
(c2) ($\omega_v = 10.3455$ (rad/s)) Noise-corrupted Accelerance FRF plots for the SDOF systems (linear scale) (1st row: amplitude ratio between the input and output responses; 2nd row: phase angle between the input and output responses)

Figure 6.22 The spectral density and the Accelerance FRF plots for the verification of influence of the nondimensionalised spatial frequency. From left to right, mode 1 to mode 5

While in case (c), as ω_v has changed, the frequency ratios κ_k will change as well. When we substitute the new vehicle frequency into the relation defined between the bridge's natural frequency and the vehicle's natural frequency, i.e., $\kappa_k = \omega_1 k^2 / \omega_v = k^2 \kappa_1$, we obtain a new set of frequency ratios κ_k, i.e., 0.9745, 3.8979, 8.7703, 15.5916 and 24.3619 for each mode, respectively. Therefore, compared to the ϖ calculated in Table 6.10, we can only identify the first three modes.

Noise level and mass ratios

For the effect of the noise level, we can easily observe it by comparing the noise-free and noise-contaminated identification results we discussed in the last section for the moving load system and the quarter car system. However, the effect of the noise is not as interesting as the influence factors we mentioned above. If white noise is assumed, then its effect is very straightforward and universal (i.e., all the modes will be affected at the same time). Therefore, we will not spend more time on explaining the noise effect.

As for the mass ratios between the two subsystems, from Eq. (6.41), we see that it only affects the amplitude of the Accelerance FRF. For low vehicle mass, the modal mass ratio $\mu_k = M_v/M_k$ for each mode will be small. The forcing and the response will be lower, as a result, the amplitude of the FRF will be small, then the noise will be more important.

Furthermore, the change of the vehicle mass will affect the vehicle natural frequency. In turn, it will change the frequency ratios κ_k between the two subsystems and the nondimensionalised spatial frequency. Also, if the mass of the bridge is changed, it will result in the change of the bridge modal mass for each mode. Then the bridge's natural frequencies change accordingly if the stiffness of the bridge remains the same. As a result, the frequency ratios κ_k will change. In this case, even if the spatial frequency remains the same, the identification result could be very different.

In short, anything that could change the configuration of the two subsystems, such as the bridge length, the stiffness of the bridge, and stiffness of the vehicle, are all influence factors, as they can affect the relative location of the tipping point defined by the nondimensionalised spatial frequency.

6.5 Conclusion

In this chapter, we verified the proposed two-stage method with a moving load system and the quarter car system. The identification results were discussed in terms of both noise-free and noise-contaminated cases for the moving load system. While for the quarter car system, we not only considered the noise-free and noise-corrupted cases but also included the road roughness impact in our consideration.

We found that the estimated mode shapes from the first stage, nondimensionalised spatial frequency, noise level, and the modal mass ratios between the two subsystems are the main influence factors for the bridge system identification via our proposed two-stage method. Among all these factors, the effect from the nondimensionalised spatial frequency is dominant. By comparing the nondimensionalised spatial frequency to the frequency ratios between the two subsystems in terms of the singular value plot, we can make inferences about the identification results. When the interested frequency ratios are smaller than the nondimensionalised spatial frequency, then we can successfully identify all the modal parameters we are interested in, namely, frequency ratios, damping ratio. Whereas the nondimensionalised spatial frequency is smaller than one of the frequency ratios, then all the modes with frequency ratios larger than the nondimensionalised spatial frequency cannot be identified.

In practice, when we do the modal analysis of a bridge via moving vehicle excitation, we can utilise the influence factors we discussed in this chapter to control the input spectrum, or more precisely the nondimensionalised spatial frequency, to excite the higher order modes. Such as by increasing the

vehicle speed to increase the value of the nondimensionalised spatial frequency. As the larger the nondimensionalised spatial frequency, the more modes can be identified.

7. Conclusion and prospects

7.1 Conclusions

We have accomplished essentially three main objectives in this thesis. The first two objectives are the main content in chapter 3 and chapter 4, respectively. As for the third objective, it is mainly discussed in chapter 5 and chapter 6.

We first proposed a concept of the real-valued spectral density matrix and proved its usefulness in estimating real mode shapes with better accuracy in chapter 3. Then we redefined the Frequency Domain Decomposition (FDD) technique and extended its application to a certain kind of nonstationary process in chapter 4. In the same chapter, we proposed a new technique by using a density operator from quantum mechanics to identify the close modes mode shapes. While in chapter 5, we proposed a two-stage system identification methodology to identify the bridge modal parameters, i.e., natural frequencies, damping ratios, modal masses, based on the measured moving force acceleration response and the bridge acceleration responses measured from several measurement locations on the bridge. And we verified this two-stage method in chapter 6 with simulated case studies.

Except for the essential work we have done from chapter 3 to chapter 6, we summarised some fundamental ideas in system identification. Continuously, our discussion starts from the general form of a time invariant linear differential equation with constant coefficients to the state-space models. We also covered the concepts of the transfer function and the frequency response function in the Laplace domain and the frequency domain, respectively. Discretely, we gave a brief discussion of some ARMA models and introduced the linear Gauss-Markov model, which is a basic model used in Kalman filtering. Apart from this, we overviewed some of the traditional identification techniques used in both EMA and OMA. Special attention was given to the Data-Driven Stochastic Subspace Identification (SSI-DATA) method and the Polyreference method. For the Polyreference method, a more rigorous theoretical background is proposed based on the modified version of the Yule-Walkers equation. While for the SSI-DATA method, we gave it a more accessible interpretation in terms of the state estimate equation used in the Kalman Filter process. Both the Polyreference method and the SSI-DATA method were used in a comparative case study in chapter 4 to do close mode analysis.

Since chapter 3 to chapter 6 contain most of our important conclusion, we will discuss them separately in the next few subsections.

7.1.1 Chapter 3: Real-valued and one-sided spectral density matrix

In this chapter, a real-valued one-sided spectral density matrix is proposed based on the Parseval's theorem in the hope that the absence of the imaginary part of a spectral density matrix will have no negative impact on the estimated modal parameters. This new concept is verified by conducting a comparative simulation case study with the classical complex-valued spectral density matrix.

In the case study, we considered a numerical model with five DOF's, where three modes are active, while the other two modes are inactive (i.e., by forcing the corresponding mode shapes to be zero). For the three active modes, two of them have fixed frequencies, while the other one is moving between them from the lower frequency to the higher frequency. The identification process for the two different spectral density matrices is realised by using the Polyreference method we introduce in chapter 2.

The simulation results clearly show that both approaches give the same level of accuracy for the measured frequencies and damping ratios. For the mode shapes, the error measured as the angles between the exact and the estimated mode shapes seem to be smaller than when we use the real-valued spectral density matrix. Upon the numerical analysis, we also provided a simple theoretical explanation for this new concept based on the Periodogram. Therefore, we concluded that we could remove the complex part of the spectral density matrix without losing quality in the identification process, provided that the mode shapes are real or close to real.

7.1.2 Chapter 4: On the theory of the FDD identification technique and the close modes estimation via a density operator

Chapter 4, without a doubt, is the most important chapter in this thesis, as we have achieved two main goals in this chapter. First, we redefined the FDD method and showed the possibility of using it to estimate modal parameters from some nonstationary processes. The reason why for doing this is because the original theoretical background of the FDD was weak and inconsistent with its application, which, as a result, largely restricted its application. As for the second achievement, we proposed a new way of identifying the mode shapes when two modes are very close to each other by evaluating a density operator. These seemly unrelated topics are related to a power spectral density (PSD) estimator, which is the essential objective the FDD works on. Therefore, the theoretical background given in this chapter was essentially revolving around a PSD estimator.

To be more specific, the theoretical background comes in two parts. In the first part, our goal is to define a PSD estimator, while in the second part, we aimed to understand the properties of that estimator and investigated the possible ways to analyse it.

Technically speaking, there are two ways to define a PSD estimator, one is derived by taking the Fourier transform of a sample correlation function is known as a Correlogram, the other one is estimated in terms of its original definition with a limit, and an expectation notion is known as a Periodogram. In this thesis, we adopted the second approach because the definition of a Periodogram can help us to interpret the PSD estimator as a covariance matrix, while the Correlogram cannot provide us with the same level of understanding. By reckoning a PSD estimator as a covariance matrix in the frequency domain, we can use the idea of PCA to explain the FDD as a variance-based method. Based on this, we explained the behaviour the meaning of singular values and pointed out that the FDD can be used to detect the close modes. However, when the two modes are very close to each other, it fails to identify the mode shapes of the two close modes. It is because its conventional procedure requires us to identify the mode shapes from the first singular vectors, which correspond to the peak values of the first singular value plot. When two modes fall into the spectral resolution, the two peaks are inseparable in the frequency domain.

To tackle the mode shape estimation problem for two close modes, we proposed a new way to analyse a PSD estimator by converting it into a density operator. In order to do this, we spent quite a lot of space discussing the properties of the PSD estimator and recognised it as a self-adjoint compact operator living in a finite-dimensional Hilbert space, which is complete. Hence, it bears a Spectral decomposition, and every vector in such a Hilbert space can be expressed by the linear combinations of the basis vectors. Upon which we explained the reason why we can use the FDD to estimate the mode shapes in a combination of the mode superposition. Finally, by normalising a self-adjoint PSD estimator, we obtained a density operator, which satisfies all the requirements given in quantum mechanics.

With a valid density operator, we can either use the purity or the Von Neumann entropy to estimate the close modes mode shapes. When the purity is close to 1, it means we are almost in a pure state. If it is less than 1, it implies that we are in a mixed state, while the Von Neumann entropy quantifies the departure of the system from a pure state for a given finite system. When the entropy is close to zero, it means the system is almost in a pure state, otherwise, it is in a mixed state. Under the classical mechanics setting, in a pure state means that we can assume the system is governed by a single mode, whereas a mixed state means that more than one mode is active. Both purity and Von Neumann entropy tell us how mixed the system states are. Thus, either of them can be used to estimate the mode shapes.

Apart from our discussion of the properties of a PSD estimator, we also explain the relationship between the Spectral theorem and the Singular Value Decomposition (SVD). And we pointed out that they are both widely used techniques in PCA.

Having redefined the FDD, we investigated the capability of the FDD to be used to deal with some nonstationary processes with a simple case study. This extended discussion of the nonstationary random process by using the FDD is unconventional, but it is a necessary preparation for its application in

chapter 6 for the vehicle induced bridge vibration analysis. There are many kinds of nonstationary processes. The application of a technique like the FDD built upon a stationary theory to a nonstationary process is quite limited. Fortunately, the nonstationary problem we encountered in chapter 6 is the one the FDD can solve.

As for the close mode analysis, the same dynamic system we used in chapter 3 was used again in this chapter to investigate the close mode identification problem. Since the natural frequencies and damping ratio cannot be obtained by analysing a density operator alone, a two-stage method, namely, Enhanced Polyreference, is designed to identify them. As its name implies, it is a combination of the Polyreference method. In the first stage, we analysed a density operator to get the mode shape estimation, then use it to decompose the normal coordinates into a set of modal coordinates; in the second stage, we applied the Polyreference method to those modal coordinates to extract the natural frequencies, and damping ratios. To evaluate the performance of our proposed method, a comparison study is conducted by applying the Polyreference method and the data-driven Stochastic Subspace Identification (SSI-DATA) method to the same data set.

Three possible influence factors, namely, the Modal Assurance Criteria (MAC) value between the two close modes, damping ratios, and the noise level, are considered in our study. According to those three factors, 8 cases, which are the different combinations of those factors, are investigated in our case study. The results show that the Enhanced Polyreference improves the estimation accuracy of the natural frequencies and the damping ratios in comparison with the Polyreference, while the SSI-DATA can provide the best damping ratios estimation. When it comes to the mode shape identification, Enhanced Polyreference uses a density operator in the first stage to estimate the mode shapes gives the most robust performance than the comparison methods. Although the mode shape estimation procedure proposed in this work does not work perfectly when two modes are extremely close or identical, when the disparity between the two close modes are slightly larger, it gives comparable identification precision of the mode shapes estimates when compared to Polyreference and SS-DATA. Most importantly, the proposed method can work in a frequency band where the FDD fails.

Particularly, our analysis showed that the best mode shape estimates are not necessarily given by the peak frequencies of the FDD. Rather, when two modes are very close to each other in the case that the first singular value plot of the FDD is unable to resolve it, the best mode shape estimates are corresponding to the frequency locations where a system approaches a pure state. This is a significant result, as it reveals the protentional bias in using the FDD to estimate the mode shapes and the natural frequencies.

It is true that different methods have their own merits. The FDD is good at detecting the close modes, but it is unable to resolve the mode shape and natural frequency identification problem when two modes are inseparable in the first singular value plot. Although good estimation results of the mode shape can

be obtained by analysing a density operator when two modes are very close but not extremely close or identical, the dependence on the peak picking of the purity plot will cripple it when a system has large damping ratios. Nevertheless, the proposed method is noise insensitive, and for well-separated modes, it can also be used to estimated natural frequencies, as the peak frequencies in a purity plot converge to their FDD counterparts. Apart from this, according to our discussion, different correlation relations between the two close modes will generally give a different view of the entropy or the purity plot. Such information can be used to make a reference to the orthogonality of mode shapes. Additionally, when used in combination with the Polyreference method, the physical pole determination problem associated with the original method is eliminated, which largely simplifies the identification procedure.

7.1.3 Chapter 5 and Chapter 6: Two-stage method and its verification

The last object of this thesis is presented in chapter 5 and chapter 6. In chapter 5, a two-stage system identification method was proposed to tackle the bridge modal parameters estimation problem under the excitation of a moving vehicle. The main idea is to use the estimated mode shapes we obtained from the first stage identification based on the bridge output only responses to decompose the bridge system in the second stage identification.

To be more specific, the mode shapes are identified in the first stage by applying an output only system identification technique to the bridge output-only responses, and the technique we used in this study is the FDD. Once we have the estimated mode shapes, we can use them to decouple the bridge responses and the vehicle response into modal coordinates. With the measured moving force vibrational information and the bridge acceleration response, a series of FRFs can be constructed. Then the modal parameters of the bridge can then be estimated through an optimisation procedure with the objective function formed in terms of the theoretical FRF expression, and this procedure forms the second stage identification. To utilise all the information and alleviate the smearing and leakage problem, the full-length record of the bridge was used in both stages in this work.

In general, the proposed two-stage method has the following merits. First and foremost, it can be used to identify the modal masses of the system, which is an important modal parameter in forced response analysis, and it can be used to scale the mode shapes. For such an important parameter, unfortunately, none of the current OMA techniques is able to identify it. Second, theoretically, it can produce a better natural frequency identification compared to the ones we obtained from the ambient vibration test, as the influence of the input is separated from the bridge system according to the FRF expression given by Eq. (5.58). More specifically, by separating the extra mass on the bridge, the estimated natural frequencies will be larger than its OMA counterparts but close to the actual values. Third, the experiment setup is simpler and cheaper than the traditional forced vibration test. This is helpful when one needs to conduct dynamic experiments on some structures such as a long-span bridge. Also, because

of the moving nature of the force, the excitation of the higher modes of the system can be easily achieved. In this case, as one of the benefits of using OMA, that all DOFs can be excited by assuming white noise input along the structure, will be dwarfed by the proposed two-stage method devised for the moving load case. Apart from this, due to the simplicity of the measurement procedure, we can run the vehicle on the bridge multiple times, which allows a statistical analysis of the results.

Finally, in chapter 6, the proposed two-stage method is verified with two nondimensionalised VBI systems, i.e., a moving load and the quarter car traversing a simply supported Euler-Bernoulli beam. The identification results are discussed in terms of both noise-free and noise-contaminated cases. We found that the estimated mode shapes from the first stage, nondimensionalised spatial frequency, noise level, and the modal mass ratios between the two subsystems are the main influence factors for the bridge system identification. Among all these factors, the effect from the nondimensionalised spatial frequency is the most important one.

7.2 Prospects

For future work, it can be wrapped into the following aspects:

- For the close mode analysis

  According to our analysis in chapter 4, we see that there is no problem for us to use the FDD or any other techniques to identify the modal parameters when the modes are well-separated. However, when two modes are very close to each other, and its frequency is inseparable under the analysing resolution, then FDD is unable to identify the close mode frequencies correctly. As a result, it cannot provide us proper estimates of the mode shapes for the two close modes. However, we saw in our analysis that the time domain method and the proposed purity or the Von Neumann entropy of a density operator measure are capable of estimating the mode shapes when two modes are very close to each other. When compared to the time domain methods, the proposed method via the density operator is more robust and user-friendly. Most importantly, it does not need any reference information. When combining it with the Polyreference method, we can easily automate the identification procedure.

  Nevertheless, none of the methods we studied in this thesis is able to deal with the identical mode case. We believe that the problem lies in three aspects. The first one is the identification objects. When the two modes are extremely close to each other or identical, they are interacting in the subspace. Especially when the two modes are identical, identify either of them individually is meaningless, as any two linear combinations of them are valid basis vectors for those two modes. Because of this, we need a technique to identify a subspace rather than the individual mode shape vectors. While in the case of two mode shapes, our goal is to identify a

hyperplane. So, the second problem is the identification technique. We could potentially use the first two singular vectors at the resonance frequency in the FDD. However, the two vectors only give us two random points in the vector space. Without any prior information about the mode shape vectors, we are unable to determine if those two points fall onto the hyperplane we are looking for. Then there comes the third problem, which is about the assessment measure. It is obvious that the conventional MAC, which works on individual vector, will not work in this case.

Now, our discussion of the FDD and the density operator provided us an easy way to analyse the behaviours of each mode frequency by frequency. It is possible that we can extend our work to tackle the identical mode estimation problem. Apart from this, the estimation uncertainty analysis has not been done in this study yet. As an important topic, we shall investigate it in future work.

- For the proposed two-stage system identification methodology

This thesis presents a preliminary work for the proposed method, which needs to be further verified with experiments in the lab or on a real bridge. Also, since the case study illustrated in chapter 6 only comprises a single point moving load and a quarter car model, it is unclear if the proposed method is still valid if a more supplicated car model is used. Thus, further numerical analysis is needed.

Besides, it must be pointed out that the proposed method is not restricted by the bridge equation of motion for the simply supported beam. It means that this two-stage method can be applied to other cases with a single moving force input.

References

[1]   F. Lanata and A. Del Grosso, "Damage detection and localization for continuous static monitoring of structures using a proper orthogonal decomposition of signals," *Smart Mater. Struct.*, vol. 15, no. 6, pp. 1811–1829, 2006.

[2]   J. T. Kim, Y. S. Ryu, H. M. Cho, and N. Stubbs, "Damage identification in beam-type structures: Frequency-based method vs mode-shape-based method," *Eng. Struct.*, vol. 25, no. 1, pp. 57–67, 2003.

[3]   J. P. Amezquita-Sanchez and H. Adeli, "Signal Processing Techniques for Vibration-Based Health Monitoring of Smart Structures," *Arch. Comput. Methods Eng.*, vol. 23, no. 1, pp. 1–15, 2016.

[4]   C. Neves, "Structural Health Monitoring of Bridges: Data-based damage detection method using Machine Learning," KTH Royal Institute of Technology, 2020.

[5]   C. C. Comisu, N. Taranu, G. Boaca, and M. C. Scutaru, "Structural health monitoring system of bridges," in *Procedia Engineering*, 2017, vol. 199, pp. 2054–2059.

[6]   O. S. Salawu and C. Williams, "Review of full-scale dynamic testing of bridge structures," *Eng. Struct.*, vol. 17, no. 2, pp. 113–121, 1995.

[7]   J. Sanchez and H. Benaroya, "Review of force reconstruction techniques," *J. Sound Vib.*, vol. 333, pp. 2999–3018, 2014.

[8]   E. J. OBrien, P. J. Mcgetrick, and A. González, "A drive-by inspection system via vehicle moving force identification," *Smart Struct. Syst.*, vol. 13, no. 5, pp. 821–848, 2014.

[9]   S. Q. Wu and S. S. Law, "Vehicle axle load identification on bridge deck with irregular road surface profile," *Eng. Struct.*, vol. 33, no. 2, pp. 591–601, 2011.

[10]   M. Tanio, Y. Hirata, and H. Suzuki, "Reconstruction of driving forces through recurrence plots," *Phys. Lett. Sect. A Gen. At. Solid State Phys.*, vol. 373, no. 23–24, pp. 2031–2040, 2009.

[11]   R. Lalthlamuana and S. Talukdar, "Obtaining vehicle parameters from bridge dynamic response: a combined semi-analytical and particle filtering approach," *J. Mod. Transp.*, vol. 23, no. 1, pp. 50–66, 2015.

[12]   S. S. Law, T. H. T. Chan, and Q. H. Zeng, "Moving force identification — a frequency and time

domains analysis," *J. Dyn. Syst. Meas. Control. Trans. ASME*, vol. 121, no. 3, pp. 394–401, 1999.

[13] A. González, E. J. Obrien, and P. J. McGetrick, "Identification of damping in a bridge using a moving instrumented vehicle," *J. Sound Vib.*, vol. 331, no. 18, pp. 4115–4131, 2012.

[14] Y. Chen, M. Q. Feng, and C.-A. Tan, "Modeling of Traffic Excitation for System Identification of Bridge Structures," *Comput. Civ. Infrastruct. Eng.*, vol. 21, no. 1, pp. 57–66, 2006.

[15] Y. Chen, M. Q. Feng, and C.-A. Tan, "Bridge Structural Condition Assessment Based on Vibration and Traffic Monitoring," *J. Eng. Mech.*, vol. 135, no. 8, pp. 747–758, 2009.

[16] Y. Yang, Y. Zhu, L. L. Wang, B. Y. Jia, and R. Jin, "Structural damage identification of bridges from passing test vehicles," *Sensors (Switzerland)*, vol. 18, no. 11, 2018.

[17] Z. H. Li and F. T. K. Au, "Damage detection of a continuous bridge from response of a moving vehicle," *Shock Vib.*, vol. 2014, no. 2, pp. 1–7, 2014.

[18] M. F. Green, "Dynamic response of short-span highway bridges to heavy vehicle loads," University of Cambridge, 1991.

[19] H. Wang, T. Nagayama, and D. Su, "Vehicle Parameter Identification through Particle Filter using Bridge Responses and Estimated Profile," in *Procedia Engineering*, 2017, vol. 188, pp. 64–71.

[20] P. J. Mcgetrick, C.-W. Kim, A. Gonzalez, and & Obrien, "Dynamic Axle Force and Road Profile Identification Using a Moving Vehicle," *Int. J. Archit. Eng. Constr.*, vol. 2, no. 1, pp. 1–16, 2013.

[21] W. Fauriat, C. Mattrand, N. Gayton, A. Beakou, and T. Cembrzynski, "Estimation of road profile variability from measured vehicle responses," *Veh. Syst. Dyn.*, vol. 54, no. 5, pp. 585–605, 2016.

[22] E. P. Carden and P. Fanning, "Vibration Based Condition Monitoring: A Review," *Struct. Heal. Monit. An Int. J.*, vol. 3, no. 4, pp. 355–377, 2004.

[23] X. Q. Zhu and S. S. Law, "Recent developments in inverse problems of vehicle–bridge interaction dynamics," *J. Civ. Struct. Heal. Monit.*, vol. 6, no. 1, pp. 107–128, 2016.

[24] Q. Zhang, Ł. Jankowski, and Z. Duan, "Simultaneous identification of moving masses and structural damage," *Struct. Multidiscip. Optim.*, vol. 42, no. 6, pp. 907–922, 2010.

[25] Q. Zhang, Ł. Jankowski, and Z. Duan, "Simultaneous identification of moving vehicles and bridge damages considering road rough surface," *Math. Probl. Eng.*, vol. 2013, 2013.

[26] X. Q. Zhu and S. S. Law, "Damage Detection in Simply Supported Concrete Bridge Structure Under Moving Vehicular Loads," *J. Vib. Acoust.*, vol. 129, no. 1, p. 58, 2007.

[27] D. Feng, H. Sun, and M. Q. Feng, "Simultaneous identification of bridge structural parameters

and vehicle loads," *Comput. Struct.*, vol. 157, pp. 76–88, 2015.

[28]   X. Q. Zhu, "Recent developments in inverse problems of vehicle–bridge interaction dynamics," *J. Civ. Struct. Heal. Monit.*, vol. 6, 2016.

[29]   Y. B. Yang, C. W. Lin, and J. D. Yau, "Extracting bridge frequencies from the dynamic response of a passing vehicle," *J. Sound Vib.*, vol. 272, no. 3–5, pp. 471–493, 2004.

[30]   Y. B. Yang and C. W. Lin, "Vehicle-bridge interaction dynamics and potential applications," *J. Sound Vib.*, vol. 284, no. 1–2, pp. 205–226, 2005.

[31]   Y.-B. Yang, Z. Yao, and Y. S. Wu, *Vehicle-bridge interaction dynamics : with applications to high-speed railways*. World Scientific, 2004.

[32]   P. J. McGetrick, A. González, and E. J. OBrien, "Theoretical investigation of the use of a moving vehicle to identify bridge dynamic parameters," *Insight Non-Destructive Test. Cond. Monit.*, vol. 51, no. 8, pp. 433–438, 2009.

[33]   A. Miyamoto and A. Yabe, "Development of practical health monitoring system for short- and medium-span bridges based on vibration responses of city bus," *J. Civ. Struct. Heal. Monit.*, vol. 2, no. 1, pp. 47–63, 2012.

[34]   E. J. O'Brien and A. Malekjafarian, "On the use of a passing vehicle for bridge health monitoring," in *8th International Conference on Bridge Maintenance, Safety & Management*, 2016.

[35]   A. Malekjafarian, P. J. McGetrick, and E. J. Obrien, "A review of indirect bridge monitoring using passing vehicles," *Shock Vib.*, vol. 2015, 2015.

[36]   C. W. Lin and Y. B. Yang, "Use of a passing vehicle to scan the fundamental bridge frequencies: An experimental verification," *Eng. Struct.*, vol. 27, no. 13, pp. 1865–1878, 2005.

[37]   Y. B. Yang and K. C. Chang, "Extraction of bridge frequencies from the dynamic response of a passing vehicle enhanced by the EMD technique," *J. Sound Vib.*, vol. 322, no. 4–5, pp. 718–739, 2009.

[38]   D. M. Siringoringo and Y. Fujino, "Estimating Bridge Fundamental Frequency from Vibration Response of Instrumented Passing Vehicle: Analytical and Experimental Study," *Adv. Struct. Eng.*, vol. 15, no. 3, pp. 417–433, 2012.

[39]   Y. B. Yang, Y. C. Li, and K. C. Chang, "Constructing the mode shapes of a bridge from a passing vehicle: A theoretical study," *Smart Struct. Syst.*, vol. 13, no. 5, pp. 797–819, 2014.

[40]   A. Malekjafarian and E. J. OBrien, "Identification of bridge mode shapes using Short Time Frequency Domain Decomposition of the responses measured in a passing vehicle," *Eng. Struct.*, vol. 81, pp. 386–397, 2014.

Reference

[41]   Y.-B. Yang, C. W. Lin, and J. D. Yau, "Extracting bridge frequencies from the dynamic response of a passing vehicle," *J. Sound Vib.*, vol. 272, pp. 471–493, 2004.

[42]   Y. Zhang, S. T. Lie, and Z. Xiang, "Damage detection method based on operating deflection shape curvature extracted from dynamic response of a passing vehicle," *Mech. Syst. Signal Process.*, vol. 35, no. 1–2, pp. 238–254, 2013.

[43]   Y. Oshima, K. Yamamoto, and K. Sugiura, "Damage assessment of a bridge based on mode shapes estimated by responses of passing vehicles," *Smart Struct. Syst.*, vol. 13, no. 5, pp. 731–753, 2014.

[44]   J. Kim, J. P. Lynch, J.-J. Lee, and C.-G. Lee, "Truck-based mobile wireless sensor networks for the experimental observation of vehicle–bridge interaction," *Smart Mater. Struct.*, vol. 20, no. 6, 2011.

[45]   J. Oskarbski and D. Kaszubowski, "Implementation of Weigh-in-Motion System in Freight Traffic Management in Urban Areas," in *Transportation Research Procedia*, 2016, vol. 16, pp. 449–463.

[46]   Y. Yu, C. Cai, and L. Deng, "State-of-the-art review on bridge weigh-in-motion technology," *Adv. Struct. Eng.*, vol. 19, no. 9, pp. 1514–1530, 2016.

[47]   M. Atluri, M. Chowdhury, N. Kanhere, R. Fries, W. Sarasua, and J. Ogle, "Development of a sensor system for traffic data collection," *J. Adv. Transp.*, vol. 43, no. 1, pp. 1–20, 2009.

[48]   V. P. Gil Jiménez and M. J. Fernández-Getino García, "Simple design of wireless sensor networks for traffic jams avoidance," *J. Sensors*, vol. 2015, 2015.

[49]   Y. Chen, C.-A. Tan, M. Q. Feng, and Y. Fukuda, "A video assisted approach for structural health monitoring of highway bridges under normal traffic," *Struct. Mater.*, vol. 6174, no. IV, pp. 1–18, 2006.

[50]   M. Fraser, A. Elgamal, X. He, and J. P. Conte, "Sensor Network for Structural Health Monitoring of a Highway Bridge," *J. Comput. Civ. Eng.*, vol. 24, no. 1, pp. 11–24, 2010.

[51]   J. Kim and J. P. Lynch, "Experimental analysis of vehiclebridge interaction using a wireless monitoring system and a two-stage system identification technique," *Mech. Syst. Signal Process.*, vol. 28, pp. 3–19, 2012.

[52]   R. Brincker, L. Zhang, and P. Andersen, "Modal identification from ambient responses using frequency domain decomposition," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 4062. 2000.

[53]   R. Brincker, C. E. Ventura, and P. Andersen, "Damping estimation by frequency domain decomposition," *Time*, vol. 1. pp. 698–703, 2001.

[54]   L. Ljung, *System identification : theory for the user*. Pearson Prentice Hall, 1999.

[55]   K. P. Murphy, *Machine learning : a probabilistic perspective*. MIT Press, 2012.

[56]   S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*. Cambridge University Press, 2013.

[57]   L. Zhang, Y. Yao, and M. Lu, "An improved time domain polyreference method for modal identification," *Mech. Syst. Signal Process.*, vol. 1, no. 4, pp. 399–413, 1987.

[58]   B. Peeters and G. De Roeck, "Reference-Based Stochastic Subspace Identification for Output-Only Modal Analysis," *Mech. Syst. Signal Process.*, vol. 13, no. 6, pp. 855–878, 1999.

[59]   L. Fu and P. Li, "The Research Survey of System Identification Method," in *2013 5th International Conference on Intelligent Human-Machine Systems and Cybernetics*, 2013, pp. 397–401.

[60]   Ding Feng, "System identification-Part A: Introduction to the identification," *J. Nanjing Univ. Inf. Sci. Technol. Nat. Sci. Ed.*, vol. 3, pp. 1–22, 2011.

[61]   D. Ewins, *Modal Testing: Theory, Practice, And Application*. Wiley, 2000.

[62]   Lennart Ljung, *System identification: Theory for the User*. Pearson Prentice Hall, 1999.

[63]   R. Brincker and C. E. Ventura, *Introduction to Operational Modal Analysis*. Chichester, UK: John Wiley & Sons, Ltd, 2015.

[64]   M. M. Maia and M. . N. Silva, *Theoretical and Experimental Modal Analysis*. Research Studies Press, 1998.

[65]   S. Smith, *Digital Signal processing: A practical guide for engineers and scientists*, vol. 1. Newnes, 2003.

[66]   J. Nocedal and S. J. Wright, *Numerical optimization*. Springer, 1999.

[67]   L. Fu and P. Li, "The Research Survey of System Identification Method," in *2013 5th International Conference on Intelligent Human-Machine Systems and Cybernetics*, 2013, pp. 397–401.

[68]   G. Weipeng, H. Qiwei, and Y. Zhengtao, "A Survey on Method of System Identification," in *International Conference on Mechanics Design, Manufacturing and Automation*, 2016.

[69]   J. McCall, "Genetic algorithms for modelling and optimisation," *J. Comput. Appl. Math.*, vol. 184, no. 1, pp. 205–222, Dec. 2005.

[70]   K. Sastry, D. Goldberg, and G. Kendall, "Genetic Algorithms," in *Search Methodologies*, Boston, MA: Springer US, 2005, pp. 97–125.

Reference

[71] C. W. Entemann, "Fuzzy Logic: Misconceptions and Clarifications," *Artif. Intell. Rev.*, vol. 17, no. 1, pp. 65–84, 2002.

[72] D. H. Hodges and G. A. Pierce, *Introduction to structural dynamics and aeroelasticity, second edition*. Cambridge University Press, 2011.

[73] D. J. Ewins, *Modal testing: theory, practice, and application*. Wiley, 2000.

[74] M. Richardson and B. Schwarz, "Modal Parameter Estimation from Operating Data," *Sound Vib.*, 2003.

[75] V. Slivinskas and V. Šimonyte, "On the Foundation of Prony's Method," *IFAC Proc. Vol.*, vol. 19, no. 5, pp. 121–126, 1986.

[76] D. R. J. Allemang and D. D. L. Brown, "Experimental Modal Analysis and Dynamic Component Synthesis VOL III - Modal Parameter Estimation," 1987.

[77] W. Zhou and D. Chelidze, "Generalized Eigenvalue Decomposition in Time Domain Modal Parameter Identification," *J. Vib. Acoust.*, vol. 130, 2008.

[78] G. H. J. Ill, T. G. Carrie, and J. P. Lauffer, "The Natural Excitation Technique (NExT) for Modal Parameter Extraction From Operating Wind Turbines," 1993.

[79] J.M. Caicedo, "Practical Guidelines for the Natural Excitation Technique (NExT) and the Eigensystem Realization Algorithm (ERA) for Modal Identification Using Ambient Vibration," *Dyn. Test. Civ. Eng. Struct. Ser.*, 2011.

[80] J.-N. Juang and R. S. Pappa, "An eigensystem realization algorithm for modal parameter identification and model reduction," *J. Guid. Control. Dyn.*, vol. 8, no. 5, pp. 620–627, 1985.

[81] M. Chang and S. N. Pakzad, "Modified Natural Excitation Technique for Stochastic Modal Identification," *J. Struct. Eng.*, vol. 139, no. 10, pp. 1753–1762, 2013.

[82] J. M. Caicedo, S. J. Dyke, and E. A. Johnson, "Natural Excitation Technique and Eigensystem Realization Algorithm for Phase I of the IASC-ASCE Benchmark Problem: Simulated Data," *J. Eng. Mech.*, vol. 130, no. 1, pp. 49–60, 2004.

[83] B. Peeters and G. De Roeck, "Stochastic System Identification for Operational Modal Analysis: A Review," *J. Dyn. Syst. Meas. Control*, vol. 123, no. 4, pp. 659–667, 2001.

[84] E. Reynders, K. Maes, G. Lombaert, and G. De Roeck, "Uncertainty quantification in operational modal analysis with stochastic subspace identification: Validation and applications," *Mech. Syst. Signal Process.*, vol. 66–67, pp. 13–30, 2016.

[85] B. Peeters and G. De Roeck, "Reference-Based Stochastic Subspace Identification for Output-Only Modal Analysis," *Mech. Syst. Signal Process.*, vol. 13, no. 6, pp. 855–878, 1999.

[86]   L. Zhang, "From Traditional Experimental Modal Analysis (EMA) to Operational Modal Analysis (OMA): an Overview," in *The 5th International Operational Modal Analysis Conference*, 2013.

[87]   A. Bajric, C. T. Georgakis, and R. Brincker, "Evaluation of damping using time domain OMA techniques," *Proc. 2014 SEM Fall Conf. Int. Symp. Intensive Load. Its Eff.*, pp. 1–11, 2014.

[88]   P. Mohanty and D. J. Rixen, "A modified Ibrahim time domain algorithm for operational modal analysis including harmonic excitation," *J. Sound Vib.*, vol. 275, no. 1–2, pp. 375–390, 2004.

[89]   N.-J. Jacobsen, P. Andersen, and R. Brincker, "Using Enhanced Frequency Domain Decomposition as a Robust Technique to Harmonic Excitation in Operational Modal Analysis," in *Proceedings of ISMA2006 : International Conference on Noise & Vibration Engineering*, 2006.

[90]   S. Gade, N. B. Møller, H. Herlufsen, and H. Konstantin-Hansen, "Frequency Domain Techniques for Operational Modal Analysis," in *The First International Operational Modal Analysis Conference*, 2005, pp. 261–271.

[91]   L. Zhang, T. Wang, and Y. Tamura, "A frequency-spatial domain decomposition (FSDD) method for operational modal analysis," *Mech. Syst. Signal Process.*, vol. 24, no. 5, pp. 1227–1239, 2010.

[92]   B. H. Kim, N. Stubbs, and T. Park, "A new method to extract modal parameters using output-only responses," *J. Sound Vib.*, vol. 282, no. 1–2, pp. 215–230, 2005.

[93]   B. H. Kim, J. Lee, and D. H. Lee, "Extracting modal parameters of high-speed railway bridge using the TDD technique," *Mech. Syst. Signal Process.*, vol. 24, no. 3, pp. 707–720, Apr. 2010.

[94]   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*, 2005, pp. 179–190.

[95]   R. Brincker, "Some Elements of Operational Modal Analysis," *Shock Vib.*, vol. 2014, pp. 1–11, 2014.

[96]   M. Batel, B. & Kjaer, and G. Norcross, "Operational Modal Analysis – Another Way of Doing Modal Testing," *Sound Vib.*, pp. 22–27, 2002.

[97]   C. Rainieri and G. Fabbrocino, *Operational modal analysis of civil engineering structures : an introduction and guide for applications*, Illustrate. Springe, 2014.

[98]   L. Brand, "The Companion Matrix and Its Properties," *Am. Math. Mon.*, vol. 71, no. 6, p. 629, 1964.

[99]   I. Lankham, B. Nachtergaele, and A. Schilling, "The Fundamental Theorem of Algebra." 2007.

Reference

[100] J. G. Proakis and D. G. Manolakis, *Digital signal processing*. Pearson Prentice Hall, 2007.

[101] K. H. Lundberg, H. R. Miller, and D. L. Trumper, "Initial conditions, generalized functions, and the laplace transform troubles at the origin," *IEEE Control Syst.*, vol. 27, no. 1, pp. 22–35, 2007.

[102] R. H. Shumway and D. S. Stoffer, *Time series analysis and its applications : with R examples*. Springer, 2017.

[103] M. Phan, J.-N. Juang, and R. W. Longman, "On Markov Parameters in System Identification," 1991.

[104] Z. Chen and E. Brown, "State space model," *Scholarpedia*, vol. 8, no. 3. Scholarpedia, 2013.

[105] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Dover, 2005.

[106] Anthony O'Hagan, "The Bayesian Approach to Stochastics," in *Probability Theory in Research Methodology*, .

[107] H. Vold, J. Kundrat, G. T. Rocklin, and R. Russell, "A Multi-Input Modal Estimation Algorithm for Mini-Computers," *SAE Transactions*, vol. 91. SAE International, pp. 815–821, 1982.

[108] B. De Moor, P. Van Overschee, and W. Favoreel, "Algorithms for Subspace State-Space System Identification: An Overview," in *Applied and Computational Control, Signals, and Circuits*, Birkhäuser Boston, 1999, pp. 247–311.

[109] H. Vold, J. Kundrat, G. T. Rocklin, and R. Russell, "A Multi-Input Modal Estimation Algorithm for Mini-Computers," *SAE Transactions*, vol. 91. SAE International, pp. 815–821, 1982.

[110] M. Pastor, M. Binda, and T. Harþarik, "Modal Assurance Criterion," *Procedia Eng.*, vol. 48, pp. 543–548, 2012.

[111] H. C. Tsai, "Modal superposition method for dynamic analysis of structures excited by prescribed support displacements," *Comput. Struct.*, vol. 66, no. 5, pp. 675–683, 1998.

[112] M. Johansson, "The Hilbert transform," Växjö University, 2012.

[113] R. Brincker and M. Lopez-Aenlle, "Mode shape sensitivity of two closely spaced eigenvalues," *J. Sound Vib.*, vol. 334, pp. 377–387, 2015.

[114] S.-K. Au, *Operational modal analysis : modeling, bayesian inference, uncertainty laws*. Springer, 2017.

[115] R. Brincker, L. Zhang, and P. Andersen, "Modal identification of output-only systems using frequency domain decomposition," *Smart Mater. Struct.*, vol. 10, no. 3, pp. 441–445, 2001.

[116] R. Brincker and P. Andersen, "Method for vibration analysis," United States patent no. US 6,779,404 B1, 03-Nov-2004.

[117] D. J. Griffiths and D. F. Schroeter, *Introduction to Quantum Mechanics*. Cambridge University Press, 2018.

[118] S. Axler, *Measure, Integration & Real Analysis*. Springer, 2020.

[119] Billingsley Patrick, *Convergence of Probability Measures*, 2nd Editio. Wiley, 1999.

[120] O. Knill, *Probability and Stochastic Processes with Applications*. Overseas Press India Private Limited, 2009.

[121] M. B. Priestley, *Spectral Analysis and Time Series*. Academic Press, 1982.

[122] Chatfield Chris, *The Analysis of Time Series: An Introduction*, Sixth Edit. Routledge, 2003.

[123] S. L. Miller and D. Childers, *Probability and Random Processes: With Applications to Signal Processing and Communications*, Second Edition. Elsevier Inc., 2012.

[124] J. S. Bendat and A. G. Piersol, *Random data : analysis and measurement procedures*. Wiley, 2000.

[125] A. V Oppenheim and G. C. Verghese, *Signals, Systems and Inference*. Pearson Prentice Hall, 2010.

[126] P. Stoica and R. L. Moses, *Spectral analysis of signals*. Pearson/Prentice Hall, 2005.

[127] I. E. Papadakis and A. Lawrence, "Improved methods for power spectrum modelling of red noise," *R. Astron. Soc.* , pp. 612–624, 1993.

[128] A. Labuda, "Daniell method for power spectral density estimation in atomic force microscopy," *Rev. Sci. Instrum.*, vol. 87, no. 3, 2016.

[129] B. L. Edge and P. C. Liu, "Comparing Power Spectra Computed by Blackman-Tukey and Fast Fourier Transform," *Water Resour. Res.*, vol. 6, no. 6, pp. 1601–1610, 1970.

[130] S. Axler, *Linear Algebra Done Right*. Springer, 2015.

[131] C. Heil, *Metrics, Norms, Inner Products, and Operator Theory*. Springer, 2018.

[132] G. Strang, *Linear algebra and its applications*. Thomson, Brooks/Cole, 2006.

[133] G. Livan, M. Novaes, and P. Vivo, *Introduction to Random Matrices Theory and Practice*. Springer, 2018.

[134] John A. Gubner, *Probability and Random Processes for Electrical and Computer Engineers*. Cambridge University Press, 2006.

[135] H. Cho, D. Venturi, and G. E. Karniadakis, "Karhunen-Loève expansion for multi-correlated stochastic processes," *Probabilistic Eng. Mech.*, vol. 34, pp. 157–167, 2013.

[136] A. Alexanderian, "A brief note on the Karhunen-Loeve expansion," *arXiv*, vol. arXiv:1509,

2015.

[137] I. T. Jollife and J. Cadima, "Principal component analysis: A review and recent developments," *R. Soc.*, vol. 374, no. 2065, 2016.

[138] R. M. Bell and Y. Koren, "Scalable collaborative filtering with jointly derived neighborhood interpolation weights," in *Proceedings - IEEE International Conference on Data Mining, ICDM*, 2007, pp. 43–52.

[139] D. Garber and E. Hazan, "Fast and Simple PCA via Convex Optimization," *arXiv*, vol. arXiv:1509, Sep. 2015.

[140] P. E. Smaldino, "Measures of individual uncertainty for ecological models: Variance and entropy," *Ecol. Modell.*, vol. 254, pp. 50–53, 2013.

[141] R. W. Clough and J. Penzien, *Dynamics of structures*. McGraw-Hill, 1993.

[142] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, 2010.

[143] J. Von Neumann, *Mathematical foundations of quantum mechanics* . Princeton University Press, 2018.

[144] G. Borot and C. Nadal, "Purity distribution for generalized random Bures mixed states," *J. Phys. A Math. Theor.*, vol. 45, no. 7, p. 43, 2012.

[145] C. E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. Tech. J.*, vol. 27, pp. 623–656, 1948.

[146] R. Dahlhaus, "Fitting time series models to nonstationary processes," *Ann. Stat.*, vol. 25, no. 1, pp. 1–37, 1997.

[147] Leon Cohen, "Time-Frequency Distribution --A Review," *Proc. IEEE*, vol. 77, pp. 941–981, 1989.

[148] D. Gabor, "Theory of communication. Part 1: The analysis of information," *J. Inst. Electr. Eng. - Part III Radio Commun. Eng.*, vol. 93, no. 26, pp. 429–441, 1946.

[149] C. H. Page, "Instantaneous power spectra," *J. Appl. Phys.*, vol. 23, no. 1, pp. 103–106, 1952.

[150] W. Martin and P. Flandrin, "Wigner-Ville Spectral Analysis of Nonstationary Processes," *IEEE Trans. Acoust.*, vol. 33, no. 6, pp. 1461–1470, 1985.

[151] M. B. Priestley, "Evolutionary Spectra and Non-Stationary Processes," *J. R. Stat. Soc. Ser. B*, vol. 27, no. 2, pp. 204–229, 1965.

[152] B. Jawerth and W. Sweldens, "Overview of wavelet based multiresolution analyses," *SIAM Rev.*,

vol. 36, no. 3, pp. 377–412, 1994.

[153] Y. F. Meyer, "Wavelets and Applications," in *Proceedings of the International Congress of Mathematicians*, 1990, pp. 1619–1626.

[154] I. Daubechies, *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, 1992.

[155] N. E. Huang and Z. Wu, "A review on Hilbert-Huang transform: Method and its applications to geophysical studies," *Reviews of Geophysics*, vol. 46, no. 2. 2008.

[156] I. Holmström, "Analysis of time series by means of empirical orthogonal functions," *Tellus*, vol. 22, no. 6, pp. 638–647, Dec. 1970.

[157] A. H. Monahan, J. C. Fyfe, M. H. P. Ambaum, D. B. Stephenson, and G. R. North, "Empirical orthogonal functions: The medium is the message," *J. Clim.*, vol. 22, no. 24, pp. 6501–6514, 2009.

[158] A. A. Adebiyi, A. O. Adewumi, and C. K. Ayo, "Comparison of ARIMA and artificial neural networks models for stock price prediction," *J. Appl. Math.*, vol. 2014, 2014.

[159] M. J. Cassidy and W. D. Penny, "Bayesian Nonstationary Autoregressive Models for Biomedical Signal Analysis," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 10, 2002.

[160] R. Dahlhaus, "Locally Stationary Processes," in *Handbook of Statistics*, vol. 30, Elsevier B.V., 2012, pp. 351–413.

[161] A. Papoulist, "Levinson's Algorithm, Wold's Decomposition, and Spectral Estimation," *Soc. Ind. Appl. Math.*, vol. 27, no. 3, pp. 405–441, 1985.

[162] R. M. Lin and M. K. Lim, "Modal analysis of close modes using perturbative sensitivity approach," *Eng. Struct.*, vol. 19, no. 6, pp. 397–406, 1997.

[163] Y. C. Zhu, S. K. Au, and S. Jones, "Identification uncertainty of close modes in operational modal analysis," in *12th International Conference on Applications of Statistics and Probability in Civil Engineering*, 2015.

[164] S. K. Au and J. M. W. Brownjohn, "Asymptotic identification uncertainty of close modes in Bayesian operational modal analysis," *Mech. Syst. Signal Process.*, vol. 133, 2019.

[165] S. K. Au, J. M. W. Brownjohn, B. Li, and A. Raby, "Understanding and managing identification uncertainty of close modes in operational modal analysis," *Mech. Syst. Signal Process.*, vol. 147, 2021.

[166] S. K. Au, B. Li, and J. M. W. Brownjohn, "Achievable precision of close modes in operational modal analysis: Wide band theory," *Mech. Syst. Signal Process.*, vol. 147, 2021.

[167] R. Brincker, A. Skafte, M. López-Aenlle, A. Sestieri, W. D'Ambrogio, and A. Canteli, "A local

correspondence principle for mode shapes in structural dynamics," *Mech. Syst. Signal Process.*, vol. 45, no. 1, pp. 91–104, 2014.

[168] J. E. Mottershead, C. Mares, and S. James, "Fictitious modifications for the separation of close modes," *Mech. Syst. Signal Process.*, vol. 16, no. 5, pp. 741–755, 2002.

[169] W.-H. Wu, S.-W. Wang, C.-C. Chen, and G. Lai, "Modal parameter identification for closely spaced modes of civil structures based on an upgraded stochastic subspace methodology," *Struct. Infrastruct. Eng.*, vol. 15, no. 3, pp. 296–313, 2019.

[170] J. Chen, "Application of empirical mode decomposition in structural health monitoring: Some experience," *Adv. Adapt. Data Anal.*, vol. 1, no. 4, pp. 601–621, 2009.

[171] A. Srikantha Phani and J. Woodhouse, "Viscous damping identification in linear vibration," *J. Sound Vib.*, vol. 303, no. 3–5, pp. 475–500, 2007.

[172] C. Rainieri and G. Fabbrocino, "Performance assessment of selected OMA techniques for dynamic identification of geotechnical systems and closely spaced structural modes," *J. Theor. Appl. Mech.*, vol. 49, no. 3, pp. 825–839, 2011.

[173] S. A. Hosseini Kordkheili, S. H. Momeni Massouleh, S. Hajirezayi, and H. Bahai, "Experimental identification of closely spaced modes using NExT-ERA," *J. Sound Vib.*, vol. 412, pp. 116–129, 2018.

[174] B. Yan and A. Miyamoto, "A comparative study of modal parameter identification based on wavelet and Hilbert-Huang transforms," *Comput. Civ. Infrastruct. Eng.*, vol. 21, no. 1, pp. 9–23, 2006.

[175] J. Bin Tan, Y. Liu, L. Wang, and W. G. Yang, "Identification of modal parameters of a system with high damping and closely spaced modes by combining continuous wavelet transform with pattern search," *Mech. Syst. Signal Process.*, vol. 22, no. 5, pp. 1055–1060, 2008.

[176] C. Bao, H. Hao, Z. X. Li, and X. Zhu, "Time-varying system identification using a newly improved HHT algorithm," *Comput. Struct.*, vol. 87, no. 23–24, pp. 1611–1623, 2009.

[177] S. Li, J. W. Pan, G. H. Luo, and J. T. Wang, "Improvements in the HHT for the Modal Parameter Identification of Structures with Closely Spaced Modes," *J. Earthq. Eng.*, pp. 1–26, 2019.

[178] Y. S. Kim and L. Q. Chen, "Separation of closely spaced modes by combining complex envelope displacement analysis with method of generating intrinsic mode functions through filtering algorithm based on wavelet packet decomposition," *Appl. Math. Mech. (English Ed.*, vol. 34, no. 7, pp. 801–810, 2013.

[179] Z. Wang and G. Chen, "Analytical mode decomposition with Hilbert transform for modal parameter identification of buildings under ambient vibration," *Eng. Struct.*, vol. 59, pp. 173–

184, 2014.

[180] Z. C. Wang, D. Zhang, W. X. Ren, and G. Da Chen, "Structural modal parameter identification from forced vibration with analytical mode decomposition," *Adv. Struct. Eng.*, vol. 17, no. 8, pp. 1129–1143, 2014.

[181] S. I. McNeill and D. C. Zimmerman, "A framework for blind modal identification using joint approximate diagonalization," *Mech. Syst. Signal Process.*, vol. 22, no. 7, pp. 1526–1548, 2008.

[182] Y. Yang and S. Nagarajaiah, "Blind modal identification of output-only structures in time-domain based on complexity pursuit," *Earthq. Eng. Struct. Dyn.*, vol. 42, no. 13, pp. 1885–1905, 2013.

[183] Y. Yang and S. Nagarajaiah, "Output-only modal identification with limited sensors using sparse component analysis," *J. Sound Vib.*, vol. 332, no. 19, pp. 4741–4765, 2013.

[184] Y. Guo and A. Kareem, "System identification through nonstationary data using Time-Frequency Blind Source Separation," *J. Sound Vib.*, vol. 371, pp. 110–131, 2016.

[185] F. C. A. Fernandas, R. L. C. Van Spaendonck, and C. S. Burrus, "A new framework for complex wavelet transforms," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1825–1837, 2003.

[186] R. J. Allemang, "The modal assurance criterion - Twenty years of use and abuse," *Sound Vib.*, vol. 37, no. 8, pp. 14–21, 2003.

[187] E. Reynders, J. Houbrechts, and G. De Roeck, "Fully automated (operational) modal analysis," *Mech. Syst. Signal Process.*, vol. 29, pp. 228–250, 2012.

[188] S. K. Au, "Fast Bayesian ambient modal identification in the frequency domain, Part I: Posterior most probable value," *Mech. Syst. Signal Process.*, vol. 26, no. 1, pp. 60–75, 2012.

[189] S. K. Au, "Fast Bayesian ambient modal identification in the frequency domain, Part II: Posterior uncertainty," *Mech. Syst. Signal Process.*, vol. 26, no. 1, pp. 76–90, 2012.

[190] W. D'Ambrogio and A. Fregolent, "Higher-order MAC for the correlation of close and multiple modes," *Mech. Syst. Signal Process.*, vol. 17, no. 3, pp. 599–610, 2003.

[191] H. Van Der Auweraer and B. Peeters, "Discriminating physical poles from mathematical poles in high order systems: Use and automation of the stabilization diagram," in *Conference Record - IEEE Instrumentation and Measurement Technology Conference*, 2004, vol. 3, pp. 2193–2198.

[192] E. Reynders, J. Houbrechts, and G. De Roeck, "Automated interpretation of stabilization diagrams," in *Conference Proceedings of the Society for Experimental Mechanics Series*, 2011, vol. 3, pp. 189–201.

[193] F. Magalhães, Á. Cunha, and E. Caetano, "Online automatic identification of the modal

parameters of a long span arch bridge," *Mech. Syst. Signal Process.*, vol. 23, no. 2, pp. 316–329, 2009.

[194] D. C. Lay, *Linear algebra and its applications*. Pearson/Addison-Wesley, 2006.

[195] C. Lucas, "Algorithms for Cholesky and QR factorizations, and the semidefinite generalized eigenvalue problem," The University of Manchester, 2004.

[196] G. Lombaert and J. P. Conte, "Random Vibration Analysis of Dynamic Vehicle-Bridge Interaction Due to Road Unevenness," *J. Eng. Mech.*, vol. 138, no. 7, pp. 816–825, 2012.

[197] K. Matsuoka, K. Kaito, and M. Sogabe, "Bayesian time–frequency analysis of the vehicle–bridge dynamic interaction effect on simple-supported resonant railway bridges," *Mech. Syst. Signal Process.*, vol. 135, p. 106373, Jan. 2020.

[198] R. M. Lewis, "Direct Search Methods: Then and Now," *Nasa / Icase*, vol. 124, pp. 191–207, 2000.

[199] J. Nocedal and S. J. Wright, *Numerical optimization*. Springer, 2006.

[200] E. Baeyens, A. Herreros, and J. R. Perán, "A direct search algorithm for global optimization," *Algorithms*, vol. 9, no. 2, pp. 1–22, 2016.

[201] J. A. Nelder and R. Mead, "A Simplex Method for Function Minimization," *Comput. J.*, vol. 7, no. 4, pp. 308–313, 1965.

[202] A. R. Yaakub and D. J. Evans, "Fourth order Runge-Kutta RK(4,4) method with error control," *Int. J. Comput. Math.*, vol. 71, no. 3, pp. 383–411, 1999.

[203] P. Paultre, O. Chaallal, and J. Proulx, "Bridge dynamics and dynamic amplification factors - a review of analytical and experimental findings," *Canadian Journal of Civil Engineering*, vol. 19, no. 2. pp. 260–278, 1992.

[204] S. S. Law and X. Q. Zhu, "Bridge dynamic responses due to road surface roughness and braking of vehicle," *J. Sound Vib.*, vol. 282, no. 3–5, pp. 805–830, 2005.

[205] *ISO 8608:2016 Mechanical vibration-Road surface profiles-Reporting of measured data*. International Organization for Standardization, 2016.

[206] Y. B. Yang, Y. C. Lee, and K. C. Chang, *Effect of road surface roughness on extraction of bridge frequencies by moving vehicle*. Springer Vienna, 2014.

[207] M. Rak, D. Mestrovic, and A. Nizic, "Damping and its importance to structure," *Earthq. Resist. Eng. Struct. V*, vol. 81, pp. 387–393, 2005.

[208] A. González, E. J. Obrien, and P. J. Mcgetrick, "Identification of damping in a bridge using a moving instrumented vehicle," *J. Sound Vib.*, vol. 331, no. 18, pp. 4115–4131, 2012.

[209] J.R. Maguire and T.A. Wyatt, *Dynamics: An introduction for civil and structural engineers ICE design and practice guides*, vol. 2, no. 1. Thomas Telford, 2000.

[210] S. Tao, "Dynamic Response Analysis for a Time-Varying Vehicle-Bridge System based on Matlab," Central South University, 2009.