Structural damage assessment through parametric and nonparametric models

Marcello Morgantini

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2020

Abstract

Structural damage assessment through parametric

and nonparametric models

Marcello Morgantini

The main purpose of Structural Health Monitoring (SHM) is the assessment of structural conditions in aerospace, mechanical and civil systems. In structural engineering, damage is defined as any permanent change in the structural and geometric properties of a system caused by an external action. Vibration-based damage assessment methods rely on the use of sensors that record the structural dynamic response of a system that is determined by its structural and geometric properties. External disturbances and environmental conditions in which the system operates cause fluctuations of these properties and might hide the change in signature induced by damage. To handle the uncertainties in the determination of the structure's characteristics, a statistical pattern recognition approach is presented in this thesis. Any statistical approach relies on the statistics of some features that provide a compact representation of the structural properties and that are sensitive to damage. Such features are called damage sensitive features and are extracted from the dynamic response of the structure: their statistical distribution is then analyzed to assess the occurrence of damage. This dissertation focuses on the analysis of the statistical distribution of damage sensitive features which are extracted through parametric and nonparametric algorithms. Cepstral coefficients are features defined in the field of acoustics and, in this thesis, they have been adapted to SHM analyses in order to develop compact damage sensitive features whose extraction requires a low computational effort. In this thesis, cepstral coefficients have been mathematically transformed through a Principal Component Analysis in order to generate damage sensitive features that are barely sensitive to measurement noise, environmental conditions and different excitation sources. In an attempt to develop an automated strategy for structural damage assessment, the search for damage sensitive features has been extended to the estimation of structural mode characteristics obtained through an output-only version of the Inner Product Vector methodology, e.g. considering only the structural response time histories. This new damage assessment procedure requires low computational effort and

is capable to identify both the presence of damage and its location. However, one of the critical points of the proposed procedure consists in the manual evaluation of the spectral content of the dynamic responses that requires the user's intervention. To automatize this procedure, a Bayesian clustering algorithm and a classifier have been successfully implemented and tested. Finally, the robustness of Bayesian regression algorithms to overfitting led us to consider their applicability to the field of system identification in order to provide a reliable estimate of the structural modal parameters that can be used as damage sensitive features. In fact, one of the main problems of system identification algorithms is that they rely on a regression algorithm that tends to overfit data producing unreliable results. Results provided by the Bayesian regression based system identification algorithm are obtained and compared with the ones coming from standard system identification algorithms.

Contents

List of Figures

List of Tables

Acknowledgments

This thesis consists of part of the research work I've been conducting at Columbia University. For almost four years, I had the pleasure to work side by side with brilliant and enthusiastic students, researchers and professors who constantly supported and motivated me. I will always remember my incredible experience at Columbia, when I started I was just a student, full of curiosity and eager to test myself, today I feel to be much more. I spent most of my time in the office, doing my research, I travelled across the U.S. to join conferences and make personal and professional experience, but I was never alone. I can certainly affirm that the support I received played a major role in the academic path I've been going through.

I want to thank my advisor, Professor Raimondo Betti, who always believed in me since the first time I met him at Columbia. I want to thank Professor Betti for being an example for me from both the academic and personal points of view and for being my North star in my times of difficulties. I want also to acknowledge the members of my dissertation defence committee, Professors George Deodatis, Raimondo Betti, Richard Longman, Maria Feng and Ioannis Kougioumtzoglou for reviewing my thesis and for their helpful observations. I would like to thank Dr. Francesco Vicario for his remarkable support and Professor Alessandro Fasana, from Polytechnic University of Turin, who encouraged me starting this adventure. I would like to thank Professor Keith Worden and Dr. Nikolaos Dervilis from University of Sheffield and Professors Cecilia Surace and Rosario Ceravolo from Polytechnic University of Turin for sharing their knowledge with me.

I want to acknowledge all the other graduate students who made my life happy at Columbia. I truly thank Dr. Eleonora Tronci for being a constant support during this period, for sharing with me anxiety and happiness. I would like to thank Dr. Patrick Alrassy for being always by my side as a true friend. A special thanks goes to my brother Dr. Ioannis Petromichelakis and to my sister Theano Dimitrakis who shared their life with me, I'll always be proud of having been part of their family during these years. I would like to thank Dr. Enrica Leccisi and our common friends for sharing their spare time with me. I thank my friends Andrea, Erica and Gaetano for being part of my team in our Italy vs Greece basketball matches in NY. I thank my friends Francesco, Gigi, Giandomenico and Straffo who called me every day from Italy to make me feel home even though I was far away.

I thank my parents who always did their best to make my dreams become truth and my sister Amalia who travelled across the world more than once to see her elder brother. Finally, I want to thank my aunt Carmela who has always been there for me, my aunt Anna and my uncle Guglielmo whose door was always open for me, and my cousins Sara and Emanuela.

To my family

Chapter 1

1. Introduction

1.1. Background

As our infrastructure system rapidly ages and deteriorates, maintenance and rehabilitation operations constitute an increasingly large economic effort. These activities cannot be performed daily for both logistic and economic reasons. In fact, direct visual inspections may often result in costly and ineffective operations: unforecasted events may suddenly compromise the service conditions of a structure and urgent repair may be needed (the waiting time before the next scheduled inspection may result into a high risk factor). In addition, the outcomes of the inspections might be dependent on the experience and ability of the inspector, often resulting in biased results. Nowadays, thanks to incredible advances in sensor and computer technologies, monitoring techniques based on data acquired from sensors embedded in a structure not only offer the advantage of reducing operational costs, but also guarantee a continuous and effective assessment of the structural condition. Among those techniques, vibration-based monitoring methods and technologies allow engineers to gather data in real time and to assess online whether a structure is damaged or not. Based on these experimental data, the development of a mathematical model, which represents the structure in its current conditions, is the key for a successful damage assessment analysis.

Assessing the structural condition and, eventually, the presence of damage has always been a challenge in the Structural Health Monitoring (SHM) community [1, 2]. In the last few decades, researchers have focused their efforts on the development of mathematical models that accurately represent dynamic systems. The conventional design of such systems allows us to define simple mathematical models that, however, present some drawbacks compared to those models relying on experimental data [3]. In fact, the correct estimation of coefficients representative of the material properties and of the structure's geometry can result in a challenging task. Furthermore, the structural properties may vary when the system is subjected to different external conditions, so that the system itself might present different dynamic responses when operating in different environmental conditions. One of the most advanced techniques for the development of mathematical models is the Finite Element Method (FEM) which finds applications in several fields of engineering. The FEM consists in the analysis of a system which is

divided into small and simple parts, the so called 'finite elements', which are then reassembled together by imposing equilibrium and compatibility. This procedure is addressed by a particular space discretization of the volume of the structure: a mesh of the system is created so to generate a solution based on a finite number of points. The FEM represents one of the most effective methods for the design of a model whose mechanical and geometric parameters are set 'a priori' (with no knowledge of experimental data). However, in real applications, these parameters can be quite different from the initial selection and so, recently, efforts have been made in the development of FEM models whose parameters can be estimated 'a posteriori', from data collected in the field. *Model updating* is an interesting area of SHM aiming to generate a mathematical model (usually an FEM model) to represent the dynamic characteristics of the real system [4, 5] based on real time data that are employed to update the parameters of the model (usually initialized 'a priori') so that its response accurately reproduces the measured one. The variation of the model's mechanical and geometric parameters is a key factor for the damage assessment: in fact, the occurrence of damage can be inferred by looking at changes in some physical parameters (e.g. a change in stiffness). However, there are some issues that need to be addressed in a successful model updating strategy. First, the complexity of the structure of interest impacts on the performance of the method: while model updating algorithms work well with simple structural models, they become cumbersome when dealing with more sophisticated models. Second, model updating strategies become hard to implement when dealing with model's nonlinearities. For these reasons, in the last few years, research in SHM has been moving towards damage assessment strategies oriented to the development of models that purely rely on the statistical analysis of experimental data.

1.2. Parametric and nonparametric models

Models representative of dynamic systems can be designed based on the system's structural properties and geometry and expressed in different forms: mass, damping and stiffness matrices, state-space representation, transfer functions, etc. In structural dynamics, the definition of a system through the *equations of motion* is one of the most common approaches due to the simplicity in designing the parameters representing structural properties and geometries. The structural response is obtained as the result of a set of differential equations whose parameters (such as mass, stiffness, etc.) are representative of the system. Another representation of a dynamic system could be through *state space models*, which consist in a set of input, output and

state variables linked together through first-order differential (or difference) equations. Thanks to their mathematical simplicity and to the fact that they lend themselves to digital implementation, state space models are largely used in control engineering. A different representation of a dynamic system, or of one of its components, is through the *transfer functions* which can be intended as a function that generates dependent outputs from independent inputs. In structural dynamics, models defined by some parameters representing the structural and geometric properties (e.g. mass, stiffenss, damping, state-space models, etc.) are referred to as parametric models. On the contrary, models that define input-output relationships with no assumptions about any structural information and that cannot be represented by specific parameters (e.g. transfer functions) are called nonparametric models.

The distinction between parametric and nonparametric models finds its own definition in the field of statistics [6]. Parametric models are based on some assumptions which can greatly simplify the learning process of the model, but can also limit the accuracy of the estimated parameters and the reliability of the model itself. Following the definition provided in [7], "a learning model that summarizes data with a set of parameters of fixed size (independent of the number of training examples) is called a parametric model. No matter how much data you throw at a parametric model, it won't change its mind about how many parameters it needs". Algorithms that are able to create a parametric model of the system are called parametric algorithms. These algorithms rely on two steps: 1) to define the form of the model to be estimated, i.e. the parameters or features characterizing the model, 2) to estimate these features representing the model from data. An example of parametric models is given by the AutoRegressive (AR) linear models which are usually used to represent a system supposed to be linear. Hence, the first step of the parametric algorithm consists in the definition of a linear model and of the number of coefficients to consider. Consequently, the second step focuses on the use of an algorithm to estimate these coefficients. The main problem is that the real system under consideration might not be linear and so the basic assumption of a linear model might lead to poor identification results. Parametric algorithms are used to develop parametric models that are easy to understand and that lead to results which are easily interpretable. Furthermore, compared to other models, parametric models usually require low learning computational effort and do not need large training datasets.

On the contrary, nonparametric methods allow us to develop nonparametric models without making strong assumptions about the mapping function generating the parameters or features

of the model. The mapping function is learnt based on the training dataset. "Nonparametric methods are good when you have a lot of data and no prior knowledge, and when you don't want to worry too much about choosing just the right features." [7].

Principal component analysis (PCA) represents an example of nonparametric algorithms: it creates a linear mapping function that allows to extract the features characterizing the model from data. Other examples of nonparametric machine learning algorithms are k-nearest neighbours (KNN), Decision trees and Support Vector Machines (SVM) [8, 9]. Generally, the advantages of nonparametric algorithms are that: 1) they can approximate different functional forms, 2) they do not need strong assumptions about the model. On the other hand, compared with parametric methods, they require large training datasets and more computational effort.

1.3. Statistical pattern recognition

Parametric and nonparametric models can also be analyzed in a *pattern recognition* framework. Pattern recognition focuses on the discovery of regularities in data so that we can use these regularities to classify the data into different categories [10]. The objective of pattern recognition is that of providing a compact representation of these regularities in data referred to as *patterns* so that these patterns can be classified into categories. Pattern recognition consists of two phases: a training phase and a test phase. During the training phase, patterns are extracted from training data in order to provide a model associating patterns with the respective category. During the test phase, new patterns are extracted from the test dataset so that they can be classified into the category (*class*) they belong to [11].

In SHM, pattern recognition approaches can be used to: 1) assess the presence of damage; 2) detect the location of damage; 3) quantify the severity of damage [1, 12]. Some patterns can be identified from the structural dynamic response so to be representative of either the *healthy* or *damaged* conditions of the structure. The structural conditions (healthy and damaged) can be associated to two categories in which data are classified. If typical patterns for each of these two categories can be learnt in the training phase, the pattern recognition procedure is referred to as *supervised pattern recognition*. Contrarily, if patterns defining only one of these two categories can be used to generate the training model, the pattern recognition procedure is called *unsupervised pattern recognition*. In SHM of typical civil engineering structures, buildings and bridges, unsupervised pattern recognition is certainly the most appealing of the two approaches for two reasons. First, experimental datasets associated with the structural dynamic response in

damaged conditions are rarely available. Second, despite datasets that represent the structure in damaged conditions can be generated by a properly designed mathematical model, such a model is likely to be inaccurate and the number of damage scenarios (representing the classes or categories) associated to any possible simulated damage occurring in the structure too large for an effective analysis. For these reasons, this thesis focuses on the application of unsupervised pattern recognition approaches for damage identification.

In a damage assessment strategy based on pattern recognition, patterns are represented by the behaviour of some features which are sensitive to damage and which play a major role in the damage assessment process. These features, extracted directly from the recorded dynamic structural response, are generally referred to as *damage sensitive features* [13, 14]. According to a pattern recognition approach, the variation of the damage sensitive features define the classes of membership identified as *healthy* or *damaged* state, with damage commonly interpreted as any change of the structure's geometry and/or material property [1]. It is important to note that structural properties may also be swayed by other external factors (e.g.wind, temperature, traffic...) [15–19]. Whereas external disturbances have a temporary effect on the structural response, a damage induced variation in the system's dynamics yields an irreversible effect on the response. Therefore, it becomes crucial to borrow tools from statistics to deal with the fluctuations of structural characteristics induced by external disturbances during structural damage assessment operations [2, 12, 20, 21]. In order to account for these uncertainties, the damage sensitive features are considered as random variables characterized by a statistical distribution. In this framework, the concept of *statistical pattern recognition* can be formulated as follows: based on the statistical distribution of the damage sensitive features, a training statistical model of the damage sensitive features can be defined and newly collected test features can be investigated through an *outlier analysis*. The objective of the outlier analysis is to state whether the features extracted from the test dataset are likely to be realizations of the training model or not.

1.4. Parametric models and system identification

The development of parametric models representing the structural conditions constitutes a significant part of this thesis. The analysis of the spectral structural response can lead to the estimation of modal parameters defining the model representative of the structural conditions. The recently developed Bayesian algorithms allow us to cluster and extract the structural modal information directly from the spectral response, opening the door to the possibility of designing

an effective and automatized algorithm for the extraction of the parameters characterizing the model. Bayesian techniques are particularly appealing in various fields of engineering for their robustness to overfitting.

Parametric models representing dynamic systems can be extracted through system identification algorithms. The goal of system identification is that of developing a parametric model representative of the dynamic system by using data collected in an experiment. The advantage of such an approach consists in the fact that system identification methods are able to overcome the difficulties associated with the assignment of the correct values of the structural parameters. In fact, it can be difficult 'a priori' to correctly estimate the coefficients representative of the material properties and of the geometry of the structural system, especially because such properties may vary when subjected to different external conditions. On the contrary, system identification methods can provide an 'a posteriori' estimation of the parameters of the structure in its operational conditions; however, since they fundamentally rely on regression models there is the inevitable problem that they can overfit the data introducing some external uncertainties in the definition of the model. This problem is addressed in the analysis conducted in the last chapter of this thesis.

1.5. Research motivation

The work presented in this dissertation aims to the development of *output-only* damage assessment algorithms able to assess the presence and location of structural damage. The entire thesis focuses on the analysis of raw signals collected from the dynamic response of a structure subjected to unknown input excitation sources. Although measurements of the input are never considered, some assumptions on the type of excitation, which are commonly used in SHM, have been made to account for rich output signals. In the theoretical formulations presented, the input has been assumed as either unit pulse or Gaussian white noise (stationary and ergodic process) so to excite the whole spectral content of the system's dynamic response. Based solely on the measured structural response, damage sensitive features have been extracted to assess the presence of damage. Emphasis has been given to those techniques that can be set in an automatized strategy work plan so to provide fast damage assessment and to minimize the user's intervention. As mentioned, it is common practice, in SHM, to extract the modal parameters of a structure from the dynamic structural response and consider their permanent variations as indicative of the presence of damage. For real applications, the correct estimation

of the structural modal parameters may present several obstacles: 1) difficulties in the extraction of the modal parameters due to the complexity of the system, 2) biased estimation of the modal parameters due to unrealistic assumptions at the base of the parametric model used, 3) high computational effort, 4) fluctuation of the modal parameters due to environmental conditions. Hence, the use of structural modal parameters (natural frequencies, mode shapes etc.) as damage sensitive features in an automated damage assessment procedure is limited by the difficulties in the estimation of the parameters themselves. For this reason, new damage sensitive features, the cepstral coefficients, have been introduced and investigated in this thesis.

Cepstral coefficients are features used in many fields of acoustics, in particular in the field of speech and speaker recognition. The intuition about the connection between acoustic waves and mechanical vibrations is the reason behind their application in an SHM context. Cepstral coefficients represent a meeting point between the two fields: features originally developed for the analysis of acoustic signals can be adapted to the structural dynamics domain. Hence, part of the originality and innovation of this work is given by the analytical representation of the cepstral coefficients expressed as function of the structural parameters. Analytical and numerical investigations about the cepstral coefficient proved that they can be interpreted as the sum of two terms, one which does not vary over the monitoring location (it is the same at any monitored location) and another term which is characterized by the local dynamic response. For this intrinsic characteristic, it has been proposed to manipulate the extracted cepstral coefficients through a PCA transformation so to extract a nonparametric model representative of the common term, whose statistical variation has been used as damage indicator. Furthermore, the nonparametric model developed through PCA has also been used to reduce the variance of the identified damage sensitive features due to: 1) external disturbances, 2) different excitation sources, 3) environmental conditions.

In this thesis, still within a framework of an automated damage assessment process, the theory behind the Inner Product Vector (IPV) has been extended to the output-only analysis. Originally though for an input-output process, the IPV-based damage assessment algorithm has been shown able to detect the presence and location of damage. The IPV allows to extract information about a specific structural mode through the cross-correlation of the monitored acceleration response time histories of the structure at the various locations. In its original formulation, the IPV requires a manual inspection of the frequency content of the structural dynamic response so to recognize and isolate the spectral energy contribution of a specific

mode before and after the occurrence of damage. In this framework, the user plays a major role in the analysis. In order to reduce the dependence of the damage assessment algorithm performance from the user expertise, an automatized version of the IPV method has been developed, proposing a Bayesian clustering algorithm and a classifier that are able to automatically extract the information of a specific structural mode before and after the occurrence of damage. This methodology represents a remarkable improvement, especially for those cases in which the occurrence of damage causes new structural modes to contribute to the dynamic response of the system and the identification of a specific mode before and after damage may result in a challenging task.

Investigations about output-only strategies to develop damage assessment algorithms led us to expand our analysis to system identification methods used for the extraction of parametric models. Most of the current system identification methods rely on regression algorithms that provide regression models from which the modal parameters, commonly used as damage sensitive features, can be extracted. However, the performance of the regression model heavily impacts on the results of the system identification algorithm. One of the main problems of regression models is that they can lead to overfitting of data. On the contrary, Bayesian regression models have been proven to be extremely robust to overfitting and, for this reason, a Bayesian regression-based system identification algorithm has been developed and presented in the last chapter of this thesis.

1.6. Thesis organization

This thesis focuses on the development of parametric and nonparametric models of structures in an output-only context. These models are representative of the structural healthy condition and any change of such a condition is analyzed in a statistical pattern recognition framework. In chapter 2, the features referred to as 'cepstral coefficients' are extracted from the structural dynamic response time histories. The mathematical manipulation of these features leads to the development of some damage sensitive features whose distribution defines a nonparametric model which is indicative of some structural conditions. The presence of damage is assessed in a novelty detection approach based on the outlier analysis. The goal of chapter 3 is that of developing an output-only cross-correlation based algorithm which is able to extract the structural modal parameters defining a parametric model. The analysis of the model allows to assess the occurrence of structural damage and to identify the structural damaged area(s). In chapter 4, a

clustering Bayesian model has been developed so to automatize the extraction of the parametric model defined by the structural modal parameters in chapter 3. Finally, chapter 5 introduces the Output-only Observer/Kalman filter Identification algorithm (O^3KID). As mentioned, the conventional implementation of such an algorithm relies on an OLS regression. In this chapter, a Bayesian regression model is used instead: the comparison between the modal parameters extracted by applying the two regression methods is presented.

Chapter 2

2. Structural damage assessment through features in quefrency domain

2.1. Introduction

Among all the possible features (e.g. natural frequencies, mode shapes, stiffness coefficients, AutoRegressive (AR) coefficients, etc.), those extracted from the structural response time history through digital signal processing are certainly the most appealing for their extraction computational efficiency and for the low level of required user expertise. However, when using such 'data based' features, the intuitive physical relationship to the structural properties such as natural frequencies and mode shapes is lost and damage assessment operation becomes more challenging. One of such 'data based' features are the power cepstrum coefficients, obtained from the time history of the structural response simply by taking the Inverse Discrete Fourier Transform (IDFT) of the logarithm of the squared magnitude of the Discrete Fourier Transform (DFT) of the signal. The power cepstrum was introduced in 1963 at Bell's Laboratory by Bogert et al. in an attempt to coin a method able to identify the presence of an echo in a sound signal [22]. Almost contemporary to the work of Bogert and coworkers and independently from the Bell's Laboratory group, in 1965 Oppenheim [23] proposed the complex cepstrum, i.e. the IDFT of the logarithm of the DFT of the signal. An important field of application of complex cepstrum analysis is fault diagnosis of gearbox systems. By noting that the DFT is a particular case of the z-transform, in [23] Oppenheim and Schafer gave a representation of the cepstrum in terms of the poles and zeros of the transfer function of multi-degrees of freedom mechanical systems. Such representation was exploited by Randall and Gao in [24] to recover poles and zeros of the Frequency Response Function of a beam from its response autospectra, by curve-fitting the analytical expression of the complex cepstrum with the cepstrum given in [23], extracted from the measured response itself. In [25], Gao and Randall proposed another method to identify poles and zeros of the transfer function by exploiting again the analytical expression of the complex cepstrum given in [23], in combination with the Ibrahim Time Domain method. Finally Tigli, in [26], extended the method proposed in [25] to the case of multiple input sources. The use of cepstrum-based features in structural damage assessment can be attributed to the work by Zhang et al. [27] in a study on the delamination of concrete bridge decks: there, they used mel-frequency cepstral coefficients (MFCCs) to investigate the

bridge deck acoustic response to ultrasonic pulses. In 2013, Balsamo et al. [13, 28] proposed a structural health monitoring strategy based on a novelty detection analysis of the MFCC using the vibrational response of buildings and bridges. In this study, the power spectrum of the acceleration response time histories was filtered though a triangular filter bank and the inverse discrete cosine transform (IDCT) was used to extract features consisting of uncorrelated MFCCs. The statistical distribution of those features was then analyzed for damage assessment purpose in a novelty detection approach. In 2019, Randall et al. [29] reaffirmed the potential of cepstrum-based features in structural dynamics to extract structural modal models and remove excitation components from the structural response time histories. In late 2019, Civera et al. [30] investigated the Teager-Kaiser Energy Cepstral Coefficients (TECCs) as an alternative to the MFCCs.

Once these cepstrum-based features become available, they can be statistically analyzed so to recognize their pattern. One way to do so is to reduce the dimensionality of such features emphasizing those that have, for example, the maximum variance. Among such techniques of dimensionality reduction, PCA, introduced by Karl Pearson in 1910 and developed and named by Harold Hotelling in the 1930s, is still at the core of modern statistical analysis algorithms and widely used in vibration-based damage assessment methods. Over the past years, PCA has found applications in structural vibrations for reduced-order modeling [31, 32], modal analysis [33], parameter identification and model updating of nonlinear systems [34] and in eliminating environmental effects in damage detection [35–37]. In addition, taking advantage of its ability in reducing model dimensions, PCA has also been used for the extraction of low-dimensional uncorrelated cepstrum-based features. An example is the work by Dackermann et al. [38] who analyzed progressive structural damage scenarios by training an Artificial Neural Network (ANN) through cepstral features. To reduce the dimensionality of those features and break their correlation to each other, a PCA was conducted before training the ANN.

The objective of this chapter is to develop a vibration-based damage assessment technique in the *quefrency* domain, domain in which cepstral coefficients are defined. For this purpose an analytical representation of the cepstral coefficients as function of the structural properties is first developed. This is the first attempt to link, analytically, modal properties of structural systems (e.g. natural frequencies, damping ratios and mode shapes) to cepstrum-based features. Because of their nature, these coefficients depict a mathematical representation of the structural dynamic properties as well as of the excitation, including external disturbances. Principal

Component Analysis can then be used as a valuable method, based on a linear mapping of the cepstral coefficients, to extract the components more likely influenced by the structural properties than by the excitation source and environmental agents. These (uncorrelated) components are those that contribute to the low-variance content of the datasets and, in this chapter, they are considered as damage sensitive features. Here, the cepstral coefficient sequences to be projected through PCA in order to generate the low-variance features have been extracted using two different datasets: 1) by considering signals recorded at multiple structural locations to reduce the variance due to local zeros, 2) by considering signals recorded at the same structural location to reduce the variance due to different excitation sources or external disturbances. In order to resort to the aforementioned statistical pattern recognition based damage detection approach, the Squared Mahalanobis Distance (SMD) [39–44] of such features rather than the Euclidean distance and the cumulative distribution function of the damage sensitive features have been considered. Hence, since the Euclidean distance is equal to the SMD for the particular case in which the covariance matrix of the sampled data is the identity matrix [10], the SMD represents, in general, a more appropriate metric for the outlier analysis of multivariate probability distributions. In [45] a comparison between these two metrics is presented in the analysis of a cracked beam through damage assessment tests relying on a statistical pattern recognition approach. The mismatch between the distributions of the damage sensitive features is then statistically detected to assess the presence of a possible damage in the system. In order to validate the method, results from both numerical and experimental tests have been analyzed. The numerical test consists in the analysis of the response of an 8DOF shear type model, whereas the experimental test is conducted on data available from the monitoring of the Z24 Bridge (Switzerland), largely used in the literature [46–49].

2.2. The use of the Power Cepstrum in a data-based framework

As previously mentioned, cepstrum coefficients have been employed with good results in a variety of fields, ranging from image and sound processing to damage assessment in rotary machine. These coefficients can be analyzed in problems typical of structural health monitoring where the information on the structural conditions of a system could be extracted by its response output. Of all the methods to solve the damage detection problem within a data-based framework, outlier analysis is one of the most robust and efficient. In this chapter, for a proper treatment of outliers, the Squared Mahalanobis Distance (SMD) is used as the metric apt to

distinguish between damaged and undamaged instances of the damage sensitive feature.

It is worth to reiterate that, in statistical pattern recognition, the SMD is used to measure the distance of the damage sensitive features extracted from the response of the system in unknown conditions from those extracted from the response of the system in known conditions: the larger the value of the SMD, the larger the departure of the new damage sensitive features from the population of realizations of damage sensitive features representative of undamaged conditions, and hence larger the probability that the new damage sensitive features are representative of a damage state of the structure.

2.2.1. Analytical expression of the Power Cepstrum from structural acceleration

Let us consider the equations of motion of an N degrees of freedom (DOF) linear time-invariant system:

$$\mathscr{M}\ddot{\boldsymbol{y}}(t)+\mathscr{C}\dot{\boldsymbol{y}}(t)+\mathscr{K}\boldsymbol{y}(t)=\boldsymbol{u}(t) \tag{2.1}$$

where $\mathscr{M}$, $\mathscr{C}$ and $\mathscr{K}$ represent the mass, damping and stiffness matrices, respectively, each of order $N\times N$. The vector $\boldsymbol{y}(t)\in\mathbb{R}^{N\times 1}$ represents the nodal displacement vector, $\dot{\boldsymbol{y}}(t)\in\mathbb{R}^{N\times 1}$ the nodal velocity vector, and $\ddot{\boldsymbol{y}}(t)\in\mathbb{R}^{N\times 1}$ the nodal acceleration vector. Finally, $\boldsymbol{u}(t)\in\mathbb{R}^{N\times 1}$ is the input vector, containing the values of the external excitation at time t.

At first, in order to simplify the algebra, let us assume that all degrees of freedom of the structure are excited by the same input time history (this assumption will be removed later). Equation (2.1) can then be rewritten as

$$\mathscr{M}\ddot{\boldsymbol{y}}(t)+\mathscr{C}\dot{\boldsymbol{y}}(t)+\mathscr{K}\boldsymbol{y}(t)=\boldsymbol{B}_2u(t) \tag{2.2}$$

where $\boldsymbol{B}_2\in\mathbb{R}^{N\times 1}$ is a vector of ones while $u(t)$ represents the time-history of the input excitation. Assuming zero initial conditions, the Laplace transform of Equation (2.2) yields:

$$(\mathscr{M}s^2+\mathscr{C}s+\mathscr{K})\mathbf{Y}(s)=\boldsymbol{B}_2U(s) \tag{2.3}$$

where $\mathbf{Y}$(s) represents the vector of the Laplace transform of nodal displacement responses and $U(s)$ represents the Laplace transform of the input time history applied at each degree of freedom.

Let us now denote by $\mathbf{\Phi}$ the mode shape matrix resulting from the solution of the generalized undamped eigenvalue problem

$$(\mathscr{K} - \lambda_i \mathscr{M})\boldsymbol{\phi}_i = 0, \tag{2.4}$$

where λ_i represents the i^{th} eigenvalue and $\boldsymbol{\phi}_i$ (the i^{th} column of the mode shape matrix $\mathbf{\Phi}$) its associated eigenvector. By making use of the well known mass-orthogonality and stiffness-orthogonality properties of the eigenvector matrix $\mathbf{\Phi}$, here onwards assumed to be mass normalized, and assuming that the system is classically damped, so that $\mathbf{\Phi}^T \mathscr{C} \mathbf{\Phi}$ represents a diagonal matrix, the expression for the vector of Laplace transform of the displacement time histories is obtained as

$$\mathbf{Y}(s) = \mathbf{\Phi}[\mathbf{I}s^2 + \mathbf{\Xi}s + \mathbf{\Lambda}]^{-1}\mathbf{\Phi}^T \mathbf{B}_2 U(s), \tag{2.5}$$

where:

1. $\mathbf{I}$ is the identity matrix of order N;
2. $\mathbf{\Xi}$ is a diagonal matrix of order N: $\mathbf{\Xi} = diag\{2\xi_i\omega_i\}$, for $i = 1,...,N$, where ξ_i and ω_i are the damping ratio and natural frequency associated with the i^{th} mode, respectively;
3. $\mathbf{\Lambda}$ is a diagonal matrix of order N: $\mathbf{\Lambda} = diag\{\omega_i^2\}$ for $i = 1,...,N$.

The matrix $\mathbf{H_d}(s) = \mathbf{\Phi}[\mathbf{I}s^2 + \mathbf{\Xi}s + \mathbf{\Lambda}]^{-1}\mathbf{\Phi}^T$, also known as the *receptance* matrix, represents the matrix of the transfer functions from the inputs, represented by the vector $\mathbf{B}_2 U(s)$ to the displacement outputs in $\mathbf{Y}(s)$. It is possible to express the $(i,j)^{th}$ term of the receptance matrix as follows:

$$H_d(s)_{i,j} = \sum_{l=1}^{N} \frac{\phi_{i,l}\phi_{j,l}}{m_l(s-\lambda_l)(s-\lambda_l^*)} \tag{2.6}$$

where the complex eigenvalue λ_l and its complex conjugate λ_l^* are given by:

$$\lambda_l, \lambda_l^* = -\omega_l\xi_l \pm j\omega_l\sqrt{1-\xi_l^2} \tag{2.7}$$

while the l^{th} modal mass, m_l, is equal to 1, by virtue of the normalization property assumed for the eigenvector matrix. By recalling the properties of the Laplace transform, it is possible to get the expression for the $(i,j)^{th}$ element of the transfer function from the j^{th} input to the i^{th}

acceleration output, i.e. the $(i,j)^{th}$ element of the inertance matrix:

$$H_a(s)_{i,j} = \sum_{l=1}^{N} \frac{\phi_{i,l}\phi_{j,l}s^2}{(s-\lambda_l)(s-\lambda_l^*)}. \tag{2.8}$$

Since, in SHM, it is customary to deal with digital signals, it is convenient to operate in the z-domain. To do so, it is possible to manipulate Equation (2.8) to obtain the $(i,j)^{th}$ element of the inertance matrix in the z-domain:

$$H_a(z)_{i,j} = \sum_{l=1}^{N} \frac{\phi_{i,l}\phi_{j,l}(1-z^{-1})(1-P_{a,l}z^{-1})}{(1-e^{\lambda_l T}z^{-1})(1-e^{\lambda_l^* T}z^{-1})} \tag{2.9}$$

where:

1. T denotes the sampling time;
2. $P_{a,l} = \frac{e^{-\xi_l \omega_l T} cos(\omega_{d,l}T - \xi_l)}{\sqrt{1-\xi_l^2}}$
3. $w_{d,l} = w_l\sqrt{1-\xi_l^2}$.

Equation (2.9) may be written in terms of products as follows:

$$H_a(z)_{i,j} = \frac{(1-z^{-1})\prod_{l=1}^{2N-1}(1-Z_l^{(ij)}z^{-1})}{\prod_{l=1}^{N}(1-e^{\lambda_l T}z^{-1})(1-e^{\lambda_l^* T}z^{-1})} \tag{2.10}$$

where $Z_l^{(ij)}$ $(l = 1,...,2N-1)$ are the roots of the following equation:

$$\sum_{l=1}^{N}\phi_{i,l}\phi_{j,l}(1-P_{a,l}z^{-1})\prod_{\substack{k=1\\k\neq l}}^{N}(1-e^{\lambda_k T}z^{-1})(1-e^{\lambda_k^* T}z^{-1}) = 0. \tag{2.11}$$

Since the assumption is made that all degrees of freedom are excited by the same input time history, the z-transform of the acceleration time history at the i^{th} DOF yields

$$A(z)_i = \sum_{j=1}^{N} H_a(z)_{i,j}U(z). \tag{2.12}$$

which can be rewritten in terms of products using Equation (2.10) and (2.11)

$$A(z)_i = \frac{(1-z^{-1})\prod_{l=1}^{2N-1}(1-Z_l^{(i)}z^{-1})}{\prod_{l=1}^{N}(1-e^{\lambda_l T}z^{-1})(1-e^{\lambda_l^* T}z^{-1})}U(z), \tag{2.13}$$

where $Z_l^{(i)}$ ($l = 1,...,2N-1$) are the roots of the following equation:

$$\sum_{j=1}^{N}\sum_{l=1}^{N}\phi_{i,l}\phi_{j,l}(1-P_{a,l}z^{-1})\prod_{\substack{k=1\\k\neq l}}^{N}(1-e^{\lambda_k T}z^{-1})(1-e^{\lambda_k^* T}z^{-1}) = 0. \tag{2.14}$$

It is interesting to point out that the dependance of the z-transform of the acceleration from the location where the acceleration is recorded is confined only in the zeros $Z_l^{(i)}$ while the other parameters (λ_l and λ_l^*) depend on the overall structural properties (e.g. frequencies and damping ratio). At this point, the process of extracting the cepstral coefficients requires the evaluation of the logarithm of the power spectrum of the sampled response time history. This represents a crucial step in the determination of the cepstral coefficients but can be circumvented by exploiting the properties of the complex logarithm. By definition, the complex logarithm is the inverse of the complex exponential function [50]. Hence, given a complex variable x_c, whose polar form is given by $Xe^{i\psi}$, where X denotes the magnitude of x_c while ψ its phase, the complex logarithm of x_c is given by:

$$ln(x_c) = ln(X) + i\psi \tag{2.15}$$

Then, the real part of the complex logarithm of x_c is the natural logarithm ln of its magnitude X. According to that, we can write:

$$ln(|A(z)_i|^2) = 2\mathscr{R}\{ln(A(z)_i)\} \tag{2.16}$$

where $\mathscr{R}\{.\}$ represents the real part of the quantity in brackets. Thus, the logarithm of the power spectrum can be easily extracted by the complex logarithm of $A(z)_i$. Equation (2.13) provides the z-transform of the acceleration time history at the i^{th} degree of freedom in terms of products, so we can use the complex logarithm to isolate each single contribution. In this case the complex logarithm of the z-transform of the acceleration time history recorded at the

i^{th} degree of freedom is given by:

$$ln[A(z)_i] = ln[U(z)] + ln(1-z^{-1}) + \sum_{l=1}^{2N-1} ln(1-Z_l^{(i)}z^{-1}) + \\ - \sum_{l=1}^{N} ln(1-e^{\lambda_l T}z^{-1}) - \sum_{l=1}^{N} ln(1-e^{\lambda_l^* T}z^{-1}). \quad (2.17)$$

Let's now focus on the last four terms of Equation (2.17): they are all summations of logarithmic functions $ln(1-\alpha z^{-1})$. It is then appropriate to make use of the McLaurin complex series expansion for the logarithmic function $ln(1-\alpha z^{-1})$:

$$ln(1-\alpha z^{-1}) = -\sum_{n=1}^{\infty} \frac{\alpha^n}{n} z^{-n}. \quad (2.18)$$

The complex series in Equation (2.18) converges for all complex numbers αz^{-1} having magnitude less than or equal to one. If we evaluate the z-transform on the unit circle, and denote the magnitude of the complex number α as $|\alpha|$ and its phase as ψ, the condition of convergence for the series in Equation (2.18) can be expressed as

$$|\alpha z^{-1}| = ||\alpha| e^{i\psi} e^{\frac{-i2\pi k}{N}}| = |\alpha| \leq 1. \quad (2.19)$$

Systems typically considered in civil and mechanical engineering applications are stable, i.e. the poles of their discrete transfer function are all contained within the unit circle, so that their magnitude is indeed less than one. In particular, the magnitude of all poles of the pulse transfer function in Equation (2.13) is equal to $e^{-\xi_l \omega_l T}$ $(l = 1,...,N)$, which is a positive quantity less than one, since the values of ξ_l,ω_l $(l = 1,...,N)$ are always positive for typical civil and mechanical structures. Therefore, the condition represented by Equation (2.19) applies and the series expansion in Equation (2.18) can be employed to expand the functions $ln(1-e^{\lambda_l T}z^{-1})$ and $ln(1-e^{\lambda_l^* T}z^{-1})$. On the contrary, looking at the term $ln(1-Z_l^{(i)}z^{-1})$ it might happen that the magnitude of some of the zeros $Z_l^{(i)}$ could be greater than one implying that, for such zeros, said series will not converge. In such a case, the transfer function in Equation (2.13) is said to represent a non-minimum phase system. However, any transfer function can be made minimum phase by 'reflecting' all zeros $Z_l^{(i)}$, for which $|Z_l^{(i)}| > 1$, inside the unit circle, i.e. by replacing the non minimum phase zero $Z_l^{(i)}$ with $Z_l^{(i)-1}$. It is important to note that this replacement will not alter the magnitude of the power spectrum, which is the only part of the spectrum we are

concerned with in the computation of the logarithm of the power cepstrum. Consequently, we can assume that the magnitude of all zeros $Z_l^{(i)}$ is less than one, so that the McLaurin series in Equation (2.18) will converge also for the $2N-1$ terms $ln(1-Z_l^{(i)}z^{-1})$.

By combining the last four terms of Equation (2.17) and recalling Equation (2.7) and Equation (2.18), the following result is obtained:

$$\begin{aligned} & ln(1-z^{-1}) + \sum_{l=1}^{2N-1} ln(1-Z_l^{(i)}z^{-1}) - \sum_{l=1}^{N} ln(1-e^{\lambda_l T}z^{-1}) - \sum_{l=1}^{N} ln(1-e^{\lambda_l^* T}z^{-1}) = \\ & = \sum_{q=1}^{\infty} \frac{1}{q} \left[\sum_{l=1}^{N} 2e^{-\xi_l \omega_l T q} cos(\omega_{d,l} T q) - 1 - \sum_{l=1}^{2N-1} Z_l^{(i)q} \right] z^{-q} = \\ & = \mathscr{Z} \left\{ \frac{1}{q} \left[\sum_{l=1}^{N} 2e^{-\xi_l \omega_l T q} cos(\omega_{d,l} T q) - 1 - \sum_{l=1}^{2N-1} Z_l^{(i)q} \right] \right\}, \quad \forall q > 0 \end{aligned} \tag{2.20}$$

where the symbol $\mathscr{Z}\{.\}$ indicates the z-transform of the term within brackets. It can then be concluded that the last four terms of Equation (2.17) represent the z-transform of the real and causal sequence in brackets in Equation (2.20). It is noteworthy that all the quantities here are only functions of the dynamic characteristics of the systems, including the zeros $Z_l^{(i)}$, while the effects of the input force are confined in the term $ln[U(z)]$. This will not be true when we will consider different forces acting at various degrees of freedom; in this case the zeros $Z_l^{(i)}$ will also account for zeros and pole of the input forces. In order to evaluate the real part of Equation (2.20), one can thus take recourse to some properties of the z-transform of real (the zeros $Z_l^{(i)}$ are real or complex conjugates) and causal ($q > 0$) signals. The real part of the z-transform of a causal real-valued function $h[q]$ is then equal to half of the z-transform of the function itself:

$$\mathscr{R}\{\mathscr{Z}\{h[q]\}\} = \frac{1}{2}\{\mathscr{Z}\{h[q]\}\}. \tag{2.21}$$

The real part of the last four terms of Equation (2.17) is then expressed using Equations (2.20) and (2.21) as:

$$\begin{aligned} & \mathscr{R} \left\{ \mathscr{Z} \left\{ \frac{1}{q} \left[\sum_{l=1}^{N} 2e^{-\xi_l \omega_l T q} cos(\omega_{d,l} T q) - 1 - \sum_{l=1}^{2N-1} Z_l^{(i)q} \right] \right\} \right\} = \\ & = \tfrac{1}{2} \mathscr{Z} \left\{ \frac{1}{q} \left[\sum_{l=1}^{N} 2e^{-\xi_l \omega_l T q} cos(\omega_{d,l} T q) - 1 - \sum_{l=1}^{2N-1} Z_l^{(i)q} \right] \right\}. \end{aligned} \tag{2.22}$$

and, consequently, the real part of the complex logarithm of $A(z)_i$ can be rewritten as:

$$\mathscr{R}\{ln[A(z)]\} = \frac{1}{2}\mathscr{Z}\left\{\frac{1}{q}\left[\sum_{l=1}^{N} 2e^{-\xi_l \omega_l T q} cos(\omega_{d,l} T q) - 1 - \sum_{l=1}^{2N-1} Z_l^{(i)q}\right]\right\} + \mathscr{R}\{ln[U(z)]\}. \tag{2.23}$$

By considering Equations (2.16) and (2.23), the expression of the logarithm of the power spectrum of the acceleration response time history becomes:

$$\begin{aligned} ln\left\{|A(z)_i|^2\right\} &= 2\mathscr{R}\{ln[A(z)_i]\} = \\ &= \mathscr{Z}\left\{\frac{1}{q}\left[\sum_{l=1}^{N} 2e^{-\xi_l \omega_l T q} cos(\omega_{d,l} T q) - 1 - \sum_{l=1}^{2N-1} Z_l^{(i)q}\right]\right\} + 2\mathscr{R}\{ln[U(z)]\}. \end{aligned} \tag{2.24}$$

By taking the inverse z-transform of Equation (2.24), the expression for the coefficients of the power cepstrum extracted from the acceleration response measured at the i^{th} degree of freedom of an MDOF system excited at all DOF by an input $u(t)$ is finally obtained as

$$c_i[q]\begin{cases} 0 & \text{for} \quad q<0 \\ \hat{u}[q] & \text{for} \quad q=0 \\ \frac{1}{q}\left[\sum\limits_{l=1}^{N} 2e^{-\xi_l \omega_l T q} cos(\omega_{d,l} T q) - 1 - \sum\limits_{l=1}^{2N-1} Z_l^{(i)q}\right] + \hat{u}[q] & \text{for} \quad q>0 \end{cases} \tag{2.25}$$

where, recalling Equation (2.16), $\hat{u}[q]$ is the inverse z-transform of the logarithm of the power spectrum of the sampled input $u[n]$. In view of Equation (2.25), it is interesting to compare these formulations with those presented in the work by Gao and Randall [25]. In Equation (10) in [25], they presented an analytical formulation for the coefficients of the complex spectrum for the case of an impulse excitation. The analytical representation of the coefficients of the power cepstrum in Equation (2.25) is referred to the case of a general external excitation and is consistent with the representation of the coefficients of the complex cepstrum presented in [25] when considering the general input as an impulse. By comparing the equations, it can be observed that: 1) both the sequences of coefficients represent a casual signal, 2) the term at $(q=0)$ depends only on the natural logarithm of the power spectrum of the sampled input $u[n]$ and 3) the coefficients for $q>0$ decay to zero depending only on the structural properties if the system is excited by an impulsive source.

Equation (2.25) shows a very important result: for $q > 0$ it is evident that the cepstral coefficients of the acceleration response time history of a classically damped MDOF system can be obtained as a linear combination of the discrete poles and zeros of the pulse transfer function separated from the input contribution. Basically, in those coefficients, it will be possible to separate the contribution from the structural properties which will remain constant if the structure does not change, from the contribution related to the input excitation, which will vary depending on the type of excitation. Therefore, the q^{th} cepstral coefficient ($q > 0$) for the acceleration at the i^{th} degree of freedom can be expressed as the sum of 3 contributions:

$$c_{i,q} = c_i[q] = \frac{1}{q}\left[\sum_{l=1}^{N} 2e^{-\xi_l \omega_l T q} cos(\omega_{d,l} T q) - 1\right] + \frac{1}{q}\left[-\sum_{l=1}^{2N-1} Z_l^{(i)q}\right] + \hat{u}[q] \tag{2.26}$$

$$c_{i,q} = \theta_q + \gamma_{i,q} + \beta_q \tag{2.27}$$

where:

$$\begin{cases} \theta_q = \frac{1}{q}\left[\sum_{l=1}^{N} 2e^{-\xi_l \omega_l T q} cos(\omega_{d,l} T q) - 1\right] \\ \gamma_{i,q} = -\frac{1}{q}\sum_{l=1}^{2N-1} Z_l^{(i)q} \\ \beta_q = \hat{u}[q]. \end{cases} \tag{2.28}$$

Equations (2.27) and (2.28) lead to some important observations. First, it is clear that the component θ_q depends only on the structural properties (frequencies and damping ratios) of the overall structure and so it is independent from the monitored position. This means that this component will be present in any q^{th} cepstral coefficient extracted from the various time histories recorded at different locations. It is also clearly shown that the component $\gamma_{i,q}$, being linked to the roots of Equation (2.14) is the only term that depends on the i^{th} monitored location (through the components of the mode shapes) as well as on the overall structural dynamic properties. The term β_q depends only on the excitation source and, by the assumption of equal excitation at every degree of freedom, does not change with respect to the i^{th} degree of freedom. It is noteworthy that the assumption of a MDOF system subjected to exactly the same input excitation at each degree of freedom is restrictive and unrealistic, but allows to easily validate Equation (2.13). Actually, such equation can be shown to be valid also for different excitations at different DOF as long as the z-transform of the input at each degree of freedom is the same: this is the case of different Gaussian white noise input excitations at the various degrees of freedom but all having the same variance. The sequence of cepstral coefficients provided in

Equation (2.27) is an analytical representation of the output in the quefrency domain which is, by definition, the inverse Fourier transform of the logarithm of the squared magnitude of the Fourier transform of the acceleration signals.

2.2.2. Numerical validation: SDOF system perturbed by a unit pulse

To verify the validity of the approximation of the logarithm with a McLaurin series, let's consider the case of a single degree of freedom system subjected to a unit pulse excitation and compare the analytical expressions of the cepstral coefficients in Equation (2.28) with the numerical values obtained from the power cepstrum of the unit pulse response. The advantage of using that kind of excitation is that, in this case, it is possible to easily obtain a close-form solution for the coefficients in Equation (2.28). The root provided by Equation (2.14) for a SDOF system is given by:

$$Z_1^{(1)} = P_{a,1} = \frac{e^{-\xi_1 \omega_1 T} cos(\omega_{d,1} T - \xi_1)}{\sqrt{1-\xi_1^2}} \tag{2.29}$$

so that we can rewrite Equation (2.28) as:

$$\begin{cases} \theta_q = \frac{1}{q}\left[\sum_{l=1}^{1} 2e^{-\xi_l \omega_l T q} cos(\omega_{d,l} T q) - 1\right] = \frac{1}{q}\left[2e^{-\xi_1 \omega_1 T q} cos(\omega_{d,1} T q) - 1\right] \\ \gamma_{1,q} = -\frac{1}{q}\sum_{l=1}^{1} Z_l^{(1)q} = -\frac{1}{q} Z_1^{(1)q} = -\frac{1}{q}(P_{a,1})^q \\ \beta_q = 0 \end{cases} \tag{2.30}$$

Figure 2.1 shows the first 60 values of the cepstral coefficients ($q = 1, \ldots, 60$) for an SDOF system, characterized by mass, stiffness and damping respectively of 1kg, 500N/m and 1%. The analytical solution in quefrency domain is represented by circles while the one generated numerically by applying the definition of power cepstrum to the unit pulse response is marked by crosses. It is evident that the analytical solution accurately approximates the numerical solution. In addition, in Figure 2.1, the contribution of the term γ to the cepstral coefficients is also plotted with the intent to show its contribution to the various coefficients. The behavior of γ is particularly interesting since, contrarily to θ, it is a local parameter. It is clearly shown that γ decays exponentially to zero and its contribution is predominant in the very first values of the cepstral coefficients which are then, intuitively, sensitive to the location of a potential structural damage. However, it is noteworthy to underline that a structural damage affects γ

as well as θ (e.g. changes in frequencies, damping ratios and mode shapes) and so all the cepstral coefficient sequences are sensitive to the damage, no matter which degree of freedom is monitored.

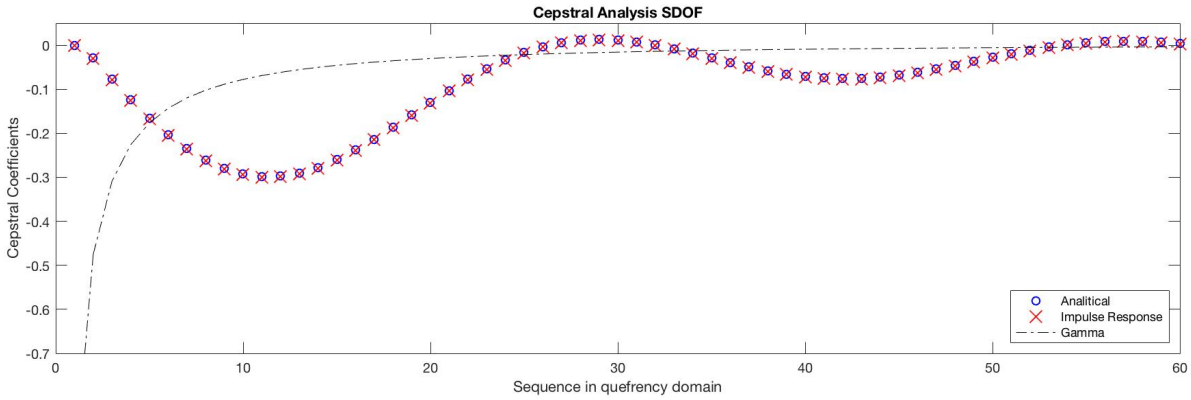


Fig. 2.1. *Cepstral coefficients: analytical vs numerical solution for a unit pulse response. The analytical representation of the Gamma ($\gamma_{1,q}$) component is provided.*

2.2.3. Generic input excitation

For the case of different excitation sources at the various DOF, Equation (2.12) turns into:

$$A(z)_i = \sum_{j=1}^{N} H_a(z)_{i,j} U_j(z) \tag{2.31}$$

where $U_j(z)$ is the z-transform of the input excitation $u_j(t)$ applied at the j^{th} degree of freedom. As before such transform can then be rewritten in terms of products as

$$A(z)_i = \frac{(1-z^{-1}) \prod_{l=1}^{M} (1 - Z_l'^{(i)} z^{-1})}{\prod_{l=1}^{N} (1 - e^{\lambda_l T} z^{-1})(1 - e^{\lambda_l^* T} z^{-1})}, \tag{2.32}$$

where M is the number of roots of the following equation:

$$\sum_{j=1}^{N} U_j(z) \sum_{l=1}^{N} \phi_{i,l} \phi_{j,l} (1 - P_{a,l} z^{-1}) \prod_{\substack{k=1 \\ k \neq l}}^{N} (1 - e^{\lambda_k T} z^{-1})(1 - e^{\lambda_k^* T} z^{-1}) = 0. \tag{2.33}$$

It is important to note that now, in Equation (2.33), the zeros $Z_l'^{(i)}$ $l = 1, \ldots, M$ account for the contributions of zeros and poles of the different input excitations while, previously, these zeros were functions only of the structural properties. This will still allow us to decompose

the cepstral coefficients into a part θ_q that is function only of the structural properties and a part $\gamma_{i,q}$ that depends on both structural properties and input excitations. Proceeding with the calculations along the same line of the previous analysis, the analytical expression of the cepstral coefficients can be found to be:

$$c_{i,q} = c_i[q] = \frac{1}{q}\left[\sum_{l=1}^{N} 2e^{-\xi_l \omega_l T q} cos(\omega_{d,l} T q) - 1 - \sum_{l=1}^{M} Z_l^{'(i)q}\right] \quad for \quad q > 0 \tag{2.34}$$

Just like in the previous section, our interest is focused on the cepstral coefficients for $q > 0$ at every degree of freedom $i = 1, \ldots, N$. Thus, we are able to write:

$$c_{i,q} = \theta_q + \gamma_{i,q} \tag{2.35}$$

where, for the first Q coefficients ($q = 1, 2, \ldots, Q$):

$$\begin{cases} \theta_q = \frac{1}{q}\left[\sum\limits_{l=1}^{N} 2e^{-\xi_l \omega_l T q} cos(\omega_{d,l} T q) - 1\right] \\ \gamma_{i,q} = -\frac{1}{q}\sum\limits_{l=1}^{M} Z_l^{'(i)q} \end{cases} \tag{2.36}$$

As anticipated, each of the Q elements of the vector $\boldsymbol{c}_i = [c_{i,1}, \ldots, c_{i,Q}]^T$ depends on two terms: θ_q fully determined by the structural properties and $\gamma_{i,q}$ that depends on the structural properties, the degree of freedom of interest and the input excitation provided to the system. Given a set of observations of the acceleration response time histories of a certain system in an identical state (either healthy or damaged), the variance between the data is given by $\gamma_{i,q}$. The projection of such data onto a space minimizing the variance introduced by $\gamma_{i,q}$ reduces the effects on the response due to the generic input source, but is not able to reduce the variance potentially due to the drop in stiffness of the structural elements affecting both θ_q and $\gamma_{i,q}$. Thus, as long as the excitation sources have similar statistics before and after a structural damage occurs, cepstral coefficients (and their projections) lend themselves to be used in damage assessment strategies.

In the next section, the benefits and the advantages of a linear projection of the cepstral coefficient sequences, based on the formulations in Equations (2.35) and (2.36), through the Principal Component Analysis are illustrated: a preliminary analysis on the projections will lead to the ultimate damage assessment algorithm presented in section 2.6.

2.3. Principal Component Analysis for minor components extraction

In a data analysis framework, Principal Component Analysis is widely used for several application such as dimensionality reduction, lossy data compression, feature extraction and data visualization [10]. Principal Component Analysis is a statistical procedure that uses a linear orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables (*principal components*). This transformation is a linear mapping of the starting observations onto a new space in which the first principal component has the largest variance. All the succeeding components are set to have the maximum variance allowed under the constraint that they are orthogonal to the precedent components.

2.3.1. Maximum variance formulation

Let's assume that we have a dataset of N_{ob} observations of the system's dynamic response (e.g. time histories of the accelerations) at a certain i^{th} degree of freedom, given N the number of monitored DOFs ($i = 1,\dots,N$). For each observation, it will be possible to extract a column vector $\boldsymbol{c}_i \in \mathbb{R}^{Q\times 1}$ containing the sequence of the first Q cepstral coefficients defined in the Q-dimensinal Euclidean space. By considering the cepstral coefficient sequences at this specific i^{th} location, it is possible to indicate with $\boldsymbol{c}_i^{(j)} \in \mathbb{R}^{Q\times 1}$ the j^{th} observation of $\boldsymbol{c}_i$ for $j = 1,\dots,N_{ob}$. The goal is to project the dataset of N_{ob} observed $\boldsymbol{c}_i^{(j)}$ vectors onto a lower dimensional space with dimensionality $Q' < Q$, by maximizing the variance of the projected data. For sake of simplicity, let's now assume $Q' = 1$ and consider a vector $\boldsymbol{v}_1 \in \mathbb{R}^{Q\times 1}$ set to be a unit vector so that $\boldsymbol{v}_1^T \boldsymbol{v}_1 = 1$ (T stands for the transpose ot the vector). Every vector $\boldsymbol{c}_i^{(j)}$ can be linearly mapped onto a new space via multiplication by the vector $\boldsymbol{v}_1$. The product $\boldsymbol{v}_1^T \boldsymbol{c}_i^{(j)}$ is a scalar value representing the projection of the vector $\boldsymbol{c}_i^{(j)}$ along the direction $\boldsymbol{v}_1$. The mean value of such projection can be expressed as $\boldsymbol{v}_1^T \bar{\boldsymbol{c}}_i$, where $\bar{\boldsymbol{c}}_i$ is given by:

$$\bar{\boldsymbol{c}}_i = \frac{1}{N_{ob}} \sum_{j=1}^{N_{ob}} \boldsymbol{c}_i^{(j)} \tag{2.37}$$

and represents the vector of mean values of the Q cepstral coefficients. The variance Σ of the projected data is provided by:

$$\Sigma = \frac{1}{N_{ob}} \sum_{j=1}^{N_{ob}} (\boldsymbol{v}_1^T \boldsymbol{c}_i^{(j)} - \boldsymbol{v}_1^T \bar{\boldsymbol{c}}_i)^2 = \boldsymbol{v}_1^T \mathbf{S} \boldsymbol{v}_1 \tag{2.38}$$

where $\mathbf{S}$ is the covariance matrix of the observations:

$$\mathbf{S} = \frac{1}{N_{ob}} \sum_{j=1}^{N_{ob}} (\boldsymbol{c}_i^{(j)} - \bar{\boldsymbol{c}}_i)(\boldsymbol{c}_i^{(j)} - \bar{\boldsymbol{c}}_i)^T \tag{2.39}$$

As previously mentioned, the final objective is to select the vector $\boldsymbol{v}_1$ so to maximize the variance of the observations in the projected space, meaning that $\boldsymbol{v}_1$ should be picked so to maximize $\boldsymbol{v}_1^T \mathbf{S} \boldsymbol{v}_1$. At the same time, the constraint $\boldsymbol{v}_1^T \boldsymbol{v}_1 = 1$ should be taken into account. This leads to the definition of a new functional by introducing a Lagrange multiplier $\bar{\lambda}_1$ that multiplies the constrain equation, and so the new functional to be maximized will become:

$$F(\boldsymbol{v}_1, \bar{\lambda}_1) = \boldsymbol{v}_1^T \mathbf{S} \boldsymbol{v}_1 + \bar{\lambda}_1 (1 - \boldsymbol{v}_1^T \boldsymbol{v}_1). \tag{2.40}$$

By setting the derivative of $F(\boldsymbol{v}_1, \bar{\lambda}_1)$ with respect to $\boldsymbol{v}_1$ equal to zero, the following relation is provided:

$$\mathbf{S} \boldsymbol{v}_1 = \bar{\lambda}_1 \boldsymbol{v}_1. \tag{2.41}$$

Equation (2.41) implies that the vector $\boldsymbol{v}_1$ must be an eigenvector of $\mathbf{S}$ and $\bar{\lambda}_1$ its corresponding eigenvalue. By left-multiplying both side of Equation (2.41) by $\boldsymbol{v}_1^T$ and recalling that $\boldsymbol{v}_1^T \boldsymbol{v}_1 = 1$, the maximum value of the Σ variance is given by Σ_1:

$$\Sigma_1 = \boldsymbol{v}_1^T \mathbf{S} \boldsymbol{v}_1 = \bar{\lambda}_1 \tag{2.42}$$

indicating that the variance Σ is maximized when we set $\boldsymbol{v}_1$ to be the eigenvector associated with the largest eigenvalue $\bar{\lambda}_1$. This eigenvector is defined as the *first principal component*. Until now we discussed the special case of $Q' = 1$. If more principal components are needed ($1 \leq Q' \leq Q$), after the extraction of the first principal component, it is possible to choose each new direction for the remaining principal components by maximizing the projected variance $\boldsymbol{\Sigma}$ along this new direction but with the constrain that this direction be orthogonal to all the components previously computed. For the generic i^{th} principal component, the variance Σ will be given as:

$$\Sigma_i = \boldsymbol{v}_i^T \mathbf{S} \boldsymbol{v}_i = \bar{\lambda}_i \tag{2.43}$$

with $\boldsymbol{v}_i \perp \boldsymbol{v}_1 \perp \boldsymbol{v}_2 \perp, \ldots, \perp \boldsymbol{v}_{i-1}$. By considering the most general case $Q' = Q$, the optimal linear projection for which the variance Σ is maximized is defined by the matrix $\mathbf{V}_p = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_Q]$ containing the eigenvectors of the covariance matrix $\mathbf{S}$ corresponding to the Q largest eigenvalues $\bar{\lambda}_1, \ldots, \bar{\lambda}_Q$. Furthermore, since we know that $\bar{\lambda}_1 \geq \bar{\lambda}_2 \geq \ldots \geq \bar{\lambda}_Q$, from Equations (2.38) and (2.42) it is obvious that $\Sigma_1 \geq \Sigma_2 \geq \ldots \geq \Sigma_Q$. The latter equation provides interesting information about the projected space which will prove to be key factors in the proposed damage assessment strategy. In fact, since the variance of the last principal component Σ_Q is the lowest, Q can be set so that all the elements projected along such direction have the dispersion around the mean and form a cluster on the projected space. This observation leads to the so-called *Minor Component Analysis* which, contrary to the classical PCA, searches for those components with lower variance.

2.3.2. A Minor Component Analysis for variance reduction

Given a structure, let's consider a training dataset associated with the structure in its undamaged conditions and a test dataset associated with the structure in unknown conditions (potentially damaged). The training dataset consists of N_{ob}^{tr} observations of the acceleration response time histories at the i^{th} structural DOFs ($i = 1, \ldots, N$) from which the training cepstral coefficient sequences $\mathbf{c}_{tr,i}^{(j)}$ are extracted ($j = 1, \ldots, N_{ob}^{tr}$). Analogously, the test dataset consists of N_{ob}^{te} observations of acceleration response time histories at the i^{th} DOF from which the test cepstral coefficient sequences $\mathbf{c}_{te,i}^{(j')}$ are extracted ($j' = 1, \ldots, N_{ob}^{te}$).

Let's now focus on the training dataset. It is possible to assemble all the Q cepstral coefficients extracted from each of the N_{ob}^{tr} acceleration response time histories recorded at the N structural locations into a matrix $\mathbf{C}^{tr} \in \mathbb{R}^{N^{tr} \times Q}$ where $N^{tr} = N \cdot N_{ob}^{tr}$ so that

$$\mathbf{C}^{tr} = [\mathbf{c}_{tr,1}^{(1)}, \ldots, \mathbf{c}_{tr,1}^{(N_{ob}^{tr})}, \mathbf{c}_{tr,2}^{(1)}, \ldots, \mathbf{c}_{tr,2}^{(N_{ob}^{tr})}, \ldots, \mathbf{c}_{tr,N}^{(1)}, \ldots, \mathbf{c}_{tr,N}^{(N_{ob}^{tr})}]^T \tag{2.44}$$

where the term T denotes the transpose of the matrix consisting of column elements vectors $\mathbf{c}_{tr,i}^{(j)} \in \mathbb{R}^{Q \times 1}$. Let's also define the matrix $\bar{\mathbf{C}} \in \mathbb{R}^{N^{tr} \times Q}$ containing the means of the columns of $\mathbf{C}^{tr}$:

$$\bar{\mathbf{C}} = [\bar{\boldsymbol{c}}, \ldots, \bar{\boldsymbol{c}}]^T \tag{2.45}$$

where

$$\bar{\boldsymbol{c}} = \frac{1}{N^{tr}} \sum_{i=1}^{N} \sum_{j=1}^{N_{ob}^{tr}} \boldsymbol{c}_{tr,i}^{(j)}. \tag{2.46}$$

Obviously, as the rows of $\bar{\mathbf{C}}$ are equal to each other, so the rank of $\bar{\mathbf{C}}$ is equal to 1. If we subtract the matrix $\bar{\mathbf{C}}$ from the matrix $\mathbf{C}^{tr}$ (an operation that transforms the original cepstral coefficient values to zero mean values), this new matrix can be projected onto a new maximum variance based space through a Principal Component Analysis as:

$$[\mathbf{C}^{tr} - \bar{\mathbf{C}}]\mathbf{V}_{pro} = \mathbf{C}_{pro}^{tr} \tag{2.47}$$

where $\mathbf{V}_{pro} \in \mathbb{R}^{Q \times Q}$ is the rotational matrix projecting the matrix $[\mathbf{C}^{tr} - \bar{\mathbf{C}}] \in \mathbb{R}^{N^{tr} \times Q}$ onto a new space in which $\mathbf{C}_{pro}^{tr} \in \mathbb{R}^{N^{tr} \times Q}$ is defined. The columns of the projecting matrix $\boldsymbol{V}_{pro}$ are the eigenvectors of the covariance matrix $\mathbf{S}^{tr}$ which, according to Equation (2.39), is given by:

$$\mathbf{S}^{tr} = \frac{1}{N^{tr}} \sum_{i=1}^{N} \sum_{j=1}^{N_{ob}} (\boldsymbol{c}_{tr,i}^{(j)} - \bar{\boldsymbol{c}})(\boldsymbol{c}_{tr,i}^{(j)} - \bar{\boldsymbol{c}})^T. \tag{2.48}$$

Equation (2.47) can be rewritten as the difference of the projections of the matrices $\mathbf{C}$ and $\bar{\mathbf{C}}$:

$$\mathbf{C}_{pro}^{tr} = \mathbf{C}^{tr}\mathbf{V}_{pro} - \bar{\mathbf{C}}\mathbf{V}_{pro}. \tag{2.49}$$

All the rows of $\bar{\mathbf{C}}$ are identical each others, consequently the projection $\bar{\mathbf{C}}\mathbf{V}_{pro}$ maps the mean vector of the N^{tr} different observations to the same point on the projected space. Multiplying the matrix $\mathbf{C}^{tr}$ by the first column vector of $\mathbf{V}_{pro}$ generates a distribution of projected data points with the maximum variance, multiplying the matrix $\mathbf{C}^{tr}$ by the second column vector of $\mathbf{V}_{pro}$ the variance of the new projected data is reduced. It is understood that, by multiplying the matrix $\mathbf{C}^{tr}$ by the remaining column vectors of the matrix $\mathbf{V}_{pro}$ ordered from 3 to Q the variance of the projected data will decrease. The Q^{th} column vector of $\mathbf{V}_{pro}$ is referred to as *minor component*; the variance of the training cepstral coefficient sequences projected onto this component is the lowest provided by PCA and the projected points, following the N distributions (one for each monitored DOF), overlap making the distributions collapse into one single clustered distribution. Having the cepstral coefficients projected onto a lower dimensional space where they are closely clustered will facilitate the assessment of structural

damage. In fact, changes in the structural properties will cause a variation in the distribution of the cepstral coefficients vectors in the projected space, variation that is more easily caught if the data are clustered together. This section presents a first approach to the *minor component analysis* that will rely on the space provided by the $Q-1^{th}$ and the Q^{th} column vectors of the matrix $\mathbf{V}_{pro}$.

2.3.3. Linear projection for data clustering

Until now, we have presented a procedure to extract the observations of the cepstral coefficient sequences from the measurements of the structural response and to project them on a space with low values of the variance. Now we want to compare the distributions of these projected training (structure in healthy conditions) and test (structure in potentially damaged conditions) datasets. The minor component analysis allows to define a space in which the distribution of the projected data of the matrix $\mathbf{C}^{tr}$ (training dataset) can be represented by a unique clustered probability distribution. The damage assessment strategy based on statistical pattern recognition requires two steps: 1) the training phase and 2) the test phase. In the training phase, a model representative of the structure in its baseline conditions (assumed undamaged) is obtained based on the statistics of the projection of the training cepstral coefficient sequences onto the minor components. In the test phase, a test dataset representative of the structure in unknown conditions (potentially damaged) is used for the extraction of the test cepstral coefficient sequences which are projected onto the minor components (defined in the training phase). The training and test distributions in the minor components are compared and the presence of damage is assessed. During the training phase the following tasks have been accomplished:

1. gathering observations of the cepstral coefficients vectors at every i^{th} monitoring location ($i = 1,\ldots,N$);
2. computing the matrices $\mathbf{C}^{tr}$, $\bar{\mathbf{C}}$ and $\mathbf{V}_{pro}$ which describe the transformation provided by PCA;
3. projecting the cepstral coefficients vectors onto the new space so to obtain $\mathbf{C}^{tr}_{pro}$.

Once the matrices $\bar{\mathbf{C}}$ and $\mathbf{V}_{pro}$ are obtained, it is possible to test whether the system is damaged or not by projecting new cepstral coefficient vectors extracted in the test phase onto the space defined by $\bar{\mathbf{C}}$ and $\mathbf{V}_{pro}$.

Let's now consider the test dataset and the matrix $\mathbf{C}^{te} \in \mathbb{R}^{N^{te}\times Q}$ containing the N^{te}_{ob} observations of cepstral coefficient sequences $\mathbf{c}^{(j')}_{te,i}$ ($j' = 1,\ldots,N^{te}_{ob}$) at each monitored i^{th} location ($N^{te} =$

$N \cdot N_{ob}^{te}$). Analogously to Equation (2.47), $\mathbf{C}^{te}$ can be written as:

$$\mathbf{C}^{te} = [\mathbf{c}_{te,1}^{(1)}, \ldots, \mathbf{c}_{tr,1}^{(N_{ob}^{te})}, \mathbf{c}_{te,2}^{(1)}, \ldots, \mathbf{c}_{te,2}^{(N_{ob}^{te})}, \ldots, \mathbf{c}_{te,N}^{(1)}, \ldots, \mathbf{c}_{te,N}^{(N_{ob}^{te})}]^T. \tag{2.50}$$

The projection of the matrix $\mathbf{C}^{te}$ onto the new space provided by the PCA is performed, according to Equation (2.47). However, the number of training observations N_{ob}^{tr} may be different from the number of test observations N_{ob}^{te} and, consequently the matrix $[\mathbf{C}^{te} - \bar{\mathbf{C}}]$ cannot be computed. By recalling that each of the row vectors of the matrix $\bar{\mathbf{C}}$ is equal to the vector $\bar{\boldsymbol{c}}$ transposed, it is possible to define a matrix $\bar{\mathbf{C}}^{te}$ whose row vectors are equal to the vector $\bar{\boldsymbol{c}}$ transposed and whose dimensions match the ones of $\mathbf{C}^{te}$, i.e. $\bar{\mathbf{C}}^{te} \in \mathbb{R}^{N^{te} \times Q}$. This procedure allows to obtain the projection of the test cepstral coefficient sequences onto the space represented by the principal components, defined in the training phase, as:

$$[\mathbf{C}^{te} - \bar{\mathbf{C}}^{te}]\mathbf{V}_{pro} = \mathbf{C}_{pro}^{te}. \tag{2.51}$$

Before focusing on the minor components, let's consider the elements in the matrix $[\mathbf{C}^{te} - \bar{\mathbf{C}}^{te}]$. By recalling Equation (2.35), each element of the j^{th} observed cepstral coefficient sequence extracted at the i^{th} location can be represented by the sum of two terms: $\gamma_{i,q}^{(j)}$ and θ_q ($q = 1, \ldots, Q$). In order to distinguish these terms computed in the training phase from those computed in the test phase, it is convenient to use the superscript 'tr' and 'te' so that $\gamma_{i,q}^{tr,(j)}$ and $\theta_q^{tr,(j)}$ will be referred to the training phase and $\gamma_{i,q}^{te,(j')}$ and $\theta_q^{te,(j')}$ to the test phase. Anyway, each of the $q = 1, \ldots, Q$ terms $\theta_q^{tr,(j')}$ depends only on the structural properties which are assumed to be constant for the training dataset ($\theta_q^{tr,(j)} = \theta_q^{tr}$ for $j = 1, \ldots, N_{ob}^{tr}$) and for the test dataset ($\theta_q^{te,(j')} = \theta_q^{te}$ for $j' = 1, \ldots, N_{ob}^{te}$).

Considering the training dataset, the q^{th} cepstral coefficient of the j^{th} observation of the cepstral coefficient sequence at the i^{th} monitored DOF can be represented as $c_{i,q}^{tr,(j)} = \gamma_{i,q}^{tr,(j)} + \theta_q^{tr}$. Analogously, considering the test dataset, the q^{th} cepstral coefficient of the j'^{th} observation of the cepstral coefficient sequence at the i^{th} monitored DOF is given by $c_{i,q}^{te,(j')} = \gamma_{i,q}^{te,(j')} + \theta_q^{te}$. By using this new formulation to represent the cepstral coefficient sequences, the matrix $\mathbf{C}^{tr}$ in

Equation (2.44) can be rewritten as:

$$\mathbf{C}^{tr} = \begin{bmatrix} \gamma_{1,1}^{tr,(1)} + \theta_1^{tr} & \dots & \gamma_{1,Q}^{tr,(1)} + \theta_Q^{tr} \\ \vdots & \ddots & \vdots \\ \gamma_{1,1}^{tr(N_{ob}^{tr})} + \theta_1^{tr} & \dots & \gamma_{1,Q}^{tr,(N_{ob}^{tr})} + \theta_Q^{tr} \\ \vdots & \vdots & \vdots \\ \gamma_{N,1}^{tr,(1)} + \theta_1^{tr} & \dots & \gamma_{N,Q}^{tr,(1)} + \theta_Q^{tr} \\ \vdots & \ddots & \vdots \\ \gamma_{N,1}^{tr,(N_{ob}^{tr})} + \theta_1^{tr} & \dots & \gamma_{N,Q}^{(tr,N_{ob}^{tr})} + \theta_Q^{tr} \end{bmatrix} \tag{2.52}$$

and, by recalling Equation (2.45):

$$\bar{\mathbf{C}} = \begin{bmatrix} \frac{1}{N^{tr}} \sum_{j=1}^{N_{ob}^{tr}} \sum_{i=1}^{N} \gamma_{i,1}^{tr,(j)} + \theta_1^{tr} & \dots & \frac{1}{Ntr} \sum_{j=1}^{N_{ob}^{tr}} \sum_{i=1}^{N} \gamma_{i,Q}^{tr,(j)} + \theta_Q^{tr} \\ \vdots & \ddots & \vdots \\ \frac{1}{N^{tr}} \sum_{j=1}^{N_{ob}^{tr}} \sum_{i=1}^{N} \gamma_{i,1}^{tr,(j)} + \theta_1^{tr} & \dots & \frac{1}{N^{tr}} \sum_{j=1}^{N_{ob}^{tr}} \sum_{i=1}^{N} \gamma_{i,Q}^{tr,(j)} + \theta_Q^{tr} \\ \vdots & \vdots & \vdots \\ \frac{1}{N^{tr}} \sum_{j=1}^{N_{ob}^{tr}} \sum_{i=1}^{N} \gamma_{i,1}^{tr,(j)} + \theta_1^{tr} & \dots & \frac{1}{N^{tr}} \sum_{j=1}^{N_{ob}^{tr}} \sum_{i=1}^{N} \gamma_{i,Q}^{tr,(j)} + \theta_Q^{tr} \\ \vdots & \ddots & \vdots \\ \frac{1}{N^{tr}} \sum_{j=1}^{N_{ob}^{tr}} \sum_{i=1}^{N} \gamma_{i,1}^{tr,(j)} + \theta_1^{tr} & \dots & \frac{1}{N^{tr}} \sum_{j=1}^{N_{ob}^{tr}} \sum_{i=1}^{N} \gamma_{i,Q}^{tr,(j)} + \theta_Q^{tr} \end{bmatrix}. \tag{2.53}$$

By substituting the relations obtained in Equations (2.52) and (2.53) into Equation (2.47), the

projected matrix $\mathbf{C}_{pro}^{tr}$ can be written as:

$$\mathbf{C}_{pro}^{tr} = \begin{bmatrix} \gamma_{1,1}^{tr,(1)} - \frac{1}{N^{tr}} \sum_{j=1}^{N_{ob}^{tr}} \sum_{i=1}^{N} \gamma_{i,1}^{tr,(j)} & \cdots & \gamma_{1,Q}^{tr,(1)} - \frac{1}{Ntr} \sum_{j=1}^{N_{ob}^{tr}} \sum_{i=1}^{N} \gamma_{i,Q}^{tr,(j)} \\ \vdots & \ddots & \vdots \\ \gamma_{1,1}^{tr(N_{ob}^{tr})} - \frac{1}{N^{tr}} \sum_{j=1}^{N_{ob}^{tr}} \sum_{i=1}^{N} \gamma_{i,1}^{tr,(j)} & \cdots & \gamma_{1,Q}^{tr,(N_{ob}^{tr})} - \frac{1}{N^{tr}} \sum_{j=1}^{N_{ob}^{tr}} \sum_{i=1}^{N} \gamma_{i,Q}^{tr,(j)} \\ \vdots & \vdots & \vdots \\ \gamma_{N,1}^{tr,(1)} - \frac{1}{N^{tr}} \sum_{j=1}^{N_{ob}^{tr}} \sum_{i=1}^{N} \gamma_{i,1}^{tr,(j)} & \cdots & \gamma_{N,Q}^{tr,(1)} - \frac{1}{N^{tr}} \sum_{j=1}^{N_{ob}^{tr}} \sum_{i=1}^{N} \gamma_{i,Q}^{tr,(j)} \\ \vdots & \ddots & \vdots \\ \gamma_{N,1}^{tr,(N_{ob}^{tr})} - \frac{1}{N^{tr}} \sum_{j=1}^{N_{ob}^{tr}} \sum_{i=1}^{N} \gamma_{i,1}^{tr,(j)} & \cdots & \gamma_{N,Q}^{(tr,N_{ob}^{tr})} - \frac{1}{N^{tr}} \sum_{j=1}^{N_{ob}^{tr}} \sum_{i=1}^{N} \gamma_{i,Q}^{tr,(j)} \end{bmatrix} \mathbf{V}_{pro}. \quad (2.54)$$

Equation (2.54) shows a very important result that will be fundamental for the damage assessment procedure. In fact, the Q components θ_q, which depend purely on the structural properties, disappear and, as a consequence of the PCA, the last two columns of the matrix $\mathbf{C}_{pro}^{tr}$ represent the projection of the training cepstral coefficient sequences onto the $Q-1^{th}$ and Q^{th} principal components (minor components). The projection of the training data onto the minor components aims to dramatically reduce the variance between cepstral coefficient sequences extracted at different structural locations from dynamic response generated by different excitation sources. In this way, as long as the system persists in its undamaged conditions, the projection of newly observed cepstral coefficient sequences onto the minor components can be represented by a unique clustered probability distribution.

The projection of the test data (N_{ob}^{te} observations of the test cepstral coefficient sequences) onto the principal components represented by the matrix $\mathbf{C}_{pro}^{te}$, such a matrix $\mathbf{C}_{pro}^{te}$ can be written as:

$$
\begin{aligned}
\mathbf{C}_{pro}^{te} &= [\mathbf{C}^{te} - \bar{\mathbf{C}}^{te}]\mathbf{V}_{pro} \\
&= \begin{bmatrix} \theta_1^{te} - \theta_1^{tr} & \dots & \theta_Q^{te} - \theta_Q^{tr} \\ \vdots & \ddots & \vdots \\ \theta_1^{te} - \theta_1^{tr} & \dots & \theta_Q^{te} - \theta_Q^{tr} \\ \vdots & \vdots & \vdots \\ \theta_1^{te} - \theta_1^{tr} & \dots & \theta_Q^{te} - \theta_Q^{tr} \\ \vdots & \ddots & \vdots \\ \theta_1^{te} - \theta_1^{tr} & \dots & \theta_Q^{te} - \theta_Q^{tr} \end{bmatrix} \mathbf{V}_{pro} \\
&+ \begin{bmatrix} \gamma_{1,1}^{te,(1)} - \frac{1}{N^{tr}} \sum_{j=1}^{N_{ob}^{tr}} \sum_{i=1}^{N} \gamma_{i,1}^{tr,(j)} & \cdots & \gamma_{1,Q}^{te,(1)} - \frac{1}{Ntr} \sum_{j=1}^{N_{ob}^{tr}} \sum_{i=1}^{N} \gamma_{i,Q}^{tr,(j)} \\ \vdots & \ddots & \vdots \\ \gamma_{1,1}^{te(N_{ob}^{te})} - \frac{1}{N^{tr}} \sum_{j=1}^{N_{ob}^{tr}} \sum_{i=1}^{N} \gamma_{i,1}^{tr,(j)} & \cdots & \gamma_{1,Q}^{te,(N_{ob}^{te})} - \frac{1}{N^{tr}} \sum_{j=1}^{N_{ob}^{tr}} \sum_{i=1}^{N} \gamma_{i,Q}^{tr,(j)} \\ \vdots & \vdots & \vdots \\ \gamma_{N,1}^{te,(1)} - \frac{1}{N^{tr}} \sum_{j=1}^{N_{ob}^{tr}} \sum_{i=1}^{N} \gamma_{i,1}^{tr,(j)} & \cdots & \gamma_{N,Q}^{te,(1)} - \frac{1}{N^{tr}} \sum_{j=1}^{N_{ob}^{tr}} \sum_{i=1}^{N} \gamma_{i,Q}^{tr,(j)} \\ \vdots & \ddots & \vdots \\ \gamma_{N,1}^{te,(N_{ob}^{te})} - \frac{1}{N^{tr}} \sum_{j=1}^{N_{ob}^{tr}} \sum_{i=1}^{N} \gamma_{i,1}^{tr,(j)} & \cdots & \gamma_{N,Q}^{(te,N_{ob}^{te})} - \frac{1}{N^{tr}} \sum_{j=1}^{N_{ob}^{tr}} \sum_{i=1}^{N} \gamma_{i,Q}^{tr,(j)} \end{bmatrix} \mathbf{V}_{pro}.
\end{aligned}
\tag{2.55}
$$

By looking at Equation (2.55) some considerations about the projected matrix $\mathbf{C}_{pro}^{te}$ can be pointed out. Such a matrix can be represented by the sum of two terms. If the structural system (in the unknown state) is working in its standard (undamaged) conditions, then the first term disappears ($\theta_q^{te} = \theta_q^{tr}$ for $q = 1, \dots, Q$) and the second term provides a projected matrix analogous to the one in Equation (2.54): the projection of the cepstral coefficient sequences is still clustered in the minor components. On the contrary, if damage occurs in the structure ($\theta_q^{te} \neq \theta_q^{tr}$ for $q = 1, \dots, Q$), both the terms in Equation 2.55 cause a variation of the distribution of the projected cepstral coefficient sequences in the minor components (last two column vectors of the matrix $\mathbf{C}_{pro}^{te}$). Whereas the first term in Equation (2.55) has an heavy impact on $\mathbf{C}_{pro}^{te}$, the second term provides a projection similar to the one obtained in Equation (2.54) since the distribution of the zeros in $\gamma_{i,q}$ changes only slightly because of the damage. By recalling that the

column vectors of the projection matrix $\mathbf{V}_{pro}$ are the eigenvectors of the covariance matrix $\mathbf{S}^{tr}$, the multiplication of the training cepstral coefficient sequences by its last two columns (projection onto the minor components) is a linear transformation that makes all the N distributions of $\gamma_{i,q}^{(tr)}$ collapse into one. When damage occurs in the system, the column vectors of the matrix $\mathbf{V}_{pro}$ do not represent the eigenvectors for the covariance matrix of the test cepstral coefficient sequences and they can be interpreted as simple linear projection vectors. Hence, because of the second term in Equation (2.55), the projection of the cepstral coefficient sequences onto the minor components is not guaranteed anymore to follow a single clustered probability distribution. Depending on the entity of damage, the first term in Equation (2.55) causes a common shift of the projected cepstral coefficient sequences in the minor components. Consequently, the second term in Equation (2.55) makes the *N* distributions diverge from each other.

The effects of damage on the minor components (last two components of $\mathbf{C}_{pro}^{te}$) are investigated in the numerical simulations presented in the following sections.

2.4. Numerical analysis of 2 minor components

The proposed damage assessment methodology relies on the analysis of the projection of the training and test cepstral coefficient sequences onto the minor components. When no damage occurs in the system, the distribution of the projected test cepstral coefficient sequences onto the minor components is assumed to be Gaussian as well as the the distribution of the projected training cepstral coefficient sequences. In this section, the Gaussianity of the distributions is verified through the Kolmogorov-Smirnov test (ks-test). In addition, since the occurrence of damage is linked to a variation in the clustered distribution, the Squares Mahalanobis Distance, accounting for the variances in multivariate analysis, is considered a valid tool to detect the presence of the damage in the data.

2.4.1. 8-DOF lumped mass model

The system analyzed in this section is a lumped mass model of an eight-story shear-type building, shown in figure 2.2, that is instrumented at every level, e.g. floor or degree of freedom (DOF), with an accelerometer (full set of sensors). In its baseline condition (undamaged), the model is characterized by horizontal springs of stiffness $k_i = 25000$ N/m and by eight mass elements $m_i = 1$ kg ($i = 1,\ldots,8$). The system is modelled as having modal damping with a damping factor of $\xi = 1\%$ for each of the 8 vibration modes. The excitation is provided by

external forces applied at various locations via zero-order-hold (ZOH) for 100 sec duration with a time sampling of 0.01 seconds. Such forces are zero-mean Gaussian white noise signals (standard normal distribution) whose magnitudes are scaled, at each floor, by a random factor given by a uniform probability distribution in the range 0-100 N. The training dataset consists of $N_{ob}^{tr} = 500$ observations of the acceleration response time histories (100 sec duration with a time sampling of 0.01 seconds) recorded at the $N = 8$ DOFs so that $N^{tr} = 500 \cdot 8 = 4000$ records in total are available.

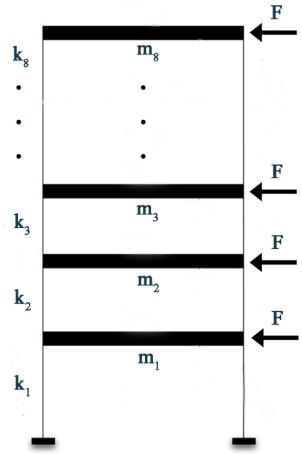


Fig. 2.2. *8-DOFs shear type system*

The cepstral coefficient sequences are extracted from each of the 4000 records: the first cepstral coefficient of the sequence ($q = 0$) is discarded since it's related only to the input excitation and only the following $Q = 50$ cepstral coefficients are considered. Even though the number of cepstral coefficients to be considered is arbitrary, the user still has to make sure that only coefficients providing redundant information during this dimensionality reduction are eliminated.

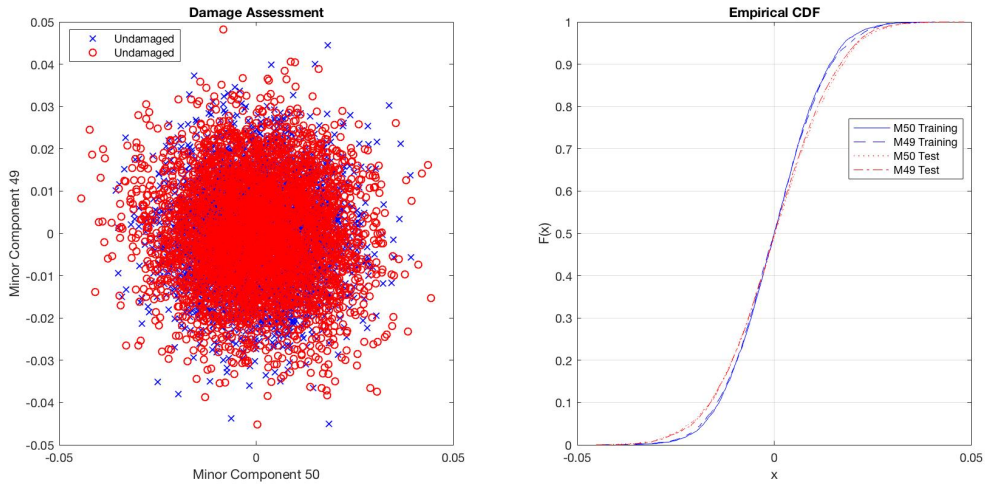
Structural damage assessment based on two minor components

The space represented by the minor components is defined during the training phase relying on the 4000 records of cepstral coefficient sequences related to the system in undamaged configuration. Out of 50 components, the components 49 and 50 have been arbitrarily selected as the minor components. The reason behind this choice is simply given by the fact that those are the ones having the lowest variance among all. Furthermore, considering two components instead one yields a better visualization and interpretation of the projected data. The training phase ends with a 2-D representation of the projected training cepstral coefficient sequences forming

a unique cluster in the minor components.

Several structural damage scenarios for the 8-DOF system are analyzed and commented. Those damage scenarios are meant to represent the system under increasing levels of damage. Damage has been simulated through a drop of stiffness of the spring elements connecting two adjacent degrees of freedom. Each damage scenario is associated with a different test phase. During the test phase, new cepstral coefficients are generated and projected in order to compare the projected test data with the projected training data.

The first test is performed in order to confirm that, if no damage occurs in the structure, the distributions of the projected training and test cepstral coefficient sequences onto the minor components match each other. The test dataset consists in $N_{ob}^{te} = 500$ observations of the 8 acceleration response time histories from which the cepstral coefficient sequences are extracted when the structure is in its baseline condition (undamaged). As expected, Figure 2.3 (a) shows that the distribution of the cepstral coefficient sequences computed in the training phase and projected onto the minor components (blue crosses) forms one single cluster. Since no damage occurs in the structure, also the distribution of the projected cepstral coefficient sequences for the test dataset (red circles) forms a unique cluster overlapping the one defined in the training phase.



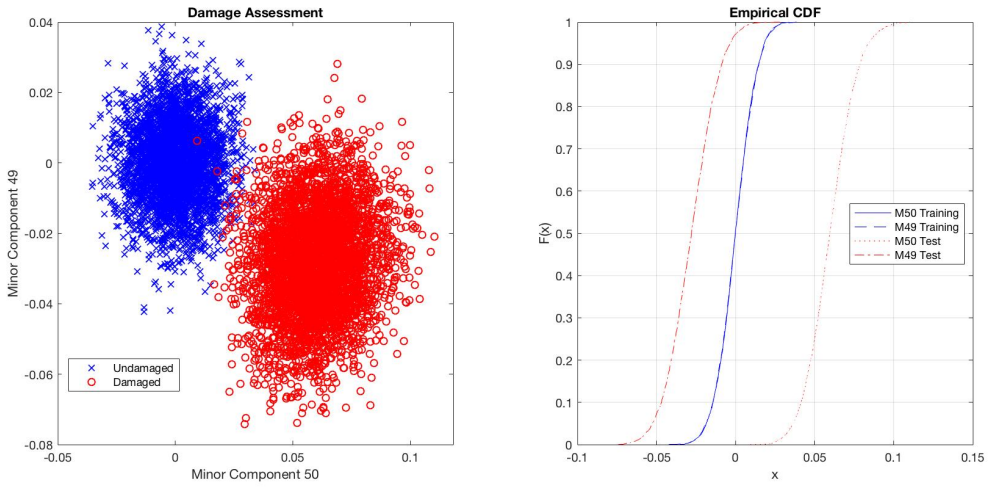
(a) **(b)**

Fig. 2.3. *Undamaged System Test: Undamaged Training and Undamaged Test Distributions (a), Minor components 49 and 50 cumulative distributions for Training and Test (b)*

Figure 2.3 (b) shows the cumulative distribution function (CDF) for the training and test datasets by using the variable 'x' on the horizontal axis to represent the domains of both component 49 and component 50 in Figure 2.3 (a). Those plots suggest the Gaussianity of the

distributions (confirmed by a ks-test at 5% of significant level) and lead to the conclusion that no damage occurs in the structure (the training and test distributions match each other).

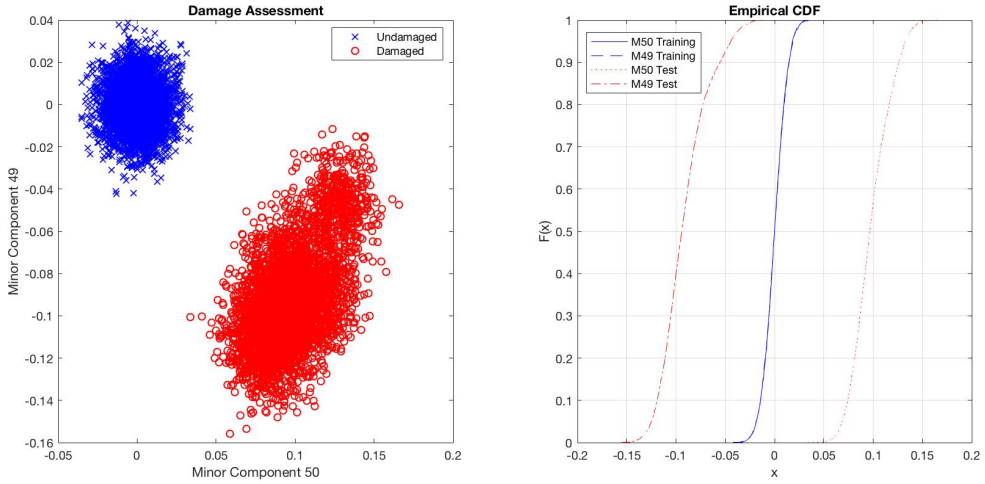
A new damage scenario is presented. A local damage has been simulated through a 10% drop of stiffness of the spring element connecting the DOFs 2 and 3. A set of $N_{ob}^{te} = 500$ observations of the cepstral coefficient sequences is generated in the test phase and projected onto the minor components. The distributions of the training and test projected cepstral coefficient sequences are shown in Figure 2.4. Contrarily to the previous case, the training and test distributions are not overlapping and the test CDFs in the CDF plot diverge from the training CDFs. Consequently, the presence of the damage can be assessed through the Squared Mahalanobis Distance in a novelty detection approach.



(a) **(b)**

Fig. 2.4. *Undamaged vs Damaged - 10% stiffness drop*

A new test dataset of $N_{ob}^{te} = 500$ observations of the cepstral coefficient sequences has been generated after introducing a 20% drop of the stiffness of the spring element connecting DOFs 2 and 3. The training and test cepstral coefficient sequences projected onto the minor components are shown in Figure 2.5.



(a) **(b)**

Fig. 2.5. *Undamaged vs Damaged - 20% stiffness drop*

Figure 2.5 confirms that, as the damage increases, the test projected cepstral coefficient sequences distribution diverge further from the training clustered distribution. Furthermore, as consequence of the damage, the $N = 8$ test distributions coming from the 8 different locations do not collapse into one; as expected, the test distribution is not guaranteed to be represented by a single clustered distribution. This behaviour is emphasized in the last damage scenario in which the drop in stiffness (between degrees of freedom 2 and 3) is set to be at 25% to simulate a considerable local damage. A test dataset of $N_{ob}^{te} = 500$ cepsral coefficients sequences is generate at every DOF. Figure 2.6 shows the two clusters, each generated by the projection of the cepstral coefficient sequences extracted at the $N = 8$ DOFs. The cluster generated by the projection of the cepstral coefficient sequences of the training dataset (undamaged) is shown to be Gaussian and can be statistically represented by mean and standard deviation. On the contrary, the distribution arising from the projection of the test cepstral coefficient sequences (damaged) is not Gaussian (ks-test). Figure 2.6 represents a remarkable example of the effects of damage on the projection of the cepstral coefficient sequences onto the minor components. Hence, Equation (2.55) can be analyzed to show how the theoretical formulation matches the numerical results in Figure 2.6. The first term in Equation (2.55) implies that the presence of structural damage causes each of the $N = 8$ test distributions to diverge from the training distribution by the same quantity. The second term in Equation (2.55) causes the $N = 8$ test distributions to diverge from each others. Figure 2.6 confirms that the common shift of the 8 test distributions from the 8 training distributions can be used as a valid indication of the

presence of damage in a novelty detection approach based on the SMD.

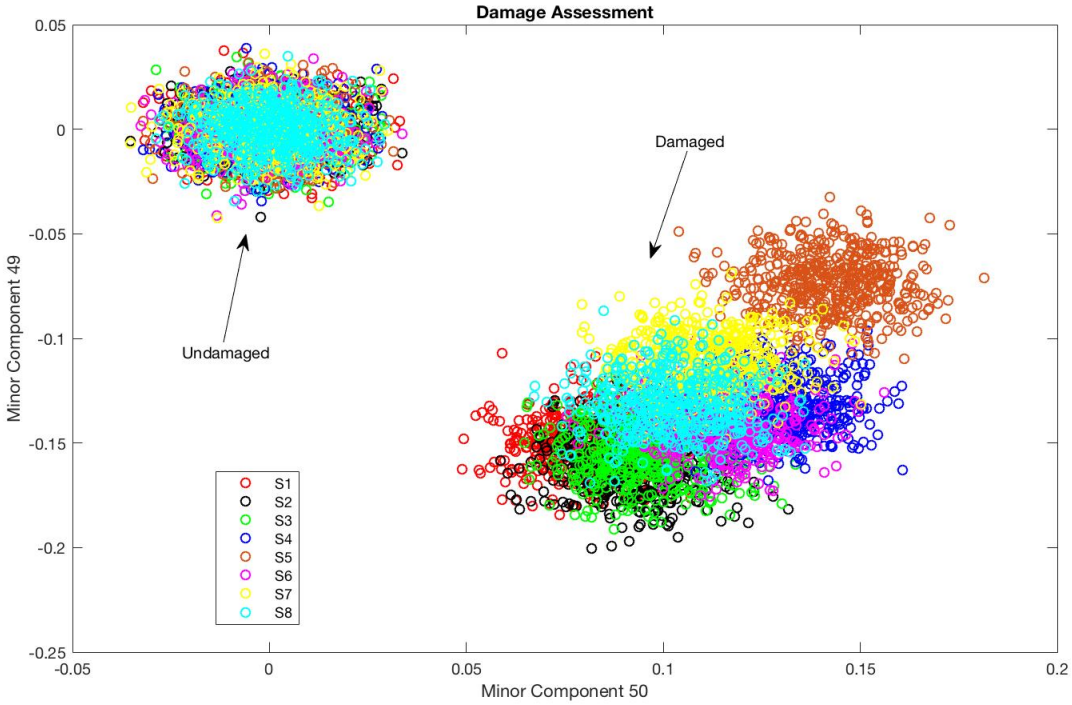


Fig. 2.6. *Projected data distributions belonging to sensors S*

Finally, the Squared Mahalanobis Distance between the training and test distributions over the 50^{th} principal component (minor component) is computed. For each box in Figure 2.7, the central red line represents the median of the SMD, the blue edges of the box represent the 25^{th} and 75^{th} percentiles and the black whiskers represent the range covering the 99% of the observations of the test dataset. A set of $N_{ob}^{te} = 500$ observations of the cepstral coefficient sequences has been generated as three test datasets for the structure in undamaged condition and the SMD has been computed exhibiting similar values. A total of $N_{ob}^{te} = 500$ observations of the cepstral coefficient sequences have been generated as test datasets simulating damage through a reduction of the stiffness between DOFs 2 and 3 varying from 5% to 25%. The SMD is computed and presented in Figure 2.7. As the simulated damage increases, the median value of the Squared Mahalanobis Distance increases too.

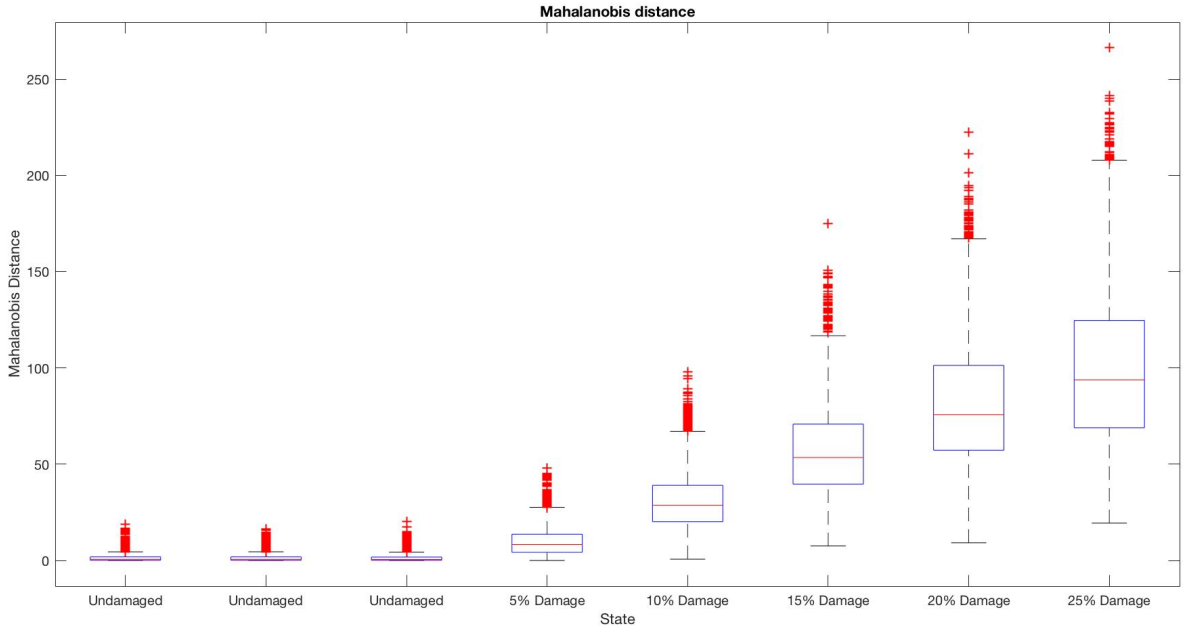


Fig. 2.7. *Squared Mahalanobis Distance vs Damage*

The information about the statistics of the distributions over different damage scenarios for the 49^{th} and 50^{th} principal components are reported in Table 2.1. As long as the test dataset is generated from the dynamic response for the system in undamaged state, the ks-test identifies the distributions of the projected cepstral coefficient sequences as Gaussians. As discussed, this is not guaranteed for test datasets generated from the system in damaged conditions. In fact, as the drop in stiffness between DOFs 2 and 3 reaches the 15%, the distributions for both component 49 and component 50 are no longer Gaussians. Furthermore, the distributions of the projection of the test cesptral coefficients sequences extracted for the system in undamaged conditions are theoretically zero-mean. The presence of the damage not only makes their means vary, but also causes a slight increment in their standard deviations. In the last column of Table 2.1, it can be observed how the median of the SMD for the 50^{th} principal component (minor component) dramatically increases because of damage.

State	$kstest_{49}$	$mean_{49}$	std_{49}	$kstest_{50}$	$mean_{50}$	std_{50}	Mahal
Training	Gaussian	0	0.01137	Gaussian	0	0.01099	0.45799
Undamaged	Gaussian	0.00047	0.01312	Gaussian	-0.00006	0.01311	0.65049
Undamaged	Gaussian	0.00013	0.01274	Gaussian	0.00032	0.01308	0.65736
Undamaged	Gaussian	-0.00003	0.01257	Gaussian	0.00095	0.01272	0.60125
5% Damage	Gaussian	-0.01063	0.01361	Gaussian	0.03212	0.01351	8.30698
10% Damage	Gaussian	-0.02798	0.01491	Gaussian	0.05985	0.01455	28.75355
15% Damage	No Gaussian	-0.05562	0.01882	No Gaussian	0.0822	0.01704	53.47868
20% Damage	No Gaussian	-0.09113	0.024	No Gaussian	0.09839	0.01965	75.721
25% Damage	No Gaussian	-0.13174	0.02972	No Gaussian	0.10893	0.02239	93.7746

Table 2.1. *Statistics over different damage scenarios.*

2.4.2. Effect of measurement noise

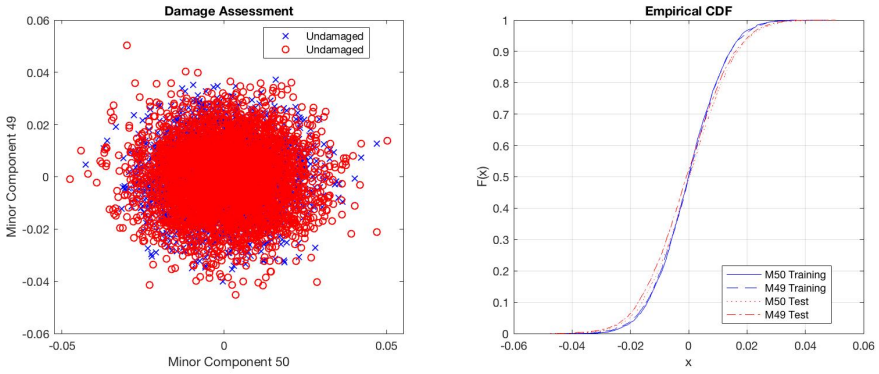
A recurrent problem, when dealing with real data, is represented by the presence of external disturbances. Hence, in reality, accelerometers are constantly subjected to electrical noise, environmental disturbances etc.. For this reason, in this section, the robustness of the proposed damage sensitive features to external noise is tested.

According to the standard assumptions in literature, disturbances have been modeled as zero-mean Gaussian white noise. Although external disturbances are never white, this modelization represents a good compromise between the mathematical tractability and the representation of reality. In the following simulations, the RMS (root mean square) of the Gaussian white noise has been set equal to a certain percentage of the RMS of the output. Such percentage has been chosen to be the same for both the training and test datasets.

The system is represented by the 8 DOF shear-type model in section 2.4.1. The training dataset consists of $N_{ob}^{tr} = 500$ observations of the acceleration response time histories which have been generated from the model in its baseline conditions, according to section 2.4.1. Analogously, $N_{ob}^{te} = 500$ observations of the acceleration response time histories have been simulated for three different damage scenarios: 1) no damage 2) damage simulated by lowering the stiffness of the spring element connecting DOFs 2 and 3 by 5% 3) damage simulated by lowering the stiffness of the spring element connecting DOFs 2 and 3 by 15%. Based on these four datasets (1 training and 3 test), two levels of measurement noise have been analyzed.

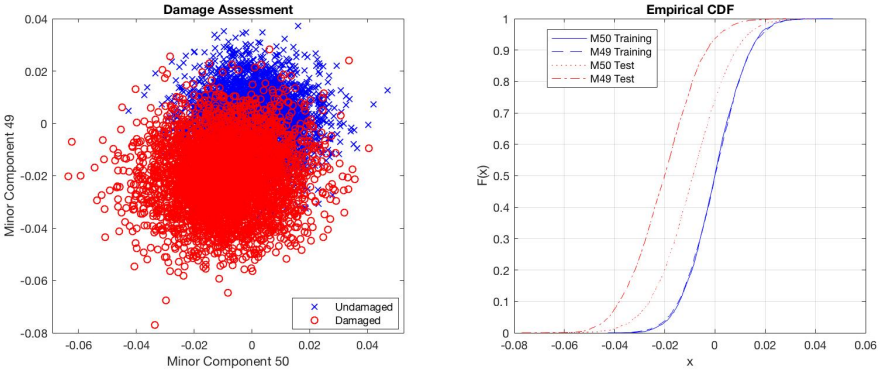
A 5% RMS measurement noise has been applied to the acceleration response time histories from which the cepstral coefficient sequences are extracted. The projection of the cepstral coefficient sequences on the 49^{th} and 50^{th} principal components of the training dataset is shown in Figure 2.8 for the damage scenario 1), in Figure 2.9 for the damage scenario 2), in Figure

2.10 for the damage scenario 3).



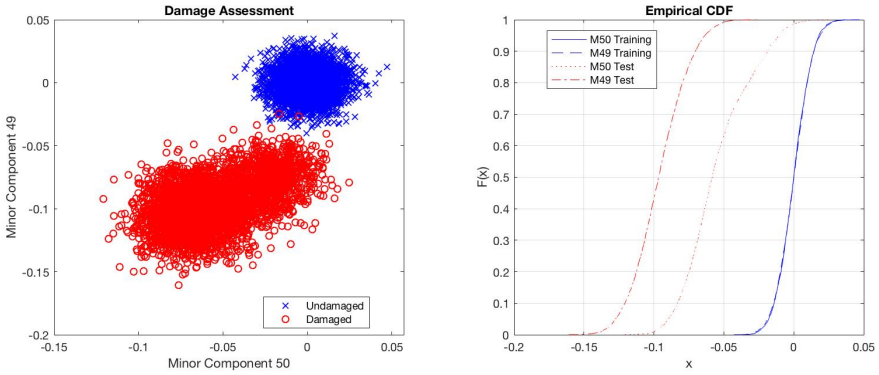
(a) **(b)**

Fig. 2.8. *Undamaged System Test - 5% Noise*



(a) **(b)**

Fig. 2.9. *Undamaged vs Damaged - 5% stiffness drop - 5% Noise*



(a) **(b)**

Fig. 2.10. *Undamaged vs Damaged - 15% stiffness drop - 5% Noise*

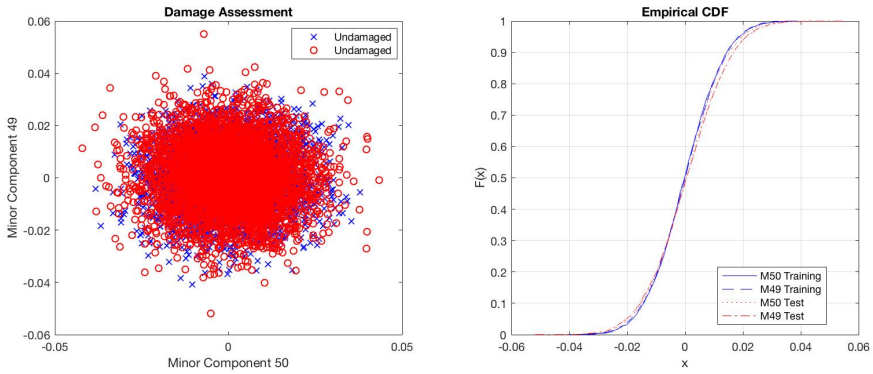
For sake of completeness, multiple damage scenarios ($N_{ob}^{te} = 500$ observations for each of them) have been tested and the statistics of the projected cepstral coefficient sequences distributions on the 49^{th} and 50^{th} principal components of the training dataset have been reported

in Table 2.2. The first column in Table 2.2 describes the state (conditions) of the structure in presence of a 5% RMS measurement noise. For each damage scenario, damage has been simulated by a drop of the stiffness of the spring element connecting DOFs 2 and 3 according to the description in the first column of Table 2.2. Results are similar to the ones provided in Table 2.1.

State	$kstest_{49}$	$mean_{49}$	std_{49}	$kstest_{50}$	$mean_{50}$	std_{50}	Mahal
Training	Gaussian	0	0.01135	Gaussian	0	0.01106	0.45767
Undamaged	Gaussian	-0.00064	0.01299	Gaussian	0.00039	0.01284	0.62332
Undamaged	Gaussian	-0.00064	0.01273	Gaussian	0.0006	0.01298	0.6785
Undamaged	Gaussian	-0.00092	0.01289	Gaussian	0.00095	0.01295	0.63943
5% Damage	Gaussian	-0.02032	0.01351	Gaussian	-0.00885	0.01367	1.0052
10% Damage	Gaussian	-0.05375	0.01536	No Gaussian	-0.02909	0.01704	7.28188
15% Damage	Gaussian	-0.09608	0.0193	No Gaussian	-0.05473	0.02393	27.83511
20% Damage	No Gaussian	-0.1454	0.02513	No Gaussian	-0.08317	0.03254	65.91477
25% Damage	No Gaussian	-0.18804	0.03136	No Gaussian	-0.10928	0.03892	111.74751

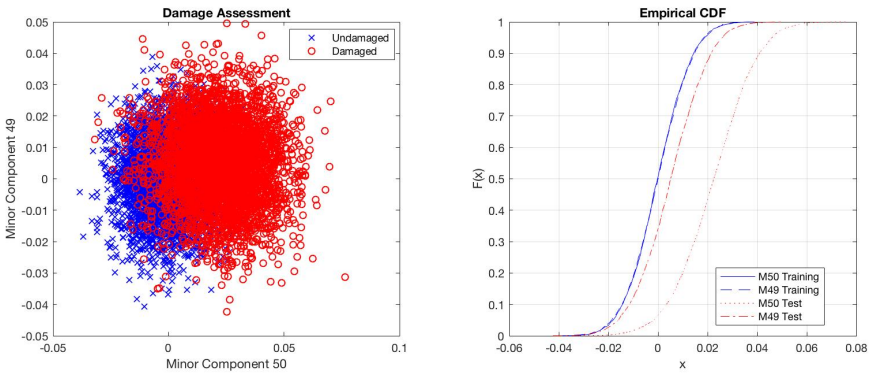
Table 2.2. *Statistics over different damage scenarios - 5% Noise*

A 15% RMS measurement noise has been applied to the acceleration response time histories from which the cepstral coefficient sequences are extracted. The projection of the cepstral coefficient sequences on the 49^{th} and 50^{th} principal components of the training dataset is shown in Figure 2.11 for the damage scenario 1), in Figure 2.12 for the damage scenario 2), in Figure 2.13 for the damage scenario 3).



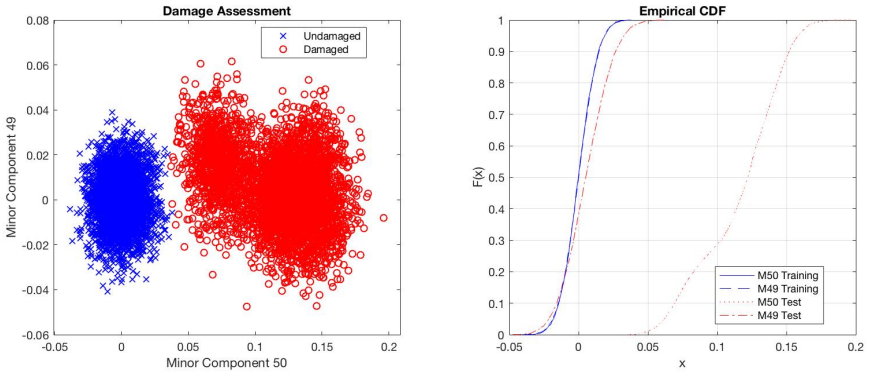
(a) **(b)**

Fig. 2.11. *Undamaged System Test - 15% Noise*



(a) **(b)**

Fig. 2.12. *Undamaged vs Damaged - 5% stiffness drop - 15% Noise*



(a) **(b)**

Fig. 2.13. *Undamaged vs Damaged - 15% stiffness drop - 15% Noise*

Once again, multiple damage scenarios ($N_{ob}^{te} = 500$ observations for each of them) for a measurement noise of 15% have been tested and the statistics of the distributions of the projected cepstral coefficient sequences onto the 49^{th} and 50^{th} principal components of the training dataset have been reported in Table 2.3.

State	kstest$_{49}$	mean$_{49}$	std$_{49}$	kstest$_{50}$	mean$_{50}$	std$_{50}$	Mahal
Training	Gaussian	0	0.01136	Gaussian	0	0.01118	0.43108
Undamaged	Gaussian	0.0006	0.01263	Gaussian	0.00019	0.0119	0.49641
Undamaged	Gaussian	0.00105	0.01271	Gaussian	0.00044	0.01215	0.53669
Undamaged	Gaussian	0.00096	0.01251	Gaussian	0.00054	0.01231	0.53854
5% Damage	Gaussian	0.00508	0.01284	Gaussian	0.0223	0.01467	4.04444
10% Damage	Gaussian	0.00634	0.01379	No Gaussian	0.06281	0.02082	34.59648
15% Damage	Gaussian	0.00533	0.01653	No Gaussian	0.11596	0.03059	119.56165
20% Damage	No Gaussian	0.00396	0.02215	No Gaussian	0.17749	0.04216	273.96921
25% Damage	No Gaussian	0.00286	0.02806	No Gaussian	0.23698	0.05143	478.19296

Table 2.3. *Statistics over different damage scenarios - 15% Noise*

At this point, it is worthy to make some remarkable observations. Let's focus on the struc-

tural state marked as 'Training' in Table 2.2 and Table 2.3. It is clear how the statistical properties described in the two tables are really similar to each other as well as the Squared Mahalanobis Distance indicated in the last column of the tables. This phenomenum can be easily explained. The increment in the RMS of the output noise randomly affects the cepstral coefficient sequence at each degree of freedom of the system. However, the principal components 49 and 50 (minor components) are set by the PCA to be those components with the lowest variance and so the least affected by the external disturbances. For that reason, the effect due to the increment of the level of measurement noise is more evident in the principal components rather than in the minor components.

2.4.3. Comparison with traditional methods

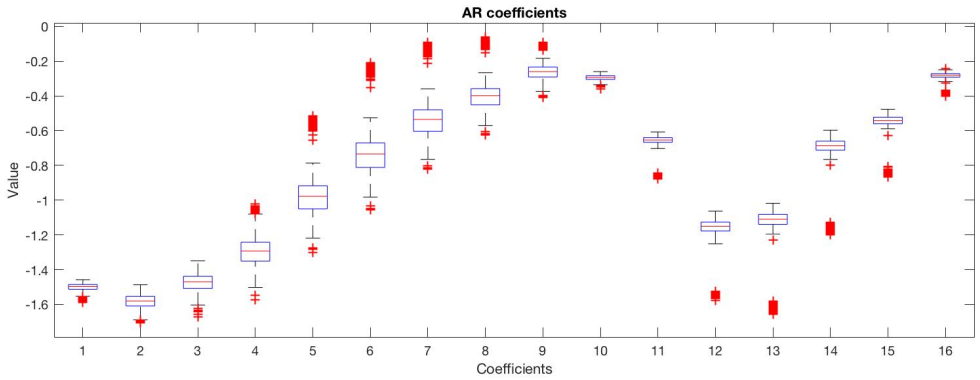
The Squared Mahalanobis Distance plays a major role in the field of novelty detection. The applicability of the SMD for a novelty detection approach relies on the hypothesis that the training dataset can be represented by a multivariate Gaussian Distribution [51]. This is a reasonable assumption since, in general, the dynamic response of civil infrastructures is theoretically assumed to be generated by excitation sources characterized by similar statistics over time, i.e. Gaussian white noise. Nonetheless, it may happen that during their working hours, some structures are subjected to excitation sources whose statistics may largely vary over time. In such circumstances, the training data can be represented by a mixture model (mixture of two or more distributions) and the concept of Squared Mahalanobis Distance should be revisited. Let's consider, for instance, that the observations of cepstral coefficient sequences extracted at a certain DOF do not follow a Gaussian distribution, but their distribution can be approximated by fitting a Gaussian Mixture Model [52]. In this case, the Squared Mahalanobis Distance of the points in the test distribution from each of the training mixture components (mixture distributions) needs to be computed and the lowest value is considered (lowest Squared Malanobis Distance) and monitored over time. When it overtakes a given threshold, the presence of structural damage is assessed. Anyway, the entire procedure is not simple; the Gaussian Mixture Model requires the knowledge of the number of clusters of the training data to fit provided by mean of some algorithm (i.e. EM) [53]. In the previous sections we presented an algorithm relying on the projection of training data through PCA to extract the minor components. This methodology allows to extract some clustered variables (i.e. the minor components) from the training data set. By projecting the training data onto the minor components, the variance be-

tween the accelerometers monitoring different locations of the structure subjected to any type of input is dramatically reduced so that the projected data distribution can be represented by a single Gaussian distribution. Then, this approach allows to avoid dealing with mixture models. The theoretical foundation hitherto discussed is followed in practice.

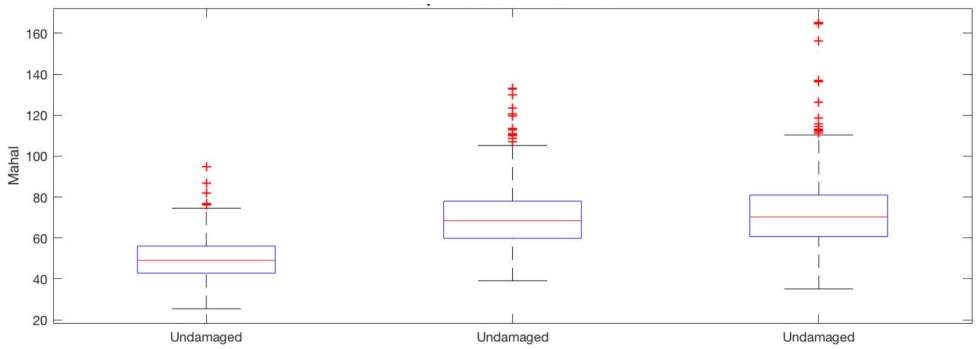
An example of the advantage given by the projection of the damage sensitive features (cepstral coefficient sequences) onto the minor components in a novelty detection based structural damage assessment is presented in this section. For this purpose, a novelty detection analysis based on the AutoRegressive coefficients (AR) is performed on the 8-DOF shear type system described in section 2.4.1. The AutoRegressive coefficients (AR) have been largely adopted as damage sensitive features in the analysis of linear systems [18, 39, 54, 55]. The methodology commonly used for their extraction is presented in Appendix 8.1.

By considering the 8-DOF shear type system in section 2.4.1 in its baseline (undamaged) conditions, it is possible to generate a dataset of $N_{ob}^{tr} = 500$ observations of the acceleration response time histories. The excitation force setup has been designed so that the training dataset cannot be represented by a multivariate Gaussian distribution. The 80% of the N_{ob}^{tr} observations of acceleration response time histories has been generated when the system is subjected to Gaussian white noise excitations whose standard deviation is 100 N at the top 4 floors and 10 N at the bottom 4 floors. Contrarily, the remaining 20% of the N_{ob}^{tr} observations of acceleration response time histories has been generated when the system is subjected to Gaussian white noise excitations whose standard deviation is 10 N at the top 4 floors and 100 N at the bottom 4 floors. The AR coefficients are extracted from the acceleration response time histories. The number of AR coefficients to take into account (model order) has been calculated according to the Akaike Information Criterion (AIC) [56]. Figure 2.14 (a) shows the distribution of the AR coefficients extracted from the training dataset at DOF 1 of the shear-type system. As expected, most of the AR coefficients are extremely sensitive to the variation of the excitation source. In fact, the 20% of the training data contributes to form outliers clustered away from the suggested values (median). Three test datasets consisting of $N_{ob}^{te} = 500$ observations of acceleration response time histories are generated for the structure in undamaged conditions. In the first dataset, the dynamic response is obtained by considering the same excitation setup used for the training dataset. The second and third test datasets are obtained by simulating the structural response of the system subjected to Gaussian white noise excitations whose standard deviation is 10 N at the top 4 floors and 100 N at the bottom 4 floors (consistently to the 20% of

the N_{ob}^{tr} observations of the training dataset). Figure 2.14 (b) shows the distribution of the SMD of the test data from the training data for each of the three test scenarios for the AR coefficients obtained at DOF 1. Despite no damage occurs in the structure, by comparing the median of the SMD for the first test dataset with those for the second and third test datasets, the presence of a damage is suggested. Thus, since the assumption at the core of the SMD is violated, it cannot be used as a valid damage sensitive feature.



(a) Autoregressive coefficients



(b) Mahalanobis distance over different states

Fig. 2.14. *Features at DOF 1.*

By considering the training dataset and the third test dataset just described, the distributions of the projection of the training and test cepstral coefficient sequences (computed at each DOF) onto the 49^{th} and 50^{th} principal components (minor components) are shown in Figure 2.15. The training and test distributions are consistent to each other confirming that no damage occurs in the structure. One of the advantages of the minor component analysis is here clear. It allows to overtake the obstacles due to different excitation and provides features not sensitive to the input configuration, but only to the damage.

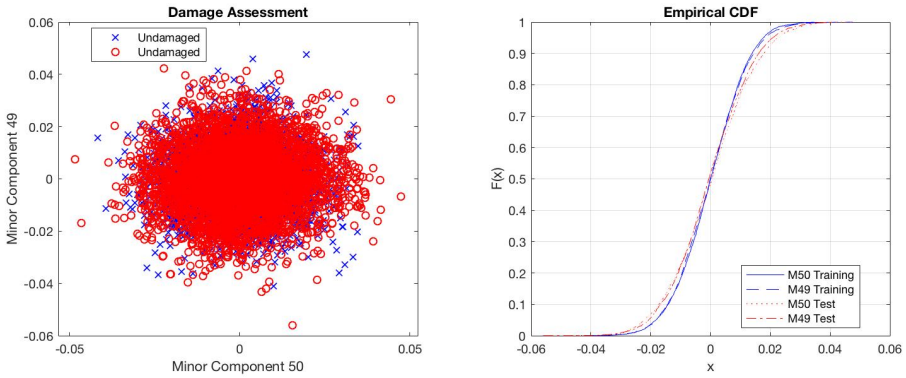


Fig. 2.15. *Undamaged states under different excitation sources.*

2.5. Principal Components of the Minor Components

The effectiveness of the minor components as damage sensitive features for a novelty detection based structural damage assessment method has been proven in the previous sections. However, all the numerical simulations hitherto presented rely on a small number of minor components arbitrarily selected; out of 50 components, only the last two are taken into account. Some further considerations in opposition to that choice can be made. First, some components might be more sensitive to the damage than others and an arbitrary selection could not be proper. Second, the total number of components provided by the PCA is equal to the length of the cepstral coefficient sequence Q. Then, by increasing Q the total number of components increases as well as the number of those that can be considered minor components. Potentially, we may have a significant amount of minor components that can be more or less sensitive to the damage and an arbitrary and random selection of two of them could be a contestable policy. For such a reason, in this section, the novelty detection analysis is conducted so that we can have a low dimensional representation of the distributions of the minor components even when they are more than 2.

A visual representation of more than 3 minor components

The statistical pattern recognition based damage assessment method relies on a training phase and a test phase leading to a novelty detection analysis of the minor components. The issue regarding the definition of the minor components arises during the training phase. Hence, the PCA provides the principal components as well as the minor components (those principal components contributing to the lowest variance of the datasets). Are we really able to distinguish the formers from the latters? The answer to that question can be found by looking at the PCA

as a dimensionality reduction method. In fact, the dimensionality reduction is carried out by considering a cut-off threshold based on the contribution of the single components to the total variance of the data. In this framework, according to the standard assumptions in literature, those components contributing to less than the 1% of the variance can be claimed to be minor components.

In the test phase, new observations are projected onto the space defined by the minor components. As long as more than 3 minor components are selected, a visual representation cannot be provided unless further mathematical manipulations. Furthermore, the mismatch between the training and test distributions projected onto the minor components can be emphasized in a new space. Hence, a new PCA can be helpful to develop a low dimensional representation of those multivariate distributions. Whereas the PCA performed in the training phase aims to find a low variance space and cluster the training data distributions, the new PCA is intended to define the space maximizing the variance between the projections of the training and test cepstral coefficient sequences datasets. The former PCA is oriented to the features extraction, the latter to the dimensionality reduction.

Structural damage assessment relying on several minor components

The analysis of the minor components is reported for the 8-DOFs shear type system described in section 2.4.1. The cepstral coefficient sequences used for the training and test phases are exactly the same used in the training dataset and in the first two damage scenarios reported in section 2.4.1. Let's recall that in the first damage scenario the unknown state of the system is undamaged, whereas in the second the system is damaged and the damage has been simulated through a drop in stiffness of the 10% of the spring element connecting the degrees of freedom 2 and 3. As the training dataset (cepstral coefficient sequences) is the same for both the tests, also the minor components given by the PCA are the same.

As mentioned in the last section, in the training phase we need to define the minor components by mean of a threshold that is set to 1% of their contribution to the total variance. Such contribution is displayed in Figure 2.16 and let us observe that the components 9 to 50 can be interpreted as minor components. Therefore we will consider the distribution of the minor components in a 42-dimensional space.

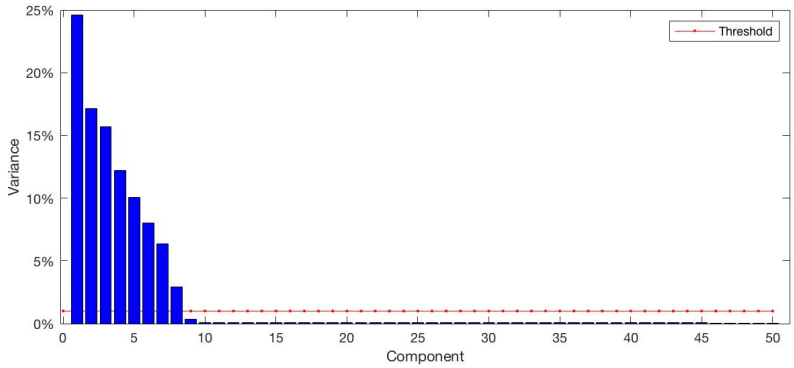


Fig. 2.16. *Variance contribution for each component.*

In the test phase, cepstral coefficient sequences for the structure in unknown conditions are projected on the 42-dimensional space and the second PCA, considering the training and test projected data points, is performed. The selection of the number of principal components to consider defines the dimensionality of the low dimensional representation of the distributions. In this section, a 2-dimensional representation will be provided.

The first test dataset (structure in undamaged conditions) is analyzed to remark that the distributions of the projected training and test cepstral coefficient sequences match each other. Figure 2.17 (a) shows the 2-D representation of the training and test distributions in the space defined by the principal components of the second PCA. The training and the test data are marked in blue and red respectively. It is clear that the distributions overlap. The second test dataset (structure in damaged conditions) is performed to show the mismatch between the training and test distributions due to damage. Figure 2.17 (b) shows the distributions of the principal components for an undamaged state (blue) and for the damaged state (red). For sake of clarity it is worth to point out that, despite the training minor components are the same, their projection shown in Figure 2.17 (b) is on a space which is not the same as the one in Figure 2.17 (a), as the projecting matrices of the second PCA also depend on the projected test data.

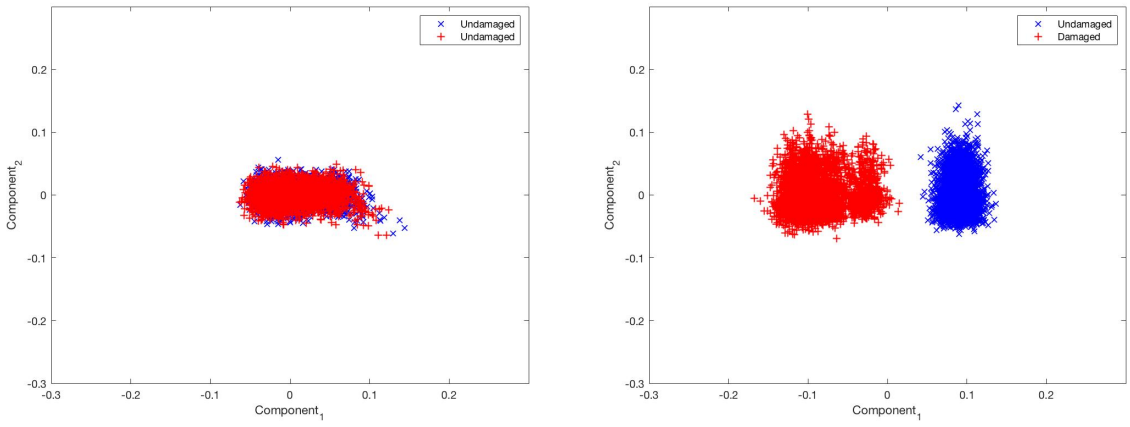


Fig. 2.17. *Principal Components, Undamaged vs Undamaged (a) and Undamaged vs Damaged,* 10% *stiffness drop (b).*

Compared with the damage assessment based on two minor components only, this approach offers the advantage of a compact low dimensional representation of more damage sensitive features avoiding to the user the arbitrary selection of the minor components.

The next section presents the proposed pattern recognition based damage assessment algorithm in a newly formulation so to extend the analysis to damage sensitive features which can be represented by either the cepstral coefficients extracted from the acceleration response time histories or the projection of the cepstral coefficient sequences onto the minor components.

2.6. A data-based damage detection algorithm using Cepstral Coefficients

In the field of statistical pattern recognition, the common approach used in damage detection analyses is to define a model of the damage sensitive features which is representative of a healthy state of the system and compare new instances of the damage sensitive features extracted from new data obtained from the system in an unknown state against those originally computed. The final goal is to be able to state whether those new instances are novel, and so diverging from the realizations of the healthy state, or not. Following the terminology used in the current literature, determining the statistical distribution (pattern) of the damage sensitive features from the data obtained from the system in its original state is called the "training phase" and the corresponding statistical distribution the "training model". Once the training phase is complete and the training model determined, then, when new data from the structure in unknown state becomes available, the damage sensitive features are extracted from the new recorded data and compared with the training model. This step of the analysis is called the

"testing phase" and ends with a binary decision: the newly extracted damage sensitive features fit the statistical distribution of the training model and so the structure is still considered healthy or they represent an outlier of the training model, implying that damage might have occurred. The assumption at the base of this analysis is that the extracted damage sensitive features are multivariate normally distributed variables: given N_{ob} observations of independent identically distributed (i.i.d.) variables, their distribution approaches a normal distribution as N_{ob} approaches infinite. While the AR coefficients introduced in the last section are commonly assumed to follow a normal distribution and perfectly satisfy this requirement, the same is not exactly true for the cepstral coefficients. However, two considerations about the cepstral coefficients can be made: 1) the distribution of the zeros $Z^{(i)}$ in Equation (2.25) is biased since there is not a linear dependence between such zeros and the structural properties, 2) there is a strong correlation between the cepstral coefficients and so they cannot be claimed to be independent variables. Anyway, through a mathematical manipulation, e.g. performing a Principal Component Analysis, the projection of the cepstral coefficients leads to the extraction of independent and identically distributed damage sensitive features.

In the case that the distribution of the damage sensitive features follows a multiple mixture model, we proposed to project the training distribution onto a space where those mixtures are clustered together. This has been done by using a Principal Component Analysis (PCA) and looking at the lower order components, an approach that has been shown to be particularly effective when dealing with cepstral coefficients.

In this chapter, following a multivariate outlier detection approach, the Squared Mahalanobis Distance was proposed as the damage index. The analytical formulation of the SMD is here provided. Given a multivariate normal distribution of the damage sensitive features from the training phase, whose mean is $\boldsymbol{\mu} \in \mathbb{R}^{d\times 1}$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$, the Squared Mahalanobis Distance of a new feature vector $\boldsymbol{x} \in \mathbb{R}^{d\times 1}$ from such distribution is defined as follows:

$$\mathscr{D}^2(\boldsymbol{x}) = (\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}). \tag{2.56}$$

It is noteworthy that, as long as the distribution of the training features can be represented by a normal distribution, such distribution implicitly contains all the variability due to external factors and disturbances.

In an outlier analysis, the Squared Mahalanobis Distance of an instance from the training distribution is a scalar that has to be compared with a threshold to establish whether such instance

belongs to the system in a healthy state or not. The definition of the threshold completely depends on the training distribution.

In machine learning, a common approach to determine this threshold is to use the cross validation method, splitting the training data into k-folds and using one fold at a time to compute the Squared Mahalanobis Distance from the remaining training population and then by adding the mean of those k-values and their standard deviation multiplied by a safety coefficient. Alternatively, it has been proven that the distribution of the Squared Mahalanobis Distance of d-variate independent and identically distributed variables asymptotically converges to a χ^2-distribution with degree of freedom d. Furthermore, Ververidis and Kotropoulos [57] proved that the distribution of the Squared Mahalanobis Distance of a d-variate point not involved in the estimation of the sample mean and covariance follows a scaled F-distribution with degrees of freedom d and $n_{ob}-d$, where n_{ob} is the number of observations used to generate the sample distribution.

Using this framework to address a damage assessment problem, the proposed damage assessment algorithm, schematically represented in Figure 2.18, will consist of two phases: the training phase and the test phase. During the training phase, N_{ob}^{tr} observations of the structural response acceleration time histories are recorded at N locations (again here the assumption is that all the degrees of freedom are instrumented, but it is not a necessary condition). At each location $i = 1,2,\ldots,N$, the j^{th} acceleration time history ($j = 1,2\ldots N_{ob}^{tr}$) will be analyzed and the first nc cepstral coefficients (except for the first one $q = 0$) will be taken into account. A newly defined vector $\boldsymbol{c}_{tr}^{(j)} \in \mathbb{R}^{d\times 1}$ can be arbitrarily set to contain the damage sensitive features considered (i.e. cepstral coefficients or their projections). For instance, it can contain either all the observations of cepstral coefficients computed at any location so that $d = nc \cdot N$ (global analysis) or all the observations computed at a certain position $i = 1,2,\ldots,N$, yielding $d = nc$ (local analysis). The sample mean vector over N_{ob}^{tr} training realizations of the damage sensitive features $\boldsymbol{c}_{tr}^{(j)} \in \mathbb{R}^{d\times 1}$ for $j = 1,\ldots,N_{ob}^{tr}$ is given by:

$$\boldsymbol{\mu}_{tr} = \frac{1}{N_{ob}^{tr}} \sum_{j=1}^{N_{ob}^{tr}} \boldsymbol{c}_{tr}^{(j)} \tag{2.57}$$

and the covariance matrix is evaluated as:

$$\boldsymbol{\Sigma}_{tr} = \frac{1}{(N_{ob}^{tr}-1)} \sum_{j=1}^{N_{ob}^{tr}} (\boldsymbol{c}_{tr}^{(j)} - \boldsymbol{\mu}_{tr})(\boldsymbol{c}_{tr}^{(j)} - \boldsymbol{\mu}_{tr})^T. \tag{2.58}$$

They define the statistical distribution representing the healthy state of the structure. Once the trained model is finally available, a number of N_{ob}^{te} new instances $\boldsymbol{c}_{te}^{(j')} \in \mathbb{R}^{d\times 1}$ for $j' = 1,\dots,N_{ob}^{te}$ of the damage sensitive features of the system in unknown conditions are collected in the test phase. The Squared Mahalanobis Distance of $\boldsymbol{c}_{te}^{(j')}$ from the training distribution is evaluated:

$$\mathscr{D}^2(\boldsymbol{c}_{te}^{(j')}) = (\boldsymbol{c}_{te}^{(j')} - \boldsymbol{\mu}_{tr})^T [\boldsymbol{\Sigma}_{tr}]^{-1} (\boldsymbol{c}_{te}^{(j')} - \boldsymbol{\mu}_{tr}). \tag{2.59}$$

so that we obtain N_{ob}^{te} different scalar values for $\mathscr{D}^2(\boldsymbol{c}_{te}^{(j')})$ ($j' = 1,\dots,N_{ob}^{te}$) at each DOF and consider their *median* value as a local damage index $\mathscr{D}^2$. The estimated damage index is compared against the threshold value of the scaled F-distribution in order to assess the presence of the damage. As long as the damage sensitive features extracted in the test phase (unknown state) belong to the same distribution type of those extracted in the training phase (healthy state), the Squared Mahalanobis Distance of $\boldsymbol{c}_{te}^{(j')}$ follows a scaled F-distribution:

$$\frac{N_{ob}^{tr}(N_{ob}^{tr}-d)}{((N_{ob}^{tr})^2-1)d} \mathscr{D}^2(\boldsymbol{c}_{te}^{(j')}) \sim F_{d,N_{ob}^{tr}-d}. \tag{2.60}$$

In this chapter, the value of the threshold Γ is set to the 0.99-quantile of the training scaled $F_{d,N_{ob}^{tr}-d}$ distribution. As long as the median of the Squared Mahalanobis Distance, scaled by the factor $N_{ob}^{tr}(N_{ob}^{tr}-d)/((N_{ob}^{tr})^2-1)d$, is lower than the threshold Γ, the system is proclaimed undamaged.

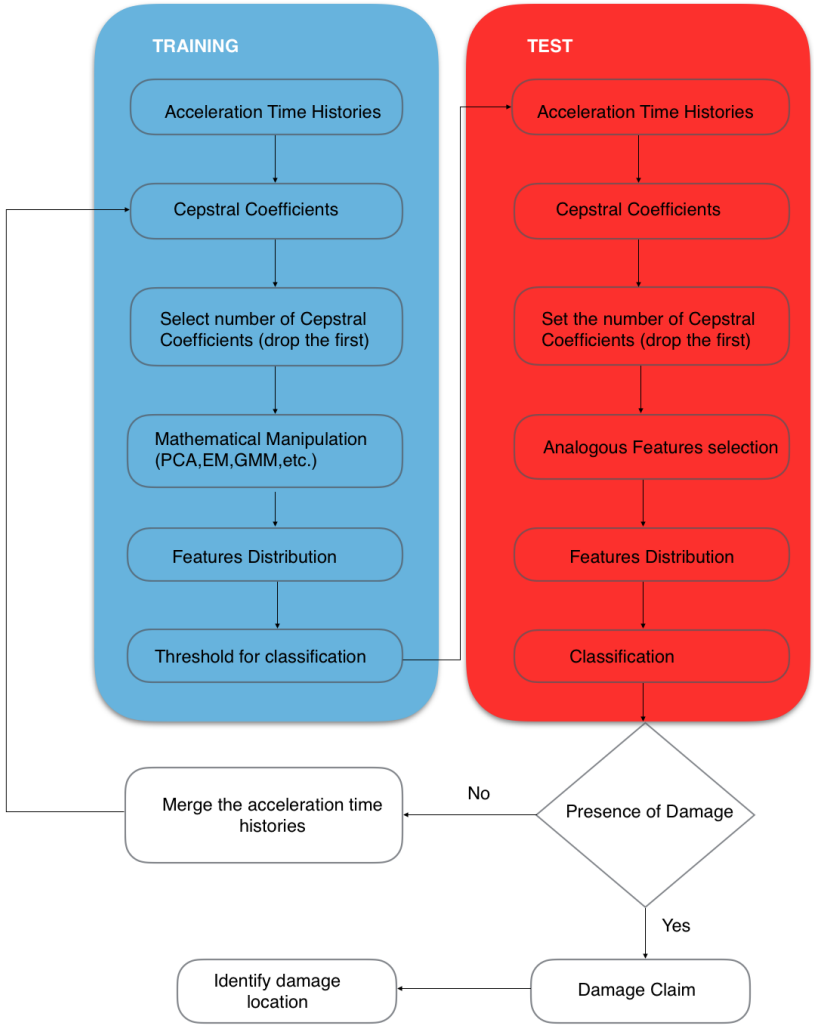


Fig. 2.18. *Damage Assessment Algorithm.*

2.7. Structural damage detection

Three case studies are presented to examine the performance of the damage assessment algorithm based on different damage sensitive features. In the first two cases, the acceleration response time histories are numerically simulated from an 8DOF shear-type system, while in the third case, the time histories of the structural response recorded through accelerometers strategically positioned on a real structure are considered.

2.7.1. 8-DOF shear-type - Case I

The structural system analyzed in this case is a lumped mass model of an 8-DOF shear-type system. As shown in Figure 2.19 the mass and spring elements are numbered in ascending order from the ground constraint to the top. The system in its baseline condition is characterized by

horizontal springs of stiffness $k_i = 25000$ N/m ($i = 1,\ldots,8$), and each mass is equal to $m_i = 1$ kg ($i = 1,\ldots,8$). The frame is supposed to have modal damping with a damping factor of $\xi = 1\%$ for each of the 8 vibration modes.

Sixteen different structural conditions are considered in the analysis (Table 2.4). In addition to the first state, that corresponds to the baseline condition of the system, there are 8 additional states that correspond to slight changes of the stiffness at various floors: these cases simulate fluctuations of the structural properties due to environmental conditions (e.g. temperature, humidity, etc.). For the remaining 7 cases, a drop in stiffness of certain spring elements will simulate different damage scenarios.

For each scenario, the excitation is provided by external forces applied at every DOF via zero-order-hold (ZOH) for 100 sec duration with a sampling period of 0.01 seconds. Such forces are represented via zero-mean Gaussian white noise signals (standard normal distribution) whose magnitudes are scaled at each floor, by a factor of 100. The output dataset is represented by the time-histories of the structural acceleration recorded at every level (full set of sensors) corrupted by a 10% RMS Gaussian white noise to simulate the effect of measurement disturbances.

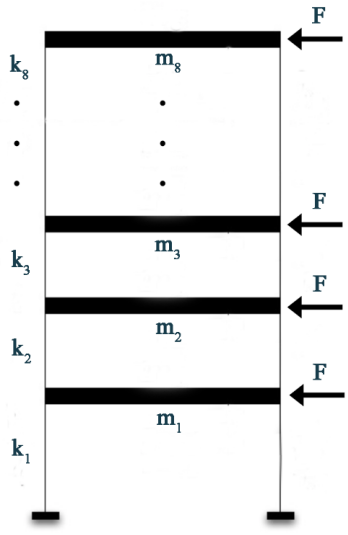


Fig. 2.19. *8-DOF shear type system*

Scenario	Condition	Anomalies
1	Undamaged	Baseline
2	Undamaged	$k_i = 0.98k_i^0$ for $i = 5,\ldots,8$
3	Undamaged	$k_i = 0.99k_i^0$ for $i = 5,\ldots,8$
4	Undamaged	$k_i = 1.01k_i^0$ for $i = 5,\ldots,8$
5	Undamaged	$k_i = 1.02k_i^0$ for $i = 5,\ldots,8$
6	Undamaged	$k_i = 0.98k_i^0$ for $i = 1,\ldots,4$
7	Undamaged	$k_i = 0.99k_i^0$ for $i = 1,\ldots,4$
8	Undamaged	$k_i = 1.01k_i^0$ for $i = 1,\ldots,4$
9	Undamaged	$k_i = 1.02k_i^0$ for $i = 1,\ldots,4$
10	Damaged	$k_i = 0.9k_i^0$ for $i = 1$
11	Damaged	$k_i = 0.9k_i^0$ for $i = 3$
12	Damaged	$k_i = 0.9k_i^0$ for $i = 5$
13	Damaged	$k_i = 0.9k_i^0$ for $i = 7$
14	Damaged	$k_i = 0.85k_i^0$ for $i = 7$
15	Damaged	$k_i = 0.9k_i^0$ for $i = 3,7$
16	Damaged	$k_i = 0.9k_i^0$ for $i = 2,8$

Table 2.4. *Damage scenarios.*

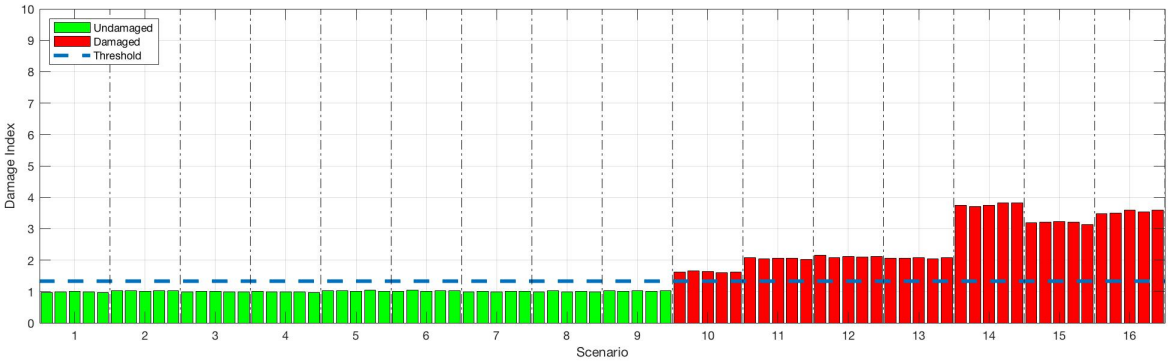
First, let's address the performance of cepstral coefficients in assessing whether the system has suffered structural damage or not. Later, we will look at their ability to identify the location

of damage. When dealing with cepstral coefficients as damage sensitive features, let us recall that their analytical representation showed that they converge to zero relatively fast and hence it is reasonable to assume that the local information about the structural properties is concentrated in the first few elements of the cepstral coefficient sequence. For that reason, the first nc elements of the cepstral coefficient sequences computed at each of the $N = 8$ DOF are retained. Since our goal is to assess whether or not there is damage, all the retained cepstral coefficients can be gathered in a damage sensitive feature vector $\boldsymbol{c}_{tr}^{(j)} \in \mathbb{R}^{nc \cdot N \times 1}$ for $j = 1, \ldots, N_{ob}^{tr}$ that accounts for all the available sensors and that is used to develop the training model (distribution). In order to introduce additional uncertainty factors, the values of the stiffness elements for each damage scenario are perturbed by a random value between -1% and $+1\%$. The training dataset consists in 100 realizations of the 9 undamaged scenarios reported in Table 2.4, so that $N_{ob}^{tr} = 900$. To test the effectiveness of the cepstral coefficients as damage sensitive features, the same damage assessment strategy is applied using AutoRegressive coefficients as damage sensitive features and the results are compared. The number of AR coefficients to be considered in the analysis is determined using the Akaike Information Criterion on the first realization of the undamaged scenario in baseline condition, resulting in an order $p = 18$. Consequently, an equal number of cepstral coefficients was adopted ($nc = 18$): later, this value will be varied to show its effect on the damage detection process.

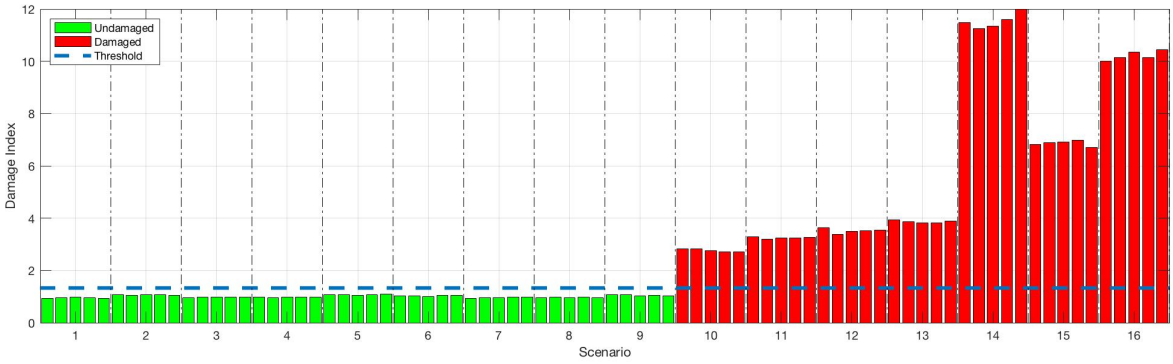
Looking at the computational efficiency, in the training phase the time required to obtain the $N_{ob}^{tr} = 900$ realizations of the damage sensitive features is 7.99 sec for the cepstral coefficients and 32.23 sec for the AR coefficients (Mac Pro, 2.8 GHz Intel Core i5). In determining the threshold limit from the training model, both the damage sensitive features have identical scaling factor (0.0058) and threshold value (1.33) for a confidence level of 99%.

The damage detection test is performed to predict the health conditions of the system. Once the training model and the corresponding scaling factor and threshold have been found, the test phase begins. Here for each of the scenarios of Table 2.4, 5 sets of 100 realizations of the acceleration responses have been considered so that the total number of instances is 8000. The median value of the Squared Mahalanobis Distance is used as damage index and compared against the threshold. The results are shown in Figure 2.20. By looking at Figure 2.20, it is clear that the cepstral coefficients are quite successful in identifying healthy and damaged conditions. Their accuracy matches very well with that of the AR coefficients but at a much reduced computational effort. In fact, considering the results in Figure 2.20, the following

considerations can be made: 1) the results from the AR model are obtained using an *optimal* AR model where the number of coefficients ($nc = 18$) has been chosen according to the Akaike Information Criterion, while the equal number of cepstral coefficients ($nc = 18$) is not necessarily the *optimal* model order, as shown in Figure 2.22 (a), and 2) the order of the AR model to consider affects the value of the AR coefficients and so, for different model orders, new regression models need to be evaluated while, on the contrary, cepstral coefficients can be computed only once, independently from nc. Furthermore, when dealing with experimental data that may contain, even slightly, nonlinear effects (outside the purpose of this chapter), the cepstral coefficients will be proven to be much more stable than the AR coefficients in damage assessment.



(a)



(b)

Fig. 2.20. *Damage Index over 16 scenarios. Cepstral coefficients (a) and AR coefficients (b) as damage sensitive features.*

Figure 2.21 (b) shows the same results in terms of confusion matrices. Here, each matrix shows the performance of the classification algorithm considering the Squared Mahalanobis Distance of the test realizations individually, assigning a '0' if the system is classified as undamaged or a '1' if damaged. The elements along the diagonal of the confusion matrix represent

the instances when the damage sensitive features correctly classify the condition of the system. The confusion matrix for the cepstral coefficients in Figure 2.21 (b) points out that, for 0.7% of the 8000 instances, the algorithm leads to classify the undamaged system as damaged (Type I error) and only for the 0.4%, misclassifies the damaged system (Type II error). An interesting result is shown in the first and last confusion matrices obtained by changing the number of cepstral coefficients considered in the analysis as for $nc = 7$ in Figure 2.21 (a) and $nc = 28$ in Figure 2.21 (c). The number of Type I and Type II errors are quite similar (Type I: from 0.4% to 0.9%, Type II: from 0.5% to 0.7%) and so the accuracy (from 98.6% to 98.9%) for the three values of nc considered ($nc = 7, 18, 28$).

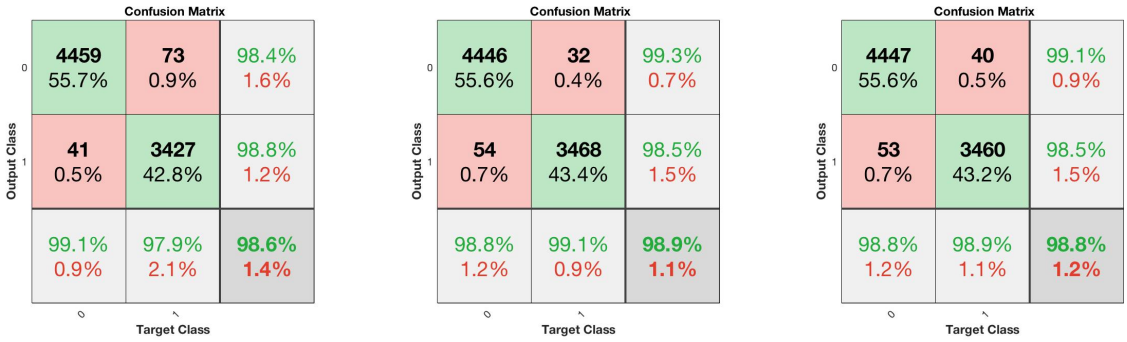
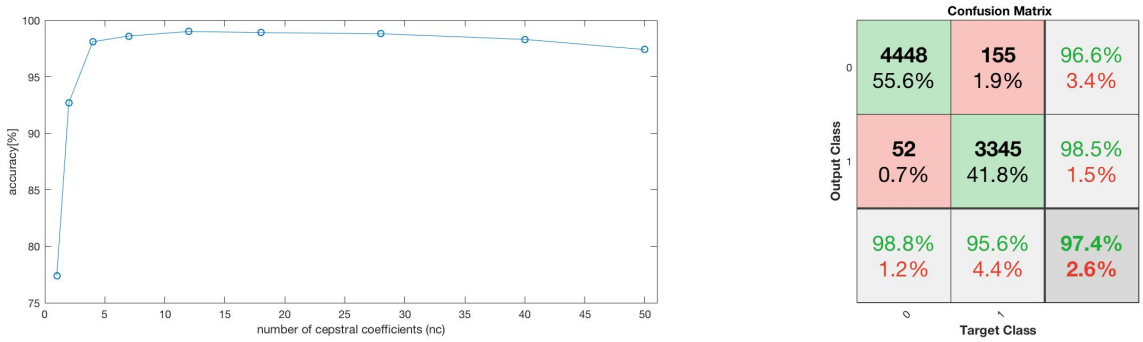


Fig. 2.21. *Confusion Matrix for different damage sensitive features. a) Cepstral Coefficients nc = 7, b) Cepstral Coefficients nc = 18 , c) Cepstral Coefficients nc = 28.*

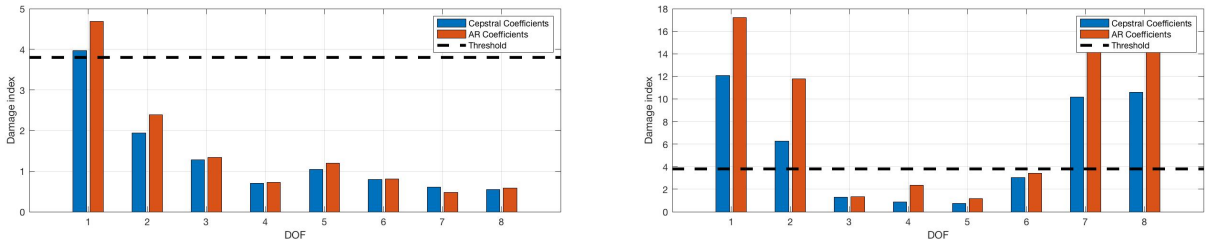
An investigation about the performance of the proposed damage assessment algorithm by varying the number of components nc considered shows that using a number of cepstral coefficients beyond a certain range leads to a decrease in accuracy. This can be seen in Figure 2.22 (a) where the accuracy of the proposed algorithm is plotted as function of the number of cepstral coefficients nc. When considering few coefficients, the number of damage sensitive features does not contain enough information about the structural system leading to inaccurate estimations. When dealing with large numbers of cepstral coefficients, the part of the coefficients linked to the structural characteristics tends to zero (see Figure 2.1 for a SDOF), while the remaining part linked to the external noise and excitation sources becomes predominant leading to a decrease of accuracy in the damage assessment. Figure 2.22 (b) shows the confusion matrix for the case $nc = 50$ where the Type II error increases to 1.9%.



(a) **(b)**

Fig. 2.22. *Accuracy over nc a), confusion matrix for Cepstral Coefficients nc* $= 50$ *b).*

Finally, one last consideration about this case study can be made. Until now, the damage sensitive feature was represented by a unique vector containing the first *nc* cepstral coefficients (or the first p AR coefficients) extracted from the acceleration response time histories at every DOF. An alternative approach is to consider each monitoring location separately so to obtain multiple local damage indexes. As pointed out in section 2.2.2, the first values of the cepstral coefficient sequences are the ones embedding most of the local information of the structural response. For such a reason, once the cepstral coefficients (or the AR coefficients) are locally extracted, only the first 3 elements of the sequences are considered ($nc = p = 3$). Both the threshold Γ and the scaling factor are common for all the distributions given by the damage sensitive features computed at different DOF. For this example the threshold Γ is calculated to be equal to 3.81 and the scale factor is 0.322. For sake of brevity, the local damage indexes are reported in Figure 2.23 only for the damage scenarios 10 (a) and 16 (b). The bar plots contain the damage indexes extracted through both cepstral coefficients and AR coefficients.



(a) **(b)**

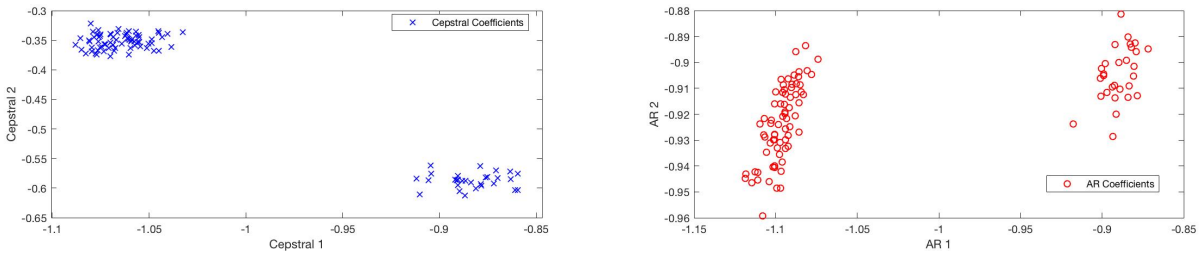
Fig. 2.23. *Comparison between local damage index using cepstral coefficients and AR coefficients as damage sensitive features for Scenario 10 (a) and Scenario 16 (b)*

In both the reported scenarios (a) and (b), a strong correlation between the damage indexes provided by the two damage sensitive features is clearly visible. Only the damage indexes com-

puted at DOF 1 overtake the threshold when the damage is simulated between DOF 1 and the base (Figure 2.23 (a)) while, when the damage is between DOFs 1-2 and 7-8, only the damage indexes relative to those DOFs go beyond the threshold. Hence, the cepstral coefficients, when applied to each single DOF separately, appear capable of detecting not only the occurrence of damage, but also its location.

2.7.2. 8-DOF shear-type - Case II

Here, the attention is focused on those cases where the cepstral coefficients (as well as the AR coefficients) cannot be assumed normally distributed but can be, for instance, a mixture of many distributions. This could be the case in which the excitation is not the same or when it has different variance at various DOFs. A simple case where this can be easily shown is represented by the same structural system from Case I where the excitation force is applied either at the bottom mass (1^{st}) or at the top mass (8^{th} DOF). The training dataset consists of $N_{ob}^{tr} = 100$ realizations of the structural acceleration, each consisting in 8 time-histories, for each of 9 undamaged representations of the undamaged system. In each realization, the acceleration time-histories are simulated by exciting the system with just one force applied at DOF 1 or DOF 8 for the entire duration, with a probability of the force to act at DOF 1 or DOF 8 equal to 70% and 30% respectively. The distribution of the first two cepstral coefficients is shown in Figure 2.24. This variability can be explained by recalling the analytical expression of the cepstral coefficients: while the θ_q component was the same at every location, the components $\gamma_{i,q}$ varied from sensor to sensor, depending on the locations of both the sensor and the actuator.

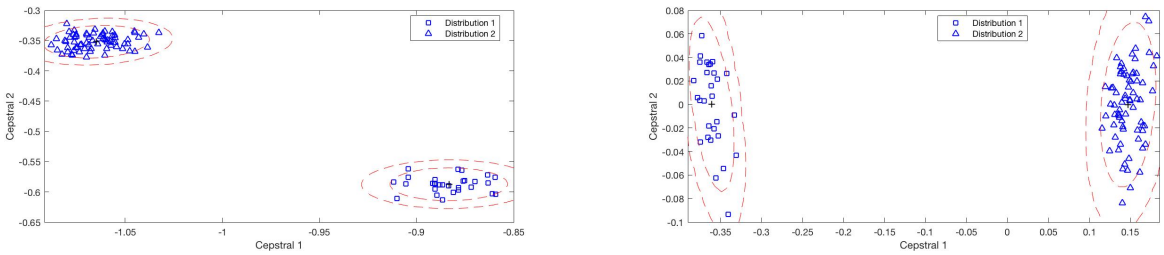


(a) **(b)**

Fig. 2.24. *Distribution of the first and second cepstral coefficients (a) and AR coefficients (b) for the scenario 1 at DOF 1.*

It can be easily stated that, looking at the distribution of these cepstral coefficients, they do not follow a normal distribution, a requirement needed for the direct application of the damage detection algorithm. Actually these data points, being representative of the cepstral coeffi-

cients from the training datasets, can be considered sampled from a mixture of distributions. In practice, it is common to approximate those training distributions and, in the testing phase, to perform an outlier analysis on the new test data using such approximate distributions. Several algorithms are available in literature to estimate these training distributions (e.g. Expectation Maximization (EM) algorithm, Variational Inference, Gibbs sampling, etc.) [58–61]. In Figure 2.25 an EM algorithm has been used to evaluate the training mixture model. However, one word of caution in using such algorithm is that, since an EM algorithm requires to specify in advance the number of mixtures to be estimated, this could be a difficult task: not in this case (clearly 2 mixtures in the training data) but in many real applications it could be difficult to assess this number. In evaluating the number of mixtures in a distribution, a principal component representation of the cepstral coefficients from the training dataset is quite helpful (Figure 2.25 (b)) as well as the Akaike Information Criterion statistics.



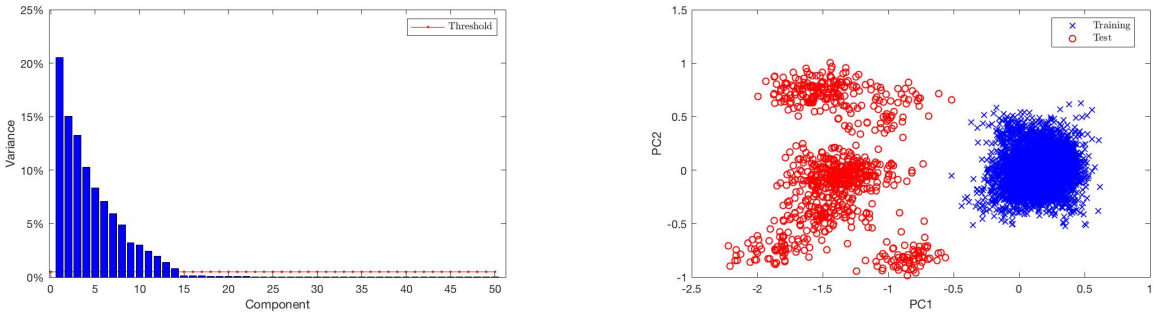
(a) **(b)**

Fig. 2.25. *Gaussian mixture models for the training distribution. Distribution of the first and second cepstral coefficients (a) and first and second principal components of the cepstral coefficients (b) for the scenario 1 at DOF 1. Maximum likelihood (+) and covariances (dashed lines for* 2σ *and* 3σ*) of the two Gaussian mixtures.*

According to the proposed strategy for the case of multiple mixture models (section 2.6), we conducted a PCA on the training dataset in order to select as minor components all those components contributing to less than 0.5% of the total variance of the entire dataset. It is noteworthy that, for noisy signals, some of those components could tend to overfit the signals, including noise, and so it would be better to impose also a lower threshold to remove such components.

Looking at the specific example of the 8DOF system with forces at either the DOF 1 or the DOF 8, the training dataset consists of 900 realizations (9 undamaged cases, each with 100 realizations). Each realization contains 8 time histories of the structural acceleration recorded at various DOFs. Consequently, for each realization, 8 sequences of 50 cepstral coefficients ($nc = 50$) are extracted: the value of nc has been set large in order to get, more likely, a

large number of minor components. Hence, a total of 7200 cepstral coefficient sequences are available and each of them will represent a point in the 50-dimensional space. It's important to remark that the 7200 sequences are projected onto the same space, common for each sequence, no matter its monitoring location. Figure 2.26 (a) shows the contribution to the total variance by the single components. Keeping the threshold of 0.5% of the total variance, the components 15 to 50 can be considered minor components.



(a) **(b)**

Fig. 2.26. *Variance contribution for each component (a). Projection of the cepstral coefficients on the first two principal components, Scenario 16 (b).*

When a new dataset (8 new time histories of the structural acceleration) obtained from the system in unknown conditions becomes available (testing phase), 8 new sequences of 50 cepstral coefficients will be extracted and projected onto the space identified by the previously defined minor components. Again, the process of whitening scales the distributions along their components. At this point, since some minor components may not be sensitive to damage, a new PCA, based this time on both the training and the new testing features, can be performed in order to maximize the variance and so the evidence of the damage. Figure 2.26 (b) shows the projections of the cepstral coefficients on the new principal components for scenario 16 on Table 2.4: it is clearly shown that the feature corresponding to the training data form a unique cluster (crosses), whereas those corresponding to the testing data (circles) is a mixture of distributions. Consequently, the projection of the cepstral coefficients on the 2 principal components can be defined as the new damage sensitive feature and the previously defined damage assessment criterion, based on the Squared Mahalanobis Distance, can be applied. Thus, for this specific case, $\boldsymbol{c}_{tr}^{(j)} \in \mathbb{R}^{2\times 1}$ in Equation (2.57) where $j = 1,2,\ldots,N \cdot N_{ob}^{tr}$ and $\boldsymbol{c}_{te}^{(j')} \in \mathbb{R}^{2\times 1}$ where $j' = 1,2,\ldots,N \cdot N_{ob}^{te}$.

Similarly to Case I, in Case II the test phase comprises of 5 sets of 100 realizations each, for each of the 16 scenarios and the relative damage index is reported in Figure 2.27. According to

the scaled F-distribution, the threshold for this case is 4.62 and the damage index is scaled by a factor equal to 0.49.

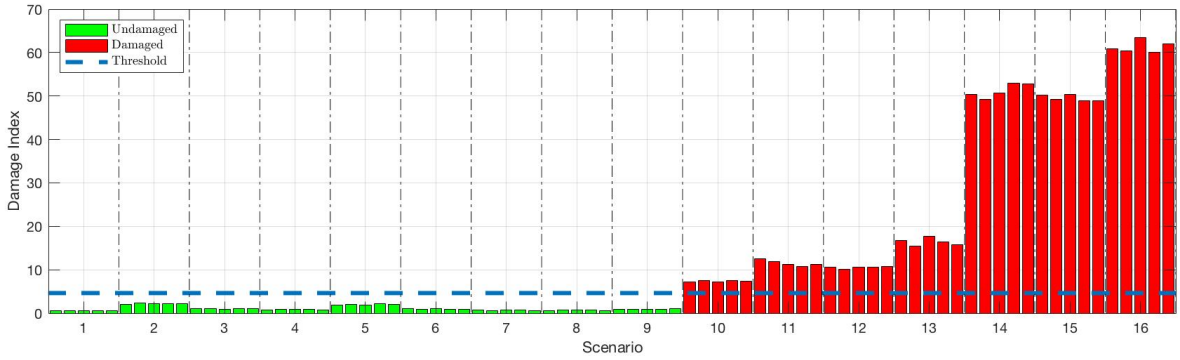


Fig. 2.27. *Damage Index over 16 scenarios, damage sensitive features extracted from the minor components.*

The results shown in Figure 2.27 are consistent with those obtained for Case I demonstrating the ability of the proposed algorithm to detect structural damage, even in the case of an arbitrary force distribution. Hence, it can be concluded that the PCA and whitening play a crucial role in the extraction of the damage sensitive features, in line with the assumptions of the damage detection algorithm.

2.7.3. The Z24 Benchmark - Case III

The third study case is considered in order to evaluate the performance of the proposed damage detection algorithm when dealing with real measurement data.

The Z24 bridge is a very well known case study, used by many researchers in their work. Part of the roadway link connecting the towns of Koppigen and Utzeenstorf, in the canton of Bern, Switzerland, the bridge was a post-tensioned concrete box girder bridge, with a main span of 30 m and two 14 m side spans. After 36 years of activity, the bridge was demolished in 1998 to be replaced with a new, larger bridge. The unique feature of this bridge is that, before its demolition, it was decided to monitor the dynamic behaviour of this bridge for a period of 10 months (from November 10, 1997 to September 10, 1998) together with environmental parameters such as local temperature, rain, wind speed, humidity, traffic, etc. In addition, towards the end of the monitoring period, some progressive damage was induced in the bridge structure (e.g. progressing lowering of the pier, spalling of concrete, etc.) and the response of the damaged structure recorded. Table 2.5 shows a chronological overview of the different structural conditions.

Before and after each damage occurrence, not having the traffic excitation anymore, the bridge was subjected to a series of forced and ambient vibration tests. Two vertical shakers were placed on the bridge deck so to excite the structure through a fairly flat spectrum between 3 and 30 Hz. A network of 16 accelerometers recorded structural accelerations at strategic locations on the bridge structure: for every hour, a total number of 65,536 samples (with a sampling time of 0.01 s) were recorded by each accelerometer, using an antialiasing filter with 30-Hz cutoff frequency.

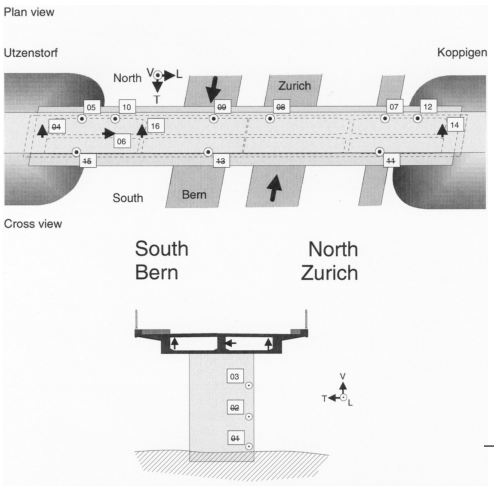


Fig. 2.28. *Z24 Bridge, sensors setup.*

Date (1998)	Scenario
10-17 July	**Undamaged Condition**
4 August	**Undamaged Condition**
9 August	Installation of pier settlement sys
10 August	Lowering of pier, 20 mm
12 August	Lowering of pier, 40 mm
17 August	Lowering of pier, 80 mm
18 August	Lowering of pier, 95 mm
19 August	Lifting of pier, tilt of foundation
20 August	New reference condition
25 August	Spalling of concrete at soffit, $12m^2$
26 August	Spalling of concrete at soffit, $24m^2$
27 August	**Landslide of 1 m at abutment**
31 August	**Failure of concrete hinge**
2 September	**Failure of 2 anchor heads**
3 September	**Failure of 4 anchor heads**

Table 2.5. *Chronological overview of applied scenarios; only the states in bold have been considered in this study.*

This dataset is one of the most comprehensive datasets available to study the impact of environmental conditions as well as progressive damage on a real bridge structure and provides a valid benchmark for testing theories and algorithms. Several studies [46–49] have been carried out using the data from the Z-24 bridge, mainly focused on the identification of the modal properties of the bridge. Among the recent ones, in 2017 Langone et al. [62] extracted natural frequencies, damping ratios and mode shapes using an update Finite Element model of the bridge: for example, using the measurement data recorded from August 20^{th} to August 25^{th}, they were able to identify six natural frequencies and damping ratios for the mode shapes as reported in Table 2.6.

Mode Number	Type	Frequency [Hz]	Damping ratio [%]
1	Bending	3.86	0.8
2	Torsional	4.90	1.4
3	Torsional	9.76	1.4
4	Torsional	10.30	1.3
5	Bending	12.42	2.8
6	Bending	13.22	3.4

Table 2.6. *Estimated modal parameters estimated.*

A complete analysis in the frequency domain requires an enormous computational effort and, above all, a certain level of experience by the analyst. In fact, the accuracy of the results is strongly dependent on the values of the parameters selected in the analysis. Let us consider, for example, one of the data-driven, output-only Stochastic Subspace Identification (SSI) algorithms available in literature. Here, the user selected parameters are: 1) the number of block rows and 2) the total number of columns used in the Hankel matrix, 3) the subpartition of the Hankel matrix in past and future output, and 4) the order of the state space matrices to be used for the estimation of the natural frequencies, damping ratios and mode-shapes. Changing these parameters will provide substantial variation in the results that will require extensive calculations to be able to extrapolate some meaningful results. As an example, using only the acceleration time histories recorded at sensor 05, 07, 10 and 12, the natural frequencies obtained from one set of records from the Z-24 bridge by varying the 4 parameters previously mentioned are identified and presented in Figure 2.29 following the methodology presented in the work by Tronci [63].

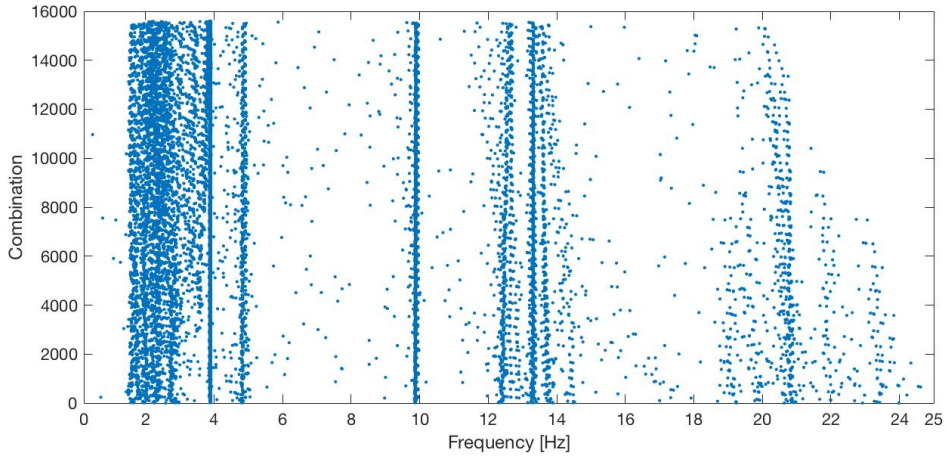


Fig. 2.29. *Z24 Bridge, eigenfrequencies for different SSI parameters.*

From this diagram, an estimation of the first few natural frequencies (below 15 Hz) leads to the values listed in Table 2.7

SSI Identified Frequencies [Hz]				
mean	3.8782	4.8294	9.8574	13.1333
std	0.0663	0.0334	0.1499	0.4524

Table 2.7. *First 4 natural frequencies identified.*

which obviously do not match the values estimated in Table 2.6 [62]. Not only the numerical values do not match but also the number of identified frequencies (4 vs 6). This variability of the results reduces the reliability of the entire damage assessment analysis when using the natural frequencies.

Using the identified cepstral coefficients as initial features, the proposed damage assessment strategy is conducted using the recorded data from the first two scenarios in Table 2.5 as training datasets (July 10^{th} through 17^{th} and August 4^{th} through 9^{th}) while the data corresponding to the other scenarios (Table 2.5) are used as test data. Only the data recorded by accelerometers 05, 07, 10, 12 have been considered in the analysis since the other records either are not available or present some abnormalities. The hourly records are framed into two time histories of equal length (30 min each) so to increase the number of realizations. Consequently, the training dataset consists of 48 time histories per sensor per day for a total number of 672 observations per sensor. The two datasets used in the training phase were acquired while the bridge was subjected to different excitation sources (traffic for the July 10^{th}-17^{th} period and shakers during the August 4^{th}-9^{th} period) when the environmental conditions (e.g. temperature, humidity, wind) were quite similar.

The identified frequencies over the entire set of observations (training and test) are shown in Figure 2.30 following the procedure in Tronci [63]. Using a lengthy operation that involved first the use of an SSI algorithm over a range of parameters and then a clustering analysis, four natural frequencies have been detected, even though for some observations, the algorithm has not been able to pinpoint reliable values for some of them. The vertical dashed lines separate the observations belonging to the different damage scenarios. The frequencies from observation 1 to 672 are representative of the structural system in an undamaged state, while the following observations (from 673 to 1008) were collected when the system was subjected to the cumulative damage as indicated in Table 2.5. The occurrence of damage introduces a slight drop in

some of the natural frequencies, as indicated in Figure 2.30. Figure 2.31 shows a comparison between the variation of the second natural frequency of the bridge and the variation of the 6^{th} cepstral coefficient from accelerometer 12.

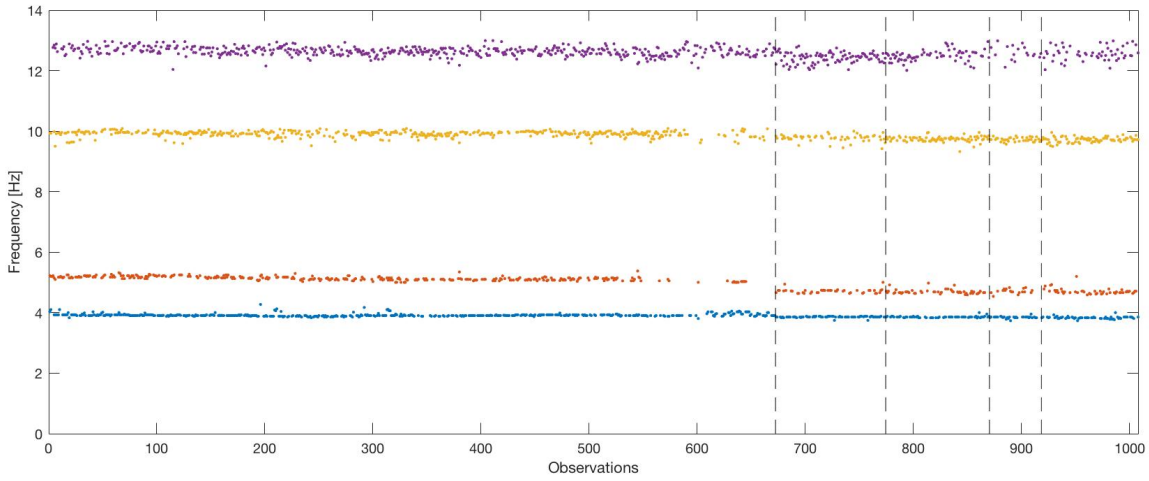
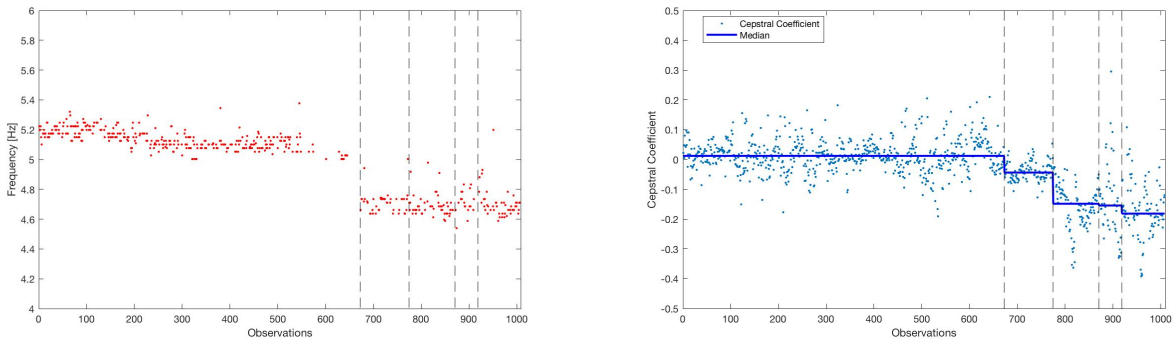


Fig. 2.30. *Eigenfrequencies obtained through SSI.*

Each of the observations in Figure 2.30 is related to a set of 4 acceleration time histories (4 sensors) 30 min long. From those acceleration time histories, the power cepstrum can be locally extracted at each monitored location providing the 4 cepstral coefficient sequences. As the cepstral coefficient sequences converge to zero relatively fast (at least for unit pulse and white noise excitations), only the first 50 values of the sequence are considered for the analysis. Subsequently, the first value for each sequence has been discarded, according to the theory. Finally, 4 sequences of $nc = 49$ cepstral coefficients are extracted for each of the observations belonging to the training (observations 1 to 672) and test (observations 673 to 1008) datasets. A comparison of the second natural frequency shown in Figure 2.30 against one of the cepstral coefficients extracted from accelerometer 12 (the 6^{th}) is provided in Figure 2.31. It is worthy remarking that only some of the cepstral coefficients are sensitive to the structural damage, depending on what part of the quefrency domain is more affected by specific variations of the structural properties. Among the natural frequencies, the second natural frequency seems to be more sensitive to the structural damage, even though it doesn't vary over the 4 damaged scenarios considered. It is worth to remember that the system identification algorithm providing the natural frequencies, SSI, adopts a time-varying kalman filter, so the identified modal parameters (in this case natural frequencies) are generated from filtered data. Contrarily, cepstral coefficients are generated from rough data and their trend appears more noisy (Figure 2.31).

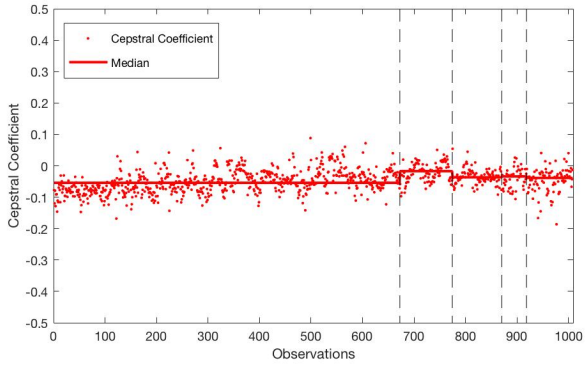
By the way, the projection of the training data onto the minor components aims to get rid of the variance due to different factors (excitation sources, temperature, humidity...).



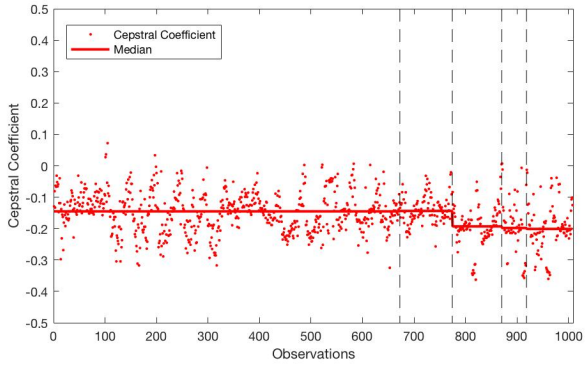
(a) **(b)**

Fig. 2.31. *Sensitivity to the damage. Second eigenfrequency identified (a),* 6^{th} *cepstral coefficient for accelerometer* #12 *(b).*

In this analysis, it is important to account for all the cepstral coefficients because they have different sensitivity to damage. Figure 2.32 (a) shows the variation of the 11^{th} cepstral coefficient for accelerometer 12: it is evident that the 11^{th} coefficient shows a quite different sensitivity to damage than the 6^{th} coefficient. In addition, it is recommended to consider cepstral coefficients from different accelerometers because they could also show a different sensitivity to the damage: for example Figure 2.32 (b) shows the same cepstral coefficient as in Figure 2.31 (b) (the 6^{th} coefficient) obtained from a different accelerometer (sensor 7). It is clear that the sensitivity to the damage is quite different and this depends on the sensor location relative to the damaged area. For this reason the records from all available sensors should be considered.



(a)



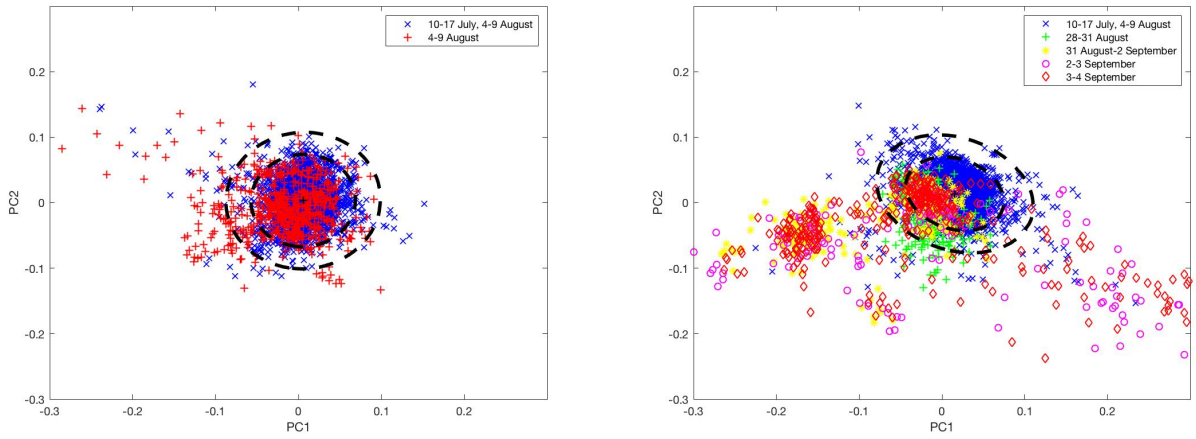
(b)

Fig. 2.32. *Sensitivity to the damage.* 11^{th} *cepstral coefficient for accelerometer* #12 *(a),* 6^{th} *cepstral coefficient for accelerometer* #7 *(b).*

In order to validate the proposed damage assessment strategy using the cepstral coefficients, two approaches are presented here. In the first approach (a "global" approach), the extraction of a "global" damage index consists of the projection along the principal directions of the cepstral coefficient sequences from all the observations for all the sensors, with the goal of providing information only on whether or not damage has occurred. Similarly to Case II, the idea behind this approach is to reduce the variance of the coefficients due to the local term γ so to provide a clustered training distribution (small variance) in the minor components. The second approach (a "local" approach) considers the projection of the cepstral coefficient sequences for each sensor individually so to obtain a "local" damage index, providing local information not only on the presence of the damage, but also on the location. Basically, the N_{ob}^{tr} cepstral sequences will be projected onto a different space depending on the sensor they belong to.

In the first approach, the dataset corresponding to the system in undamaged conditions consists of 672 observations ($N_{ob}^{tr} = 672$) of the structural acceleration recorded at the 4 sensor locations (sensors 05, 07, 10 and 12). From each time history, 50 cepstral coefficients were

extracted and, as previously explained, the first one was discarded ($nc = 49$). In order to test the system in undamaged conditions, 75% of the 672 observations were used in the training phase while the remaining 25% was used in the testing phase. Keeping only the minor components that contribute to less than 0.5% of the overall variance and performing the second PCA on the overall dataset, the proposed damage assessment algorithm is able to recognize that both the training dataset as well as the testing dataset are representative of the structure in the undamaged conditions. This is clearly shown in Figure 2.33 (a) where the two distributions of the projected features (the training and testing one) basically overlap. To test the system in a damaged state, the training, as well as the test datasets have to be redefined. The training dataset consists now of all the 672 observations while the testing dataset contains the observations collected during the period August 28^{th}-Sepstember 4^{th}. Looking at Figure 2.33 (b) it is evident that now the projected feature distributions do not overlap, clearly highlighting the presence of damage. However, some of the test observations fall under the percentile of two standard deviation of the Gaussian Mixture Model (GMM) fitting the training distribution (dashed lines in Figure 2.33 to indicate 2σ and 3σ) and this will imply that the damage index, computed as the median of the Squared Mahalanobis Distance, has a low probability to lead to false positive classifications.



(a) **(b)**

Fig. 2.33. *Training and Test distributions. Undamaged scenarios (a), cumulatively damaged scenarios (b).*

It is also interesting to look at each damage scenario case individually. Five different structural conditions (1 undamaged and 4 damaged) have been defined, as shown in Table 2.8.

Test Damage Scenario	Training Dataset	Test Dataset
Undamaged	10-17 July, 4-9 August (90%)	10-17 July, 4-9 August (10%)
Damaged	10-17 July, 4-9 August	28-31 August
Damaged	10-17 July, 4-9 August	31 August-2 September
Damaged	10-17 July, 4-9 August	2-3 September
Damaged	10-17 July, 4-9 August	3-4 September

Table 2.8. *Training and test datasets.*

The undamaged state serves as a reference to highlight the differences with the damaged states. For the undamaged case, three different sets of training and testing datasets have been created. In each one, following a cross-validation logic, 90% of the initial 672 observations have been randomly selected as representative of the training datasets while the remaining 10% constituted the testing dataset. For the damaged configurations of the system, the training and the testing datasets were created by considering the time histories recorded during the time periods reported in Table 2.8. Also in this case, considering the 2 principal components as damage sensitive features, the damage index has been defined as the median of the Squared Mahalanobis Distance of the test distribution from the training distribution. The results are presented in box plots in Figure 2.34.

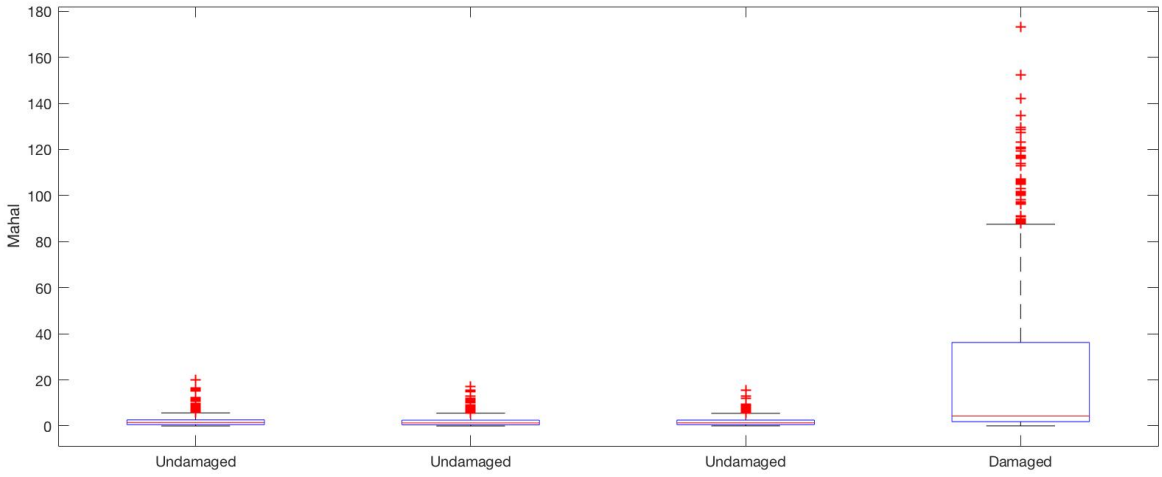
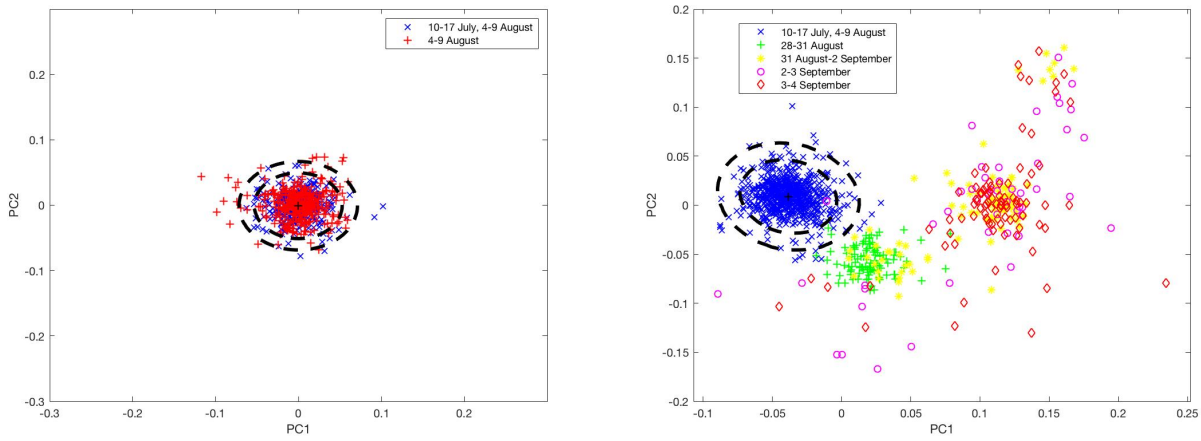


Fig. 2.34. *Squared Mahalanobis Distance over the damage scenarios.*

Of the three scenarios marked 'undamaged', the first is consistent with the representation provided by the components in Figure 2.33 (a) while the other two prove that the Squared Mahalanobis Distance of the projected test observations is an unbiased indicator. The distributions (and the median) of the Squared Mahalanobis Distance have similar values as long as the sys-

tem remains in the undamaged state. The occurrence of damage induces substantial changes in the distributions and their values.

In the second (local) approach, attention is now placed on the use of the cepstral coefficient sequences to provide local damage sensitive features. The training and testing datasets are the same as the ones from Table 2.8 but now the cepstral coefficients extracted at each sensor are treated individually. Thus, the damage sensitive features, defined by the projections on the first two principal components, are obtained at each sensor location: this should help not only in assessing the presence of damage but also in pinpointing its location. Figure 2.35 shows the distribution of the two principal components obtained processing the data only from sensor 12, testing the system in its undamaged state (Figure 2.35 (a)) or in the four damaged states (Figure 2.35(b)). Again, the minor components have been selected according to the variance threshold of 0.5% of the overall variance.



(a) **(b)**

Fig. 2.35. *Sensor 12: Training and Test distributions. Undamaged scenarios (a), damaged scenarios (b).*

As expected, the training and testing distributions cluster together when the testing dataset belongs to the system in its undamaged state while clearly diverge when the structure is damaged. Using the extracted damage sensitive features in the proposed damage assessment strategy allow us to estimate the damage index based on the Squared Mahalanobis Distance with a threshold now estimated equal to 4.63. Figure 2.36 shows the values of the proposed damage index at the four sensor locations for the 4 damage scenarios. From these results, it appears that accelerometers 10 and 12 seem to be the most sensitive to the presence of structural damage, even though some observations are missing for accelerometer 10 in the first damage scenario.

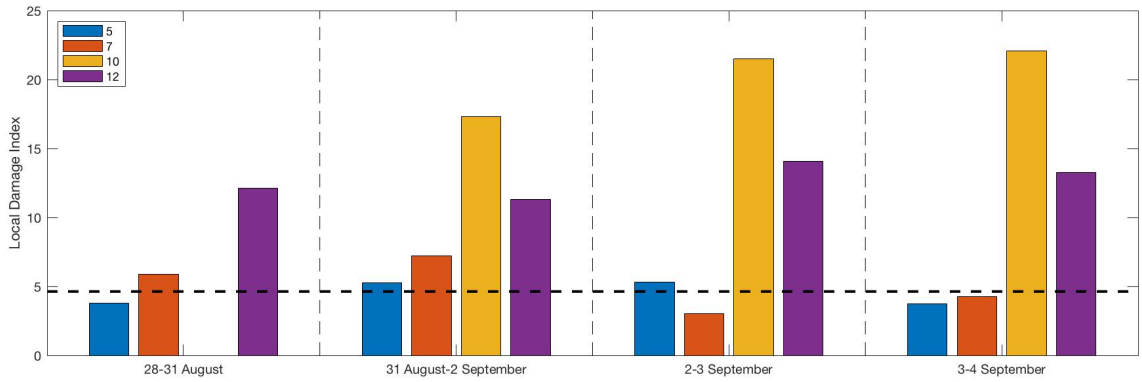


Fig. 2.36. *Local Damage Index over 4 damage scenarios. Each monitoring sensor is represented by a different colour in the bar plot.*

It is noteworthy that the data used in the entire analysis are referred to records from July 1998 to September 1998 so that the environmental conditions can be considered constant over the time period and no special filtering of data is needed (e.g. the temperature variation can be considered stationary over time). If the data recordings extend over multiple seasons, special techniques, e.g. cointegration [63, 64], and/or special features, e.g. frequency ratios [65], can be used to remove environmental effects. However, it is important to point out that, in the proposed methodology, the damage assessment relies on the use of the minor components which are the least affected by the environmental effects. This represents a great advantage since this technique could handle data recorded in different environmental conditions just in one analysis.

2.8. Cepstral Coefficients: the LANL Benchmark

The proposed damage assessment methodology has been investigated on data collected from the dynamic response of a 3-DOF shear-type system. Datasets have been provided by the Engineering Institute (EI) at Los Alamos National Laboratory (LANL) [66–68]. In its baseline condition, the system consists of four aluminum columns ($17.7 \times 2.5 \times 0.6$ cm) connected at the top and bottom to aluminum plates ($30.5 \times 30.5 \times 2.5$ cm) [69], forming a structure consisting of 3 floors and a sliding base. The excitation is provided by an electromagnetic shaker that acts at the center line of the base floor of the structure. Both the structure and the shaker are fixed on a base plate ($76.2 \times 30.5 \times 2.5$ cm). Four accelerometers with a nominal sensitivity of 1000 mV/g are attached at the center of the side of each floor at the opposite side from shaker to measure the response of each plate. The random excitation applied at the sliding base is band

limited in the range of 20-150 Hz to avoid rigid body modes of the structure.

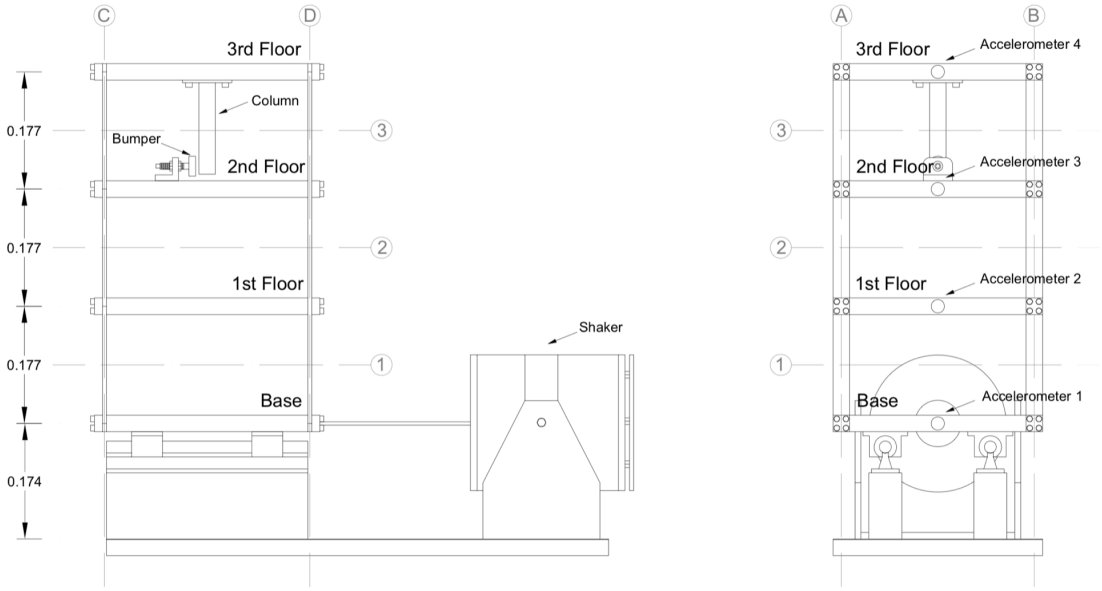


Fig. 2.37. *LANL benchmark: 4 DOFs shear-type system.*

2.8.1. Training and test datasets

The acceleration response time histories collected at each floor are signals of 8192 time steps with a sampling time of 0.0031 sec leading to records 25.3952 sec long. Multiple damage scenarios have been tested and, for each of them, 9 observations of acceleration response time histories have been collected. In order to generate richer a dataset consisting of a large number of observations, a framing procedure, according to [13] has been performed: the 8192 time steps long time histories have been framed into 142 time histories, partially overlapping, which are 1000 time steps long. This procedure led to a total of $142 \cdot 9 = 1278$ observations (frames) at each monitored structural location. For each structural DOF (floors 1 to 3), a training dataset that consists of $N_{ob}^{tr} = 1278$ observations (frames) from which the cepstral coefficient sequences are extracted. In a statistical pattern recognition framework, the distribution of the first nc cepstral coefficients of the cepstral coefficient sequences is considered to develop a training model. According to the formulation of the damage assessment methodology presented in 2.6, $N_{ob}^{tr} = 1278$ observations of the vector $\boldsymbol{c}_{tr}^{(j)} \in \mathbb{R}^{d\times 1}$ $(j = 1,\ldots,N_{ob}^{tr})$ containing the first $d = nc$ cepstral coefficients in the sequences are defined.

Three damage scenarios, reported in Table 2.9, are analyzed to investigate the effectiveness of the damage assessment methodology, based on the statistical distribution of the cesptral coefficients sequences, for local structural damage identification. For each damage scenarios,

a test dataset of $N_{ob}^{te} = 1278$ observations of framed acceleration response time histories are obtained by following the same procedure as the one used to generated the training dataset. Thus, $N_{ob}^{te} = 1278$ observations of the vector $\boldsymbol{c}_{te}^{(j')} \in \mathbb{R}^{d\times 1}$ ($j' = 1,\ldots,N_{ob}^{te}$) containing the first $d = nc$ cepstral coefficients in the sequences are obtained.

Label	Damaged State Condition
1	25% column stiffness reduction between base and 1^{st} floor
2	25% column stiffness reduction between 1^{st} and 2^{nd} floors
3	25% column stiffness reduction between 2^{nd} and 3^{rd} floors

Table 2.9. *Test damage scenarios.*

The median of the SMD of the test distribution of the $N_{ob}^{te} = 1278$ vectors $\boldsymbol{c}_{te}^{(j')}$ from the training distributions of the $N_{ob}^{tr} = 1278$ vectors $\boldsymbol{c}_{tr}^{(j)}$ is used as a damage index indicator so to suggest the presence of a local damage. Recalling that the very first cepstral coefficients of the cepstral coefficient sequences are those more sensitive to the local zeros of the structural response, i.e. to the structural local behaviour (section 2.2.2), different values of nc are analyzed ($nc = 1,\ldots,4$). Figure 2.38 shows the local damage index (at each floor), for the three presented damage scenarios: it can be observed that, given a damage scenario and considering nc cepstral coefficients, the value of the damage index is significantly larger when considering data collected at a position which is adjacent to the damage location. In section 2.7.1, it has been shown that, by considering very few coefficients ($nc = 1,2$), the accuracy of the damage detection algorithm is unacceptable. Hence, by focusing on the second damage scenario (column stiffness reduction between 1^{st} and 2^{nd} floors), for $nc = 1,2$ the damage index associated with the 3^{rd} floor (DOF) is similar to those computed at the 1^{st} and 2^{nd} floors (DOFs). Contrarily, for $nc = 3,4$, the damage index is always able to detect the local damage.

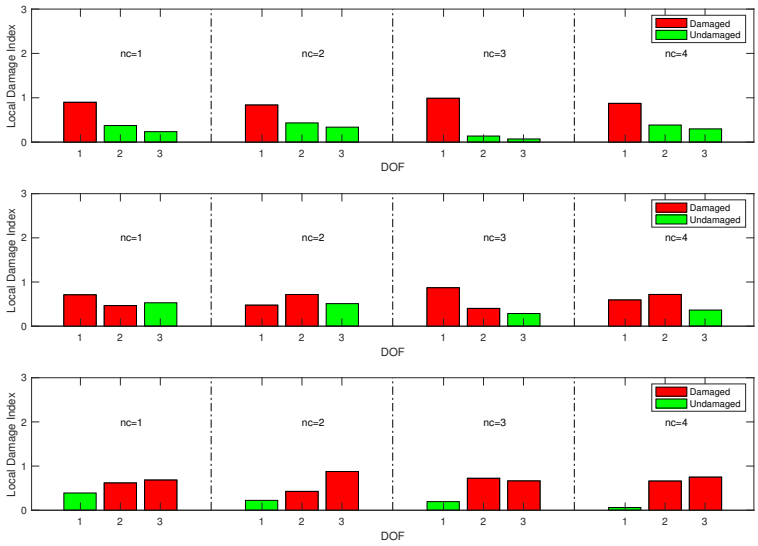


Fig. 2.38. *Damage index at every DOF over nc for each damage scenario. Damage scenarios from 1 to 3 in Table 2.9 are ordered from the bottom to the top.*

A visual representation of the distribution of the first $nc = 3$ elements in the vectors $\boldsymbol{c}_{tr}^{(j)}$ and $\boldsymbol{c}_{te}^{(j')}$ is provided by Figure 2.39. The first damage scenario (column stiffness reduction between base and 1^{st} floor) has been considered. The terms along the axes, $c_{i,q}$, indicate the q^{th} element of the cepstral coefficient sequence vectors (either $\boldsymbol{c}_{tr}^{(j)}$ or $\boldsymbol{c}_{te}^{(j')}$) extracted at the i^{th} DOF. The training (blue circles) and test (red circles) distributions are significantly separated for the datasets collected at DOF1, location that is the closest to the local damage.

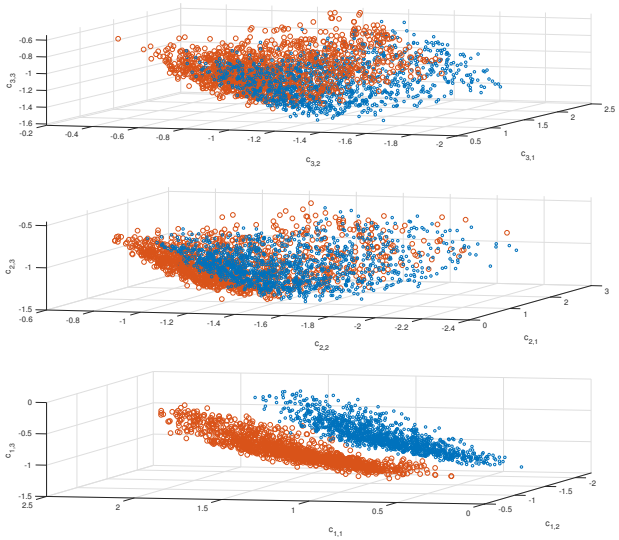


Fig. 2.39. *Cepstral coefficient distributions.*

2.8.2. Cepstral coefficients for nonlinear structural damage assessment

The structural damage assessment methodology based on the statistics on the cepstral coefficients distribution has been investigated by analyzing new test datasets consisting of acceleration response time histories generated by the LANL shear-type system subjected to nonlinearities. A vertical column is positioned at the center of the third floor of the structure and a bumper is installed on the second floor at a certain distance from the column. When the structure is subjected to the shaker excitation, the contact between the dumper and the column affects the structural dynamic response, simulating some nonlinearities of the system. Different damage scenarios are reported in Table 2.10: the gap between the column and the bumper is indicative of the entity of the nonlinearity introduced in the structure.

Two experiments have been conducted by training two different models relying on the acceleration response time histories collected by the sensors placed at the 3 floors and at the sliding base (4 monitored locations). For the first experiment, the training dataset consists of $N_{ob}^{tr} = 1278$ observations of the cepstral coefficients extracted from the acceleration response time histories of the system in its baseline condition (linear). For each of the state conditions labeled as $3, 4$ and 5 in Table 2.10, a test datasets of $N_{ob}^{te} = 1278$ observations of cepstral coefficient sequences has been collected. Let's focus on the training dataset. By selecting $nc = 7$ cepstral coefficients, each observation of the cepstral coefficient sequences leads to a training vector $\boldsymbol{c}_{tr}^{(j)} \in \mathbb{R}^{d\times 1}$ ($j = 1,\dots,N_{ob}^{tr}$). The vectors $\boldsymbol{c}_{tr}^{(j)}$ are obtained gathering the $nc = 7$ first cepstral coefficients extracted at each of the 3 floors (DOF) and at the sliding base so that $d = nc \cdot 4$. Analogously, for each test dataset, $N_{ob}^{te} = 1278$ test vectors $\boldsymbol{c}_{te}^{(j')} \in \mathbb{R}^{d\times 1}$ ($j' = 1,\dots,N_{ob}^{te}$) are generated for the considered damage scenarios. By following the same procedure, the training and test vectors are computed by considering the first nc AR coefficients to provide a comparison between cepstral coefficients and AR coefficients. Figure 2.41 shows the damage index computed as the median of the SMD of the distribution of the test datasets, associated with the presented damage scenarios, from the training distribution. Figure 2.41 (a) shows the damage indexes over the damage scenarios by considering the cepstral coefficients as damage sensitive features. For each damage scenario, the presence of structural damage is confirmed. Analogously, in Figure 2.41 (b) the damage index based on the AR coefficients allows to easily detect the damaged condition of the structure. Anyway, as mentioned in section 2.4.3, AR coefficients are damage sensitive features particularly performant for linear systems. These coefficients are generated by performing a linear regression of the dataset, so their estimation for nonlinear

models is biased. Figure 2.41 (b) shows that the damage indexes obtained by considering AR coefficients as damage sensitive features are strongly affected by that bias leading to numerically unstable values.

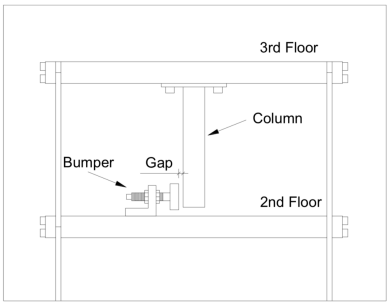


Fig. 2.40. *LANL Bumper between floors 2 and 3.*

Label	State condition
1	Baseline condition
2	Gap=0.20 mm
3	Gap=0.13 mm
4	Gap=0.10 mm
5	Gap=0.05 mm

Table 2.10. *Damage scenarios.*

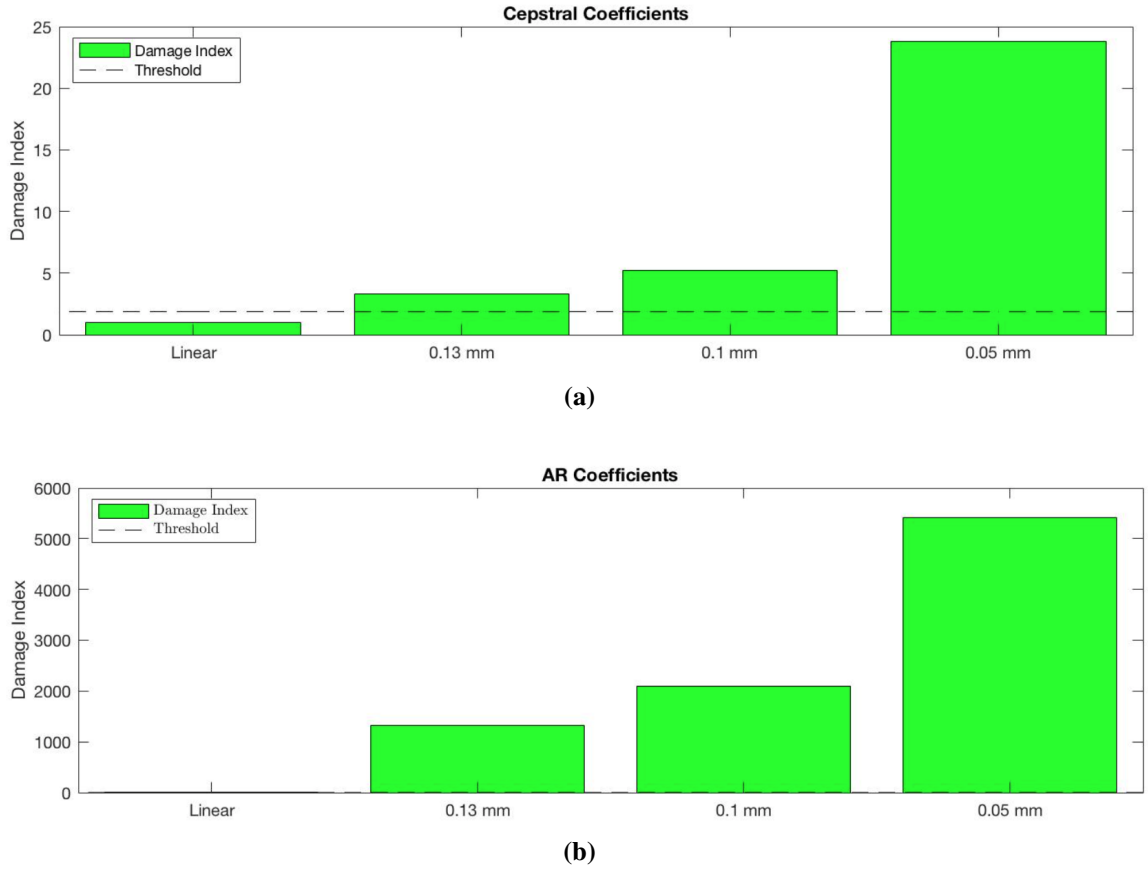


Fig. 2.41. *Damage Index for nonlinear damage scenarios. Cepstral coefficients (a) and AR coefficients (b) as damage sensitive features.*

The second experiment relies on a procedure analogous to the one used to obtain the results in Figure 2.41. In this experiment, the training dataset consists of $N_{ob}^{tr} = 1278$ observations of the acceleration response time histories from which cepstral coefficient sequences are extracted

when the structure is in the state condition labeled as 2 in Table 2.10. Thus, the model is trained on a nonlinear dynamic response of the structure. In this case, the AR coefficients extracted from both the training and test datasets are biased. Also in this case, as shown in Figure 2.42, the distribution of the cepstral coefficients leads to stable values of the damage index. Contrarily, the damage index obtained by considering the AR coefficients as damage sensitive features presents reasonable values for the damage scenarios 3 and 4 in Table 2.10 and dramatically increases for damage scenario 5.

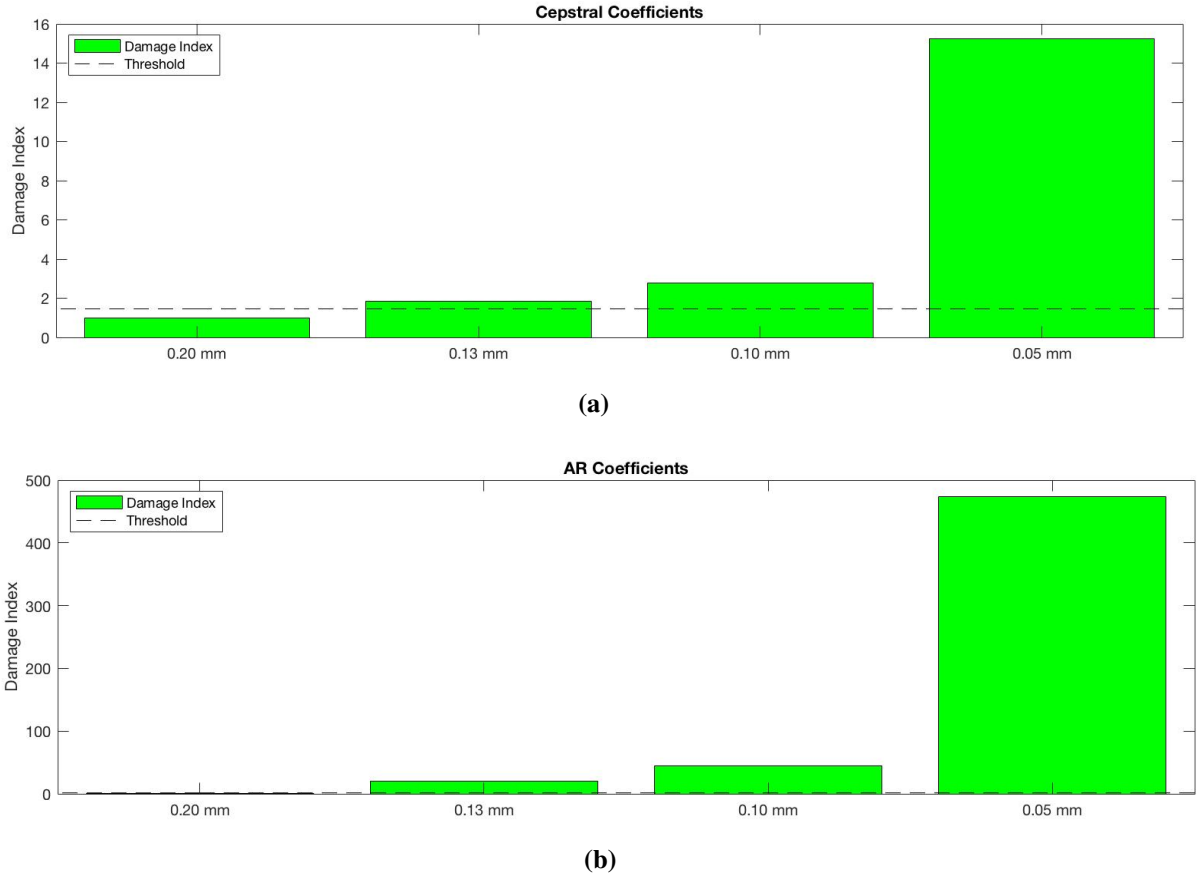


Fig. 2.42. *Damage Index for nonlinear damage scenarios. Cepstral coefficients (a) and AR coefficients (b) as damage sensitive features.*

The two experiments presented in this section lead to the conclusion that, compared with the AR coefficients, the cepstral coefficients are damage sensitive features which lead to more stable values of the damage index when the analyzed structure is subjected to nonlinear effects.

2.9. Conclusions

In this chapter the reliability of the power ceptrum of structural accelerations of mechanical systems for damaged sensitive features extraction was investigated. The analytical expression for the cepstral coefficients and their projection on a new space were obtained and used in a damage detection strategy. Three case studies were presented: 1) an 8DOF system excited through Gaussian white noise excitations having the same RMS at each DOF, 2) an 8DOF system excited by just one input force applied at either the 1^{st} DOF or the 8^{th} DOF, 3) the Z24 bridge in Switzerland. The excitation source which has been set for the test described in the first study case was meant to generate cepstal coefficients that follow an approximately Gaussian distribution. Different damage scenarios have been tested. An outlier analysis has been performed by considering the cepstral coefficients as damage sensitive feature and has successfully detected the presence of the damage. A correlation between the damage index defined by the Squared Mahalanobis Distance and the damage location has been shown. In the second case study the distribution of the cepstral coefficients was far from Gaussian and the Squared Mahalanobis Distance as metric indicative of the damage was not directly applicable. To circumvent this problem, a principal component analysis, aiming to obtain the minor components, was the key to extract damage sensitive features, less sensitive to external disturbances and random effects affecting the dynamic response, and mainly depend on a variation in the structural properties. It has been shown how those minor components are able to make different distributions collapse into one single clustered distribution leading to a possible analysis through the SMD indicating an alternative to the conventional methods based on mixture models. In the last case study the performance of the damage assessment method has been investigated on data from an actual structure subjected to a variety of external disturbances caused by the environment and excited by various excitations for part of the monitoring period. The outlier analysis of the projection of the cepstral coefficients successfully detected the presence of the damage. Furthermore, it is noteworthy to remark that the proposed approach drastically reduces the computational time required by a complete analysis in frequency domain. The data compression performed by cepstral coefficients may be a valuable option not only for storage and computing requirement, but also for their convenient mathematical representation.

Acknowledgements

The Z24 data have been collected within the Brite EuRam Programme BE-3157 SIMCES with a financial contribution from the European Commission. The data have been kindly provided by the project coordinator, the Structural Mechanics Section of KU Leuven. The authors gratefully acknowledge the Brite EuRam Programme BE-3157 SIMCES, the European Commission and the the Structural Mechanics Section of KU Leuven for gathering and sharing the data of the Z24 bridge.

Chapter 3

3. The Inner Product Vector as an output-only cross-correlation-based feature to structural damage assessment

Among the damage sensitive features, the Inner Product Vector (IPV) has shown great promise for applications in a damage assessment strategy. In the original formulation of the IPV (Wang et al. [70]), a carefully designed input excitation, used to excite a specific structural mode, was applied to the structure and the dynamic response at different locations directly used to compute their cross-correlation. The main advantage of using these features is that no computation of the modal parameters is required: they only rely on a data analysis. This approach has been successfully used in numerous publications [71–75]. The shortcoming of this approach is that it requires two tests on the structure: the first test is used to determine the optimal frequency range to design the input excitation that is going to be used in the second test. This is a condition that can be satisfied in a laboratory environment, but not in a typical field application where only the results from one test are available and the excitation cannot be controlled. It is for these reasons that the methodology proposed in this chapter is addressing the calculations of the IPV in an output-only framework.

In this chapter, the theoretical formulation of the IPV using the time histories of the structural response has been derived for both the cases of unit impulse and white noise excitations. The identified IPVs are obtained through the cross-correlation of the properly filtered structural response at various locations and used in a damage index vector for damage assessment and localization. Different reference points in the calculation of the cross-correlation of the response have been considered for validation.

Numerical simulations on a 8-DOF shear-type and on a 100-DOF 2-D structural models have shown the effectiveness of the proposed damage assessment methodology, accounting also for the effects due to additional disturbances (measurement noise, environmental conditions, unidentified modes etc.). The effectiveness of the proposed methodology has been also validated by considering experimental results from a 3-DOF shear-type laboratory system.

3.1. Inner Product Vector

Let's consider a generic dynamic system discretized in N lumped mass elements whose acceleration response time histories are monitored. The cross-correlation between the acceleration time histories recorded at 2 different locations is investigated in order to obtain valuable damage sensitive features. For that purpose, the N acceleration time histories (one for each DOF) are gathered in an array $\{\ddot{x}(t)\} = \{\ddot{x}_1(t), \ddot{x}_2(t), ..., \ddot{x}_N(t)\}^T$. The cross-correlation between the accelerations can be evaluated pairwise for a generic time lag T. As it will be proved in the next sections, computing the cross-correlation for $T = 0$ (which represents the Inner Product Vector) and making some reasonable assumptions lead to the extraction of reliable damage sensitive features from the acceleration time histories.

3.1.1. Single input case

A generic dynamic N-DOF system can be represented through the equations of motion for a linear time-invariant model as follows:

$$[M]\{\ddot{x}(t)\} + [C]\{\dot{x}(t)\} + [K]\{x(t)\} = \{f(t)\} \tag{3.1}$$

where $[M], [C], [K] \in \mathbb{R}^{N\times N}$ are respectively the mass matrix, the damping matrix and the stiffness matrix. The term $\{f(t)\} \in \mathbb{R}^{N\times 1}$ represents the vector of the forcing functions applied on the lumped masses. The arrays $\{\ddot{x}(t)\}$, $\{\dot{x}(t)\}$, $\{x(t)\} \in \mathbb{R}^{N\times 1}$ are respectively the acceleration, the velocity and the displacement vectors. As a result of the linear modal analysis, $\{x(t)\}$ can be written as a linear combination of mode shape vectors $\{\Phi_r\} \in \mathbb{R}^{N\times 1}$ for $r = 1, 2, \ldots, N$, gathered in the modal matrix $[\Phi]$=$[\{\Phi_1\}, \{\Phi_2\}, ..., \{\Phi_N\}]$, multiplied by some scalar function of time $q_r(t)$ ($r = 1, 2, \ldots, N$):

$$\{x(t)\} = [\Phi]\{q(t)\} = \sum_{r=1}^{N} \{\Phi_r\} q_r(t). \tag{3.2}$$

Using the orthogonality property of the mode shapes, the equations of motions can then be decoupled by substituting Equation (3.2) into Equation (3.1) and pre-multiplying all the members by $[\Phi]^T$. The dynamic system can then be represented as a set of N scalar equations in the

modal coordinates $q_r(t)$.

$$\ddot{q}_r(t) + 2\xi_r\omega_{nr}\dot{q}_r(t) + \omega_{nr}^2 q_r(t) = \frac{1}{m_r}\{\Phi_r\}^T\{f(t)\}. \tag{3.3}$$

The coefficients ξ_r, ω_{nr} and m_r in Equation (3.3) are respectively the damping ratio, the undamped natural frequency and the modal mass for the r^{th} mode shape. Assuming zero initial conditions, the solution for $q_r(t)$, is provided by the Duhamel Integral:

$$q_r(t) = \int_{-\infty}^{t} \{\Phi_r\}^T\{f(\tau)\}g_r(t-\tau)d\tau \tag{3.4}$$

where $g_r(t)$ is the unit pulse response function related to the r^{th} mode. Then, substituting Equation (3.4) into Equation (3.2) leads to:

$$\{x(t)\} = \sum_{r=1}^{N}\{\Phi_r\}\int_{-\infty}^{t}\{\Phi_r\}^T\{f(\tau)\}g_r(t-\tau)d\tau. \tag{3.5}$$

To simplify the derivation of the Inner Product Vector, let us first consider the case of a single input force applied at k^{th} location. In this case, the displacement at i^{th} position induced by an arbitrary force at the k^{th} position can be computed as:

$$x(t)_{i,k} = \sum_{r=1}^{N}\Phi_{i,r}\Phi_{k,r}\int_{-\infty}^{t} f_k(\tau)g_r(t-\tau)d\tau. \tag{3.6}$$

Note that $\Phi_{i,r}$ denotes the i^{th} component of the r^{th} mode shape. Recalling the formula of the unit-pulse response for the displacement:

$$g_r(t) = \begin{cases} 0 & t < 0 \\ \frac{e^{-\xi_r\omega_{nr}t}}{m_r\omega_{dr}}sin(\omega_{dr}t) & t \geq 0 \end{cases} \tag{3.7}$$

where ω_{dr} is the damped natural frequency of the r^{th} mode, the unit-pulse response for velocity and acceleration can be obtained by simply taking the first and second derivative of $g_r(t)$ with respect to time of Equation (3.7), obtaining respectively:

$$\dot{g}_r(t) = \frac{e^{-\xi_r\omega_{nr}t}}{m_r\omega_{dr}}\left[-\xi_r\omega_{nr}sin(\omega_{dr}t) + \omega_{dr}cos(\omega_{dr}t)\right] \qquad t > 0 \tag{3.8}$$

and

$$\ddot{g}_r(t) = \frac{e^{-\xi_r \omega_{nr} t}}{m_r \omega_{dr}} [\xi_r^2 \omega_{nr}^2 sin(\omega_{dr} t) - 2\xi_r \omega_{nr} \omega_{dr} cos(\omega_{dr} t) - \omega_{dr}^2 sin(\omega_{dr} t)] \qquad t > 0. \tag{3.9}$$

By considering Equation (3.6) and applying the Leibniz formula, the velocity at the i^{th} location induced by a force at the k^{th} location can be obtained as:

$$\begin{aligned} \dot{x}(t)_{i,k} &= \sum_{r=1}^{N} \Phi_{i,r} \Phi_{k,r} \left[\int_{-\infty}^{t} f_k(\tau) \dot{g}_r(t-\tau) d\tau + f_k(t) g_r(0) \right] \\ &= \sum_{r=1}^{N} \Phi_{i,r} \Phi_{k,r} \int_{-\infty}^{t} f_k(\tau) \dot{g}_r(t-\tau) d\tau. \end{aligned} \tag{3.10}$$

while the acceleration at the i^{th} location induced by a force at the k^{th} location can be expressed as:

$$\begin{aligned} \ddot{x}(t)_{i,k} &= \sum_{r=1}^{N} \Phi_{i,r} \Phi_{k,r} \left[\int_{-\infty}^{t} f_k(\tau) \ddot{g}_r(t-\tau) d\tau + f_k(t) \dot{g}_r(0^+) \right] \\ &= \sum_{r=1}^{N} \Phi_{i,r} \Phi_{k,r} \left[\int_{-\infty}^{t} f_k(\tau) \ddot{g}_r(t-\tau) d\tau + \frac{f_k(t)}{m_r} \right]. \end{aligned} \tag{3.11}$$

The term $\dot{g}_r(0^+)$ is due to the fact that the unit-pulse response for displacement g_r is not derivable in 0. The ratio $\frac{f_k(t)}{m_r}$ inside of Equation (3.11) can be rewritten as:

$$\frac{f_k(t)}{m_r} = 2 \int_{-\infty}^{t} \frac{f_k(\tau)}{m_r} \delta(t-\tau) d\tau = \int_{-\infty}^{t} f_k(\tau) \left[\frac{2}{m_r} \delta(t-\tau) \right] d\tau \tag{3.12}$$

so that Equation (3.11) becomes:

$$\begin{aligned} \ddot{x}(t)_{i,k} &= \sum_{r=1}^{N} \Phi_{i,r} \Phi_{k,r} \left[\int_{-\infty}^{t} f_k(\tau) \ddot{g}_r(t-\tau) d\tau + \int_{-\infty}^{t} f_k(\tau) \left[\frac{2}{m_r} \delta(t-\tau) \right] d\tau \right] \\ &= \sum_{r=1}^{N} \Phi_{i,r} \Phi_{k,r} \int_{-\infty}^{t} f_k(\tau) \ddot{g}'_r(t-\tau) d\tau \end{aligned} \tag{3.13}$$

where $\ddot{g}'_r(t-\tau) = \left[\ddot{g}_r(t-\tau) + \frac{2}{m_r} \delta(t-\tau) \right]$. So far no assumption has been made about the nature of the excitation $f_k(t)$.

3.1.2. Unit pulse and white noise input

Referring to Equation (3.13), the cross-correlation between two acceleration time-histories (one recorded at the i^{th} position and one at the j^{th} position) induced by an external excitation applied at the k^{th} location can be computed as the expected value $\mathbb{E}$ of the product of such signals delayed by a time lag T as follows:

$$R_{i,j,k}(T) = \mathbb{E}[\ddot{x}(t+T)_{i,k}\ddot{x}(t)_{j,k}] \tag{3.14}$$

such that, using Equation (3.13):

$$R_{i,j,k}(T) = \sum_{r=1}^{N}\sum_{s=1}^{N}\Phi_{i,r}\Phi_{k,r}\Phi_{j,s}\Phi_{k,s}\int_{-\infty}^{t}\int_{-\infty}^{t+T}\mathbb{E}[f_k(\sigma)f_k(\tau)]\ddot{g}'_r(t-\sigma+T)\ddot{g}'_s(t-\tau)d\tau d\sigma \tag{3.15}$$

The terms in Equation (3.15) are related to the deterministic parameters of the system as well as to the force f_k. If we assume that the force $f_k(t)$ represents a *unit pulse* excitation, defined by the Dirac Delta $\delta(t)$, then this leads to:

$$\mathbb{E}[f_k(\sigma)f_k(\tau)] = \delta(\sigma-\tau) \tag{3.16}$$

Analogously, if $f_k(t)$ is a *white noise* excitation, the term $\mathbb{E}[f_k(\sigma)f_k(\tau)]$ in Equation (3.15) becomes:

$$\mathbb{E}[f_k(\sigma)f_k(\tau)] = \alpha_k\delta(\sigma-\tau) \tag{3.17}$$

where α_k is a positive coefficient depending on the statistics of the force acting at location k. Equation (3.16) provides the same result of Equation (3.17) given $\alpha_k = 1$. Therefore an analysis about the cross-correlation between signals under white noise excitation will be automatically valid for the unit pulse excitation. Equation (3.15) and Equation (3.17) combined yield:

$$R_{i,j,k}(T) = \sum_{r=1}^{N}\sum_{s=1}^{N}\alpha_k\Phi_{i,r}\Phi_{k,r}\Phi_{j,s}\Phi_{k,s}\int_{-\infty}^{t}\ddot{g}'_r(t-\tau+T)\ddot{g}'_s(t-\tau)d\tau. \tag{3.18}$$

Then, substituting $\lambda = t-\tau$ into Equation (3.18), the cross-correlation becomes:

$$R_{i,j,k}(T) = \sum_{r=1}^{N}\sum_{s=1}^{N}\alpha_k\Phi_{i,r}\Phi_{k,r}\Phi_{j,s}\Phi_{k,s}\int_{0}^{\infty}\ddot{g}'_r(\lambda+T)\ddot{g}'_s(\lambda)d\lambda. \tag{3.19}$$

Assuming that we are able to separate the contribution of a certain mode r' from those of the other modes for both $\ddot{x}(t+T)_{i,k}$ and $\ddot{x}(t)_{j,k}$, we can first calculate $\ddot{g}_{r'}(\lambda)$ from Equation (3.9) and combine it with Equation (3.19) to obtain the expression for $R_{i,j,k,r'}(0)$:

$$R_{i,j,k,r'}(0) = \alpha_k \Phi_{i,r'} \Phi_{j,r'} \Phi_{k,r'}^2 \int_0^\infty \ddot{g}_{r'}^{\prime 2}(\lambda) d\lambda. \tag{3.20}$$

$R_{i,j,k,r'}(0)$ represents the cross-correlation, at time lag $T = 0$, of the r'^{th} components of the structural accelerations recorded at the i^{th} and j^{th} locations, induced by an external excitation at the k^{th} location. $R_{i,j,k,r'}(0)$ can be computed for different positions $i = 1,2,\ldots,N$ and $j = 1,2,\ldots,N$, considering the N acceleration time histories due to the same force at position k exciting the r'^{th} mode. Considering the location j as the reference point, then, the elements $R_{1,j,k,r'}(0), R_{2,j,k,r'}(0) \ldots R_{N,j,k,r'}(0)$ can be gathered into a cross-correlation vector $\{R_{j,k,r'}(0)\} \in \mathbb{R}^{N\times 1}$ defined as Inner Product Vector (IPV). In order to simplify the notation in Equation (3.20), all the positive terms can be gathered into a positive coefficient $\psi_{k,r'}$:

$$\psi_{k,r'} = \alpha_k \Phi_{k,r'}^2 \int_0^\infty \ddot{g}_{r'}^{\prime 2}(\lambda) d\lambda \geq 0. \tag{3.21}$$

This can be zero only when the k^{th} component of the r'^{th} mode, $\phi_{k,r'}$, is zero (e.g. the r'^{th} mode has a node at the k^{th} location). Hence, the cross-correlation vector at time lag $T = 0$ is thus obtained:

$$\{R_{j,k,r'}(0)\} = \Phi_{j,r'} \psi_{k,r'} \{\Phi_{r'}\}. \tag{3.22}$$

The dependance of the IPV on k vanishes via L-2 normalization. In fact, the normalized IPV $\{\hat{R}_{jr'}(0)\}$ can be obtained:

$$\begin{aligned} \{\hat{R}_{j,r'}(0)\} &= \frac{\{R_{j,k,r'}(0)\}}{||\{R_{j,k,r'}(0)\}||_2} = \frac{\Phi_{j,r'} \psi_{k,r'} \{\Phi_{r'}\}}{|\Phi_{j,r'} \psi_{k,r'}| \sqrt{\Phi_{1,r'}^2 + \Phi_{2,r'}^2 + .. + \Phi_{N,r'}^2}} \\ &= \frac{sign(\Phi_{j,r'})}{\sqrt{\Phi_{1,r'}^2 + \Phi_{2,r'}^2 + .. + \Phi_{N,r'}^2}} \{\Phi_{r'}\} \end{aligned} \tag{3.23}$$

indicating that the IPV associated with a given mode (in this case r') and with a given reference point (e.g. j) is proportional to the given mode with a sign corresponding to the component of the given mode at the reference point. It is noteworthy to remark that, by normalizing the

vector $\{R_{j,k,r'}(0)\}$, it is automatically assumed $\Phi_{j,r'} \neq 0$ and $\Phi_{k,r'} \neq 0$ meaning that none of the j^{th} and k^{th} components at the corresponding locations is a node for the r' mode. The vector $\{\bar{\Phi}_{r'}\} = \{\Phi_{r'}\}/\sqrt{\Phi_{1,r'}^2 + \Phi_{2,r'}^2 + .. + \Phi_{N,r'}^2}$ represents the normalized r'^{th} mode, so that:

$$\sqrt{\bar{\Phi}_{1,r'}^2 + \bar{\Phi}_{2,r'}^2 + .. + \bar{\Phi}_{N,r'}^2} = 1 \tag{3.24}$$

and, considering that $sign(\Phi_{j,r'}) = sign(\bar{\Phi}_{j,r'})$, we obtain:

$$\{\hat{R}_{j,r'}(0)\} = sign(\bar{\Phi}_{j,r'})\{\bar{\Phi}_{r'}\}. \tag{3.25}$$

Equation (3.25) shows that there is a direct connection between the cross-correlation vector $\{\hat{R}_{j,r'}(0)\}$ and the normalized mode shape r'. In order to assess the evolution of the r' mode shape of the system passing from an undamaged condition to an unknown (potentially damaged) condition, a damage index vector can be defined as the difference between the IPVs corresponding to the different situations:

$$\{D_{j,r'}\} = \{\hat{R}_{j,r'}^d(0)\} - \{\hat{R}_{j,r'}^u(0)\} = sign(\bar{\Phi}_{j,r'}^d) \begin{pmatrix} \bar{\Phi}_{1,r'}^d \\ \bar{\Phi}_{2,r'}^d \\ \vdots \\ \bar{\Phi}_{N,r'}^d \end{pmatrix} - sign(\bar{\Phi}_{j,r'}^u) \begin{pmatrix} \bar{\Phi}_{1,r'}^u \\ \bar{\Phi}_{2,r'}^u \\ \vdots \\ \bar{\Phi}_{N,r'}^u \end{pmatrix} \tag{3.26}$$

where $\{D_{j,r'}\}$ indicates the damage index vector for the mode r' with a reference location at point j. If we can assume now that the reference point j is not a node for the mode r' and that the occurring damage does not change the sign of the element j of the mode r', so that $sign(\bar{\Phi}_{j,r'}^d) = sign(\bar{\Phi}_{j,r'}^u)$, then:

$$\{D_{j,r'}\} = sign(\bar{\Phi}_{j,r'}^u) \left(\begin{pmatrix} \bar{\Phi}_{1,r'}^d \\ \bar{\Phi}_{2,r'}^d \\ \vdots \\ \bar{\Phi}_{N,r'}^d \end{pmatrix} - \begin{pmatrix} \bar{\Phi}_{1,r'}^u \\ \bar{\Phi}_{2,r'}^u \\ \vdots \\ \bar{\Phi}_{N,r'}^u \end{pmatrix} \right). \tag{3.27}$$

Equation (3.27) implies that the reference location j, arbitrarily chosen, defines only the sign of the vector $\{D_{j,r'}\}$. Thus, the selection of a reference location j related to an element $\bar{\Phi}_{j,r'}^u$ of the vector $\{\bar{\Phi}_{r'}^u\}$ yields a damage index vector defined as $\{D_{j,r'}\}$. Contrarily, by choosing

a reference location $j' \neq j$ related to an element $\bar{\Phi}^u_{j',r'}$ of the vector $\{\bar{\Phi}^u_{r'}\}$ that has an opposite sign to $\bar{\Phi}^u_{j,r'}$, the damage index vector is given by $\{-D_{j,r'}\}$.

As specified in [76–80] it is known that the local damage occurring between two lumped mass elements causes a discontinuity in the difference between the damaged and undamaged eigenvectors. The relation between the damage index vector $\{D_{j,r'}\}$ and the normalized r'^{th} mode shown in Equation (3.27), let us conclude that the elements of the former can be considered as local damage sensitive features and help us locate the damaged area(s).

It is important to point out that the formulation provided by Equation (3.26) is consistent with the one proposed by Wang et al. [70]. However, in addition to the fact that in this study, contrarily to [70], only output information is considered, the original approach in [70] accounted only for the contributions of the first vibrational mode in the calculation of the IPVs. Instead, the additional assumptions made in this study open the door to the analysis of IPVs from different structural modes and this represents one of the novelties of the proposed approach. This freedom will allow to consider low frequency modes other than the first mode, whose extraction might be difficult because of external (e.g. measurement noise, thermal effects, etc.) and internal (e.g. rigid body modes, aliasing, etc.) disturbances.

3.1.3. Multiple input case

The general solution for the displacement vector computed through the Duhamel Integral, Equation (3.5) is at the core of the approach to multiple input analysis treated in this section. The analogous relation for accelerations is given by Equation (3.28).

$$\{\ddot{x}(t)\} = \sum_{r=1}^{N} \{\Phi_r\} \int_{-\infty}^{t} \{\Phi_r\}^T \{f(\tau)\} \ddot{g}'_r(t-\tau) d\tau \tag{3.28}$$

where now $\{f(\tau)\}$ indicates a force vector containing the N time histories of the external excitation (some of them could be zeros). Let's now focus on a system excited at all the DOFs such that the Equation (3.28) above can be rewritten as:

$$\{\ddot{x}(t)\} = \sum_{r=1}^{N} \sum_{k=1}^{N} \Phi_{k,r} \{\Phi_r\} \int_{-\infty}^{t} f_k(\tau) \ddot{g}'_r(t-\tau) d\tau. \tag{3.29}$$

Thus, the acceleration at the i^{th} location can be expressed as:

$$\ddot{x}_i(t) = \sum_{r=1}^{N}\sum_{k=1}^{N} \Phi_{k,r}\Phi_{i,r} \int_{-\infty}^{t} f_k(\tau)\ddot{g}'_r(t-\tau)d\tau. \tag{3.30}$$

By taking into account the contribution of the r' mode to the acceleration at the i^{th} location:

$$\ddot{x}_{i,r'}(t) = \sum_{k=1}^{N} \Phi_{k,r'}\Phi_{i,r'} \int_{-\infty}^{t} f_k(\tau)\ddot{g}'_{r'}(t-\tau)d\tau \tag{3.31}$$

and considering the same contribution for the acceleration at the j^{th} location, the cross-correlation between these two contributions can be expressed as

$$\begin{aligned} R_{i,j,r'}(T) &= \mathbb{E}[\ddot{x}_{i,r'}(t+T)\ddot{x}_{j,r'}(t)] \\ &= \sum_{k=1}^{N}\sum_{k'=1}^{N} \Phi_{i,r'}\Phi_{j,r'}\Phi_{k,r'}\Phi_{k',r'} \int_{-\infty}^{t}\int_{-\infty}^{t+T} \mathbb{E}[f_k(\sigma)f_{k'}(\tau)]\ddot{g}'_{r'}(t-\sigma+T)\ddot{g}'_{r'}(t-\tau)d\sigma d\tau. \end{aligned} \tag{3.32}$$

If the input force is represented by *white noise* excitation, then, when $k = k'$, $\mathbb{E}[f_k(\sigma)f_k(\tau)]$ represents the autocorrelation function of the excitation and can be expressed by Equation (3.17). On the other hand, when $k \neq k'$, the cross-correlation is null if the excitations at the k^{th} and k'^{th} locations are uncorrelated with each other. Those observations lead to a further simplification of Equation (3.32):

$$R_{i,j,r'}(T) = \sum_{k=1}^{N} \Phi_{i,r'}\Phi_{j,r'}\Phi_{k,r'}\Phi_{k,r'} \int_{-\infty}^{t}\int_{-\infty}^{t+T} \mathbb{E}[f_k(\sigma)f_k(\tau)]\ddot{g}'_{r'}(t-\sigma+T)\ddot{g}'_{r'}(t-\tau)d\sigma d\tau. \tag{3.33}$$

Setting the time lag $T = 0$ and solving for the integral lead us to an expression analogous to Equation (3.20) with a linear superposition for the N excitations acting simultaneously:

$$R_{i,j,r'}(0) = \sum_{k=1}^{N} \alpha_k \Phi_{i,r'}\Phi_{j,r'}\Phi_{k,r'}^2 \int_0^{\infty} \ddot{g}'^2_{r'}(\lambda)d\lambda. \tag{3.34}$$

The cross-correlation in Equation (3.34) can be decomposed as the sum of N cross-correlation elements from Equation (3.20) having the common multiplier $\Phi_{i,r'}\Phi_{j,r'}$.

$$R_{i,j,r'}(0) = \Phi_{i,r'}\Phi_{j,r'} \sum_{k=1}^{N} \left(\alpha_k \Phi_{k,r'}^2 \int_0^{\infty} \ddot{g}'^2_{r'}(\lambda)d\lambda \right) = \Phi_{i,r'}\Phi_{j,r'}\psi'_{r'} \tag{3.35}$$

where $\psi'_{r'} = \sum_{k=1}^{N} \left(\alpha_k \Phi_{k,r'}^2 \int_0^\infty \ddot{g}_{r'}^{\prime 2}(\lambda) d\lambda \right)$ indicates a sum of either zero or positive terms. It can then be concluded that the vector $\{R_{i,j,r'}(0)\}$, containing the cross-correlations of the r' contributions of all the acceleration time histories with that at the reference j^{th} location, can be rewritten as:

$$\{R_{j,r'}(0)\} = \Phi_{j,r'} \psi'_{r'} \{\Phi_{r'}\}. \tag{3.36}$$

The relation obtained is analogous to the one provided by Equation (3.22). Thus, we can proceed to the normalization of the cross correlation vector $\{R_{j,r'}(0)\}$ yielding $\{\hat{R}_{j,r'}(0)\}$ (Equation (3.25)) and finally provide the damage index vector $\{D_{j,r'}\}$ defined in Equation (3.27).

3.2. Damage detection through a local damage index vector

Until now, the definition of a valid damage index vector has been the object of many studies conducted by researchers like Wang et al. [71], Trendafilova and Manoach [72], Kim and Stubbs [73]. In this chapter, the damage index vector is defined according to Wang et al. [71] even though, as previously mentioned, Wang's experiment relies on specific hypothesis about the input which has been strategically designed. According to Equation (3.26), the damage index vector can be written as:

$$\{D_{j,r'}\} = \{\hat{R}_{j,r'}^d(0)\} - \{\hat{R}_{j,r'}^u(0)\} \tag{3.37}$$

where $\{D_{j,r'}\} \in \mathbb{R}^{N \times 1}$ for a full sensors setup, when the acceleration response time history is monitored at every DOF of the system. A more general formulation of Equation (3.37) can be expressed as:

$$\{D_{IPV,j,r'}\} = \{\hat{R}_{IPV,j,r'}^d(0)\} - \{\hat{R}_{IPV,j,r'}^u(0)\} \tag{3.38}$$

where the vectors $\{D_{IPV,j,r'}\}$, $\{\hat{R}_{IPV,j,r'}^d(0)\}$ and $\{\hat{R}_{IPV,j,r'}^u(0)\}$ are subsets of the vectors $\{D_{j,r'}\}$, $\{\hat{R}_{j,r'}^d(0)\}$ and $\{\hat{R}_{j,r'}^u(0)\}$ respectively and have dimension $Q \leq N$, given a sensors setup that monitors the structure at only Q locations. An additional consideration about the dimensionality Q of the vectors in Equation (3.38) has to be pointed out. Despite the fact that it can be represented by a properly discretized model, any real (continuous) dynamic system has an infinite number of degrees of freedom. For such a reason, the restriction $Q \leq N$ doesn't affect

in practice the number of sensors we can use to obtain the damage index vector.

A graphical representation of the damage index vector $\{D_{IPV,j,r'}\}$ is provided by plotting its elements over their respective monitored locations (or DOFs). For sake of clarity, let's consider the particular case of an 8-DOF system whose acceleration response time histories have been collected in both a damaged and an undamaged state at every DOF so to compute the corresponding damage index vector. In this example, the damaged state represents the condition of the system with a reduction of stiffness at a given location. Because of the localized damage, a local abrupt change in the damage index vector is expected.

Depending on the structural boundary conditions (structural constrains) and material properties, the damage index vector can present different types of jump discontinuity and so a *local damage index vector* $\{L_{IPV,j,r'}\}$ can be defined based on the type of jump discontinuity presented in the plot of the damage index vector. Three possible cases are here reported, considering the way that the structural damage affects the mechanical characteristics of the structure:

1. If all the DOFs but the closest to the damage location are more sensitive to the structural constrains than to the occurrence of damage, the damage index is approximatively null at those locations away from the damaged one. Figure 3.1 (a) shows a possible configuration plot of the damage index vector for the case in which, by considering a local damage between DOFs 3 and 4, only the 4^{th} element of $\{D_{IPV,j,r'}\}$ shows an appreciable variation (in this specific case, it is arbitrarily set equal to 1 to provide a clear graphical representation). The plot of the damage index vector is similar to one representing an *impulse change*. For such a specific case, the local damage index $\{L_{IPV,j,r'}\}$ is defined as the damage index vector itself so that $\{L_{IPV,j,r'}\} = \{D_{IPV,j,r'}\}$.
2. It might happen that, for specific boundary conditions and material properties, the presence of a local damage may induce a *step change* as jump discontinuity in the trend of the plotted elements of the damage index vector. An example is shown in Figure 3.1 (b) where the two trends are represented by arbitrarily setting the elements of the vector to zeros and ones. Again, the damage has been introduced between the DOFs 3 and 4. In this case, the local damage index vector $\{L_{IPV,j,r'}\}$ is provided by the first order difference of the damage index vector, $\{D'_{IPV,j,r'}\}$. The elements of $\{D'_{IPV,j,r'}\}$ related to the position halfway between two adjacent monitored positions i and $i+1$ are given

by:

$$D'_{IPV,(i+0.5),j,r'} = D_{IPV,(i+1),j,r'} - D_{IPV,(i),j,r'} \tag{3.39}$$

where the index $(i+0.5)$ of the element $D'_{IPV,(i+0.5),j,r'}$ indicates the location halfway between the i^{th} and $i+1^{th}$ elements of the vector $\{D_{IPV,j,r'}\}$ indicated by $D_{IPV,(i),j,r'}$ and $D_{IPV,(i+1),j,r'}$. In this particular case $i = 1,2,\ldots,Q-1$ so that $\{D'_{IPV,j,r'}\} \in \mathbb{R}^{Q-1\times 1}$. In this case, the local damage index vector is defined as $\{L_{IPV,j,r'}\} = \{D'_{IPV,j,r'}\}$.

3. Finally, for some boundary conditions and material properties, it is possible that all the elements of the damage index vector are sensitive to the local damage, and their plot looks like a *weak impulse*. This is the case depicted by Figure 3.1 (c) in which a damage between the DOFs 4 and 5 is represented. For this particular case to locate the damaged area, the local damage index vector $\{L_{IPV,j,r'}\}$ can be defined as the second order difference of the damage index vector, $\{D''_{IPV,j,r'}\}$. The elements of the vector $\{D''_{IPV,j,r'}\}$ are computed as follows:

$$D''_{IPV,(i+1),j,r'} = D_{IPV,(i+2),j,r'} - 2D_{IPV,(i+1),j,r'} + D_{IPV,(i),j,r'} \tag{3.40}$$

for $i = 0,1,\ldots,Q-1$, setting, for convenience, $D_{IPV,(0),j,r'} = D_{IPV,(Q+1),j,r'} = 0$ so that $\{D''_{IPV,j,r'}\} \in \mathbb{R}^{Q\times 1}$.

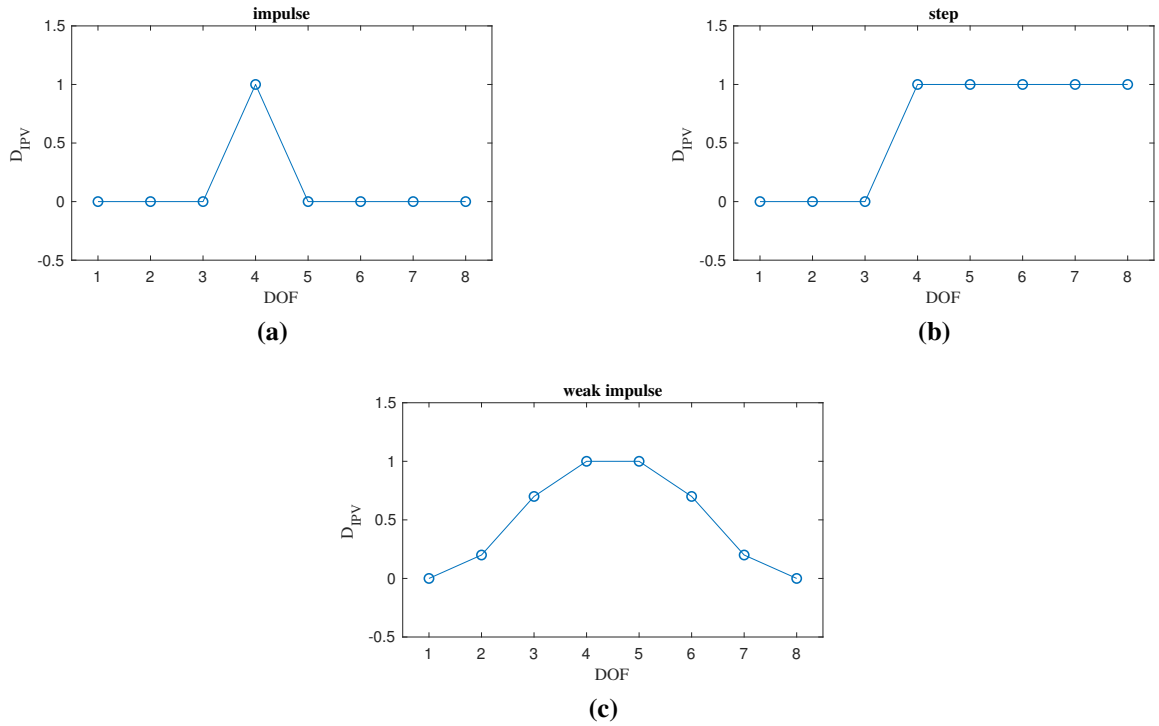


Fig. 3.1. Impulse change (a), step change (b), weak change (c)

In summary, Figure 3.1 represents the three abrupt changes in the damage index vector, each of which is related to particular structural cases. For each of them it is possible to derive a local damage index vector $\{L_{IPV,j,r'}\}$ linked to the change of the damage index vector: this operation is usually performed manually, but classification algorithms that rely on cross-correlation or classifiers can be designed in order to automate such process.

The final goal of this damage assessment algorithm is to detect abrupt changes in the elements of the local damage index vector $\{L_{IPV,j,r'}\}$ to assess the presence of locations of potential damage. In order to quantify the entity of these abrupt changes, the introduction of a threshold value for the elements of the local damage index vector is necessary, as shown in the next section.

3.2.1. Damage threshold for the local damage index vector

The most suitable approach for the definition of a threshold for local damage index vector is the one proposed by Wang et al. [6, 20] based on the statistics (mean and the standard deviation) of the elements of such a vector. The upper and lower threshold boundaries, t_u and t_l, are defined

as:

$$t_u = \mu_D + \beta_c \sigma_D \tag{3.41}$$

$$t_l = \mu_D - \beta_c \sigma_D \tag{3.42}$$

where μ_D and σ_D are respectively the mean and the standard deviation of the local damage index vector $\{L_{IPV,j,r'}\}$. The term β_c is a constant value set to define a confidence interval and is commonly assumed to be $1 \leq \beta_c \leq 1.8$. For instance, in case of normal distribution of the values of the elements of the local damage index vector, the choice of $\beta_c = 1.2$, $\beta_c = 1.5$ and $\beta_c = 1.8$ leads to a confidence interval respectively of 76.99%, 86.64% and 92.81%. When the threshold is overcome by some values of the elements of the local damage index vector, the structure is claimed to be damaged and the location of the local damage is detected. A brief example is reported in Figure 3.2.

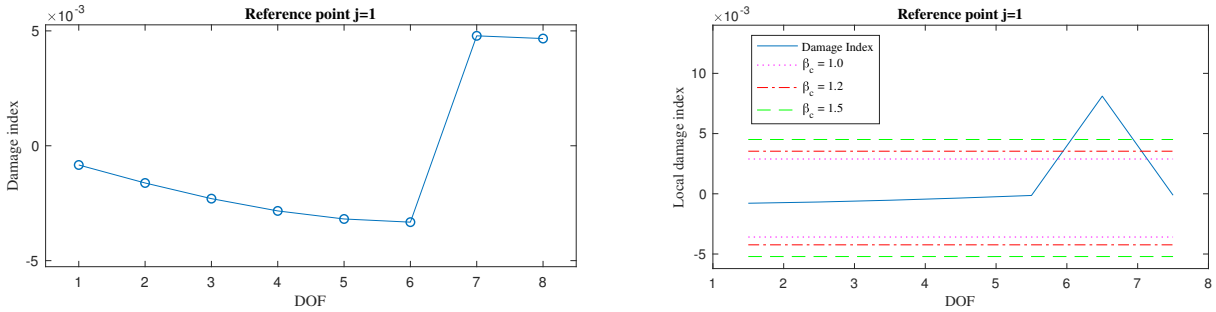


Fig. 3.2. Damage index vector $\{D_{IPV,j,r'}\}$ classified as step change (a) and local damage index vector $\{L_{IPV,j,r'}\}$ (b).

Let's focus, for instance, on an 8-DOF system whose damaged state is characterized by a localized damage between the DOFs 6 and 7. By looking at Figure 3.2 (a) the damage index vector $\{D_{IPV,j,r'}\}$ for $j = 1$ and $r' = 1$ can be easily classified in the *step change* class (Figure 3.1 (b)) while the corresponding local damage index vector $\{L_{IPV,j,r'}\}$, is plotted in Figure 3.2 (b). Also in Figure 3.2 (b), the thresholds are shown for $\beta_c = 1.0$, $\beta_c = 1.2$ and $\beta_c = 1.5$. It is evident that the plot of the local damage index vector indicates that damage has occurred between the DOFs 6 and 7.

3.3. The IPV in an output-only framework

The effectiveness of the IPV algorithm has been exhaustively proven by Wang et al. In their study [70] they analyzed a lumped element mock up excited by a low-pass (LP) filtered input: this input was designed so to excite exclusively the first natural frequency of the undamaged system. Obviously, this step requires a preliminary laboratory test in order to identify the natural frequencies of the system. After recording the acceleration response time histories for the system in undamaged conditions, local damage was introduced in the structure and, using a band-passed input with the same cut-off frequencies as the one used in the undamaged structure, the acceleration response of the frame in damaged conditions was recorded. By applying the IPV theory based on the cross-correlation of the signals for the undamaged and damaged states, the algorithm accurately identified the location of the damage (*IPV* in Figure 3.3).

As mentioned, the condition of a properly designed band-pass (BP) filtered signal to induce a specific system excitation is a valid option for laboratory tests only: it implies that the structure be subjected to two tests, one to determine the frequency range of interest and the other to collect data to be used in the identification of the IPV. In the methodology presented in this chapter, only one test is required. In fact, instead of filtering the input excitation, it is proposed that the filtering procedure be performed on the acceleration response time histories from the only test. Such a filtering procedure is allowed as long as all the theoretical assumptions at the base of the proposed methodology (e.g. richness and independence of the input forces) are respected. Hence, the approach presented in this chapter will then be referred to as *output-only IPV* (Figure 3.4): it allows us select the response contribution of the r'^{th}structural mode to be isolated and examine it directly from the spectrum of the response output. After the selected component of the acceleration response has been isolated, the IPV method can be applied to obtain the local damage index vector.

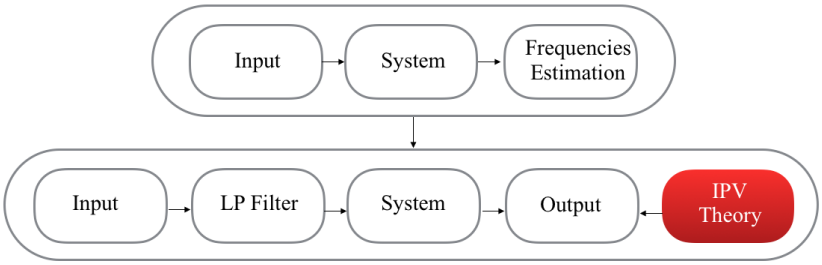


Fig. 3.3. *IPV* (Wang et al.)

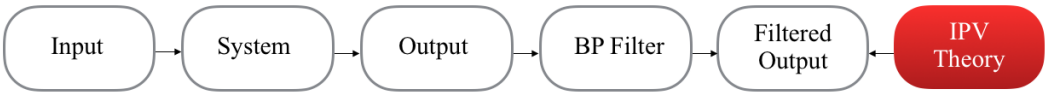


Fig. 3.4. *Output-only IPV* (Filtered Output)

The proposed methodology has some advantages with respect to the original IPV-based approach [70] and to the conventional parametric system identification based approaches (e.g. Output-Only Observer Kalman filter [81], Stochastic Subspace Identification [82], etc.) for damage assessment purposes. This method extends the boundaries of the original IPV approach by allowing an output-only analysis using information from higher modes than the first one. The possibility of using multiple reference points allows us to indirectly extract information about the analyzed structural mode shape by looking at the characteristics of the identified IPVs. Another main advantage of the proposed method is its easiness in handling large datasets from dense sensor networks: traditional identification methods rely on regression models and might suffer from the curse of dimensionality when dealing with large covariance matrices. Instead, the calculations associated with the cross-correlation vectors for the IPVs are much more efficient in terms of computational efforts. However, there are also some drawbacks in the proposed methodology: the main one is the undesired contribution of other structural modes that can be difficult to be filtered out from the one considered and may induce undesired noise effects.

3.4. Analysis of the results

To evaluate the performance of the proposed output-only IPV method in assessing and locating structural damage, numerical simulations as well as experimental data have been considered.

3.5. Numerical simulation: 8-DOF shear-type

The model of the structure is an 8-DOF shear-type model matching the one employed by [75] and is shown in Figure 3.5 (a). In its baseline undamaged conditions, the system is characterized by springs of stiffness $k_i = 25000$ N/m and masses $m_i = 1$ kg for $i = 1,2,\ldots,8$. The frame is characterized by modal damping with a damping factor of $\xi_i = 1\%$ for each of the 8 vibration modes. The force excitation is applied horizontally on the 8-DOF model via zero-order-hold (ZOH) with a time sampling of 0.01 seconds: it is a zero-mean Gaussian signal with standard

deviation $\sigma = 1$ N and is applied at the top of the model (Figure 3.5 (a)). The input/output time histories are 100 seconds long leading to $10,000$ time steps for each signal. In order to simulate a local damage, the 8-DOF system has been weakened by decreasing the stiffness of the spring between the 4^{th} and 5^{th} floors by 20%.

Once the acceleration response time histories have been simulated, a spectral analysis can be carried out in order to select a suitable mode to be investigated. In this first example, in order to show the effectiveness of the proposed approach, let us assume that there is no measurement noise in the response signals (it will be included at a later stage).

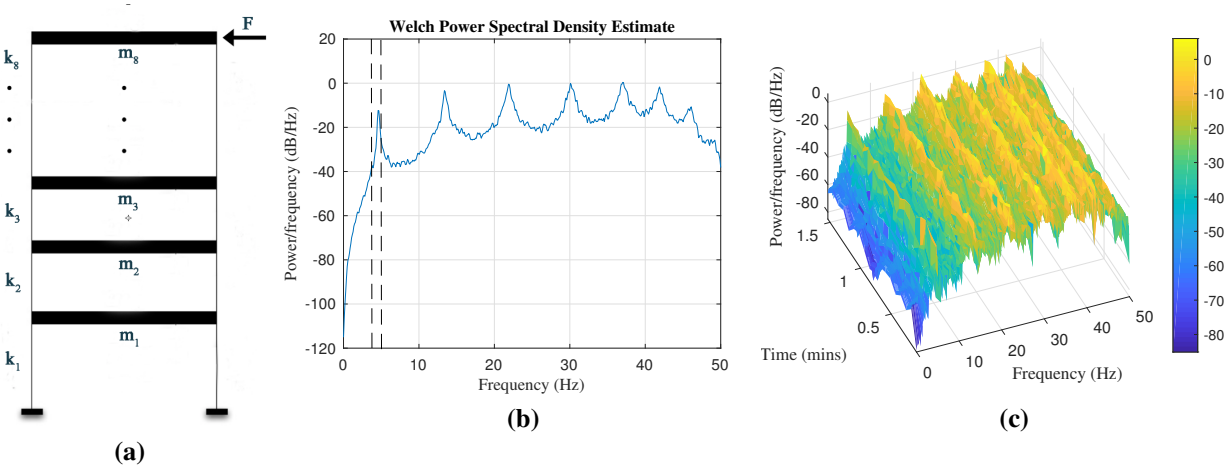


Fig. 3.5. 8-DOF shear type system (a). Spectral analysis for the 1^{st} floor: Power Spectral Density (b).

In this case, all the modes appear to be well excited and separated, thus we can select one of them and filter out the contributions of the other modes. It is noteworthy to point out that generally, for civil engineering applications, e.g. buildings, it is common practice to explore lower frequencies rather than the higher ones because they have better resolution and they are more easily excited. Therefore, the proposed algorithm is applied to data obtained by considering only the contribution of the *first* mode. Based on the spectral analysis, the lower and upper cut-off frequencies of the band pass filter have been set to 4.3 Hz and 5 Hz respectively (see Figure 3.5 (b)).

In the proposed methodology, the computation of the IPVs (Equation (3.25)) and, consequently, of the damage index $\{D_{j,r'}\}$ (Equation (3.27)) requires the selection of an arbitrary j^{th} reference point. The components of the damage index vector $\{D_{IPV,j,1}\}$ for any choice of the reference point $j = 1,2,\ldots,8$ are reported in Figure 3.6.

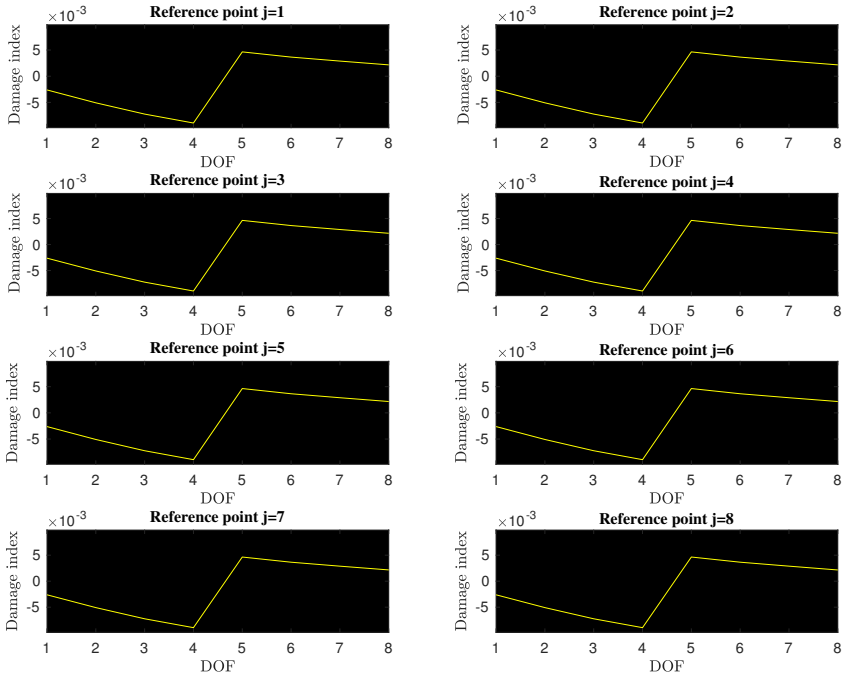


Fig. 3.6. Damage index vector for $j = 1, 2, ...8$ as defined in Equation (3.27), $r' = 1$.

Looking at Figure 3.6, it can be concluded that, no matter what j^{th} reference position is selected, the damage index vector always shows a jump discontinuity between the lumped elements bounding the damaged spring. Such kind of discontinuity let the damage index vector $\{D_{IPV,j,1}\}$ be classified into the step change class (Figure 3.1 (b)). Thus, the elements of the local damage index vector $\{L_{IPV,j,r'}\}$ can be obtained using Equation (3.39). The local damage index vector for $j = 1$ is plotted over the monitored DOFs in Figure 3.7 (b) with the threshold set for three different values of β_c (1.0, 1.2 and 1.5). Clearly, from the analysis of the results, it appears that the proposed algorithm is successful in detecting the damage between the 4^{th} and 5^{th} floors for any of the three values of β_c. Figure 3.7 (a) shows an analogous result for a damage simulated by 10% drop in stiffness between the DOFs 4 and 5.

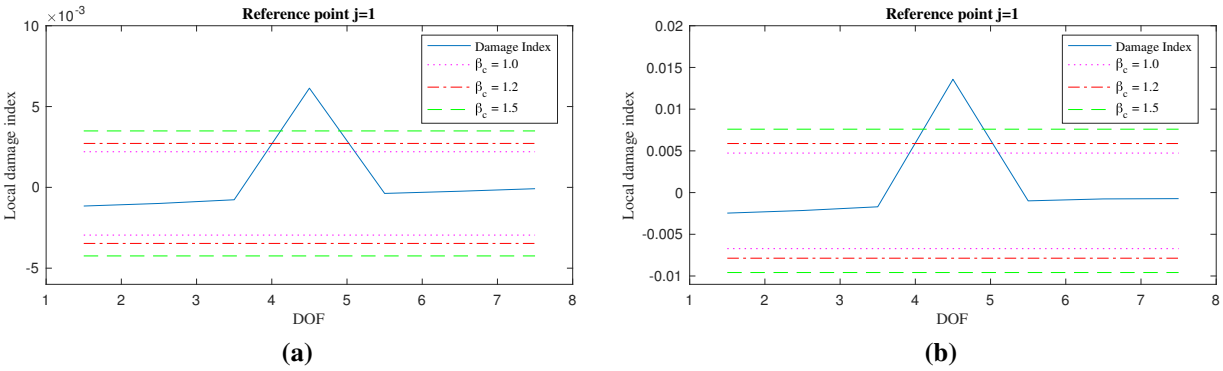


Fig. 3.7. Local damage index vector, 10% Damage (a), 20% Damage (b), $r' = 1$.

These numerical simulations based on an output-only approach can be compared with those obtained following the approach by Wang et al. [70] which rely on input-output information. By considering this noise free experiment, a band pass filter applied to the acceleration response time histories leads to the same results obtained by filtering the excitation source. It is worthy to recall that, in this specific case, the frequency contribution provided by modes other than the first one ($r' \neq 1$) have been considered negligible in the filtered frequency band. Of course, this condition implies that the natural frequencies related to those neglected modes should be far enough from the one of the analyzed mode or, at least, the contribution of those modes to the total response should not be significant in the vicinity of such a frequency.

As the spectral analysis suggests, also the mode related to the second peak appearing in Figure 3.5 (b) ($r' = 2$) can be isolated and used in the estimation of the IPVs taking advantage of the flexibility of the proposed methodology. For this purpose, the lower and upper cut-off frequencies of the Finite Impulse Response (FIR) filter have been set equal to 13 Hz and 14.5 Hz respectively. Again, the damage index vector $\{D_{IPV,j,2}\}$ is computed for any reference point $j = 1,2,...8$ and such vectors are presented in Figure 3.8. A jump discontinuity, similar to the one already shown in Figure 3.6, appears between the DOFs 4 and 5, revealing the presence of damage. However, an interesting observation can be made by looking at the plots in Figure 3.8: it appears that the damage index vectors computed for $j = 1,\ldots,5$ are 'mirror' images of those obtained for $j = 6,\ldots,8$. The reason why two different types of damage index vectors appear is due to the fact that, recalling Equation (3.27), the sign of the damage index vector depends on the sign of the component of the mode in question at the j^{th} reference location ($sign(\bar{\Phi}^{u}_{j,r'})$). Thus, moving from the reference point $j = 5$ to the reference point $j = 6$, the damage index changes its sign because there is a sign change between the 5^{th} and 6^{th} components of the second mode. Since this is a peculiarity of the mode analyzed (in this case the second), it can be concluded that the damage index vectors also provide information about the selected mode shape.

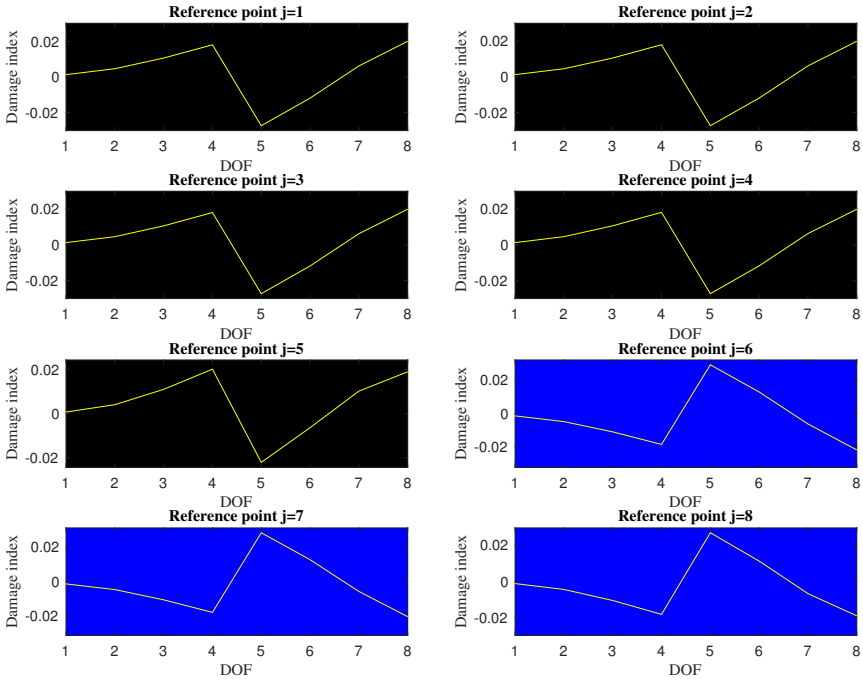


Fig. 3.8. Damage index vector for $j = 1,2,...8$ ad defined in Equation (3.27). $r' = 2$

Analogously to the previous analysis ($r' = 1$), also in this case the damage index vector $\{D_{IPV,j,2}\}$ is claimed to belong to the class defined as a step change and the local damage index vector $\{L_{IPV,1,2}\}$ (for $j = 1$) is shown in Figure 3.9 for a drop in stiffness of the 10% (a) and for a drop in stiffness of the 20% (b).

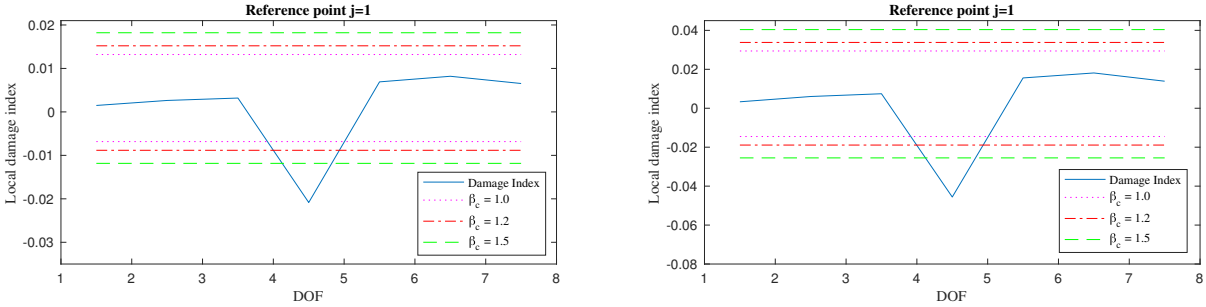


Fig. 3.9. Local damage index vector, 10% Damage (a), 20% Damage (b), $r' = 2$.

To show the effect of the sign of the component of the selected mode at the reference location on the damage index vector, the sign of the first two mode shapes $sign(\bar{\Phi}^u_{j,r'})$ ($r' = 1$ and $r' = 2$) for reference locations $j = 1,2,\ldots,8$ is shown in Figure 3.10 together with the corresponding mode shapes. The bars, providing an estimate of $sign(\bar{\Phi}^u_{j,1})$ given by changes in the sign of the damage index vector $\{D_{IPV,j,1}\}$, have been arbitrarily set equal to +0.5 when $sign(\bar{\Phi}^u_{j,1})$ is positive and equal to -0.5 when it is negative. According to Figure 3.6, for $r' = 1$

the damage index vector $\{D_{IPV,j,1}\}$ never changes its sign over different reference locations since the components of the first mode have all the same sign (Figure 3.10 (a)). On the contrary, looking at Figure 3.8, the damage index vector $\{D_{IPV,j,2}\}$ obtained for $r' = 2$ changes its sign moving from the reference location $j = 5$ to $j = 6$, in agreement with the second mode shape (for both damaged and undamaged conditions).

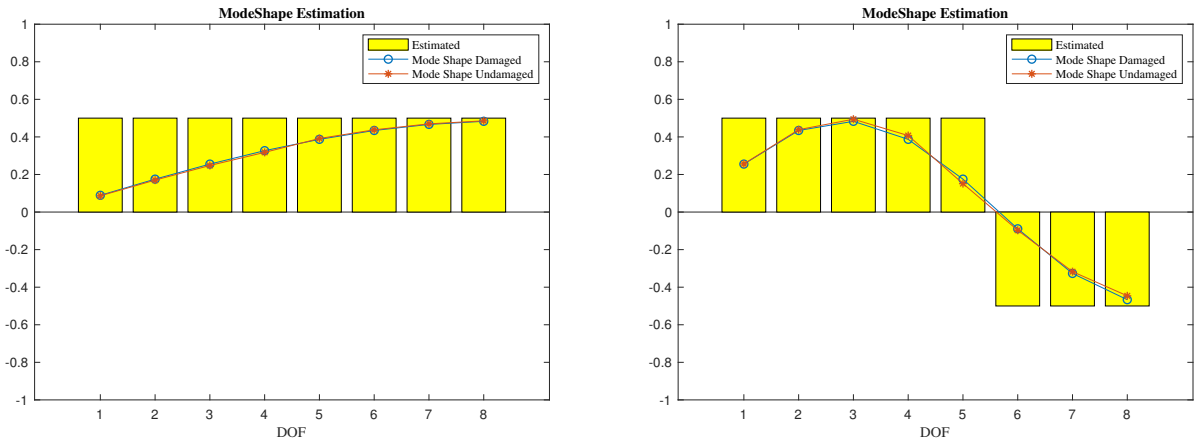


Fig. 3.10. First mode shape estimation (a), Second mode shape estimation (b)

A remarkable peculiarity of the IPV is that, even if the acceleration response time histories provided by some sensors are missing, still the IPV theory can be applied based on the available data. For example, let's now assume that the dataset of the monitored 8-DOF system lacks information from sensor 4 and sensor 7. The local damage index vectors $\{L_{IPV,j,1}\}$ and $\{L_{IPV,j,2}\}$ are computed and, for sake of brevity, only those with reference point $j = 1$ are reported in Figure 3.11.

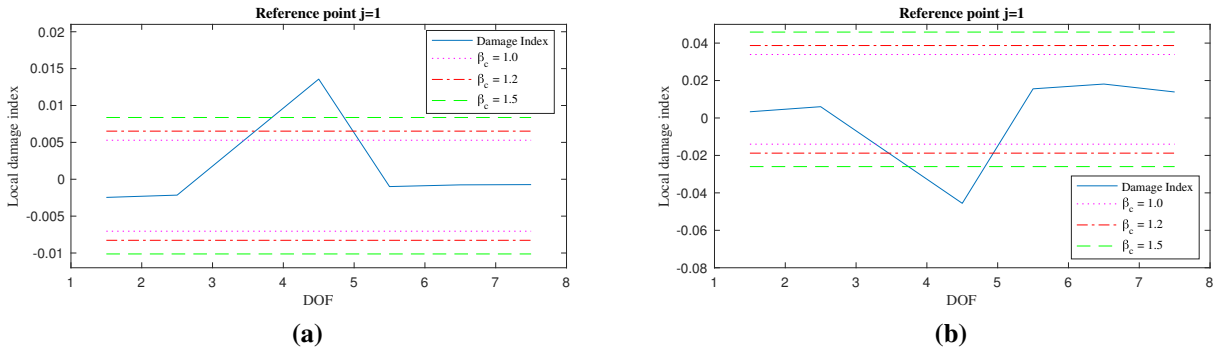
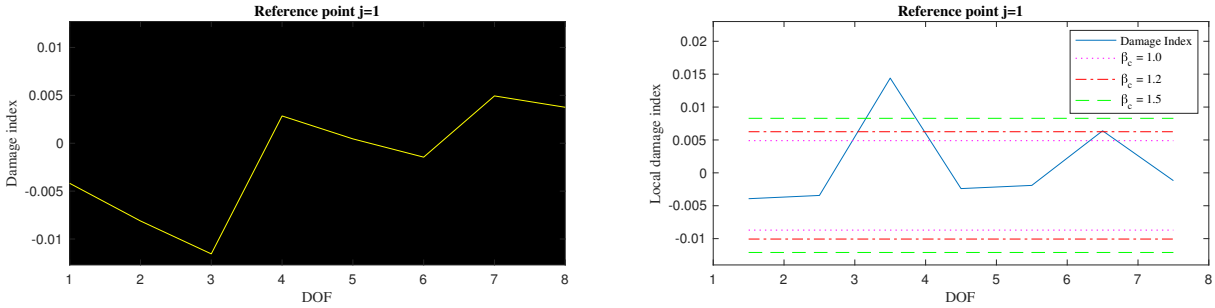


Fig. 3.11. Local damage index vectors: $r' = 1$ (a), $r' = 2$ (b), missing sensors at DOFs 4 and 7

It is clear that both local damage index vectors (either $\{L_{IPV,1,1}\}$ or $\{L_{IPV,1,2}\}$) are able to detect the occurrence of the damage between accelerometers 3 and 5, even if some sensors are

not available. However, this lack of information reduces the accuracy of the damage localization.

It is important to mention that, even in the case of multiple damage locations, the proposed methodology is successful in locating the damaged areas. Figure 3.12 shows the damage index vector $\{D_{IPV,1,1}\}$ and the corresponding local damage index vector $\{L_{IPV,1,1}\}$ for a double damage occurrence between the DOFs 3 and 4 and the DOFs 6 and 7.



(a) **(b)**

Fig. 3.12. Double local damage: damage index vector (a), local damage index vector (for step change) (b)

Finally it should be remarked that, as long as the single mode's contribution to the acceleration responses can be isolated, the normalization in Equation (3.23) removes any effect of the input. Consequently, any change in the position and/or magnitude of the input from the undamaged to the damaged configuration doesn't affect the final solution. This is one of the advantages of the proposed methodology because tests are generally performed under different excitation configurations and so removing the requirement of identical testing conditions from the undamaged and damaged tests free engineers from unnecessary constrains.

3.5.1. Fully excited system: effects of measurement noise

To look at the impact of external disturbances on the accuracy of the results, let's consider the same 8-DOF shear-type system subjected to an external excitation at every DOF. Each input force is represented by a zero-mean Gaussian white noise signal, with standard deviation of 1 N, and it is uncorrelated with the others. The disturbance representing measurement noise has been modelled as an additional zero-mean white noise signal, having a root mean square (RMS) equal to a certain percentage of the RMS of the output, and added to the output signals. Because of the stochastic nature of the external white noise excitation, a statistical approach

based on Monte Carlo simulations has been used to highlight the effect of noise disturbances on the damage index vector.

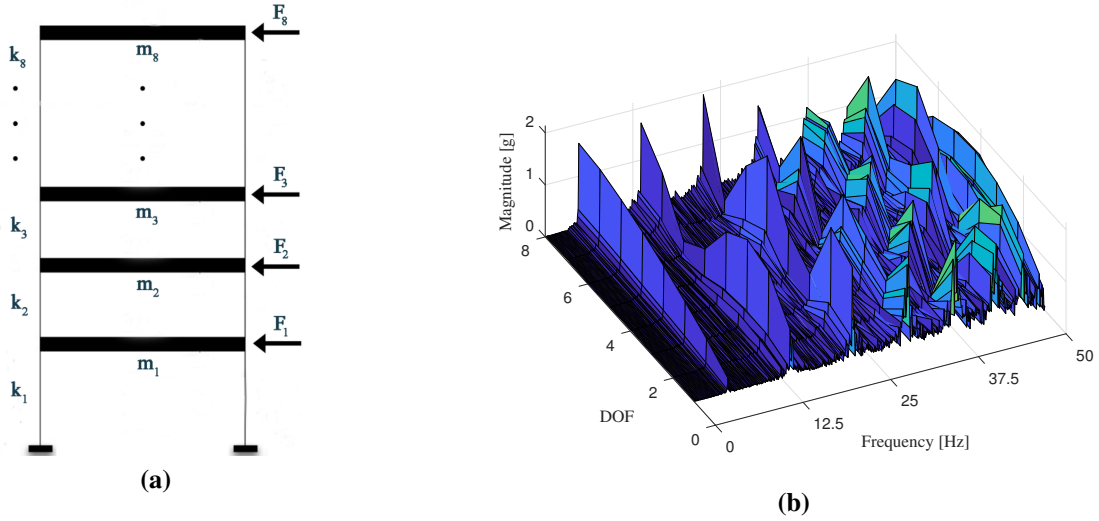


Fig. 3.13. Excitation setup (a), Waterfall plot (b).

The baseline configuration of the structure and the excitation setup are shown in Figure 3.13 (a), whereas the waterfall plot of the magnitude of the spectrum of the acceleration time histories, at each DOF, is presented in Figure 3.13 (b). Also in this case, the first mode ($r' = 1$) seems to be the perfect candidate for the structural damage assessment through the proposed methodology. Through Monte Carlo simulations, 50 realizations of the cross-correlation vectors $\{\hat{R}^d_{IPV,j,1}(0)\}$ and $\{\hat{R}^u_{IPV,j,1}(0)\}$ have been generated in order to obtain the damage index vector $\{D_{IPV,j,1}\}$. Damage has been simulated by introducing a 20% stiffness reduction between the DOFs 4 and 5. Local damage index vectors $\{L_{IPV,j,1}\}$ for a noise with RMS of 1%, 5%, 10% and 20% are reported in Figure 3.14. These plots show that, as long as the assumptions behind of the theory of the IPV are fully respected, the damage index shows a remarkable robustness to white noise disturbances and remains a good indicator of damaged areas.

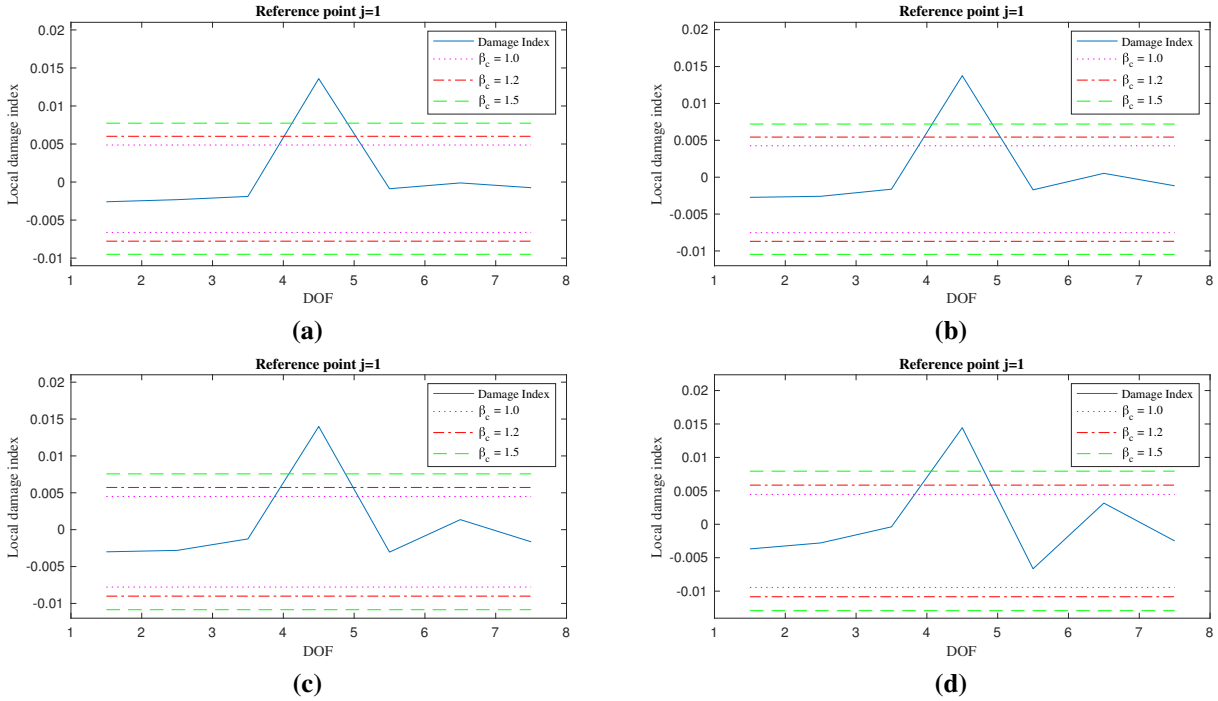


Fig. 3.14. Local damage index vector for varying RMS of noise: 1% (a), 5% (b), 10% (c), 20% (d).

3.5.2. Partially excited system: effects of measurement noise

Let's now consider the 8-DOF shear-type system excited by just one Gaussian white noise input force applied at the DOF 1. Before proceeding with the calculation of the damage index, let's run a preliminary spectral analysis of the response of the system. In vibration based methods, the basic requirement for a successful spectral analysis is that the recorded structural response be rich of modal information so to clearly highlight the contributions of the single modes. For the system under investigation, Figure 3.15 shows the spectral magnitude of the structural response recorded at the 8 DOFs, plotted in a "waterfall plot", for the case of the input force applied at the DOF 1 (Figure 3.15 (a)) and for the case of the force applied at the DOF 8 (Figure 3.15 (b)) before applying measurement noise with RMS of 10%.

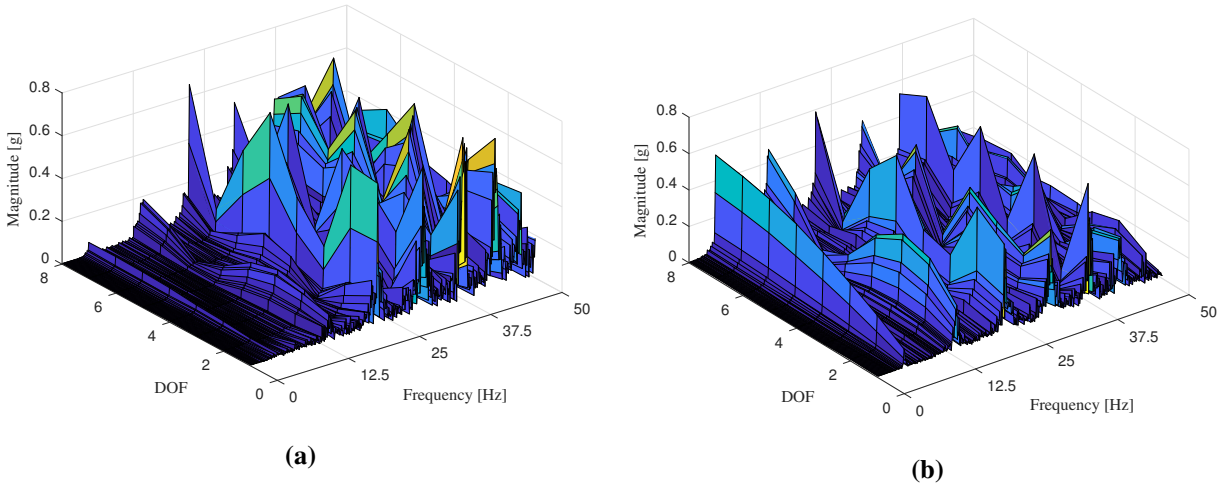


Fig. 3.15. Spectral Magnitude over DOFs. Excitation at DOF 1 (a) and DOF 8 (b).

By comparing those two figures, it is apparent the difference in the contributions of the first mode to the spectrum at various DOFs. For example, the magnitude of the power spectrum of the response in correspondence of the first mode increases from a value of 0.055 g (Figure 3.15 (a)) to 0.541 g (Figure 3.15 (b)). It is reasonable to think that, in the case described in Figure 3.15 (a), the contribution from the first mode can be easily covered by other disturbances as effects of the filtering technique or the excitation of other modes. Thus, for the case of the input force applied at the DOF 1, extracting the information from the first mode may lead to some numerical difficulties. On the other hand, one can observe that there is an important energy contribution of the second mode at 13.77 Hz in both cases.

In order to assess the difficulties arising from modes with low energy contributions, let us consider the case of the input force applied at the DOF 1 (Figure 3.5 (a)). The damage scenario is simulated by decreasing the stiffness between the DOFs 6 and 7 by 20% of its original (undamaged) value. Such a damage scenario corresponds to a jump discontinuity of the damage index between those two DOFs. Here the proposed approach has been applied twice, first considering the contribution of the first mode ($r' = 1$) and then the contribution of the second mode ($r' = 2$) so to evaluate and compare the impact of measurement noise with RMS 10% on both of them. Considering the contribution of the first mode, Figure 3.16 (a) and (b) show the resulting damage index vector using as reference points DOFs 3 ($j = 3$) and 4 ($j = 4$).

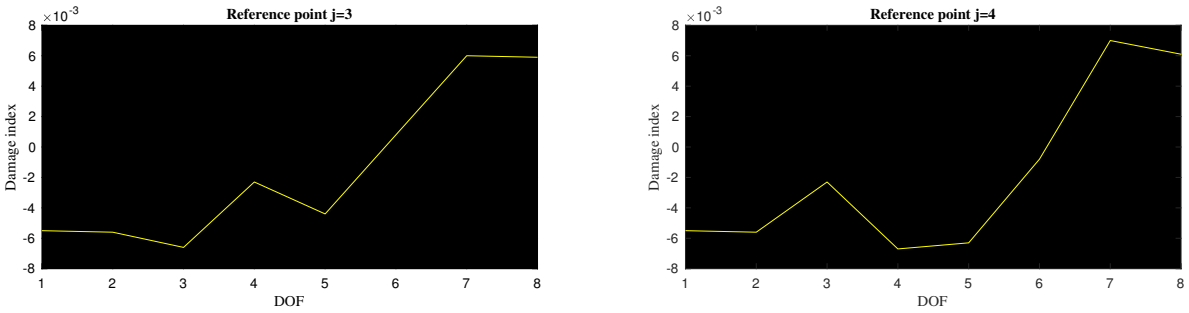


Fig. 3.16. Damage index vector for $r' = 1$: reference point $j = 3$ (a), reference point $j = 4$ (b).

It is clear from these plots that the jump discontinuity does not appear only between the expected DOFs (6 and 7), but also between DOFs 5 and 6 and between 3 and 4, so the two sets of results (for $j = 3$ (a) and $j = 4$ (b)) are not consistent with each other, even though the two IPVs are highly correlated. Such an inaccuracy of the results is due to the fact that the first mode, in the case of the input force applied on the DOF 1, is weakly excited and is strongly affected by noise. Instead, when the proposed approach is applied on the information obtained from the second mode ($r' = 2$, Figure 3.17 (a) and (b)), not only the plots are consistent, but also the predicted jump discontinuity between DOFs 6 and 7 is clearly defined showing great resistance to measurement noise.

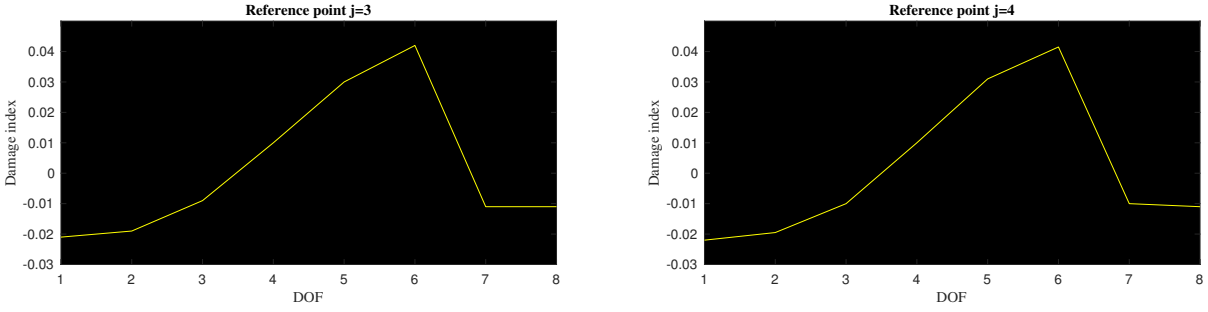
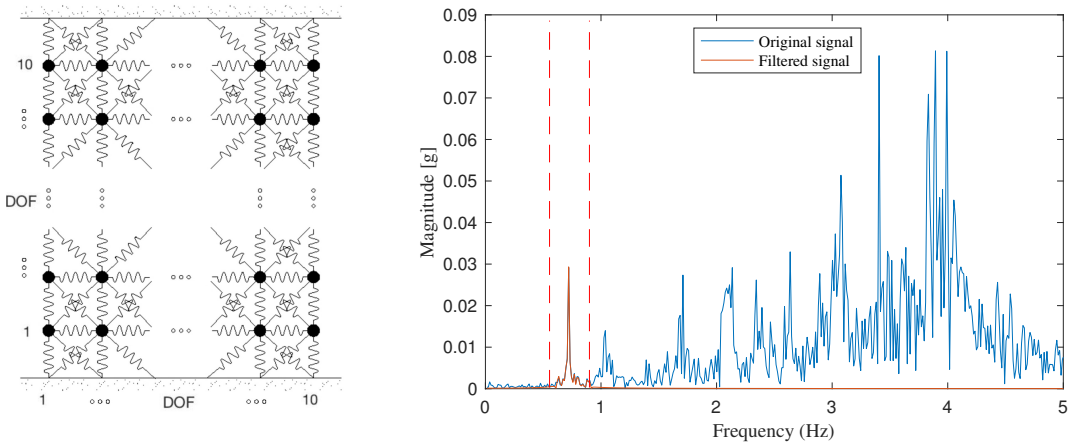


Fig. 3.17. Damage index vector for $r' = 2$: reference point $j = 3$ (a), reference point $j = 4$ (b).

3.6. Numerical simulation: 100-DOF model

This section extends the investigation about the applicability of the proposed IPV-based damage index vector to a more complex system, (e.g. a plate) represented by a two-dimensional frame (Figure 3.18 (a)). The structure is a 2-D square grid of 10×10 lumped masses of 1 kg each connected by spring elements placed horizontally, vertically and diagonally, each one having stiffness of 1000 N/m, 900 N/m and 800 N/m respectively. The modal damping has been set to

$\xi = 1\%$ for all vibration modes. The structure is doubly-fixed at the top and at the bottom and a set of excitation forces acts perpendicular to the plane of structure. Using the assumption of zero-order-hold (ZOH) with a time sampling of 0.01 seconds, these forces are represented by zero-mean Gaussian signals (uncorrelated to each other) with standard deviation of $\sigma = 1$ N and a length of 100 seconds.

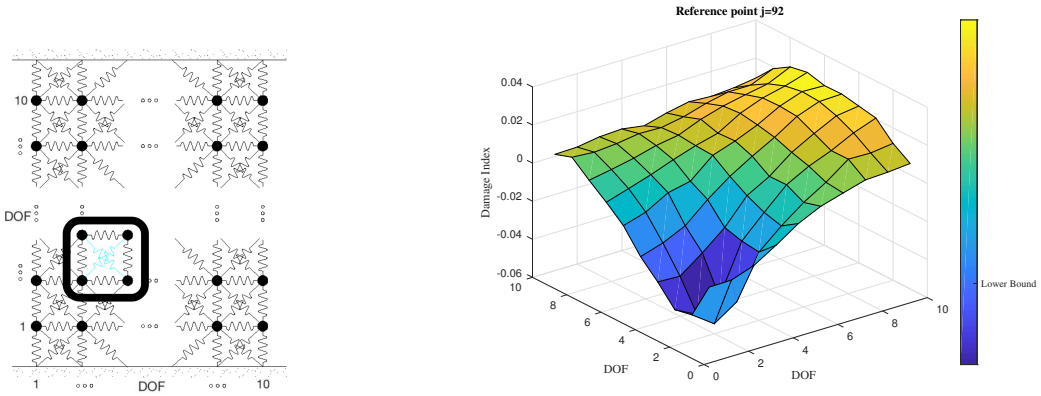


(a) **(b)**

Fig. 3.18. 100 DOF mock up (a), Spectrum (b)

Figure 3.18 (b) shows the spectral magnitude of the response of the system at DOF 50, i.e. 5^{th} row from the bottom, 10^{th} column from left of the model shown in Figure 3.18 (a). The spectral analysis suggests that the first mode contribution is the most suitable to be analyzed through the proposed IPV-based approach since it appears well isolated by the other structural modes (0.72 Hz).

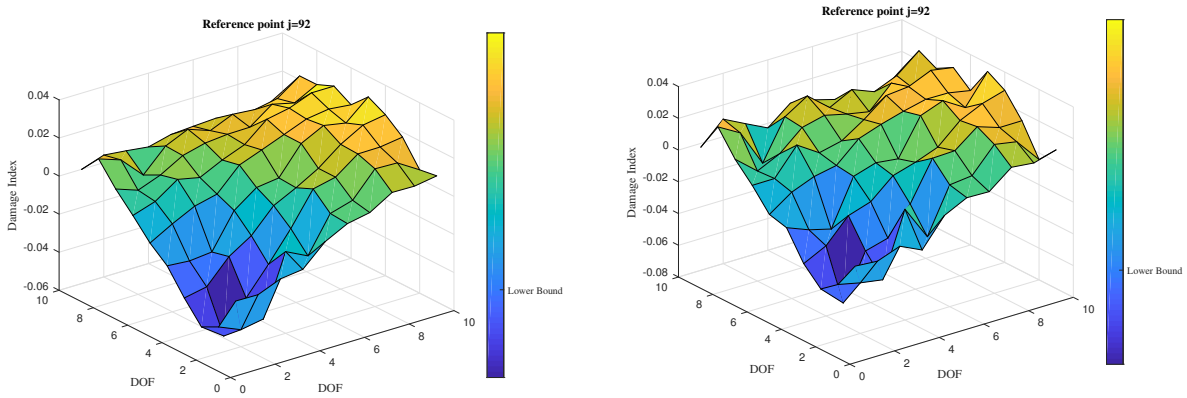
The damage is simulated through a drop in stiffness of 25% for two of the diagonal springs, as shown in Figure 3.19 (a). After checking that the damage index vector for each j^{th} reference point of the 100-DOFs is consistent with the others, Figure 3.19 (b) shows the value of the computed damage index vector for a randomly picked reference point $j = 92$, i.e. 10^{th} row from the bottom, 2^{nd} column from left, far from the damage location. It is clear that the damage location can be identified by just looking at the plot of the damage index vector $\{D_{IPV,92,1}\}$. The threshold for the damage has been set accordingly to the mean and the standard deviation of the elements of the damage index vector so that, for a value of the standard deviation multiplier $\beta_c = 1.8$, the 'Upper Bound' is set at 0.0334 and the 'Lower Bound' at -0.0356. The latter is reported in the colorbar of Figure 3.19 (b).



(a) **(b)**

Fig. 3.19. Simulated damage spot (a), damage index vector, 2-D representation (b)

Increasing the level of measurement noise in the response output signals to 5% RMS and 10% RMS does not prevent the proposed IPV-based approach to find the damage location, as seen in Figure 3.20. In this case, the corresponding 'Upper Bound' and 'Lower Bound' are respectively 0.0347 and -0.0375, for the case in Figure 3.20 (a), and 0.0362 and -0.0387, for the case in Figure 3.20 (b).



(a) **(b)**

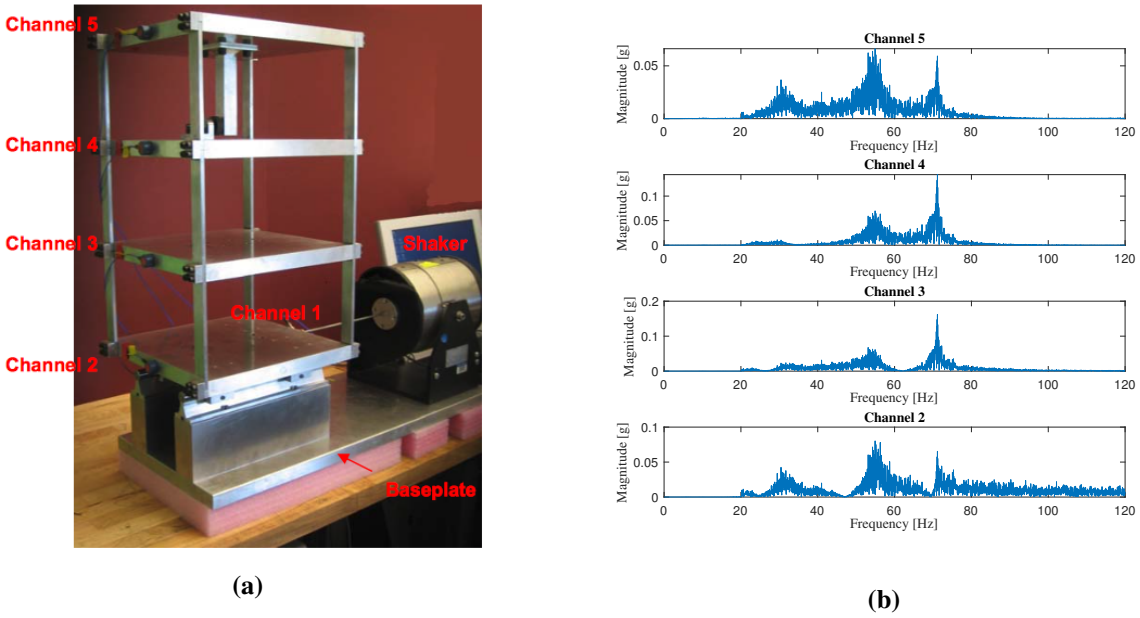
Fig. 3.20. Damage index vector, 2-D representation: 5% RMS (a), 10% RMS (b)

These numerical tests confirm the effectiveness of the proposed IPV-based methodology for damage identification and localization even in the case of more complicated structural models.

3.7. Experimental test: LANL 3-DOF shear-type

The proposed approach has also been tested on experimental test results obtained from a 3-DOF shear-type system shown in Figure 3.21 (a). Test data have been provided by the Engineering

Institute (EI) at Los Alamos National Laboratory (LANL) [66–68]. The system consists of four aluminum columns ($17.7 \times 2.5 \times 0.6$ cm) connected at the top and bottom to aluminum plates ($30.5 \times 30.5 \times 2.5$ cm) [69], forming a structure consisting of 3 floors and a sliding base. The excitation is provided by an electromagnetic shaker that acts at the center line of the base floor of the structure. Both the structure and the shaker are fixed on a base plate ($76.2 \times 30.5 \times 2.5$ cm). Four accelerometers with a nominal sensitivity of 1000 mV/g are attached at the center of the side of each floor at the opposite side from shaker to measure the response of each plate. The random excitation applied at the sliding base is band limited in the range of 20-150 Hz to avoid rigid body modes of the structure. Even if the structure was initially supposed to behave linearly, some non-linear effects due to the sliding rails have been noted [83].



(a) **(b)**

Fig. 3.21. LANL 3-DOF shear-type (a), channels 2-5 FFT Magnitude (b)

Figure 3.21 (b) shows the magnitude of the output response spectrum computed through Fast Fourier Transform (FFT) for each floor in undamaged conditions. From previous studies [68], the natural frequency of the first mode has been determined to be 30.7 Hz. However, the information obtained through the FFT shows that such a natural frequency is not well excited at the 2^{nd} and 3^{rd} floors by the selected input.

The LANL database supplies data about force and accelerations recorded for three different structural conditions (different damage scenarios), other than the original (baseline or healthy) condition. The three damage conditions have been imposed through stiffness reduction of the columns connecting the floors. The damage scenarios considered in this dataset are the

following:

- 50% stiffness reduction between floors 1-2
- 50% stiffness reduction between floors 2-3
- 50% stiffness reduction between floors 3-4

Since the sliding plate (floor 1) can move along the sliding rails when subjected to the shaker action, such a motion will be considered as an additional DOF (DOF 1) so that the model can be analyzed as a 4-DOF system.

3.7.1. 50% stiffness reduction between floors 1-2 and 2-3

As previously done, the first step is the spectral analysis of the output response in order to evaluate the contribution of the various modes for both the damaged and undamaged configurations. The magnitude of the spectral response of the system in the undamaged and damaged configuration for damage case 1 is shown in Figure 3.22 (b) where the spectra of the structural accelerations at each DOF are superimposed. It appears that, after the occurence of damage, a huge drop of the stiffness between floors 1-2 induces the first natural frequency to shift of almost 2 Hz. For such a reason, a band pass filter with cut-off frequencies at 29 and 33 Hz has been adopted in undamaged conditions whereas in damaged configuration the cut-off frequencies have been changed to 27 and 31 Hz.

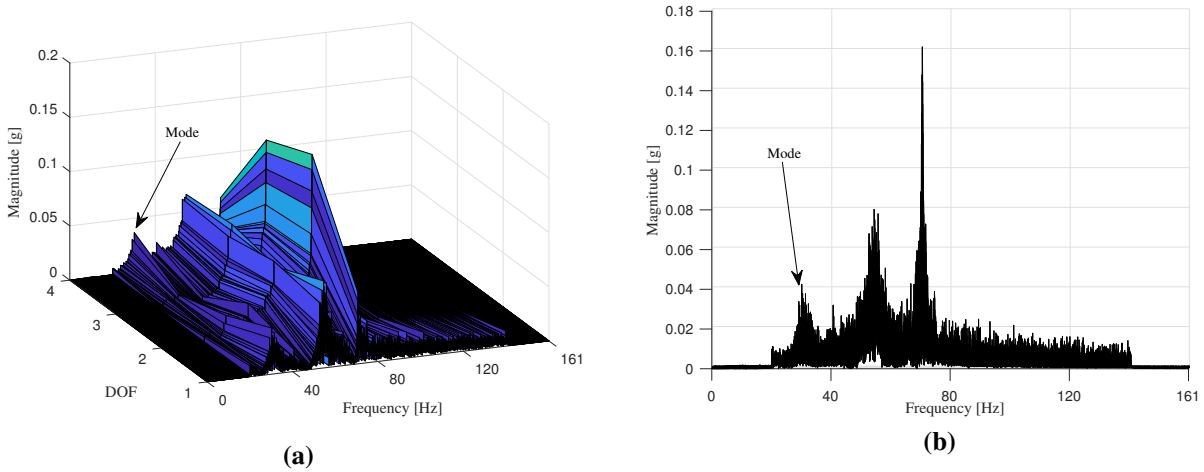


Fig. 3.22. Undamaged structure: Magnitude of the FRFs in a waterfall plot

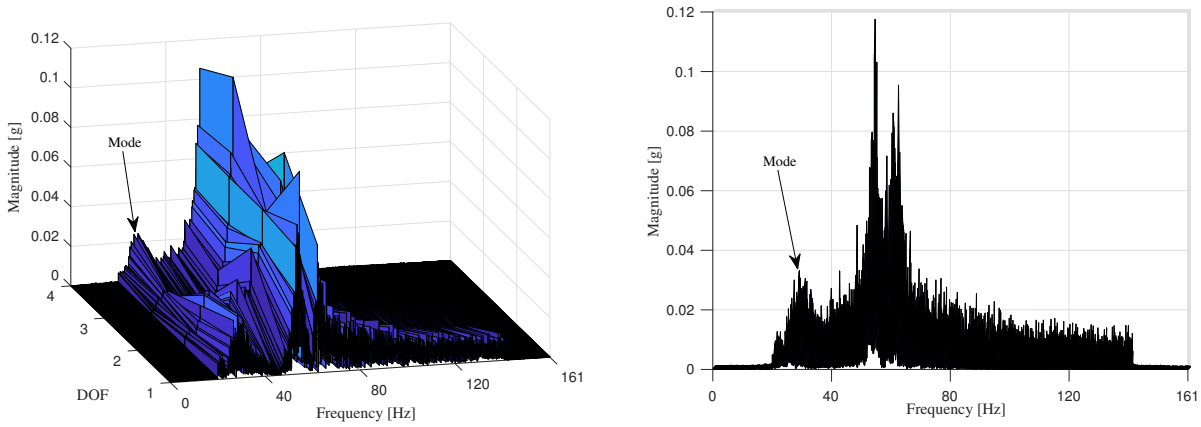


Fig. 3.23. Damaged structure: Magnitude of the FRF, waterfall plot

Figure 3.24 shows the plots of the calculated damage index vector as function of the DOFs for different j^{th} reference points ($j = 1, \ldots, 4$). As seen in the numerical example, when moving from reference point $j = 2$ to $j = 3$, the values of the damage index change sign because of the sign change between the two corresponding modal components in Equation (3.27). This indicates that the mode considered in the damage assessment analysis is not the first mode: this would have prevented the application of the original formulations of the IPV based approach (Wang et al. [70]). However, this obstacle has been overcome with the proposed methodology.

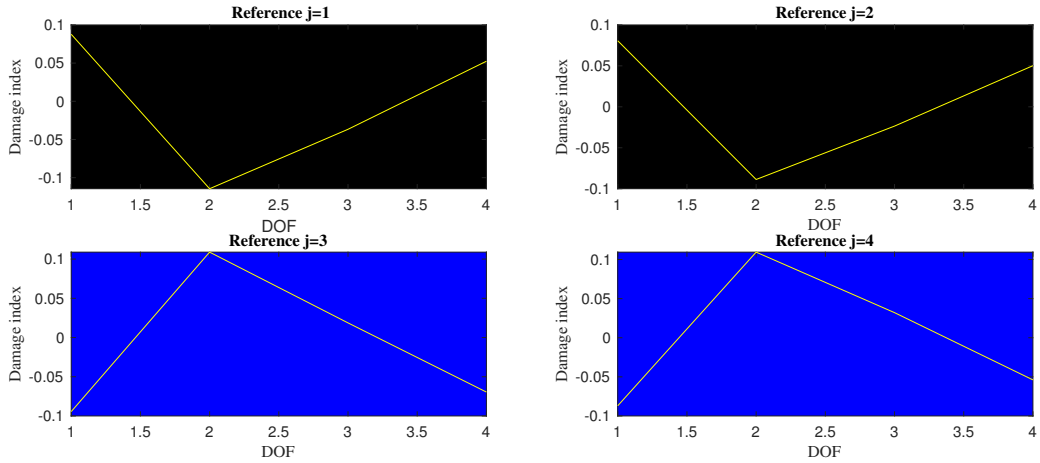


Fig. 3.24. Damage index vector over different reference points.

Usually the threshold for the damage localization relies on calculations of the mean and standard deviation of the elements of the local damage index vector: however, monitoring only 4 positions, the damage index vector contains 4 values only (i.e. Figure 3.24) and its first derivative (representing the local damage index vector) only 3 (i.e. Figure 3.25). Hence, finding

outliers among just 3 points can be difficult when using only mean and standard deviation. However, in this case, having only 3 points does not seem to be a limiting factor: in fact, even with just 3 points, the algorithm is capable to locate the damaged columns as shown in Figure 3.25 (a) for case 1 (damage between DOF 1 and DOF 2) and in Figure 3.25 (b) for case 2 (damage between DOF 2 and DOF 3).

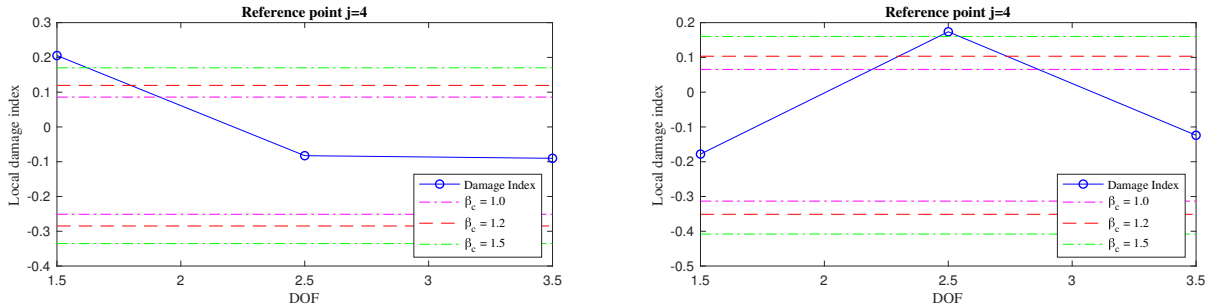


Fig. 3.25. Local damage index vector for case 1 (a) and for case 2 (b). Reference point $j = 4$.

3.7.2. 50% stiffness reduction between floors 3-4

In this case, even though the analysis of this damage scenario is conducted along the same line as the previous ones, something different happens. As shown in Figure 3.26, where the damage index vectors for $j = 1,2,3,4$ are plotted, an interesting behaviour is shown for the case of reference point $j = 3$ with the corresponding damage index vector showing an inconsistent pattern with respect to the other vectors.

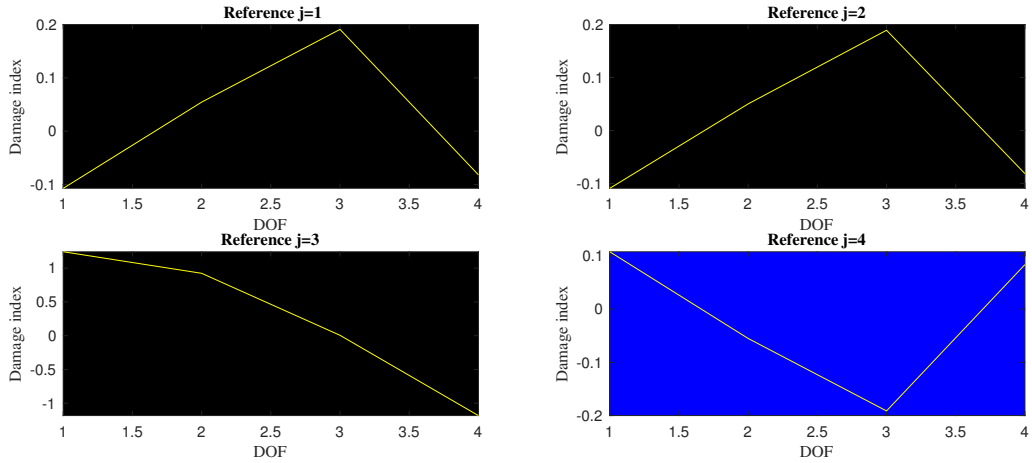
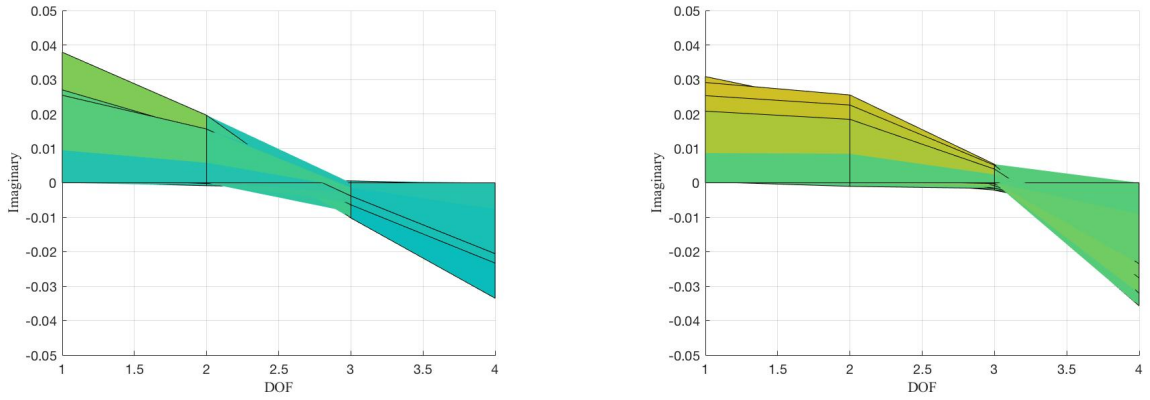


Fig. 3.26. Damage index vector over different reference points.

As discussed in [84], the imaginary part of the Frequency Response Function (FRF) provides information about the shape of a particular mode. In Figure 3.27, the imaginary part of

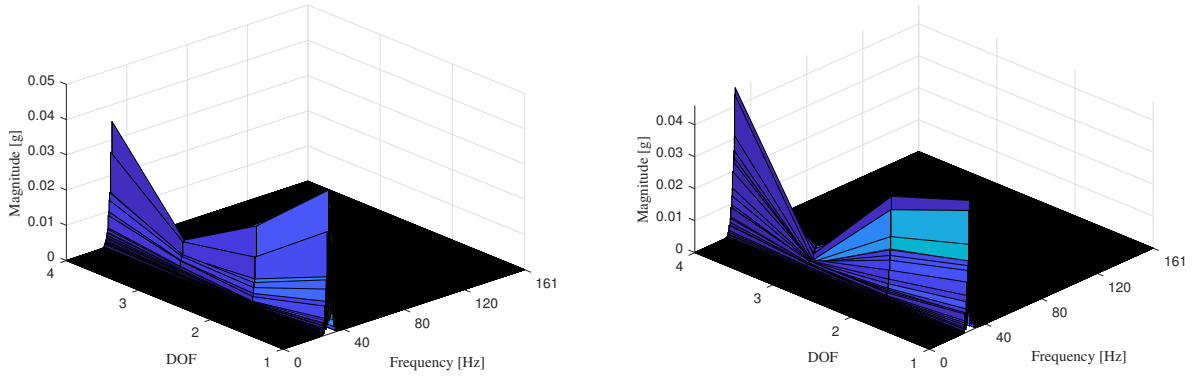
the response spectra in the frequency range in which the signal is filtered is plotted over the DOFs for the undamaged (a) and damaged (b) conditions. It can be noted that passing from the former (undamaged) to the latter (damaged), the imaginary part for the FRF for DOF 3 changes its sign (from negative -undamaged- to positive -damaged-) violating the theoretical assumption allowing to pass from Equation (3.26) to Equation (3.27). As a consequence, the reference point $j = 3$ does not constitute a reliable reference point. All the other DOFs ($j = 1, 2, 4$) are valid reference points, leading to the correct solution.



(a) **(b)**

Fig. 3.27. FRF, imaginary part. Undamaged configuration (a), damaged configuration (b).

Furthermore, looking at the magnitude of the spectrum of the filtered acceleration response time histories in undamaged condition (Figure 3.28 (a)) the analyzed mode is excited at every DOF. Contrarily, in Figure 3.28 (b) the spectral analysis of the system in damaged conditions shows a really low energy content at DOF 3 leading to the conclusion that DOF 3 can be assumed to be a node for the damaged system.



(a) **(b)**

Fig. 3.28. Waterfall plot. Undamaged configuration (a), damaged configuration (b).

Using the damage index vectors in Figure 3.27, it is possible to compute the corresponding local damage index vector $\{L_{IPV,j,1}\}$ and successfully locate the damaged area, Figure 3.29 (a).

It is interesting to observe that, for this particular study case, the presence of a rigid body mode does not allow to use the mode related to the first natural frequency of the system in the calculation of the IPVs. Since the first mode has a very low natural frequency, its contribution is mixed with that of the rigid body mode and a low pass filter (0-20 Hz) applied to the input excitation helps removing such a contribution in the original acceleration response time histories. This would impair the use of the original formulation of the IPV approach presented by Wang et al. [70]; instead, it does not represent a problem for the current formulation because of its ability to handle higher modes. In addition, the next mode, the one with the lowest natural frequency in the spectrum, presents a structural node at the third floor for the structure in damaged condition (damage between DOFs 3 and 4) and so the arbitrary selection of the reference point might lead to some numerical inaccuracies (Figure 3.26, Reference $j = 3$). It is then recommended, when using the proposed methodology, to compute the damage index vector for multiple reference points in order to verify their compatibility and correctly assess the presence of local damage.

Finally, as a further validation of the effectiveness of the proposed method in assessing the damage location, the case of a unit pulse excitation is analyzed. Considering the acceleration response time histories for case 3 (damage between floors 3-4), the unit pulse response can be obtained by considering the Markov parameters extracted through an input-output identification algorithm, e.g. Observer Kalman Identification [83, 85]. By using the system's Markov parameters sequences as unit pulse responses into the proposed IPV-based methodology, it is possible to obtain the corresponding local damage index vector. A comparison between the local damage index vectors obtained using a Gaussian white noise excitation (a) and a unit pulse response (b) is shown in Figure 3.29. The results are consistent and a successful damage localization is achieved.

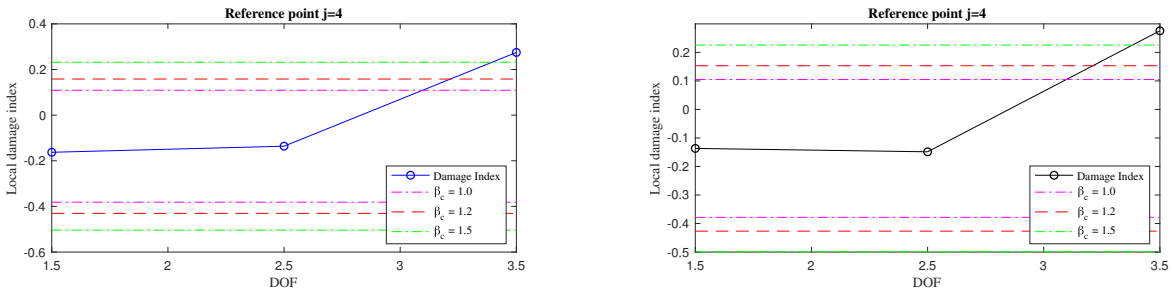


Fig. 3.29. Local damage index vector for white noise (a) and unit pulse (b) excitation.

3.8. Conclusions

In this chapter, a new formulation of a IPV based approach to damage detection is proposed. Differently from the original IPV formulation, this methodology relies on information extracted only from the dynamic response of the structure, without knowledge of the input excitation, e.g. in an output-only context. With respect to the original input-output formulation, there is no need to design an input excitation that is properly tailored to the structural characteristics, reducing the number of necessary tests. In addition, the proposed methodology allows for the analysis of higher modal contributions, eliminating the constrain to account only for the first mode as in the original formulation. The only requirement is the design of a band-pass filter needed to extract the energy contribution of a specific mode in the structural response. The validity of the proposed output-only IPV based theory has been proven for unit impulse and white noise excitations. Numerical simulations on small and large system models as well as experimental data analyses have confirmed the effectiveness of the methodology.

From an engineering point of view, the proposed methodology has many advantages with respect to both the original input-output formulation and the conventional system identification damage assessment strategies. Compared with the original IPV based formulation, the proposed approach requires fewer tests since it does not need a preliminary test to properly design a filter to extract the first mode contribution from the dynamic response. Furthermore, the analysis of structural modes other than the first one extends the applicability of the methodology to more general laboratory test investigations where it is difficult to excite and correctly extract low frequency modal contributions. Another major advantage of the proposed methodology is its ability to handle datasets from large sensor networks: when using system identification methods for structural damage assessment, large datasets might lead to huge computation efforts that result in inaccurate estimations of the damage and its locations. Contrarily, the pro-

posed approach can easily handle large dataset since the estimation of the cross correlations between filtered responses of the system requires much less computation than the estimation of the covariance matrices. Hence, because of its capabilities to handle large sensor networks, the resolution of the damage assessment and localization increases.

Chapter 4

4. A Bayesian clustering approach as support for the Inner Product Vector in structural damage assessment

4.1. Introduction

The purpose of this chapter is the development of a damage assessment algorithm able to assess the presence of a potential structural damage and to localize the damaged area(s). Following a pattern recognition approach, the first step consists in the definition of a valuable damage sensitive feature. Modal parameters (frequency, damping ratios, modes) are usually used as damage sensitive features. Generally, damping ratios are difficult to estimate and their values are not numerically accurate. On the other hand, natural frequencies can usually be estimated with an acceptable level of resolution as well as structural modes. Many system identification algorithms (like SSI, O^3KID, etc.) available in literature are used for this purpose [81, 82, 86]. Anyway, most of the times these methods need to be supported by the so called stabilization diagrams to extract correctly the modal parameters.

Both natural frequencies and modes provide information about the presence of a structural damage. Hence, when a structure presents a structural damage, an analysis purely based on the natural frequencies can provide information about the damage location [78], but the implementation of a mapping method based on laboratory tests and/or numerical simulations on a Finite Element Method (FEM) model is required [87, 88]. With regards to structural modes, it is well known that a local damage causes a local abrupt change in the structural modes [76, 77, 80].

The IPV [70–72, 75, 89] is a damage sensitive feature extracted by computing the cross-correlation of signals representing dynamic response time histories. Under certain circumstances and specific pre-processing procedures, the IPV can be proven to be strictly correlated to a specific structural mode [90].

A Bayesian clustering method is discussed in this chapter and implemented in order to automatize and optimize the extraction of the damage sensitive feature (IPV). The goal is to extract the IPV so that the information embedded in this feature are purely representative of a specific structural mode and not affected by any effect due to changes in natural frequencies and damping ratios and by any structural mode other than the one of interest.

This chapter presents the implementation of a strategy for the correct extraction of the IPV

relying on Bayesian Variational-Inference (VI) [91] for clustering purpose and an Artificial Neural Network (ANN) as classifier. The Bayesian method allows to collect the most likely parameters which automatically set the frequency range to consider for the extraction of a potentially valid damage sensitive feature vector (IPV). The number of parameters taken into account is directly proportional to the number of potential damage sensitive features. These parameters are extracted considering the training dataset first and the test dataset secondly. The implementation of the classifier allows to obtain a pairwise coupling of similar potential damage sensitive feature vectors (IPV) which are strictly correlated, even though some elements different due to the local damage. Among those potential damage sensitive feature vectors, the ones related to the lowest natural frequencies are considered. Based on those features, a local damage index is computed to assess the presence of the damage and identify the damaged area(s).

The validity of the method for the extraction of the IPV and the effectiveness of the presented damage assessment algorithm is proven through mathematical simulations first, and secondly by testing a real 3-DOF shear-typer structure, a laboratory model at Los Alamos National Laboratory (LANL) [67–69, 79].

4.2. Problem statement

The proposed damage assessment algorithm relying on the IPV aims to assess the presence of a damage and to localize the damaged area(s) by isolating and evaluating the energy contribution of the structural modes to the dynamic response of a system. The application of Finite Impulse Response (FIR) filters to the dynamic response time histories is the key to extract signals containing exclusively the energy contribution of a specific mode, reducing as much as possible the effects of any other mode on the dynamic response. Basically, the characteristics of those filters depend on the modal parameters which are supposed to remain the same as long as the healthy state of the structure persists. The occurrence of a local damage causes changes in the modal parameters (frequencies, damping factors and modes) making the most likely parameters characterizing the filters change. Because of the damage, not only the natural frequencies related to existing modes change, but new modes of different nature (bending, torsion) may appear and contribute to the dynamic response influencing the Frequency Response Function (FRF) of the system. By comparing the dynamics of a structure in undamaged state with those of the same structure in damaged state, we can assume that a specific mode just changes because of the

damage.

The extraction of the IPV needs a pre-processing phase that is carried out by accomplishing two tasks: 1) a strategy in order to set the FIR filters to isolate and evaluate the energy contribution to the response due to different modes has to be found; 2) similar energy contributions to the dynamic response when the system is in undamaged and damaged state have to be identified.

4.2.1. Structural damage assessment algorithm

The IPV is a damage sensitive feature that is extracted by operating on the dynamic response time histories (i.e. displacements, velocities, accelerations). Despite its computation fully relies on manipulations of signals defined in time domain, the filtering procedure (part of the pre-processing phase) is carried out by considering exclusively the responses in the frequency domain. The energy contribution of each structural mode to the dynamic signals is observable in their power spectra. Each of the structural modes can be associated to a certain natural frequency that occupies a specific position in the frequency domain explored by the power spectrum. The FIR filters are set properly and effectively if they are able to filter a specific portion of the frequency domain so to catch most of the energy contribution due to a specific mode and reduce the effects of the other modes to the dynamic response. The damage assessment algorithm here proposed can be summarized in five key points:

1. Cluster different areas of the frequency domain based on the power spectrum of the training acceleration time histories;
2. Define a characteristic feature for each of these clustered areas of the frequency domain representing a specific class;
3. Cluster different areas of the frequency domain based on the power spectrum of the test acceleration time histories;
4. Extract the characteristic features according to 2. and classify the clusters in order to couple training and testing clusters;
5. Apply the IPV based on FIR filters properly set according to the selected pairwise coupled training and testing clusters.

Basically, the five steps we should go through in order to implement the damage assessment algorithm require the application of: 1) a clustering algorithm; 2) a classifier; 3) the use of the cross-correlation to identify the IPV. As mentioned, the clustering algorithms here presented

are based on a Bayesian approach. Bayesian methods have been shown to be particularly appealing for clustering purpose, especially in those cases in which the exact number of clusters is not specified a priori. Among the Bayesian clustering algorithms, two in particular are the most commonly used: the Expectation-Maximization (EM) algorithm and the Variational Inference (VI) algorithm. The EM algorithm is a technique aiming to identify statistical distributions by setting the parameters defining them in order to maximize the natural logarithm of the likelihood (the model). Differently, the VI algorithm is a Maximum A Posteriori (MAP) technique based on a random initialization of the parameters of an auxiliary distribution that converges to the posterior by updating the parameters over some iterations [92, 93]. The posterior distribution is so called to emphasize the fact that it represents an estimate of the parameters distribution after some observations have been collected.

The classification task is performed through an Artificial Neural Network (ANN) which is trained with the scope of detecting compatible clusters and create their pairwise coupling. Thus, a specific cluster in undamaged configuration can be coupled to one new cluster defined in unknown (potentially damaged) configuration. The correctness of the coupling is verified by the Modal Assurance Criterion (MAC). By considering the parameters relative to a certain couple of clusters, the FIR filters (one for each of the two structural conditions) can be designed, and the IPV method applied.

In the next sections we present the Variational Inference algorithm implemented and its applicability to the spectral content to the dynamic signals considered.

4.3. Variational Inference algorithm

Let's assume we are able to collect n observations $\{x_1, x_2, \ldots, x_n\}$ of a random variable into a vector $\boldsymbol{X} \in \mathbb{R}^{1\times n}$ which can be assumed to be sampled from a probability distribution defined by m hidden parameters $\{\omega_1, \omega_2, \ldots, \omega_m\}$ in the vector $\boldsymbol{\omega} \in \mathbb{R}^{1\times m}$. Generally, the objective of Variational Inference algorithms is to estimate the posterior probability distribution $p(\boldsymbol{\omega}|\boldsymbol{X})$ when the prior probability distribution $p(\boldsymbol{\omega})$ is not a conjugate prior for the model function, i.e. likelihood distribution $p(\boldsymbol{X}|\boldsymbol{\omega})$. In this framework, we present a problem set-up in which the model and the prior can be written as follows:

$$\boldsymbol{X} \sim p(\boldsymbol{X}|\boldsymbol{\omega}), \quad \boldsymbol{\omega} \sim p(\boldsymbol{\omega}) \tag{4.1}$$

which indicates that the n observations x_i are sampled from a probability distribution $p(\boldsymbol{X}|\boldsymbol{\omega})$ and the m hidden parameters from the distribution $p(\boldsymbol{\omega})$. The posterior probability distribution function (pdf) can be computed through Bayes rule:

$$p(\boldsymbol{\omega}|\boldsymbol{X}) = \frac{p(\boldsymbol{X}|\boldsymbol{\omega})p(\boldsymbol{\omega})}{p(\boldsymbol{X})}. \tag{4.2}$$

If the prior pdf is not a conjugate prior for the likelihood, the numerator in Equation (4.2) can be represented by a distribution that could be difficult to interpret. Consequently, being the denominator obtained from the integration of the numerator over all ω_i

$$p(\boldsymbol{X}) = \int p(\boldsymbol{X}, \boldsymbol{\omega}) d\boldsymbol{\omega} = \int p(\boldsymbol{X}|\boldsymbol{\omega})p(\boldsymbol{\omega}) d\boldsymbol{\omega}. \tag{4.3}$$

the denominator becomes intractable. Thus, the posterior distribution cannot be computed by applying the Bayes rule, and an alternative way has to be found. The mathematic formulation of the Bayes rule can be rewritten as function of the posterior distribution:

$$p(\boldsymbol{X}, \boldsymbol{\omega}) = p(\boldsymbol{\omega}|\boldsymbol{X})p(\boldsymbol{X}) \tag{4.4}$$

where the left hand side represents the joint probability distribution of $\boldsymbol{X}$ and $\boldsymbol{\omega}$. By applying the natural logarithm to both hand sides of Equation (4.4) leads to:

$$ln(p(\boldsymbol{X}, \boldsymbol{\omega})) = ln(p(\boldsymbol{\omega}|\boldsymbol{X})) + ln(p(\boldsymbol{X})). \tag{4.5}$$

which can be rewritten as:

$$ln(p(\boldsymbol{X})) = ln(p(\boldsymbol{X}, \boldsymbol{\omega})) + ln\left(\frac{1}{p(\boldsymbol{\omega}|\boldsymbol{X})}\right) \tag{4.6}$$

We can now introduce an auxiliary distribution $q(\boldsymbol{\omega})$. By adding and subtracting $\ln(q(\boldsymbol{\omega}))$ on the right hand side leads to:

$$ln(p(\boldsymbol{X})) = ln(p(\boldsymbol{X}, \boldsymbol{\omega})) - ln(q(\boldsymbol{\omega})) + ln\left(\frac{1}{p(\boldsymbol{\omega}|\boldsymbol{X})}\right) + ln(q(\boldsymbol{\omega})) \tag{4.7}$$

which can be rewritten

$$ln(p(\boldsymbol{X})) = ln\left(\frac{q(\boldsymbol{\omega})}{p(\boldsymbol{\omega}|\boldsymbol{X})}\right) + ln\left(\frac{p(\boldsymbol{X},\boldsymbol{\omega})}{q(\boldsymbol{\omega})}\right). \tag{4.8}$$

It is now clear how the distribution $q(\boldsymbol{\omega})$ links $p(\boldsymbol{X},\boldsymbol{\omega})$ and $p(\boldsymbol{\omega},\boldsymbol{X})$. We can multiply Equation (4.8) by the distribution $q(\boldsymbol{\omega})$ and integrate out $\boldsymbol{\omega}$ on its domain and obtain:

$$\int q(\boldsymbol{\omega})ln(p(\boldsymbol{X}))d\boldsymbol{\omega} = \int q(\boldsymbol{\omega})ln\left(\frac{q(\boldsymbol{\omega})}{p(\boldsymbol{\omega}|\boldsymbol{X})}\right)d\boldsymbol{\omega} + \int q(\boldsymbol{\omega})ln\left(\frac{p(\boldsymbol{X},\boldsymbol{\omega})}{q(\boldsymbol{\omega})}\right)d\boldsymbol{\omega} \tag{4.9}$$

Since the distribution on the left hand side $p(\boldsymbol{X})$ is independent from $\boldsymbol{\omega}$, its logarithm can be pulled out of the integral. Since $q(\boldsymbol{\omega})$ is a probability distribution, its integral $\int q(\boldsymbol{\omega})d\boldsymbol{\omega} = 1$ by definition and so we obtain:

$$ln(p(\boldsymbol{X})) = \int q(\boldsymbol{\omega})ln\left(\frac{q(\boldsymbol{\omega})}{p(\boldsymbol{\omega}|X)}\right)d\boldsymbol{\omega} + \int q(\boldsymbol{\omega})ln\left(\frac{p(\boldsymbol{X},\boldsymbol{\omega})}{q(\boldsymbol{\omega})}\right)d\boldsymbol{\omega} \tag{4.10}$$

Let's now analyze the three terms in Equation (4.10).

- The term $ln(p(\boldsymbol{X}))$ is constant over the parameters in $\boldsymbol{\omega}$.
- The term $\int q(\boldsymbol{\omega})ln\left(\frac{q(\boldsymbol{\omega})}{p(\boldsymbol{\omega}|\boldsymbol{X})}\right)d\boldsymbol{\omega}$ is the so called *Kullback-Leibler (KL) divergence*, always ≥ 0. As the auxiliary distribution $q(\omega)$ approaches the posterior distribution $p(\boldsymbol{\omega}|\boldsymbol{X})$, their ratio tends to one and the KL divergence approaches zero. Thus, if we want our auxiliary distribution to approximate the posterior distribution we need to seek for the parameters in $\boldsymbol{\omega}$ that minimize the KL divergence. Since we don't have any information about either $q(\boldsymbol{\omega})$ or $p(\boldsymbol{\omega}|\boldsymbol{X})$, the KL divergence is matematically intractable.
- The term $\int q(\boldsymbol{\omega})ln\left(\frac{p(\boldsymbol{X},\boldsymbol{\omega})}{q(\boldsymbol{\omega})}\right)d\boldsymbol{\omega}$ is referred to as the *VI objective function*. This function depends on the joint distribution $p(\boldsymbol{X},\boldsymbol{\omega})$ which is easy to compute, and on $q(\boldsymbol{\omega})$ arbitrarily defined.

Since the left hand side term in Equation (4.10) is constant over $\boldsymbol{\omega}$, the sum of the VI objective function and of the KL divergence must be constant. Hence minimizing the KL divergence means maximizing the VI objective function over the parameters in $\boldsymbol{\omega}$.

The VI objective function $\mathscr{L}$ can be rewritten as:

$$\mathscr{L} = \int q(\boldsymbol{\omega})ln\left(\frac{p(\boldsymbol{X},\boldsymbol{\omega})}{q(\boldsymbol{\omega})}\right)d\boldsymbol{\omega} = \int q(\boldsymbol{\omega})ln\left(p(\boldsymbol{X},\boldsymbol{\omega})\right)d\boldsymbol{\omega} - \int q(\boldsymbol{\omega})ln\left(q(\boldsymbol{\omega})\right)d\boldsymbol{\omega}. \tag{4.11}$$

Before proceeding, additional assumptions referred to as mean-field assumptions [92] have to be considered. The 'mean-field' assumptions were first introduced in Physics and, in this framework, can be stated as:

- the m parameters in the vector $\boldsymbol{\omega}$ can be split in groups;
- a vector $\boldsymbol{\eta} = \{\eta_1, \ldots, \eta_m\}$, whose parameters can be associated each with its corresponding parameter ω_i, is introduced to define the family of the probability distribution $q(\omega_i|\eta_i)$; this classification is done to allow different parameters of $\boldsymbol{\omega}$ to refer to different types of distribution;
- the auxiliary distribution $q(\boldsymbol{\omega})$ can be factorized as $q(\boldsymbol{\omega}) = \prod_{i=1}^{m} q(\omega_i|\eta_i)$ so that the distributions $q(\omega_i|\eta_i)$ for $i = 1, \ldots, m$ are statistically independent.

In order to simplify the calculations, let's now implicitly assume $q(\omega_i) = q(\omega_i|\eta_i)$. These new considerations lead to a new formulation of the VI objective function $\mathscr{L}$:

$$\begin{aligned} \mathscr{L} &= \int \prod_{i=1}^{m} q(\omega_i) ln\left(p(\boldsymbol{X}, \omega_1, \ldots, \omega_m)\right) d\omega_1 \ldots d\omega_m \\ &- \int \prod_{i=1}^{m} q(\omega_i) ln\left(\prod_{i=1}^{m} q(\omega_i)\right) d\omega_1 \ldots d\omega_m \end{aligned} \tag{4.12}$$

The second term on the right hand side of Equation (4.12) can also be rewritten as:

$$\begin{aligned} &\int \prod_{i=1}^{m} (q(\omega_i)) ln\left(\prod_{i=1}^{m} q(\omega_i)\right) d\omega_1 \ldots d\omega_m \\ &= \int \prod_{i=1}^{m} (q(\omega_i)) \sum_{i=1}^{m} ln\left(q(\omega_i)\right) d\omega_1 \ldots d\omega_m \\ &= \int \prod_{i=1}^{m} q(\omega_i) ln\left(q(\omega_1)\right) d\omega_1 \ldots d\omega_m + \ldots + \int \prod_{i=1}^{m} q(\omega_i) ln\left(q(\omega_m)\right) d\omega_1 \ldots d\omega_m \\ &= \int q(\omega_1) ln\left(q(\omega_1)\right) d\omega_1 + \ldots + \int q(\omega_m) ln\left(q(\omega_m)\right) d\omega_m \\ &= \sum_{i=1}^{m} \int q(\omega_i) ln\left(q(\omega_i)\right) d\omega_i \end{aligned} \tag{4.13}$$

By substituting Equation (4.13) into Equation (4.12) leads to:

$$\mathscr{L} = \int \prod_{i=1}^{m} q(\omega_i) ln\left(p(\boldsymbol{X}, \omega_1, \ldots, \omega_m)\right) d\omega_1 \ldots d\omega_m - \sum_{i=1}^{m} \int q(\omega_i) ln\left(q(\omega_i)\right) d\omega_i \tag{4.14}$$

Let's try to analyze the dependance of $\mathscr{L}$ from a specific distribution $q(\omega_i)$. By considering

the first term on the right hand side of Equation (4.14):

$$\begin{aligned} & \int \prod_{i=1}^{m} q(\omega_i) ln\left(p(\boldsymbol{X}, \omega_1, \ldots, \omega_m)\right) d\omega_1 \ldots d\omega_m \\ = & \int q(\omega_i) \left[\int \prod_{l \neq i}^{m} (q(\omega_l)) \, ln\left(p(\boldsymbol{X}, \omega_1, \ldots, \omega_m)\right) d\omega_1 \ldots d\omega_m \right] d\omega_i \end{aligned} \tag{4.15}$$

By recalling the definition of the expectation operator $\mathbb{E}[\mathscr{T}(x)] = \int q(x)\mathscr{T}(x)dx$ for an arbitrary function $\mathscr{T}$ of a generic variable x, Equation (4.15) can be rewritten as:

$$\begin{aligned} & \int \prod_{i=1}^{m} q(\omega_i) ln\left(p(\boldsymbol{X}, \omega_1, \ldots, \omega_m)\right) d\omega_1 \ldots d\omega_m \\ = & \int q(\omega_i) \mathbb{E}_{q_{l \neq i}} \left[ln\left(p(\boldsymbol{X}, \omega_1, \ldots, \omega_m)\right) \right] d\omega_i. \end{aligned} \tag{4.16}$$

The second term on the right hand side of Equation (4.14) can be decomposed into:

$$-\sum_{i=1}^{m} \int q(\omega_i) ln\left(q(\omega_i)\right) d\omega_i = -\int q(\omega_i) ln\left(q(\omega_i)\right) d\omega_i - \sum_{l \neq i}^{m} \int q(\omega_l) ln\left(q(\omega_l)\right) d\omega_l \tag{4.17}$$

so that by substituting Equation (4.16) and Equation (4.17) into Equation (4.14) a new formulation for the objective function $\mathscr{L}$ is obtained:

$$\begin{aligned} \mathscr{L} &= \int q(\omega_i) \mathbb{E}_{q_{l \neq i}} \left[ln\left(p(\boldsymbol{X}, \omega_1, \ldots, \omega_m)\right) \right] d\omega_i - \int q(\omega_i) ln\left(q(\omega_i)\right) d\omega_i - \sum_{l \neq i}^{m} \int q(\omega_l) ln\left(q(\omega_l)\right) d\omega_l \\ &= \int q(\omega_i) ln(e^{\mathbb{E}_{q_{l \neq i}}[ln(p(\boldsymbol{X}, \omega_1, \ldots, \omega_m))]}) d\omega_i - \int q(\omega_i) ln\left(q(\omega_i)\right) d\omega_i - \sum_{l \neq i}^{m} \int q(\omega_l) ln\left(q(\omega_l)\right) d\omega_l \\ &= \int q(\omega_i) ln \frac{e^{\mathbb{E}_{q_{l \neq i}}[ln(p(\boldsymbol{X}, \omega_1, \ldots, \omega_m))]}}{q(\omega_i)} d\omega_i - \sum_{l \neq i}^{m} \int q(\omega_l) ln\left(q(\omega_l)\right) d\omega_l \end{aligned} \tag{4.18}$$

Since we want to maximize $\mathscr{L}$ over a specific parameter ω_i, we realize that the summation of the integrals in Equation (4.18) does not depend on ω_i and so it can be considered constant:

$$\mathscr{L} = \int q(\omega_i) ln \frac{e^{\mathbb{E}_{q_{l \neq i}}[ln(p(\boldsymbol{X}, \omega_1, \ldots, \omega_m))]}}{q(\omega_i)} d\omega_i + \text{const.} \tag{4.19}$$

At this point, let's introduce a new quantity Z, defined as:

$$Z = \int e^{\mathbb{E}_{q_{l \neq i}}[ln(p(\boldsymbol{X}, \omega_1, \ldots, \omega_m))]} d\omega_i \tag{4.20}$$

which is a constant term. By adding and subtracting $ln(Z)$ on the right side of Equation (4.20), the final formulation of $\mathscr{L}$ can be expressed as:

$$\mathscr{L} = \int q(\omega_i) ln \frac{\frac{1}{Z} e^{\mathbb{E}_{q_{l\neq i}}[ln(p(\boldsymbol{X},\omega_1,\ldots,\omega_m))]}}{q(\omega_i)} d\omega_i + \text{const.} \tag{4.21}$$

The term $\frac{1}{Z} e^{\mathbb{E}_{q_{l\neq i}}[ln(p(\boldsymbol{X},\omega_1,\ldots,\omega_m))]}$ can be interpreted as a probability distribution function so that the objective function $\mathscr{L}$ can be seen, following the definitions previously introduced, as a negative KL divergence:

$$\mathscr{L} = \int q(\omega_i) ln \frac{\frac{1}{Z} e^{\mathbb{E}_{q_{l\neq i}}[ln(p(\boldsymbol{X},\omega_1,\ldots,\omega_m))]}}{q(\omega_i)} d\omega_i = -\text{KL}(q(\omega_i) || \frac{1}{Z} e^{\mathbb{E}_{q_{l\neq i}}[ln(p(\boldsymbol{X},\omega_1,\ldots,\omega_m))]}). \tag{4.22}$$

By recalling that KL ≥ 0 then VI objective function $\mathscr{L} \leq 0$. Minimizing the positive KL divergence means maximizing the objective function $\mathscr{L}$. From Equation (4.22) we finally know that we should set:

$$q(\omega_i) = \frac{1}{Z} e^{\mathbb{E}_{q_{l\neq i}}[ln(p(\boldsymbol{X},\omega_1,\ldots,\omega_m))]} \tag{4.23}$$

Equation (4.23) points out a remarkable result setting a mathematical connection between the auxiliary function $q(\omega_i)$ and the joint probability distribution $p(\boldsymbol{X},\omega_1,\ldots,\omega_m)$. Furthermore, it is crucial to observe that by setting the auxiliary distributions $q(\omega_i)$ according to Equation (4.23) we are minimizing the KL divergence without using the gradient descending method. This observation is at the core of the VI method. This process is repeated for every parameter ω_i so to obtain a set of m distributions $q(\omega_i)$. These distributions $q(\omega_i)$ are claimed to be statistically independent (given the parameters characterizing them), respecting the mean field assumptions. Anyway, there is an inference between the parameters of the prior distributions which are updated in an iterative process to minimize the KL divergence on the right hand side of Equation (4.22) leading to an approximation of the posterior distribution, i.e. $q(\boldsymbol{\omega}) = \prod_{i=1}^{m} q(\omega_i)$.

4.4. VI for data clustering

Let's assume a random variable x that has a probability distribution that can be represented by a mixture model with K components. Let's now consider a set of observations of this random variable x, x_i $(i = 1,\ldots,n)$ which can be stored in a vector $\boldsymbol{X} = \{x_1, x_2, \ldots, x_n\} \in \mathbb{R}^{1\times n}$. For

each observation x_i we associate a label c_i ($i = 1,\dots,n$) which indicates the mixture component the observation x_i is referred to. Given a number $K \in \mathbb{N}$ of mixture components, $c_i \in (1,\dots,K)$. Consider now the case of a set of observations $\boldsymbol{X} = \{x_1, x_2, \dots, x_n\}$ where each $x_i \in (1,\dots,N)$ can be assumed to be generated from a mixture of Binomial distributions. Given the mixture component c_i, the observation x_i can be considered sampled by a Binomial distribution as:

$$x_i | c_i \sim \text{Binomial}(N, \theta_{c_i}) \tag{4.24}$$

where θ_{c_i}, $\theta_{c_i} \in (\theta_1,\dots,\theta_K)$, is a characteristic parameter of the distribution related to the c_i mixture component. Let's recall that the model from which the set of x_i is sampled is a mixture of K Binomial distributions, each of them defined by one of the K parameters θ_{c_i}. The mixture component label c_i is sampled from a Discrete (Categorical) prior distribution as follows:

$$c_i \sim \text{Discrete}(\boldsymbol{\pi}) \tag{4.25}$$

whose parameters are the elements of the vector $\boldsymbol{\pi} = \{\pi_1,\dots,\pi_K\}$. The elements of $\boldsymbol{\pi}$ are probabilities whose sum is equal to 1.

According to the VI approach, the objective is to create a prior distribution for every parameter of the distributions in Equation (4.24) and Equation (4.25) and, through inference of the parameters defining the prior distributions, to update them so that the posterior distribution can be approximated. Therefore, we need to define some prior distributions for the vector $\boldsymbol{\pi}$ and for each of the K θ_{c_i}, distributions that, for mathematical convenience, are set respectively equal to a Dirichlet distribution and to Beta distributions. This choice is dictated by the fact that the Dirichlet distribution is a conjugate prior for the Discrete (Categorical) distribution defined by $\boldsymbol{\pi}$ while the Beta distribution is a conjugate prior for the Binomial distributions defined by θ_{c_i}: Recalling that $c_i \in (1,\dots,K)$, $\theta_{c_i} = \{\theta_{c_1},\dots,\theta_{c_n}\}$ and $\theta_{c_i} \in (\theta_1,\dots,\theta_K)$, only K prior distributions, one for each θ_j ($j = 1,...,K$), have to be defined.

$$\boldsymbol{\pi} \sim \text{Dirichlet}(\boldsymbol{\alpha}), \quad \theta_j \sim \text{Beta}(a_j, b_j). \tag{4.26}$$

In order to simplify the notation, we can group all the parameters θ_j in a vector $\boldsymbol{\theta} = \{\theta_1,\dots,\theta_K\}$. At the initial step, the parameters in the prior distributions (Equation (4.26)) are initialized so that the elements of the vector $\boldsymbol{\alpha} = \{\alpha_1,\dots,\alpha_K\}$ are set to be constant: $\alpha_j = \alpha$ for any

$j = 1, \ldots, K$; similarly all the K Beta distributions will have the initial parameters set equal to two constants a and b ($a_j = a$, $b_j = b$ for $j = 1, \ldots, K$). The initialization of the numerical values for the parameters of these priors is supported by empirical investigations: a proper choice is to set $\alpha = 0.1$, $a = 0.5$ and $b = 0.5$.

The analytical representation of likelihood and prior distributions is here briefly presented:

- $p(\boldsymbol{X}|\boldsymbol{\theta}, \forall c_i = j) = \prod_{i=1}^{n} \binom{N}{x_i} \theta_j^{x_i} (1-\theta_j)^{(N-x_i)}$ (Binomial distribution)
- $p(c_i|\boldsymbol{\pi}) = \prod_{j=1}^{K} \pi_j^{\mathbb{1}[c_i=j]}$ (Categorical distribution)
- $p(\boldsymbol{\pi}) = \frac{\Gamma(K\alpha)}{K\Gamma(\alpha)} \prod_{j=1}^{K} \pi_j^{\alpha-1}$ (Dirichlet distribution)
- $p(\theta_j) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta_j^{a-1}(1-\theta_j)^{b-1}$ (Beta distribution)

The term $\mathbb{1}[c_i = j]$ is called indicator and is equal to 1 if $c_i = j$, to 0 otherwise.

We now want to approximate the posterior distribution by taking advantage of the auxiliary distribution defined in the previous section:

$$p(\boldsymbol{\pi}, \boldsymbol{\theta}, \boldsymbol{c}|\boldsymbol{X}) \sim q(\boldsymbol{\pi}, \boldsymbol{\theta}, \boldsymbol{c}) \tag{4.27}$$

where $\boldsymbol{c} = \{c_1, \ldots, c_n\}$. The mean-field assumptions allow us factorize the auxiliary distribution as follows:

$$q(\boldsymbol{\pi}, \boldsymbol{\theta}, \boldsymbol{c}) = q(\boldsymbol{\pi}) \left[\prod_{j=1}^{K} q(\theta_j)\right] \left[\prod_{i=1}^{n} q(c_i)\right]. \tag{4.28}$$

By recalling Equation (4.23), we know that the auxiliary distribution converges to the posterior distribution if each of the distributions in the right hand side of Equation (4.28) is iteratively computed as:

$$q(\omega_i) = \frac{1}{Z} e^{\mathbb{E}_{q_{l\neq i}}[ln(p(X,\omega_1,\ldots,\omega_m))]} \tag{4.29}$$

where ω_i represents the generic i^{th} parameter. For this specific case, all the parameters can be allocated in the vector $\boldsymbol{\omega} = \{\omega_1, \ldots, \omega_{K+n+1}\} = \{\boldsymbol{\pi}, \theta_1, \ldots, \theta_K, c_1, \ldots, c_n\}$. Applying the Bayes rule leads to an expression of the joint probability distribution that can be written as:

$$p(\boldsymbol{X}, \boldsymbol{\pi}, \boldsymbol{\theta}, \boldsymbol{c}) = p(\boldsymbol{X}|\boldsymbol{\pi}, \boldsymbol{\theta}, \boldsymbol{c}) p(\boldsymbol{c}|\boldsymbol{\pi}) p(\boldsymbol{\pi}) p(\boldsymbol{\theta}). \tag{4.30}$$

By taking the natural logarithm to Equation (4.30) and recalling that $p(\boldsymbol{X},\boldsymbol{\pi},\boldsymbol{\theta},\boldsymbol{c}) = p(x_1,\ldots,x_n,\boldsymbol{\pi},\boldsymbol{\theta},\boldsymbol{c})$ we can finally write the logarithm of the joint probability distribution in a factorized form:

$$\begin{aligned} ln(p(\boldsymbol{X},\boldsymbol{\pi},\boldsymbol{\theta},\boldsymbol{c})) &= ln(p(\boldsymbol{X}|\boldsymbol{\pi},\boldsymbol{\theta},\boldsymbol{c})) + ln(p(\boldsymbol{c}|\boldsymbol{\pi})) + ln(p(\boldsymbol{\pi})) + ln(p(\boldsymbol{\theta})) \\ &= \sum_{i=1}^{n} ln(p(x_i|\boldsymbol{\theta},c_i)) + \sum_{i=1}^{n} ln(p(c_i|\boldsymbol{\pi})) + ln(p(\boldsymbol{\pi})) + ln(p(\boldsymbol{\theta})). \end{aligned} \tag{4.31}$$

This will be beneficial in computing the auxiliary distribution $q(\omega_i)$ according to Equation (4.29) where the logarithm of the joint probability distribution appears in the exponent.

4.4.1. Auxiliary distribution factorization

Let's substitute Equation (4.31) into Equation (4.29) and let's now focus on obtaining the distributions $q(\boldsymbol{\pi})$, $q(\theta_j)$ and $q(c_i)$.

1. $q(\boldsymbol{\pi})$

   Referring to Equation (4.29), the auxiliary distribution $q(\boldsymbol{\pi})$ is proportional to

   $$q(\boldsymbol{\pi}) \propto e^{\mathbb{E}_{\boldsymbol{\theta},\boldsymbol{c}}\left[\sum_{i=1}^{n} ln(p(x_i|\boldsymbol{\theta},c_i)) + \sum_{i=1}^{n} ln(p(c_i|\boldsymbol{\pi})) + ln(p(\boldsymbol{\pi})) + ln(p(\boldsymbol{\theta}))\right]} \tag{4.32}$$

   where the normalization constant Z can be neglected. The operator $\mathbb{E}_{\boldsymbol{\theta},\boldsymbol{c}}$ represents the expectation over all the auxiliary distributions q defined by the parameters in $\boldsymbol{\theta}$ and $\boldsymbol{c}$ but not in $\boldsymbol{\pi}$. However, since the probability distribution $q(\boldsymbol{\pi})$ has to be defined over the parameters in $\boldsymbol{\pi}$, only the terms function of $\boldsymbol{\pi}$ in Equation (4.32) are kept. Thus, Equation (4.32) can be rewritten as follows:

   $$\begin{aligned} q(\boldsymbol{\pi}) &\propto e^{\mathbb{E}_{\boldsymbol{\theta},\boldsymbol{c}}\left[\sum_{i=1}^{n} ln(p(c_i|\boldsymbol{\pi})) + ln(p(\boldsymbol{\pi}))\right]} \\ &\propto e^{\sum_{i=1}^{n} \mathbb{E}_{\boldsymbol{c}}[ln(p(c_i|\boldsymbol{\pi}))] + ln(p(\boldsymbol{\pi}))}. \end{aligned} \tag{4.33}$$

   In Equation (4.33) there are no terms function of $\boldsymbol{\theta}$ and so the expected value with respect to $\boldsymbol{\theta}$ ($\mathbb{E}_{\boldsymbol{\theta}}[\cdot]$) can be neglected. The term $\mathbb{E}_{\boldsymbol{c}}[\cdot]$ represents the expectation over the n distributions $q(c_i)$. By definition, the expectation over a single distribution $q(c_i)$ for a discrete distribution can be analytically represented as

   $$\mathbb{E}_{c_i}[ln(p(c_i|\boldsymbol{\pi}))] = \sum_{j=1}^{K} q(c_i = j) ln(p(c_i = j|\boldsymbol{\pi}) = \sum_{j=1}^{K} \phi_i(j) ln(\pi_j) \tag{4.34}$$

by setting, for simplicity, $\phi_i(j) = q(c_i = j)$. With respect to the second term in the exponent of Equation (4.33), substituting the expression for $p(\boldsymbol{\pi})$ given by the Dirichlet distribution and ignoring constant terms, leads to an expression of $ln(p(\boldsymbol{\pi})) \propto \sum_{j=1}^{K}(\alpha - 1)ln(\pi_j)$. Hence, the new expression for $q(\boldsymbol{\pi})$ can be rewritten as:

$$\begin{aligned} q(\boldsymbol{\pi}) &\propto e^{\sum_{i=1}^{n}\sum_{j=1}^{K}\phi_i(j)ln(\pi_j)+\sum_{j=1}^{K}(\alpha-1)ln(\pi_j)} \\ &\propto e^{\sum_{j=1}^{K}\left[\alpha-1+\sum_{i=1}^{n}\phi_i(j)\right]ln(\pi_j)} \\ &\propto \prod_{j=1}^{K}\pi_j^{\left[\alpha+\sum_{i=1}^{n}\phi_i(j)\right]-1}. \end{aligned} \tag{4.35}$$

Equation (4.35) points out that $q(\boldsymbol{\pi})$ is a Dirichlet distribution with parameter $\alpha'_j = \alpha + \sum_{i=1}^{n}\phi_i(j)$, i.e.:

$$\boldsymbol{\pi} \sim \text{Dirichlet}(\{\alpha'_1, \ldots, \alpha'_K\}).$$

This result was expected because, in the initial distribution selection, the prior $p(\boldsymbol{\pi})$ has been intentionally chosen so to be a conjugate prior for $p(c_i|\boldsymbol{\pi})$. Hence, their product, or the sum of their logarithms, leads to a Dirichlet distribution.

2. $q(\theta_j)$

Let's now focus on the auxiliary distribution function of θ_j. Considering again Equation (4.29) and ignoring the parameters in the exponent different from θ_j, the auxiliary distribution $q(\theta_j)$ becomes proportional to

$$\begin{aligned} q(\theta_j) &\propto e^{\mathbb{E}_{\boldsymbol{\pi},\boldsymbol{c}}\left[\sum_{i=1}^{n} ln(p(x_i|\boldsymbol{\theta},c_i=j))+ln(p(\theta_j))\right]} \\ &\propto e^{\mathbb{E}_{\boldsymbol{c}}\left[\sum_{i=1}^{n} ln(p(x_i|\boldsymbol{\theta},c_i=j))\right]+ln(p(\theta_j))} \\ &\propto e^{\sum_{i=1}^{n}\phi_i(j)ln(\binom{N}{x_i}\theta_j^{x_i}(1-\theta_j)^{(N-x_i)})+ln(\theta_j^{a-1}(1-\theta_j)^{b-1})} \\ &\propto e^{\sum_{i=1}^{n}\phi_i(j)\left[x_i ln(\theta_j)+(N-x_i)ln(1-\theta_j)\right]+(a-1)ln(\theta_j)+(b-1)ln(1-\theta_j)} \\ &\propto e^{(\sum_{i=1}^{n}\phi_i(j)x_i+(a-1))ln(\theta_j)+(\sum_{i=1}^{n}\phi_i(j)(N-x_i)+(b-1))ln(1-\theta_j)} \\ &\propto (\theta_j)^{((a+\sum_{i=1}^{n}\phi_i(j)x_i)-1)}(1-\theta_j)^{((b+\sum_{i=1}^{n}\phi_i(j)(N-x_i))-1)}. \end{aligned} \tag{4.36}$$

It is noteworthy that in the first logarithm term in Equation (4.36), the term $p(x_i|\boldsymbol{\theta}, c_i = j)$ is equivalent to $p(x_i|\theta_j)$. Equation (4.36) points out that the auxiliary distribution $q(\theta_j)$ is a Beta distribution whose parameters are $a'_j = a + \sum_{i=1}^{n}\phi_i(j)x_i$ and $b'_j = b +$

$\sum_{i=1}^{n} \phi_i(j)(N - x_i)$, i.e.:

$$\theta_j \sim \text{Beta}\left(a'_j, b'_j\right).$$

Also in this case, the family of the distributions $q(\theta_j)$ could have been predicted just realizing that, in the argument of $\mathbb{E}_{\boldsymbol{\pi},\boldsymbol{c}}[\cdot]$ in the first line of Equation (4.36), a Binomial distribution $\prod_{i=1}^{n} p(x_i|\boldsymbol{\theta}, c_i = j)$ is multiplied by a Beta distribution $p(\theta_j)$ (the multiplication is represented by the sum of the logarithms) and the latter is a conjugate prior for the former.

3. $q(c_i)$

Lastly, let's consider the auxiliary distribution $q(c_i)$ and prove that it is a Categorical distribution. From Equation (4.29), $q(c_i)$ is directly proportional to

$$\begin{aligned} q(c_i) &\propto e^{\mathbb{E}_{\boldsymbol{\pi},\boldsymbol{\theta}}\left[\sum_{i=1}^{n} ln(p(x_i|\boldsymbol{\theta},c_i))+ln(p(c_i|\boldsymbol{\pi}))+ln(p(\boldsymbol{\pi}))+ln(p(\boldsymbol{\theta}))\right]} \\ &\propto e^{\mathbb{E}_{\boldsymbol{\pi},\boldsymbol{\theta}}[ln(p(x_i|\boldsymbol{\theta},c_i))+ln(p(c_i|\boldsymbol{\pi})))]}. \end{aligned} \tag{4.37}$$

The logarithmic terms in the exponent, $ln(p(x_i|\boldsymbol{\theta}, c_i))$ and $ln(p(c_i|\boldsymbol{\pi}))$, can be written respectively as

$$ln(p(x_i|\boldsymbol{\theta}, c_i)) = \sum_{j=1}^{K} ln(p(x_i|\boldsymbol{\theta}, c_i = j)^{\mathbb{1}[c_i=j]}) \qquad ln(p(c_i|\boldsymbol{\pi})) = \sum_{j=1}^{K} ln(p(c_i = j|\boldsymbol{\pi})^{\mathbb{1}[c_i=j]})$$

yielding:

$$\begin{aligned} &\propto e^{\mathbb{E}_{\boldsymbol{\theta}}\left[\left(\sum_{j=1}^{K} ln(p(x_i|\boldsymbol{\theta},c_i=j))^{\mathbb{1}[c_i=j]}\right)\right]+\mathbb{E}_{\boldsymbol{\pi}}\left[\sum_{j=1}^{K} ln(p(c_i=j|\boldsymbol{\pi}))^{\mathbb{1}[c_i=j]}\right]} \\ &\propto e^{\mathbb{E}_{\boldsymbol{\theta}}\left[\left(ln(p(x_i|\boldsymbol{\theta},c_i=1))^{\mathbb{1}[c_i=j]}\right)\right]+\mathbb{E}_{\boldsymbol{\pi}}\left[ln(p(c_i=1|\boldsymbol{\pi}))^{\mathbb{1}[c_i=j]}\right]} \ldots e^{\mathbb{E}_{\boldsymbol{\theta}}\left[\left(ln(p(x_i|\boldsymbol{\theta},c_i=K))^{\mathbb{1}[c_i=K]}\right)\right]+\mathbb{E}_{\pi}\left[ln(p(c_i=K|\boldsymbol{\pi}))^{\mathbb{1}[c_i=K]}\right]} \\ &\propto \prod_{j=1}^{K}\left[e^{\mathbb{E}_{\boldsymbol{\theta}}[(ln(p(x_i|\boldsymbol{\theta},c_i=j)))]+\mathbb{E}_{\boldsymbol{\pi}}[ln(p(c_i=j|\boldsymbol{\pi}))]}\right]^{\mathbb{1}[c_i=j]} \\ &\propto \prod_{j=1}^{K}\left[e^{\mathbb{E}_{\boldsymbol{\theta}}\left[\left(ln(p(x_i|\theta_j))\right)\right]+\mathbb{E}_{\boldsymbol{\pi}}[ln(p(c_i=j|\boldsymbol{\pi}))]}\right]^{\mathbb{1}[c_i=j]} \end{aligned} \tag{4.38}$$

which is the common form of a Categorical distribution. Equation (4.39) shows that the auxiliary distribution $q(c_i)$ is proportional to the product of K distributions $q(c_i = j)$, each of which expressed as:

$$\begin{aligned} q(c_i = j) &\propto e^{\mathbb{E}_{\theta_j}\left[ln\left(\binom{N}{x_i}\theta_j^{x_i}(1-\theta_j)^{(N-x_i)}\right)\right]+\mathbb{E}_{\boldsymbol{\pi}}\left[ln(\pi_j)\right]} \\ &\propto e^{x_i\mathbb{E}_{\theta_j}\left[ln(\theta_j)\right]+(N-x_i)\mathbb{E}_{\theta_j}\left[ln(1-\theta_j)\right]+\mathbb{E}_{\boldsymbol{\pi}}\left[ln(\pi_j)\right]}. \end{aligned} \tag{4.39}$$

In order to proceed with the calculations we need to estimate the three expectations in Equation (4.39), by considering the results presented in [94].

- $\mathbb{E}_{\boldsymbol{\pi}}\left[ln(\pi_j)\right]$

  Since $q(\boldsymbol{\pi})$ has been proven to be a Dirichlet distribution, it is possible to verify that:

  $$\mathbb{E}_{\boldsymbol{\pi}}\left[ln(\pi_j)\right] = \psi(\alpha'_j) - \psi\left(\sum_{j=1}^{K}\alpha'_j\right) \tag{4.40}$$

  where $\psi(\cdot)$ is the digamma function and $\alpha'_j = \alpha + \sum_{i=1}^{n}\phi_i(j)$ are the parameters defining the Dirichlet distribution $q(\boldsymbol{\pi})$ in Equation (4.35).

- $\mathbb{E}_{\theta_j}\left[ln(\theta_j)\right]$

  Similarly, since $q(\theta_j)$ has been proven to be a Beta distribution, it is easy to verify that:

  $$\mathbb{E}_{\theta_j}\left[ln(\theta_j)\right] = \psi(a'_j) - \psi(a'_j + b'_j) \tag{4.41}$$

  where $a'_j = a + \sum_{i=1}^{n}\phi_i(j)x_i$ and $b'_j = b + \sum_{i=1}^{n}\phi_i(j)(N-x_i)$ are the parameters defining the K Beta distributions $q(\theta_j)$ in Equation (4.36).

- $\mathbb{E}_{\theta_j}\left[ln(1-\theta_j)\right]$

  By definition of the Expected Value, $\mathbb{E}_{\theta_j}\left[ln(1-\theta_j)\right] = \int_0^1 ln(1-\theta_j)\frac{\theta_j^{a'_j-1}(1-\theta_j)^{b'_j-1}}{\beta(a'_j,b'_j)}d\theta_j$.

  By substituting $\theta'_j = 1-\theta_j$ we obtain $\int_0^1 ln(\theta'_j)\frac{(1-\theta'_j)^{a'_j-1}(\theta'_j)^{b'_j-1}}{\beta(a'_j,b'_j)}d\theta'_j$ that can be represented as $\mathbb{E}_{\theta'_j}\left[ln(\theta'_j)\right]$. Hence,

  $$q(\theta'_j) = \frac{(1-\theta'_j)^{a'_j-1}(\theta'_j)^{b'_j-1}}{\beta(a'_j,b'_j)}$$

  is a Beta distribution $\theta'_j \sim \text{Beta}(b'_j,a'_j)$. According to [94], $\mathbb{E}_{\theta_j}\left[ln(1-\theta_j)\right]$ is

computed as:

$$\mathbb{E}_{\theta_j}\left[ln(1-\theta_j)\right] = \mathbb{E}_{\theta'_j}\left[ln(\theta'_j)\right] = \psi(b'_j) - \psi(a'_j + b'_j). \tag{4.42}$$

By plugging Equation (4.40), Equation (4.41), Equation (4.42) into Equation (4.39) we obtain:

$$q(c_i = j) \propto e^{x_i\left[\psi(a'_j)-\psi(a'_j+b'_j)\right]+(N-x_i)\left[\psi(b'_j)-\psi(a'_j+b'_j)\right]+\psi(\alpha'_j)-\psi\left(\sum_{j=1}^{K}\alpha'_j\right)} \tag{4.43}$$

and defining three terms as:

- $t_{1j} = \psi(a'_j) - \psi(a'_j + b'_j)$
- $t_{2j} = \psi(b'_j) - \psi(a'_j + b'_j)$
- $t_{3j} = \psi(\alpha'_j) - \psi\left(\sum_{j=1}^{K}\alpha'_j\right)$

we can rewrite Equation (4.43) as:

$$q(c_i = j) \quad \propto \quad e^{x_i t_{1j}+(N-x_i)t_{2j}+t_{3j}}. \tag{4.44}$$

In order to evaluate the distribution $q(c_i = j)$, Equation (4.44) has to be factorized:

$$q(c_i = j) = \phi_i(j) = \frac{e^{x_i t_{1j}+(N-x_i)t_{2j}+t_{3j}}}{\sum_{k=1}^{K} e^{x_i t_{1k}+(N-x_i)t_{2k}+t_{3k}}} \tag{4.45}$$

so that $\sum_{k=1}^{K}\phi_i(k) = 1$.

At this point we obtained the analytical expressions for the auxiliary distributions $q(\boldsymbol{\pi})$, $q(\theta_j)$ and $q(c_i)$ and so we are ready to proceed to the computation of the objective function $\mathscr{L}$.

4.4.2. Objective function $\mathscr{L}$

Although the auxiliary distribution approximating the posterior distribution has been estimated, the analytical expression for the objective function $\mathscr{L}$ has not been developed yet. Recalling

Equation (4.11) we can write:

$$\begin{aligned}
\mathscr{L} &= \mathbb{E}[ln(p(\boldsymbol{X},\boldsymbol{\theta},\boldsymbol{\pi},\boldsymbol{c}))] - \mathbb{E}[ln(q(\boldsymbol{\theta},\boldsymbol{\pi},\boldsymbol{c}))] \\
&= \mathbb{E}[ln(p(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{c})p(\boldsymbol{c}|\boldsymbol{\pi})p(\boldsymbol{\pi})p(\boldsymbol{\theta}))] - \mathbb{E}[ln(q(\boldsymbol{\theta})q(\boldsymbol{\pi})q(\boldsymbol{c}))] \\
&= \mathbb{E}\left[\sum_{i=1}^{n}[ln(p(x_i|\boldsymbol{\theta},c_i)) + ln(p(c_i|\boldsymbol{\pi}))] + ln(p(\boldsymbol{\pi})) + ln(p(\boldsymbol{\theta}))\right] + \\
&- \mathbb{E}[ln(q(\boldsymbol{\pi}))] - \mathbb{E}\left[ln\left(\prod_{j=1}^{K} q(\theta_j)\right)\right] - \mathbb{E}\left[\left(\prod_{i=1}^{n} ln(q(c_i))\right)\right]. \qquad (4.46)
\end{aligned}$$

Let's now focus on the first of the four terms on the right hand side of Equation (4.46) and consider the expected values only for those parameters that appear in each term:

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i=1}^{n}[ln(p(x_i|\boldsymbol{\theta},c_i)) + ln(p(c_i|\boldsymbol{\pi}))] + ln(p(\boldsymbol{\pi})) + ln(p(\boldsymbol{\theta}))\right] \\
= &\sum_{i=1}^{n}\left(\mathbb{E}_{\boldsymbol{\theta}}[\mathbb{E}_{\boldsymbol{c}}[ln(p(x_i|\boldsymbol{\theta},c_i))]] + \mathbb{E}_{\boldsymbol{\pi}}[\mathbb{E}_{\boldsymbol{c}}[ln(p(c_i|\boldsymbol{\pi}))]]\right) + \mathbb{E}_{\boldsymbol{\pi}}[ln(p(\boldsymbol{\pi}))] + \mathbb{E}_{\boldsymbol{\theta}}[p(\boldsymbol{\theta})] \qquad (4.47)
\end{aligned}$$

Recalling the expectations computed in Equation (4.34) and Equation (4.36), Equation (4.47) becomes:

$$\begin{aligned}
&\sum_{i=1}^{n}\mathbb{E}_{\boldsymbol{\theta}}\left[\sum_{j=1}^{K}\phi_i(j)[ln(p(x_i|\boldsymbol{\theta},c_i=j))] + \sum_{j=1}^{K}\phi_i(j)\mathbb{E}_{\boldsymbol{\pi}}[ln(p(c_i=j|\boldsymbol{\pi}))]\right] + \mathbb{E}_{\boldsymbol{\pi}}[ln(p(\boldsymbol{\pi}))] + \mathbb{E}_{\boldsymbol{\theta}}[p(\boldsymbol{\theta})] \\
= &\sum_{i=1}^{n}\mathbb{E}_{\boldsymbol{\theta}}\left[\sum_{j=1}^{K}\phi_i(j)\left(ln\binom{N}{x_i} + x_i[ln\theta_j] + (N-x_i)[ln(1-\theta_j)] + \mathbb{E}_{\boldsymbol{\pi}}[ln(\pi_j)]\right)\right] + \mathbb{E}_{\boldsymbol{\pi}}[ln(p(\boldsymbol{\pi}))] \\
&+\mathbb{E}_{\boldsymbol{\theta}}[p(\boldsymbol{\theta})] \qquad (4.48)
\end{aligned}$$

Taking the expected values over $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$ into the summations and substituting the expressions of the Dirichlet distribution for $p(\boldsymbol{\pi})$ and Beta distributions for the K $p(\theta_j)$ into Equation (4.48) leads to:

$$\begin{aligned}
&= \sum_{i=1}^{n}\sum_{j=1}^{K}\phi_i(j)\left[ln\binom{N}{x_i} + x_i\mathbb{E}_{\theta_j}[ln(\theta_j)] + (N-x_i)\mathbb{E}_{\theta_j}[ln(1-\theta_j)] + \mathbb{E}_{\boldsymbol{\pi}}[ln(\pi_j)]\right] \\
&+ \mathbb{E}_{\boldsymbol{\pi}}\left[ln\left(\frac{\Gamma(K\alpha)}{K\Gamma(\alpha)}\prod_{j=1}^{K}\pi_j^{\alpha-1}\right)\right] + \mathbb{E}_{\boldsymbol{\theta}}\left[ln\left(\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\prod_{j=1}^{K}\theta_j^{a-1}(1-\theta_j)^{b-1}\right)\right]
\end{aligned}$$

$$
\begin{aligned}
= & \sum_{i=1}^{n}\sum_{j=1}^{K}\phi_i(j)\left[ln\binom{N}{x_i}+x_i\mathbb{E}_{\theta_j}[ln(\theta_j)]+(N-x_i)\mathbb{E}_{\theta_j}[ln(1-\theta_j)]+\mathbb{E}_{\boldsymbol{\pi}}[ln(\pi_j)]\right] \\
& +(\alpha-1)\sum_{j=1}^{K}\mathbb{E}_{\boldsymbol{\pi}}[ln(\pi_j)]+(a-1)\sum_{j=1}^{K}\mathbb{E}_{\theta_j}[ln(\theta_j)]+(b-1)\sum_{j=1}^{K}\mathbb{E}_{\theta_j}[ln(1-\theta_j)] \\
& +ln\left(\frac{\Gamma(K\alpha)}{K\Gamma(\alpha)}\right)+ln\left(\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right).
\end{aligned}
\tag{4.49}
$$

By substituting Equation (4.40), Equation (4.41) and Equation (4.42) into Equation (4.49):

$$
\begin{aligned}
= & \sum_{i=1}^{n}\sum_{j=1}^{K}\phi_i(j)\left[ln\binom{N}{x_i}+x_i\left[\psi(a'_j)-\psi(a'_j+b'_j)\right]+(N-x_i)\left[\psi(b'_j)-\psi(a'_j+b'_j)\right]+\psi(\alpha'_j)-\psi\left(\sum_{j=1}^{K}\alpha'_j\right)\right] \\
+ & (\alpha-1)\sum_{j=1}^{K}\left[\psi(\alpha'_j)-\psi\left(\sum_{j=1}^{K}\alpha'_j\right)\right]+(a-1)\sum_{j=1}^{K}\left[\psi(a'_j)-\psi(a'_j+b'_j)\right]+(b-1)\sum_{j=1}^{K}\left[\psi(b'_j)-\psi(a'_j+b'_j)\right] \\
+ & ln\left(\frac{\Gamma(K\alpha)}{K\Gamma(\alpha)}\right)+ln\left(\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right)
\end{aligned}
\tag{4.50}
$$

and, recalling the definition of the three terms t_{1j}, t_{2j} and t_{3j}, we can conclude that the analytical expression for the first component of the objective function $\mathscr{L}$, given in Equation (4.46), can be reduced to:

$$
\begin{aligned}
& \mathbb{E}\left[\sum_{i=1}^{n}[ln(p(x_i|\boldsymbol{\theta},c_i))+ln(p(c_i|\boldsymbol{\pi}))]+ln(p(\boldsymbol{\pi}))+ln(p(\boldsymbol{\theta}))\right] \\
= & \sum_{i=1}^{n}\sum_{j=1}^{K}\phi_i(j)\left[ln\binom{N}{x_i}+x_it_{1j}+(N-x_i)t_{2j}+t_{3j}\right]+(\alpha-1)\sum_{j=1}^{K}t_{3j}+(a-1)\sum_{j=1}^{K}t_{1j}+(b-1)\sum_{j=1}^{K}t_{2j} \\
+ & ln\left(\frac{\Gamma(K\alpha)}{K\Gamma(\alpha_j)}\right)+ln\left(\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right).
\end{aligned}
\tag{4.51}
$$

Let's now focus on the last three terms on the right hand side of Equation (4.46). Each term is, by definition, the entropy of the corresponding distribution.

- The distribution $q(\boldsymbol{\pi})$ has been shown to be a Dirichlet distribution and its entropy is computed as:

$$
-\mathbb{E}_{\boldsymbol{\pi}}[ln(q(\boldsymbol{\pi}))]=\left[ln\left(\frac{\prod_{j=1}^{K}\Gamma(\alpha'_j)}{\Gamma(\sum_{j=1}^{K}\alpha'_j)}\right)+\left(\sum_{j=1}^{K}\alpha'_j-K\right)\psi\left(\sum_{j=1}^{K}\alpha'_j\right)-\sum_{j=1}^{K}(\alpha'_j-1)\psi(\alpha'_j)\right].
\tag{4.52}
$$

- The distributions $q(\theta_j)$ have been shown to be Beta distributions. The entropy of $q(\theta_j)$

is computed as:

$$
\begin{aligned}
- \quad & \mathbb{E}_{\boldsymbol{\theta}}\left[ln\left(\prod_{j=1}^{K} q(\theta_j)\right)\right] \\
= \quad & \sum_{j=1}^{K}\left[ln\left(\frac{\Gamma(a'_j)\Gamma(b'_j)}{\Gamma\left(a'_j+b'_j\right)}\right) - (a'_j-1)\psi(a'_j) - (b'_j-1)\psi(b'_j) + (a'_j+b'_j-2)\psi(a'_j+b'_j)\right].
\end{aligned}
\tag{4.53}
$$

- Last, the distributions $q(c_i)$ have been shown to be a Discrete distribution whose entropy is given by definition as:

$$
-\mathbb{E}_{\boldsymbol{c}}\left[ln\left(\prod_{i=1}^{n} q(c_i)\right)\right] = -\sum_{i=1}^{n}\sum_{j=1}^{K}\phi_i(j)ln(\phi_i(j)). \tag{4.54}
$$

Finally, the objective function $\mathscr{L}$ can be then computed through the sum of the contributes given by Equation (4.51), Equation (4.52), Equation (4.53) and Equation (4.54):

$$
\begin{aligned}
&\mathscr{L} = \sum_{i=1}^{n}\sum_{j=1}^{K}\phi_i(j)\left[ln\binom{N}{x_i} + x_i t_{1j} + (N-x_i)t_{2j} + t_{3j}\right] + (\alpha-1)\sum_{j=1}^{K} t_{3j} + (a-1)\sum_{j=1}^{K} t_{1j} + (b-1)\sum_{j=1}^{K} t_{2j} + \\
&ln\left(\frac{\prod_{j=1}^{K}\Gamma(\alpha'_j)}{\Gamma\left(\sum_{j=1}^{K}\alpha'_j\right)}\right) + \left(\sum_{j=1}^{K}\alpha'_j - K\right)\psi\left(\sum_{j=1}^{K}\alpha'_j\right) - \sum_{j=1}^{K}(\alpha'_j-1)\psi(\alpha'_j) + Kln\left(\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right) + ln\left(\frac{\Gamma(K\alpha)}{\Gamma(\alpha)}\right) + \\
&+\sum_{j=1}^{K}\left[ln\left(\frac{\Gamma(a'_j)\Gamma(b'_j)}{\Gamma\left(a'_j+b'_j\right)}\right) - (a'_j-1)\psi(a'_j) - (b'_j-1)\psi(b'_j) + (a'_j+b'_j-2)\psi(a'_j+b'_j)\right] - \sum_{i=1}^{n}\sum_{j=1}^{K}\phi_i(j)ln(\phi_i(j)).
\end{aligned}
\tag{4.55}
$$

This expression of the objective function $\mathscr{L}$ can now be used in an iterative process to assess the convergence of the parameters defining the auxiliary function. According to the theoretical foundation of the VI algorithm, this objective function is expected to monotonically converge to a negative constant value.

4.4.3. VI Algorithm: pseudocode

Once the analytical expressions for the distributions $q(\boldsymbol{\pi})$, $q(\theta_j)$ and $q(c_i)$ have been derived and the inference between their parameters has been shown, it is possible to use these distributions in an iterative algorithm that updates the distributions' parameters. The correct implementation and convergence of the algorithm is validated by the objective function $\mathscr{L}$ monotonically increasing before converging to negative constant value.

To explain the algorithm let us assume a dataset $\boldsymbol{X} = \{x_1, x_2 \ldots x_n\}$ with $x_i \in (1, \ldots, N)$.

The first step is to choose the maximum number of clusters K (i.e. the number of Binomial distributions in the mixture) we expect to find in the data. In practice the parameter K can be chosen arbitrarily large, overestimating the number of mixture components in the real model: unnecessary clusters will not be considered by the algorithm in the iterative process.

The implementation of the VI algorithm relies on three steps. Indicating with T the index of iteration ($T = 1, \ldots, N_T$):

1. Inizialization at $T = 0$: select an initial set of parameters $\boldsymbol{\alpha} = \{\alpha_1^0, \ldots, \alpha_K^0\}$ and $(a_j'^0, b_j'^0)$ for $j = 1, \ldots, K$. In this step the initial parameters for the distributions $q(\boldsymbol{\pi})$ and $q(\theta_j)$ are selected. Note that it would be possible also to initialize the parameters for $q(c_i)$, but they will be update at the beginning of the next iteration.
2. At the generic T^{th} iteration:
   - Update the parameters $\phi_i(j)$ according to Equation (4.45), for $i = 1, \ldots, n$ and for $j = 1, \ldots, K$ using the parameters α_j', a_j', b_j' from the previous iteration:

$$\phi_i^T(j) = \frac{e^{x_i t_{1j}^T + (N - x_i) t_{2j}^T + t_{3j}^T}}{\sum_{j=1}^K e^{x_i t_{1j}^T + (N - x_i) t_{2j}^T + t_{3j}^T}} \tag{4.56}$$

     where:

$$\begin{aligned} t_{1j}^T &= \psi(a_j'^{(T-1)}) - \psi(a_j'^{(T-1)} + b_j'^{(T-1)}) \\ t_{2j}^T &= \psi(b_j'^{(T-1)}) - \psi(a_j'^{(T-1)} + b_j'^{(T-1)}) \\ t_{3j}^T &= \psi(\alpha_j'^{(T-1)}) - \psi\left(\sum_{j=1}^K \alpha_j'^{(T-1)}\right). \end{aligned}$$

   - Use the new $\phi_i(j)$ to update the parameters of the distributions $q(\boldsymbol{\pi})$ and $q(\theta_j)$ for $j = 1, \ldots, K$:

$$\begin{aligned} \alpha_j'^T &= \alpha + \sum_{i=1}^n \phi_i^T(j) \\ a_j'^T &= a + \sum_{i=1}^n \phi_i^T(j) x_i \\ b_j'^T &= b + \sum_{i=1}^n \phi_i^T(j)(N - x_i). \end{aligned}$$

   - Compute the value of the objective function $\mathscr{L}^T$ at the T^{th} iteration using Equation (4.55) and compare it with the value corresponding to the previous iteration.
3. Check convergence of $\mathscr{L}^T$ to a constant value. Once this convergence is achieved, that value of $\mathscr{L}^T$ corresponds to the final value of $\mathscr{L}$.

As the objective function $\mathscr{L}$ converges to a constant value, the auxiliary distribution ($q(\boldsymbol{\pi}, \boldsymbol{\theta}, \boldsymbol{c})$) given in Equation (4.28) approaches the posterior distribution $p(\boldsymbol{\pi}, \boldsymbol{\theta}, \boldsymbol{c}|\boldsymbol{X})$ (Equation (4.27)).

It is noteworthy that VI algorithms are, by nature, computationally intensive. In fact, the

practical implementation may become cumbersome when in the presence of an excessively large numerical value in the observations (N large): this is because in the computation of the objective function $\mathscr{L}$, the term $\binom{N}{x_i}$ in Equation (4.55) becomes intractable as N approaches large values.

4.4.4. Clustering procedure for VI

Recalling that the terms $\phi_i(j)$ represent the probability that the observation x_i belongs to the j^{th} mixture component, $q(c_i = j)$, as the convergence of the algorithm is assessed, they need to be closely considered for clustering purposes since they drive the assignment of the n observations to the most likely mixture component. The sets of observations in $\boldsymbol{X}$ belonging to the same mixture component are claimed to be a cluster. Hence, each observation x_i will be assigned to the cluster associated with the largest $\phi_i(j)$ among all possible mixture components $j = 1,\ldots,K$.

A useful application of a VI-based clustering algorithm is in the analysis of complex response power spectra obtained from vibration measurements (e.g. accelerations) recorded at different locations in a structure. Let's recall from structural dynamics that the power spectrum of the acceleration response time histories reveals the energy distribution of the response in the frequency domain. At every sensor, several observations of the structure in a given state (healthy or damaged) can be collected to compose training or test datasets and a local average spectrum of the response can be computed at every location to provide a local spectral representation of the energy contribution. Furthermore, the contribution to the spectral energy associated with a given structural mode may vary from record to record (from location to location) and so, it might be appropriate to consider an average spectrum obtained by the measurements from all sensors. By averaging the different spectra coming from the different locations within a structure, local information are blended together and this might require additional tools to identify the contribution of the different modes. It is in this framework that the proposed VI-based clustering algorithm has to be considered for classification purpose: it will help assigning the spectral contributions to the most likely mixture component (e.g. structural mode).

In order to apply the proposed VI clustering procedure to the problem associated with the spectral analysis, let's recall that the dataset is represented by the vector $\boldsymbol{X} = \{x_1, x_2, \ldots, x_n\}$ with $x_i \in (1,\ldots,N)$. In a spectral framework, given the sampling period dt of the signals, the

frequency domain is discretized with frequency step $\Delta f = \frac{1}{dt}$ so that the first N samples can be written as $f_\xi = \xi \times \Delta f$ for $\xi = 1,\ldots,N$. The integer values ξ can be considered as those values that a generic observation x_i can assume, $(\xi = 1,\ldots,N)$, indicating the ξ^{th} frequency sample, which is associated to a specific value of the average power spectrum. Such a value of the average power spectrum can be interpreted as the 'frequency of occurrence' of the observations $x_i = \xi$ in the dataset. The frequency of occurrence establishes the total number of observations $x_i = \xi$, approximated to the closest integer, appearing in the dataset $\boldsymbol{X}$. For example, given $\Delta f = 0.2$ Hz and considering the fifth frequency step ($\xi = 5$), let's assume that the value of the average power spectrum in correspondence of $f_5 = 1$ Hz is equal to 3.1 g^2. Then, this means that there are exactly 3 observations for $x_i = \xi = 5$ appearing in the observation vector $\boldsymbol{X}$, i.e. $\boldsymbol{X} = \{x_1, x_2, \ldots, x_s = 5, x_{s+1} = 5, x_{s+2} = 5, \ldots, x_n\}$.

At this point, it is important to highlight some numerical difficulties that arise when dealing with real applications. One of the main stumbling blocks is linked to the length of the time history signals using the analysis. For example, dealing with time histories 100 sec long, sampled at 0.01 sec, leads to vectors of 10000 elements. By using an optimized p-points Fast Fourier Transform (FFT) to convert such signals into the frequency domain leads to a frequency spectrum defined by a vector of $p = 4096$ elements, hence, $N = 4096$. From a practical point of view, the computation of the term $\binom{4096}{x_i}$ in Equation (4.55) is numerically intractable and so the objective function $\mathscr{L}$ cannot be numerically calculated for large values of N. There are possible numerical implementations that allow us to overcome this limitation. One way is to use downsampling techniques through a filter bank of triangular filters inspired to the Mel-scale filter bank [13]. Each triangular filter spans over a frequency interval and has a magnitude equal to 1 at the center frequency, decreasing linearly to 0 at the ends of the frequency interval. The filter bank is created by overlapping the triangular filters so that each filter spans over half of the frequency interval of each of the two adjacent filters. These filters ensure that the spectrum, after downsampling, conserves its total energy, providing a 'smoother' spectrum at the expenses of a loss of resolution. In this chapter, the frequency bandwidth for each filter has been set equal to $6\Delta f$ and the distance between peaks of two adjacent triangles to $3\Delta f$.

The procedure just presented is illustrated in Fig. 4.1. An 8-DOF shear-type system, with frequencies ranging from 0 to 50 Hz, is excited by a Gaussian white noise excitation. The time histories of the structural accelerations at every floor have been sampled at 0.01 sec for a duration of 100 sec. In the spectral analysis a frequency range between 0 and 50 Hz has been

considered. The downsampled average power spectrum is the result of the application of the triangular filter bank to the average power spectrum.

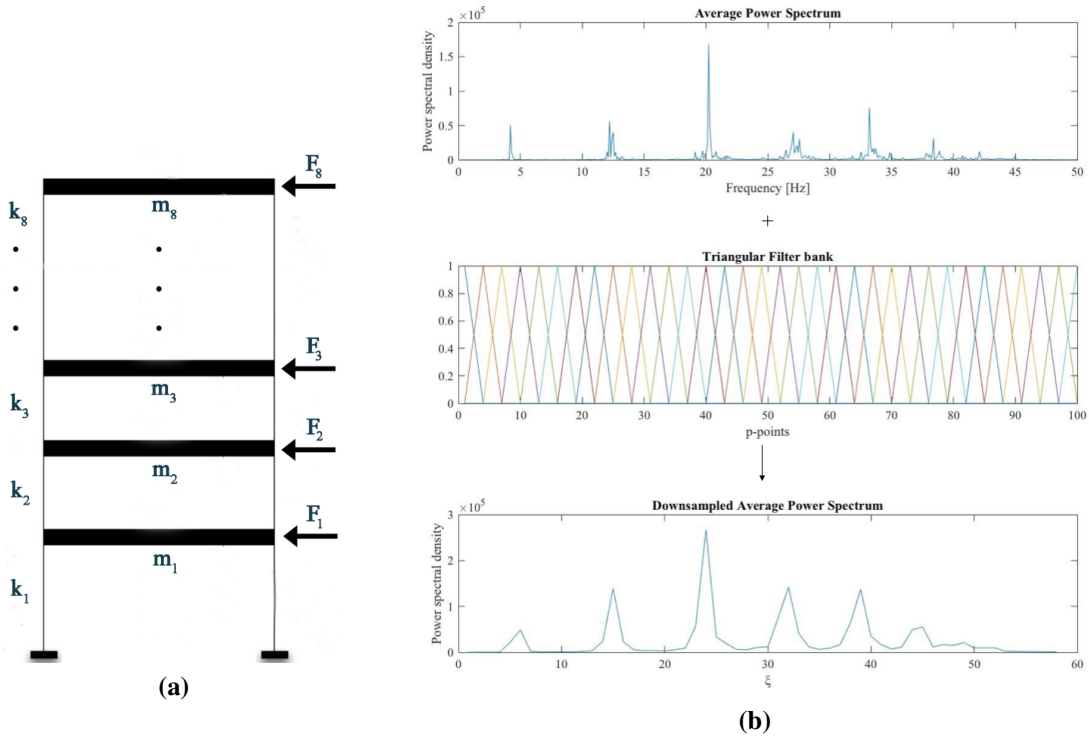


Fig. 4.1. Fully excited shear type system (a). Extraction of the Downsampled Average Power Spectrum (b).

At this point, the average downsampled spectrum is represented by a vector $\boldsymbol{S} \in \mathbb{R}^{1\times D}$ with $D \leq N$. Since the number of observations x_i for $i = 1,\ldots,n$ depends on the numerical values of the downsampled spectrum, it may be convenient, to ensure a rich dataset $\boldsymbol{X}$, to scale up the numerical values of the downsampled spectrum. One way to do so is by dividing the values of the spectrum by its smallest element:

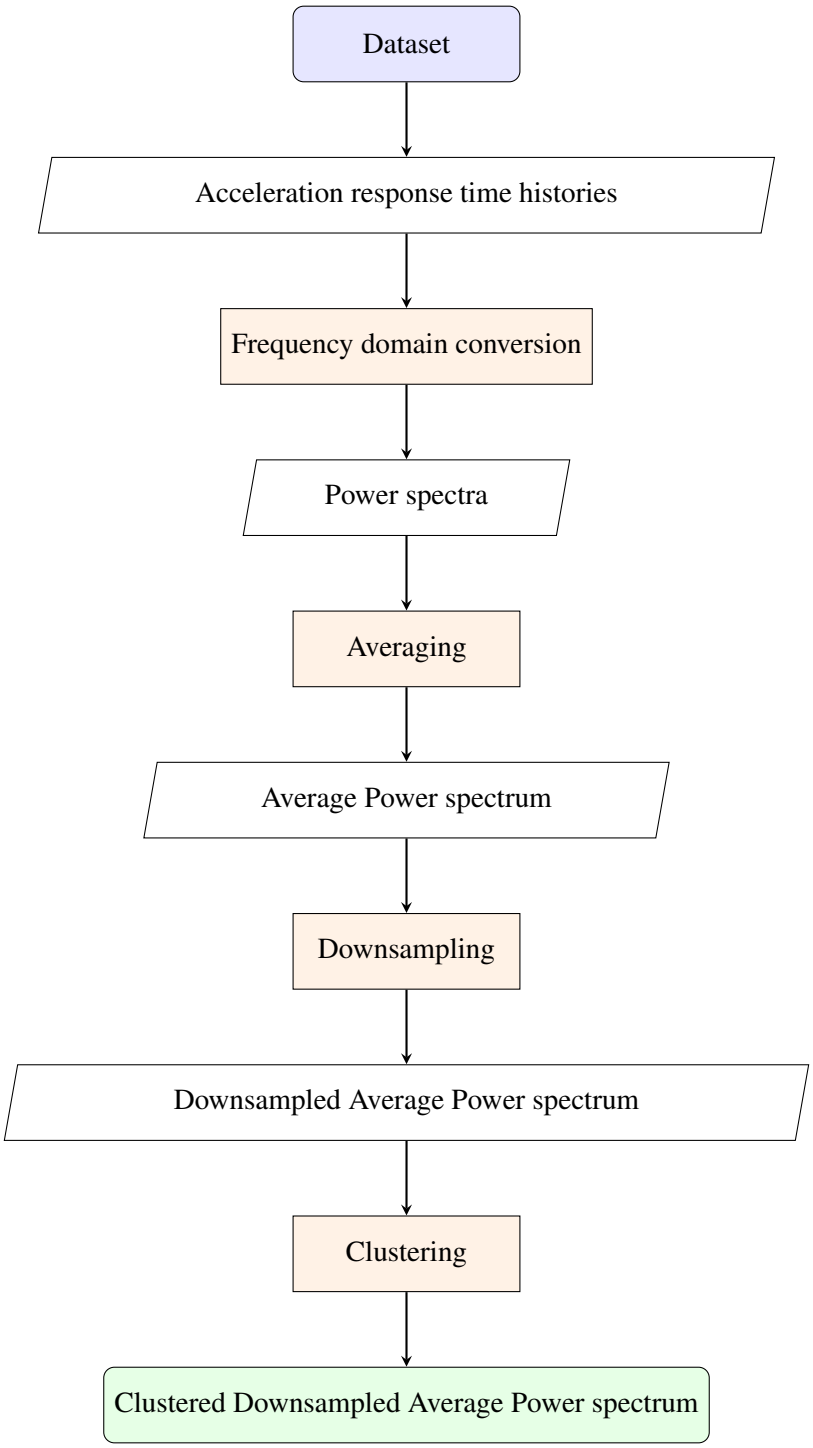
$$\boldsymbol{S}^s = \frac{\boldsymbol{S}}{\min(\boldsymbol{S})} \tag{4.57}$$

where $\boldsymbol{S}^s = \{\mathscr{S}_1^s,\ldots,\mathscr{S}_D^s\}$ is a vector of D elements. The generic value $\mathscr{S}_k^s$ represents the number of observations in the dataset $\boldsymbol{X}$ that have values equal to $\frac{f_k}{\Delta f}$. Hence:

$$n = \sum_{i=1}^{D} \mathscr{S}_i^s. \tag{4.58}$$

Once the dataset $\boldsymbol{X} = \{x_1,\ldots,x_n\}$ is available, the VI clustering algorithm can be applied. The

procedure to extract the Clustered Downsampled Average Power spectrum is shown in the flow chart below.



4.4.5. Feature vectors and classification

To include the proposed VI clustering algorithm into a pattern recognition-based damage assessment strategy, let's first consider the case when multiple sets of response time histories

corresponding to a given condition of the structure are available.

Once the dataset $\boldsymbol{X}$ has been obtained from the downsampled average spectrum, the VI clustering algorithm can be applied for the identification of the individual clusters. As the objective function $\mathscr{L}$ converges, the clusters contained in the dataset $\boldsymbol{X}$ become progressively better defined leading to a 'clustered' downsampled average spectrum. For each of the identified K' clusters ($K' \leq K$), it is then possible to obtain the mean μ'_k and standard deviation σ'_k ($k = 1, \ldots, K'$) for the values of the parameter ξ within the given cluster. Assuming a normal distribution for such values, the range $[\mu'_k - \sigma'_k \quad \mu'_k + \sigma'_k]$, corresponding to a confidence level of 68.27%, represents the range attributed to the specific cluster.

After the intervals have been identified, an upsampling of the clustered average spectrum is performed so to map from the intervals $[\mu'_k - \sigma'_k \quad \mu'_k + \sigma'_k]$ in the ξ domain to the corresponding intervals $[\mu_k - \sigma_k \quad \mu_k + \sigma_k]$ in the original frequency domain. In this way, the definition of the frequency domain interval associated with each cluster is necessary to define the upper and lower cut-off frequencies of the FIR filters that will be used to isolate the contribution of each structural mode. It is worthy to remark that the entire process is performed automatically by the VI clustering algorithm without any input by the user.

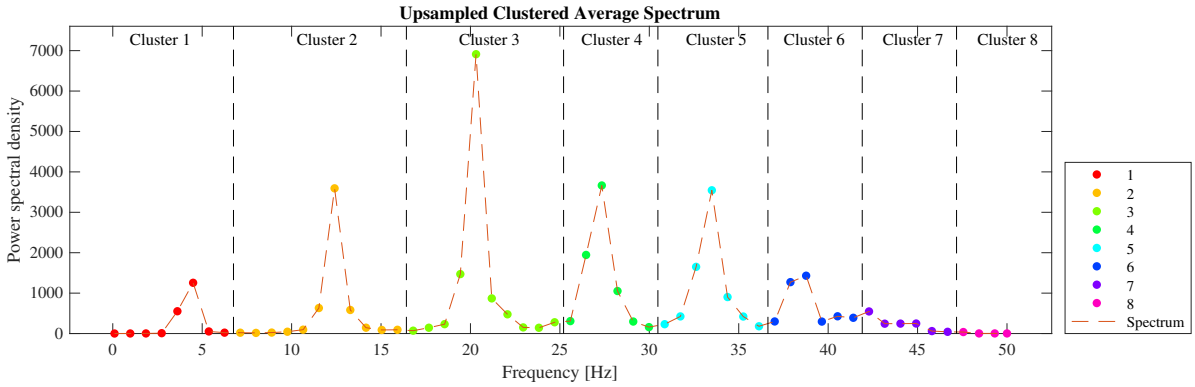


Fig. 4.2. Example of clustered average spectrum.

Figure 4.2 shows an example of a clustered average spectrum (upsampled to the frequency domain) as the result of the VI clustering algorithm applied to the downsampled average power spectrum in Fig. 4.1. A total of $K' = 6$ clusters have been identified in the frequency range [0 50] Hz.

In a pattern recognition framework, the idea is to use a certain set of time histories to generate a model representative of the structure in a baseline condition (usually the healthy condition) and to adopt a novelty detection approach when new data from the structure in unknown condi-

tions become available. The first stage to generate the baseline model is usually referred to as 'training phase', while the novelty detection strategy goes under the name of 'test phase'. Using the data from the training dataset, the clustering algorithm leads to the identification of K'_{tr} clusters. For the generic k_{tr} cluster ($k_{tr} = 1, \ldots, K'_{tr}$), the interval $[\mu_{tr,k_{tr}} - \sigma_{tr,k_{tr}} \quad \mu_{tr,k_{tr}} + \sigma_{tr,k_{tr}}]$ represents the frequency range in which the energy contribution associated with the k_{tr} cluster from the training dataset is condensed. Based on this interval, it is possible to define a FIR filter that can be used, for any spectrum at every monitored location, to isolate the energy contribution associated with a given vibrational mode.

As the occurrence of a local damage causes abrupt changes in a structure, these changes have an impact, more or less evident, on the structural mode shapes and consequently on their energy contribution to the dynamic response of the structure. However, in the presence of damage, it is reasonable to assume that the mode shapes corresponding to the undamaged and damaged conditions, and their energy contributions are somehow correlated with each other. A high level of correlation indicates small changes in the structural modes, while low level of correlation is indicative of large differences. In dealing now with two datasets, one used in the training phase and one in the test phase, the difficult task is to properly match corresponding modes so to test their correlation: a common tool to help with this operation is a classifier that can be used to assess the degree of correlation between a pair of modes, one from the training and one from the test. This classification can be based on a variety of characteristic quantities: in this chapter, this classification is based on the energy content of the contributing modes. The basic idea is to associate the spectral energy contribution of each mode in the undamaged state with the energy contribution of the same mode in an unknown state.

To accomplish this, let's assume that the dynamic response is monitored at N_{DOF} locations in the structure and that a number of N_{ob}^{tr} tests are conducted as a part of the training session. For each of the N_{ob}^{tr} tests, a spectrum of the structural response can be obtained at any of the N_{DOF} locations, following the procedure presented earlier, and the contribution from the k_{tr} cluster, obtained from the interval $[\mu_{tr,k_{tr}} - \sigma_{tr,k_{tr}} \quad \mu_{tr,k_{tr}} + \sigma_{tr,k_{tr}}]$, can be highlighted from each of them. At this point, for each observation $i = 1, \ldots, N_{ob}^{tr}$, it is possible to define a feature vector $\mathscr{G}_{tr,k_{tr}}^{(i)} \in \mathbb{R}^{N_{DOF} \times 1}$, characterizing the specific k_{tr} cluster for the i^{th} observation, whose components represent the amount of vibrational energy in the interval $[\mu_{tr,k_{tr}} - \sigma_{tr,k_{tr}} \quad \mu_{tr,k_{tr}} + \sigma_{tr,k_{tr}}]$ in the spectrum computed at any location.

Having a consistent number of samples of acceleration response time histories in the train-

ing dataset (N_{ob}^{tr}) allows us to set up and train a classic Artificial Neural Network (ANN) classifier. The input dataset consists of the $N_{ob}^{tr} \times K_{tr}'$ feature vectors $\mathscr{G}_{tr,k_{tr}}^{(i)}$ ($i = 1,\ldots,N_{ob}^{tr}$ and $k_{tr} = 1,\ldots,K_{tr}'$), and the output of their associated classes defined by the k_{tr} cluster ($k_{tr} = 1,\ldots,K_{tr}'$). For cross-validation purposes, the samples of the entire dataset are split into two subsets: 1) training data (80%) and 2) validation data (20%).

Figure 4.13 shows the architecture of an ANN classifier for the 8-DOF system previously analyzed. The number of neurons in the hidden layer has been arbitrarily set equal to 10 and sigmoid functions have been selected as activation functions. For the output, 8 neurons, with softmax functions as activation functions, have been considered.

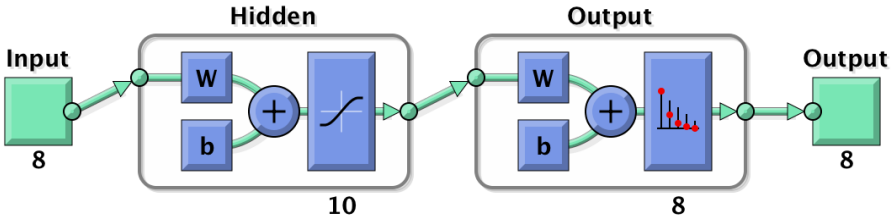


Fig. 4.3. ANN architecture (10 hidden neurons) for an 8DOF system and $K' = 8$ identified clusters in the average spectrum.

Once the network is trained, it can be used to place the modal information coming from the analysis of new test data into the clusters obtained from the training data. This will allow us to check the degree of correlation between modes extracted from the test data and modes identified through the training data. By considering a test dataset of N_{ob}^{te} observations of acceleration response time histories, it is possible to identify K_{te}' clusters following the same procedure used in the training phase. Let $\mathscr{G}_{te,k_{te}}^{(j)} \in \mathbb{R}^{N_{DOF}\times 1}$ indicate the feature vector characterizing the k_{te} cluster ($k_{te} = 1,\ldots,K_{te}'$) for the j^{th} observation ($j = 1,\ldots,N_{ob}^{te}$) in the test dataset. Each of these $N_{ob}^{te} \times K_{te}'$ feature vectors $\mathscr{G}_{te,k_{te}}^{(j)}$ can now be used as input in the trained ANN classifier so that it can be classified into one of the categories (clusters) defined in the training phase of the ANN. The final goal is to associate each of the K_{te}' test clusters to one of the training clusters K_{tr}'. For the generic k_{te} cluster, N_{ob}^{te} feature vectors $\mathscr{G}_{te,k_{te}}^{(j)}$ are classified into one of the K_{tr}' training clusters by choosing the k_{tr} cluster that has been associated with the majority of the N_{ob}^{te} feature vectors $\mathscr{G}_{te,k_{te}}^{(j)}$ tested. Finally, the mean of the N_{ob}^{tr} observed feature vectors $\bar{\mathscr{G}}_{tr,k_{tr}} = \frac{1}{N_{ob}^{tr}} \sum_{i=1}^{N_{ob}^{tr}} \mathscr{G}_{tr,k_{tr}}^{(i)}$ and the mean of the N_{ob}^{te} vectors $\bar{\mathscr{G}}_{te,k_{te}} = \frac{1}{N_{ob}^{te}} \sum_{j=1}^{N_{ob}^{te}} \mathscr{G}_{te,k_{te}}^{(j)}$ are computed. The coupling is claimed to be valid if the MAC number for the feature vectors $\bar{\mathscr{G}}_{tr,k_{tr}}$ and $\bar{\mathscr{G}}_{te,k_{te}}$ is larger than a threshold value set to 0.9.

It is worthy pointing out that the order in which the elements are positioned into the vectors $\mathscr{G}^{(i)}_{tr,k_{tr}} \in \mathbb{R}^{N_{DOF}\times 1}$ doesn't affect the performance of the classification process as long as the ordering criteria follows the same logic for the vectors $\mathscr{G}^{(j)}_{te,k_{te}} \in \mathbb{R}^{N_{DOF}\times 1}$.

Up to this point, the entire analysis has been conducted on features extracted from representations of the signals (acceleration response time histories) in the frequency domain. It is now possible to extract the information about the frequency energy content of the structural responses so to generate time domain signals which can be used in determining damage sensitive features, referred to as Inner Product Vectors (IPVs), at the core of the proposed structural damage assessment method.

4.4.6. IPV for a cross-correlation based damage detection

The use of Inner Product Vectors (IPVs) has been proven to be successful in assessing the presence and location of damage in a structure by just considering vibrational data. The theoretical formulation of the IPV features has been presented in the work by Le Wang et al. [70, 89] and by Morgantini and Betti [90]. Such features have been proven to be reliable damage sensitive features when the system is subjected to unit pulses and Gaussian white noise excitations. The focal point of the methodology is that it relies on the determination of the cross-correlation between acceleration response time histories, generated exclusively by a specific structural mode, recorded at different structural locations. While Le Wang et al. obtained this contribution by finely selecting a proper input excitation, Morgantini and Betti rely on the use of a proper filter to separate the contribution of a specific mode from the general response: it is in this framework that the proposed methodology helps selecting the proper contributions. In this chapter, the main points of the IPV based damage assessment methodology are reported: for details the reader is referred to the work by Morgantini and Betti [90].

The proposed VI-based methodology allows us to identify the portion of the frequency domain that needs to be considered for extracting the contribution of single structural modes from the original recorded acceleration time histories. To do so, Finite Impulse Response (FIR) filters can be designed to act as band pass filters for the frequency interval associated with each cluster, both for the training and test datasets: the upper and lower cut-off frequencies of such filters are determined accordingly to the bounds of the identified intervals $[\mu_{tr,k_{tr}} - \sigma_{tr,k_{tr}} \quad \mu_{tr,k_{tr}} + \sigma_{tr,k_{tr}}]$ and $[\mu_{te,k_{te}} - \sigma_{te,k_{te}} \quad \mu_{te,k_{te}} + \sigma_{te,k_{te}}]$. It is noteworthy that the order of these FIR filters has to be set sufficiently high to emphasize the spectral content at the center of

these intervals and drastically reduce any other energy contribution, from outside this range, to the original acceleration response time histories.

As a result of the proposed VI approach, each k_{te} clusters is coupled to one of the K'_{tr} clusters so that each interval $[\mu_{te,k_{te}} - \sigma_{te,k_{te}} \quad \mu_{te,k_{te}} + \sigma_{te,k_{te}}]$ is coupled to one of the intervals $[\mu_{tr,k_{tr}} - \sigma_{tr,k_{tr}} \quad \mu_{tr,k_{tr}} + \sigma_{tr,k_{tr}}]$. By applying FIR filters on these two coupled intervals, it is possible to isolate the energy contribution of the r' structural mode from the training and test datasets and convert them into the time domain. At this point, these two newly generated time histories can be used into the IPV methodology.

In this process, a critical point is represented by the selection of the mode r' to be considered. It is known that, in dynamic analyses, structural modes related to lower frequencies offer many advantages over higher frequency modes: they present a lower number of nodes, they have a better resolution and are generally more easily excited [90]. For these reasons, it is convenient to consider a structural mode r' associated with the k_{te} cluster, successfully coupled the k_{tr} cluster, with the lowest frequency interval. This restricts the entire analysis to just one of the identified couples k_{te} and k_{tr}.

The IPV based damage assessment algorithm requires the definition of a reference point w which can arbitrarily selected among the monitored N_{DOF} structural locations ($w \in 1,\ldots,N_{DOF}$).

In order to obtain the cross-correlation vectors for the training and test datasets, let's first consider the training dataset. For each of the N_{ob}^{tr} observations, a number of N_{DOF} acceleration response time histories generated by considering exclusively the r' mode can be obtained by using the filtering procedure described above. The cross-correlation for zero time lag between each of these time histories and the one associated with the reference point w yields a total of N_{DOF} values which can be gathered into a cross-correlation vector $\{R_{w,r'}^{tr,(i)}(0)\} \in \mathbb{R}^{N_{DOF}\times 1}$. The vector $\{R_{w,r'}^{tr,(i)}(0)\}$ is the cross-correlation vector for the i^{th} observation ($i = 1,\ldots,N_{ob}^{tr}$) of the training dataset (tr) obtained by considering a reference point w and the structural mode r'. Each of these $\{R_{w,r'}^{tr,(i)}(0)\}$ vectors can be normalized by its 2-norm to provide a normalized vector $\{\hat{R}_{w,r'}^{tr,(i)}(0)\}$ and it is then possible to represent the average of these N_{ob}^{tr} normalized vectors as $\{\hat{R}_{w,r'}^{tr}(0)\} = \frac{1}{N_{ob}^{tr}}\sum_{i=1}^{N_{ob}^{tr}}\{\hat{R}_{w,r'}^{tr,(i)}(0)\}$.

According to the theory in [morgan], the vector $\{\hat{R}_{w,r'}^{tr}(0)\}$ can be expressed as:

$$\{\hat{R}_{w,r'}^{tr}(0)\} = \text{sign}(\bar{\Phi}_{w,r'}^{tr})\{\bar{\Phi}_{r'}^{tr}\} \tag{4.59}$$

where $\bar{\Phi}_{r'}^{tr}$ represents the r' normalized mode shape and $\text{sign}(\bar{\Phi}_{w,r'}^{tr})$ is the sign of the normalized

mode r' at the w location for the structural conditions corresponding to the training dataset.

Following the same procedure for the test dataset allows us to determine a normalized vector $\{\hat{R}^{te}_{w,r'}(0)\}$:

$$\{\hat{R}^{te}_{w,r'}(0)\} = \text{sign}(\bar{\Phi}^{te}_{w,r'})\{\bar{\Phi}^{te}_{r'}\} \tag{4.60}$$

where $\bar{\Phi}^{te}_{r'}$ and $\text{sign}(\bar{\Phi}^{te}_{w,r'})$ are the corresponding quantities as in Equation (4.59) for the structural conditions represented in the test dataset.

In order to assess the evolution of the r' mode shape of the system passing from an undamaged condition ($\{\bar{\Phi}^{tr}_{r'}\}$) to an unknown (potentially damaged) condition ($\{\bar{\Phi}^{te}_{r'}\}$), a damage index vector $\{D_{w,r'}\}$ can be defined as [morgan]:

$$\{D_{w,r'}\} = \{\hat{R}^{te}_{w,r'}(0)\} - \{\hat{R}^{tr}_{w,r'}(0)\} = sign(\bar{\Phi}^{te}_{w,r'}) \begin{pmatrix} \bar{\Phi}^{te}_{1,r'} \\ \bar{\Phi}^{te}_{2,r'} \\ \vdots \\ \bar{\Phi}^{te}_{N_{DOF},r'} \end{pmatrix} - sign(\bar{\Phi}^{tr}_{wr'}) \begin{pmatrix} \bar{\Phi}^{tr}_{1,r'} \\ \bar{\Phi}^{tr}_{2,r'} \\ \vdots \\ \bar{\Phi}^{tr}_{N_{DOF},r'} \end{pmatrix} \tag{4.61}$$

where $\{D_{w,r'}\} \in \mathbb{R}^{N_{DOF}\times 1}$ indicates the damage index for the mode r' with a reference location at point w and $\{\hat{R}^{te}_{w,r'}(0)\}$ and $\{\hat{R}^{tr}_{w,r'}(0)\}$ are the inner product vectors respectively for potentially damaged (test) and undamaged (training) state.

A local abrupt change of the nomalized modes due to the local damage [76–80] will have a remarkable effect on the damage index vector $\{D_{w,r'}\}$. On the bases of the numerical values of the elements of the damage index vector, it is possible to define a *local damage index vector* and, based on its statistics (mean and standard deviation), to set an upper and lower bound for the values of the elements inside of the local damage index [71–73]. The presence of elements in the local damage index vector which are above the upper threshold or below the lower threshold will confirm the occurrence and location of a local damage. The definition of the local damage index vector and of the thresholds adopted in this chapter is reported in [90]. For details the reader is referred to the work by Morgantini and Betti [90].

4.5. Numerical example

In this section, two numerical simulations (8 and 100 DOF systems) and an experimental test (4 DOF system) are performed in order to validate the proposed damage assessment strategy.

One of the goals of the first numerical simulation is that of providing the numerical expression of the objective function $\mathscr{L}$ over the iterations of the VI clustering algorithm: the multiplication of the average spectrum by the filter bank is performed until the spectrum, or at least its first natural frequencies, is represented by less than 60 points. Hence $N = 60$ is an acceptable value aiming to compute the term $\binom{N}{x_i}$ in Equation (4.55). The restriction of containing at least the first natural frequencies in the first $N \leq 60$ points is imposed due to the fact that, in a civil engineering framework, the lower frequencies are generally the ones with the best resolution in the spectrum and the ones whose investigation through IPV damage assessment algorithm leads to the best performance in terms of damage location identification. [90].

In the second numerical simulation (100-DOFs), the convergence of the VI clustering algorithm will be assessed by computation of the clusters at every iteration. Thus, N is not constrained to assume values larger than 60 so to avoid an excessive (for this particular case) downsampling of the average spectrum. If the assignment of the n observations x_i to their clusters doesn't change over 10 consecutive iterations, the VI algorithm is claimed to converge and the clusters are extracted. Anyway, also in this case the triangular filter bank is used once to take benefit of its 'smoothing' effect on the average spectrum which leads average spectrum to be more likely represented by a mixture of Binomial distributions.

Finally, the experimental test confirms the validity of the presented damage assessment strategy based on an empirical datasets.

4.5.1. 8-DOF shear-type model

The numerical simulation is performed on an 8-DOF shear-type model. The structural parameters have been set to simulate the structure in undamaged conditions. In its baseline conditions, the system is characterized by springs of stiffness $k_i = 25000$ N/m, and each mass is equal to $m_i = 1$ kg for $i = 1, 2, \ldots, 8$. The frame is supposed to be affected by modal damping with a damping factor of $\xi_i = 1\%$ for each of the 8 vibration modes. The excitation source is applied horizontally on the 8-DOF model via zero-order-hold (ZOH) with a sampling period of 0.01 seconds. The force is a zero-mean Gaussian signals with standard deviation $\sigma = 1$ N applied to the top level of the mock up. The input/output time history is 30 seconds long. The val-

ues of the stiffness elements k_i are subject to a variation sampled from a uniform distribution in the range $\pm 2\%$ to account for environmental conditions influencing the dynamic structural response. In its damaged state, the damage has been simulated through a 20% drop of the stiffness of the spring element connecting the 6^{th} and 7^{th} DOF (counting from the bottom). All the other structural parameters, the external conditions and the nature of the excitation source remains the same as in the baseline condition. The training dataset (undamaged conditions) consists in $N_{ob}^{tr} = 100$ instances each counting $N_{DOF} = 8$ acceleration response time histories. The test dataset (damaged) consists in a set of $N_{ob}^{te} = 50$ instances.

4.5.2. Objective function validation

A representation of the shear-type model is provided in Fig. 4.4 (a). The dynamic behaviour of the structure in its baseline condition is depicted by the power spectrum of the acceleration response time histories. The energy content represented by the spectrum at DOF 1 is reported in Fig. 4.4 (b) as well as in the spectrogram Fig. 4.4 (c).

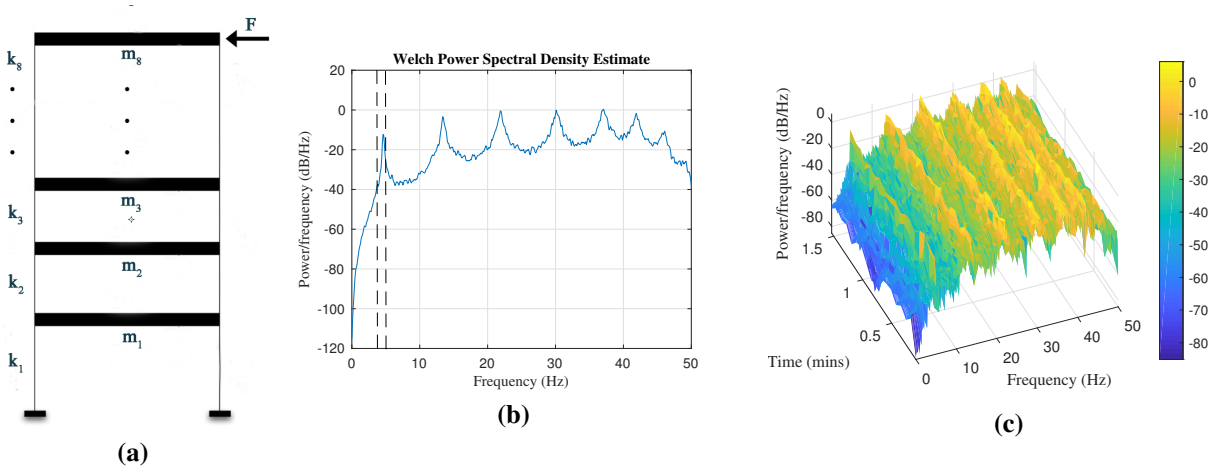
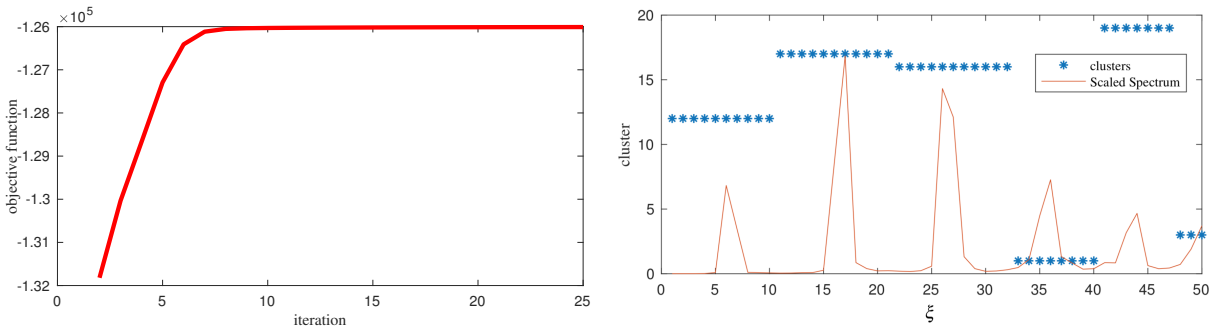


Fig. 4.4. Mock up (a). Spectral analysis for the 1^{st} floor. Power Spectrum (b) and Spectrogram (c) in baseline condition.

A total of 25 iterations for the VI algorithm are performed in order to let the objective function $\mathscr{L}$ converge and cluster the downsampled power spectrum. According to the theory [95], $\mathscr{L}$ is negative and increases monotonically (Fig. 4.5 (a)). The downsampled (scaled) power spectrum is reported in Fig. 4.5 (b) as well as the clusters assigned to each of its samples. Despite the number of clusters has been defined as $K = 20$, only $K'_{tr} = 6$ of them are assigned. The values indicating the identified clusters fully depends on the random initialization of the parameters for the probability distributions in the VI algorithm.



(a) **(b)**

Fig. 4.5. Objective function (a). Averaged and scaled power spectrum, observations assignment c_i (b).

The downsampled spectrum is then upsampled, scaled and represented in the frequency domain. Fig. 4.6 shows the normalized average spectrum $\boldsymbol{S}_{tr}^{s}$ as defined in Equation (4.57). The frequency domain is divided into 6 areas, each indicated by a different cluster. The clusters k_{tr} are reordered from 1 to 6 according to Fig. 4.6.

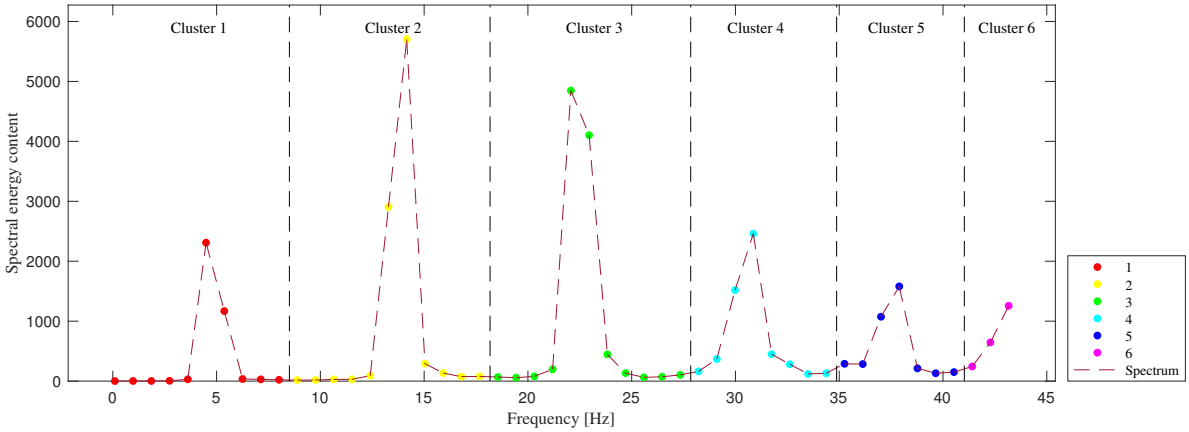
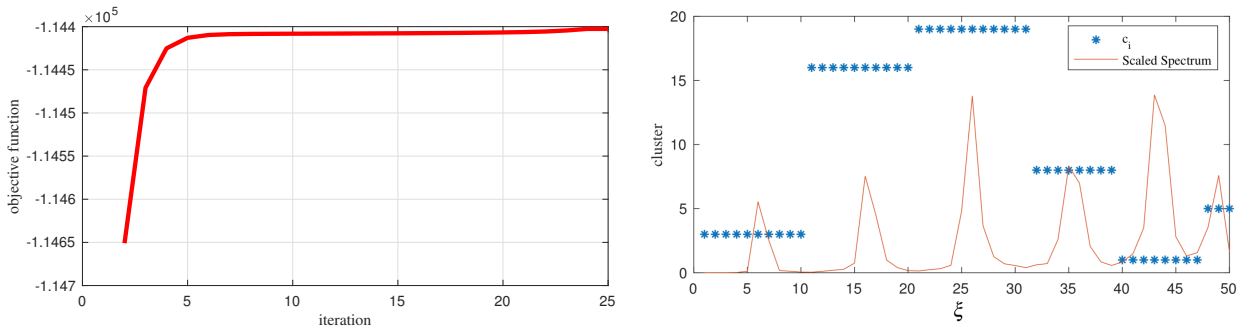


Fig. 4.6. Train data: clustered energy content.

Once the intervals of interest $[\mu_{tr,k_{tr}} - \sigma_{tr,k_{tr}} \quad \mu_{tr,k_{tr}} + \sigma_{tr,k_{tr}}]$ for any $k_{tr} = 1,\ldots,K'_{tr}$ are obtained, the feature vectors $\mathscr{G}_{tr,k_{tr}}$ for the classification task are computed for each of the $N_{ob}^{tr} = 100$ observations in the training dataset. The ANN is trained based on the $N_{ob} \times K' = 600$ feature vectors $\mathscr{G}_{tr,k_{tr}}$.

As a new dataset of acceleration response time histories is generated from the system in damaged conditions (unknown a priori), the clustering procedure is performed. The objective function $\mathscr{L}$ for this case is shown in Fig. 4.7 (a) and the downsampled (scaled) power spectrum and the cluster assigned to each of its samples in Fig. 4.7 (b).



(a) **(b)**

Fig. 4.7. Objective function (a). Averaged and scaled power spectrum, observations assignment c_i (b).

By upsampling and scaling the downsampled spectrum the normalized average spectrum $\boldsymbol{S}^s_{te}$ is computed and shown in Fig. 4.8. Again, the clusters k_{te} are reordered from 1 to 6.

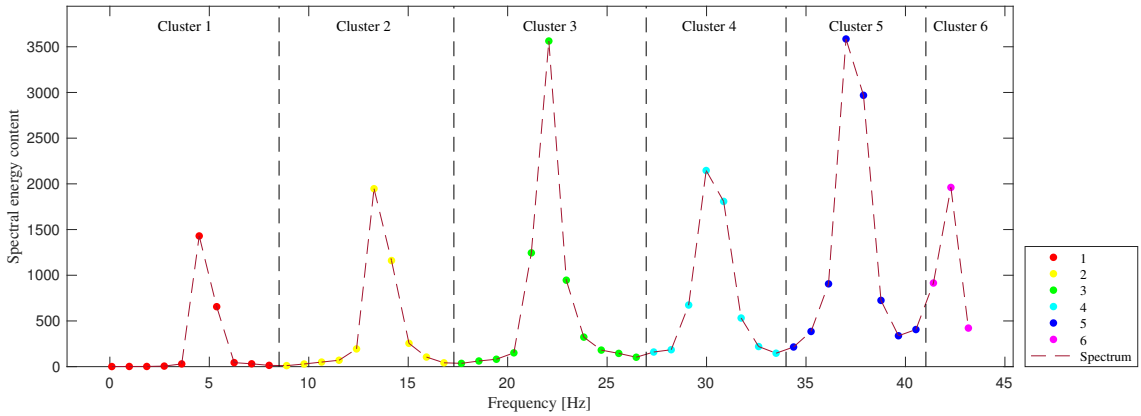


Fig. 4.8. Test data: clustered energy content.

The feature vectors $\mathcal{G}_{tr,k_{tr}}$ for classification purpose are computed for the $N^{te}_{ob} = 50$ observations. These vectors are used as inputs to the trained ANN and the results are gathered into a confusion matrix shown in Fig. 4.9. Such matrix reports the association of the classes in damaged condition (y-axis) with the classes defined in undamaged conditions (x-axis). The non-null elements in the matrix are positioned along the diagonal providing a pairwise association of the damaged and undamaged clusters.

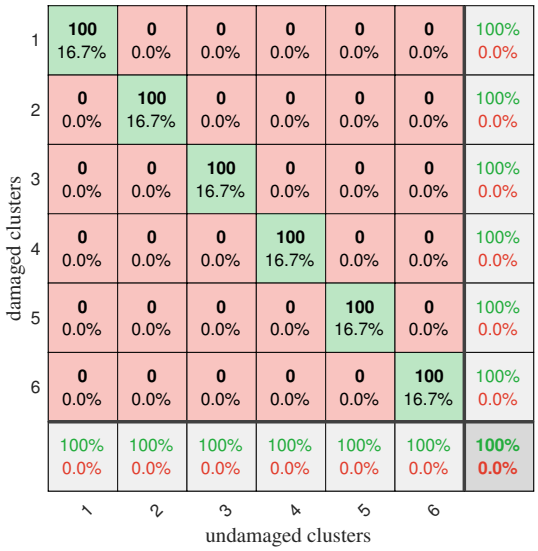
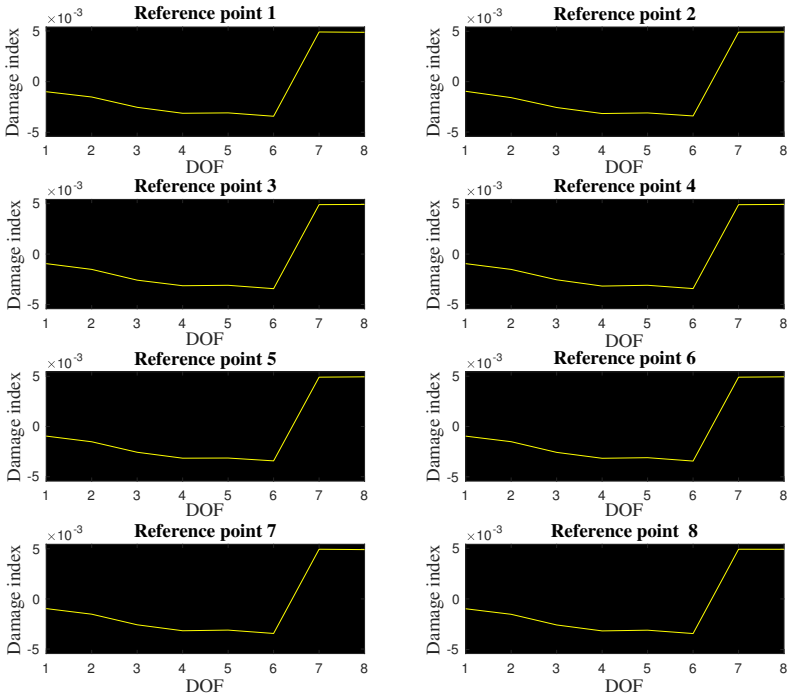
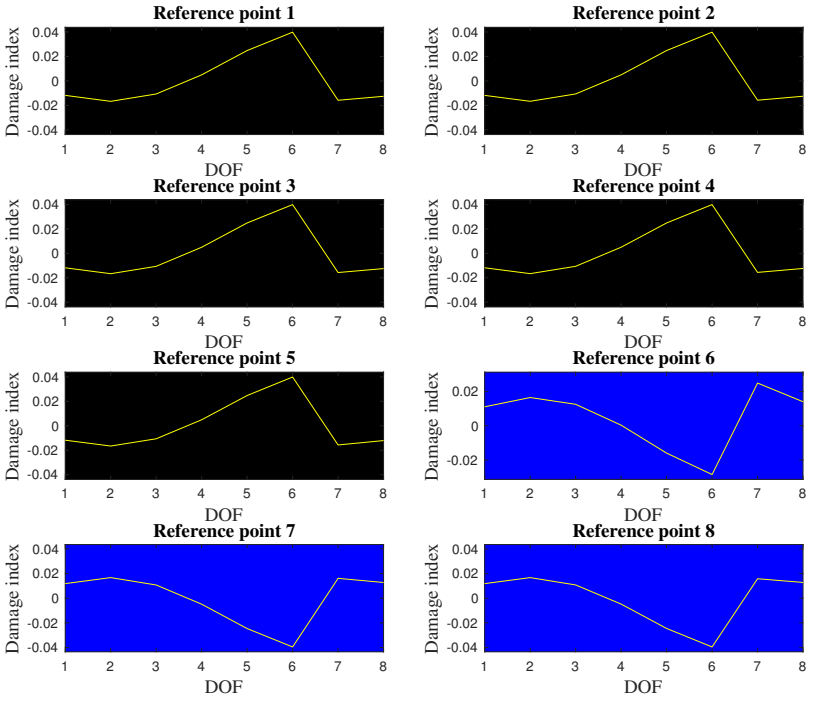


Fig. 4.9. Confusion Matrix.

As mentioned, structural modes related to lower frequencies offer many advantages over the modes related to higher frequencies [90]. Generally they present the lower number of nodes, they have a better resolution and are more easily excited. For these reasons, in this numerical test we focus on the energy contribution of the modes associated to the first ($r' = 1$) and second ($r' = 2$) clusters. For the couple $k_{tr} = 1$ and $k_{te} = 1$, the intervals $[\mu_{tr,1} - \sigma_{tr,1} \quad \mu_{tr,1} + \sigma_{tr,1}]$ and $[\mu_{te,1} - \sigma_{te,1} \quad \mu_{te,1} + \sigma_{te,1}]$ correspond respectively to the $[4.20 \quad 5.37]$ Hz and $[4.10 \quad 5.37]$ Hz. The FIR filters are designed according to these intervals and the vector $\{\hat{R}^{tr}_{w,r'}(0)\}$ is computed for any of the 100 samples in the training dataset. The average of these vectors is considered. Analogously the vector $\{\hat{R}^{te}_{w,r'}(0)\}$ is computed for any of the $N^{te}_{ob} = 50$ samples in the test dataset to obtain the average vector. The damage index vector $\{D_{w,r'}\}$ is then calculated for any reference point w and shown in Fig. 4.10 (a).



(a)



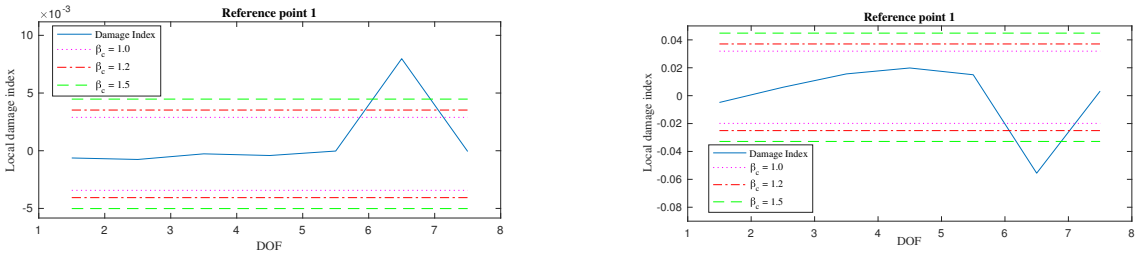
(b)

Fig. 4.10. Damage index vector for different reference ponts w, $r' = 1$ (a) and $r' = 2$ (b).

According to Fig. 4.10, the damage indexes for different reference points w are consistent

to each other proving that choosing different w doesn't affect the result of the analysis. We can arbitrarily consider one of the damage index vectors computed and obtain the local damage index vector $\{D'_{IPV,w,r'}\}$ (first derivative of $\{D_{IPV,w,r'}\}$, jump discontinuity). The local damage index vector showed in Fig. 4.11 (a) accounts for the reference point $w = 1$ and correctly identifies the presence and location of the damage for any considered threshold value.

Let's now focus on the modal energy content associated to the second cluster ($r' = 2$). For the couple $k_{tr} = 2$ and $k_{te} = 2$, the intervals $[\mu_{tr,2} - \sigma_{tr,2} \quad \mu_{tr,2} + \sigma_{tr,2}]$ and $[\mu_{te,2} - \sigma_{te,2} \quad \mu_{te,2} + \sigma_{te,2}]$ correspond respectively to the intervals $[13.09 \quad 14.65]$ Hz and $[12.59 \quad 14.55]$ Hz. By following a procedure analogous to the one used for the first couple of clusters, the damage index vector $\{D_{w,r'}\}$ is calculated for any reference point w and shown in Fig. 4.10 (b). The local damage index vector for $w = 1$ is reported in Fig. 4.11 (b).



(a) **(b)**

Fig. 4.11. Local damage index, $r' = 1$ (a) $r' = 2$ (b).

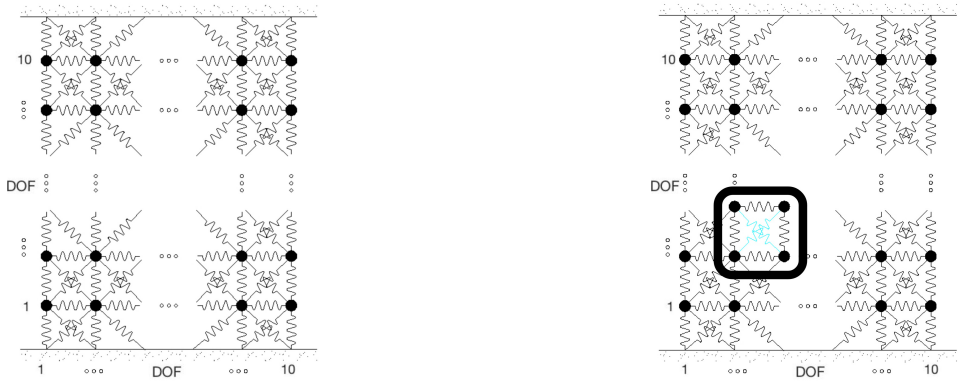
4.5.3. High order systems

This section extends the investigation about the applicability of the IPV-based damage assessment method to a more complex system, (e.g. a plate) represented by a two-dimensional frame (Fig. 4.12 (a)). Taking recourse to the presented damage assessment algorithm is recognized as ideal to effectively account for the extremely high order of the system. Central in the implementation of system identification algorithms (SSI, O^{3}KID) is the definition of some parameters depending on the system order. Since an 'a priori' knowledge of the system order is never provided, the estimation of these parameters may result in a challenging task. Nonetheless, the implementation of large values for those parameters leads to a huge computational effort and relatively high numerical uncertainties in the estimated modal parameters. The aim of this analysis is that of showing the presented damage assessment algorithm as a valid alternative to the conventional system identification algorithms in those cases in which the system order is relatively large.

4.5.4. 100-DOFs

In its baseline conditions, the 100-DOF structure is a 2-D square grid of 10×10 lumped masses of 1 kg each connected by spring elements placed horizontally, vertically and diagonally, each one having stiffness of 1000 N/m, 900 N/m and 800 N/m respectively. The modal damping has been set to $\xi = 1\%$ for all vibration modes. The structure is doubly-fixed at the top and at the bottom and set of excitation forces act perpendicular to the plane of structure on each of the masses. Using the assumption of zero-order-hold (ZOH) with a time sampling of 0.01 seconds, these forces are zero-mean Gaussian signals (uncorrelated to each other) with standard deviation $\gamma = 10$ N providing input/output time histories 100 second long. The acceleration perpendicular to the plane of the structure is recorded at each of the 100 DOFs reproducing a dense sensor network setup applied to the system. Also in this case, the values of the stiffness elements k_i are subject to a variation sampled from a uniform distribution in the range $\pm 2\%$.

In its 'damaged' state the structure presents a variation in two of the stiffness elements. Hence, the local damage is simulated through a drop in stiffness of 25% for two of the diagonal springs as shown in Figure 4.12 (b).



(a) **(b)**

Fig. 4.12. 100-DOF system: undamaged (a), damaged (b).

4.5.5. Converging clusters validation

As mentioned, the extraction of the modal parameters of high order structures may result in an extremely challenging task. Hence, the conventional system identification algorithms like O^3KID rely on recurrent regression models whose order has to be set much higher than the system order. Likewise, other system identification algorithms (SSI) rely on the projections of

large matrices whose dimensions are difficult to estimate, but are theoretically required to be way larger than the system order. Basically, these considerations let problems like the computational effort and the uncertainties about the setting parameters come to mind. Hence, because of the uncertainty about the system order, it is usual practise to use the so called 'stabilization diagrams'. Fig. 4.13 represents a stabilization diagram for the baseline 100-DOF system computed through O^{3}KID. In a state-space representation the order of the 100-DOF system is equal to 200. For such a reason, the order of the regression model (linear least squares) has been set to 300. By using the Eigensystem Realization Algorithm (ERA), which is part of the O^{3}KID algorithms, the modal parameters of the system are extracted assuming different system orders. The considered system orders are reported on the vertical axis of the stabilization diagram, whereas the identified natural frequencies of the system are on horizontal axis. For each each of the considered system orders along the vertical axis, the modal parameters are computed. If by increasing the system order some of the modal parameters are repeated a star appears in the diagram. Its position depends on the system order and the identified natural frequency. The stabilization diagram shows also these cases in which modal parameters are partially repeated over the system order:

- if only natural frequency and damping factors are repeated a 'd' appears in the diagram;
- if only natural frequency and modes are repeated, a 'v' appears in the diagram;
- if only natural frequency is repeated, an 'f' appears in the diagram;

Obviously, a tolerance range has been set for each of the modal parameters: 1% for the natural frequencies, 10% for the damping ratios, 5% for the Modal Assurance Criterion (MAC numbers).

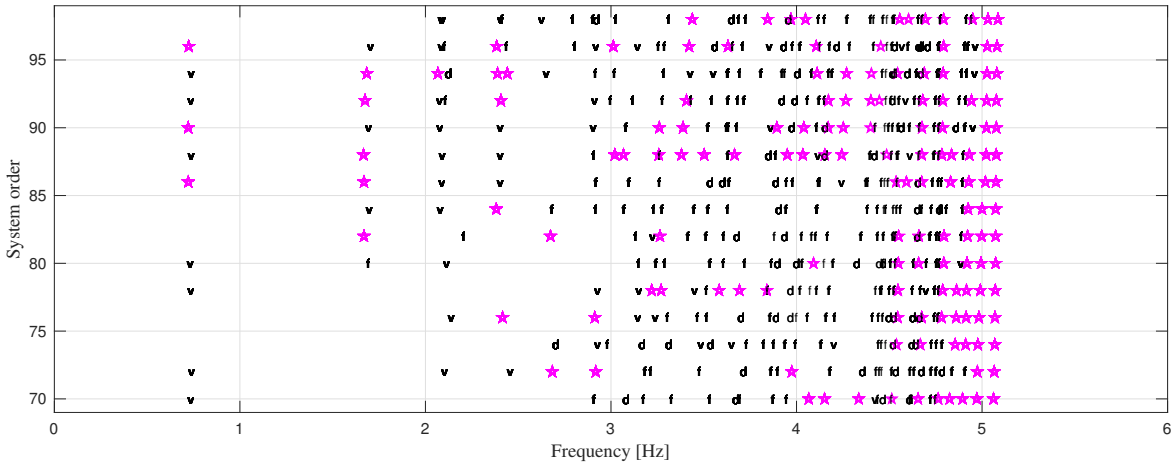


Fig. 4.13. Stabilization diagram, baseline configuration (undamaged), O^{3}KID.

By looking at the stabilization diagram, we need to consider high orders of the system to let the natural frequencies appear, and sometimes, for some system orders, they are not even displayed. In this section we show that the damage assessment algorithm presented represents a valid alternative to the conventional system identification methods for structural damage assessment purpose.

A total of $N_{ob} = 50$ observations are collected for the system in its baseline state. The average spectrum is computed and the triangular filter bank applied once to the average spectrum. The clustering algorithm is applied in the range [0 3] Hz. The resulting clustered average spectrum is shown in Fig. 4.14. A total of $K'_{tr} = 7$ clusters are identified. The ANN classifier is trained on the 50 generated features $\mathscr{G}_{tr,k_{tr}} \in \mathbb{R}^{100\times 1}$ for $k_{tr} = 1,\ldots,7$ (350 in total).

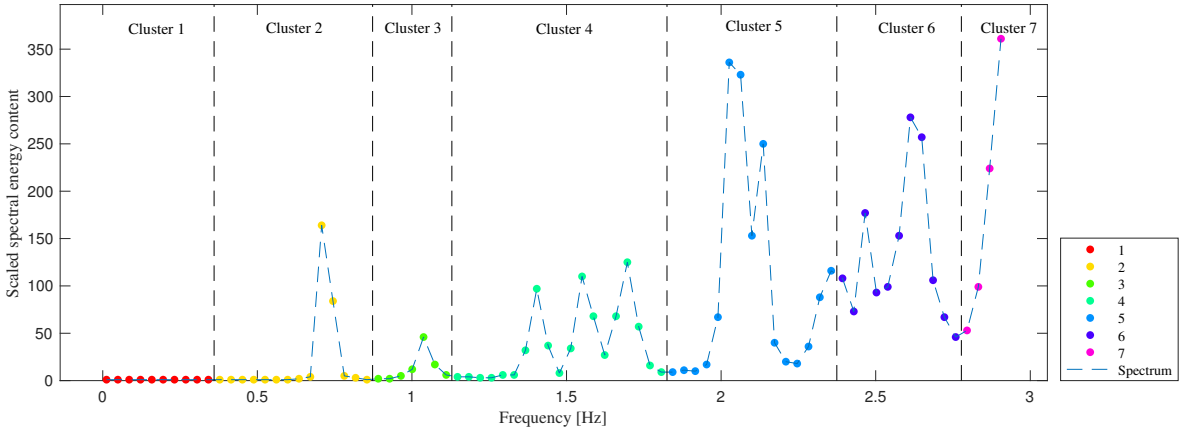


Fig. 4.14. Clustered average spectrum, baseline condition.

After the ANN is trained, the clustering procedure is performed on the 50 test observations (structure in 'damaged' state). The clustered average spectrum is shown in Fig. 4.15. 10 clusters are identified.

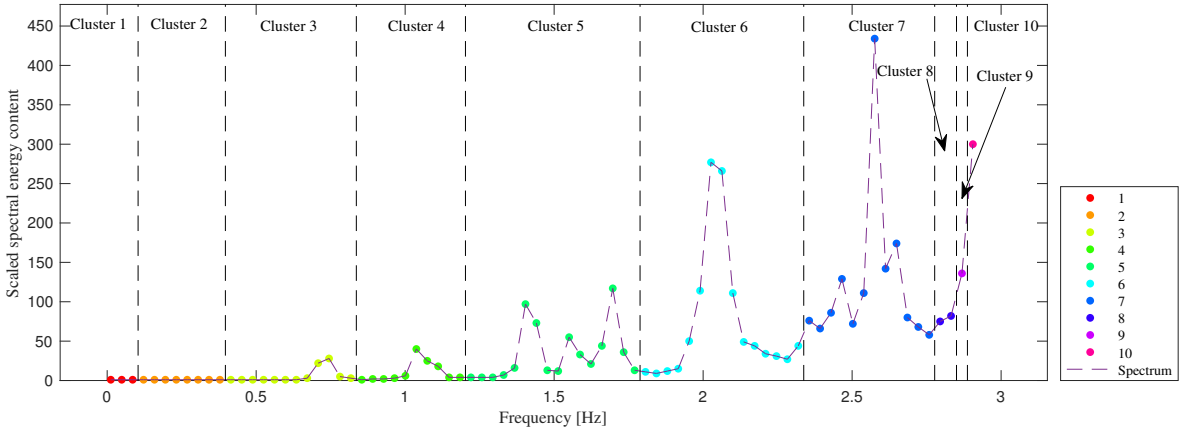
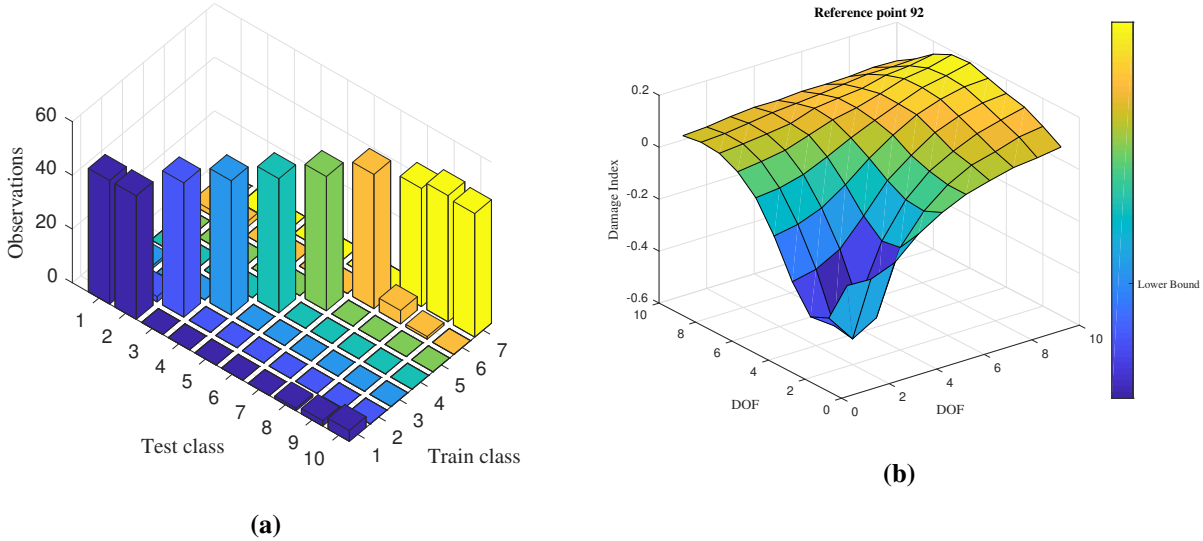


Fig. 4.15. Clustered average spectrum, damaged condition.

The ANN classifier is tested on the 50 observed features $\mathscr{G}_{te,k_{te}} \in \mathbb{R}^{100\times 1}$ for each $k_{te} = 1,\ldots,10$. For each test class k_{te} the 50 observations are classified into one of the $K'_{tr} = 7$ train class. The pairwise clusters coupling is imposed based on the largest number of observations which, for a specific test class are associated to a train classes. Fig. 4.16 (a) shows a 3-D graph: 50 test observations for each test class are classified into a train class. The observations belonging to test classes $k_{te} = 1,2$ are all classified into the $k_{tr} = 1$ train class, but their MAC number is below the threshold 0.9 (Table 4.1). The observations belonging to test classes $k_{te} = 3$ are all classified into the $k_{tr} = 2$ train class. By setting this coupling, we can proceed with the algorithm to develop the damage index shown in Fig. 4.16 (b). The reference point for the computation of the IPV is arbitrarily set to $w = 92$. Anyway, results are consistent to each otherfor any choice of the reference point. Based on the statistics, the upper and lower bounds for the damage index vector are respectively 0.0334 and -0.0356.



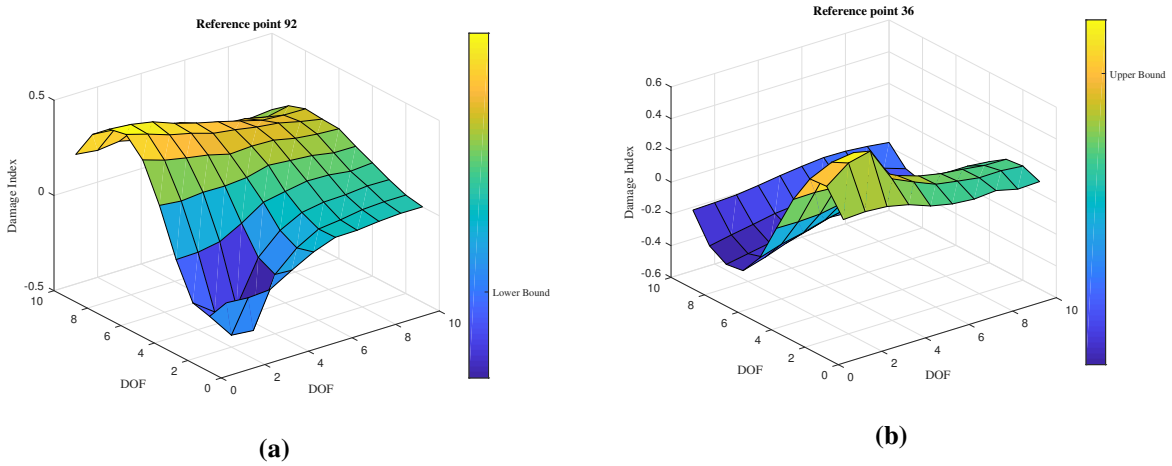
(a) (b)

Fig. 4.16. Test data classification (a). 2-D damage index, Train class: 2, Test Class: 3 (b).

Test - Train class	1 - 1	2 - 1	3 - 2	4 - 3	5 - 4	6 - 5	7 - 6	8 - 7	9 - 7	10 - 7
MAC number	0.644	0.661	0.998	0.997	0.993	0.991	0.988	0.954	0.954	0.901

Table 4.1. MAC number for pairwise coupled clusters.

By looking at Fig. 4.16 (a), we can also notice that the observations belonging to test classes $k_{te} = 4$ are all classified into the $k_{tr} = 3$ train class. By considering that coupling, the damage assessment algorithm is applied. The IPV has been computer for two different reference points arbitrarily picked. The choice of the reference point $w = 92$ yields a damage index (Fig. 4.17 (a)) that is the mirrored image of the one provided by setting $w = 36$ (b). The upper and lower bounds are respectively 0.4390 and -0.2312 for $w = 92$ and 0.3599 and -0.4870 for $w = 36$.



(a) **(b)**

Fig. 4.17. 2-D damage index: Train class: 3, Test Class: 4. Reference point: $w = 92$ (a), $w = 36$ (b).

4.6. Experimental test: LANL 3-DOF shear-type

The proposed automatized damage assessment algorithm has also been tested by considering the dynamic response of a 3-DOF shear-type system shown in Fig. 4.18. Data have been provided by the Engineering Institute (EI) at Los Alamos National Laboratory (LANL) [66]. The system consists of four aluminum columns ($17.7 \times 2.5 \times 0.6$ cm) connected at the top and bottom to aluminum plates ($30.5 \times 30.5 \times 2.5$ cm) [69], forming a structure consisting of 3 floors and a sliding base. An electromagnetic shaker that acts at the center line of the base floor of the structure provides the excitation to the dynamic system. The structure and the shaker are fixed on a base plate ($76.2 \times 30.5 \times 2.5$ cm). Four accelerometers with a nominal sensitivity of 1000 mV/g are positioned at the center of the side of each floor at the opposite side from shaker to measure the response of each plate. The random excitation is band limited in the range of 20-150 Hz to avoid rigid body modes of the structure. Even if the structure was initially supposed to behave linearly, a certain non-linear tendency due to the sliding rails has been noted [83].

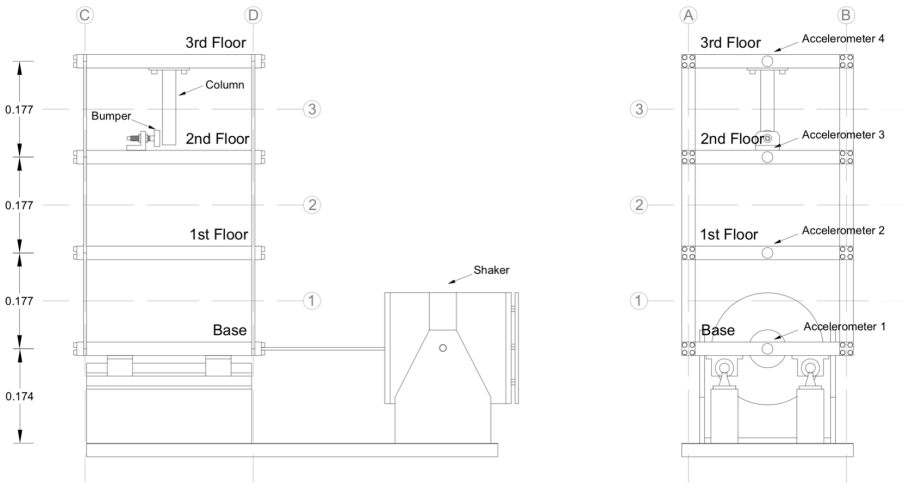


Fig. 4.18. LANL structure.

The LANL database supplies data about force and accelerations recorded while the structure is in different conditions. When the standard 3-DOF model previously described is considered, the structure is claimed to be in "baseline condition". Three damage conditions have been investigated in such a test setup, with damage imposed through stiffness reduction of the columns connecting the floors. The damage scenarios considered in this dataset are the following:

- 50% stiffness reduction between floors 1-2
- 50% stiffness reduction between floors 2-3
- 50% stiffness reduction between floors 3-4

These damage scenarios have been already analyzed through a damage assessment algorithm in [90]. It has been shown that the third damage scenario presents an unreliable damage index for the reference point $w = 3$ that, once the damage occurs becomes a node for the first structural mode. In this chapter, every step of the damage assessment algorithm is reported exclusively for that specific case.

For each state of the structure (damaged or not) a total of $N_{ob} = 9$ realizations of the dynamic response time histories are considered. The acceleration response time histories consist in 8192 time steps sampled at 0.0031 sec. Since the excitation source has been filtered in the interval [0 20] Hz and the frequency content on the right side of 20 may be affected by the filter, we will consider the spectral content in the frequency range [24 77] Hz. The clustering procedure for the system led to the clustered average spectrum reported in Fig. 4.19. A total of $K'_{tr} = 5$ clusters are identified. The ANN classifier is trained on the 9 generated features $\mathscr{G}_{tr,k_{tr}} \in \mathbb{R}^{4\times 1}$ for $k_{tr} = 1,\ldots,5$ (45 in total). It is noteworthy to specify that, in this particular case, the number

of observations in the training dataset may be poor, anyway results provided by the ANN are validated by the MAC numbers.

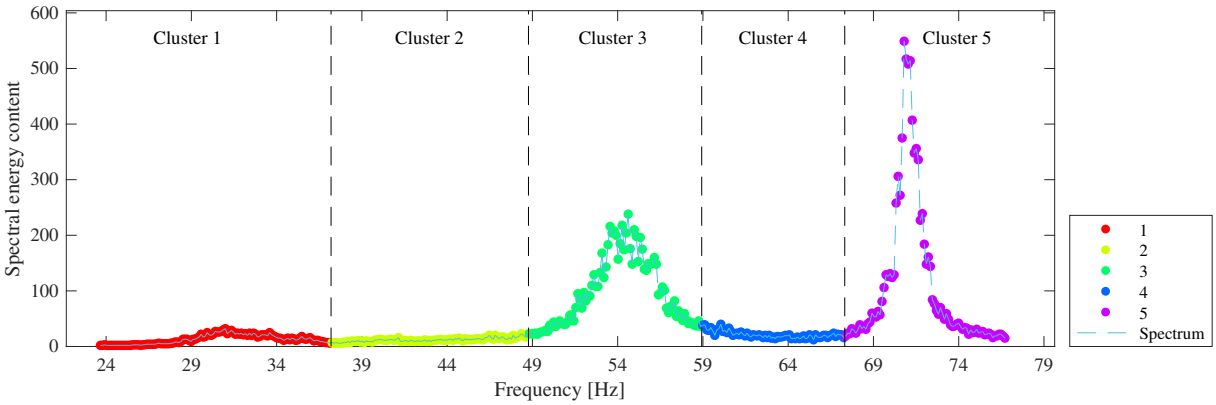


Fig. 4.19. Clustered average spectrum, undamaged condition.

Once the ANN is trained, the clustering procedure is performed on the 9 test observations (structure in 'damaged' state). The clustered average spectrum is shown in Fig. 4.20. $K'_{te} = 6$ clusters are identified.

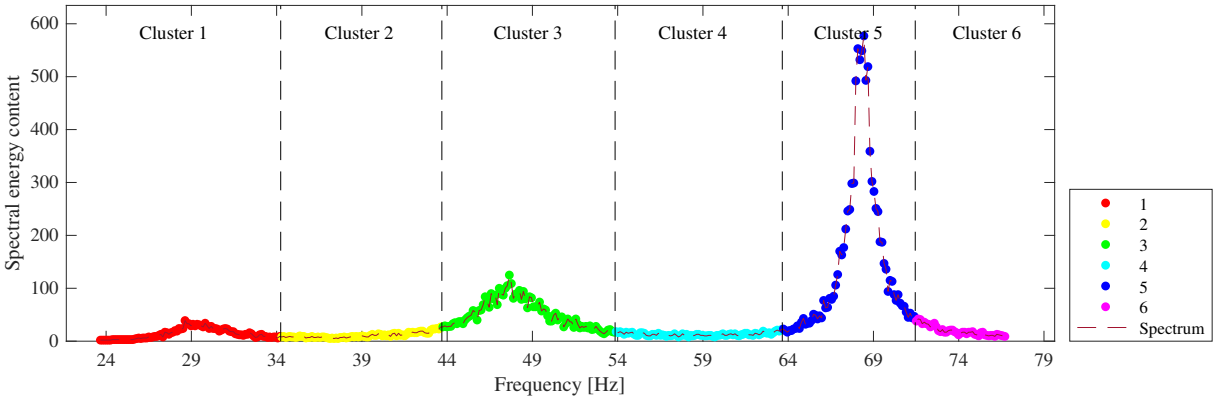


Fig. 4.20. Clustered average spectrum, damaged condition.

The ANN classifier is tested on the 9 observed features $\mathscr{G}_{te,k_{te}} \in \mathbb{R}^{4\times 1}$ for each $k_{te} = 1,\ldots,6$. For each test class k_{te}, the 9 observations are classified into one of the $K'_{tr} = 5$ train class. The pairwise clusters coupling is imposed based on the largest number of observations which, for a specific test class are associated to a train classes. Fig. 4.21 shows the 3-D bar graph: 9 test observations for each test class are classified into a train class. The observations belonging to test classes $k_{te} = 1$ are all classified into the $k_{tr} = 1$ train class and the MAC number is reported in Table 4.2. By setting this coupling, we can proceed with the algorithm to develop the damage index vector shown in Fig. 4.22.

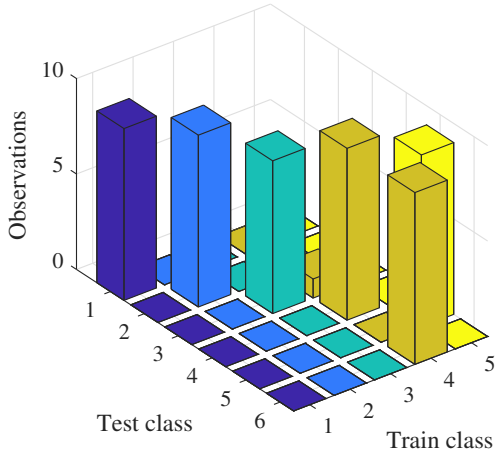


Fig. 4.21. Test data classification.

Test - Train class	1 - 1	2 - 2	3 - 3	4 - 4	5 - 5	6 - 5
MAC number	0.989	0.934	0.938	0.907	0.973	0.848

Table 4.2. MAC number for pairwise coupled clusters.

As mentioned, the damage index vector for the reference point $w = 3$ is consistent with the other three ($w = 1,2,4$) due to the fact that $w = 3$ is a node for the damaged structure. For the other reference points, results are consistent to each other.

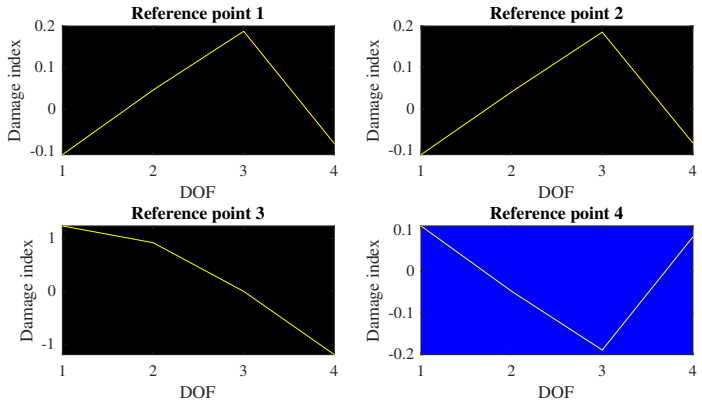


Fig. 4.22. Damage index vectors for different reference points $w = 1,2,3,4$.

Finally, the local damage index is computed for the reference point $w = 4$ and shown in Fig. 4.23. The value of the local damage index at 3.5 is above the upper thresholds confirming the presence of a local damage between DOFs 3 and 4.

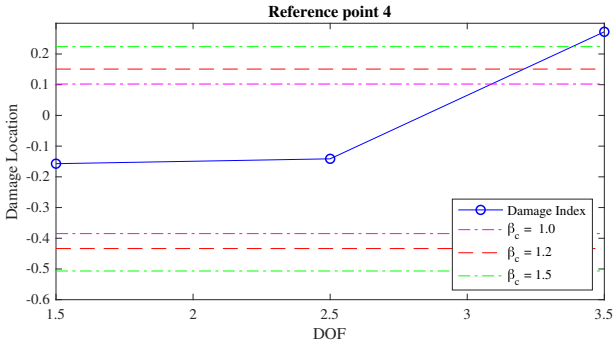


Fig. 4.23. Local damage index vector for reference point $w = 4$.

4.7. Conclusions

This chapter presents a damage assessment method based on two key features. The first is the implementation of a Bayesian clustering method (VI) that aims to identify the portions of the frequency domain containing the energy contribution of a specific structural mode to the spectral dynamic response. The second key feature is represented by a classification algorithm. The objective is to label the energy contributions given by similar structural modes under the same class, so that the pattern recognition analysis can be performed for a properly selected class. The fist and the second steps are combined together in order to opportunely extract the damage sensitive features. In a statistical pattern recognition framework, the damage index is developed by comparing the cross-correlation vector (IPV) in undamaged state with the one in unknown state. The damage index provides a reliable indication about the location of the damage. Of course, one more time it is worth to recall that the whole algorithm relies on restrictive assumptions.

Numerical simulation and experimental tests pointed out the effectiveness of the method, which lends itself especially well to the analysis of shear-type systems, but also to generic structures [90].

Chapter 5

5. A Bayesian approach to the Output-only observer/Kalman filter identification

5.1. Introduction

In this chapter, a Bayesian regression algorithm is presented and integrated into the Output-only Observer/Kalman filter IDentification (O^3KID), an output-only system identification algorithm, so to provide a more accurate representation of the system in the presence of measurement noise. This leads to a more accurate estimation of the system's natural frequencies and mode shapes and of the model's order in the case when signals of the response are affected by noise substantially improving the performance of damage assessment algorithms based on modal characteristics.

State space deterministic models consist of two sets of equations: the state equation, that describes the evolution over time of state of the system, and the output equation, in which the desired output is related to the state of the system. In order to account for the uncertainties due to process and measurement noise, instead of using a deterministic state space model of the system, a state observer is used. The state observer is a system that estimates the internal state of a state space system that can rely on either an input-output or output-only dataset.

In the first section of this chapter, the state space formulation of a state observer is reviewed. Previous studies proved that the estimation of the state observer matrices provided by the Ordinary Least Squares (OLS) solution allows us to obtain an estimation of the Markov parameters of a Kalman filter [83]. Hence, the Gauss-Markov theorem states that OLS algorithm is the Best Linear Unbiased Estimator (BLUE) and, considering that the Kalman filter is the (only) linear system that minimizes the expected value of the residuals squared, the residuals of the OLS solution are the Kalman output residuals. Such a statement is valid in the case in which the dataset contains extremely long time histories of the dynamic response (theoretically infinite time steps) [83]. The presence of external disturbances can lead OLS algorithms to overfit data, especially when the order of the system, and consequently of its state space representation, is large [96]. An alternative to the OLS algorithm is presented by the partial least squares (PLS) [97] but its use also requires extremely large dataset.

A Bayesian regression method can be used to reduce overfitting of the noisy structural response [98, 99], and is used in this chapter to provide an approximation of the Markov pa-

rameters of a Kalman filter. Although the application of Bayesian regression methods leads to remarkable benefits in terms of overfitting, it is important to point out that the solution of the Bayesian regression algorithm provides an *approximation* of the Markov parameters of a Kalman filter. This is because the minimization criteria of a Bayesian regression algorithm relies on the minimization of a Kullback-Leibler divergence, rather than the squared error of the residuals, and so the retrieved Markov parameters do not fit the definition of the Kalman filter.

5.2. State space representation of an input-output system

State space models are one of the most popular mathematical representations of dynamic systems. A discrete-time linear-time-invariant state space model can be expressed as:

$$\boldsymbol{x}(k+1) = \mathbf{A}\boldsymbol{x}(k) + \mathbf{B}\boldsymbol{u}(k) \tag{5.1}$$

$$\boldsymbol{y}(k) = \mathbf{C}\boldsymbol{x}(k) + \mathbf{D}\boldsymbol{u}(k) \tag{5.2}$$

where k is the time step, $\boldsymbol{x} \in \mathbb{R}^{n\times 1}$ is the state vector, $\boldsymbol{u} \in \mathbb{R}^{m\times 1}$ is the input vector, $\boldsymbol{y} \in \mathbb{R}^{q\times 1}$ is the output vector, $\mathbf{A}$ is the system matrix, $\mathbf{B}$ is the input matrix, $\mathbf{D} \in \mathbb{R}^{q\times m}$ is the direct input matrix and $\mathbf{C} \in \mathbb{R}^{q\times n}$ is the output matrix. Equations (5.1) and (5.2) are the state equation and the observation equation respectively. The deterministic model needs to be modified so to account for process and measurement noise, so Equations (5.1) and (5.2) are modified as:

$$\boldsymbol{x}(k+1) = \mathbf{A}\boldsymbol{x}(k) + \mathbf{B}\boldsymbol{u}(k) + \boldsymbol{w}'_p(k) \tag{5.3}$$

$$\boldsymbol{y}(k) = \mathbf{C}\boldsymbol{x}(k) + \mathbf{D}\boldsymbol{u}(k) + \boldsymbol{w}'_m(k) \tag{5.4}$$

where the vectors $\boldsymbol{w}'_p \in \mathbb{R}^{n\times 1}$ and $\boldsymbol{w}'_m \in \mathbb{R}^{q\times 1}$ represent the process and measurement noise. The vectors $\boldsymbol{w}'_p(k)$ and $\boldsymbol{w}'_m(k)$ are commonly assumed to be zero mean, white and uncorrelated with $\boldsymbol{u}(k)$ and $\boldsymbol{y}(k)$. An additional assumption is that the covariance matrices $\mathbf{R}'$ and $\mathbf{Q}'$ of the vectors $\boldsymbol{w}'_p(k)$ and $\boldsymbol{w}'_m(k)$ respectively are time invariant, i.e.constant over time.

5.3. Output-Only Observer/Kalman filter identification

The development of the so called Output-Only Observer/Kalman filter identification (O^3KID) system identification algorithm is exhaustively discussed in [81]. Let's now consider a dataset

of l time steps of acceleration response time histories:

$$\{\boldsymbol{y}(k)\} = \{\boldsymbol{y}(1), \boldsymbol{y}(2), \ldots, \boldsymbol{y}(l)\}. \tag{5.5}$$

measured from the system subjected to an unknown excitation source, the objective is to identify a state space model representing the dynamic of the system. When the input excitation $\boldsymbol{u}(k)$ is unknown, Equations (5.3) and (5.4) can be rewritten as:

$$\begin{aligned} \boldsymbol{x}(k+1) &= \mathbf{A}\boldsymbol{x}(k) + \boldsymbol{w}_p(k) && (5.6) \\ \boldsymbol{y}(k) &= \mathbf{C}\boldsymbol{x}(k) + \boldsymbol{w}_m(k) && (5.7) \end{aligned}$$

where $\boldsymbol{w}_p \in \mathbb{R}^{n\times 1}$ and $\boldsymbol{w}_m \in \mathbb{R}^{q\times 1}$ are zero-mean white stationary processes including not only the original process and measurement noise, but also the effect of the unknown input on the state equation, $\mathbf{B}\boldsymbol{u}(k)$ in Equation (5.1), and on the measurement equation, $\mathbf{D}\boldsymbol{u}(k)$ in Equation (5.2). It is worthy to notice that the presence of the input in both the terms $\boldsymbol{w}_p$ and $\boldsymbol{w}_m$ makes them be correlated. Let's denote the auto-covariance matrices of $\boldsymbol{w}_p$ and $\boldsymbol{w}_m$ as $\mathbf{R}$ and $\mathbf{Q}$ respectively. Let's now consider the steady-state Kalman filter for the system in Equations (5.6) and (5.7):

$$\begin{aligned} \hat{\boldsymbol{x}}(k+1) &= \mathbf{A}\hat{\boldsymbol{x}}(k) + \mathbf{K}(\boldsymbol{y}(k) - \hat{\boldsymbol{y}}(k)) && (5.8) \\ \hat{\boldsymbol{y}}(k) &= \mathbf{C}\hat{\boldsymbol{x}}(k) && (5.9) \end{aligned}$$

where $\hat{\boldsymbol{x}} \in \mathbb{R}^{n\times 1}$ and $\hat{\boldsymbol{y}} \in \mathbb{R}^{q\times 1}$ are respectively the observer state and the output and the matrix $\mathbf{K}$ is the observer gain. By substituting Equation (5.9) into Equation (5.8), Equation (5.10) is obtained. Furthermore, given $\boldsymbol{\varepsilon}(k) = \boldsymbol{y}(k) - \hat{\boldsymbol{y}}(k)$ and recalling Equation (5.9) it is possible to write:

$$\begin{aligned} \hat{\boldsymbol{x}}(k+1) &= \bar{\mathbf{A}}\hat{\boldsymbol{x}}(k) + \mathbf{K}\boldsymbol{y}(k) && (5.10) \\ \boldsymbol{y}(k) &= \mathbf{C}\hat{\boldsymbol{x}}(k) + \boldsymbol{\varepsilon}(k) && (5.11) \end{aligned}$$

where $\bar{\mathbf{A}} = \mathbf{A} - \mathbf{KC}$. It it worthy to note that the observer in Equation (5.10) is a one step-ahead predictor and that the matrices in the system are linear time-invariant matrices. Equation (5.10) can be then expanded providing an expression which can be interpreted as a recursive

model. By propagating Equation (5.10) p steps backward, it is possible to obtain:

$$\hat{\boldsymbol{x}}(k+1) = \bar{\mathbf{A}}^{p}\hat{\boldsymbol{x}}(k-p) + \mathbf{T}\boldsymbol{v}(k) \tag{5.12}$$

$$\tag{5.13}$$

where the vector $\boldsymbol{v}(k)$ and the matrix $\mathbf{T}$ are given by:

$$\boldsymbol{v}(k) = \begin{bmatrix} \boldsymbol{y}(k-1) \\ \boldsymbol{y}(k-2) \\ \vdots \\ \boldsymbol{y}(k-p) \end{bmatrix} \tag{5.14}$$

and

$$\mathbf{T} = \begin{bmatrix} \mathbf{K} & \bar{\mathbf{A}}\mathbf{K} & \dots & \bar{\mathbf{A}}^{p-2}\mathbf{K} & \bar{\mathbf{A}}^{p-1}\mathbf{K} \end{bmatrix}. \tag{5.15}$$

Since Equation (5.10) is valid for a steady-state Kalman filter, according to the theory [83], $\mathbf{A}^{p} \sim 0$ for $p >> n$. Thus, the observer state vector $\hat{\boldsymbol{x}}$ can be obtained as:

$$\hat{\boldsymbol{x}} = \mathbf{T}\boldsymbol{v}(k). \tag{5.16}$$

Equation (5.16) can be plugged into Equation (5.11) to provide:

$$\boldsymbol{y}(k) = \boldsymbol{\Phi}\boldsymbol{v}(k) + \boldsymbol{\varepsilon}(k) \tag{5.17}$$

where

$$\boldsymbol{\Phi} = \begin{bmatrix} \mathbf{C}\mathbf{K} & \mathbf{C}\bar{\mathbf{A}}\mathbf{K} & \dots & \mathbf{C}\bar{\mathbf{A}}^{p-2}\mathbf{K} & \mathbf{C}\bar{\mathbf{A}}^{p-1}\mathbf{K} \end{bmatrix}. \tag{5.18}$$

Equation (5.17) can be written for $k = p, p+1, \dots, l$ providing:

$$\boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{V}(k) + \mathbf{E}(k) \tag{5.19}$$

where:

$$\mathbf{Y} = \begin{bmatrix} \boldsymbol{y}(p) & \boldsymbol{y}(p+1) & \dots & \boldsymbol{y}(l) \end{bmatrix} \tag{5.20}$$

$$\mathbf{V} = \begin{bmatrix} \boldsymbol{v}(p) & \boldsymbol{v}(p+1) & \dots & \boldsymbol{v}(l) \end{bmatrix} \tag{5.21}$$

$$\mathbf{E} = \begin{bmatrix} \boldsymbol{\varepsilon}(p) & \boldsymbol{\varepsilon}(p+1) & \dots & \boldsymbol{\varepsilon}(l) \end{bmatrix} \tag{5.22}$$

Equation is known to be the core equation of O^{3}KID. The matrices $\mathbf{Y}$ and $\mathbf{V}$ are known by measurement, $\mathbf{\Phi}$ and $\mathbf{E}$ are not. Since $l - p > qp$ the solution can be obtained via least squares. The matrix $\tilde{\mathbf{\Phi}} \sim \mathbf{\Phi}$ is given by:

$$\tilde{\mathbf{\Phi}} = \mathbf{Y}\mathbf{V}^T \left(\mathbf{V}\mathbf{V}^T\right)^{-1} \tag{5.23}$$

Analogously, the estimation of the corresponding residuals $\tilde{\mathbf{E}} \sim E$ can be computed as:

$$\tilde{\mathbf{E}} = \mathbf{Y} - \tilde{\mathbf{\Phi}}\mathbf{V}. \tag{5.24}$$

By post-multipling Equation (5.24) by $\mathbf{V}^T$:

$$\tilde{\mathbf{E}}\mathbf{V}^T = \mathbf{Y}\mathbf{V}^T - \tilde{\mathbf{\Phi}}\mathbf{V}\mathbf{V}^T \tag{5.25}$$

so that substituting $\tilde{\mathbf{\Phi}}$ from Equation (5.23):

$$\tilde{\mathbf{E}}\mathbf{V}^T = \mathbf{Y}\mathbf{V}^T - \mathbf{Y}\mathbf{V}^T \left(\mathbf{V}\mathbf{V}^T\right)^{-1} \mathbf{V}\mathbf{V}^T. \tag{5.26}$$

Thus we can conclude that $\tilde{\mathbf{E}}\mathbf{V}^T = 0$ and so that:

$$\sum_{k=p}^{l} \tilde{\boldsymbol{e}}(k)\boldsymbol{y}^T(k-j) = 0 \qquad j = 1, 2, \dots, p \tag{5.27}$$

where $\tilde{\boldsymbol{e}}(k)$ is the k^{th} column vector of the matrix $\tilde{\mathbf{E}}$. Since by assumption the process is stationary and ergodic, the solution provided by least squares leads to an approximation of the Markov parameters and output residuals of the Kalman filter related to the system matrices

$\mathbf{A}$, $\mathbf{C}$ and noise statistics which generated the output sequence $\{\boldsymbol{y}(k)\}$. Hence:

$$\tilde{\mathbf{E}} \sim \mathbf{E} \tag{5.28}$$

$$\tilde{\mathbf{\Phi}} \sim \mathbf{\Phi} \tag{5.29}$$

This result is extremely important as it points out that the solution of the *autoregressive* model in Equation (5.19) given by Equation (5.23) provides an estimate of the Markov parameters $\mathbf{CK}, \mathbf{C\bar{A}K}, \ldots, \mathbf{C\bar{A}}^{p-1}\mathbf{K}$ of the observer. The matrix $\mathbf{\Phi}$ can be manipulated so to provide the system Markov parameters $\mathbf{CK}, \mathbf{CAK}, \ldots, \mathbf{CA}^{p-1}\mathbf{K}$. At this point, the Eigensystem Realization Algorithm (ERA) can be used so to identify the matrices $\mathbf{A}$, $\mathbf{C}$ and $\mathbf{K}$.

5.4. State space system estimation through a Bayesian strategy

In statistics, the Gauss-Markov theorem states that considering a linear regression model in which the errors are uncorrelated and have equal variances and expectation zero, the ordinary least squares provides the best linear unbiased estimator (BLUE) of the coefficients. As mentioned in the previous section, the value of p is required to be $p >> n$ in order to make the Kalman filter stable ($\bar{\mathbf{A}}^p \sim 0$).

Equation (5.23) shows that the matrix $\mathbf{\Phi}$ can be approximated by the matrix $\tilde{\mathbf{\Phi}}$ provided by OLS regression. One of the most challenging tasks in the definition of an autoregressive model is the choice of the number of coefficients to consider. Among all the possible approaches to this problem, the Akaike Information Criterion (AIC method) is one of the most used in literature [56]. Its objective is to select a number of autoregressive coefficients so that the regression model provides a good mathematical representation of the system, avoiding problems of overfitting. As previously pointed out, the validity of Equation (5.10) relies on the fact that the Kalman filter is assumed to be steady-state and so large values of p are needed. For very large values of p, the overfitting problem represents the real threat to the system identification approach.

Bayesian methods have been proved to be extremely robust to overfitting [98, 99]. Among those methods, Variational Inference (VI) is presented in this chapter as a valid alternative to the conventional OLS regression model. The implementation of this method is discussed in the next section.

5.5. Recurrent Variational Inference model

Variational inference recurrent regression models rely on the use of a Bayesian approach to create linear regression models which have been shown to be significantly robust to overfitting. In this chapter, the procedure used to develop the Bayesian autoregressive model is conducted similarly to the one in section 4.4.1. For each i^{th}monitored DOF, a linear regression model of the acceleration can be defined as function of the acceleration response time histories collected at every DOF. According to the notation used in the previous section, it is possible to set the k^{th} element of the acceleration response time history at the i^{th} DOF as the k^{th} output observation $y_k^{(i)}$ for the linear regression model so that $y_k^{(i)} = y_i(k)$. The observation $y_k^{(i)}$ is generated by the linear regression model using the k^{th} observed input vector $\boldsymbol{x}_k$ containing the elements of all acceleration time histories at $k-1,\ldots,k-p$ steps so that $\boldsymbol{x}_k = [\boldsymbol{y}^T(k-1),\ldots,\boldsymbol{y}^T(k-p)]^T$ for $\boldsymbol{x}_k \in \mathbb{R}^{d\times 1}$ $(d = N \cdot p)$.

It is noteworthy that, for each monitored DOF, the Bayesian linear regression model is equivalent to one row vector of the matrix $\boldsymbol{\Phi}$ containing the AR coefficients generated by Least Squares (LS) solution of Equation (5.23). The Bayesian approximation of the matrix $\boldsymbol{\Phi}$ can be obtained by properly arranging the features characterizing the Bayesian regression models for all the DOFs.

The complete dataset, for each i^{th} monitored DOF, is of the form $\{y_k^{(i)}, \boldsymbol{x}_k\}_{k=p}^{l}$ and can be split into output features $\mathbf{Y} = \{y_p^{(i)},\ldots,y_l^{(i)}\}$ and input features $\mathbf{X} = \{\boldsymbol{x}_p,\ldots,\boldsymbol{x}_l\}$ so that the i^{th} regression model is estimated based on the total of $l-p+1$ features $(\boldsymbol{x}_k^T, y_k^{(i)})$. At this point, it is important to define the probabilistic model for the acceleration at the time step k monitored at the i^{th} DOF. Let's assume that the scalar quantity $y_k^{(i)}$ is sampled from a Normal distribution whose mean is generated by the product of the vector $\boldsymbol{x}_k^T$ by the vector $\boldsymbol{w} \in \mathbb{R}^{d\times 1}$ and whose variance is λ^{-1}. It is worthy to observe that the elements of the vector $\boldsymbol{w}$ are weights of the components of the vector $\boldsymbol{x}_k$ and can be interpreted as the coefficients of an AR model for $y_k^{(i)}$. Let's assume that the vector $\boldsymbol{w}$ is sampled from a multivariate zero mean Normal distribution whose covariance matrix is given by $\mathrm{Diag}(1/\alpha_1,\ldots,1/\alpha_d)$. The term $\mathrm{Diag}(\cdot)$ represents a diagonal square matrix whose elements along the diagonal are those inside of the parenthesis. In order to define the probability distributions to sample the parameters α_j and λ, two Gamma distributions have been defined as conjugate priors for the Normal distributions. The parameters α_j $(j = 1,\ldots,d)$, defining the covariance matrix of the multivariate Normal distribution from which $\boldsymbol{w}$ is extracted, are sampled from a Gamma distribution whose parameters are a_0 and

b_0. The parameter λ, defining the Normal distribution from which $y_k^{(i)}$ is extracted, is sampled from a Gamma distribution whose parameters are e_0 and f_0. The probability density functions of the model and prior distributions are represented as:

$$y_k^{(i)} \sim \text{Normal}(\boldsymbol{x}_k^T \boldsymbol{w}, \lambda^{-1}), \qquad \boldsymbol{w} \sim \text{Normal}(\mathbf{0}, \text{Diag}(1/\alpha_1, \ldots, 1/\alpha_d)),$$
$$\alpha_j \sim \text{Gamma}(a_0, b_0), \qquad \lambda \sim \text{Gamma}(e_0, f_0).$$

Let's represent the joint probability distribution as $p(\mathbf{Y}, \mathbf{X}, \boldsymbol{w}, \lambda, \boldsymbol{\alpha})$. By using the chain rule, the natural logarithm (ln) of the joint probability distribution can be written as:

$$ln(p(\mathbf{Y}, \mathbf{X}, \boldsymbol{w}, \lambda, \boldsymbol{\alpha})) = ln(p(\mathbf{Y}|\mathbf{X}, \boldsymbol{w}, \lambda, \boldsymbol{\alpha})) + ln(p(\boldsymbol{w}|\boldsymbol{\alpha})) + ln(p(\boldsymbol{\alpha})) + ln(p(\lambda)). \quad (5.30)$$

Following a procedure similar to the one in section 4.3, the objective function $\mathscr{L}$ can be expressed as:

$$\mathscr{L}(\mathbf{Y}, \mathbf{X}, \boldsymbol{w}, \lambda, \boldsymbol{\alpha}) = \int q(\boldsymbol{w}, \lambda, \alpha_1, \ldots, \alpha_d) ln\left(\frac{p(\mathbf{Y}, \boldsymbol{w}, \lambda, \alpha_1, \ldots, \alpha_d|\mathbf{X})}{q(\boldsymbol{w}, \lambda, \alpha_1, \ldots, \alpha_d)}\right) d\boldsymbol{w} d\lambda d\alpha_1 \ldots d\alpha_d. \quad (5.31)$$

Let's now focus on the probability distributions inside of the integral of Equation (5.31). By considering the mean-field assumption in section 4.3, the auxiliary distribution can be factorized as $q(\boldsymbol{w}, \lambda, \alpha_1, \ldots, \alpha_d) \approx q(\lambda) q(\boldsymbol{w}) \prod_{j=1}^{d} q(\alpha_j)$. The factored terms are analyzed in the next section. For sake of simplicity, since the procedure has to be considered for each i^{th} DOF, it is convenient to remove the superscript (i) so that $y_k^{(i)}$ becomes y_k, with the understanding that the procedure is applied for the i^{th} DOF ($i = 1, \ldots, N$).

5.5.1. Auxiliary distribution factorization

Recalling section 4.4.1, the probability distributions $q(\alpha_j)$, $q(\lambda)$, $q(\boldsymbol{w})$ factorizing the auxiliary probability distribution can be estimated by plugging the joint distribution in Equation (5.30) into Equation (4.23). Let's now analyze each of them singularly.

- $q(\alpha_j)$

Referring to Equation (4.23), the auxiliary distribution $q(\alpha_j)$ is proportional to

$$q(\alpha_j) \propto \exp(\mathbb{E}_{q(\lambda,\boldsymbol{w},\alpha_{j'\neq j})}[ln(p(\mathbf{Y}|\mathbf{X},\boldsymbol{w},\lambda,\boldsymbol{\alpha}))+ln(p(\boldsymbol{w}|\boldsymbol{\alpha}))+ln(p(\boldsymbol{\alpha}))+ln(p(\lambda))]) \tag{5.32}$$

$$q(\alpha_j) \propto \exp(\mathbb{E}_{q(\lambda,\boldsymbol{w},\alpha_{j'\neq j})}[ln(p(\boldsymbol{w}|\boldsymbol{\alpha}))+ln(p(\alpha_j))]) \tag{5.33}$$

where $\mathbb{E}_{q(\lambda,\boldsymbol{w},\alpha_{j'\neq j})}[\cdot]$ represents the expectation operator over all the auxiliary distributions $q(\cdot)$ defined by the parameters $\boldsymbol{w}$, λ and $\alpha'_j \neq \alpha_j$. The terms $ln(p(\mathbf{Y}|\mathbf{X},\boldsymbol{w},\lambda,\boldsymbol{\alpha}))$ and $ln(p(\lambda))$ in Equation (5.32) are independent from α_j so they are not considered in Equation (5.33). The model $p(\boldsymbol{w}|\boldsymbol{\alpha})$ and the prior distribution $p(\alpha_j)$ in Equation (5.33) can be expressed as:

$$p(\boldsymbol{w}|\boldsymbol{\alpha}) = (2\pi)^{-\frac{d}{2}}\det(\mathrm{Diag}(1/\boldsymbol{\alpha}))^{-\frac{1}{2}}e^{-\boldsymbol{w}^T\frac{\mathrm{Diag}(\boldsymbol{\alpha})}{2}\boldsymbol{w}} \tag{5.34}$$

$$p(\alpha_j) = \frac{b_0^{a_0}}{\Gamma(a_0)}\alpha_j^{a_0-1}e^{-b_0\alpha_j} \tag{5.35}$$

where $\det(\cdot)$ is the determinant operator, the terms $\mathrm{Diag}(\boldsymbol{\alpha})$ and $\mathrm{Diag}(1/\boldsymbol{\alpha})$ represent two diagonal matrices whose elements along the diagonal are respectively $\alpha_1,\ldots,\alpha_d$ and $1/\alpha_1,\ldots,1/\alpha_d$. By plugging Equations (5.34) and (5.35) into Equation (5.33) $q(\alpha_j)$ can be written as:

$$q(\alpha_j) \propto \det(\mathrm{Diag}(1/\boldsymbol{\alpha}))^{-\frac{1}{2}}e^{-\frac{1}{2}\mathbb{E}_{q(\boldsymbol{w},\alpha_{j'\neq j})}[\boldsymbol{w}^T\mathrm{Diag}(\boldsymbol{\alpha})\boldsymbol{w}]}\frac{b_0^{a_0}}{\Gamma(a_0)}\alpha_j^{a_0-1}e^{-b_0\alpha_j}. \tag{5.36}$$

By considering only the terms depending on α_j:

$$q(\alpha_j) \propto \alpha_j^{\frac{1}{2}}e^{-\frac{1}{2}\mathbb{E}_{q(\boldsymbol{w},\alpha_{j'\neq j})}[w_j^2\alpha_j]}\alpha_j^{a_0-1}e^{-b_0\alpha_j} \tag{5.37}$$

where

$$\mathbb{E}_{q(\boldsymbol{w},\alpha_{j'\neq j})}[w_j^2\alpha_j] = \mathbb{E}_{q(\boldsymbol{w})}[w_j^2]\alpha_j. \tag{5.38}$$

So, finally, we have:

$$q(\alpha_j) \propto \alpha_j^{a_0-1/2}e^{-\alpha_j(b_0+\frac{1}{2}\mathbb{E}_{q(\boldsymbol{w})}[w_j^2])}. \tag{5.39}$$

Equation (5.39) shows that the distributions $q(\alpha_j)$ are Gamma(a',b'_j) distribution, whose

parameters are $a' = a_0 + 1/2$ and $b'_j = b_0 + 1/2\mathbb{E}_{q(\boldsymbol{w})}[w_j^2]$.

- $q(\lambda)$

  Let's now focus on the auxiliary distribution $q(\lambda)$. By substituting the joint distribution in Equation 5.30 inside of Equation (4.23) yields:

$$q(\lambda) \propto exp(\mathbb{E}_{q(\boldsymbol{\alpha},\boldsymbol{w})}[ln(p(\mathbf{Y}|\mathbf{X},\boldsymbol{w},\lambda,\boldsymbol{\alpha})) + ln(p(\lambda))]) \tag{5.40}$$

  where the term $\mathbb{E}_{q(\boldsymbol{\alpha},\boldsymbol{w})}[\cdot]$ represents the expectation operator over the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{w}$. Since, given the parameters $\boldsymbol{w}$ and λ and the data in $\mathbf{X}$, the term $p(\mathbf{Y}|\mathbf{X},\boldsymbol{w},\lambda,\boldsymbol{\alpha})$ becomes independent of $\boldsymbol{\alpha}$, it can be factorized as:

$$p(\mathbf{Y}|\mathbf{X},\boldsymbol{w},\lambda,\boldsymbol{\alpha}) = p(\mathbf{Y}|\mathbf{X},\boldsymbol{w},\lambda) = \prod_{k=p}^{l} p(y_k|\boldsymbol{x}_k,\boldsymbol{w},\lambda). \tag{5.41}$$

  Recalling the initial assumptions on the probability distributions, the distributions $p(y_k|\boldsymbol{x}_k,\boldsymbol{w},\lambda)$ are Normal probability distributions:

$$p(y_k|\boldsymbol{x}_k,\boldsymbol{w},\lambda) = \sqrt{\frac{\lambda}{2\pi}} e^{-\frac{\lambda}{2}(y_k - \boldsymbol{x}_k^T\boldsymbol{w})^2} \tag{5.42}$$

  and $p(\lambda)$ is a Gamma probabiliy distribution:

$$p(\lambda) = \frac{f_0^{e_0}}{\Gamma(e_0)} \lambda^{e_0-1} e^{-f_0\lambda}. \tag{5.43}$$

  Equations (5.42), (5.43) and (5.41) can be plugged into Equation (5.40) yielding:

$$\begin{aligned}
q(\lambda) &\propto exp\left(\mathbb{E}_{q(\boldsymbol{\alpha},\boldsymbol{w})}\left[ln\left(\sqrt{\frac{\lambda}{2\pi}}\right)^{l-p+1} - \frac{\lambda}{2}\sum_{k=p}^{l}(y_k - \boldsymbol{x}_k^T\boldsymbol{w})^2 + ln\left(\frac{f_0^{e_0}}{\Gamma(e_0)}\lambda^{e_0-1}e^{-f_0\lambda}\right)\right]\right) \\
q(\lambda) &\propto \left(\sqrt{\frac{\lambda}{2\pi}}\right)^{l-p+1} e^{-\frac{\lambda}{2}\sum_{k=p}^{l}\mathbb{E}_{q(\boldsymbol{w})}[(y_k - \boldsymbol{x}_k^T\boldsymbol{w})^2]} \frac{f_0^{e_0}}{\Gamma(e_0)}\lambda^{e_0-1}e^{-f_0\lambda} \\
q(\lambda) &\propto \lambda^{(e_0 + \frac{l-p+1}{2})-1} e^{-\lambda(f_0 + \frac{1}{2}\sum_{k=p}^{l}\mathbb{E}_{q(\boldsymbol{w})}[(y_k - \boldsymbol{x}_k^T\boldsymbol{w})^2])}
\end{aligned} \tag{5.44}$$

  The distribution $q(\lambda)$ is a Gamma distribution, Gamma(e', f'), whose parameters are $e' = e_0 + \frac{l-p+1}{2}$ and $f' = f_0 + \frac{1}{2}\sum_{k=p}^{l}\mathbb{E}_{q(\boldsymbol{w})}[(y_k - \boldsymbol{x}_k^T\boldsymbol{w})^2]$.

- $q(\boldsymbol{w})$

  Let's now focus on the auxiliary distribution $q(\boldsymbol{w})$. The joint distribution in Equation

5.30 can be plugged into Equation (4.23) leading to:

$$q(\boldsymbol{w}) \propto exp(\mathbb{E}_{q(\boldsymbol{\alpha},\lambda)}[ln(p(\mathbf{Y}|\mathbf{X},\boldsymbol{w},\lambda,\boldsymbol{\alpha}))+ln(p(\boldsymbol{w}|\boldsymbol{\alpha}))]) \tag{5.45}$$

where the term $\mathbb{E}_{q(\boldsymbol{\alpha},\lambda)}[\cdot]$ represents the expectation operator over the parameters $\boldsymbol{\alpha}$ and λ. The probability distribution $p(\mathbf{Y}|\mathbf{X},\boldsymbol{w},\lambda,\boldsymbol{\alpha})$ can be obtained by plugging Equation (5.42) into Equation (5.41). The probability distribution $p(\boldsymbol{w}|\boldsymbol{\alpha})$ is the multivariate Normal distribution in Equation (5.34). By substituting Equations (5.34), (5.41) and (5.42) into Equation (5.45):

$$q(\boldsymbol{w}) \propto e^{\mathbb{E}_{q(\boldsymbol{\alpha},\lambda)}[ln\left(\frac{\lambda}{2\pi}\right)^{\frac{l-p+1}{2}}-\sum_{k=p}^{l}\frac{\lambda}{2}(y_k-\boldsymbol{x}_k^T\boldsymbol{w})^2+ln((2\pi)^{-\frac{d}{2}}\det(\mathrm{Diag}(1/\boldsymbol{\alpha}))^{\frac{1}{2}})-\boldsymbol{w}^T\frac{\mathrm{Diag}(\boldsymbol{\alpha})}{2}\boldsymbol{w}]}$$

$$\begin{aligned} q(\boldsymbol{w}) &\propto e^{-\mathbb{E}_{q(\lambda)}[\frac{\lambda}{2}]\sum_{k=p}^{l}(y_k-\boldsymbol{x}_k^T\boldsymbol{w})^2}e^{-\boldsymbol{w}^T\frac{\mathbb{E}_{q(\boldsymbol{\alpha})}[\mathrm{Diag}(\boldsymbol{\alpha})]}{2}\boldsymbol{w}} \\ q(\boldsymbol{w}) &\propto e^{-\frac{1}{2}\left(\mathbb{E}_{q(\lambda)}[\lambda]\sum_{k=p}^{l}(y_k-\boldsymbol{x}_k^T\boldsymbol{w})^2+\sum_{j=1}^{d}w_j^2\mathbb{E}_{q(\alpha_j)}[\alpha_j]\right)}. \end{aligned} \tag{5.46}$$

Equation (5.46) shows that $q(\boldsymbol{w})$ is actually a multivariate Normal probability distribution, i.e. $q(\boldsymbol{w}) \sim \mathcal{N}(\boldsymbol{\mu}', \Sigma')$. The mean vector $\boldsymbol{\mu}'$ and covariance matrix Σ' can be found by analyzing the exponent in Equation (5.46):

$$\begin{aligned} & \mathbb{E}_{q(\lambda)}[\lambda]\sum_{k=p}^{l}(y_k-\boldsymbol{x}_k^T\boldsymbol{w})^2+\boldsymbol{w}^T\mathbb{E}_{q(\boldsymbol{\alpha})}[Diag(\boldsymbol{\alpha})]\boldsymbol{w} \\ = \;& \mathbb{E}_{q(\lambda)}[\lambda]\sum_{k=p}^{l}\left(y_k^2-2y_k\boldsymbol{x}^T\boldsymbol{w}+\boldsymbol{x}_k^T\boldsymbol{w}\boldsymbol{w}^T\boldsymbol{x}_k\right)+\boldsymbol{w}^T\mathbb{E}_{q(\boldsymbol{\alpha})}[Diag(\alpha)]\boldsymbol{w} \\ = \;& \sum_{k=p}^{l}\left(\mathbb{E}_{q(\lambda)}[\lambda]y_k^2-2y_k\mathbb{E}_{q(\lambda)}[\lambda]\boldsymbol{x}_k^T\boldsymbol{w}\right)+\boldsymbol{w}^T\left(\mathbb{E}_{q(\lambda)}[\lambda]\sum_{k=p}^{l}\boldsymbol{x}_k\boldsymbol{x}_k^T\right)\boldsymbol{w}+\boldsymbol{w}^T\mathbb{E}_{q(\boldsymbol{\alpha})}[Diag(\boldsymbol{\alpha})]\boldsymbol{w} \\ = \;& \sum_{k=p}^{l}\left(\mathbb{E}_{q(\lambda)}[\lambda]y_k^2-2y_k\mathbb{E}_{q(\lambda)}[\lambda]\boldsymbol{x}_k^T\boldsymbol{w}\right)+\boldsymbol{w}^T\left(\mathbb{E}_{q(\boldsymbol{\alpha})}[Diag(\boldsymbol{\alpha})]+\mathbb{E}_{q(\lambda)}[\lambda]\sum_{k=p}^{l}\boldsymbol{x}_k\boldsymbol{x}_k^T\right)\boldsymbol{w} \end{aligned} \tag{5.47}$$

By comparison with the conventional multivariate Normal probability distribution, we can conclude that the covariance $\Sigma' = (\mathbb{E}_{q(\boldsymbol{\alpha})}[\mathrm{Diag}(\boldsymbol{\alpha})]+\mathbb{E}_{q(\lambda)}[\lambda]\sum_{k=p}^{l}\boldsymbol{x}_k\boldsymbol{x}_k^T)^{-1}$. From the general expansion of the argument in the exponent of a multivariate Normal probability distribution $q(\boldsymbol{w}) \sim \mathcal{N}(\boldsymbol{\mu}', \Sigma')$, the term linearly depending on $\boldsymbol{w}$, i.e. $2\boldsymbol{w}^T\Sigma'^{-1}\boldsymbol{\mu}'$, is equal to the one in Equation (5.47) so that:

$$2\boldsymbol{w}^T\Sigma'^{-1}\boldsymbol{\mu}' = \sum_{k=p}^{l}2y_k\mathbb{E}_{q(\lambda)}[\lambda]\boldsymbol{x}_k^T\boldsymbol{w} \tag{5.48}$$

leading to:

$$\boldsymbol{\mu}' = \Sigma'(\mathbb{E}_{q(\lambda)}[\lambda] \sum_{k=p}^{l} y_k \boldsymbol{x}_k). \tag{5.49}$$

Finally, once the $q(\alpha_j)$, $q(\lambda)$ and $q(\boldsymbol{w})$ have been identified, it is possible to compute the expectations. Using the conventional formulas for the Gamma and multivariate Normal distributions it is possible to obtain:

$$\mathbb{E}_{q(\alpha_j)}[\alpha_j] = \frac{a'}{b'_j} \tag{5.50}$$

$$\mathbb{E}_{q(\lambda)}[\lambda] = \frac{e'}{f'} \tag{5.51}$$

$$\mathbb{E}_{q(\boldsymbol{w})}[w_j^2] = (\Sigma' + \boldsymbol{\mu}'\boldsymbol{\mu}'^T)_{j,j} \tag{5.52}$$

The last term to evaluate is $\mathbb{E}_{q(\boldsymbol{w})}[(y_k - \boldsymbol{x}_k^T \boldsymbol{w})^2]$:

$$\mathbb{E}_{q(\boldsymbol{w})}[(y_k - \boldsymbol{x}_k^T \boldsymbol{w})^2] = y_k^2 - 2y_k \boldsymbol{x}_k^T \mathbb{E}_{q(\boldsymbol{w})}[\boldsymbol{w}] + \boldsymbol{x}_k^T \mathbb{E}_{q(\boldsymbol{w})}[\boldsymbol{w}\boldsymbol{w}^T]\boldsymbol{x}_k. \tag{5.53}$$

By recalling the conventional equality $\Sigma' = \mathbb{E}_{q(\boldsymbol{w})}[\boldsymbol{w}\boldsymbol{w}^T] - \boldsymbol{\mu}'\boldsymbol{\mu}'^T$ for a multivariate Normal distribution, Equation (5.53) can be expressed as:

$$\begin{aligned} \mathbb{E}_{q(\boldsymbol{w})}[(y_k - \boldsymbol{x}_k^T \boldsymbol{w})^2] &= y_k^2 - 2y_k \boldsymbol{x}_k^T \boldsymbol{\mu}' + \boldsymbol{x}_k^T [\Sigma' + \boldsymbol{\mu}'\boldsymbol{\mu}'^T]\boldsymbol{x}_k \\ &= (y_k - \boldsymbol{x}_k^T \boldsymbol{\mu}')^2 + \boldsymbol{x}_k^T \Sigma' \boldsymbol{x}_k. \end{aligned} \tag{5.54}$$

It is now possible to conclude that the computation of the auxiliary distribution $q(\boldsymbol{w}, \boldsymbol{\alpha}, \lambda)$ fully depends on the values of the parameters a'. b'_j, e', f', $\boldsymbol{\mu}'$ and Σ'. Through an iterative process, the value of these parameters will be updated (iteratively) so to provide the optimal estimation of the auxiliary distribution.

At this point we obtained the analytical expressions for the auxiliary distributions $q(\boldsymbol{w})$, $q(\lambda)$ and $q(\boldsymbol{\alpha})$ and so we are ready to proceed to the computation of the objective function $\mathscr{L}$.

5.5.2. Objective function

By applying the chain rule, it is possible to write the logarithm of the probability distribution $p(\mathbf{Y}, \boldsymbol{w}, \lambda, \alpha_1, \ldots, \alpha_d | \boldsymbol{X})$ inside of the integral in Equation (5.31) as:

$$ln\left(p(\mathbf{Y}, \boldsymbol{w}, \lambda, \alpha_1, \ldots, \alpha_d | \boldsymbol{X})\right) = ln\left(p(\mathbf{Y} | \boldsymbol{w}, \lambda, \boldsymbol{X}) p(\lambda) p(\boldsymbol{w} | \boldsymbol{\alpha}) p(\boldsymbol{\alpha})\right). \tag{5.55}$$

By using the logarithm properties, the term inside the integral in Equation (5.31) can be expressed as:

$$\begin{aligned} & q(\boldsymbol{w}, \lambda, \alpha_1, \ldots, \alpha_d) ln\left(\frac{p(\mathbf{Y}, \boldsymbol{w}, \lambda, \alpha_1, \ldots, \alpha_d | \boldsymbol{X})}{q(\boldsymbol{w}, \lambda, \alpha_1, \ldots, \alpha_d)}\right) \\ & = q(\boldsymbol{w}, \lambda, \alpha_1, \ldots, \alpha_d) ln\left(p(\mathbf{Y}, \boldsymbol{w}, \lambda, \alpha_1, \ldots, \alpha_d | \boldsymbol{X})\right) - q(\boldsymbol{w}, \lambda, \alpha_1, \ldots, \alpha_d) ln\left(q(\boldsymbol{w}, \lambda, \alpha_1, \ldots, \alpha_d)\right). \end{aligned} \tag{5.56}$$

The two logarithms in Equation (5.56) can be decomposed into sums of logarithms. The first term leads to

$$\begin{aligned} ln\left(p(\mathbf{Y}, \boldsymbol{w}, \lambda, \alpha_1, \ldots, \alpha_d | \boldsymbol{X})\right) &= ln\left(p(\mathbf{Y} | \boldsymbol{w}, \lambda, \boldsymbol{X}) p(\boldsymbol{w} | \boldsymbol{\alpha}) p(\boldsymbol{\alpha}) p(\lambda)\right) \\ &= ln\left(p(\mathbf{Y} | \boldsymbol{w}, \lambda, \mathbf{X})\right) + ln(p(\boldsymbol{w} | \boldsymbol{\alpha})) + ln(p(\boldsymbol{\alpha})) + ln(p(\lambda)) \end{aligned} \tag{5.57}$$

and the second to

$$ln(q(\boldsymbol{w}, \lambda, \alpha_1, \ldots, \alpha_d)) = ln(q(\boldsymbol{w})) + ln(q(\lambda)) + ln(q(\boldsymbol{\alpha})) \tag{5.58}$$

where $ln(q(\boldsymbol{\alpha})) = \sum_{j=1}^{d} ln(q(\alpha_j))$. Thus, the integral in Equation (5.31), which leads to the formulation of the objective function, can be decomposed into the sum of 7 integrals: the 4 terms in Equations (5.57) and the 3 terms in Equation (5.58). At this point, the 7 terms can be singularly analyzed.

1. $p(\mathbf{Y} | \boldsymbol{w}, \lambda, \boldsymbol{X})$

   Let's compute the integral of the first term in Equation (5.57) contributing to the objective function $\mathscr{L}$:

$$\int \ldots \int q(\lambda, \boldsymbol{w}, \alpha_1, \ldots, \alpha_d) ln(p(\mathbf{Y} | \boldsymbol{w}, \mathbf{X}, \lambda)) d\boldsymbol{w} d\lambda d\alpha_1 \ldots d\alpha_d. \tag{5.59}$$

Since the term $ln(p(\mathbf{Y}|\boldsymbol{w},\mathbf{X},\lambda))$ is independent from $\boldsymbol{\alpha}$ it is possible to write:

$$\begin{aligned}
&\int \dots \int q(\lambda,\boldsymbol{w}) ln(p(\mathbf{Y}|\boldsymbol{w},\mathbf{X},\lambda)) d\boldsymbol{w} d\lambda \\
&= \mathbb{E}_{q(\lambda,\boldsymbol{w})}\left[\sum_{k=q}^{l} ln(p(y_k|\boldsymbol{w},\boldsymbol{x}_k,\lambda))\right] \\
&= \mathbb{E}_{q(\lambda,\boldsymbol{w})}\left[\sum_{k=q}^{l} ln\left(\sqrt{\frac{\lambda}{2\pi}} e^{-\frac{\lambda}{2}(y_k-\boldsymbol{x}_k^T\boldsymbol{w})^2}\right)\right] \\
&= \sum_{k=p}^{l} \frac{1}{2}\mathbb{E}_{q(\lambda)}[ln((\lambda))] - \sum_{k=p}^{l} \frac{1}{2}\mathbb{E}_{q(\lambda)}[\lambda]\,\mathbb{E}_{q(\boldsymbol{w})}\left[(y_k-\boldsymbol{x}_k^T\boldsymbol{w})^2\right] + \text{const.} \quad (5.60)
\end{aligned}$$

By considering that the logarithmic expectation of the $q(\lambda)$ Gamma distribution is $\mathbb{E}_{q(\lambda)}[ln(\lambda)] = \psi(e') - ln(f')$ and recalling the expectations expressed in Equations (5.51) and (5.54), Equation (5.60) becomes:

$$\frac{l-p+1}{2}(\psi(e') - ln(f')) - \sum_{k=p}^{l} \frac{1}{2}\frac{e'}{f'}[(y_k - \boldsymbol{x}_k^T\boldsymbol{\mu}')^2 + \boldsymbol{x}_k^T\Sigma'\boldsymbol{x}_k] + \text{const} \quad (5.61)$$

2. $p(\boldsymbol{w}|\boldsymbol{\alpha})$

The integral of the second term in Equation (5.57) contributing to the objective function $\mathscr{L}$ is expressed as:

$$\int \dots \int \int q(\lambda)q(\boldsymbol{w}) \prod_{i=1}^{d} q(\alpha_i) ln\left(p(\boldsymbol{w}|\alpha_1,\dots,\alpha_d)\right) d\boldsymbol{w} d\lambda d\alpha_1 \dots d\alpha_d \quad (5.62)$$

We can observe that $ln(p(\boldsymbol{w}|\boldsymbol{\alpha}))$ is independent from λ. By substituting Equation (5.34) into Equation (5.62):

$$\begin{aligned}
&\mathbb{E}_{q(\boldsymbol{w},\boldsymbol{\alpha})}\left[\left(-\frac{1}{2}\right)\left(ln(\det(2\pi \text{Diag}(1/\boldsymbol{\alpha}))) + \boldsymbol{w}^T \text{Diag}(\boldsymbol{\alpha})\boldsymbol{w}\right)\right] \\
= \quad & -\frac{1}{2}\sum_{j=1}^{d} \mathbb{E}_{q(\alpha_j)}[ln(\frac{2\pi}{\alpha_j})] - \frac{1}{2}\sum_{j=1}^{d} \mathbb{E}_{q(\boldsymbol{w})}[w_j^2]\mathbb{E}_{q(\alpha_j)}[\alpha_j]. \quad (5.63)
\end{aligned}$$

The expectations in Equation (5.63) can be computed as $\mathbb{E}_{q(\alpha_j)}[ln\left(\frac{2\pi}{\alpha_j}\right)] = ln(2\pi) - (\psi(a') - ln\left(b'_j\right))$, $\mathbb{E}_{q(\boldsymbol{w})}[w_j^2] = [\Sigma' + \boldsymbol{\mu}'\boldsymbol{\mu}'^T]_{j,j}$ and $\mathbb{E}_{q(\alpha_j)}[\alpha_j] = \frac{a'}{b'_j}$ leading to:

$$\frac{1}{2}\sum_{j=1}^{d}(\psi(a') - ln(b'_j) - \frac{a'}{b'_j}[\Sigma' + \boldsymbol{\mu}'\boldsymbol{\mu}'^T]_{j,j}) + \text{const.} \quad (5.64)$$

3. $p(\boldsymbol{\alpha})$

   The integral of the third term in Equation (5.57) can be expressed as:

$$\int \dots \int q(\lambda) q(\boldsymbol{w}) \prod_{j=1}^{d} q(\alpha_j) \sum_{j=1}^{d} ln(p(\alpha_j)) d\boldsymbol{w} d\lambda d\alpha_1 \dots d\alpha_d \tag{5.65}$$

   The term $\sum_{j=1}^{d} ln(p(\alpha_j))$ is independent from λ and $\boldsymbol{w}$, so we can write Equation (5.65) as:

$$\begin{aligned} & \mathbb{E}_{q(\boldsymbol{\alpha})} \left[\sum_{j=1}^{d} ln(p(\alpha_j)) \right] \\ = & \sum_{j=1}^{d} \mathbb{E}_{q(\alpha_j)} \left[ln(p(\alpha_j)) \right] \end{aligned} \tag{5.66}$$

   Which can be rewritten as:

$$\sum_{j=1}^{d} \left[(a_0 - 1)(\psi(a') - ln(b_j)) - b_0 \frac{a'}{b'_j} \right] + \text{const.} \tag{5.67}$$

4. $p(\lambda)$

   Let's compute the integral of the fourth term in Equation (5.57) contributing to the objective function $\mathscr{L}$:

$$\int \dots \int q(\lambda) q(\boldsymbol{w}) \prod_{j=1}^{d} q(\alpha_j) ln(p(\lambda)) d\boldsymbol{w} d\lambda d\alpha_1 \dots d\alpha_d \tag{5.68}$$

   The term $p(\lambda)$ is independent from $\boldsymbol{w}$ and $\boldsymbol{\alpha}$, so Equation (5.68) can be written as:

$$\mathbb{E}_{q(\lambda)} \left[ln(p(\lambda)) \right] \tag{5.69}$$

   and, by analyzing the distribution, it is possible to obtain:

$$(e_0 - 1)(\psi(e') - ln(f')) - f_0 \frac{e'}{f'} + \text{const.} \tag{5.70}$$

   At this point the expression for the integral of the first four terms in Equation (5.57) has been derived. Next, we will evaluate the integrals of the three terms in Equation (5.58) which are the so called 'entropies'.

5. $q(\boldsymbol{w})$

   Recalling the definition of entropy as the negative of the expected value of the natural logarithm of a distribution with respect to itself, the negative integral of the first term in Equation (5.58) can be interpreted as the entropy of the distribution $q(\boldsymbol{w})$:

$$-\int \ldots \int q(\lambda)q(\boldsymbol{w}) \prod_{j=1}^{d} q(\alpha_j) ln(q(\boldsymbol{w})) d\boldsymbol{w} d\lambda d\alpha_1 \ldots d\alpha_d. \tag{5.71}$$

   Since $q(\boldsymbol{w})$ is a multivariate normal distribution, the entropy is given by:

$$-\mathbb{E}_{q(\boldsymbol{w})}\left[ln(q(\boldsymbol{w}))\right] = \frac{1}{2} ln(\det(2\pi e \Sigma')) = \frac{1}{2} ln(\det(\Sigma')) + \text{const.} \tag{5.72}$$

6. $q(\lambda)$

   The entropy of the second term in Equation (5.58) contributing to the objective function $\mathscr{L}$ is given by:

$$-\int \ldots \int q(\lambda)q(\boldsymbol{w}) \prod_{j=1}^{d} q(\alpha_j) ln(q(\lambda)) d\boldsymbol{w} d\lambda d\alpha_1 \ldots d\alpha_d \tag{5.73}$$

   The distribution $q(\lambda)$ is a Gamma(e', f') distribution and its entropy is given by:

$$-\mathbb{E}_{q(\lambda)}\left[ln(q(\lambda))\right] = e' - ln(f') + ln(\Gamma(e')) + (1-e')\psi(e'). \tag{5.74}$$

7. $q(\boldsymbol{\alpha})$

   The entropy of the distribution represented by the third term in Equation (5.58) contributing to the objective function $\mathscr{L}$ can be written as:

$$-\int \ldots \int q(\lambda)q(\boldsymbol{w}) \prod_{j=1}^{d} q(\alpha_j) ln(q(\boldsymbol{\alpha})) dw d\lambda d\alpha_1 \ldots d\alpha_d. \tag{5.75}$$

   The entropy given by Equation (5.75) can be computed by factorizing the distribution $q(\boldsymbol{\alpha})$ so to obtain

$$-\int \ldots \int q(\lambda)q(\boldsymbol{w}) \prod_{j=1}^{d} q(\alpha_j) \sum_{j=1}^{d} ln(q(\alpha_j)) d\boldsymbol{w} d\lambda d\alpha_1 \ldots d\alpha_d \tag{5.76}$$

leading to:

$$-\mathbb{E}_{q(\boldsymbol{\alpha})}\left[\sum_{j=1}^{d} ln(q(\alpha_j))\right]. \tag{5.77}$$

Since $q(\alpha_j)$ is a Gamma(a', b'_j) distribution, the sum of the d entropies is given by:

$$\begin{aligned} -\sum_{j=1}^{d} \mathbb{E}_{q(\alpha_j)}\left[ln(q(\alpha_j))\right] &= \sum_{j=1}^{d} (a' - ln(b'_j) + ln(\Gamma(a')) + (1-a')\psi(a')) \\ &= d(a' + ln(\Gamma(a')) + (1-a')\psi(a')) - \sum_{j=1}^{d} ln(b'_j). \end{aligned} \tag{5.78}$$

Finally, we can compute the objective function $\mathscr{L}$ by considering the terms in Equations (5.61), (5.64), (5.67), (5.70), (5.72), (5.74) and (5.78):

$$\begin{aligned} \mathscr{L}(a', \boldsymbol{b}', e', f', \boldsymbol{\mu}', \Sigma') &= \frac{l-p+1}{2}(\psi(e') - ln(f')) - \sum_{k=p}^{l} \frac{1}{2}\frac{e'}{f'}[(y_k - \boldsymbol{x}_k^T \boldsymbol{\mu}')^2 + \boldsymbol{x}_k^T \Sigma' \boldsymbol{x}_k] \\ &+ \frac{1}{2}\sum_{j=1}^{d}(\psi(a') - ln(b'_j) - \frac{a'}{b'_j}[\Sigma' + \boldsymbol{\mu}'\boldsymbol{\mu}'^T]_{j,j}) \\ &+ \sum_{j=1}^{d}\left[(a_0 - 1)(\psi(a') - ln(b_j)) - b_0\frac{a'}{b'_j}\right] \\ &+ (e_0 - 1)(\psi(e') - ln(f')) - f_0\frac{e'}{f'} \\ &+ \frac{1}{2}ln(det(\Sigma')) \\ &+ e' - ln(f') + ln(\Gamma(e')) + (1-e')\psi(e') \\ &+ d(a' + ln(\Gamma(a')) + (1-a')\psi(a')) - \sum_{i=1}^{d} ln(b'_j) \\ &+ \text{const} \end{aligned} \tag{5.79}$$

where $\boldsymbol{b}' = \{b'_1, \ldots, b'_d\}$.

5.5.3. Practical implementation of the Variational Inference regression algorithm

Once the analytical expression fo the distributions $q(\boldsymbol{w})$, $q(\boldsymbol{\alpha})$ and $q(\lambda)$ have been derived and the inference between their parameters a', b'_j, e', f', $\boldsymbol{\mu}'$ and Σ' has been shown, it is possible to

use these distributions in an iterative algorithm that updates the distributions' parameters. The correct implementation and convergence of the auxiliary distribution parameters is validated by the objective function which is always negative and monotonically increasing before converging to a constant value. The implementation of the VI recurrent regression algorithm relies on three steps. Let's indicate with t the superscript of iteration $t = 1,\ldots,N_t$:

1. Initialization at time step $t = 0$: select an initial set of parameters $a'_0, b'_{0,j}, e'_0, f'_0$ for $j = 1,\ldots,d$. In this step the initial parameters for the distributions $q(\boldsymbol{\alpha})$ and $q(\lambda)$ are selected.
2. At the generic $t+1^{th}$ iteration:
   - Update the parameters $\boldsymbol{\mu}'$ and Σ' according to:

$$\Sigma'_{t+1} = (\text{Diag}(\frac{a'_t}{b'_{t,j}}) + \frac{e'_t}{f'_t}\sum_{k=p}^{l} \boldsymbol{x}_k \boldsymbol{x}_k^T)^{-1} \tag{5.80}$$

$$\boldsymbol{\mu}'_{t+1} = \Sigma'_{t+1}(\frac{e'_t}{f'_t})\sum_{k=p}^{l} y_k \boldsymbol{x}_k. \tag{5.81}$$

   - Update the parameters a', b'_j, e', f' according to:

$$a'_{t+1} = a_0 + \frac{1}{2} \tag{5.82}$$

$$b'_{t+1,j} = b_0 + \frac{1}{2}[\Sigma'_{t+1} + \boldsymbol{\mu}'_{t+1}\boldsymbol{\mu}'^T_{t+1}]_{j,j} \tag{5.83}$$

$$e'_{t+1} = e_0 + \frac{l-p+1}{2} \tag{5.84}$$

$$f'_{t+1} = f_0 + \frac{1}{2}\sum_{k=p}^{l}(y_k - \boldsymbol{x}_k^T \boldsymbol{\mu}'_{t+1})^2 + \boldsymbol{x}_k^T \Sigma'_t \boldsymbol{x}_k. \tag{5.85}$$

   - Compute the value of the objective function $\mathscr{L}^t$ at the t^{th} iteration using Equation 5.31 and compare it with the value corresponding to the previous iteration.
3. Check convergence of $\mathscr{L}^t$ to a constant value. Once this convergence is achieved, that value of $\mathscr{L}^t$ corresponds to the final value of $\mathscr{L}$.

As the objective function $\mathscr{L}$ converges to a constant value, the auxiliary distribution $q(\boldsymbol{w}, \boldsymbol{\alpha}, \lambda)$ approaches the posterior distribution $p(\boldsymbol{w}, \boldsymbol{\alpha}, \lambda | \boldsymbol{X})$.

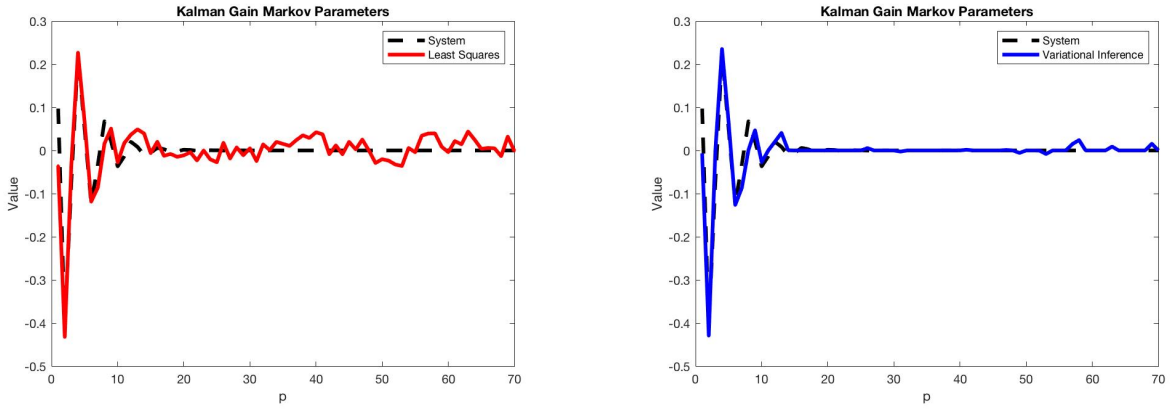
5.6. Numerical Simulations

A first comparison between the LS and Bayesian regression approaches is provided by analyzing a dynamic oscillator. The Kalman gain Markov parameters are extracted and compared. Secondly, the acceleration response time histories are generated by an 8-DOF shear type system. The coefficients inside of the matrix $\tilde{\Phi}$ are extracted through a conventional OLS regression according to Equation (5.23) and through Bayesian regression models according to the procedure just presented.

5.6.1. Dynamic oscillator

The system is represented by a linear dynamic oscillator. The spring element has stiffness equal to $k = 2500$ N/m and the mass is equal to 1 kg. The damping ratio has been set to $\xi = 1\%$. The excitation source is a gaussian zero mean force whose standard deviation σ is equal to 1 N. The force acts so to excite the oscillator horizontally. The measurement noise has been simulated by a gaussian white noise signal whose root mean square has been set at the 30% of the output response.

The system Kalman gain Markov parameters are provided in Figure 5.1. The theoretical Kalman Gain Markov parameters $\mathbf{CK}, \mathbf{CAK}, \ldots, \mathbf{CA}^p\mathbf{K}$ have been extracted by monitoring both the input $\boldsymbol{u}$ and the output $\boldsymbol{y}$ according to theoretical formulations in Appendix 8.1. The theoretical Kalman Gain Markov parameters have been than compared with those extracted by the linear regression computed through OLS, Figure 5.1 (a), and with those computer by the linear regression performed through Bayesian regression, Figure 5.1 (b).



(a) **(b)**

Fig. 5.1. *System Kalman Gain Markov parameters: comparison with OLS (a) and Bayesian (b) Kalman Gain Markov parameters.*

Figure 5.1 shows that the theoretical system Kalman gain Markov parameters $\mathbf{CK}, \mathbf{CAK}, \ldots, \mathbf{CA}^p\mathbf{K}$ converge to zero, whereas the ones obtained by OLS solution do not because of overfitting. Contrarily, the system Kalman gain Markov parameters obtained through Bayesian regression converge to zero as well as the theoretical ones.

It is interesting to note that, by increasing the level of measurement noise, the system Kalman gain Markov parameters computed by the linear regression models converge to the theoretical system Kalman gain Markov parameters. That fact can be explained by considering that the theoretical formulation of the Kalman filter relies on the assumption that the process noise $\boldsymbol{w}_p$ and the measurement noise $\boldsymbol{w}_m$ are uncorrelated each other. In absence of external disturbances, the covariance matrices of $\boldsymbol{w}_p$ and $\boldsymbol{w}_m$ are respectively $\mathbf{Q} = \mathbf{B}\mathbf{B}^T\sigma^2$ and $\mathbf{R} = \mathbf{D}\mathbf{D}^T\sigma^2$ and so correlated. By introducing external measurement disturbances (stationary and ergodic), the covariance matrix of $\boldsymbol{w}_m$ is given by $\mathbf{R} = \mathbf{D}\mathbf{D}^T\sigma^2 + \mathbf{R}_d$ where $\mathbf{R}_d$ is the covariance matrix of the external disturbance. It is obvious that, by increasing the level of measurement noise, the correlation between $\boldsymbol{w}_p$ and $\boldsymbol{w}_m$ is reduced and the system Kalman gain Markov parameters obtained by the regression models converge to the one obtained by the theoretical analysis.

5.6.2. 8-DOF shear type system

The system is an 8-DOF shear-type system. The mass and spring elements are numbered in ascending order from the bottom to the top (Figure 5.2). The system is characterized by horizontal springs of stiffness $k_i = 25000N/m$ $(i = 1, \ldots, 8)$, and each mass is equal to $m_i = 1kg$ $(i = 1, \ldots, 8)$. The frame is subjected to modal damping with a damping factor of $\xi = 1\%$ for each of the 8 vibration modes. The system is excited by an external force applied at the first DOF via zero-order-hold (ZOH) for 100 sec duration with a time sampling of 0.01 seconds. Such force is a zero-mean Gaussian white noise signal (standard normal distribution) whose magnitude is scaled by a factor of 100. The output dataset consists of the time-histories of the structural acceleration recorded at every level (full set of sensors) corrupted by different levels of Gaussian white noise, in terms of root mean square (RMS), to simulate the effect of measurement disturbances.

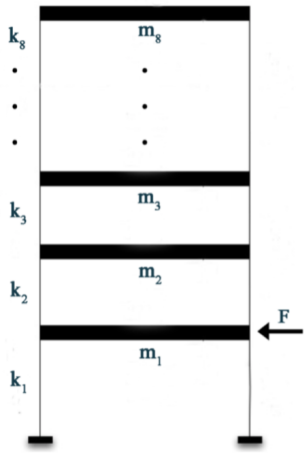


Fig. 5.2. *8-DOF shear type system*

For the 8-DOF system, the state space model order is $n = 16$, so the parameter p has been set equal to 70. The regression methods (OLS and Bayesian regression) have been used so to provide an estimate of the matrix $\mathbf{\Phi}$ in the core equation of the O^{3}KID algorithm. Once the matrix $\mathbf{\Phi}$ has been estimated, and the unwrapping procedure has been used to generate the system Kalman gain Markov parameters, the ERA algorithm can be used to estimate the system matrices $\mathbf{A}$ and $\mathbf{C}$. The extraction of these matrices relies the singular value decomposition (SVD) of the Henkel matrix of the system Kalman gain Markov parameters: the non-zero singular values generated by the SVD identify the order of the system and, consequently, of the system matrix $\mathbf{A}$. If the dynamic response in the dataset has been generated from the system subjected to external disturbances, distinguish the non-zero singular values may result in a challenging task. Since the system is an 8-DOFs, the order of the system in its state space representation is $n = 16$. According to the system order, only the largest 16 singular values have been considered for the extraction of the system matrix $\mathbf{A}$.

At this point, one important consideration has to be made. By considering a linear transformation (matrix), it is well known the direct connection between the singular values and the eigenvalues. Analogously to the eigenvalues, the singular values are representative of the entity of the transformation of their associated eigenvectors. Let's consider again the Henkel matrix of the system Kalman gain Markov parameters for a state space system of order $n = 16$. The 16 largest singular values are supposed to be associated to the modal dynamic response, anyway, some of them may be both corrupted by measurement noise or associated to the mathematical representation of the noise. In conclusion, due to measurement noise, the singular values associated to the modal dynamic response might present lower values than the first 16.

Three cases are presented for different values of the RMS of the gaussian white noise: 0%, 5% and 10% of the acceleration time histories RMSs. $N_{ob} = 50$ observations of the acceleration response time histories have been generated, for each case, so to extract $N_{ob} = 50$ observations of the natural frequencies and their standard deviations. The first case is a measurement noise free test (RMS=0%). The identified natural frequencies and the relative standard deviations are reported in Tables 5.1 and 5.2 respectively.

1. $0\%Noise$

State	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8
True	4.4054896	13.0664453	21.2824388	28.7736844	35.2850766	40.594878	44.5222694	46.9335083
ERA	4.4046614	13.0626293	21.2844583	28.7842344	35.2771006	40.5805771	44.4945392	46.9202584
ERA-dc	4.4046614	13.0626293	21.2844583	28.7842344	35.2771006	40.5805771	44.4945392	46.9202584
VI-ERA	4.4099553	12.9121514	21.0469581	28.6682095	35.2538091	40.6605491	44.4832105	46.9363397
VI-ERA-dc	4.4099553	12.9121514	21.0469581	28.6682095	35.2538091	40.6605491	44.4832105	46.9363397

Table 5.1. *8-DOFs: identified natural frequencies, no noise.*

State	std_1	std_2	std_3	std_4	std_5	std_6	std_7	std_8
True	0	0	0	0	0	0	0	0
ERA	0.0186575	0.0248458	0.0359533	0.0471506	0.0540776	0.0578433	0.0737128	0.0756137
ERA-dc	0.0186575	0.0248458	0.0359533	0.0471506	0.0540776	0.0578433	0.0737128	0.0756137
VI-ERA	0.0146066	0.088025	0.1218259	0.0916647	0.1016702	0.1678594	0.2347524	0.1594731
VI-ERA-dc	0.0146066	0.088025	0.1218259	0.0916647	0.1016702	0.1678594	0.2347524	0.1594731

Table 5.2. *8-DOFs: identified natural frequencies' standard deviation, no noise.*

In absence of noise the effects of the overfitting are not considered and the conventional OLS regression model provides better estimations than the Bayesian regression model.

By introducing a 5% RMS output noise, the problem due to overfitting appears. As shown by the Tables 5.3 and 5.4 the first natural frequency at 4.405 Hz cannot be identified by the OLS regression considering only 16 singular values, but more than 16 singular values need to be considered. Contrarily, the Bayesian regression is still able to identify the first natural frequency even though results are not as accurate as those in Tables 5.1 and 5.2.

2. $5\%Noise$

State	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8
True	4.4054896	13.0664453	21.2824388	28.7736844	35.2850766	40.594878	44.5222694	46.9335083
ERA	12.8754566	21.1565444	28.5426141	33.7359569	35.6370032	40.4416679	44.8922532	45.9223153
ERA-dc	12.8754566	21.1565444	28.5426141	33.7359569	35.6370032	40.4416679	44.8922532	45.9223153
VI-ERA	4.5211871	13.2938669	21.3115372	28.7928925	35.5220416	40.9760416	44.4419339	45.9708162
VI-ERA-dc	4.5211871	13.2938669	21.3115372	28.7928925	35.5220416	40.9760416	44.4419339	45.9708162

Table 5.3. *8-DOFs: identified natural frequencies, 5% measurement noise.*

State	std_1	std_2	std_3	std_4	std_5	std_6	std_7	std_8
True	0	0	0	0	0	0	0	0
ERA	1.7689673	1.1646737	1.0488671	2.9301714	0.4238344	0.6897958	1.5655008	1.2764975
ERA-dc	1.7689673	1.1646737	1.0488671	2.9301714	0.4238344	0.6897958	1.5655008	1.2764975
VI-ERA	1.3584474	1.4971053	0.0604535	0.0631223	2.2320067	2.617175	2.0481248	1.1296203
VI-ERA-dc	1.3584474	1.4971053	0.0604535	0.0631223	2.2320067	2.617175	2.0481248	1.1296203

Table 5.4. *8-DOFs: identified natural frequencies' standard deviation, 5% measurement noise.*

Finally, a 10% RMS output noise has been used as measurement noise. Again, by considering only 16 singular values, the conventional OLS regression model is not able to correctly identify the first natural frequency and a stabilization diagram is required.

3. 10%*Noise*

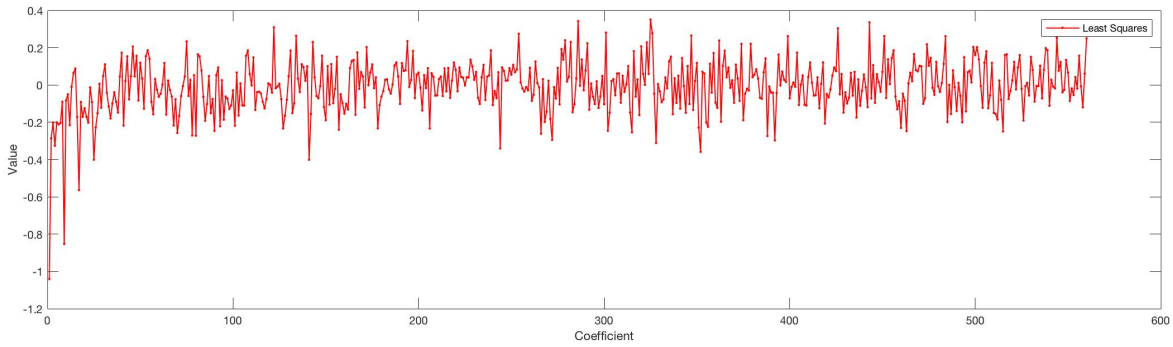
State	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8
True	4.4054896	13.0664453	21.2824388	28.7736844	35.2850766	40.594878	44.5222694	46.9335083
ERA	10.1868155	19.1090574	26.8000132	32.5106314	35.4456892	40.2191345	43.6276533	45.9906089
ERA-dc	10.1868155	19.1090574	26.8000132	32.5106314	35.4456892	40.2191345	43.6276533	45.9906089
VI-ERA	4.3539136	13.0815869	21.2980052	28.7850883	35.2717249	40.5541808	44.4473212	46.7643273
VI-ERA-dc	4.3539136	13.0815869	21.2980052	28.7850883	35.2717249	40.5541808	44.4473212	46.7643273

Table 5.5. *8-DOFs: identified natural frequencies, 10% measurement noise.*

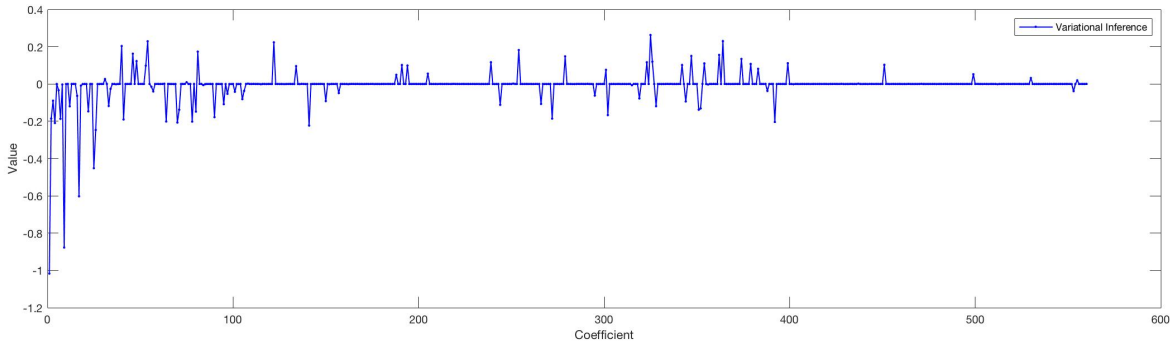
State	std_1	std_2	std_3	std_4	std_5	std_6	std_7	std_8
True	0	0	0	0	0	0	0	0
ERA	5.2597368	3.7625974	3.2829824	2.3986221	0.6933501	2.1061994	1.9021779	1.1614249
ERA-dc	5.2597368	3.7625974	3.2829824	2.3986221	0.6933501	2.1061994	1.9021779	1.1614249
VI-ERA	0.5350292	0.0327529	0.0467729	0.067903	0.0756994	0.0702979	0.14049	0.3354919
VI-ERA-dc	0.5350292	0.0327529	0.0467729	0.067903	0.0756994	0.0702979	0.14049	0.3354919

Table 5.6. *8-DOFs: identified natural frequencies' standard deviation, 10% measurement noise.*

A comparison between the elements of the first row of the matrix $\mathbf{\Phi}$ which have been estimated through LS (a) and Bayesian (b) regression and shown in Figure 5.3. The result is consistent to the one provided in Figure 5.1 showing the system Kalman gain Markov parameters.



(a)



(b)

Fig. 5.3. *Estimated regression coefficients (first row of* $\mathbf{\Phi}$*)*

The estimation of the mode shapes extracted through the Bayesian regression, for the last case presented (10% RMS output noise), is shown and compared with the real mode shapes in Figure 5.4.

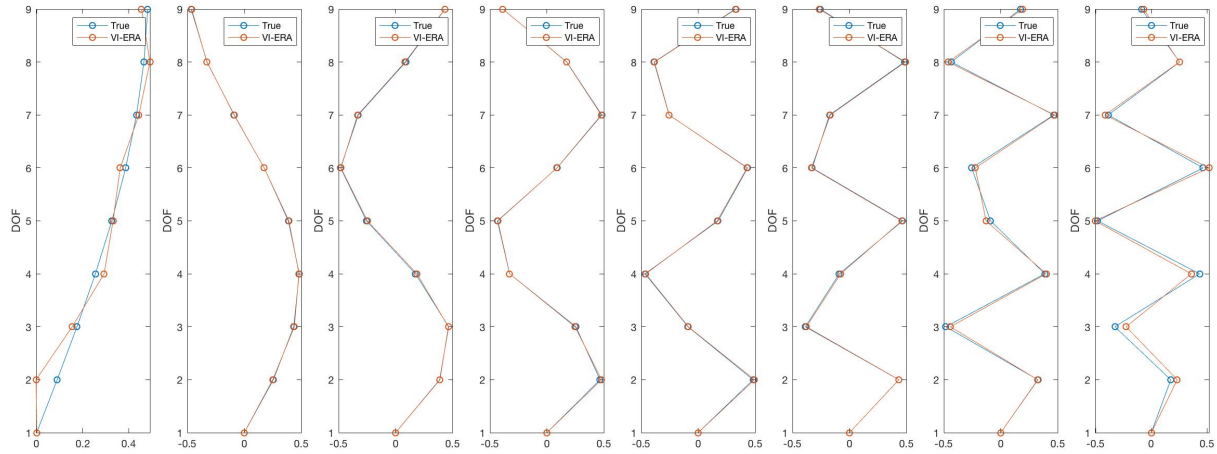


Fig. 5.4. *8-DOF shear type system mode shapes.*

In conclusion, the advantage of the Bayesian regression model over the conventional OLS

solution has been shown. The robustness to overfitting of the Bayesian model allows to correctly estimate the modal parameters relying on less singular values than those needed by the conventional OLS solution. In this context, by using a stabilization diagram to properly identify the modal parameters, the system identification algorithm relying on the Bayesian regression model allows to identify the stabilized natural frequencies for lower orders (number of singular values considered).

5.7. Conclusions

In this chapter, an alternative formulation of O^3KID relying on a Bayesian regression model has been presented. The advantages of a Bayesian regression model over the conventional OLS regression model has been shown in the numerical simulations subjected to different levels of measurement noise. Conventional OLS regression models are used to extract regression coefficients whose estimation is affected by overfitting problems in presence of external disturbances. The robustness of the Bayesian models to external noise has been proven and shown by means of two numerical simulations. In the first numerical simulation, it has been proven that the system Kalman gain Markov parameters extracted through the Bayesian approach approximate the theoretical values. In the second simulation, the benefits of the Bayesian regression model led to a better estimation of the modal parameters by considering low orders of the system model.

Chapter 6

6. Conclusions and future directions

In this thesis, output-only structural damage assessment methods based on both parametric and nonparametric models have been developed and presented.

The damage assessment methodology proposed in Chapter 2 is conducted within a statistical pattern recognition framework. In this context, two phases have to be defined: 1) a training phase, in which data are collected from a structure assumed to operate in its standard conditions, so to obtain a training model, 2) a test phase, in which new data are collected and tested, so to assess the presence of damage following a novelty detection approach. A training model is developed based on the statistical distribution of the damage sensitive features extracted from the training dataset. The first part of this work consists in the formulation of the cepstral coefficients as functions of the structural parameters and in their adaptation as damage sensitive features in SHM. Cepstral coefficients represent valuable damage sensitive features since they are able to provide a compact representation of data requiring a low user expertise. The damage assessment algorithm using cepstral coefficients has been implemented based on simulated data and a comparison of the results with those obtained using AutoRegressive coefficients is provided. Compared with the AutoRegressive coefficients, the cepstral coefficients require less computational time and are stable to the presence of nonlinearities in the data. The analytical formulation of the cepstral coefficients showed that any term of the cepstral coefficient sequence can be written as the sum of two terms, one which does not vary over the monitoring location (it is the same at any monitored location) and another term which is characterized by the local dynamic response. The variability induced by the second term can be reduced through a PCA process. The application of the PCA to generate a nonparametric model represents a relevant part of this study since it allows us to reduce the variance of data collected at different sensors and subjected to external noise, different excitations and environmental conditions. The performance of the damage assessment method has been presented through numerical and experimental tests. In particular, the analysis of the data collected from the Z24 bridge in Switzerland presents an enlightening comparison between modal parameters analysis and statistical pattern recognition, between local and global cepstral analysis, in a context in which PCA is extremely useful to reduce effects of external disturbances.

The implementation of an output-only damage assessment algorithm based on the extraction of modal parameters solely based on the knowledge of the structural dynamic response has been investigated in chapter 3. The Inner Product Vector (IPV) based methodology has been adapted to output-only analysis so that an effective estimation of the structural modes is obtained through the cross-correlation of manipulated dynamic signals. The proposed damage assessment strategy has been shown effective to assess the presence of structural damage and its location within the structure. The robustness of the methodology to measurement noise has been tested through numerical simulations and its performance validated on experimental data measured on a steel frame at the Los Alamos National Laboratory. The presented damage assessment algorithm is particularly effective for those cases in which a large amount of sensors is used to monitor the structure's dynamic response: the larger the number of sensors used, the better the resolution of the damage location. However, one of the main limitations of the proposed IPV-based approach is that it still requires the manual estimation of the structural modes' contributions to the frequency content of the dynamic response and this represents a critical point in an automatized damage assessment algorithm since it depends on the user expertise. For such a reason, in chapter 4, the implementation of an algorithm that automatizes this procedure has been presented.

In order to implement an automated version of the IPV method, chapter 4 focuses entirely on the design and implementation of a clustering Bayesian algorithm and of a classifier that is able to automatically distinguish the contribution of specific mode to the dynamic response spectra before and after the occurrence of damage. The contributions of different structural modes to the spectral content have been interpreted as statistical distributions and, consequently, the spectra as mixtures of these statistical distributions. Numerical simulations proved that a Variational-Inference based approach allows to automatically detect the structural mode's contributions to the spectra and to extract some features that characterize each of the modes so to recognize the same mode before and after the occurrence of damage. This methodology is particularly effective for those cases in which, because of damage, new structural modes contribute to the dynamic response of the system. Once this procedure is implemented, it becomes possible to assess the presence and location of damage through the IPV-based damage assessment algorithm previously developed. The proposed methodology has been tested and validated through the analysis of both numerical and experimental data.

Finally, in chapter 5, the data overfitting problem affecting regression algorithms at the

core of conventional system identification algorithms is addressed. The Ordinary Least Squares (OLS) method is conventionally used to extract the parameters of regression models from which the structural modal parameters are estimated. The reasons for using OLS are associated to its simplicity in the implementation and to the fact that it represents the Best Linear Unbiased Estimator (BLUE). In this chapter it has been confirmed that, according to the theory, OLS might lead to overfitting problems when the order of the regression model is large. In this framework, Variational Inference regression models represent a valid alternative to the conventional OLS because of their robustness to overfitting. A Bayesian Variational Inference regression algorithm has been designed and implemented on O^3KID and the benefits of such integration have been shown in numerical examples. Although the estimation of the modal parameters through the Bayesian approach is less accurate than the one provided by the OLS approach in absence of external disturbances, when measurement noise is introduced in system, the Bayesian approach provides better estimates. This result led us to consider the newly developed Bayesian system identification algorithm as a valid alternative to the conventional algorithm when dealing with noisy measurements.

The adaptation of the cepstral coefficients to SHM represents one of the first approaches to the field of *domain adaptation*. In this thesis, cepstral coefficients, features largely used in acoustics, have been successfully adapted to structural engineering as damage sensitive features for structural damage assessment purpose. Along this line, future research could be conducted to develop damage assessment algorithms that explore the applicability of new damage sensitive features commonly used in other fields. For example, *transfer learning* is a particularly interesting topic in the field of SHM, where most of datasets obtained during operational conditions are used as representative of the healthy conditions and few information about the system in its damaged condition are available. In this framework, transfer learning consists in learning the procedure for the extraction of damage sensitive features based on data provided by a source domain and in transferring and adapting the learnt damage sensitive features to the target domain.

In the analysis of the Z24 bridge, the nonparametric model has been obtained through PCA based on datasets presenting similar external conditions, hence, the acceleration response time histories considered have been recorded during the summer period. For this reason, future research could be conducted so to develop a damage assessment algorithm based on the cepstral coefficients able to filter more effectively the fluctuations of the structural properties due to dif-

ferent environmental conditions. Furthermore, PCA requires the use of large training datasets and the computation of the covariance matrix of data might require a significant computational effort, thus future research can be focused on the seeking of alternative methods for the damage sensitive features extraction.

7. Bibliography

References

[1] C. R. Farrar and K. Worden, *Structural health monitoring: a machine learning perspective*. John Wiley & Sons, 2012.

[2] K. Worden, C. R. Farrar, G. Manson, and G. Park, "The fundamental axioms of structural health monitoring," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 463, no. 2082, pp. 1639–1664, 2007.

[3] H. Sohn, C. R. Farrar, F. M. Hemez, and J. J. Czarnecki, "A review of structural health review of structural health monitoring literature 1996-2001." Los Alamos National Laboratory, Tech. Rep., 2002.

[4] J. M. Brownjohn, P. Q. Xia, H. Hao, and Y. Xia, "Civil structure condition assessment by FE model updating:: methodology and case studies," *Finite elements in analysis and design*, vol. 37, no. 10, pp. 761–775, 2001.

[5] M. Friswell and J. E. Mottershead, *Finite element model updating in structural dynamics*. Springer Science & Business Media, 2013, vol. 38.

[6] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*. Springer, 2013, vol. 112.

[7] P. R. Norvig and S. A. Intelligence, *A modern approach*. Prentice Hall, 2002.

[8] J. L. Rojo-Álvarez, M. Martínez-Ramón, A. R. Figueiras-Vidal, A. García-Armada, and A. Artés-Rodríguez, "A robust support vector algorithm for nonparametric spectral analysis," *IEEE Signal Processing Letters*, vol. 10, no. 11, pp. 320–323, 2003.

[9] F. Burba, F. Ferraty, and P. Vieu, "k-Nearest Neighbour method in functional nonparametric regression," *Journal of Nonparametric Statistics*, vol. 21, no. 4, pp. 453–469, 2009.

[10] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.

[11] K. Worden, G. Manson, and N. R. Fieller, "Damage detection using outlier analysis," *Journal of Sound and Vibration*, vol. 229, no. 3, pp. 647–667, 2000.

[12] L. Balsamo, "Statistical pattern recognition based structural health monitoring strategies," Ph.D. dissertation, Columbia University, 2015.

[13] L. Balsamo, R. Betti, and H. Beigi, "A structural health monitoring strategy using cepstral features," *Journal of Sound and Vibration*, vol. 333, no. 19, pp. 4526–4542, 2014.

[14] K. Balafas and A. S. Kiremidjian, "Extraction of a series of novel damage sensitive features derived from the continuous wavelet transform of input and output acceleration measurements," in *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2014*, vol. 9061. International Society for Optics and Photonics, 2014, p. 90611L.

[15] C. R. Farrar, T. A. Duffey, S. W. Doebling, and D. A. Nix, "A statistical pattern recognition paradigm for vibration-based structural health monitoring," *Structural Health Monitoring*, vol. 2000, pp. 764–773, 1999.

[16] Y. Lei, A. Kiremidjian, K. Nair, J. Lynch, K. Law, T. Kenny, E. Carryer, and A. Kottapalli, "Statistical damage detection using time series analysis on a structural health monitoring benchmark problem," in *Proceedings of the 9th international conference on applications of statistics and probability in civil engineering*, 2003, pp. 6–9.

[17] M. Gul and F. N. Catbas, "Damage assessment with ambient vibration data using a novel time series analysis methodology," *Journal of Structural Engineering*, vol. 137, no. 12, pp. 1518–1526, 2011.

[18] E. Figueiredo, G. Park, C. R. Farrar, K. Worden, and J. Figueiras, "Machine learning algorithms for damage detection under operational and environmental variability," *Structural Health Monitoring*, vol. 10, no. 6, pp. 559–572, 2011.

[19] H. Sohn, K. Worden, and C. R. Farrar, "Statistical damage classification under changing environmental and operational conditions," *Journal of intelligent material systems and structures*, vol. 13, no. 9, pp. 561–574, 2002.

[20] C. R. Farrar and K. Worden, "An introduction to structural health monitoring," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 365, no. 1851, pp. 303–315, 2007.

[21] H. Sohn, D. W. Allen, K. Worden, and C. R. Farrar, "Structural damage classification using extreme value statistics," *J. Dyn. Sys., Meas., Control*, vol. 127, no. 1, pp. 125–132, 2005.

[22] B. P. Bogert, "The quefrency alanysis of time series for echoes; Cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking," *Time series analysis*, pp. 209–243, 1963.

[23] A. V. Oppenheim, "Superposition in a class of nonlinear systems." 1965.

[24] R. Randall and Y. Gao, "Extraction of modal parameters from the response power cepstrum," *Journal of sound and vibration*, vol. 176, no. 2, pp. 179–193, 1994.

[25] Y. Gao and R. Randall, "Determination of frequency response functions from response measurements-I. Extraction of poles and zeros from response cepstra," *Mechanical systems and signal processing*, vol. 10, no. 3, pp. 293–317, 1996.

[26] O. F. Tigli, "Estimation of zeros from response measurements," Ph.D. dissertation, Northeastern University, 2008.

[27] G. Zhang, R. S. Harichandran, and P. Ramuhalli, "Application of noise cancelling and damage detection algorithms in NDE of concrete bridge decks using impact signals," *Journal of Nondestructive Evaluation*, vol. 30, no. 4, pp. 259–272, 2011.

[28] L. Balsamo, R. Betti, and H. Beigi, "Structural damage detection using speaker recognition techniques," in *Proceedings of the 11th international conference on structural safety and reliability (ICOSSAR), Columbia University, New York, NY, USA*, 2013.

[29] R. Randall, J. Antoni, and W. Smith, "A survey of the application of the cepstrum to structural modal analysis," *Mechanical Systems and Signal Processing*, vol. 118, pp. 716–741, 2019.

[30] M. Civera, M. Ferraris, R. Ceravolo, C. Surace, and R. Betti, "The Teager-Kaiser Energy Cepstral Coefficients as an Effective Structural Health Monitoring Tool," *Applied Sciences*, vol. 9, no. 23, p. 5064, 2019.

[31] M. Azeez and A. F. Vakakis, "Proper orthogonal decomposition (POD) of a class of vibroimpact oscillations," *Journal of Sound and vibration*, vol. 240, no. 5, pp. 859–889, 2001.

[32] G. Kerschen, B. Feeny, and J. C. Golinval, "On the exploitation of chaos to build reduced-order models," *Computer Methods in Applied Mechanics and Engineering*, vol. 192, no. 13-14, pp. 1785–1795, 2003.

[33] S. Han and B. Feeny, "Enhanced proper orthogonal decomposition for the modal analysis of homogeneous structures," *Journal of Vibration and Control*, vol. 8, no. 1, pp. 19–40, 2002.

[34] V. Lenaerts, G. Kerschen, and J. C. Golinval, "Identification of a continuous structure with a geometrical non-linearity. Part II: Proper orthogonal decomposition," *Journal of Sound and vibration*, vol. 262, no. 4, pp. 907–919, 2003.

[35] A. M. Yan, G. Kerschen, P. De Boe, and J. C. Golinval, "Structural damage diagnosis under varying environmental conditions-part i: a linear analysis," *Mechanical Systems and Signal Processing*, vol. 19, no. 4, pp. 847–864, 2005.

[36] A. M. Yan, G. Kerschen, P. De Boe, and J. C. Golinval, "Structural damage diagnosis under varying environmental conditions-part II: local PCA for non-linear cases," *Mechanical Systems and Signal Processing*, vol. 19, no. 4, pp. 865–880, 2005.

[37] H. Sohn, "Effects of environmental and operational variability on structural health monitoring," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 365, no. 1851, pp. 539–560, 2007.

[38] U. Dackermann, W. A. Smith, M. M. Alamdari, J. Li, and R. B. Randall, "Cepstrum-based damage identification in structures with progressive damage," *Structural Health Monitoring*, vol. 18, no. 1, pp. 87–102, 2019.

[39] M. Gul and F. Catbas, "Statistical pattern recognition for structural health monitoring using time series modeling: Theory and experimental verifications," *Mechanical Systems and Signal Processing, 23*, pp. 2192–2204, 2009.

[40] V. Rothamsted, T. Lewis, and V. Barnett, "Outliers in statistical data," 1996.

[41] E. Cross, G. Manson, K. Worden, and S. Pierce, "Features for damage detection with insensitivity to environmental and operational variations," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 468, no. 2148, pp. 4098–4122, 2012.

[42] Y. Q. LIN and W. X. REN, "Stochastic state space model-based damage detection of engineering structures [j]," *Journal of Vibration Engineering*, vol. 6, 2007.

[43] A. Mosavi, D. Dickey, R. Seracino, and S. Rizkalla, "Identifying damage location under ambient vibrations utilizing vector autoregressive models on mahalanobis distances," *Mechanical Systems and Signal Processing*, vol. 26, no. 1, pp. 254–267, 2012.

[44] K. Worden, G. Manson, and D. Allman, "Experimental validation of a structural health monitoring methodology: Part I. Novelty detection on a laboratory structure," *Journal of sound and vibration*, vol. 259, no. 2, pp. 323–343, 2003.

[45] C. Surace, K. Worden, and G. Tomlinson, "A novelty detection approach to diagnose damage in a cracked beam," in *Proceedings-SPIE The International Society For Optical Engineering*. SPIE International Society for Optical, 1997, pp. 947–953.

[46] G. De Roeck, "The state-of-the-art of damage detection by vibration monitoring:the SIMCES experience," *Journal of Structural Control*, vol. 10, pp. 127–143, 2003.

[47] J. Maeck and G. De Roeck, "Description of Z24 benchmark," *Mechanical Systems and Signal Processing 17(1):127-131*, 2003.

[48] E. Reynders and G. De Roeck, "Continuous vibration monitoring and progressive damage testing on the Z24 bridge," *Encyclopedia of structural health monitoring*, 2009.

[49] E. Reynders and G. De Roeck, "Vibration-based damage identification:the Z24 benchmark," *Encyclopedia of earthquake engineering*, pp. 1–8, 2014.

[50] D. Sarason, *Complex function theory*. American Mathematical Soc., 2007, no. 49.

[51] K. Worden and G. Manson, "The application of machine learning to structural health monitoring," *Philosophical Transactions of the Royal Society A:Mathematical, Physical and Engineering Sciences*, pp. 515–537, 2006.

[52] M. E. Tipping, "Deriving cluster analytic distance functions from gaussian mixture models," *IET*, 1999.

[53] D. Ververidis and C. Kotropoulos, "Gaussian mixture modeling by exploiting the Mahalanobis distance," *IEEE transactions on signal processing*, vol. 56, no. 7, pp. 2797–2811, 2008.

[54] H. Sohn, C. Farrar, N. Hunter, and K. Worden, "Structural Health Monitoring using statistical pattern recognition techniques," *Transactions of the ASME 123:706-711*, 2001.

[55] P. Omenzetter and J. M. W. Brownjohn, "Application of time series analysis for bridge monitoring," *Smart Materials and Structures*, vol. 15, no. 1, p. 129, 2006.

[56] R. Shibata, "Selection of the order of autoregressive model by Akaike's Information Criterion," *Biometrika*, vol. 63, no. 1, pp. 117–126, 1976.

[57] D. Ververidis and C. Kotropoulos, "Information loss of the Mahalanobis distance in high dimensions: application to feature selection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 2275–2281, 2009.

[58] T. Rogers, K. Worden, R. Fuentes, N. Dervilis, U. Tygesen, and E. Cross, "A Bayesian non-parametric clustering approach for semi-supervised Structural Health Monitoring," *Mechanical Systems and Signal Processing*, vol. 119, pp. 100–119, 2019.

[59] K. Krishnan Nair and A. S. Kiremidjian, "Time series based structural damage detection algorithm using Gaussian mixtures modeling," *Journal of dynamic systems, measurement, and control*, vol. 129, no. 3, pp. 285–293, 2007.

[60] E. Figueiredo and E. Cross, "Linear approaches to modeling nonlinearities in long-term monitoring of bridges," *Journal of Civil Structural Health Monitoring*, vol. 3, no. 3, pp. 187–194, 2013.

[61] J. Kullaa, "Structural health monitoring under nonlinear environmental or operational influences," *Shock and Vibration*, vol. 2014, 2014.

[62] R. Langone, E. Reynders, and S. Mehrkanoon, "Automated structural health monitoring based on adaptive kernel spectral clustering," *Mechanical Systems and Signal Processing*, vol. 94, pp. 64–78, 2017.

[63] E. M. Tronci, "Damage sensitive features. From classic parameters to new indicators," Ph.D. dissertation, Sapienza University of Rome, 2019.

[64] E. J. Cross and K. Worden, "Approaches to nonlinear cointegration with a view towards applications in SHM," in *Journal of Physics: Conference Series*, vol. 305, no. 1. IOP Publishing, 2011, p. 012069.

[65] C. Surace and A. Bovsunovsky, "The use of frequency ratios to diagnose structural damage in varying environmental conditions," *Mechanical Systems and Signal Processing*, vol. 136, p. 106523, 2020.

[66] E. Figueiredo and E. Flynn, "Three-story building structure to detect nonlinear effects," *Report SHMTools data description*, 2009.

[67] M. R. Hernandez-Garcia, S. F. Masri, R. Ghanem, E. Figueiredo, and C. R. Farrar, "An experimental investigation of change detection in uncertain chain-like systems," *Journal of Sound and Vibration*, vol. 329, no. 12, pp. 2395–2409, 2010.

[68] E. Figueiredo, G. Park, J. Figueiras, C. Farrar, and K. Worden, "Structural health monitoring algorithm comparisons using standard data sets," Los Alamos National Lab.(LANL), Los Alamos, NM (United States), Tech. Rep., 2009.

[69] M. R. Hernandez-Garcia, S. F. Masri, R. Ghanem, E. Figueiredo, and C. R. Farrar, "A structural decomposition approach for detecting, locating, and quantifying nonlinearities in chain-like systems," *Structural Control and Health Monitoring*, vol. 17, no. 7, pp. 761–777, 2010.

[70] L. Wang, Z. Yang, T. Waters, and M. Zhang, "Theory of inner product vector and its application to multi-location damage detection," in *Journal of Physics: Conference Series*, vol. 305, no. 1. IOP Publishing, 2011, p. 012003.

[71] S. Wang, Q. Ren, and P. Qiao, "Structural damage detection using local damage factor," *Journal of Vibration and Control*, vol. 12, no. 9, pp. 955–973, 2006.

[72] I. Trendafilova and E. Manoach, "Vibration-based damage detection in plates by using time series analysis," *Mechanical Systems and Signal Processing*, vol. 22, no. 5, pp. 1092–1106, 2008.

[73] J. T. Kim and N. Stubbs, "Crack detection in beam-type structures using frequency data," *Journal of sound and vibration*, vol. 259, no. 1, pp. 145–160, 2003.

[74] L. Wang and Z. C. Yang, "Structural damage detection using inner product vector and low pass filter technique," in *Applied Mechanics and Materials*, vol. 204. Trans Tech Publ, 2012, pp. 2942–2946.

[75] M. Zhang, R. Schmidt, and B. Markert, "Structural damage detection methods based on the correlation functions," in *Proceedings of the 9th International Conference on Structural Dynamics*, 2014, pp. 2435–2441.

[76] K. Roy and S. Ray-Chaudhuri, "Fundamental mode shape and its derivatives in structural damage localization," *Journal of Sound and Vibration*, vol. 332, no. 21, pp. 5584–5593, 2013.

[77] F. Iezzi, D. Spina, and C. Valente, "Damage assessment through changes in mode shapes due to non-proportional damping," in *Journal of Physics: Conference Series*, vol. 628, no. 1. IOP Publishing, 2015, p. 012019.

[78] O. Salawu, "Detection of structural damage through changes in frequency: a review," *Engineering structures*, vol. 19, no. 9, pp. 718–723, 1997.

[79] S. W. Doebling, C. R. Farrar, M. B. Prime, and D. W. Shevitz, "Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: a literature review," Los Alamos National Lab., NM (United States), Tech. Rep., 1996.

[80] V. Dawari and G. Vesmawala, "Structural damage identification using modal curvature differences," *IOSR J Mech Civ Eng*, vol. 4, pp. 33–38, 2013.

[81] F. Vicario, M. Q. Phan, R. Betti, and R. W. Longman, "Output-only observer/Kalman filter identification (O3KID)," *Structural Control and Health Monitoring*, vol. 22, no. 5, pp. 847–872, 2015.

[82] R. Brincker and P. Andersen, "Understanding stochastic subspace identification," *Proceedings of the 24th IMAC, St. Louis*, vol. 126, 2006.

[83] F. Vicario, "OKID as a general approach to linear and bilinear system identification," Ph.D. dissertation, Columbia University, 2014.

[84] P. Avitabile, "Experimental modal analysis," *Sound and vibration*, vol. 35, no. 1, pp. 20–31, 2001.

[85] F. Vicario, M. Q. Phan, R. Betti, and R. W. Longman, "OKID as a unified approach to system identification," 2014.

[86] J.-N. Juang, M. Phan, L. G. Horta, and R. W. Longman, "Identification of observer/Kalman filter Markov parameters-Theory and experiments," *Journal of Guidance, Control, and Dynamics*, vol. 16, no. 2, pp. 320–329, 1993.

[87] Z.-X. Li and X.-M. Yang, "Damage identification for beams using ANN based on statistical property of structural responses," *Computers & structures*, vol. 86, no. 1-2, pp. 64–71, 2008.

[88] C. S. N. Pathirage, J. Li, L. Li, H. Hao, W. Liu, and P. Ni, "Structural damage identification based on autoencoder neural networks and deep learning," *Engineering Structures*, vol. 172, pp. 13–28, 2018.

[89] L. Wang, Z. Yang, and T. Waters, "Structural damage detection using cross correlation functions of vibration response," *Journal of Sound and Vibration*, vol. 329, no. 24, pp. 5070–5086, 2010.

[90] M. Morgantini and R. Betti, "The Inner Product Vector as an output-only cross-correlation-based feature to Structural Damage Assessment," *Journal of vibroengineering, in press*, 2020.

[91] D. Blei, R. Ranganath, and S. Mohamed, "Variational Inference: Foundations and Modern Methods," *Neural Information Processing Systems (NIPS) Tutorial*, 2016.

[92] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *Journal of the American statistical Association*, vol. 112, no. 518, pp. 859–877, 2017.

[93] R. Ranganath, S. Gerrish, and D. M. Blei, "Black Box Variational Inference," in *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, 2014.

[94] Y. D. Kim and S. Choi, "Bayesian binomial mixture model for collaborative prediction with non-random missing data," in *Proceedings of the 8th ACM Conference on Recommender systems*, 2014, pp. 201–208.

[95] A. Corduneanu and C. M. Bishop, "Variational Bayesian model selection for mixture distributions," in *Artificial intelligence and Statistics*, vol. 2001. Morgan Kaufmann Waltham, MA, 2001, pp. 27–34.

[96] S. Tompaidis and C. Yang, "Pricing American-style options by Monte Carlo simulation: Alternatives to ordinary least squares," *Journal of computational finance*, vol. 18, no. 1, 2014.

[97] N. Faber and R. Rajko, "How to avoid over-fitting in multivariate calibration–The conventional validation approach and an alternative," *Analytica Chimica Acta*, vol. 595, no. 1-2, pp. 98–106, 2007.

[98] C. W. Granger and Y. Jeon, "Dynamics of model overfitting measured in terms of autoregressive roots," *Journal of Time Series Analysis*, vol. 27, no. 3, pp. 347–365, 2006.

[99] Y. Gal and Z. Ghahramani, "Bayesian convolutional neural networks with Bernoulli approximate variational inference," *arXiv preprint arXiv:1506.02158*, 2015.

8. Appendix

8.1. AutoRegressive coefficients

Any time signal can be mathematically represented by an AR model of order p as follows:

$$x(k) = \sum_{j=1}^{p} a_j x(k-j) + e(k) \tag{8.1}$$

where $x(k)$ is the value of the signal at time instant $k\Delta T$, given the sampling interval ΔT, the j^{th} AutoRegressive coefficient a_j and the residual error at the k^{th} time step $e(k)$. The AR coefficients can be easily evaluated by regression methods, e.g. Ordinary Least Square. The definition of the order p represents the hardest challenge concerning the development of such a model. Obviously, the higher the order of the model, the better the approximation of the observed signal. Anyway, model orders too large should be avoided to prevent data overfitting. The Akaike Information Criterion (AIC) represents a valuable method commonly employed for the model order selection. Such method provides a model representing the best compromise between the one that best fits the data and the one that has the least number of parameters to estimate the data. The AIC for a model of order p is given by:

$$AIC(p) = n_s[ln(\sigma_p^2) + 1] + 2p \tag{8.2}$$

where n_s is the total number of data points to estimate, σ_p^2 is the mean of the sum of square residual errors, $e(k)$. The AIC value is evaluated for a range of possible orders (1-max_p), e.g. in the analysis of the 8-DOFs shear type system this range has been set to vary between 1 and 20. The optimal order is selected computing the AIC values and picking the smallest value greater than a prescribed threshold set by the user. In this thesis, the threshold has been set set at the 5% of the difference between AIC(1) and AIC(max_p). The results of the damage detection assignment when using AR models are largely affected by the user's expertise since he has to set max_p and the threshold arbitrarily.

8.2. The Riccati equation for the computation of the steady-state Kalman filter

The discrete-time linear-invariant state space representation of a dynamic system according to Equations (5.1) and (5.2):

$$\begin{aligned} \boldsymbol{x}(k+1) &= \mathbf{A}\boldsymbol{x}(k) + \mathbf{B}\boldsymbol{u}(k) \\ \boldsymbol{y}(k) &= \mathbf{C}\boldsymbol{x}(k) + \mathbf{D}\boldsymbol{u}(k) \end{aligned} \tag{8.3}$$

can be modified to account for the process and measurement noise:

$$\boldsymbol{x}(k+1) = \mathbf{A}\boldsymbol{x}(k) + \mathbf{B}\boldsymbol{u}(k) + \boldsymbol{w}'_p(k) \tag{8.4}$$

$$\boldsymbol{y}(k) = \mathbf{C}\boldsymbol{x}(k) + \mathbf{D}\boldsymbol{u}(k) + \boldsymbol{w}'_m(k) \tag{8.5}$$

according to Equations (5.3) and (5.4).

The basic assumptions are here summarized:

1. $\boldsymbol{w}'_p(k)$ is the process noise assumed to be Gaussian, zero-mean ($\mathbb{E}[\boldsymbol{w}'_p(k)] = 0$) and white with the covariance matrix ($\mathbb{E}[\boldsymbol{w}'_p(k)\boldsymbol{w}'^T_p(j)] = \mathbf{Q}'\delta(k-j)$) given $\delta(k-j) = \mathbf{I}$ for $k = j$ and $\delta(k-j) = 0$ for $k = j$;
2. $\boldsymbol{w}'_m(k)$ is the measurement noise assumed to be Gaussian, zero-mean ($\mathbb{E}[\boldsymbol{w}'_m(k)] = 0$) and white with the covariance matrix ($\mathbb{E}[\boldsymbol{w}'_m(k)\boldsymbol{w}'^T_m(j)] = \mathbf{R}'\delta(k-j)$);
3. the initial condition $\boldsymbol{x}(0)$ is unknown, only its mean $\bar{\boldsymbol{x}}(0) = 0$ and covariance matrix $\mathbf{P}_x(0)$ are known;
4. $\boldsymbol{x}(0)$, $\boldsymbol{w}'_p(j)$ and $\boldsymbol{w}'_m(k)$ are mutually uncorrelated for all j and k so that $\bar{\boldsymbol{x}}(0)$ and $\mathbf{P}_x(0)$ are independent of $\boldsymbol{w}'_p(j)$ and $\boldsymbol{w}'_m(k)$;
5. future inputs are uncorrelated from past outputs so that $\mathbb{E}[\boldsymbol{u}(k)\boldsymbol{y}^T(j)] = 0$ for $k-1 \geq j \geq 0$.

The equation of a generic linear time-varying observer can be written as:

$$\hat{\boldsymbol{x}}(k+1) = \mathbf{F}(k)\hat{\boldsymbol{x}}(k) + \mathbf{K}(k)\boldsymbol{y}(k) + \mathbf{H}(k)\boldsymbol{u}(k) \tag{8.6}$$

where $\hat{\boldsymbol{x}} \in \mathbb{R}^{n\times 1}$ is the observer vector and $\mathbf{F} \in \mathbb{R}^{n\times n}$, $\mathbf{K} \in \mathbb{R}^{n\times m}$ and $\mathbf{H} \in \mathbb{R}^{n\times r}$ are the

generic observer system matrices.

The state error can be computed as the difference between the real state value and its estimation as:

$$\boldsymbol{e}(k) = \boldsymbol{x}(k) - \hat{\boldsymbol{x}}(k) \tag{8.7}$$

so that:

$$\begin{aligned}
\boldsymbol{e}(k+1) &= \boldsymbol{x}(k+1) - \hat{\boldsymbol{x}}(k+1) \\
&= \mathbf{A}\boldsymbol{x}(k) + \mathbf{B}\boldsymbol{u}(k) + \boldsymbol{w}'_p(k) - \mathbf{F}(k)\hat{\boldsymbol{x}}(k) - \mathbf{K}(k)\boldsymbol{y}(k) - \mathbf{H}(k)\boldsymbol{u}(k) \\
&= [\mathbf{A} - \mathbf{K}(k)\mathbf{C}]\boldsymbol{e}(k) + [\mathbf{A} - \mathbf{F}(k) - \mathbf{K}(k)\mathbf{C}]\hat{\boldsymbol{x}}(k) \\
&\quad + [\mathbf{B} - \mathbf{K}(k)\mathbf{D} - \mathbf{H}(k)]\boldsymbol{u}(k) - \mathbf{K}(k)\boldsymbol{w}'_m(k) + \boldsymbol{w}'_p(k).
\end{aligned} \tag{8.8}$$

By taking the expectation of Equation (8.8), the unbiased and zero-mean estimation error can be computed as:

$$\begin{aligned}
\mathbb{E}[\boldsymbol{e}(k+1)] &= [\mathbf{A} - \mathbf{K}(k)\mathbf{C}]\mathbb{E}[\boldsymbol{e}(k)] + [\mathbf{A} - \mathbf{F}(k) - \mathbf{K}(k)\mathbf{C}]\mathbb{E}[\hat{\boldsymbol{x}}](k) \\
&\quad + [\mathbf{B} - \mathbf{K}(k)\mathbf{D} - \mathbf{H}(k)]\mathbb{E}[\boldsymbol{u}(k)].
\end{aligned} \tag{8.9}$$

Given that the process and measurement signals are zero-mean, i.e. $\mathbb{E}[\boldsymbol{w}'_m(k)] = \mathbb{E}[\boldsymbol{w}'_p(k)] = 0$. The estimation error is zero-mean only if:

$$\mathbf{F}(k) = \mathbf{A} - \mathbf{K}(k)\mathbf{C} \tag{8.10}$$

$$\mathbf{H}(k) = \mathbf{B} - \mathbf{K}(k)\mathbf{D} \tag{8.11}$$

$$\hat{\boldsymbol{x}}(k) = \mathbb{E}[\boldsymbol{x}(0)] = \hat{\boldsymbol{x}}(0). \tag{8.12}$$

So that

$$\begin{aligned}\mathbb{E}[\boldsymbol{e}(k+1)] &= [\mathbf{A}-\mathbf{K}(k)\mathbf{C}]\mathbb{E}[\boldsymbol{e}(k)] \\ &= [\mathbf{A}-\mathbf{K}(k)\mathbf{C}][\mathbf{A}-\mathbf{K}(k)\mathbf{C}]\ldots[\mathbf{A}-\mathbf{K}(k)\mathbf{C}]\mathbb{E}[\boldsymbol{e}(0)] \\ &= 0 \end{aligned} \tag{8.13}$$

given that $\mathbb{E}[\boldsymbol{e}(0)] = \mathbb{E}[\boldsymbol{x}(0) - \hat{\boldsymbol{x}}(0)] = 0$. Considering the assumptions in Equation (8.10), (8.11) and (8.12), Equation (8.8) becomes:

$$\boldsymbol{e}(k+1) = [\mathbf{A}-\mathbf{K}(k)\mathbf{C}]\mathbb{E}[\boldsymbol{e}(k)] - \mathbf{K}(k)\boldsymbol{w}'_m(k) + \boldsymbol{w}'_p. \tag{8.14}$$

It is possible to compute the covariance matrix $\mathbf{P}(k+1)$:

$$\begin{aligned}\mathbf{P}(k+1) &= \mathbb{E}[\boldsymbol{e}(k+1)\boldsymbol{e}^T(k+1)] \\ &= [\mathbf{A}-\mathbf{K}(k)\mathbf{C}]\mathbb{E}[\boldsymbol{e}(k)\boldsymbol{e}^T(k)][\mathbf{A}-\mathbf{K}(k)\mathbf{C}]^T \\ &\quad +\mathbf{K}(k)\mathbb{E}[\boldsymbol{w}'_m(k)\boldsymbol{w}'^T_m(k)]\mathbf{K}^T(k) + \mathbb{E}[\boldsymbol{w}'_p\boldsymbol{w}'^T_p] \end{aligned} \tag{8.15}$$

and, defining $\mathbb{E}[\boldsymbol{w}'_m(k)\boldsymbol{w}'^T_m(k)] = \mathbf{R}'$ and $\mathbb{E}[\boldsymbol{w}'_p\boldsymbol{w}'^T_p] = \mathbf{Q}'$, it is possible to write:

$$\mathbf{P}(k+1) = [\mathbf{A}-\mathbf{K}(k)\mathbf{C}]\mathbf{P}(k)[\mathbf{A}-\mathbf{K}(k)\mathbf{C}]^T + \mathbf{K}(k)\mathbf{R}'\mathbf{K}^T(k) + \mathbf{Q}'. \tag{8.16}$$

The Kalman Filter $\mathbf{K}(k)$ is selected so to minimize the expected value of the squared norm of $\boldsymbol{e}(k)$, which is equivalent to minimize the trace (sum of diagonal elements) of the matrix $\mathbf{P}(k)$

$$\mathbf{J}(k) = \mathbb{E}[\boldsymbol{e}^T(k)\boldsymbol{e}(k)] = trace\mathbb{E}[\boldsymbol{e}(k)\boldsymbol{e}^T(k)] = trace[\mathbf{P}(k)]. \tag{8.17}$$

Thus, the objective is to find $\mathbf{K}(k)$ which maximizes the $trace[\mathbf{P}(k+1)]$:

$$\frac{\delta[trace[\mathbf{P}(k+1)]]}{\delta\mathbf{K}(k)} = -2[\mathbf{A}-\mathbf{K}(k)\mathbf{C}]\mathbf{P}(k)\mathbf{C}^T + 2\mathbf{K}(k)\mathbf{R} = 0. \tag{8.18}$$

Solving Equation (8.18) for $\mathbf{K}(k)$:

$$\mathbf{K}(k) = \mathbf{A}\mathbf{P}(k)\mathbf{C}^T[\mathbf{R}' + \mathbf{C}\mathbf{P}(k)\mathbf{C}^T]^{-1}. \tag{8.19}$$

By plugging Equation (8.19) into Equation (8.16), and recalling that $\mathbf{P}(k) = \mathbf{P}^T(k)$, the discrete time algebraic Riccati equation can be obtained:

$$\mathbf{P}(k+1) = \mathbf{A}\mathbf{P}(k)\mathbf{A}^T - \mathbf{A}\mathbf{P}(k)\mathbf{C}^T[\mathbf{R}' + \mathbf{C}\mathbf{P}(k)\mathbf{C}^T]^{-1}\mathbf{C}\mathbf{P}(k)\mathbf{A}^T + \mathbf{Q}'. \tag{8.20}$$

Equations (8.19) and (8.20) can be iteratively used until the Kalman gain $\mathbf{K}(k)$ converges to the steady-state Kalman gain $\mathbf{K}$. This procedure has been followed to provide the theoretical formulation of the theoretical Kalman Gain Markov parameters in section 5.6.1.