Next-generation wind turbine tower modeling
Uncertainty quantification

Ivelina Dimitrova
Next-generation wind turbine tower modeling
Uncertainty quantification

MASTER OF SCIENCE THESIS

For the degree of Master of Science in Systems and Control at Delft University of Technology

Ivelina Dimitrova

January 13, 2016

Faculty of Mechanical, Maritime and Materials Engineering (3mE) · Delft University of Technology
The natural frequency of wind turbine towers can be more than 10% off from the designed frequency. This frequency is important for control because it sets bandwidth limits to some of the control loops. Improving the model will result in better control and load reduction that will in turn contribute to the decrease of the cost of wind energy.

The main focus of this project is on uncertainty quantification. Finding more accurate uncertainty quantification techniques will help determining the accuracy of models. Currently, the accuracy of the models is described based on asymptotic statistics theory. However, asymptotic uncertainty descriptions exhibit some problems. Variance results are not reliable for small numbers of measurement samples. Moreover, the true values of the parameters are present in the asymptotic variance expressions. As a result, the uncertainty of the model is often approximately quantified or unknown. When the uncertainty of the model is not known or approximately quantified the controllers have to be designed with conservative margins. Therefore, obtaining more reliable uncertainty descriptions will enable the use of less conservative margins, and thus more aggressive controllers. For this purpose, the bootstrap technique for accuracy quantification of model estimates is proposed.

Additionally, Operational Modal Analysis (OMA) is used to obtain an estimate of the natural frequency of the tower. With OMA a model of a system is identified based on the output data collected when the system is in operating conditions. The use of OMA eliminates the need to artificially excite the wind turbine tower by using artificial exciters. Another advantage of OMA over system identification approaches is that the input signal to the system is not measured and only the output vibration response of the tower is measured. In this project, the system identification method optimized Predictor Based Subspace Identification (PBSID\textsubscript{opt}) was adapted to work with output only data. Asymptotic variance expressions have been derived for the newly proposed OMA-PBSID\textsubscript{opt} method.

The identification procedure based on ambient excitation was tested on a simulation example and the results showed that it is effective. The uncertainty of the natural frequency and damping ratio of the identified system model has been quantified using both the asymptotic statistics and bootstrap techniques and a comparison of the results is made. It is shown that the bootstrap method outperforms the asymptotic variance approach in the sense that it is less sensitive to the number of measurements samples.
# Table of Contents

1 Introduction
   1-1 Background ................................................. 1
   1-2 Problem statement ......................................... 2
   1-3 Methodologies ............................................. 2
      1-3-1 Uncertainty assessment ................................. 2
      1-3-2 Wind turbine tower modeling ......................... 4
   1-4 Outline ................................................... 5

2 Identification
   2-1 Wind turbine tower modeling .............................. 7
   2-2 Operational modal analysis (OMA) ......................... 8
   2-3 The framework of Operational modal analysis .......... 8
   2-4 PBSID_{opt} method for Operational modal analysis .... 11
      2-4-1 Identification problem ................................. 11
      2-4-2 Notations ............................................... 13
      2-4-3 Data equations ......................................... 14
      2-4-4 Relation to the ARX model structure ................. 14
      2-4-5 Estimating the predictor Markov parameters ....... 15
      2-4-6 Recovery of the system matrices ..................... 15
   2-5 Concluding remarks ....................................... 17

3 Bootstrap .................................................... 19
   3-1 General motivation ........................................ 19
   3-2 The bootstrap principle ................................... 19
   3-3 A general bootstrap application .......................... 21
   3-4 Bootstrap for dynamic models and the model based resampling method ..... 23

Master of Science Thesis Ivelina Dimitrova
3-5 Computational issues .................................................. 24
3-6 Confidence regions calculation using bootstrap .......................... 26
  3-6-1 Confidence intervals generation using bootstrap tables .......... 26
  3-6-2 Confidence intervals based on bootstrap percentiles ............. 29
3-7 Bootstrap for PBSID_{opt} ............................................ 32
3-8 Concluding remarks .................................................. 32

4 Asymptotic variance .................................................. 35
  4-1 Asymptotic normality ............................................... 35
  4-2 Uncertainty propagation ........................................... 36
  4-3 Asymptotic variance of system matrices ........................... 37
    4-3-1 Notations ..................................................... 38
    4-3-2 Estimate of \( \Delta \tilde{X} \) ........................................ 39
    4-3-3 Asymptotic variance .......................................... 39
  4-4 Asymptotic variance of eigenvalues ................................ 40
    4-4-1 Analytic calculation of the eigenvalue sensitivity .......... 41
    4-4-2 Numerical calculation of the eigenvalue sensitivity ...... 42
  4-5 Asymptotic variance of the eigenfrequencies and damping ratios 43
  4-6 Confidence intervals .............................................. 44
  4-7 Asymptotic variance for PBSID_{opt} ................................ 44
  4-8 Concluding remarks ................................................ 44

5 Uncertainty assessment: a simulation example .......................... 47
  5-1 True system dynamics .............................................. 47
  5-2 Realization of the Identification ................................ 48
  5-3 Uncertainty assessment ............................................ 51
  5-4 Interpretations of the results .................................... 61
    5-4-1 Comparison between bootstrap and asymptotic variance approach 61
    5-4-2 Comparison between analytic and numerical calculation of eigenvalues covariance ............................................. 63
    5-4-3 Comparison between normal bootstrap distribution and bootstrap percentile methods for confidence intervals generation .......... 64
  5-5 Conclusions ....................................................... 64

6 Conclusions and recommendations ..................................... 65
  6-1 Conclusions .......................................................... 65
  6-2 Recommendations .................................................. 66

A Preliminaries ......................................................... 69
  A-1 Kronecker product .................................................. 69
  A-2 Singular value decomposition .................................... 69
  A-3 Eigenvalue decomposition ....................................... 70
  A-4 Hankel matrix ..................................................... 70
  A-5 Jacobian matrix ................................................... 71
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bibliography</td>
<td>73</td>
</tr>
<tr>
<td>Glossary</td>
<td>79</td>
</tr>
<tr>
<td>List of Acronyms</td>
<td>79</td>
</tr>
</tbody>
</table>
List of Figures

1-1 The natural frequency of the wind turbine tower can be more than 10% off from the designed using physical modeling principles. .................................................. 2

2-1 System configuration considered in output-only identification. With $P$ being the data generating system, $y(k)$ the measured output response, $u(k)$ the unknown ambient excitation and $e(k)$ a noise sequence .............................................. 9

2-2 State-space system configuration considered in output-only identification, with $y_k$ being the measured output and $x_k$ the unknown states. System described by matrices $A$ and $C$. $\Delta$ represents the backward-shift operator [1]. .................... 10

2-3 A block scheme of a system in a closed-loop configuration. $P$ is the LTI system and $C$ is the controller, $r(k)$ is the reference signal, $y(k)$ the output signal, $u(k)$ the input signal and $e(k)$ the noise signal. The input signal $u(k)$ is correlated with the noise signal $e(k)$ ................................................................. 12

3-1 Bootstrap process for estimating the standard error of the sample mean [2]. The bootstrap samples $(x^*_1, x^*_2, ..., x^*_B)$ are obtained from the original data set $x = (x_1, x_2, ..., x_N)$ by resampling with replacement. The bootstrap replications $x^*_B$ are calculated from the bootstrap samples. .................................................. 21

3-2 The concept of residual-based bootstrap for estimating the variance of an estimate. .................. 25

3-3 Methodologies for constructing confidence intervals with the bootstrap. .................. 26

3-4 Bootstrap normal distribution. ................................................................. 27

3-5 Critical values $z_{\alpha}$ of the standard normal distribution. ................................. 27

3-6 Bootstrap distribution that is not normal. ........................................... 28

3-7 Histogram of 1000 bootstrap replications of $\theta$. The solid line is the estimate $\hat{\theta}$. The dashed lines are the 5% and 95% percentiles of the bootstrap histogram. ... 30

4-1 Asymptotic variance calculation of the eigenfrequencies and damping ratios. ............... 37

5-1 Bode magnitude plot of the Linear Time-Invariant (LTI) second-order system, used for the simulation experiment. .................................................. 48
5-2 System configuration considered in output-only identification. With $P$ being the data generating system, $y(k)$ the measured output response, $u(k)$ the unknown ambient excitation and $e(k)$ a noise sequence.

5-3 Data set gathered from the 2nd order true system, $y_k$ is the output of the system, $u_k$ the input to the open-loop system, being a a zero-mean Gaussian white noise signal and $e_k$ is a Gaussian white noise, with $k = 1, \ldots, N$ and $N = 2000$, the sampling time is $\Delta t = 1$ s. Only the output signal $y_k$ is used for the identification.

5-4 The first 20 singular values of the matrix $\tilde{\Gamma}(f) X_{p,N}$ calculated with the OMA-PBSID$_{opt}$ identification method.

5-5 Eigenvalues of the identified system with the OMA-PBSID$_{opt}$ identification algorithm, denoted with ‘x’ and eigenvalues of the true system, denoted with ‘+'.

5-6 Estimated eigenvalues and their 99% confidence ellipsoids using $N = 2000$ data points. Identified eigenvalues of the system denoted with ‘x’ and eigenvalues of the true system, denoted with ‘+'. Estimated confidence ellipsoids represented with a plain line and approximated with Monte Carlo simulation plotted with dotted line.

5-7 Estimated eigenvalues and their 99% confidence ellipsoids using $N = 500$ data points. Identified eigenvalues of the system denoted with ‘x’ and eigenvalues of the true system, denoted with ‘+'. Estimated confidence ellipsoids represented with a plain line and approximated with Monte Carlo simulation plotted with dotted line.
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>PBSID\textsubscript{opt} subspace OMA method.</td>
<td>17</td>
</tr>
<tr>
<td>3-1</td>
<td>Bootstrap PBSID\textsubscript{opt}-OMA procedure for confidence interval calculation of parameter estimates.</td>
<td>33</td>
</tr>
<tr>
<td>4-1</td>
<td>Asymptotic variance procedure for confidence interval calculation of parameter estimates.</td>
<td>45</td>
</tr>
<tr>
<td>5-1</td>
<td>Simulation matrices. System matrices of the true system.</td>
<td>49</td>
</tr>
<tr>
<td>5-2</td>
<td>Natural frequency $f$ and damping ratio $\xi$ of the true system and the identified system, obtained with the OMA-PBSID\textsubscript{opt} identification method for the case of Chapter 5.</td>
<td>50</td>
</tr>
<tr>
<td>5-3</td>
<td>Modal parameters from Monte Carlo simulation with 1000 iterations. Where $\mu$ and $\sigma^2$ represent the mean and the variance of the 1000 estimates.</td>
<td>50</td>
</tr>
<tr>
<td>5-4</td>
<td>Covariance estimation and its standard deviation over 50 Monte Carlo simulations using the bootstrap method, asymptotic variance approach and Monte Carlo simulation with different number of measurement samples $N$. OMA-PBSID\textsubscript{opt} identification method with past window $p = 60$.</td>
<td>57</td>
</tr>
<tr>
<td>5-5</td>
<td>Covariance estimation and its standard deviation over 10 Monte Carlo simulations using the bootstrap method, asymptotic variance approach and Monte Carlo simulation with different number of measurement samples $N$. OMA-PBSID\textsubscript{opt} identification method with past window $p = 60$. The uncertainty is not estimated with the bootstrap method for the case with $N = 10000$ due to the increased computational cost.</td>
<td>58</td>
</tr>
<tr>
<td>5-6</td>
<td>Covariance estimation and its standard deviation over 10 Monte Carlo simulations, using the bootstrap method, asymptotic variance approach and Monte Carlo simulation with different number of measurement samples $N$. OMA-PBSID\textsubscript{opt} identification method with past window $p = 20$.</td>
<td>58</td>
</tr>
<tr>
<td>5-7</td>
<td>Confidence intervals estimation with the bootstrap method, asymptotic variance approach and Monte Carlo simulation using different number of measurement samples $N$. Average results over 10 Monte-Carlo simulations.</td>
<td>59</td>
</tr>
<tr>
<td>5-8</td>
<td>Confidence intervals estimation with the bootstrap method, using percentile and standard error methods for confidence intervals calculation. Average results over 10 Monte-Carlo simulations.</td>
<td>59</td>
</tr>
</tbody>
</table>
5-9 Covariance estimation and its standard deviation over 10 Monte Carlo simulations using the asymptotic variance approach with analytic and numerical calculation of the eigenvalues sensitivity. .......................... 60
Chapter 1

Introduction

1-1 Background

In recent years, there has been an increase of the use of wind turbine energy. Wind energy accounts for at least 8% of the total production capacity in Europe. By the end of 2014, the total cumulative installations stand at 128,751 MW. In addition to that, Europe targets to have renewable energy form at least 20% of its total energy production by the year 2020, of which 40 GW will be accounted for by off-shore wind turbines [3, 4]. However, the cost of wind energy is still high and must be decreased. This cost is among others related to the performance and loads of the wind turbine. Study shows that both, improvement of the performance and mitigation of loads can be achieved with better control. The improved turbine performance leads to increased power production. At the same time, lowering the loads on the wind turbine reduces the fatigue and maintenance cost of the turbine. However, the controller performance is limited by the accuracy of the model. In addition to that, when the accuracy is unknown or approximately quantified, controllers must be designed with conservative margins.

In this project, the considered case will be wind turbine towers. Typically, in the design stage a first principles method is employed to develop a detailed model of the system. However, the natural frequency of wind turbine towers can be more than 10% off from the designed frequency. This error can be caused by many different factors that are not considered in the design process. Such factors can be, for example, deviations of the actual material properties from the expected, differences in manufacturing, variations of the foundation characteristics, modeling assumptions and unmodeled sensor characteristics [5]. This frequency is important for control as it sets bandwidth limits to the pitch to generator speed loop. To avoid instability, the control system bandwidth should be smaller than the natural frequency of the wind turbine tower. Obtaining more accurate estimates of the natural frequency of the tower can improve the performance of the control system. Moreover, assigning a measure of the uncertainty of the model will enable the use of less conservative control methods, such as robust control. That way robust performance and stability of the control system can be achieved even in the presence of modeling errors. Therefore, improving the accuracy of the natural frequency
2 Introduction

Figure 1-1: The natural frequency of the wind turbine tower can be more than 10% off from the designed using physical modeling principles.

estimate will result in better control, that will in turn contribute to the decrease of wind energy cost.

1-2 Problem statement

In this project, the limitation of the performance of the wind turbine system will be addressed in two ways. First, a more accurate wind turbine tower model will be obtained using the latest techniques. Second, the accuracy of that model will be quantified with better and more accurate techniques. This accuracy quantification is the major part and contribution of this project, because it will enable the use of less conservative control methods, that will in turn improve the performance of the wind turbine. Thus, the objectives of this thesis can be summarized as follows:

- Identify a wind turbine tower model using improved identification methods.
- Develop better accuracy quantification techniques and apply them on the case.

1-3 Methodologies

In the following section, are described the challenges related to the wind turbine tower modeling and uncertainty quantification. Then we present promising methods for meeting the previously stated research objectives, given the challenges that arise.

1-3-1 Uncertainty assessment

System identification can be used to obtain a mathematical model of a system using measurement data from the real system. In principle, these measurements are corrupted with noise and the estimates always have some error. Other sources of modeling error can be,
for example, non-linearities of the system and unmodeled dynamics. Thus, after a model of a system has been identified its accuracy must be addressed. There are a number of representations of the estimate’s accuracy. For instance, the accuracy of the estimate can be represented in terms of its standard error or by constructing confidence intervals. Confidence regions are a useful tool, because they provide an interval with a probability of containing the true parameter, instead of a single estimate [6]. The size of this interval gives us an idea how close the identified estimate is to the true parameter.

In the controller design phase it is beneficial to take into account the uncertainty of the model. However, sometimes a measure of the uncertainty cannot be derived, or it is not sufficiently accurate. In such situations, engineers typically design the closed loop systems with gain margins of 3 or more combined with phase margins between 30 and 60 degrees [7]. Thus, resulting in a reasonable trade-offs between bandwidth and robustness. Nevertheless, this affects the performance of the designed controller, since a greater phase margin is considered than strictly necessary. Therefore, if the uncertainty is known, this would permit the design of a more aggressive controllers.

Typically, the uncertainty descriptions of the estimates are derived using asymptotic statistics theory. To quantify the uncertainty, this method is based on the asymptotic variance property of an estimate. Approximate expressions have to be derived for the calculation of the asymptotic variance. A portion of the research in this thesis focuses on the derivation of asymptotic variance expressions for the Operational Modal Analysis (OMA) - optimized Predictor Based Subspace Identification (PBSID\textsubscript{opt}) method.

Although asymptotic uncertainty descriptions techniques are widely used in practice, they do exhibit some problems. This type of methods relies on variance approximations, as a result variance results are not reliable for small number of measurement samples. This happens due to violation of the convergence property. The convergence property implies that the asymptotic variance converges to the parameters covariance matrix as the number of measurements samples is large [8]. For the case of wind turbine tower, the number of measurement samples available for identification and uncertainty assessment is limited. Moreover, deriving subspace uncertainty descriptions using this approach is often a complicated task and sometimes impossible. Hence, another approach to evaluation of model accuracy is necessary that is less dependent on the number of measurement samples and that does not involve derivation of difficult analytic expressions.

Therefore, a methodology for assigning a measure of the estimates accuracy will be investigated, that fulfills the following requirements:

- Less dependent on the number of measurement samples.
- Straightforward to apply.

The main contribution of this research is the development of a method for uncertainty quantification that fulfills these requirements. Such an alternative method that meets these specifications is the bootstrap method. It is a completely different approach that manages to overcome the issues related to the traditional approach based on asymptotic approximations. The bootstrap is a statistical data-based method for assigning a measure of the estimates
accuracy. This efficient approach is used to represent and propagate uncertainties in less number of measurement samples situations.

The performance of the bootstrap method and asymptotic variance approach will be tested. The performance will be reviewed for the main challenge that is generally faced when assigning a measure of wind turbine tower estimates, that is the limited amount of measurements samples, and the results will be compared.

1-3-2 Wind turbine tower modeling

In the design stage, before the wind turbine tower is built a first principles method is used to acquire a detailed model of the system. Alternatively, a more accurate and detailed model can be obtained by using system identification. System identification allows the user to obtain models from measurement data, without the requirement of any a priori knowledge.

When applying system identification, the traditional approach involves performing an identification experiment. The role of this experiment is to collect input-output measurement data from the system. To obtain accurate models, using system identification, the system has to be excited with a proper input signal [9, 8]. This input signal should excite the relevant system dynamics. For the wind turbine tower, the experiment has to be performed in closed-loop while the system is operating by either applying specially designed inputs or by using artificial exciters. However, artificially exciting such a large massive structure as the tower of the wind turbine, is in general, difficult and expensive to realize. In addition, external excitation may reduce the performance and as a result add additional costs [10]. Similarly, the design of specific input signals, also requires some knowledge for the system and affects the final result. Not properly designed input signal would not entirely excite the desired system dynamics [11].

Therefore, in this thesis OMA is employed to obtain an estimate of the natural frequency of the wind turbine tower. This approach is much easier to apply and also more attractive in terms of cost, while at the same time delivering estimates with outstanding accuracy [11, 12]. OMA is a technique which allows you to identify a model of a system on the basis of the output responses only [13, 14]. In OMA instead of artificially exciting the system, one relies on the natural loading that is acting on the system. Hence, the use of OMA eliminates the need to artificially excite the wind turbine tower using the stochastic excitation generated by the wind [15]. In addition, the input to the system does not have to be measured for OMA techniques, instead only the output vibration response of the tower is measured.

Based on the challenges arising from the wind turbine tower modeling, the following requirements for the identification method are formulated:

- Easier to apply for the considered case of wind turbine tower.
- Deliver more accurate estimate of the natural frequency of the tower than physical modeling principles.

To meet the presented requirements, the subspace system identification method PBSID\textsubscript{opt} was adapted to work with output only data. The PBSID\textsubscript{opt} algorithm is a subspace identification
method that is particularly appropriate for wind turbine tower modeling. On one hand, since it can be extended to work with Linear Parameter-Varying (LPV) systems. And on the other hand, because of its excellent performance when dealing with real data.

1-4 Outline

This thesis concerns development and research on new identification and uncertainty assessments techniques. Chapter 2 discusses the current wind turbine tower modeling techniques and presents the OMA as an alternative modeling method. Then the framework and principles of OMA are presented. The PBSID\textsubscript{opt} algorithm for OMA is introduced and explained in details.

In Chapter 3 the bootstrap method for assigning a measure of estimates accuracy is presented. The bootstrap technique is a data-based method that originates from statistics and it is mainly applied in situations when the asymptotic variance techniques deliver inaccurate results. The principle of bootstrapping is illustrated. And then it is explained how the bootstrap can be employed with the OMA-PBSID\textsubscript{opt} identification algorithm. Different techniques for confidence regions formation with the bootstrap are discussed, as well. In Chapter 4, asymptotic variance descriptions for the OMA-PBSID\textsubscript{opt} identification algorithm method are derived. Then we look at the way of confidence intervals calculation when using asymptotic statistics approach.

In Chapter 5 a comparison is made between the bootstrap and asymptotic variance approaches on a simulation study. An estimate of a system is identified using the OMA-PBSID\textsubscript{opt} method. To evaluate the performance of the two methods, confidence intervals of obtained estimates are calculated. Furthermore, the influence of the number of measurements samples available on the obtained uncertainty results is investigated.

Finally, in Chapter 6 the conclusions, and proposals for future research will be given.
Chapter 2

Identification

The role of this chapter is to investigate the challenges that are faced when identifying the natural frequency of the wind turbine tower and to suggest a method that takes them into account. First, the challenges that arise when identifying the natural frequency of the wind turbine tower are presented. After these problems have been introduced, the requirements for the identification procedure for estimation of the wind turbine tower natural frequency will be established. Following these requirements, the Operational Modal Analysis (OMA) technique will be presented as a tool for identification of the considered system. Then the proposal for the new identification method in the framework of OMA is introduced.

2-1 Wind turbine tower modeling

In this section, it is given a brief overview of the methods that are currently often used for identification of the wind turbine towers and the challenges that arise.

Usually, system identification is applied in order to obtain a more accurate model of the wind turbine tower than techniques based on physical modeling principles. System identification allows the user to obtain models from measurement data, without the requirement of any a priori knowledge. Two system identification approaches exist: Prediction Error Identification (PEI) and Subspace Model Identification (SMI).

PEI methods, first require the parametrization of a specific model structure, and subsequently the estimates of these parameters are obtained in an iterative way by minimizing a prediction error [16]. Whereas, in SMI framework a state space model is identified in a non-iterative manner by solving a number of simple linear algebra problems [9].

In practice, subspace identification is commonly used to obtain initial estimates for subsequent prediction error identification. In particular, SMI techniques are a preferred choice for modeling a wind turbine [5]. The main reason stems from the fact that, the wind turbine is a Multivariable Input, Multiple Output (MIMO) system of a high order. It is difficult to apply PEI to such type of systems. First, because it is not feasible to fit a specific model structure
in terms of physical parameters and disturbances. Second, identifying a high order MIMO system with PEI is likely to lead to a nonlinear optimization problem.

To obtain accurate model using system identification technique, the system has to be excited with a proper input signal. For the wind turbine tower, this is done by either applying specially designed inputs or by using special devices to artificially excite the system. If the latter is performed, then one has to determine where exactly to place these devices, called exciters. Because not properly situated exciters, may not excite all modes of the system. Thus, leading to a poor data set [9, 8]. Moreover, artificially exciting such a massive structure as the tower of the wind turbine, is in general, difficult and expensive to realize. The challenge related to the other approach that relies on specially designed input signals, is that it requires some knowledge for the system in order to construct these signals.

Due to all this challenges, there is a need for new ways to tackle the problem of identifying the natural frequency of wind turbine towers. Based on these challenges, the following requirements for the identification method are formulated:

- Easier to apply for the considered case of wind turbine tower.
- Deliver more accurate estimate of the natural frequency of the tower than physical modeling techniques.

An identification technique that overcomes these problems is the OMA, which will be described in the following section.

### 2-2 Operational modal analysis (OMA)

OMA is a technique for deriving a model of a system on the basis of the output responses only. This data-driven modeling tool has emerged in the field of structural dynamics, where the unavailability of input force information disallows the use of traditional system identification techniques.

Many cases in the field of structural dynamics exist where it is difficult to apply an artificial input force and then one has to rely upon available ambient excitation sources. Often, it is practically impossible to measure this ambient excitation and the output measurements are the only information that can be passed to the system identification algorithm [17, 18]. Examples of structures where artificial excitation and determination of forces exhibits a problem, are wind turbines towers, bridges, towers, suspended roofs, vehicles [19, 15]. In such situations the systems are excited with natural sources, which allows model parameters to be estimated in operational conditions, hence the name OMA. The modal parameters are natural frequencies, damping ratios and mode shapes of the system.

In the next section, the framework that is considered in OMA will be presented.

### 2-3 The framework of Operational modal analysis

In this section, the framework for output-only identification problem will be considered.
It is assumed that the plant $P$ is excited from a source of an ambient input force $u(k_u)$ that is not measured, as shown in Figure 2-1. The problem of identifying a system in OMA can be formulated in the following way. On the basis of a finite set of output measurements $y_k$, obtained from a system excited with input signal $u(k_u)$, we wish to identify a modal model. The modal model consists of eigenfrequencies, damping ratios, mode shapes and modal participation factors. The system to be identified is assumed to be a finite dimensional, linear, time-invariant system, subject to measurement and/or process noise.

Since the input loading $u_k$ is not measured, then it is considered as being part of the noise sequence. Having said this, the system in state-space representation admits the following form [20]:

$$ x_{k+1} = Ax_k + w_k, \quad (2-1a) $$

$$ y_k = Cx_k + v_k, \quad (2-1b) $$

with system matrices $A \in \mathbb{R}^{n \times n}$, $C \in \mathbb{R}^{l \times n}$. $l$ represents the number of outputs and $n$ is the order of the system. The vectors $x_k \in \mathbb{R}^n$ and $y_k \in \mathbb{R}^l$ are the state vector and the output signal, respectively. The vectors $w_k \in \mathbb{R}^l$ and $v_k \in \mathbb{R}^l$ are zero mean, white vector processes, with covariance matrix:

$$ E[(w_p^T v_p^T) (w_p^T v_p^T)] = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix} \delta_{pq}, $$

where $E$ is the expected value operator and $\delta_{pq}$ is the Kronecker delta. The vector $v_k$ is the measurement noise, typically due to sensor inaccuracy, but here also due to the unknown excitation of the system $u_k$. The vector $w_k$, usually caused by disturbances and modeling inaccuracies, but in this case also due to the unknown excitation signal $u_k$ [14]. The state-space system is presented in Figure 2-2.

It is assumed that the input signal is a realization of a white noise, implying that the power spectra of the input signal is constant and has no poles or zeroes in the frequency range of interest. We also assume that the number of input samples $k_u$ approaches the number of output samples $k$ [18].

**Assumption 1.** It is assumed that the distribution of the input signal is a realization of a stochastic process.
Assumption 2. It is assumed that the input force is being applied on the entire output sequence $y_k \in \mathbb{R}^l$.

Therefore, in output-only identification a stochastic framework should be considered, instead of a deterministic for system identification. Since in these cases, the deterministic knowledge of the input is replaced by the assumption that the input $u_k$ is a realization of a stochastic process (white noise) leading to a random response $y_k$, which is a stochastic process [14].

Modal parameters

The modal parameters can be obtained from the system matrices $A$ and $C$. To calculate these parameters, we start by performing an eigenvalue decomposition of the state matrix:

$$A = \Psi \Lambda_d \Psi^{-1},$$

where $\Psi \in \mathbb{C}^{n \times n}$ is a matrix, containing the eigenvectors and $\Lambda_d \in \mathbb{C}^{n \times n}$ is a diagonal matrix, consisting of the discrete-time eigenvalues $\lambda_i$. The eigenfrequencies $\omega_i$ and the damping ratios $\xi_i$ can be found from the following expressions:

$$\omega = e^{\lambda_i \Delta t},$$

$$\lambda_i = -\xi_i \omega_i + j \sqrt{1 - \xi_i^2 \omega_i^2},$$

$$\lambda_i^* = -\xi_i \omega_i - j \sqrt{1 - \xi_i^2 \omega_i^2},$$

where with $\Delta t$ is denoted the sampling time. The mode shapes $V \in \mathbb{C}^{l \times n}$ are calculated according to the following formula:

$$V = C \Psi. \tag{2-3}$$

In this section, the problem of identifying a system when only relying on output data was formulated. In the next section, a novel OMA technique will be presented.
2-4 PBSID\textsubscript{opt} method for Operational modal analysis

Similarly to the system identification techniques, there is a variety of identification methods for OMA. Numerous OMA identification techniques in both time domain and frequency domain have been developed so far. The major developments of output-only identification in time domain include four time domain approaches, i.e. Natural Excitation Technique (NExT)-type [21, 22, 13], Auto-Regressive Moving Average (ARMA) model-based [13, 14], stochastic realization-based [23, 14, 13, 24, 25, 20] and stochastic subspace-based methods [26, 20, 25]. The most prominent output-only identification in frequency domain are the Peak Picking technique (PP) technique [27, 13, 14, 24, 19, 18], the Frequency Domain Decomposition (FDD) methods [28, 29, 14, 24, 30, 31, 32] and the output-only Polyreference Least Squares Complex Frequency domain (P-LSCF) methods [14, 33, 12, 34, 35].

The subspace techniques are the most well-known and used time-domain OMA techniques. Here, in this thesis a new subspace OMA method is presented. This is the optimized Predictor Based Subspace Identification (PBSID\textsubscript{opt}) system identification method, that is adopted to work with output-only data.

The PBSID\textsubscript{opt} is a subspace identification method that is very appropriate for wind turbine tower modeling. On one hand, because it can be extended to Linear Parameter-Varying (LPV) systems [36]. LPV systems are a very special class of nonlinear systems, whose parameters vary with time. On the other hand, because of its excellent performance when dealing with real data.

In Predictor-Based Subspace Identification (PBSID)\textsubscript{opt} algorithm a predictor $\hat{x}_k$ for the state sequence is estimated. The PBSID\textsubscript{opt} mixes the properties of traditional PEI and SMI methods, since also a high-order Vector Auto-Regressive with eXogeneous input (VARX) model is obtained. The estimated VARX model is then used to obtain an estimate of the state sequence $\hat{x}_k$. When an estimate of the state sequence is known, the system matrices can be found directly from two least-squares problems.

Before describing the PBSID\textsubscript{opt}-OMA method, in the following sections, the identification problem will be defined, the used notations together with the data equations of the underlying system are presented.

2-4.1 Identification problem

Let us consider a finite set of output measurements $\{y_k\}_{k=0}^{N-1}$ obtained from a system operating in either closed-loop or open-loop, given as:

\begin{align}
    x_{k+1} &= Ax_k + Bu_k + Ke_k, \\
    y_k &= Cx_k + Du_k + e_k,
\end{align}

with $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times r}$, $K \in \mathbb{R}^{n \times l}$, $C \in \mathbb{R}^{l \times n}$, $D \in \mathbb{R}^{l \times r}$. The vectors $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^r$ and $y_k \in \mathbb{R}^l$ are the state vector, input signal and output signal, respectively. The vector $e_k \in \mathbb{R}^l$ denotes the zero mean white innovation process with covariance matrix $\mathbb{E}\{e_k e_k^T\} = W \delta_{kk}$, with $W \succ 0$. The matrix $K$ is the Kalman gain. A block scheme of a closed-loop configuration
Identification problem is illustrated in Figure 2-3. We wish to identify a dynamic state-space model in form 2-2. Thus, our goal is to find an estimate of the order \( n \) of the system and the associated system matrices \((A, C, K)\) up to a global similarity transformation. The system to be identified is assumed to be a finite dimensional, linear, time-invariant system, subject to measurement and/or process noise. Therefore, the system admits the innovation state-space representation [20]:

\[
\begin{align*}
x_{k+1} &= Ax_k + Ke_k, \quad (2-5a) \\
y_k &= Cx_k + e_k. \quad (2-5b)
\end{align*}
\]

In order to obtain a model of the system, the system has to be reachable and observable [9]. Therefore, we assume the pair \((A, C)\) to be observable and the pair \((A, [B \ KW])\) to be reachable. And we also have to assume that all eigenvalues of \( A - KC \) are inside the unit circle, to allow consistent estimate of the innovation sequence.

System representation 2-5, may be rewritten in one-step-ahead predictor form, by eliminating the innovation signal \( e_k \):

\[
\begin{align*}
x_{k+1} &= \tilde{A}x_k + Ky_k, \quad (2-6a) \\
y_k &= Cx_k + e_k, \quad (2-6b)
\end{align*}
\]

where \( \tilde{A} \equiv A - KC \). The predictor model 2-6 sets the foundation of the PBSID\(_{\text{opt}}\) algorithm, which involves estimation of a predictor for the state sequence \( \tilde{x}_k \). And since an invertible linear transformation of the state \( x_k \) does not change the input output behavior of a state-space system, we can estimate the system matrices, up to a similarity transformation \( T \in \mathcal{R}^{n \times n}: T^{-1}AT, CT \) and \( T^{-1}K \). This will be explained in more detailed later, in Section 2-4-6.

Figure 2-3: A block scheme of a system in a closed-loop configuration. \( P \) is the LTI system and \( C \) is the controller, \( r(k) \) is the reference signal, \( y(k) \) the output signal, \( u(k) \) the input signal and \( e(k) \) the noise signal. The input signal \( u(k) \) is correlated with the noise signal \( e(k) \).

Problem 1. **Subspace identification problem in OMA framework.** Given a finite set of output data \( \{y_k\}_{k=0}^{N-1} \), obtained from a system in the form 2-4, estimate the order \( n \) of the discrete-time system 2-5 and the associated system matrices \( A, C, K \) up to a similarity transformation.
In this section, the problem of identifying a system in operating conditions using the output signal only was formulated. In the next section, the used notations together with the data equations of the underlying system are given.

### 2.4-2 Notations

Before deriving the data equations for the considered identification problem, we introduce some notations that will be used throughout this chapter. We begin with presenting a stacked vector that is constructed from the output data \( Z_k^{(f)} \) according to:

\[
Z_k^{(f)} = \begin{bmatrix} u_k \\ y_k \end{bmatrix}.
\]

The stacked vector \( Z_k^{(f)} \) is defined as:

\[
Z_k^{(f)} = \begin{bmatrix} Z_k^T, & Z_{k+1}^T, & \ldots, & Z_{k+f-1}^T \end{bmatrix}^T,
\]

where \( f \) represents the size of the 'future window', the size of the future window has to chosen such that \( f > n \). In a similar way, the following vectors are constructed: \( U_k^{(f)} \), \( Y_k^{(f)} \) and \( E_k^{(f)} \). Let us also define the extended controllability matrix \( \tilde{K}^{(p)} \) of the innovation model:

\[
\tilde{K}^{(p)} = \begin{bmatrix} \tilde{A}^{p-1} \overline{B}, & \tilde{A}^{p-2} \overline{B}, & \ldots, & \overline{B} \end{bmatrix},
\]

(2-7)

where we have defined \( \overline{B} = [K] \) for brevity.

The extended observability matrices, have full column rank, and are given by:

\[
\tilde{\Gamma}^{(f)} = \begin{bmatrix} C \\ C\tilde{A} \\ \vdots \\ C\tilde{A}^{f-1} \end{bmatrix}, \quad \Gamma^{(f)} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{f-1} \end{bmatrix}.
\]

Finally, we denote the Hankel matrices constructed from data sequences according to:

\[
Y_{i,s,N} = \begin{bmatrix} y_i & y_{i+1} & \cdots & y_{i+N-1} \\ y_{i+1} & y_{i+2} & \cdots & y_{i+N} \\ \vdots & \cdots & \cdots & \cdots \\ y_{i+s-1} & y_{i+s} & \cdots & y_{i+N+s-2} \end{bmatrix}.
\]

\( Y_{i,s,N} \) possesses \( s \) block rows and \( N \) columns. In the same way, the Hankel matrices constructed from the input sequence \( U_{i,s,N} \) and from the innovation sequence \( E_{i,s,N} \) are defined.
2-4-3 Data equations

Having introduced the notations required for the considered subspace identification method, next the data equations will be derived. The state equation in 2-5 can be written as:

\[ x_{k+p} = \tilde{A}^p x_k + \tilde{K}^{(p)} z_{k+p} \]  

(2-8)

Based on the derived state equation 2-8, the output equation at time \( k + p \) can written as:

\[ y_{k+p} = C \tilde{A}^p x_k + f + C \tilde{K}^{(p)} z_{k+p} + e_{k+p}. \]  

(2-9)

Assuming that \( \tilde{A} \) has its eigenvalues inside the unit circle, the term \( \tilde{A}^p \) can be made arbitrary small [37]. Which is achieved by selecting the past window \( p \) sufficiently large. This allows, the first term on the right-hand side of 2-8 and 2-9 to be neglected. Taking into account that too large size of the past window \( p \) results in a larger dimension of the matrices and thus increases the computational complexity and the number of parameters that need to be estimated. Based on this assumption, the output equation for \( y_k \) up to \( y_{N-1} \) results in:

\[ Y_{p,N} = C \tilde{K}^{(p)} Z_{0,p,N}^{(p)} + E_{p,N}, \]  

(2-10)

where we have defined \( N_p = N - p \) for brevity.

In PBSID_{opt} a high-order VARX model is obtained. The estimated VARX model is then used to obtain an estimate of the state sequence \( \hat{x}_k \). When an estimate of the state sequence is known, the system matrices can be found directly from two least-squares problems. We begin by presenting the relation of data equation to the Auto-Regressive with eXogeneous input (ARX) structure, which is outlined in the next subsection.

2-4-4 Relation to the ARX model structure

The aim of this subsection is to show the relationship between the predictor form 2-6 and ARX model structure. Since, this analogy provides the basis of the PBSID_{opt} method.

Note that the data equation 2-10 has a VARX structure:

\[ A(z)y = B(z)u_k + e_k, \]  

(2-11)

with \( z^{-1} \) the unit backshift operator and:

\[ A(z) = I - a_1 z^{-1} - \cdots - a_p z^{-p} \]

\[ B(z) = b_0 - b_1 z^{-1} - \cdots - b_p z^{-p}. \]

Based on this finding, and if we work with the predictor form 2-6, it can be concluded that a high-order VARX model is fully equivalent to the predictor model 2-6. The parameters \( a_i \) and \( b_i \) can be explicitly given as the Markov parameters of the predictor form 2-6:

Ivelina Dimitrova

Master of Science Thesis
2-4 PBSID\textsubscript{opt} method for Operational modal analysis

\begin{align}
  a_i &= C \tilde{A}^{i-1} K, \quad \text{for } i = 1...p, \quad (2-12a) \\
  b_i &= C \tilde{A}^{i-1} B, \quad \text{for } i = 1...p, \quad (2-12b) \\
  b_0 &= D. \quad (2-12c)
\end{align}

These insights serve as a foundation of the PBSID\textsubscript{opt} framework. Therefore, the first step in the PBSID\textsubscript{opt} algorithm is to derive an estimate for the predictor Markov parameters 2-12. This is done in a least squares fashion and is presented in the next section.

\section*{2-4-5 Estimating the predictor Markov parameters}

In the previous section, we showed that the state sequence can be approximated with a VARX structure. Thus, in order to find the predictor Markov parameters 2-12 a least-squares problem has to be solved. Referring back to the output equation 2-10 the following least-squares problem is constructed:

\begin{equation}
  \min_{[C\tilde{K}(p) \ D]} \left\| Y_{p,N_p} - \begin{bmatrix} C\tilde{K}(p) \\ D \end{bmatrix} \begin{bmatrix} Z_{0,p,N_p} \\ U_{p,N_p} \end{bmatrix} \right\|_F^2, \quad (2-13)
\end{equation}

where \( \| \cdots \|_F \) represents the Frobenius norm [38]. Observe that if the data matrix \([z_{0,p,N_p}^T, u_{p,N_p}^T]^T\) has full-rank, the least-squares solution can be found from an RQ decomposition of the data. It can be pointed out that the uniqueness of the parameter estimate requires the data matrix \([z_{0,p,N_p}^T, u_{p,N_p}^T]^T\) to be full rank. This requirement depends on the experimental data, reference signal and the nature of the feedback controller.

Once the Markov parameters have been determined, the algorithm continues with the estimation of the state-space matrices. The mechanism for deriving the system matrices is presented in the next section. First, the 'extended observability-times-controllability' matrix is constructed. Then the system matrices can be obtained by solving two least squares problems.

\section*{2-4-6 Recovery of the system matrices}

In PBSID\textsubscript{opt}, a predictor for the state sequence is constructed. By neglecting the first term from the state equation 2-8, it can be prescribed that the product \( \tilde{K}(p) Z_{0,p,N_p} \) represents the state sequence \( X_{p,N_p} \). However, this term cannot be estimated directly. Instead, we can use the estimate of the parameters \( C\tilde{K}(p) \) to construct the 'extended observability-times-controllability' matrix \( \tilde{\Gamma}(f)\tilde{K}(p) \). This matrix is constructed using a 'future' window \( f > n \), and has the following structure:

\begin{equation}
  \tilde{\Gamma}(f)\tilde{K}(p) \approx \begin{bmatrix} C\tilde{A}^{p-1} B & C\tilde{A}^{p-2} B & \ldots & C\tilde{B} \\ 0 & C\tilde{A}^{p-1} B & \ldots & C\tilde{A} B \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & C\tilde{A}^{f-1} B \end{bmatrix}. \quad (2-14)
\end{equation}
Note that, the first block row of \( \tilde{\Gamma}(f) \tilde{\mathcal{K}}(p) \) actually contains the estimated Markov parameters \( \mathcal{C} \). Each block row in 2-14 is obtained from the previous by shifting it and filling it with zeroes [37].

Having obtained the matrix \( \tilde{\Gamma}(f) \tilde{\mathcal{K}}(p) \) we can compute the product \( \tilde{\Gamma}(f) \tilde{\mathcal{K}}(p) Z_{0,p,N_p} \) which equals by definition the extended observability matrix times the state sequence \( \tilde{\Gamma}(f) X_{p,N_p} \). By computing a Singular-Value Decomposition (SVD) of this estimate it is possible to estimate the state sequence, as well as to determine the order of the system \( n \). The SVD are given by:

\[
\tilde{\Gamma}(f) X_{p,N_p} = \tilde{\Gamma}(f) \tilde{\mathcal{K}}(p) Z_{0,p,N_p},
\]

\[\text{(2-15)}\]

\[
\tilde{\Gamma}(f) X_{p,N_p} = U_n \Sigma_n V_n^T,
\]

\[\text{(2-16)}\]

where \( \Sigma_n \) is the diagonal matrix containing the \( n \) dominant singular values, chosen by the user, and \( V \) is the corresponding row space. We distinguish the dominant \( n \) singular values, that are disturbed by the system from the remaining singular values that are due to the noise by looking at a gap between the \( n \)th and the \((n + 1)\)th singular value [9]. It is possible to include a left weighing matrix in this equation. In such manner, the variance of the estimated state sequence and consequently of the resulting identified system can be affected. The choice of such weighing matrix has been discussed in [39] and [40].

The state sequence is estimated (up to the similarity transformation) as:

\[
\hat{X}_{p,N_p} = \Sigma_n V_n^T.
\]

\[\text{(2-17)}\]

With the knowledge of the state sequence, the matrices of the state-space model can then be calculated directly from two least-squares problems:

\[
X_{p+1,N_p-1} = \begin{bmatrix} A & K \end{bmatrix} \begin{bmatrix} X_{p,N_p-1} \\ E_{p,N_p-1} \end{bmatrix},
\]

\[\text{(2-18)}\]

\[
Y_{p,N_p} = C X_{p,N_p} + E_{p,N_p}.
\]

\[\text{(2-19)}\]

First, 2-19 is solved and then its residual \( \hat{E}_{p,N_p} \) is used in the solution of 2-18.

This ends the description of the PBSID_{opt} identification algorithm. A summary of the steps of PBSID_{opt} is presented in Table 2-1.

In conclusion, we should note that the PBSID_{opt} method is considered as one of the most appealing in the class of subspace identification algorithms [41]. This is true mainly because it is consistent under closed-loop operating conditions. It also performs well for open loop data. Additionally, PBSID_{opt} is associated with very simple and computationally attractive implementation via VARX modeling [42].
Algorithm: OMA-PBSID\textsubscript{opt}

1. Construct the matrices $Y_{p,N_p}$ and $Z_{0,p,N_p}$.
2. Solve the VARX problem in 2-13 using RQ decomposition.
3. Construct $\tilde{\Gamma}(f) \tilde{K}^p$ 2-14, using the estimated Markov parameters.
4. Determine the state estimate $\tilde{X}_{p,N_p}$ using 2-17.
5. Obtain an estimate of the system matrices $A$ $C$ and $K$
   by solving the two least squares problems 2-18 and 2-19.

\begin{table}[h]
\centering
\begin{tabular}{|l|}
\hline
Table 2-1: PBSID\textsubscript{opt} subspace OMA method. \\
\hline
\end{tabular}
\end{table}

\section{Concluding remarks}

In this chapter, the OMA identification technique has been motivated and presented as a suitable tool for wind turbine tower identification. The subspace system identification algorithm PBSID\textsubscript{opt} was adapted to work with output only data. The PBSID\textsubscript{opt} method is feasible, provides consistent estimates with real data and can be extended to work with LPV systems. Therefore, it is expected that using PBSID\textsubscript{opt} to perform OMA on the wind turbine tower will lead to the identification of more accurate models. This will be investigated in the last part of the report. Before the OMA is performed, in the following chapters the available uncertainty quantification techniques will be discussed.
Chapter 3

Bootstrap

This chapter presents the bootstrap method for uncertainty quantification. In order to obtain a proper understanding of the principle of bootstrapping, the theory behind the bootstrap will be discussed first. This will allow for a better understanding of its principles of operation and characteristics. Once the framework for uncertainty assessment with bootstrap has been established, the application of the bootstrap to the Operational Modal Analysis (OMA)-optimized Predictor Based Subspace Identification (PBSID$_{opt}$) identification method will be presented. Additionally, an individual section is dedicated to the different techniques for constructing confidence regions with bootstrap.

3-1 General motivation

The bootstrap technique is a computer-based method originating from statistics for assigning measures of accuracy to estimates [2]. Thus, quantities such as bias, variance and confidence intervals of the estimate can be calculated [43]. It was originally used to calculate the standard error of estimates. The bootstrap is a data-based simulation algorithm for estimating the standard error, meaning that no analytic calculations are required. Thus, it is mainly applied in situations where the number of observations are not large enough such that asymptotic results are inaccurate. In addition, the bootstrap is extremely useful when analytic expressions for the variance are mathematically difficult to derive. Apart for determining the accuracy of parameters, bootstrapping can be applied to model selection and signal detection - as well [44].

In the next section, the principle of bootstrapping is presented using an example.

3-2 The bootstrap principle

To explain the basic idea behind the bootstrap we will discuss the principle of bootstrapping. We introduce the principle of bootstrapping with a simple example concerning the accuracy of a sample mean.
Imagine that we want to determine what the average height is of the males in The Netherlands. An experiment was conducted, and 1000 random males were measured, the results were collected in the data sample $x_1, x_2, ..., x_N$, where the number of collected samples is $N = 1000$. The distribution $F$ of the males height is considered to be unknown. The sample mean equals:

$$\bar{x}_N = \frac{1}{N} \sum_{i=1}^{N} x_i.$$  \hfill (3-1)

We would like to know how close the estimate is to the model feature it is supposed to estimate. In the given example, as model features can be considered the model distribution function $F$ and the expectation $\mu$ corresponding to $F$ [6]. For this example, the model feature is defined as the expected value of the average males height in The Netherlands. What is the probability that the sample mean $\bar{x}_N$ and the true expectation of males height $\mu$ differ more than a given tolerance $\epsilon$? After all, the estimated parameters are based on a small sample. Additionally, the sample is only one possible realization of the random sample. If we would conduct another experiment, we would obtain a different sample mean as an estimate for $\mu$. Consequently, we decide to assign a standard error to the estimate of the sample mean, in order to account for the model deviation.

The bootstrap estimate of the standard error, first proposed by Efron [45], works in the following way. A bootstrap sample $x^* = (x^*_1, x^*_2, ..., x^*_N)$ is obtained by randomly resampling with replacement $N$ times the original data $x_1, x_2, ..., x_N$. It is assumed that the distribution of the bootstrap sample $F_N$ is close to the true unknown distribution $F$. Which indeed holds since each element of the new sample $x^*_i$ has probability $\frac{N-1}{N}$ of being equal to any of the elements of the original sample $x_i$. Since all further bootstrap algorithms are based on this assumption, we introduce it formally.

**Assumption 3.** It is assumed that the distribution of the bootstrap sample $F_N$ is close to the real distribution $F$.

In this particular case, we resample from the empirical distribution $F_N$, which is known to approach $F$ as $n \to \infty$ [46].

**Remark 1.** In practice, we do not directly estimate the empirical distribution $F_N$, since choosing one of the elements $x_1, x_2, ..., x_N$ of the original dataset with equal probability $1/n$ is equal to generating $x^*$ from $F_N$.

This resampling process is repeated a large number of times $B$, and consequently $B$ bootstrap samples $x^{*1}, x^{*2}, ..., x^{*B}$ are generated, each of size $N$.

The corresponding average, called the *bootstrapped sample mean* is then calculated for each $x^{*B}$:

$$\bar{x}^*_N = \frac{1}{N} \sum_{i=1}^{N} x^*_i.$$  \hfill (3-2)

The preceding procedure is called the bootstrap principle for the sample mean. A schematic representation of the bootstrap process for estimating the mean is presented on Figure 3-1.

Ivelina Dimitrova

Master of Science Thesis
Figure 3-1: Bootstrap process for estimating the standard error of the sample mean [2]. The bootstrap samples \((x^*1, x^*2, ..., x^*B)\) are obtained from the original data set \(x = (x_1, x_2, ..., x_N)\) by resampling with replacement. The bootstrap replications \((\bar{x}^1, \bar{x}^2, ..., \bar{x}^B)\) are calculated from the bootstrap samples.

Simply put, instead of performing a new experiment, with the bootstrap the original data is reused. A new data set is created and the estimate is recomputed. This procedure is done a large number of times.

Then the distribution and the variability of the mean can be obtained from the calculated \(B\) bootstrap approximations of the mean [47]. The estimate of the standard error is the standard deviation \(\hat{\sigma}_{boot}\) of the bootstrap distribution [2]:

\[
\hat{\sigma}_{boot} = \frac{1}{\sqrt{B-1}} \left( \sum_{i=1}^{B} \left( x^{*i} - \bar{x}^{(*)} \right)^2 \right)^{1/2}
\]

\[
\bar{x}^{(*)} = \frac{1}{B} \sum_{i=1}^{B} x^{*i},
\]

where \(\bar{x}^{(*)}\) is the bootstrap mean.

In this section we have explained the principle of bootstrapping. This routine serves as a basis of the bootstrap algorithm which is introduced in the next section, including its existing variants.

3-3 A general bootstrap application

In this section we introduce the basic bootstrap and provide a review of the basic bootstrapping techniques. We start with an example that shows how bootstrap can be used to evaluate the accuracy of an estimate. Then, we discuss the non-parametric and parametric bootstrap methods.

To understand how bootstrap works, a second example is presented. It is intended to show how bootstrap method can be used to find the variance \(\sigma_{\theta}^2\) of an estimator \(\hat{\theta}\) of \(\theta\). Suppose we
are given a random sample \( x = (x_1, x_2, \ldots, x_N) \) from an unknown probability distribution \( F \). Then, we compute an estimate \( \hat{\theta} \) of the parameter of interest \( \theta = t(F) \) on the basis of the available data \( x \). The bootstrap is used to evaluate the accuracy of the obtained estimate, in terms of its standard deviation \( \sigma^2_\hat{\theta} \).

In some cases it is possible to derive explicit expressions for \( \sigma^2_\hat{\theta} \), but often this is not mathematically feasible, or it is too complex. Alternatively, one usually uses asymptotic analysis to compute an estimate \( \hat{\sigma}^2_\hat{\theta} \) of \( \sigma^2_\hat{\theta} \). However, asymptotic approximations hold only for very large values of \( N \), as it was previously discussed. A way to overcome these problems is to use the bootstrap to approximate \( \sigma^2_\hat{\theta} \) by \( \hat{\sigma}^2_\hat{\theta} \). Convergence and consistency results for the bootstrap variance \( \hat{\sigma}^2_\hat{\theta} \) can be found in [48].

With the bootstrap algorithm, we use the original random sample \( x = (x_1, x_2, \ldots, x_N) \) and generate a new sample by sampling with replacement from \( x \). In this manner we create a number \( B \) of samples \( x^{*1}, \ldots, x^{*B} \). Then from each of the obtained samples, replication of the estimate \( \hat{\theta}^{*(i)} \) is computed. The derived bootstrap estimates \( \hat{\theta}^{*(i)} \) are then used to calculate the estimate of the standard error \( \hat{\sigma}^2_\hat{\theta} \). The standard error of a statistic can be estimated using the standard deviation \( \hat{\sigma}^2_\hat{\theta} \) of the bootstrap distribution \[2\]. The bootstrap distribution is the distribution of the many bootstrap estimates \( \hat{\theta}^{*(i)} \). Therefore, the bootstrap algorithm for variance estimation is as follows:

1. Conduct an experiment and collect a random data sample \( x = (x_1, x_2, \ldots, x_N) \).
2. Obtain an estimate \( \hat{F} \) of the unknown distribution \( F \). The estimate can be either parametric or non-parametric.
3. From the distribution \( \hat{F} \) of \( x \), draw (with replacement) \( B \) samples \( x^{*(i)}, i = 1, \ldots, B \) of size \( n \).
4. Evaluate the bootstrap estimate \( \hat{\theta}^{*(i)} \) from each sample \( x^{*(i)} \).
5. Estimate the variance \( \sigma^2_\hat{\theta} \) of \( \hat{\theta} \) by equation:

\[
\hat{\sigma}^2_\hat{\theta} = \frac{1}{B-1} \sum_{i=1}^{B} \left( \hat{\theta}^{*i} - \hat{\theta}^{*(\cdot)} \right)^2, \tag{3-4a}
\]

\[
\hat{\theta}^{*(\cdot)} = \frac{1}{B} \sum_{i=1}^{B} \hat{\theta}^{*i}. \tag{3-4b}
\]

Estimating the distribution \( \hat{F} \) of the unknown distribution \( F \) is the crucial step of the bootstrap process. Non-parametric estimate \( \hat{F} \) is preferred when no a priori information on distribution \( F \) can be made. In this case \( \hat{F} \) is defined as the empirical discrete distribution, which puts equal mass \( 1/N \) at each of the data points in sample \( x \). Suppose that some information about \( F \) is available, this suggests that we can use parametric approach for the estimation of \( F \), by assuming that it belongs to some specified class of probability distribution. It can

Ivelina Dimitrova
Master of Science Thesis
then be determined from the available data the mean $\mu$ and the variance $\sigma^2$ of $\hat{F}$. Once the estimate of the distribution $\hat{F}$ is calculated from $x$, we generate $B$ samples from $\hat{F}$. Parametric bootstrap generally gives better results compared to the non-parametric bootstrap [2, 43]. Specifically, when the number of samples is small [44]. This, however, holds if the estimates distribution is chosen correct. Hence, the non-parametric approach is typically more robust [44].

Until now, we have assumed that the data set $x$ is a collection of $N$ numbers drawn at random from a completely unspecified distribution $F$, except for the parametric bootstrap. In such case, the data set $x$ is a collection of an independent and identically distributed (IID) data. This is the simplest possible case. The assumption of IID data breaks down when dealing with more complicated data structures, like time series and dynamic models. Fortunately, there exist bootstrap techniques that allow us to use not an IID data, as well. In the following section, it is described a bootstrap variation technique that can deal with problems involving more complicated data structures.

### 3-4 Bootstrap for dynamic models and the model based resampling method

The bootstrap algorithm was originally developed for IID data. Singh (1981) [49] found out that if one applies the IID bootstrap to data that are dependent, inconsistency follows. If we want to apply the bootstrap for more general data structures, like time-series and dynamic models, the algorithm has to be adapted. This is done by reformulating the problem and still using the basic resampling scheme for IID data. Different bootstrap approaches have been developed that can deal with samples which are not IID. In particular, for time-domain analysis, two main classes of methods can be distinguished [50]:

- Block resampling
- Model based resampling

The advantage of block resampling methods is that they are less dependent of the model than the model based resampling methods as we will see later [2]. However, there are many parameters to be optimized in block resampling bootstrap methods. Moreover, the applicability of block resampling bootstrap tools is limited, since their effectiveness depends on the size of the data sample. Specifically, for the moving block methods, the size of each block should be chosen large enough, such that much of the correlation structure of the data is retained. In the same time the number of blocks should also be enough to obtain a good estimate [50, 2, 44]. Some of the disadvantages of block resampling methods can be overcome by the means of the model based resampling methods. Therefore, the model based resampling method is considered in this work and is described next.

#### Model-based resampling method

In the class of model based resampling methods, the residuals are generated and resampled based on a time series model, hence the name. The residual-based bootstrap for linear
regression and autoregression was introduced and studied by Freedman [51], [52] and Efron and Tibshirani [53], [2]. Refinements of the above approach are described in Davidson and Hinkley [43] and Franke, Kreiss and Mammen [54, 55].

To present the residual based approach let us consider the problem of estimating the variance of the parameter of a first order Auto-Regressive (AR) model. We collect \( N \) observations \( x_t = x_1, x_2, ..., x_N \) from the first order autoregressive model:

\[
x_t + ax_{t-1} = e_t,
\]

where \( e_t \) is a stationary white Gaussian noise with \( E[e_t] = 0 \) and a probability distribution \( F_e \).

After centering the observed data \( x_t \), a first order AR model is fitted and an estimate \( \hat{a} \) of the parameter \( a \) is obtained. The estimate can be calculated either with least-squares or with the maximum likelihood method. Then the prediction error \( \varepsilon = y_t - \hat{y}(t|t-1) \) is calculated based on the estimated model. The idea is to draw a bootstrap sample from the prediction error \( \varepsilon \). To draw a new bootstrap sample from the prediction error \( \varepsilon \) its distribution has to be known. Assuming that the prediction error is a realization of white noise process, we use it to estimate its distribution. This can be done either in a parametric or in a non-parametric way. Next, create \( B \) bootstrap sample \( x^*_1, x^*_2, ..., x^*_N \) by drawing \( \hat{e}^*_2, \hat{e}^*_3, ..., \hat{e}^*_N \) with replacement, from the residuals \( \hat{e}_2, \hat{e}_3, ... , \hat{e}_N \). Calculate the bootstrap estimates \( \hat{a}^*_1, \hat{a}^*_2, ..., \hat{a}^*_B \) for each bootstrap replication of \( x^* \). Finally, the variance of the model using residual bootstrap is found by:

\[
\sigma^2_a = \frac{1}{B-1} \sum_{i=1}^{B} \left( \hat{a}^*_i - \frac{1}{B} \sum_{j=1}^{B} \hat{a}^*_j \right).
\]  

A similar approach can be taken to estimate the variance for various kinds of correlation structures. In the statistical literature, the main focus is on the derivation of residuals bootstrapping for AR and Auto-Regressive Moving Average (ARMA) time-series models. But the same idea can be also extended for input/output dynamic models, with slight adaptations in order to take into account the presence of the deterministic input \( u \) of the system. A modification of residuals bootstrap made to cope with time-series data is presented in the work of Bittanti and Lovera [56]. They have developed a residual bootstrap procedure for evaluating the model uncertainty in the framework of subspace identification methods, in particular for the Multivariable Output-Error State sPace (MOESP) class of methods. Specifically, they have provided a computationally efficient algorithm for estimating the standard error of identified eigenvalues and of the frequency response of identified models. In the next section, the computational cost of the bootstrap method is addressed.

### 3-5 Computational issues

To ensure good enough convergence of the estimate of the standard error or the confidence intervals, the number of replications \( B \) is usually very large. Thus, the choice of bootstrap replications \( B \) has to be made properly. In their work Shao and Tu [48] discuss the problem of choosing \( B \) and also provide some practical guidelines. For example, iterations of around 100
Figure 3.2: The concept of residual based-bootstrap for estimating the variance of an estimate.

To 200 are necessary for estimates such as standard errors and bias. Whereas, for estimating confidence intervals the required iterations are of 1000 and more [57]. Several methods have been introduced to reduce the computational cost of bootstrapping [58, 59]. Therefore, the application of bootstrap can be considered only if the model estimate can be computed in a very efficient way [56].

Bittani and Lovera propose a procedure to achieve computational savings for the ordinary MOESP methods and the PI/PO versions using residual-based bootstrap [56]. Their approach is based on the fact that the distribution of the residuals can be approximated (either in a parametric or non-parametric way), assuming that the sequence of the prediction error is a realization of a white noise process. In this way, the computation of the noise free model is performed only once. After that, in order to take into account the noise acting on the system, the model is perturbed \( B \) times.

Until now, we discussed how the estimates standard error can be calculated using the bootstrap approach. In the next section, another tool for expressing the estimates accuracy will
be studied.

### 3-6 Confidence regions calculation using bootstrap

So far, the calculation of the standard error and the standard deviation with the bootstrap was considered. The uncertainty of an estimate can be quantified in terms of confidence intervals, as well. Confidence intervals give a set of plausible values of $\theta$, having observed $\hat{\theta}$, instead of just a point estimate $\hat{\theta}$. In addition to calculating an interval of plausible values for $\theta$, it is possible to provide a statement how confident we are that the true parameter $\theta$ lies in that interval. In other words, we can state what is the probability that a certain unknown parameter $\theta$ is within confidence interval.

In this section, different techniques for constructing confidence intervals using the bootstrap are presented. Confidence intervals using the bootstrap can be calculated with two classes of methods, based on bootstrap tables and based on bootstrap percentiles, as depicted in Figure 3-3.

**Figure 3-3:** Methodologies for constructing confidence intervals with the bootstrap.

#### 3-6-1 Confidence intervals generation using bootstrap tables

Here, we present the principle of constructing confidence regions using bootstrap tables. This class of methods uses the estimates distribution and the estimated standard error to calculate the confidence intervals. The first method, is based on the assumption that the estimates distribution is normal. The second presented approach, does not rely on normal theory assumptions, but uses the data to estimate the distribution.

**Normal confidence intervals**

Under most circumstances, the bootstrap distribution of a given estimate $\hat{\theta}$ becomes normal, with the increase of bootstrap replications $B$, as showed in Figure 3-4 [2, 6]. The estimate $\hat{\theta}$ has a distribution with mean near $\theta$ and a variance around $\sigma^2$, written as $N(\theta, \sigma^2)$ and from the properties of the normal distribution it follows that:
If the variance $\sigma^2$ is known, the $(1 - 2\alpha)$ confidence interval can be constructed by using the critical values $z_\alpha$. The critical value $z_\alpha$ of a normal distribution $N(0, 1)$ is the number that corresponds to a probability $\alpha$. This number can be read from a table [6]. And the value $\alpha$ indicates the $(1 - \alpha)$th percentile point of the standard normal distribution, as shown in Figure 3-5.
The $(1 - 2\alpha)$ confidence intervals have the following form:

$$\hat{\theta} \pm z_{\alpha}\sigma.$$  

Therefore, the confidence intervals are calculated using the standard error $\sigma$. The standard error $\sigma$ of an estimate $\hat{\theta}$ can be obtained from the standard deviation of the bootstrap distribution 3-3.

Applying the theory presented so far to generate confidence intervals with probability of 95%, we have to choose $\alpha = 0.025$ and $\alpha = 0.925$. Then select from the standard distribution table, the critical values $z_{0.025} = -1.96$ and $z_{0.975} = 1.96$. And the the confidence regions with probability of 95% are given as:

$$\hat{\theta} \pm 1.96\sigma.$$  

The interval derived with this approach relies on the assumption that the estimate has a normal distribution. It makes use the estimates variance, calculated from the standard deviation of the bootstrap distribution, to build the confidence intervals. Next, we discuss another way for determining confidence intervals, without having to rely on normal theory assumptions.

**The bootstrap-t interval**

When making an assumption for a normal distribution of the estimate is not reliable, the confidence intervals can be built with the bootstrap-t interval method. An illustration of a bootstrap distribution of estimate that is not normal is shown in Figure 3-6. The bootstrap-t interval is a generalization of the Student-t method [2, 43].

![Bootstrap distribution that is not normal.](image)

With this approach, the distribution of the estimator is approximated directly from the data and a table with percentiles is generated [2]. The constructed bootstrap-t table is used in the same way as the normal table, to construct the confidence intervals. To obtain the bootstrap-t table, $B$ bootstrap samples are generated, and therefore, $B$ versions of the distribution $Z$ are calculated. The percentiles $z_{\alpha}$ of the obtained $B$ number of distributions $Z$ form the bootstrap-t table.
Here, we describe the procedure for confidence intervals calculation using the bootstrap-t method. Given a dataset $y_k$, ($k = 1, ..., N$), we generate $B$ bootstrap datasets $y_{k}^{*b}$, ($b = 1, ..., B$) and for each dataset we compute a distribution:

$$Z_b^* = \frac{\hat{\theta}_b^* - \theta}{\sigma_b^*},$$

where $\hat{\theta}_b^*$ is the value of $\hat{\theta}$, obtained from the bootstrap sample $y_{k}^{*b}$. And $\sigma_b^*$ is the standard error of $\hat{\theta}_b^*$ estimated for the bootstrap sample $y_{k}^{*b}$.

When the distribution is approximated with $Z_b^*$, then the quantiles $\tilde{z}_\alpha$ are known, and the bootstrap-t table can be constructed. The $\alpha$th percentile of $Z_b^*$ is estimated by the critical value $\tilde{z}_\alpha$. So, the $(1 - 2\alpha)$ bootstrap-t confidence interval has the following form:

$$\left[\hat{\theta} - \tilde{z}_{(1-\alpha)(B+1)}\sigma, \quad \hat{\theta} - \tilde{z}_{\alpha(B+1)}\sigma\right].$$

The bootstrap-t approach requires the sample variance $\sigma^2$ for $\hat{\theta}$ to be computed from $y_k$ at each bootstrap iteration. A difficulty arises when the standard variance $\sigma^2$ of the modal parameters have to be estimated for a single bootstrap sample. To calculate the variance for $\hat{\theta}_b^*$ from $y_{k}^{*}$ at each iteration implies the use of asymptotic theory approach. And the second drawback of the bootstrap-t method is that it may perform erratically in case of small sample size situations. In order to overcome these problems, the bootstrap estimate of the standard error can be calculated for each bootstrap iteration. This is realized with two nested levels of bootstrapping. To estimate the standard error, a number of $B = 25$ iterations could be sufficient [2]. Calculating the confidence regions requires much larger number of bootstrap iterations, such as $B = 1000$ [2]. Therefore, the total number of bootstrap iterations to perform will be $B = 25000$, which could be not a viable solution if the estimate $\hat{\theta}$ is computationally costly to estimate.

In this section, we discussed the creation of confidence intervals in the framework of bootstrapping by making use of tables. Below, another bootstrap approach for defining confidence intervals of estimates is presented.

### 3-6-2 Confidence intervals based on bootstrap percentiles

There is another class of methods that exists for creation of confidence intervals when using the bootstrap. Unlike the confidence intervals based on bootstrap samples, this class of methods for constructing confidence regions does not involves calculation of estimates standard errors. Instead, the intervals are derived based on percentiles of the bootstrap distribution of the considered estimate. Two methods belonging to this class of methods will be described in detail. Those are the percentile interval and the improved percentile intervals methods. To deal with the drawbacks of the percentile method, a number of improved percentile interval methods were developed. Next, two improved percentile methods will be described.
The percentile interval

The percentile interval method suggests employing the percentiles of the bootstrap histogram in order to define the confidence regions \([43, 2, 44]\). The \(1 - 2\alpha\) percentile interval can be determined by the \(\alpha\) and \(1 - \alpha\) of the bootstrap distribution. If \(\hat{\theta}_b^*\) is the 100th percentile percentile of the bootstrap distribution \(\hat{\theta}_b^*\), then the \(1 - 2\alpha\) confidence interval can be found as:

\[
\left[ \hat{\theta}_{\alpha}, \hat{\theta}_{(1-\alpha)}^* \right].
\]

Consequently, to determine the 95% confidence interval we have to identify the 2.5% and the 97.5% percentiles of the bootstrap histogram, as visualized in Figure 3-7. For example, if we have generated 1000 bootstrap replications \(B = 1000\) and select the level of probability \(\alpha = 0.05\), then \(\hat{\theta}_\alpha^*\) is the 50th ordered value of bootstrap replications. Specifically, \(\hat{\theta}_{(1-\alpha)}\) is the 950th value of the ordered bootstrap replications.

Figure 3-7: Histogram of 1000 bootstrap replications of \(\hat{\theta}\). The solid line is the estimate \(\hat{\theta}\). The dashed lines are the 5% and 95% percentiles of the bootstrap histogram.

The advantage of this method is that it is straightforward to apply. It gives accurate results in cases where the bootstrap distribution is approximately symmetrical and centered around the observed estimate. However, if this criteria is not met, the percentile bootstrap can deliver too narrow confidence intervals [2].

Improved percentile interval methods (\(BC_a\), \(ABC\))

The Bias-corrected and accelerated (\(BC_a\)) and the Approximate bootstrap confidence intervals (\(ABC\)) methods are improved versions of the percentile interval method. The ABC method aims at reducing the amount of computation required for the \(BC_a\) approach. First, the \(BC_a\) will be briefly explained.

Ivelina Dimitrova

Master of Science Thesis
The $1 - 2\alpha$ confidence interval of a given estimate, when applying the $BC_a$ approach, is defined as:

$$\left[ \hat{\theta}_a^{(1-\alpha_1)}, \hat{\theta}_a^{(1-\alpha_2)} \right]$$

Here, $\alpha_1$ and $\alpha_2$ are chosen to improve properties. Therefore, the difference between the percentile and $BC_a$ percentile methods is in the endpoints, used to construct the confidence regions. These coefficients are chosen in the following way:

$$\alpha_1 = \Phi \left( z_0 + \frac{z_0 + z_\alpha}{1 - \hat{a}(z_0 + z_\alpha)} \right),$$

$$\alpha_2 = \Phi \left( z_0 + \frac{z_0 + z_{1-\alpha}}{1 - \hat{a}(z_0 + z_{1-\alpha})} \right),$$

where $\Phi$ is the cumulative distribution function Cumulative Distribution Function (CDF) for the normal distribution and $z_\alpha$ is the 100\(th\) percentile point of the normal distribution. For instance, if we choose $\alpha = 0.95$, then $z_{0.95} = 1.645$ and $\Phi(1.645) = 0.95$. Furthermore, in these expressions, $\hat{z}_0$ and $\hat{a}$ are the correction coefficients [2]. Specifically, $\hat{z}_0$ is the bias-correction coefficient and $\hat{a}$ is the acceleration coefficient. When $\hat{z}_0 = 0$ and $\hat{a} = 0$ then $\alpha_1 = \alpha$ and $\alpha_2 = 1 - \alpha$. With this choice of coefficients, the $BC_a$ method is the same as the percentile method.

The bias-correction coefficient $\hat{z}_0$ is computed by:

$$\hat{z}_0 = \Phi^{-1} \left( \frac{\hat{\theta}^*(b) < \hat{\theta}}{B} \right).$$

To determine the acceleration coefficient $\hat{a}$, the following expression can be used [2]:

$$\hat{a} = \frac{\sum_{i=1}^{N} (\hat{\theta}_{(i)} - \hat{\theta})^3}{6 \left( \sum_{i=1}^{N} (\hat{\theta}_{(i)} - \hat{\theta})^2 \right)^{3/2}}.$$

The $ABC$ method is an approximation of the $BC_a$ method. Because to calculate the confidence intervals with $BC_a$ method requires the generation of a large number of bootstrap replications, the idea is to approximate it. This operation reduces by a large factor the amount of computation that the $BC_a$ method demands. Approximating the $BC_a$ intervals endpoints analytically, does not involves generation of replications. The idea of this approach is to approximate the bootstrap random sampling results by a Taylor series expansions [2].

In this section, two classes of methods were investigated for delivering confidence regions of estimates when using the bootstrap. Pointed out were the advantages and disadvantages of each methodology. From the presented methodologies the percentile interval method appears to be the most appealing. Because it does not rely on any assumptions regarding the bootstrap distribution, as the normal confidence intervals method does. It can be applied to any bootstrap distribution directly, without estimating it from the data, as in the bootstrap-t interval methodology. Estimating the bootstrap distribution is done by calculating the sample
variance, which can cause drawbacks. Such drawbacks are increase of computational time or loss of accuracy. The reason for that is that the sample variance has to be calculated using either the asymptotic theory or by using another level of bootstrapping. The percentile method is simple to apply. It employs use of the percentile of the bootstrap histogram in order to build the confidence intervals. The accuracy of the percentile is guaranteed when the bootstrap distribution is approximately symmetrical and centered around the estimate, which can be verified simply by plotting the bootstrap histogram.

Yet, we introduced the theory of standard error and confidence intervals calculation using the bootstrap approach. In the coming section, the problem of evaluating models uncertainty in the framework of subspace OMA-PBSID$_{opt}$ method with the bootstrap is considered.

### 3-7 Bootstrap for PBSID$_{opt}$

In this section we describe how the accuracy of an estimate obtained with OMA-PBSID$_{opt}$ method is determined, using the bootstrap uncertainty quantification technique. The theory, presented so far in this chapter is applied and the procedure is summarized in Table 3-1.

Considering that for the identification we are dealing with a time-series data from a dynamic model, a bootstrap method for dynamic models is necessary to be chosen. Hence, the residual-based bootstrap, described in detail in Section 3-4 is used. What is specific for this approach is that we draw a new bootstrap replication from the prediction error $\epsilon_k$. Then the identified model is perturbed with the bootstrap replication of the prediction error $\epsilon_k^*$, to obtain a bootstrap output sequence $y_k^B$. This bootstrap output sequence $y_k^B$ is used for the identification of a new estimate, called bootstrap estimate. Repeating this procedure a large number of times $B$, will deliver $B$ bootstrap estimates. The standard error of estimates is calculated from the bootstrap distribution with equation 3-5. To calculate confidence intervals for the obtained estimate, one of the methods described in the previous section 4-6 can be applied.

### 3-8 Concluding remarks

In this chapter, we discussed a data-based approach for assigning a measure of model accuracy, namely the bootstrap. The performance of the bootstrap will be explored in Chapter 5, on a simulation study. In the next chapter, we present the traditional asymptotic approach, which relies on analytical asymptotic expressions to obtain approximations of the model uncertainty.
Bootstrap-based estimates of uncertainty with OMA-PBSID\textsubscript{opt} method.

1. **Experiment.** Conduct an experiment to obtain \( y_k \) with number of samples \( N \).

2. **Calculation of the model estimate.** Using the output data \( y_k \) perform OMA - PBSID\textsubscript{opt} identification to identify a model \((\hat{A}, \hat{C})\) and compute the prediction error \( \epsilon_k \).

3. **Draw from the prediction error \( \epsilon_k \).** Generate new bootstrap output data \( y_k^B \).
   
   3.1 Perturb the identified model with the prediction error \( \epsilon_k \).
   
   3.2 Extracting from the measured output \( y_k \) the generated perturbed output to obtain noiseless output sequence.
   
   3.3 Perturb the identified model with \( \epsilon_k^* \) which is constructed by resampling from the distribution of \( \epsilon_k \).
   
   3.4 Add the generated perturbed output to the noiseless output sequence to obtain the bootstrap output data \( y_k^*B \).

4. **Calculation of the bootstrap estimate.** Using the new bootstrap output data \( y_k^*B \) perform another OMA-PBSID\textsubscript{opt} identification to identify another bootstrap model \((\hat{A}, \hat{C})\).

5. **Repetition.** Repeat steps 3 and 4 \( B \) times, to obtain a total of \( B \) bootstrap model estimates.

6. **Confidence interval.** Calculate \( 100(1 - \alpha)\% \) bootstrap confidence interval. Use one of the procedures presented in Section 4-6.

**Table 3-1:** Bootstrap PBSID\textsubscript{opt}-OMA procedure for confidence interval calculation of parameter estimates.
Chapter 4

Asymptotic variance

In this chapter, we derive the asymptotic variance expression for the novel Operational Modal Analysis (OMA) - optimized Predictor Based Subspace Identification (PBSID\textsubscript{opt}) algorithm introduced in Chapter 2. In particular, we formulate the asymptotic variance expressions for the identified matrices $A$, $C$ and $K$. In this project, we are investigating the uncertainty of estimated modal parameters. The modal parameters that we are interested are the eigenfrequencies $\omega_i$ and the damping ratios $\xi_i$, which can be calculated from the system matrices $A$ and $C$. Therefore, we obtain variance expressions for the eigenfrequencies $\omega_i$ and the damping ratios $\xi_i$, as well. In order to get the variance of modal parameters, first-order sensitivity expressions are derived for the modal parameters to perturbation of the estimated models. In such a way, the uncertainties of the modal parameters can be computed from the uncertainty of the estimated model. Finally, it is shown how the estimated variance of identified parameters can be used for the construction of confidence intervals.

4-1 Asymptotic normality

Before, deriving the asymptotic variance expressions for the OMA-PBSID\textsubscript{opt} identification method, we define the general expressions for the asymptotic normality. Because, the asymptotic normality result and the availability of parameters covariance expressions, serve as a basis for the generations of confidence interval of the parameter estimates related to a specified probability [8].

With every random variable there is a Probability Density Function (PDF) associated that completely characterizes this variable. Consequently, to understand how the identified parameter estimate $\hat{\theta}_N$ is distributed around the true parameters, denoted as $\theta^*$, its PDF is required. However, it is impossible to derive an expression for this PDF for finite values of $N$. Instead, we may compute the asymptotic PDF of the random variable $\hat{\theta}_N - \theta^*$ in the asymptotic case [16]:

$$\sqrt{N}(\hat{\theta}_N - \theta^*) \in \mathcal{N}(0, P_0).$$

(4-1)
Hence, the variable $\hat{\theta}_N - \theta^*$ converges to a Gaussian PDF with a zero mean and a covariance matrix $P_\theta$. Knowing this asymptotic PDF, it is possible to assess the model accuracy, at least for very large values of $N$. Moreover [8],

$$Cov(\sqrt{N}\hat{\theta}_N) \rightarrow P_\theta \text{ as } N \rightarrow \infty.$$  \hfill (4-2)

Then, the asymptotic covariance matrix $P_\theta$ of the estimated parameter vector $\hat{\theta}_N$ serves as an expression for the asymptotic variance. The question, therefore, is how to derive expressions for the asymptotic covariance function for the modal parameters.

In section 4-6 we directly use the statistical properties of asymptotic normality to evaluate the modeling error, by constructing confidence intervals of parameter estimates.

In the following section, we introduce the principle of uncertainty propagation. The uncertainty propagation principle is necessary in order to derive expressions for the asymptotic covariance function for the natural frequency and damping ratios. It will be shown that the variance estimation algorithm relies on computation of the first-order sensitivity of the estimated modal parameters to perturbation of the measured output data.

### 4-2 Uncertainty propagation

We already showed that to assess the accuracy of estimates, its covariance function has to be available. Here, the principle of uncertainty propagation is presented. This principle is important because it plays an important role in the derivation of covariance expressions for estimates.

Let $\theta$ is some parameter of interest and $\hat{\theta}$ is its estimate. Consider also $f(\theta)$ that is a vector valued function of the parameter $\theta$. The covariance of $\hat{\theta}$ can be approximated by:

$$\text{cov} f(\hat{\theta}) \approx J_f \text{cov}(\hat{\theta}) J_f^T,$$  \hfill (4-3)

where the sensitivity matrix $J_f$ is the first derivative of $f$.

Hence, from the covariance of an estimate, that can for instance be calculated from a sample covariance, the covariance of a function of this estimate can be obtained with equation 4-3.

Considering that the system matrices and subsequently the modal parameters are obtained from the output response, the objective is to compute the sensitivities of these parameters with respect to the output data and then to calculate their covariance based on equation 4-3.

To find a particular sensitivity matrix, it can be made use of first-order perturbations. In order to define the first-order perturbations, let $\varepsilon$ is the perturbation magnitude. Accordingly, the first-order perturbation of the parameter $\theta$ is given as [60]:

$$\Delta \theta = \varepsilon \frac{\partial \theta}{\partial \varepsilon} = \theta \varepsilon^2,$$

and the first-order perturbation of $f(\theta)$ is defined as:
Therefore, to get the desired sensitivity matrices it is necessary to calculate first-order perturbations [60].

The procedure for the estimation of the uncertainty of modal parameters (natural frequencies and damping ratios) estimates from a single experiment can be divided into three steps: starting from an estimate of the covariance of the system matrices $A$ and $C$ from measured data, calculation of eigenvalues and eigenvectors variance and computation of modal parameters asymptotic variance. These procedure steps are presented in Figure 4-1.

In the following sections, the underlying theory necessary for the computation of the covariance matrices is explained in detail. We start with the derivations of asymptotic variance expressions for the system matrices $A$ and $C$, obtained with the OMA-PBSID$_{opt}$ identification method. The purpose is to investigate how the identified system matrices differ with a perturbation of the measured output response $y_k$.

**Problem 2. Asymptotic variance of system matrices, obtained with OMA - PBSID$_{opt}$ method.** Given a finite set of output data $\{y_k\}_{k=0}^{N-1}$, obtained from a system in the form 2-4,
find the asymptotic variance of identified system matrices $\hat{A}, \hat{C}, \hat{K}$, identified with OMA-PBSID$_{opt}$ method, so:

$$\plim \Delta \Theta \Delta \Theta^T$$  \hspace{1cm} (4-4)

with:

$$\Delta \Theta = \begin{bmatrix} \text{vec}(\hat{A} - T^{-1}AT)^T, \text{vec}(\hat{C} - CT)^T, \text{vec}(\hat{K} - T^{-1}K)^T \end{bmatrix}. \hspace{1cm} (4-5)$$

4-3-1 Notations

Following the derivations of the OMA-PBSID$_{opt}$ algorithm in Chapter 2, we add the subsequent notations, that are necessary for the derivation of the asymptotic variance expressions. The estimate of the ‘extended observability-times-controllability’ matrix has the following form:

$$\hat{\Gamma}^{fKp} = \begin{bmatrix} YZ^\dagger_{z1} & YZ^\dagger_{z2} & \cdots & YZ^\dagger_{zp} \\ 0 & YZ^\dagger_{z1} & \cdots & YZ^\dagger_{zp-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & YZ^\dagger_{z1} \end{bmatrix}. \hspace{1cm} (4-6)$$

Then the error of the estimated ‘extended observability-times-controllability’ matrix is defined as:

$$\Delta \hat{\Gamma}^{fKp} = \begin{bmatrix} EZ^\dagger_{z1} & EZ^\dagger_{z2} & \cdots & EZ^\dagger_{zp} \\ 0 & EZ^\dagger_{z1} & \cdots & EZ^\dagger_{zp-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & EZ^\dagger_{z1} \end{bmatrix}. \hspace{1cm} (4-7)$$

The following equality holds:

$$\hat{\Gamma}^{fKp} = \Gamma^{fKp} + \Delta \hat{\Gamma}^{fKp}. \hspace{1cm} (4-8)$$

Since, the state contains a similarity transformation, we can also write:

$$\hat{X} = TX + \Delta \hat{X}. \hspace{1cm} (4-9)$$

By taking into account equations 2-18 and 2-19, the uncertainty of the system matrices is given by:

$$[\Delta \hat{A}, \Delta \hat{K}] = (\Delta \hat{X} - K\Delta \hat{E}) \begin{bmatrix} \hat{X}^\dagger \\ \hat{E} \end{bmatrix}, \hspace{1cm} (4-10)$$

$$\Delta \hat{C} = (-C\Delta \hat{X} - \Delta \hat{A})\hat{X}^\dagger. \hspace{1cm} (4-11)$$
4-3-2 Estimate of $\Delta \bar{X}$

To come up with asymptotic variance expressions for the system matrices $\hat{A}, \hat{C}, \hat{K}$, we first look at the error of the state estimate $\Delta \hat{X}$. By using 4-8 and 4-9, we discover that the error of the state estimate $\Delta \hat{X}$ can be written as [36]:

$$\Delta \hat{X} = S\Delta F^T \hat{K} Z,$$  \hspace{1cm} (4-12)

where $S$ is a matrix with a dimension $S \in \mathbb{R}^{n \times lf}$. In addition, we define the following matrices, for $f = p$:

$$\begin{bmatrix} Z_{z_1}^\dagger & Z_{z_2}^\dagger & \cdots & Z_{z_p}^\dagger \\ 0 & Z_{z_1}^\dagger & \cdots & Z_{z_{p-1}}^\dagger \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Z_{z_1}^\dagger \end{bmatrix} Z.$$

(4-13)

In the same manner the matrices $\bar{Q}, Q$ are defined. Then equation 4-12 can be written as:

$$\Delta \hat{X} = SE_{I_f} Q,$$  \hspace{1cm} (4-14)

where $E_{I_f} = I_f \otimes E$ and $I_f$ is the identity matrix with size $f$. In addition, it can be proofed that $\Delta \bar{E} = E\Pi_Z$, and by defining $\hat{E} = EZ^\dagger Z$. Therefore, we can rewrite the uncertainty expression for the system matrices 4-10 and 4-11, in the following form:

$$[\Delta \hat{A}, \Delta \hat{K}] = (SE_{I_f} \bar{Q} - KEZ^\dagger Z) \begin{bmatrix} \hat{X} \\ \bar{E} \end{bmatrix},$$  \hspace{1cm} (4-15)

$$\Delta \hat{C} = (-CSE_{I_f} Q - EZ^\dagger Z) \hat{X}^\dagger.$$  \hspace{1cm} (4-16)

We use these equations, to compute directly the asymptotic variance of system matrices.

4-3-3 Asymptotic variance

With the previously derived expressions in section 4-3-2, the asymptotic variance of system matrices that was defined in 4-4, can now be rewritten as [36]:

$$vec([\Delta \hat{A}, \Delta \hat{K}]) = \alpha_1 vec(E_{I_f}) + \beta_1 vec(E),$$

$$vec([\Delta \hat{C}]) = \alpha_2 vec(E_{I_f}) + \beta_2 vec(E).$$
The matrices $\alpha$ and $\beta$ are defined as:

\[
\alpha_1 = \left( Q_i \left[ \hat{X} \right]^\dagger \right)^T \otimes S_i - \left( Q_i \left[ \hat{X} \right]^\dagger \right)^T \otimes (\hat{A})S_i \\
\alpha_2 = \left( Q_i \left[ \hat{X} \right]^\dagger \right)^T \otimes (-\hat{C}S_i) \\
\beta_1 = - \left( Z^\dagger Z \left[ \hat{X} \right]^\dagger \right)^T \otimes (\hat{K}) \\
\beta_2 = - \left( \Pi Z \left[ \hat{X} \right]^\dagger \right)^T \otimes I.
\]

Then, if we define a matrix $P$ such that $vec E_{I_i} = P vec(E)$ holds true. With this definition, we can write:

\[
vec([\Delta \hat{A}, \Delta \hat{K}, \Delta \hat{C}]) = \left[ \alpha_1 P + \beta_1 \right. \\
\left. \alpha_2 P + \beta_2 \right] (E).
\]

Finally, the asymptotic variance of system matrices is given by:

\[
plim \Delta \Theta \Delta \Theta^* = \left[ \begin{array}{c}
\alpha_1 P + \beta_1 \\
\alpha_2 P + \beta_2
\end{array} \right] (I \otimes \Lambda) \left[ \begin{array}{c}
\alpha_1 P + \beta_1 \\
\alpha_2 P + \beta_2
\end{array} \right]^T,
\]

where

\[
\alpha_1 P = \sum_{i=1}^f \left( \left( Q_i \left[ \hat{X} \right]^\dagger \right)^T \otimes S_i - \left( Q_i \left[ \hat{X} \right]^\dagger \right)^T \otimes (\hat{A})S_i \right) \\
\alpha_2 P = \sum_{i=1}^f \left( \left( Q_i \left[ \hat{X} \right]^\dagger \right)^T \otimes (-\hat{C}S_i) \right).
\]

In this section, we derived asymptotic variance expressions for the identified system matrices in the framework of the OMA-PBSID_opt method. In the next section, we study the derivation of variance expressions for eigenvalues by using the variance of system matrices.

### 4-4 Asymptotic variance of eigenvalues

Just as the problem of finding the system matrices $A$ and $C$ can be sensitive to perturbations present in the measured data, the computation of the eigenvalues and eigenvectors of the state matrix is sensitive to small shifts in the entries of the system matrices. To study the extent of this sensitivity, the eigenvalue perturbation problem is considered. Usually, it consists of finding the derivative of eigenvectors and eigenvalues of a system that is perturbed.
There exist a variety of methods that can deal with the theory and computation of sensitivities
\[ \frac{\partial}{\partial a_{kk}} \left( \lambda(A(q)) \right) \] of eigenvalues of a matrix \( A(a_{kk}) \) with respect to variations of parameters entries \( a_{kk} \) of the matrix. Each methodology arrives in a different way to the solution of the problem. And the applicability of different formulas to variety of system types may vary, as well. Further, they can differ in terms of efficiency of computation, as well.

In this section, we will present two of the existing methodologies for computing the eigenvalues sensitivity.

### 4-4-1 Analytic calculation of the eigenvalue sensitivity

The results for the first-order perturbations of the eigenvalues \( \lambda_i \) of the matrix \( A \) are derived in [38]. Let \( \lambda \) is an eigenvalue of \( A \in \mathbb{C}^{n \times n} \) and that the right eigenvector \( \psi \) and left eigenvector \( \chi \) of \( A \) satisfy the following equality:

\[ A\psi = \lambda \psi \quad \chi^T A = \lambda \chi \]

and the eigenvectors are normalized \( \|\psi\|_2 = \|\chi\|_2 = 1 \). If \( X^T A \Psi = J \) is the Jordan decomposition and \( X^T = \Psi^{-1} \), then the eigenvectors \( \chi \) and \( \psi \) will be nonzero multiples of \( \Psi(:,1) \) and \( X(:,i) \). Then from \( \Psi(:,i)^T X(:,i) = 1 \) it follows that \( \chi^T \psi \neq 0 \). These properties will be used next, to derive the eigenvalues sensitivity formulas.

Consider that there exist a differentiable \( \psi(\epsilon) \) and \( \chi(\epsilon) \), where \( \epsilon \) is the perturbation, such that:

\[ (A + \epsilon F)\psi(\epsilon) = \lambda(\epsilon)\psi(\epsilon) \quad ||F||_2 = 1, \quad ||x(\epsilon)||_2 \equiv 1, \]

where for \( \epsilon = 0 \), \( \lambda(0) = \lambda \) and \( \psi(0) = \psi \). Differentiating this equation with respect to the perturbation \( \epsilon \) and then setting \( \epsilon = 0 \), we get:

\[ A\dot{\psi}(0) + F\psi = \dot{\lambda}(0)\psi + \lambda\dot{\psi}(0). \]

Applying \( \chi^T \) to both sides of this equation gives:

\[ \chi^T A\dot{\psi}(0) + \chi^T F\psi = \chi^T \dot{\lambda}(0)\psi + \chi^T \lambda\dot{\psi}(0), \]

and since \( \chi^T \psi = I \), we get:

\[ \chi^T A\dot{\psi}(0) + \chi^T F\psi = \dot{\lambda}(0) + \chi^T \lambda\dot{\psi}(0). \]

Then we divide by \( \chi^T \psi \):

\[ A\dot{\psi}(0)\psi^{-1} + F = \frac{\dot{\lambda}(0)}{\chi^T \psi} + \lambda\dot{\psi}(0)\psi^{-1}, \]
and take the absolute values we obtain, the first-order perturbation of $\lambda_i$ due to a perturbation of the elements of $A$:

$$|\dot{\lambda}(0)| = |\chi^T F \psi|.$$  

The covariances between the eigenvalues estimates $\lambda_i$ and $\lambda_j$ can be calculated as:

$$\text{cov}(\dot{\lambda}_i, \dot{\lambda}_j) = J_{\lambda_i} \text{cov}(\dot{A}) J_{\lambda_i}^T,$$

using the Kronecker algebra [61], the sensitivity matrix $J_{\lambda_i}$, is given by:

$$J_{\lambda_i} = (\psi_i \otimes \chi)^T_{re}.$$  

The subscript $re$ places the real part $\text{real}(J_{\lambda_i})$ of a matrix on top of the imaginary part $\text{imag}(J_{\lambda_i})$:

$$J_{re} = \begin{bmatrix} \text{real}(J_{\lambda_i}) \\ \text{imag}(J_{\lambda_i}) \end{bmatrix}.$$  

### 4-4-2 Numerical calculation of the eigenvalue sensitivity

A first-order eigenvalue sensitivity with respect to the elements of the state matrix $A$ can be solved numerically, as well [62]. In this case the Jacobian matrix is constructed and solved numerically. The Jacobian matrix is the matrix of all first-order partial derivatives of a vector-valued function $f$, as follows:

$$J = \frac{df}{dx} = \left[ \frac{\partial f}{\partial x_1} \cdots \frac{\partial f}{\partial x_n} \right] = \left[ \begin{array}{ccc} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{array} \right],$$  

$$J_{ij} = \frac{\partial f_i}{\partial x_j}.$$  

Therefore, the partial derivatives of the eigenvalues of the state-space system with respect to the entries of the state matrix $A$, are given as:

$$J_{\lambda} = \frac{\partial \lambda_i}{\partial a_{kk}} = \left[ \begin{array}{ccc} \frac{\partial \lambda_1}{\partial a_{11}} & \cdots & \frac{\partial \lambda_1}{\partial a_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \lambda_n}{\partial a_{n1}} & \cdots & \frac{\partial \lambda_n}{\partial a_{nn}} \end{array} \right],$$  

where $\lambda_i$ is the $i$th eigenvalue, $a_{kk}$ is the $k$th row and $k$th column of the state matrix.

In the next section, we study the variance expressions for the eigenfrequencies and damping ratios. The derived variance expressions in this section for the eigenvalues are part of the variance formulas for the modal parameters, due to the principle of uncertainty propagation.
4-5 Asymptotic variance of the eigenfrequencies and damping ratios

The natural frequencies and the damping ratios both depend on the eigenvalues of the system matrix $A$. Whereas, the mode shapes are a function of the system matrices $A$ and $C$. Therefore, the covariances of these modal parameters can be estimated from the covariances of the system matrices. The analytic expressions for the covariances of the eigenfrequencies and damping ratios have been previously derived in [63]. The covariance expressions for the mode shapes are studied in [35]. Here we study the covariances of the natural frequencies and modal shapes, using the procedure proposed in [63].

The covariance between the continuous-time system poles is obtained [64] as:

$$\text{cov}(\hat{\lambda}_{ci}, \hat{\lambda}_{cj}) = J_{\lambda ci} \text{cov}(\hat{\lambda}_i, \hat{\lambda}_j) J_{\lambda cj}^T,$$

(4-22)

with

$$J_{\lambda ci} = \frac{1}{\Delta t |\lambda_i|^2} \left[ \begin{array}{cc} \lambda_i^R & \lambda_i^I \\ -\lambda_i^I & \lambda_i^R \end{array} \right],$$

(4-23)

where $\Delta t$ is the sampling time. The eigenfrequencies $f_i$ and damping ratios $\xi_i$ are functions of the continuous-time system poles $\lambda_{ci}$ and can be calculated by [35]:

$$f_i = \frac{|\lambda_{ci}|}{2\pi} \quad \text{and} \quad \xi_i = -100 \frac{\lambda_{ci}^R}{|\lambda_{ci}|},$$

where $|.|$ is the complex modulus, and $\lambda_{ci} = \lambda_{ci}^R + i\lambda_{ci}^I$. Therefore, the covariances between two pairs of eigenfrequencies $f_i$ and damping ratios $\xi_i$ are calculated using:

$$\text{cov} \left[ \begin{array}{c} \Delta f_i \\ \Delta \xi_i \end{array} \right] = J_{f\xi i} \text{cov}(\hat{\lambda}_{ci}, \hat{\lambda}_{cj}) J_{f\xi i}^T,$$

(4-24)

and the relationship between the discrete-time poles and the continuous-time poles is given by:

$$J_{f\xi i} = \frac{1}{|\lambda_{ci}|} \left[ \begin{array}{cc} \frac{\lambda_{ci}^R}{2\pi} & \frac{\lambda_{ci}^I}{2\pi} \\ -100 \frac{\lambda_{ci}^2}{|\lambda_{ci}|^2} & 100 \frac{\lambda_{ci}^R\lambda_{ci}^I}{|\lambda_{ci}|^2} \end{array} \right],$$

(4-25)

$$\lambda_{ci} = \frac{\ln(\lambda_i)}{\Delta t}.$$

So far, we explored the derivation of variance expressions for the natural frequencies and damping ratios of a dynamical system. The generation of confidence intervals when using asymptotic variance expressions for the estimated parameters is considered next.
4-6 Confidence intervals

To determine how close the identified parameter vector $\hat{\theta}_N$ is to the true parameter vector $\theta^*$, we can form an uncertainty confidence regions, as previously discussed in section 3-6. In this section, we show how the confidence intervals of estimated parameters can be built from a single data set by the means of asymptotic variance uncertainty quantification technique.

The asymptotic normality result 4-1 and the availability of covariance matrix $\text{cov}(\hat{\theta}_N)$ provide the expressions for the confidence intervals of the estimated parameter, related to a specified probability [8, 16]. Under most circumstances, the distribution of the obtained estimates $\hat{\theta}$ becomes normal with the increase of the number of measured samples $N$. This ensures that the estimate $\hat{\theta}$ will have a distribution with mean near $\theta$ and variance around $\sigma^2$, and from the properties of the normal distribution it follows that:

$$\frac{\hat{\theta}_N - \theta^*}{\sigma} \approx N(0, 1).$$

Therefore, if the variance $\sigma^2$ is known, the $(1 - 2\alpha)$ confidence interval can be constructed using the critical values $z_\alpha$. The critical values had been already defined in section 3-6 of the previous Chapter 3. The critical value $z_\alpha$ of a normal distribution $N(0, 1)$ is the number that has probability $\alpha$. This number is read from a table and it indicates the $(1 - \alpha)$th percentile point of the standard normal distribution [6]. The $(1 - 2\alpha)$ confidence intervals have the following form:

$$\hat{\theta}_N \pm z_\alpha \sigma.$$

From which it can be concluded that the confidence intervals can be generated using the standard error $\sigma$. The standard error $\sigma$ of a an estimate $\hat{\theta}$ can be calculated as a square root of the variance $\sigma^2$. To construct confidence intervals with probability of 95%, we choose $\alpha = 0.025$ and $\alpha = 0.975$. We select from the standard distribution table $z_{0.025} = -1.96$ and $z_{0.975} = 1.96$, and the confidence intervals with probability of 95% are given as:

$$\hat{\theta}_N \pm 1.96 \sigma.$$

4-7 Asymptotic variance for PBSID$_{opt}$

Just like in the end of the previous Chapter 3, for clearance we summarize the theory for uncertainty quantification with asymptotic variance for the OMA-PBSID$_{opt}$ method. This procedure is summarized in Table 4-1.

4-8 Concluding remarks

In this chapter, we presented the asymptotic variance approach for assigning a measure of model accuracy. The performance of the considered uncertainty quantification method will
4-8 Concluding remarks

Asymptotic variance confidence interval calculation

1. **Experiment.** Conduct an experiment to obtain $y_k$ with number of samples $N$.

2. **Calculation of the model estimate.** Using the output data $y_k$ perform OMA-PBSID$_{opt}$ identification to identify a system model.

3. **Covariances.**
   3.1 Calculate $\text{cov}(\hat{A})$ and $\text{cov}(\hat{C})$.
   3.2 Use $\text{cov}(\hat{A})$ to calculate the covariances $\text{cov}(\hat{\lambda}_{i,\text{re}}, \hat{\lambda}_{j,\text{re}})$ between eigenvalue estimates $\hat{\lambda}_i$ and $\hat{\lambda}_j$.
   3.3 Using the discrete-time poles covariance $\text{cov}(\hat{\lambda}_{i,\text{re}}, \hat{\lambda}_{j,\text{re}})$ calculate continuous-time poles covariance $\text{cov}(\hat{\lambda}_{ci,\text{re}}, \hat{\lambda}_{cj,\text{re}})$.
   3.4 Calculate the covariance between two pairs of eigenfrequencies and damping ratios $\text{cov} \left( \begin{bmatrix} f_i \\ \xi_i \end{bmatrix}, \begin{bmatrix} f_i \\ \xi_i \end{bmatrix} \right)$ using $\text{cov}(\hat{\lambda}_{ci,\text{re}}, \hat{\lambda}_{cj,\text{re}})$.
   3.5 Use $\text{cov}(\hat{A})$ and $\text{cov}(\hat{C})$ to calculate the covariance matrix $\text{cov}(\varphi_{ire}, \varphi_{jre})$ of the mode shapes $\varphi_{ire}$ and $\varphi_{jre}$.

4. **Confidence interval.** Calculate $100(1-\alpha)\%$ confidence interval $(\hat{\mu}(U), \hat{\mu}(L))$ for each estimate of the modal parameters from the calculated covariance function in step 3. For 95% confidence interval we have $\alpha = 0.05$ and we get $[\hat{\mu}_i - 1.96\sigma_i; \hat{\mu}_i + 1.96\sigma_i]$

| Table 4-1: Asymptotic variance procedure for confidence interval calculation of parameter estimates. |  |

be explored in Chapter 5 on a simulation example. And its performance will be compared with the bootstrap uncertainty quantification techniques, that was explored in the previous Chapter 3.
Asymptotic variance
In this chapter we investigate the validity of the theories and the performance of the proposed identification method in Chapter 2 and uncertainty quantification techniques in Chapter 3 and Chapter 4, using a simulation experiment. The simulation experiment aims at showing the strengths of the proposed uncertainty quantification techniques. In general, the bootstrap is investigated as a viable choice for uncertainty quantification when small number of measurement samples $N$ is available, compared to the traditional asymptotic variance method. We will first start with introducing the details about the system being simulated, which is a second-order highly damped system. Then the identification and uncertainty quantification parts follow. Finally, this chapter ends with some conclusions.

5-1 True system dynamics

The true system that is simulated is a Single Input, Single Output (SISO) Linear Time-Invariant (LTI) second-order system. The dynamics of the plant can be represented with the following transfer function:

$$G(s) = \frac{\omega_n^2}{s^2 + 2\xi\omega_n s + \omega_n^2} = \frac{1}{s^2 + 0.2s + 1},$$

with natural frequency: $\omega_n = 1$ rad/s and damping ratio: $\xi = 0.1$. The magnitude Bode plot of the plant is shown in Figure 5-1.

The goal is to obtain an estimate of the natural frequency $\omega_n$ and damping ratio $\xi$ of the system 5-1, using only output data and to assign a measure of estimates accuracy. For this purpose, to identify the dynamics of the plant, the novel Operational Modal Analysis (OMA)-optimized Predictor Based Subspace Identification (PBSID$_{opt}$) identification method is applied that was introduced in Chapter 2. Once the model estimate is obtained, its accuracy will be quantified.
with the techniques presented in Chapters 3 and 4. We start with the system identification part, and describe the identification experiment in the next section.

5-2 Realization of the Identification

The first step towards performing system identification is to collect measurement data from the real plant. As was introduced in Chapter 2, in the framework of OMA only output data $y_k$ from the plant has to be measured. The system configuration in output-only identification is presented in Figure 5-2. Although, in OMA the input signal $u_k$ is not measured, two assumptions (1 and 2) are made regarding its nature. The data-generating system in state-space representation admits the following form:

$$x_{k+1} = Ax_k + Bu_k + Ke_k,$$  \hspace{1cm} (5-2a)  

$$y_k = Cx_k + Du_k + e_k,$$  \hspace{1cm} (5-2b)

with system matrices given in Table 5-1.

Figure 5-1: Bode magnitude plot of the LTI second-order system, used for the simulation experiment.

Figure 5-2: System configuration considered in output-only identification. With $P$ being the data generating system, $y(k)$ the measured output response, $u(k)$ the unknown ambient excitation and $e(k)$ a noise sequence.
Recall Problem 1, from Chapter 2, which is the subspace identification problem in OMA framework. Given a finite number of output data \( \{ y_k \}_{k=0}^{N-1} \), obtained from the system 5-2 we can estimate the order \( n \) of a discrete-time system in the following form:

\[
\begin{align*}
    x_{k+1} &= Ax_k + Ke_k, \\
y_k &= Cx_k + e_k,
\end{align*}
\] (5-3a, 5-3b)

and its system matrices \( A, C, K \) up to a similarity transformation.

We have collected \( N = 2000 \) output measurements \( y_k \) from the true system in open-loop configuration represented with equation 5-1. As an input signal we take a zero-mean Gaussian white noise signal with \( \text{cov}(u_k) = I_r = 1 \). The system is perturbed with a noise signal \( e_k \), and we take a Gaussian white noise with the variance \( \text{cov}(u_k) = I_l = 0.005 \), which corresponds to Signal-to-Noise Ratio (SNR) of approximately 40 dB. The generated signals are shown in Figure 5-3. The collected output data \( y_k \) is used to identify an LTI state-space model of the form 5-2 and the corresponding asymptotic variance.

For the identification algorithm we selected the following windows \( p = f = 20 \). As discussed in Chapter 2, the singular values of the matrix \( \tilde{\Gamma}(f)X_{p,N_p} \) calculated with the OMA-PBSID\(_{opt}\) identification method can be used to determine the order of the system. In Figure 5-4, the first 20 singular values, obtained by applying OMA-PBSID\(_{opt}\) identification to the output data \( y_k \) of the simulation plant, are presented. There appears to be a group of large and a group of small singular values, separated by a 'gap'. This could indicate that the group of small singular values are due to noise. Or it could be a non-dominant behavior (poorly controllable and observable). Regardless of its nature, we perform model reduction by truncating all but the two largest singular values. We hope that this results in a good model, that corresponds to the dynamics of the plant. Eventually, we validate the model using validation data. In essence, we only care about obtaining low-order models which can describe the true system well enough. Please notice that the case is indeed second order.

The performance of the identification algorithm is evaluated by looking at the eigenvalues and modal parameters of the identified system. On Figure 5-5, the eigenvalues of the identified system and those from the true system are plotted. From the system matrices \( A \) and \( C \), the modal parameters can be calculated, as discussed in Chapter 2. We compare the estimated eigenfrequency and damping ratio with the true parameters and the results are summarized in Table 5-2. The results indicate that the identification algorithm performs well, since the estimates are very close to the true parameters.
Figure 5-3: Data set gathered from the 2nd order true system, $y_k$ is the output of the system, $u_k$ the input to the open-loop system, being a a zero-mean Gaussian white noise signal and $e_k$ is a Gaussian white noise, with $k = 1, ..., N$ and $N = 2000$, the sampling time is $\Delta t = 1$ s. Only the output signal $y_k$ is used for the identification.

<table>
<thead>
<tr>
<th></th>
<th>$f$, Hz</th>
<th>$\xi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True system</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>Identified system</td>
<td>1.0054</td>
<td>0.0995</td>
</tr>
</tbody>
</table>

Table 5-2: Natural frequency $f$ and damping ratio $\xi$ of the true system and the identified system, obtained with the OMA-PBSID$_{opt}$ identification method for the case of Chapter 5.

To investigate the sensitivity of the novel OMA-PBSID$_{opt}$ identification algorithm with respect to the noise, a Monte Carlo simulation with 1000 runs was conducted. For each of the 1000 simulations a different realization of the input $u_k$ and noise $e_k$ is generated. Consequently, on each iteration the identification is performed using different output data $y_k$. The results of the 1000 Monte Carlo simulations are presented in Table 5-3. One can observe that the variance of the obtained 1000 estimates is very small and the mean is close to the true the parameter.

<table>
<thead>
<tr>
<th></th>
<th>True parameter</th>
<th>$\mu$</th>
<th>$\sigma^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural frequency $f$, Hz</td>
<td>1</td>
<td>1.0040</td>
<td>0.0007</td>
</tr>
<tr>
<td>Damping ratio $\xi$</td>
<td>0.1</td>
<td>0.1018</td>
<td>0.0007</td>
</tr>
</tbody>
</table>

Table 5-3: Modal parameters from Monte Carlo simulation with 1000 iterations. Where $\mu$ and $\sigma^2$ represent the mean and the variance of the 1000 estimates.

So far, the novel OMA-PBSID$_{opt}$ identification method was used to identify a model of a system, given with equations 5-1. We compared the identified system with the true system.

Ivelina Dimitrova
Master of Science Thesis
A general conclusion, based on the results, is that the OMA-PBSID\textsubscript{opt} method delivers very accurate system models. In the next section, the accuracy of the obtained estimates with the OMA-PBSID\textsubscript{opt} identification algorithm will be quantified.

5-3 Uncertainty assessment

The major role of this section is to show the quality of different uncertainty quantification methods. The results will be presented in tables. These uncertainty results will be evaluated in the next Section 5-4.

The accuracy of the estimated eigenfrequency \( \hat{f} \) and damping ratio \( \hat{\xi} \) will be assessed using the uncertainty quantification techniques described in Chapter 3 and 4. Uncertainty quantification is the process of representing the uncertainties of the models [16]. The purpose is to assign a measure of the accuracy of identified models, using a single measurement set. The uncertainty of models will be represented in terms of calculating their variance and generating confidence intervals at confidence levels. By calculating confidence intervals for each estimated parameter, we can provide a statement how confident we are that the true parameter lies in that interval with a certain level of confidence. Confidence regions are useful tool, because they provide an interval, instead of a single estimate [6].
With this simulation experiment we aim at showing the performance of the proposed uncertainty quantification techniques with respect to the number of measured data samples $N$. Because, for wind tower applications, the number of available data samples is limited; hence, the performance of the proposed methods from this perspective will be evaluated, as well. It is also interesting to study the effect on the size of past window $p$ of the OMA-PBSID$_{opt}$ identification method on the performance of the considered uncertainty quantification techniques. The PBSID$_{opt}$ method is unbiased asymptotically for $p \to \infty$, and has some small bias for finite $p$ [37, 41, 65]. Increasing the size of $p$ reduces the bias, and can affect the variance [40, 66].

To assign a measure of the uncertainty of model estimates obtained with the OMA-PBSID$_{opt}$ identification algorithm in the previous section, using the bootstrap method, we follow the procedure presented in Table 3-1. To perform the bootstrap we take $B = 1000$ bootstrap replications. The variance $\sigma^2$ is calculated directly from the bootstrap distribution, as discussed in Chapter 3. Additionally, confidence intervals for each estimate are calculated. We use the bootstrap percentile method and the normal bootstrap distribution method to generate the confidence intervals. The percentile method was presented in Section 3-6-2 and the normal bootstrap distribution method in Section 3-6-1. In order to mitigate the effect of statistical noise on calculated estimates, we run multiple Monte Carlo simulations and take the average of each estimate.

For uncertainty assessment with the asymptotic variance technique, we follow the instructions given in Table 4-1. In Chapter 4, we presented two different methodologies for eigenvalues...
sensitivity calculation. Here, we apply both of them. Similarly to the bootstrap, we use the asymptotic variance technique to calculate the variance and then to generate confidence intervals for the obtained parameter estimates. We build the confidence regions by using the method presented in Section 4-6. Again, to decrease the effect of statistical noise on calculated estimates, we run multiple Monte Carlo simulations and take the average.

In addition to calculating the estimates covariance with the bootstrap and the asymptotic variance approach, the true variance of the modal parameter estimates is approximated by means of Monte Carlo simulations, as well. The purpose is to obtain an estimate that is close to the true unknown covariance, which can be used to judge the accuracy of the covariance estimates. The Monte Carlo simulation method has been widely applied for uncertainty quantification due to its simplicity and ease of implementation [6]. With the Monte Carlo simulation, the true variance is approximated by running simulations, instead of performing many experiments on the real plant. In this example, by using 1000 Monte Carlo simulations we calculate the variance of each estimate. Performing 1000 Monte Carlo iterations, delivers 1000 estimates for each modal parameter. The estimates variance is the variance of the identified 1000 estimates. Confidence intervals are calculated, as well. The methodology for generating the confidence regions is the same that we used with the asymptotic variance technique, presented in Section 4-6 and that of the normal bootstrap distribution method 3-6-1, using the estimated covariance. These variance and confidence intervals approximations serve to validate the estimated variance and confidence intervals with the bootstrap and asymptotic variance techniques.

To assign a measure of the accuracy of covariance estimation of each technique - bootstrap, asymptotic variance and Monte Carlo simulation, we perform again Monte Carlo simulations. In such a way, instead of calculating only one estimate of estimates covariance, with Monte Carlo simulations a set of covariance estimates are obtained. From the set of covariance estimates, the mean and the standard deviation of the estimated covariance can be then calculated. This standard deviation is a measure of the quality of obtained estimates [6]. It shows the amount of variation present in the set of estimates, obtained with the Monte Carlo iterations. The smaller the standard deviation, the closer the estimates to their mean. The accuracy of covariance estimation and its standard deviation with Monte Carlo simulation depends on the number of Monte Carlo simulations. Thus, performing a larger number of Monte Carlo iterations would give us more precise estimate of the covariance mean and its standard deviation. However, to perform a larger number of Monte Carlo iterations more computational time is required. The bootstrap, because of its nature, is more computationally involved. The number of data samples used, also affects the time required for computing the uncertainty. Having these consideration in mind, the covariance and its standard deviation will be calculated with more Monte Carlo iterations, using smaller number of data samples.

In order to investigate the performance of the considered uncertainty quantification techniques with respect to the size of the past window, we identify another models using the OMA-PBSIDopt identification method with past window \( p = 60 \) and \( p = 20 \). In addition to varying the size of the past window, the length of data set used for identification and uncertainty quantification is changed, as well. Data sets with different number of data samples are generated, ranging from \( N = 100 \) to \( N = 2000 \). In addition, one experiment with \( N = 10000 \) numbers of data samples is performed and the data is used to verify the convergence properties of the asymptotic variance approach. With the collected data, the accuracy of the obtained estimates is quantified with the two uncertainty quantification methods - the bootstrap and
the asymptotic variance technique. The true variance is approximated by means of 1000 Monte Carlo simulations. To assess the accuracy of estimating the estimates covariance, we perform Monte Carlo simulations for each uncertainty quantification technique.

The covariance estimation results obtained with past window $p = 60$ and number of data samples $N = 200, N = 300, N = 400$ and $N = 500$, are presented in Table 5-4. Those results are most reliable, because 50 Monte Carlo simulations are performed. The covariance estimation results obtained with past window $p = 60$ and number of data samples $N = 2000$ and $N = 10000$, are presented in Table 5-5. The latter table, aims at showing the performance of the uncertainty quantification techniques with larger number of data samples. Finally, the covariance is estimated with past window $p = 20$ and number of data samples $N = 100, N = 250, N = 500, N = 2000$ and $N = 10000$, these results are summarized in Table 5-6.

The estimated confidence intervals with the bootstrap, asymptotic variance and Monte Carlo techniques are presented in Table 5-7. These results are obtained with past window $p = 20$ and number of data samples $N = 100, N = 250, N = 500, N = 2000$ and $N = 10000$.

It is also investigated the performance of the normal bootstrap distribution and bootstrap percentile methods for confidence intervals generation with the bootstrap, with respect to the number of data samples. Results are derived for $N = 100, N = 250, N = 500, N = 2000$ and $N = 10000$ and are presented in Table 5-8.

Two different methodologies for eigenvalues sensitivity calculation with the asymptotic variance technique were used. We compare the covariance estimation with these techniques, again for $N = 100, N = 250, N = 500, N = 2000$ and $N = 10000$ numbers of data samples and the results are presented in Table 5-9.

In addition to quantifying the accuracy of estimated modal parameters, we assign a measure of the accuracy of the estimated eigenvalues $\hat{\lambda}_{1,2}$. To capture the uncertainty in the estimated eigenvalues $\hat{\lambda}_{1,2}$ we use both the bootstrap and asymptotic variance techniques. The uncertainty in the estimated eigenvalues is represented in terms of confidence ellipsoids. If the parameter estimates have a Gaussian probability density function, the covariance matrix can be used to plot the contour lines of the density function [8]. These ellipsoids are centered around the estimated eigenvalue. Similarly to the confidence intervals, the confidence ellipsoids delivers a set of points, with a predefined probability of containing the true parameter. In Figure 5-6 and Figure 5-7, a plot with the true eigenvalues and with the identified eigenvalues and their 99% confidence ellipsoids, is shown. On the top the bootstrap results are presented, while on the bottom we see an eigenvalues plot obtained with the asymptotic variance technique.

In this section, we applied the bootstrap and the asymptotic variance uncertainty quantification techniques as a tools to assess the accuracy of obtained models with OMA-PBSID_{opt} identification method. A couple of test scenarios were considered each with different number of measurement samples in the range between $N = 100$ and $N = 10000$. In the next section, we use these results to investigate the performance of each uncertainty quantification method, with respect to the number of measurement samples $N$. Additionally, we performed another two tests scenarios with different size of the past window $p = 20$ and $p = 60$. And the effect of the past window of the OMA-PBSID_{opt} identification method on the performance of the uncertainty quantification techniques will be investigated, as well.
5-3 Uncertainty assessment

(a) Bootstrap technique.

(b) Asymptotic variance technique.

**Figure 5-6**: Estimated eigenvalues and their 99% confidence ellipsoids using N = 2000 data points. Identified eigenvalues of the system denoted with 'x' and eigenvalues of the true system, denoted with '+'. Estimated confidence ellipsoids represented with a plain line and approximated with Monte Carlo simulation plotted with dotted line.
Uncertainty assessment: a simulation example

(a) Bootstrap technique.

(b) Asymptotic variance technique.

Figure 5-7: Estimated eigenvalues and their 99% confidence ellipsoids using N = 500 data points. Identified eigenvalues of the system denoted with 'x' and eigenvalues of the true system, denoted with '+'. Estimated confidence ellipsoids represented with a plain line and approximated with Monte Carlo simulation plotted with dotted line.
5-3 Uncertainty assessment

<table>
<thead>
<tr>
<th></th>
<th>Monte Carlo</th>
<th>Bootstrap</th>
<th>Asymptotic variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N=200$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Natural frequency</td>
<td>(0.0039 ± 0.0000)</td>
<td>(0.3103 ± 0.0534)</td>
<td>(0.0014 ± 0.0003)</td>
</tr>
<tr>
<td>Damping ratio</td>
<td>(0.0037 ± 0.0000)</td>
<td>(0.0644 ± 0.0099)</td>
<td>(0.0012 ± 0.0003)</td>
</tr>
</tbody>
</table>

| $N=300$ |             |           |                     |
| Natural frequency | (0.9334 ± 0.0028) | (0.9708 ± 0.1008) | (0.5253 ± 0.0298) |
| Damping ratio | (0.8195 ± 0.0022) | (0.7412 ± 0.0701) | (0.4741 ± 0.0315) |

| $N=400$ |             |           |                     |
| Natural frequency | (0.5516 ± 0.0012) | (0.4192 ± 0.0264) | (0.3692 ± 0.0168) |
| Damping ratio | (0.4736 ± 0.0012) | (0.3640 ± 0.0197) | (0.3228 ± 0.0123) |

| $N=500$ |             |           |                     |
| Natural frequency | (0.39 ± 0.0030) | (0.2964 ± 0.0147) | (0.2799 ± 0.0112) |
| Damping ratio | (0.37 ± 0.0039) | (0.2542 ± 0.0122) | (0.2507 ± 0.0082) |

Table 5-4: Covariance estimation and its standard deviation over 50 Monte Carlo simulations using the bootstrap method, asymptotic variance approach and Monte Carlo simulation with different number of measurement samples $N$. OMA-PBSID$_{opt}$ identification method with past window $p = 60$. 

Master of Science Thesis Ivelina Dimitrova
### Table 5-5: Covariance estimation and its standard deviation over 10 Monte Carlo simulations using the bootstrap method, asymptotic variance approach and Monte Carlo simulation with different number of measurement samples $N$. OMA-PBSID\textsubscript{opt} identification method with past window $p = 60$. The uncertainty is not estimated with the bootstrap method for the case with $N = 10000$ due to the increased computational cost.

<table>
<thead>
<tr>
<th>$N$</th>
<th>Monte Carlo</th>
<th>Bootstrap</th>
<th>Asymptotic variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>(7.49 ± 0.03)$e^{-5}$</td>
<td>(5.05 ± 0.24)$e^{-5}$</td>
<td>(6.98 ± 0.25)$e^{-5}$</td>
</tr>
<tr>
<td></td>
<td>(7.25 ± 0.02)$e^{-5}$</td>
<td>(5.14 ± 0.26)$e^{-5}$</td>
<td>(6.20 ± 0.17)$e^{-5}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$N$</th>
<th>Monte Carlo</th>
<th>Bootstrap</th>
<th>Asymptotic variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>10000</td>
<td>(1.36 ± 0.005)$e^{-5}$</td>
<td>-</td>
<td>(1.39 ± 0.023)$e^{-5}$</td>
</tr>
<tr>
<td></td>
<td>(1.34 ± 0.005)$e^{-5}$</td>
<td>-</td>
<td>(1.24 ± 0.016)$e^{-5}$</td>
</tr>
</tbody>
</table>

### Table 5-6: Covariance estimation and its standard deviation over 10 Monte Carlo simulations, using the bootstrap method, asymptotic variance approach and Monte Carlo simulation with different number of measurement samples $N$. OMA-PBSID\textsubscript{opt} identification method with past window $p = 20$.

<table>
<thead>
<tr>
<th>$N$</th>
<th>Monte Carlo</th>
<th>Bootstrap</th>
<th>Asymptotic variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>(0.0030 ± 0.0000)</td>
<td>(0.0018 ± 0.0002)</td>
<td>(0.0011 ± 0.0002)</td>
</tr>
<tr>
<td></td>
<td>(0.0021 ± 0.0000)</td>
<td>(0.0013 ± 0.0002)</td>
<td>(0.0011 ± 0.0002)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$N$</th>
<th>Monte Carlo</th>
<th>Bootstrap</th>
<th>Asymptotic variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>(7.35 ± 0.05)$e^{-4}$</td>
<td>(5.45 ± 0.79)$e^{-4}$</td>
<td>(6.05 ± 0.72)$e^{-4}$</td>
</tr>
<tr>
<td></td>
<td>(6.12 ± 0.03)$e^{-4}$</td>
<td>(5.07 ± 0.69)$e^{-4}$</td>
<td>(5.35 ± 0.53)$e^{-4}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$N$</th>
<th>Monte Carlo</th>
<th>Bootstrap</th>
<th>Asymptotic variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>(2.97 ± 0.01)$e^{-4}$</td>
<td>(2.36 ± 0.24)$e^{-4}$</td>
<td>(3.07 ± 0.20)$e^{-4}$</td>
</tr>
<tr>
<td></td>
<td>(3.04 ± 0.02)$e^{-4}$</td>
<td>(2.36 ± 0.17)$e^{-4}$</td>
<td>(2.72 ± 0.16)$e^{-4}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$N$</th>
<th>Monte Carlo</th>
<th>Bootstrap</th>
<th>Asymptotic variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>(7.08 ± 0.03)$e^{-5}$</td>
<td>(4.99 ± 0.19)$e^{-5}$</td>
<td>(6.87 ± 0.15)$e^{-5}$</td>
</tr>
<tr>
<td></td>
<td>(7.42 ± 0.02)$e^{-5}$</td>
<td>(5.14 ± 0.15)$e^{-5}$</td>
<td>(6.07 ± 0.11)$e^{-5}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$N$</th>
<th>Monte Carlo</th>
<th>Bootstrap</th>
<th>Asymptotic variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>10000</td>
<td>(1.40 ± 0.006)$e^{-5}$</td>
<td>-</td>
<td>(1.41 ± 0.025)$e^{-5}$</td>
</tr>
<tr>
<td></td>
<td>(1.43 ± 0.006)$e^{-5}$</td>
<td>-</td>
<td>(1.25 ± 0.018)$e^{-5}$</td>
</tr>
<tr>
<td>N=100</td>
<td>Monte Carlo</td>
<td>Bootstrap</td>
<td>Asymptotic variance</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>-----------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Natural frequency</td>
<td>0.8932 1.1092</td>
<td>0.9023 1.0697</td>
<td>0.9204 1.0487</td>
</tr>
<tr>
<td>Damping ratio</td>
<td>0.0138 0.1971</td>
<td>0.0187 0.1572</td>
<td>0.0176 0.1439</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N=250</th>
<th>Monte Carlo</th>
<th>Bootstrap</th>
<th>Asymptotic variance</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural frequency</td>
<td>0.9489 1.0557</td>
<td>0.9497 1.0395</td>
<td>0.9502 1.0398</td>
<td>0.9934</td>
</tr>
<tr>
<td>Damping ratio</td>
<td>0.0537 0.1512</td>
<td>0.0647 0.1516</td>
<td>0.0602 0.1504</td>
<td>0.1053</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N=500</th>
<th>Monte Carlo</th>
<th>Bootstrap</th>
<th>Asymptotic variance</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural frequency</td>
<td>0.9691 1.0368</td>
<td>0.9730 1.0325</td>
<td>0.9680 1.0366</td>
<td>1.0023</td>
</tr>
<tr>
<td>Damping ratio</td>
<td>0.0687 0.1374</td>
<td>0.0803 0.1403</td>
<td>0.0769 0.1414</td>
<td>0.1092</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N=2000</th>
<th>Monte Carlo</th>
<th>Bootstrap</th>
<th>Asymptotic variance</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural frequency</td>
<td>0.9876 1.0207</td>
<td>0.9966 1.0242</td>
<td>0.9933 1.0258</td>
<td>1.0095</td>
</tr>
<tr>
<td>Damping ratio</td>
<td>0.0850 0.1188</td>
<td>0.0878 0.1159</td>
<td>0.0862 0.1167</td>
<td>0.1015</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N=10000</th>
<th>Monte Carlo</th>
<th>Bootstrap</th>
<th>Asymptotic variance</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural frequency</td>
<td>0.9968 1.0115</td>
<td>-</td>
<td>0.9949 1.0096</td>
<td>1.0022</td>
</tr>
<tr>
<td>Damping ratio</td>
<td>0.0943 0.1091</td>
<td>-</td>
<td>0.0960 0.1099</td>
<td>0.1030</td>
</tr>
</tbody>
</table>

**Table 5-7:** Confidence intervals estimation with the bootstrap method, asymptotic variance approach and Monte Carlo simulation using different number of measurement samples N. Average results over 10 Monte-Carlo simulations.

<table>
<thead>
<tr>
<th>N=100</th>
<th>Normal bootstrap distribution</th>
<th>Bootstrap percentile</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural frequency</td>
<td>0.9023 1.0697</td>
<td>0.9066 1.0734</td>
<td>0.9845</td>
</tr>
<tr>
<td>Damping ratio</td>
<td>0.0187 0.1572</td>
<td>0.0176 0.1439</td>
<td>0.0807</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N=250</th>
<th>Normal bootstrap distribution</th>
<th>Bootstrap percentile</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural frequency</td>
<td>0.9497 1.0395</td>
<td>0.9457 1.04118</td>
<td>0.9934</td>
</tr>
<tr>
<td>Damping ratio</td>
<td>0.0647 0.1516</td>
<td>0.0707 0.1576</td>
<td>0.1053</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N=500</th>
<th>Normal bootstrap distribution</th>
<th>Bootstrap percentile</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural frequency</td>
<td>0.9730 1.0325</td>
<td>0.9732 1.0339</td>
<td>1.0023</td>
</tr>
<tr>
<td>Damping ratio</td>
<td>0.0803 0.1403</td>
<td>0.0829 0.1427</td>
<td>0.1092</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N=2000</th>
<th>Normal bootstrap distribution</th>
<th>Bootstrap percentile</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural frequency</td>
<td>0.9966 1.0242</td>
<td>0.9965 1.0246</td>
<td>1.0095</td>
</tr>
<tr>
<td>Damping ratio</td>
<td>0.0878 0.1159</td>
<td>0.0884 0.1166</td>
<td>0.1015</td>
</tr>
</tbody>
</table>

**Table 5-8:** Confidence intervals estimation with the bootstrap method, using percentile and standard error methods for confidence intervals calculation. Average results over 10 Monte-Carlo simulations.
<table>
<thead>
<tr>
<th>N</th>
<th>Analytic method</th>
<th>Numerical method</th>
</tr>
</thead>
<tbody>
<tr>
<td>N=100</td>
<td>(0.0011 ± 0.0002)</td>
<td>(0.0011 ± 0.0002)</td>
</tr>
<tr>
<td></td>
<td>(0.0011 ± 0.0002)</td>
<td>(0.0011 ± 0.0002)</td>
</tr>
<tr>
<td>N=250</td>
<td>(6.05 ± 0.72)e^{-4}</td>
<td>(6.05 ± 0.72)e^{-4}</td>
</tr>
<tr>
<td></td>
<td>(5.35 ± 0.53)e^{-4}</td>
<td>(5.35 ± 0.53)e^{-4}</td>
</tr>
<tr>
<td>N=500</td>
<td>(3.07 ± 0.20)e^{-4}</td>
<td>(3.07 ± 0.20)e^{-4}</td>
</tr>
<tr>
<td></td>
<td>(2.72 ± 0.16)e^{-4}</td>
<td>(2.72 ± 0.16)e^{-4}</td>
</tr>
<tr>
<td>N=1000</td>
<td>(6.87 ± 0.15)e^{-5}</td>
<td>(6.87 ± 0.15)e^{-5}</td>
</tr>
<tr>
<td></td>
<td>(6.07 ± 0.11)e^{-5}</td>
<td>(6.07 ± 0.11)e^{-5}</td>
</tr>
<tr>
<td>N=10000</td>
<td>(1.41 ± 0.025)e^{-5}</td>
<td>(1.41 ± 0.025)e^{-5}</td>
</tr>
<tr>
<td></td>
<td>(1.25 ± 0.018)e^{-5}</td>
<td>(1.25 ± 0.018)e^{-5}</td>
</tr>
</tbody>
</table>

**Table 5-9:** Covariance estimation and its standard deviation over 10 Monte Carlo simulations using the asymptotic variance approach with analytic and numerical calculation of the eigenvalues sensitivity.
5-4 Interpretations of the results

In this section, we interpret the uncertainty quantification results obtained with the bootstrap method and asymptotic variance technique. The main objective is to compare the results from the bootstrap to those from the asymptotic variance technique and draw conclusions regarding their performance with respect to the number of measured data samples $N$ and the size of the past window $p$. Additionally, the analytic asymptotic variance technique will be compared to the numerical asymptotic variance technique for estimated covariance estimation. Finally, a comparison will be made between the normal bootstrap distribution and the bootstrap percentile methods for confidence intervals generation. The final part comprises the most important conclusions that are derived based on the analysis carried out in the current section.

5-4-1 Comparison between bootstrap and asymptotic variance approach

In order to assess the performance of the two uncertainty quantification techniques - bootstrap and asymptotic variance with respect to covariance estimation we compare the estimated variance to the approximated covariance obtained by the means of Monte Carlo simulations. In addition to that, we will also compare the estimated confidence intervals with the bootstrap and asymptotic variance technique to the approximated confidence intervals. Finally, the accuracy of calculating eigenvalues confidence ellipsoids with the bootstrap and asymptotic variance technique will be studied. The purpose is to explore the effect of changing the size of the past window and the length of the data set on the accuracy of the two uncertainty quantification techniques.

The effect of the length of the data set on the estimated variance

Here, an analysis is made of the accuracy of estimating the covariance with the bootstrap and the asymptotic variance technique for cases with different length of the data set, varying between $N = 200$ and $N = 2000$. Let us first consider the case when the past window is $p = 60$ and number of measured data samples between 200 and 500, with results given in Table 5-4. It has to be pointed out that those are the most reliable covariance estimates that resulted by performing a higher number of Monte Carlo simulations. According to the obtained results the bootstrap delivers more accurate estimates of the variance than the asymptotic variance technique, when less number of data samples is available. However, for cases with higher amount of data samples such as $N = 2000$, given in Table 5-5, the asymptotic variance technique slightly outperforms the bootstrap method. On the contrary, when the past window is decreased to $p = 20$, the asymptotic method is more accurate than the bootstrap even for cases with smaller amount of data samples available, as can be verified from Table 5-6. However, those results are less reliable, since less number of Monte Carlo simulations have been performed. Furthermore, lowering the size of the past window $p$ increases the bias of identified parameters [37]. Meanwhile, uncertainty quantification techniques assume unbiased parameter estimates [2, 67, 6]. Therefore, this scenario is hard to generalize and it is difficult to draw strong conclusions regarding the accuracy of uncertainty quantification methods. To conclude, regardless of the size of the past window selected, the asymptotic variance technique is slightly more accurate than the bootstrap when a larger number of measured data samples...
is available. Whereas, the bootstrap is more accurate in situations with smaller number of measured data samples.

We know that asymptotic estimates asymptotically tend to converge to the true estimate with increasing the number of data sample $N$. This tendency can be verified by looking at the estimated variance with $N = 10000$ in Table 5-9. These variance estimates are much closer, due to the convergence properties of the asymptotic variance technique [16].

The main implication of this analysis is that in situations with less numbers of data samples available, the bootstrap outperforms the asymptotic variance technique in terms of variance calculation. Likewise, when a larger data set is available, the asymptotic variance technique delivers slightly more accurate variance estimates than the bootstrap method. Nonetheless, in general the performance of the two uncertainty quantification techniques is quite close. In addition to that, the effect of the past windows size $p$ of the OMA-PBSID$_{opt}$ identification algorithm on the covariance estimation is higher in small number of data samples situations. This observation holds for the two uncertainty quantification techniques.

The effect of varying the size of the past window $p$ on the estimated variance

As early stated, changing the size of the past window $p$ can affect the uncertainty estimates. According to theory, uncertainty quantification methods require zero bias of identified parameters [2, 67, 6]. Among others, the accuracy of parameters identification obtained with OMA-PBSID$_{opt}$ depends on the size of the past window $p$. The best Vector Auto-Regressive with eXogeneous input (VARX) predictor is attained when the past window $p$ is chosen as large as possible. Hence, the most accurate system model is expected when $p \rightarrow N$ [68, 37, 40]. For this reason, the size of the past window $p$ is often chosen in relation to the number of available samples $N$ in order to result in optimal choice [65, 41].

To study the accuracy of uncertainty quantification with the two newly proposed techniques with respect to the size of the past window $p$, we look at the results given in Tables 5-4, 5-5 and 5-6. As we are interested in investigating the effect of varying the size of the past window $p$ used in the OMA-PBSID$_{opt}$ identification on the accuracy of uncertainty quantification with the bootstrap and asymptotic variance techniques we compare the variance estimates obtained with $p = 20$ with those with $p = 60$. From the results, it appears that the estimated covariance is affected when the past window $p$ is varied. These observations hold for all uncertainty quantification techniques the bootstrap, the asymptotic variance technique and the Monte Carlo simulation method.

Comparison of the confidence intervals calculation

In the previous part the discussion was dedicated to accuracy of covariance estimation, here we discuss the accuracy of confidence intervals calculation of each uncertainty quantification technique. Recall that the bootstrap and the asymptotic variance technique differ in principle of calculating the variance, but also in the way the confidence intervals are generated. We used two different methodologies to construct the confidence intervals with the bootstrap. For generating the confidence intervals with the bootstrap we used to percentile method and the normal bootstrap distribution method. A comparison between the two bootstrap methods for confidence intervals generation is made in Section 5-4-3. Whereas, in order to construct
the confidence regions with the asymptotic variance technique we used a method that relies on assumption of normality distribution, which is comparable to the normal bootstrap distribution method.

The normal confidence intervals make use of the estimated variance to generate the confidence regions. We already discussed the variance estimation properties of the two uncertainty quantification techniques. Therefore, the same conclusions for the accuracy of variance estimation drawn above hold for the accuracy of normal confidence intervals calculation with the asymptotic variance and bootstrap techniques, as well. The estimated confidence intervals with the normal distribution method are presented in Table 5-7. From the results, it can be verified that when less number of measurement samples \( N \) are available, the bootstrap delivers more accurate confidence intervals. While when a larger number of data samples \( N \) is measured the asymptotic variance technique is more accurate than the bootstrap. In particular, when \( N = 100, N = 250 \) and \( N = 500 \) the asymptotic variance technique produces confidence intervals that deviate more from the approximated with Monte Carlo simulation than the bootstrap. With the increase of the number of measurement samples \( N \), the difference between the confidence intervals generated with the asymptotic and bootstrap techniques becomes much smaller. In general, the confidence intervals generated with the bootstrap are tighter compared to those computed with the asymptotic variance technique.

**Comparison of the confidence ellipsoids**

Here, we discuss the performance of the two uncertainty quantification techniques in terms of confidence ellipsoids calculation of the identified eigenvalues. Given the eigenvalues covariance, confidence ellipsoids of the estimated eigenvalues can be calculated. Therefore, through the plotted confidence ellipsoids it is possible to assess the accuracy of estimating the eigenvalues covariance of each uncertainty quantification method. The accuracy of calculating confidence regions around the estimated eigenvalues will be judged by comparing it to the approximated by means of Monte Carlo simulation.

The confidence ellipsoids associated with the bootstrap are presented in Figures and 5-6a and 5-7a, respectively using \( N = 2000 \) and \( N = 500 \) numbers of measured data samples. The confidence ellipsoids associated with the asymptotic variance technique are given in Figures and 5-6b and 5-7b, respectively using \( N = 2000 \) and \( N = 500 \) numbers of measured data samples. By visual inspection of these plots, we see that when \( N = 2000 \) no significant difference can be observed between the confidence ellipsoids associated with the bootstrap and the asymptotic variance technique. When less numbers of data samples is available, such as \( N = 500 \), the bootstrap delivers more accurate estimates than the asymptotic variance technique. Moreover, the estimated confidence ellipsoids with the asymptotic variance technique are smaller than those obtained with the bootstrap method.

**5-4-2 Comparison between analytic and numerical calculation of eigenvalues covariance.**

Recall from Section 4-4, that there exist different approaches to calculate the eigenvalues covariance. We presented two techniques - an analytic, presented in Section 4-4-1 and numerical, discussed in Section 4-4-2. In view of this, a comparison between these techniques
for estimating eigenvalues covariance is made. For a comparison, we look at the calculated variance for the estimated modal parameters. From the results summarized in Table 5-9, we see that the two methodologies deliver the same results. Therefore, given these results it can be concluded that the numerical and analytic method for calculation of the eigenvalues covariance, presented in Chapter 4 have the same precision.

5-4-3 Comparison between normal bootstrap distribution and bootstrap percentile methods for confidence intervals generation.

With the bootstrap percentile method, the confidence regions are calculated directly from the bootstrap distribution. Due differences in the computation procedure, a special investigation of the estimated confidence intervals is needed. From Table 5-8, we see that increasing the length of the measured signal, the difference between the results obtained with the percentile method and the normal bootstrap distribution method becomes smaller.

5-5 Conclusions

In this chapter, firstly, we examined the novel OMA-PBSID_{opt} identification method by applying it to a second-order LTI system. By means of Monte Carlo simulation we managed to verify its performance. A general conclusion, is that the OMA-PBSID_{opt} identification method delivers accurate system models. In the second and most important part, the accuracy of the two uncertainty quantification techniques was assessed. Their performance was evaluated in terms of accuracy of variance and confidence regions calculation for the estimated system modal parameters. We applied the bootstrap and the asymptotic variance techniques, but as well as the Monte Carlo simulation to estimate these stochastic properties. The bootstrap and asymptotic variance methods are compared based on accuracy and suitability for the uncertainty quantification in the given simulation model. The results highlighted that the bootstrap method is more reliable and accurate than the asymptotic variance approach, in small number of measured data samples situations. In cases where a larger number of data samples is available, the asymptotic variance approach slightly outperforms the bootstrap method, but the difference in their performance is small.
Conclusions and recommendations

6-1 Conclusions

In this thesis, a novel Operational Modal Analysis (OMA) method and two uncertainty quantification techniques were proposed. With these methods we aim at improving the performance of a wind turbine tower system. First, by obtaining more accurate model of the wind turbine tower. And secondly, by more precisely accounting for the uncertainty in the estimated model. Because assigning a measure of the uncertainty of the model will enable the use of less conservative control methods. In addition, obtaining more accurate estimates of the tower’s natural frequency will result in better control. Improving the wind turbine control will in turn contribute to the decrease of wind energy cost, as discussed in Chapter 1.

The objective was to obtain more accurate wind turbine tower models, since the natural frequency of the tower can be more than 10% off from that, designed with first principle methods [5]. We suggested using OMA approach for data-based system modeling, as argued in Chapters 1 and 2. With OMA the system’s model is obtained in such a way that no artificial excitation is required, as with most system identification techniques. In the framework of OMA the excitation generated by the wind is used as an stochastic input to the system, which eliminates the need to artificially excite the system. Moreover, the input signal is not measured and only the output signal is measured, making the OMA identification technique simple and economical to employ. Therefore, the OMA approach is a suitable and easy to apply methodology for the wind turbine tower data driven modeling. A general contribution is that the optimized Predictor Based Subspace Identification (PBSID\textsubscript{opt}) algorithm, which is a system identification method was adapted to work with output-only data, resulting in the new OMA-PBSID\textsubscript{opt} algorithm that was defined in Chapter 2. The performance of the proposed OMA-PBSID\textsubscript{opt} algorithm has been verified in Chapter 5 in a simulation example on a second-order Linear Time-Invariant (LTI) system. The OMA-PBSID\textsubscript{opt} proved to deliver very accurate system models.

In particular, an asymptotic variance and bootstrap techniques has been presented in order to assess the accuracy of estimates. Uncertainty quantification techniques use measurement
data from the real system to estimate and quantify the uncertainty of identified models. One contribution of this research is the development of the bootstrap technique for OMA-PBSID\textsubscript{opt} method which is presented in Chapter 3. The bootstrap is a statistical data-based uncertainty quantification method that is less dependent on the number of measurement samples. For the case of wind turbine tower, the number of available measurement sample is limited, which make the bootstrap very suitable method for quantifying uncertainty. In addition to that, since the bootstrap method is a simulation-based method it is also straightforward to apply. By performing many bootstrap simulations, accurate uncertainty estimates can be obtained, at the price of increased computational time. This principle solves the problems, encountered when applying traditional asymptotic variance techniques in limited amount of measured data samples situations. The main contribution of the bootstrap approach is the formulation of a bootstrap for the OMA-PBSID\textsubscript{opt} method.

The asymptotic variance technique for uncertainty quantification with the novel OMA-PBSID\textsubscript{opt} method, was presented in Chapter 4. Here, uncertainty descriptions of the estimates are derived by relying on the asymptotic statistics theory. These methods make use of approximate expressions for the asymptotic variance. Two types of problems are related to the asymptotic variance techniques. First, variance results may be not enough accurate in cases with small number of measurement samples, because uncertainty descriptions are based on variance approximations. This is a result of violation of the convergence property [8]. The other drawback originates from the complexity of asymptotic variance expressions that need to be derived. However, a significant reduction in computation time compared to the bootstrap method has been noticed.

A comparison has been made between the two presented uncertainty quantification approaches reported in Chapter 5. Since the number of measurement samples is a factor that is likely to affect the results and is important in wind turbine tower modeling, its effect was investigated. The bootstrap methods appears to outperform the asymptotic method at small data sizes, while having only slightly less performance at large data sizes.

Comparing the two uncertainty quantification techniques in terms of computational time, it can be concluded that the asymptotic variance technique outperforms the bootstrap method. The computation time with the bootstrap approach grows when the number of measurement sample increases. This indicates an important limitation of the bootstrap approach.

Each uncertainty assessment methodology has its drawbacks and advantages, and therefore it is difficult to conclude which is the best. In general, the following observations can be formulated. For small number of measured data samples situations, the uncertainty in models can be assessed more accurately and reliably using the bootstrap approach. For situations when a large number of measured data samples are available, the asymptotic variance technique can be used to obtain accurate variance estimates in a reasonable time.

### 6-2 Recommendations

Throughout this thesis, a number of aspects have been observed that could improve the performance and computational time of the considered uncertainty quantification techniques. These observations have lead to the following recommendations for future research.
A major contribution can be made by reducing the computational complexity of the bootstrap method. Since the bottleneck of the bootstrap method is the increased computational time for situations with large number of measured data samples $N$. In fact, replicating the estimation process $B$ times could happen to be extremely time consuming even for moderate samples sizes. If one is capable of computing the required estimates in a very efficient way this would make it possible to obtain uncertainties estimates in reasonable time even for situations with larger number of data samples $N$.

One possibility for speeding up the bootstrap in the framework of the PBSID$_{opt}$ subspace identification method is to use Singular-Value Decomposition (SVD)-free techniques for solving the two least squares problems. The SVD method is expensive to compute, as it requires $2mn^2 + 11n^3$ flops, but it most applicable as it is numerically stable and can handle rank deficiencies [9].
Appendix A

Preliminaries

A-1  Kronecker product

In Chapter 4 the Kronecker algebra has been used to present an expressions for the asymptotic variance.

The Kronecker product is a kind of matrix product, that is denoted by $\otimes$. This operation is used for matrices with arbitrary size. The Kronecker product of two matrices $A \otimes B$, where $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{p \times q}$ is a block matrix with size $mp \times nq$ defined as:

\[
A \otimes B = \begin{bmatrix}
    a_{11}B & a_{12}B & \cdots & a_{1n}B \\
    a_{21}B & a_{22}B & \cdots & a_{2n}B \\
    \vdots & \vdots & \ddots & \vdots \\
    a_{m1}B & a_{m2}B & \cdots & a_{mn}B
\end{bmatrix}.
\]

A-2  Singular value decomposition

Recall that in Chapter 2 we used Singular-Value Decomposition (SVD) in the expressions for recovering the system matrices in the Operational Modal Analysis (OMA)-optimized Predictor Based Subspace Identification (PBSID$_{opt}$) identification algorithm. By computing the SVD of the estimate of the extended observability matrix times the state sequence it was possible to estimate the state sequence, as well as to determine the order of the system $n$.

The SVD is a very useful factorization of a matrix. The SVD of a matrix $A \in \mathbb{R}^{m \times n}$ has the following form:

\[
A = U \Sigma V^T,
\]

where $\Sigma \in \mathbb{R}^{m \times n}$ is a matrix with nonzero elements on the diagonal, that are the singular values $\sigma_i$ of the matrix $A$. The singular values $\sigma_i$ are ordered in a increasing order $\sigma_1 > \sigma_2 > \cdots > \sigma_n$. 

Master of Science Thesis  Ivelina Dimitrova
\[ \sigma_k, \text{ where } k = \min(m,n). \] The matrices \( U \in \mathbb{R}^{m \times m} \) and \( V \in \mathbb{R}^{n \times n} \) are orthogonal matrices. The columns \( u_i \) of the matrix \( U \) are the \( i \)th left singular vectors and the columns \( v_i \) of the matrix \( V \) are the \( i \) right singular vectors.

Considering the matrix \( A \in \mathbb{R}^{m \times n} \) has a rank \( r \), such that \( r < m \) and \( r < n \), we can partition the SVD of \( A \) in the following way:

\[
A = \begin{bmatrix}
U_1 & U_2 \\
\Sigma_1 & 0 & \vdots \\
0 & 0 & V_1^T & V_2^T 
\end{bmatrix},
\]

where \( U_1 \in \mathbb{R}^{m \times r}, U_2 \in \mathbb{R}^{m \times (m-r)}, \Sigma_1 \in \mathbb{R}^{r \times r}, V_1 \in \mathbb{R}^{n \times r}, V_2 \in \mathbb{R}^{n \times (n-r)}. \) From this partition we directly notice that the columns of the matrices \( U_1, U_2, V_1 \) and \( V_2 \) provide orthogonal bases for the four subspaces of the matrix \( A \):

\[
\begin{align*}
\text{range}(A) &= \text{range}(U_1), \\
\ker(A^T) &= \text{range}(U_2), \\
\text{range}(A^T) &= \text{range}(V_1), \\
\ker(A) &= \text{range}(V_2).
\end{align*}
\]

These are very important orthogonal projections associated with the SVD.

### A-3 Eigenvalue decomposition

In Chapter 2, we relied on eigenvalues decomposition property to derive expressions for the modal parameters. In addition, the eigenvalues decomposition served as a tool for deriving the expressions for analytic calculation of eigenvalues sensitivity in Chapter 4.

The eigenvalue decomposition is a factorization of a matrix, represented by its eigenvalues and eigenvectors. The eigenvalue decomposition of a matrix \( A \in \mathbb{R}^{N \times n} \), that has \( n \) linearly independent eigenvectors, is defined as:

\[
A = V\Lambda V^{-1},
\]

where \( \Lambda \in \mathbb{R}^{n \times n} \) is a diagonal matrix with the eigenvalues of the matrix \( A \) along its diagonal, and the columns of the matrix \( V \in \mathbb{R}^{n \times n} \) correspond to the eigenvectors of the matrix \( A \).

Only diagonalizable matrices can be decomposed in that way.

### A-4 Hankel matrix

To derive the expressions for the OMA-PBSID\(_{\text{opt}}\) identification method in Chapter 2, the Hankel type of matrix has been extensively used. Those Hankel matrices were constructed from the data sequences.
A Hankel matrix is a matrix that is symmetric and constant across the anti-diagonals. The Hankel matrix $\mathbf{H}_{(n+1),(n+1)}$, where $n$ is the order of a LTI system, constructed from the Markov parameters $h(k)$, $k = 1, ..., N$ is defined as:

$$
\mathbf{H}_{(n+1),(n+1)} = \begin{bmatrix}
h(1) & h(2) & \cdots & h(n + 1) \\
h(2) & h(3) & \cdots & h(n + 2) \\
\vdots & \vdots & \ddots & \vdots \\
h(n + 1) & h(n + 2) & \cdots & h(2n + 1)
\end{bmatrix}.
$$

A block Hankel matrix is a matrix with vector entries that are constant along the block anti-diagonals. The block Hankel matrix $\mathbf{Y}_{i,s,N}$ constructed from a sequence $y(k)$, $k = 1, ..., N$ is defined as:

$$
\mathbf{Y}_{i,s,N} = \begin{bmatrix}
y(i) & y(i + 1) & \cdots & y(i + N - 1) \\
y(i + 1) & y(i + 1) & \cdots & y(i + N - 1) \\
\vdots & \vdots & \ddots & \vdots \\
y(i + s - 1) & y(i + s) & \cdots & y(i + N + s - 2)
\end{bmatrix}.
$$

The first entry $i$ of the subscript of the Hankel matrix $\mathbf{Y}_{i,s,N}$ refers to the time index, the second entry $s$ refers to the number of block-rows, and the third entry $N$ refers to the number of columns.

**A-5 Jacobian matrix**

For numerical calculation of the eigenvalues sensitivity in Chapter 4, the Jacobian matrix had to be constructed and solved.

The Jacobian matrix is a matrix containing all first-order partial derivatives of a vector-values function. Consider a vector-values function $\mathbf{f}$ that takes as an input the vector $\mathbf{x} \in \mathbb{R}^n$ and produces the output $\mathbf{f}(\mathbf{x}) \in \mathbb{R}^m$. The Jacobian matrix $\mathbf{J} \in \mathbb{R}^{m \times n}$ of the function $\mathbf{f}$, defined in the following way:

$$
\mathbf{J} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \begin{bmatrix}
\frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n}
\end{bmatrix}.
$$


Ivelina Dimitrova Master of Science Thesis


# Glossary

## List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BC$_a$</td>
<td>Bias-corrected and accelerated</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-Noise Ratio</td>
</tr>
<tr>
<td>ABC</td>
<td>Approximate bootstrap confidence intervals</td>
</tr>
<tr>
<td>IID</td>
<td>independent and identically distributed</td>
</tr>
<tr>
<td>MOESP</td>
<td>Multivariable Output-Error State sPace</td>
</tr>
<tr>
<td>MIMO</td>
<td>Multivariable Input, Multiple Output</td>
</tr>
<tr>
<td>SISO</td>
<td>Single Input, Single Output</td>
</tr>
<tr>
<td>SVD</td>
<td>Singular-Value Decomposition</td>
</tr>
<tr>
<td>PEI</td>
<td>Prediction Error Identification</td>
</tr>
<tr>
<td>SMI</td>
<td>Subspace Model Identification</td>
</tr>
<tr>
<td>ARX</td>
<td>Auto-Regressive with eXogeneous input</td>
</tr>
<tr>
<td>VARX</td>
<td>Vector Auto-Regressive with eXogeneous input</td>
</tr>
<tr>
<td>PBSID$_{opt}$</td>
<td>optimized Predictor Based Subspace Identification</td>
</tr>
<tr>
<td>AR</td>
<td>Auto-Regressive</td>
</tr>
<tr>
<td>ARMA</td>
<td>Auto-Regressive Moving Average</td>
</tr>
<tr>
<td>PBSID</td>
<td>Predictor-Based Subspace Identification</td>
</tr>
<tr>
<td>LPV</td>
<td>Linear Parameter-Varying</td>
</tr>
<tr>
<td>LTI</td>
<td>Linear Time-Invariant</td>
</tr>
<tr>
<td>OMA</td>
<td>Operational Modal Analysis</td>
</tr>
</tbody>
</table>

Master of Science Thesis

Ivelina Dimitrova
<table>
<thead>
<tr>
<th><strong>PDF</strong></th>
<th>Probability Density Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CDF</strong></td>
<td>Cumulative Distribution Function</td>
</tr>
</tbody>
</table>