Estimation and control of large-scale systems with an application to adaptive optics for EUV lithography

Aleksandar Haber

ESTIMATION AND CONTROL OF LARGE-SCALE SYSTEMS WITH AN APPLICATION TO ADAPTIVE OPTICS FOR EUV LITHOGRAPHY

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op

dinsdag 4 februari 2014 om 12:30 uur

door

Aleksandar HABER

Mechanical Engineer, University of Belgrade, Servië
geboren te Belgrado, Servië

Dit proefschrift is goedgekeurd door de promotor:
Prof. dr. ir. M. Verhaegen

Samenstelling promotiecommisie:

Rector Magnificus,	voorzitter
Prof. dr. ir. M. Verhaegen,	Technische Universiteit Delft, promotor
Prof. ir. R.H. Munnig Schmidt,	Technische Universiteit Delft
Prof. dr. W.M.J.M. Coene,	Technische Universiteit Eindhoven and ASML
Prof. dr. ir. N.J. Doelman,	Universiteit Leiden and TNO
Prof. dr. M.R. Jovanović,	University of Minnesota
Prof. dr. A. Alessandri,	University of Genoa
Prof. dr. M. Diehl,	Katholieke Universiteit Leuven and University of Freiburg
Prof. dr. ir. H. Hellendoorn,	Technische Universiteit Delft, reservelid

This research is supported by the Dutch Ministry of the Economic Affairs and the Provinces of Noord-Brabant and Limburg in the frame of the "Pieken in de Delta" program.

ISBN: 978-94-6203-510-2

Copyright © 2014 by Aleksandar Haber.

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without written permission from the copyright owner.

Printed by Wöhrmann Print Service, Zutphen, The Netherlands.

Acknowledgments

First of all, I would like to thank my supervisor Michel Verhaegen for giving me the opportunity to demonstrate my research talents. Furthermore, I would like to thank him for teaching me how to do research and for the thrust and support that he gave me during the course of my PhD research.

Next, I would like to thank Alessandro Polo from Optics Research Group, Delft University of Technology, for teaching me optics and for being an excellent research companion during the last 4 years. Without his help, my thesis would not contain the experimental part.

I would like to thank Dr. Rufus Fraanje for being my daily supervisor during the first two years of my research. Next, I am grateful to professor Paul Urbach and professor Silvania Pereira, from Optics Research Group, Delft University of Technology, for reading my papers and for giving me useful comments. Furthermore, I would like to thank Rudolf Saathof, professor Rob Munnig Schmidt and Jo Spronck, from the department of Precision and Microsystems Engineering, Delft University of Technology, for continuously reminding me that the control theory should always be connected with real-life applications.

I would also like to thank Simon Ravensbergen and professor Nick Rosielle from Eindhoven University of Technology, for designing and providing us with a prototype of thermally actuated deformable mirror. Using this mirror we successfully demonstrated the proof of concept for predictive control of thermally induced wavefront aberrations in optical lithography machines. I would like to thank the project partners from ASML: Anton van Dijsseldonk, Jan van Schoot and Bob Streefkerk, for always being there to answer my questions about optical lithography and for steering my research in a more practical direction. I would like to thank Bernhard Kneer from Carl Zeiss for answering my questions about EUV mirrors and for attending our project meetings in the Netherlands. I am grateful to Henk Kiela from Opteq and Merijn Voets from Nadinsco, for giving me constructive criticism during our project meetings. I would like to thank professor Niek Doelman for organizing and charing our projecting meetings and for giving me useful feedback. I would also like to thank the committee members for the time that they invested to read this thesis and to give me useful feedback.

Next, I would like to thank my friends and colleagues in the Netherlands. Filipe, care, hvala ti za sve! Aydin, I am still riding your bicycle! Thank you for everything. Nikola, thank you for your support and friendship. Thank you Snezana

for giving me survival tips when I came here and for always being there to answer my questions. Thank you Pawel for your friendship that survived all the "discussions" that we had. I know that you are not a fan of pan-Slavic idea, but it is true that people from Slavic countries have many things in common. Thank you Yashar for your friendship and for inventing "the 2.25 euros donation yoke". Thank you Marco for organizing dinners and for hosting barbecues in Pijnacker. Thank you Andrea, Dang, Amol, Jianfei, Stefan and Samira for all the funny moments and for nice discussions that we had. Paolo thank you for sending me your thesis template. Max, thank you very much for translating the summary of this thesis. Pieter, thank you for translating the propositions.

Thank you Ali, Alfredo, Mohammad, Matteo, Vedran, Justin, Harsh, Mathieu, Rudy, Tamas, Alessandro, Jan-Willem, Paul, Gabriel, Arjan, Olaf, Arturo, Ivo, Roland, Amir, Zhe, Sadegh, Dieky, Pieter, Patricio, Ana, Robert, Arne, Noortje, Zulkifli, Bart, Ana, Mehdi, Subramanya, Shuai, Esmail, Bart, Sachin, Max, Yue, Ilya, Patricio, Edwin, Kim, Yihui, Le, Renshi, Ilhan, Jia and Hans Hellendoorn for being nice to me and for saying "Hi!" (with a capital letter) when we meet each other in the coffee room or in the corridor.

Ellen thank you for answering my emails and for helping me to fill in the IND form. Kitty, Saskia, Esther and Marieke, thank you very much for everything! By my opinion, you are the best part of the DCSC. I really enjoy to communicate with you.

I would also like to thank the researchers from the smart optics group. Ruxandra, thank you for helping me to write papers, for checking my propositions and for all the support in the last 4 years. Jacopo, Federico, Tope, Carlas, Jeroen, Elisabeth, Hans Verstraete and Hans Yoo, I really enjoyed working with you!

Raluca, thank you a lot for all the help and support during the last 3 years. Finally, thank you Yu Hu for being one of the best friends that I ever had.

Aleksandar Haber,
December 2013, Delft

Contents

1 CHAPTER

Introduction

In this introductory chapter we explain the main challenges of controlling thermally induced wavefront aberrations in extreme ultraviolet lithographic machines. We furthermore explain how this practical control problem is related to some fundamental, open problems in estimation and control of large-scale interconnected systems. Finally, we briefly explain our solutions to these problems and we explain the organization of the thesis.

1.1 Motivation

In systems and control and more generally in the applied mathematics field, fundamental theories and methods are usually developed by searching for solutions of real-life problems. We can say that a developed theory or a method deserves to be called "fundamental" if its application is not only limited to the specific problem that initiated the research, and if it can be applied to much broader class of real life problems. A typical example of a fundamental, numerical method that was developed to solve a real-life problem, and that is used today to solve a large variety of practical problems, is the celebrated method of least squares.[1]

Following this path, our research in the past 4.5 years was motivated by a real-life engineering problem. The problem that we tried to solve was compensation of thermally induced wavefront aberrations in the next generation of optical lithography machines[2]. As we will explain in the sequel, the solution of this problem required development of a completely novel approach to estimation and control

[1] One of the main motivations for the development of the least squares method was to calculate the orbits of heavenly bodies. This method is used today in almost every field of science. Even in this thesis, control and estimation problems are formulated either as basic least-squares problems or as their constrained versions. The mathematicians that founded the method of least squares were C. F. Gauss and A.-M. Legendre.

[2] This research is supported by the Dutch Ministry of the Economic Affairs and the Provinces of Noord-Brabant and Limburg in the frame of the "Pieken in de Delta" program.

of large-scale systems. The application of this novel approach goes beyond the field of optical lithography, and the developed estimation and control methods can be applied to a wide class of large-scale systems.

In the sequel we will explain in detail why the problem of controlling thermally induced wavefront aberrations is so interesting from the systems and control perspective, and how this problem motivated us to develop numerical methods that can be applied to a large variety of real-life problems. We first start with a brief introduction to optical lithography.

1.2 A brief introduction to optical lithography

Almost every six months there is a new version of a smart phone or a computer on the market. These new electronic devices have improved characteristics: they have more memory, they are faster and they have better graphics than their older versions. This technological progress is possible because of the continuous miniaturization of electronic components. It has been observed that the number of transistors on integrated circuits doubles approximately every 18 months.[3]. This trend is dictated by the chip manufacturers such as: Intel, Samsung Electronics, AMD and others. To make Integrated Circuits (ICs) these companies are using optical lithography machines. The world's largest producers of optical lithography machines are ASML and Nikon.

The main components of an optical lithography machine (lithographic machine) are illustrated in Fig 1.1. In lithographic machines, mask's patterns that represent an image of ICs, are optically projected (exposed) onto a light-sensitive photoresist on a semiconductor wafer.

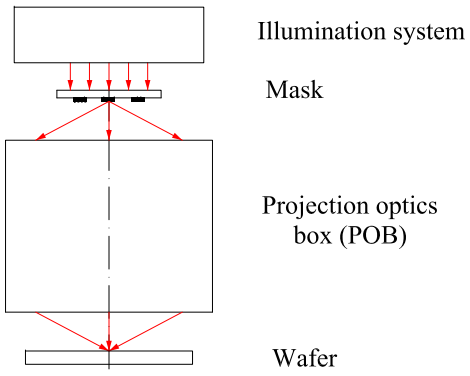


Figure 1.1: The main components of a lithographic machine.

[3]The first observation about the rate at which number of transistors on integrated circuits doubles is formulated by the Moore's law. This law dates back to 1965, and it states that the number of transistors on integrated circuits doubles each two years. Since then, it has been observed that, on average, the number of transistors on integrated circuits doubles every 18 months

The illumination system consists of a light source and optical elements (mirrors and/or lenses). Its purpose is to deliver light to the mask and consequently, to the Projection Optics Box (POB). Masks that are used in the current lithographic machines are transparent optical elements on which chip patterns have been formed [4]. As the light passes through the transparent areas of the mask it diffracts. The POB captures a portion of the diffracted light and it projects the mask image onto the wafer. The projected image is typically four times smaller than the original mask image (mask patterns). After the wafer has been exposed, it goes through the series of chemical treatments that create one layer of the IC. Once this process is finished, new material is deposited on the wafer and photoresist is added. After this has been done, a new cycle of exposure and chemical treatments starts. Before the chips are fully complete, this cycle can be repeated more than 30 times. For more information about optical lithography see [1] and references therein.

In the early days of optical lithography visible light was used to transfer (print) patterns from a mask to a wafer. Since then, light of shorter and shorter wavelengths has been used to transfer patterns on a wafer. The wavelength decreased from blue (wavelength of 436 nm) to Ultra-Violet (UV, 365 nm) and from UV to deep UV (248 nm) [1; 2]. In the current generation of optical lithography machines, the wavelength of 193 nm is used to expose wafers [3]. This decrease of wavelength is driven by the need to print smaller and smaller features. Namely, the Rayleigh resolution equation states that the smallest printable feature (the Critical Dimension (CD) or the resolution of a lithographic system) is proportional to the wavelength λ of the used light [1]:

$$CD = k_1 \frac{\lambda}{NA} \tag{1.1}$$

where k_1 is a coefficient that depends on the imaging process and NA is the Numerical Aperture of the projection system. From (1.1) we see that one of the ways to reduce the CD is to decrease λ. This is why in the new generation of optical lithography machines, 13.5 nm Extreme UltraViolet (EUV) light is used for wafer exposure [4]. In this thesis the new generation of optical lithography machines will be called the *EUV Lithography (EUVL) machines*. Because the EUV light is absorbed by most of refractive materials used in optics, refractive optical elements (lenses and transmission masks) are not used in EUVL machines. Instead, reflective optical elements (mirrors and reflective masks) are used [5; 6]. Furthermore, because air absorbs EUV radiation, the optics and stages have to operate under vacuum conditions [4].
From (1.1) it also follows that the CD can be reduced by increasing NA or by decreasing the k_1 factor. In practice this is achieved by using resolution enhancement techniques [1], such as phase shifting masks, off-axis illumination, optical proximity correction or by using immersion lithography.

Apart from being able to print ICs with smaller dimensions, each new generation of lithographic machines has higher productivity[5]. Usually this productivity is

[4] In the next generation of lithographic machines transparent masks will be replaced by the reflective masks.

[5] This productivity increase is driven by economical reasons.

expressed by the number of wafers produced per hour and it is called *the wafer throughput*. In the current optical lithography machines throughput numbers exceed 200 wafers per hour [7].

Due to the resolution enhancement techniques and increased throughput, the power transmitted through the projection optics of lithography machines constantly increases. The optical elements absorb a portion of the exposure energy. The absorbed energy transforms into heat, which induces thermoelastic deformations of optical elements [7; 8; 9; 10; 11; 12; 13; 14; 15]. Furthermore, heating creates variation of refractive index of lenses in the POB. Consequently, the heating process induces wavefront aberrations in lithographic machines. If not compensated, wavefront aberrations can seriously compromise the resolution of a system [7; 9; 10; 11; 12]. In this PhD thesis, wavefront aberrations induced by heating of optical elements will be called the *Thermally Induced Wavefront Aberrations* (TIWA). In EUVL machines degradation of resolution due to the thermoelastic deformation of optical elements is even more severe [16; 17; 18; 19]. This is mainly because each optical component in the EUVL machine absorbs around 30% of EUV radiation.
To better explain the problem of thermally induced wavefront aberrations in EUVL machines, consider a segment of a mirror used in EUVL machines that is illustrated in Fig. 1.2. The incoming wavefront is undistorted (flat). The light beam creates a non-uniform heat flux distribution over the mirror's top surface. The heat flux distribution induces a nonuniform temperature distribution and it creates thermoelastic deformations. Because the top surface of the mirror is deformed, the reflected wavefront is no longer flat. Instead, it becomes distorted (aberrated). These wavefront distortions (wavefront aberrations) compromise the resolution of the printed patterns.

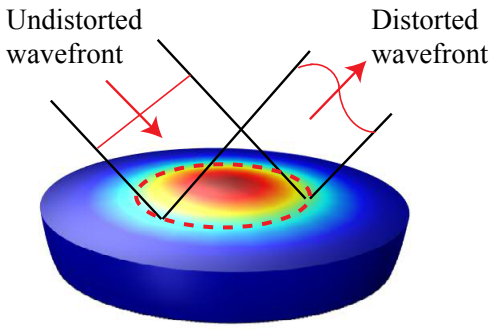


Figure 1.2: A segment of a mirror used in EUV lithography and thermally induced wavefront aberrations.

Beside optical lithography machines, TIWA can limit performance of a large variety of high power optical systems[6]. For example, in gravitational wave interfer-

[6]High power optical systems are systems that use powerful lasers or systems in which highly focused beams pass through small sections of the optics.

ometers high power lasers induce aberrations that can significantly decrease the sensitivity of the instruments [20; 21; 22]. TIWA can also degrade the beam quality of the lasers used in material processing [23; 24]. Furthermore, the performance of optical systems used in military lasers [25] can be significantly degraded by TIWA.

One of the possible solutions for compensation of TIWA in optical lithography machines, as well as in other high power optical systems, is to use the Adaptive Optics (AO) technique [2; 20; 26; 27; 28; 29; 30; 31]. Because the basic principles of the AO technique are relatively unknown to the broader control community, in the sequel we briefly explain this wavefront correction technique.

1.3 Basic principle of AO

AO is a well-established technique for correcting wavefront distortions in optical systems. The basic principle of AO is to measure wavefront aberrations using a WaveFront Sensor (WFS) and to compensate them by changing the geometry of an active optical element in the system. Widely used active optical elements in AO systems are Deformable Mirrors (DMs) and spatial light modulators.
Figure 1.3 is a simplified illustration of wavefront correction using a DM. The incoming wavefront, traveling from left to right, deviates from the flat wavefront. In order correct the distorted wavefront, the mirror contains a depression. At the moment when the distorted part of the wavefront reaches the bottom surface of the depression, the distance between the flat region of the wavefront and the DM is a. By the time the flat part of the wavefront reaches the mirror surface, the distorted part is reflected and it has traveled a distance from right to left. Because the depth of the mirror depression is two times smaller than the wavefront distortion, the reflected wavefront is completely flat.

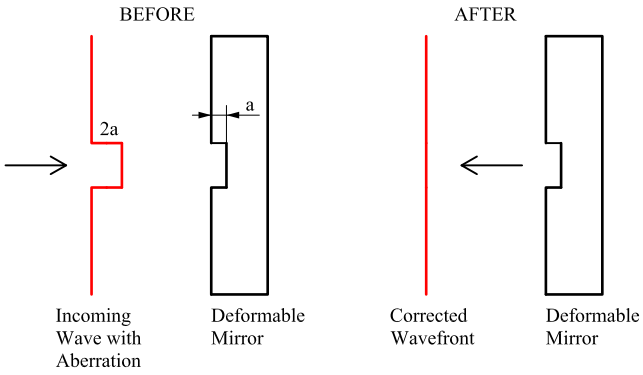


Figure 1.3: The basic principle of correcting distorted wavefront using a DM. This schematic is taken from [32].

One of the first successful, non-military applications of AO was in astronomy

[32; 33; 34] to compensate wavefront aberrations induced by atmospheric turbulence. Nowadays, the AO technique is used in the fields of microscopy [35; 36], ophthalmology [37], tomography [38], laser beam shaping [39], optical communication [40] and more recently in lithography [2; 26]. Some other applications of AO can be found in [28; 41; 42; 43; 44; 45; 46].

The main components of an AO system are illustrated in Fig. 1.4(a).

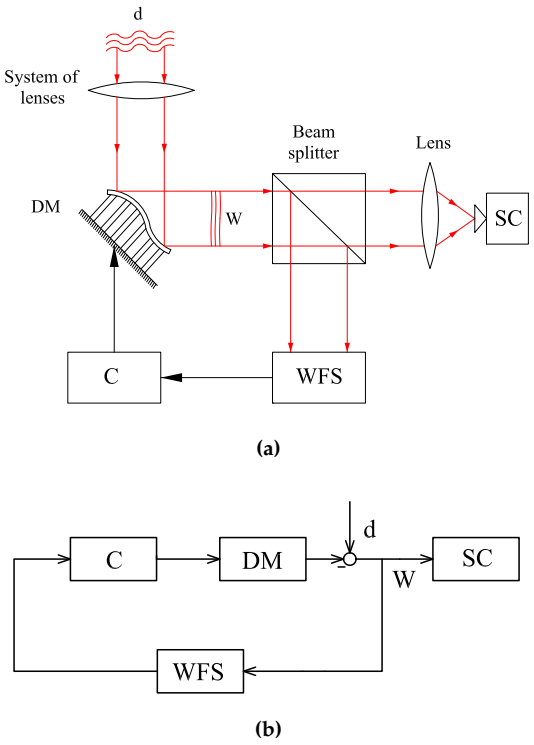


Figure 1.4: (a) Basic principle of AO; (b) Feedback loop in the AO system.

The main goal of the AO system is to correct a distorted wavefront[7] d. The light coming from an object enters into the AO system through the system of lenses[8]. The system of lenses directs the distorted wavefront d (or aberrated wavefront) towards a DM. The DM corrects the wavefront aberrations. In Fig. 1.4(a), the corrected wavefront is denoted by W. After wavefront correction, the beam splitter splits the beam of light into two beams: the reflected beam that is directed to the WFS and transmitted beam, that is directed to the Science Camera (SC). The SC

[7] From the control engineering perspective, the wavefront distortions or wavefront aberrations are disturbances that need to be suppressed.

[8] In AO for astronomy, an object can be a star or a distant galaxy. On the other hand, in microscopy an object can be a piece of human tissue. In astronomical AO, wavefront aberrations originate from the atmospheric turbulence [33; 34]. However, in industrial, medical and military applications of AO, wavefront aberrations can originate from various sources [47; 48; 49; 50].

forms an image of the object[9]. The WFS measures aberrations of the corrected wavefront and it sends its measurements to the Controller (C). On the basis of these measurements, the controller calculates and sends the control signal to the DM. Driven by the control signal, the optical surface of the DM deforms and it corrects wavefront aberrations.

The feedback loop in the AO system is shown in Fig. 1.4(b). In this thesis, the interconnection structure between the controller, the WFS and the DM that is illustrated in Fig. 1.4 will be called *the standard AO system*. One of its main characteristics is that the controller can receive WFS measurements at an arbitrary sampling time instant. Unfortunately, in optical lithography wavefront aberrations cannot be measured at an arbitrary time instant and consequently, the standard AO systems cannot be used. This motivates us to develop new AO systems that can overcome this limitation.

1.4 AO for (EUV) lithography

In the current generation of optical lithography machines, active optical elements are used for correction of wavefront aberrations [7; 12]. Because the EUV light is aborbed by refractive optical elements, in the new generation of optical lithography machines DMs need to be used for wavefront correction [2; 26].

In the current generation of optical lithography machines, wavefront aberrations are measured using wavefront sensors that are based on the lateral shearing interferometry [3; 51]. The accuracy of these sensors is in the order of $\lambda/400$, where λ is the exposure wavelength [51]. The possibility to use Hartmann sensors for wavefront measurement in EUVL machines has been investigated in [52; 53]. Furthermore, in [54; 55] measurement techniques based on Shack-Hartmann WFS (S-H WFS) have been proposed for measurement of TIWA in high power optical systems. A point diffraction interferometer and a lateral shearing interferometer suitable for operation in the EUV range, are described in [56].

In optical lithography machines, wavefront aberrations are measured at the wafer level (see Fig. 1.1). However, during the exposure process it is not possible to measure wavefront aberrations. Wavefront aberrations can be measured after each exposed wafer. Furthermore, because the measurement time reduces the wafer throughput, it should be as short as possible. One way to reduce the total measurement time is to perform measurements only after exposure of a few initial wafers.
This measurement scenario is illustrated in Fig. 1.5. During the initial measurement time, after each exposed wafer, wavefront aberrations are measured (denoted by 'x' in Fig. 1.5). The question mark in Fig. 1.5 indicates that after initial measurements are taken, wavefront aberrations should not be measured anymore. Instead, their future behavior should be predicted. A controller in an AO system should use this prediction to compensate future wavefront aberrations.

[9] In some applications of the AO, such as optical lithography, the SC is not used.

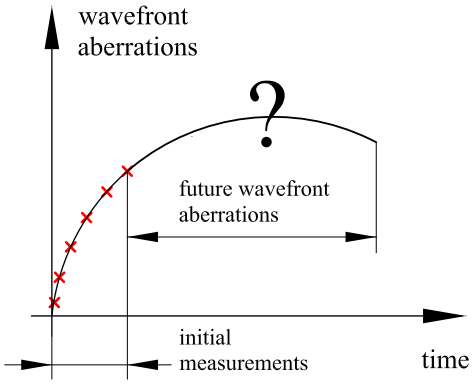


Figure 1.5: Measurement and prediction of wavefront aberrations in a lithographic machine. On the basis of the initial wavefront measurements, the future wavefront aberrations need to be predicted and compensated.

It is obvious that TIWA in lithographic machines cannot be compensated using the standard AO system illustrated in Fig. 1.4(a)-(b). The block diagram of an AO system for compensation of TIWA in lithographic machines, is shown in Fig. 1.6. In contrast to the standard AO system that is illustrated in Fig. 1.4, the WFS can measure aberrations only during the initial time period when the switch S is closed. After the initial time period is finished, the switch S is opened, and wavefront aberrations cannot be measured anymore. However, independently from that, the controller should predict and compensate wavefront aberrations in real-time.

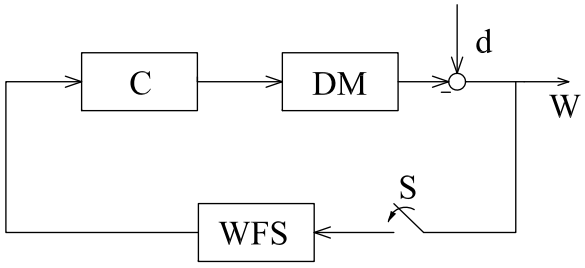


Figure 1.6: The block diagram of the AO system for compensation of TIWA in lithographic machines.

To anyone who has at least a basic understanding of control systems it should be clear that the accurate prediction of TIWA cannot be performed only on the basis of the wavefront measurements. For accurate prediction it is necessary to develop a model of TIWA. Furthermore, this model should be in the form that suitable for real-time control applications.

1.5 Model of TIWA for real-time prediction and control

The dynamical behavior of TIWA in optical lithography machines is mainly determined by *the exposure conditions*, such as: numerical aperture, source shape, reticle and mask pattern diffraction, exposure dose, throughput and resist stack [7].
The model of TIWA describes how the exposure conditions influence the dynamical behavior of the wavefront aberrations. The model of TIWA consists of the two main parts [7] that are illustrated in Fig. 1.7.

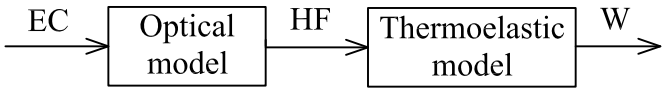


Figure 1.7: The main parts of the TIWA model.

The first part that is denoted by "Optical model" in Fig. 1.7, relates the Exposure Conditions (EC) with the Heat Flux (HF) distribution (intensity distribution or distribution of exposure energy) over the surfaces of optical elements in EUVL machines. This relation is established by computing the full mask diffraction orders which are then convoluted with the illumination source to obtain the diffraction pattern [7]. The computed diffraction pattern determines the HF distribution on the optical elements. In this thesis, the EC will be called *the inputs of the TIWA model*.
The second part, denoted by "Thermoelastic model" in Fig. 1.7, consists of thermoelastic Partial Differential Equations (PDEs) that relate the HF distribution with the temperature change and deformations of the optical elements [22; 57]. The deformations of optical elements determine the wavefront aberrations (W).

The model of TIWA can be obtained using two approaches. The first approach relies on first principles modeling. For example, the TIWA model can be derived by discretizing thermoelastic equations using the finite element method. The second modeling approach is to identify the model directly from experimental data [58]. Experimental data can be collected during the testing and calibration of a lithographic machine. However, during the imaging process the exposure conditions usually differ from the ones used in machine calibration and testing. For example, patterns of the masks used in machine testing can significantly differ from the patterns of the masks used for wafer exposure. In mathematical terms this means that during the exposure process, the inputs of the TIWA model are usually unknown and they need to be determined. There are two ways to overcome this problem. If the new exposure conditions can be modeled (in real-time) then the inputs of the TIWA model can be updated. For example, if a new mask is projected and if the model of its geometry is available, then it is possible to calculate new heat-flux distributions on the optical elements [7]. However, this is a computationally chal-

lenging problem that needs to be solved in real-time [1]. Another approach, that we propose in Chapter 10, is to estimate unknown inputs of the TIWA model or to directly estimate intensity distributions on optical elements from measurements of wavefront aberrations.

In this thesis, we will be mainly concerned with the development of the thermoelastic model that is suitable for real-time prediction and control of TIWA. The development of the optical model is left for future research.

Generally speaking, the thermoelastic model has to meet the following requirements:

1. It has to accurately describe the dynamical behavior of wavefront aberrations. As an illustration, in order to ensure a relatively good resolution of printed patterns, wavefront aberrations in EUVL machines must be kept below 1 nm root-mean-square error. This implies that the thermoelastic model should be able to predict wavefront aberrations with a very high accuracy.

2. The thermoelastic model has to be in a form that is suitable for real-time prediction, estimation and control of TIWA. In optics literature, several modeling approaches for thermoelastic deformations of optical elements have been proposed [16; 17; 18; 22; 59]. In [22; 59] analytical solutions of the thermoelastic equations have been developed using Dini's series. However, the mathematical complexity of the models developed in [22; 59] prevent us from using them for real-time prediction and control of TIWA. Advanced simulators based on the Finite Element (FE) modeling, have been used in [16; 17; 18] to study the impact of TIWA on the resolution of EUVL systems. Although the results reported in these papers can give us valuable information about dynamical behavior of TIWA, like for example, dominant time constants, steady-state value of the wavefront aberrations and etc., these papers do not address the problem of developing analytic models that can be used for control.

By discretizing the thermoelastic equations using the Finite Difference (FD) or the FE methods, large-scale state-space models can be obtained [60]. However, for accurate prediction of TIWA, the FD or the FE discretization meshes (grids) should consist of a large number of nodes. Consequently, discretized state-space models will have an extremely large number of states. For example, the FE discretization grids used in [16; 17; 18] have more than 100,000 nodes. Because each discretization node carries information about the temperature and the displacement of an optical element, the discretized FE state-space model can have more than 100,000 states.
Obviously, the design of estimators, predictors and controllers for such large-scale systems is a computationally challenging task. Furthermore, real-time implementation of the designed control algorithms is equally challenging. The main computational bottlenecks originate from cubical computational and quadratic memory complexity of estimation and control techniques.

In principle, there are two approaches for overcoming the high computational complexity of estimation and control algorithms.
The first approach is to reduce the computational complexity of control and estimation techniques by exploiting the sparsity of system matrices. Namely, the FE

or FD models of the thermoelastic equations are sparse. Furthermore, in a large number of cases, system matrices of these state-space models can be transformed into a banded form. This structure can be exploited to reduce the computational complexity of control and estimation techniques [60]. However, in the most cases, the system matrices of the designed estimators and controllers are dense. Consequently, the real-time implementation of the designed estimators and controllers might not be computationally feasible.

The second approach is to reduce the model dimensionality by using the model reduction techniques [61]. The sparsity of system matrices can be also exploited to reduce the computational complexity of the model reduction techniques [60]. In [62], computationally efficient model reduction techniques are employed to develop low-order models of reticles used in the EUV lithography. Due to their low-dimensionality, the developed models can be easily used for real-time prediction of thermoelastic deformations. However, the states of the reduced order model do not directly correspond to the physical states of the system[10]. That is, the structure of the large-scale system is not preserved in the reduced order model. A solution to this problem would be to develop structure preserving model reduction techniques, in which the most important information about the spatial structure of the system is preserved in the low-dimensional model. Some attempts to develop structure preserving model reduction techniques for large-scale systems can be found in [63; 64; 65; 66; 67; 68; 69; 70; 71; 72; 73; 74; 75].

By now it should be clear that the problem of compensating TIWA is closely related to the fundamental problems in control and estimation of large-scale systems. In the sequel, we will explain in detail how this engineering problem motivated us to search for solutions of several important problems in systems and control.

1.6 Thermoelastic model and large scale interconnected systems

Large-scale interconnected systems consist of a large number of local dynamical subsystems that are interconnected in a spatial domain. The class of large-scale interconnected systems is broad, and it includes: distributed and complex systems [76; 77; 78; 79; 80; 81; 82; 83; 84; 85; 86; 87], compartmental systems [88; 89], multi-dimensional systems [90; 91; 92; 93], biological processes and systems [94; 95] and discretized PDEs [60; 96; 97].

Because our research is motivated by problems of predicting and controlling the physical process described by the thermoelastic equations, in this thesis we will focus on large-scale systems obtained from discretization of PDEs. As it will be shown in Chapter 2, these systems can have relatively complex interconnection patterns and this fact makes their state-space models suitable for representing a large-variety of physical systems. *Consequently, theoretical results obtained by studying these systems can be used in a number of control applications such as system biology*

[10] The states of the reduced order model are a linear combination of the states of the FE model.

and control of network of dynamical systems [98; 99; 100; 101; 102; 103; 104].
For example, consider the heat equation defined on the 2D elliptical domain shown in Fig.1.8(a). The heat equation is discretized using the FE method. The FE mesh is composed of triangular elements and the time discretization is performed using the backward Euler method [105]. The resulting state space model is in the descriptor form [60; 106; 107] (for more details see Chapter 2):

$$M\underline{\mathbf{x}}(k) = G\underline{\mathbf{x}}(k-1) + \mathbf{c} \tag{1.2}$$

where M and G are state-space matrices, $\underline{\mathbf{x}}(k)$ is the vector of temperatures at the discretization nodes at the discrete time instant k and $\mathbf{c}$ is a constant vector. The matrix M is a sparse banded matrix and its structure is illustrated in Fig. 1.8(b) (the matrix G has a similar structure).

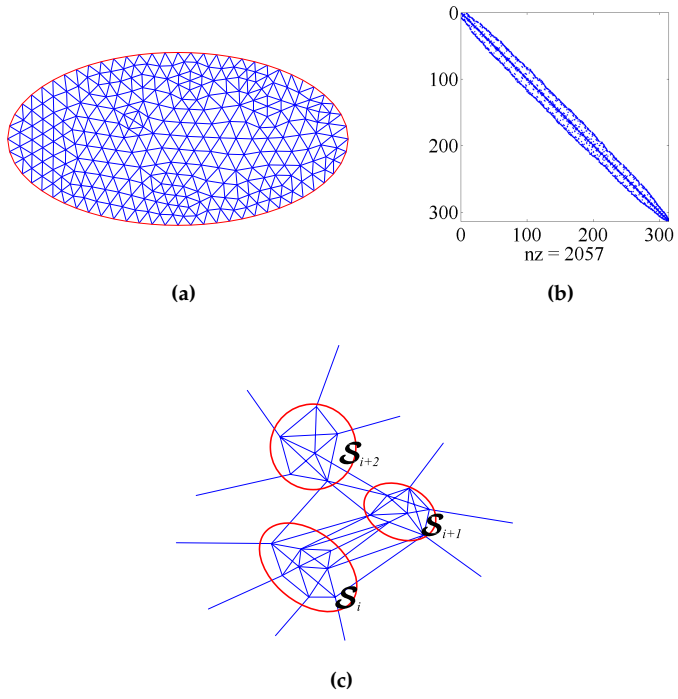


Figure 1.8: (a) The triangular mesh defined on the 2D elliptical domain that is used in the FE method; (b) The structure of the matrix M of the discretized state-space model ("nz" denotes the number of non-zero elements; (c) The FE mesh seen as an interconnection of local subsystems $\mathcal{S}_i$

The FE state-space model (1.2) can be interpreted as a network of local dynamical subsystems. Each node of a FE mesh or a group of neighboring nodes can be seen as a local subsystem, which local state consists of the temperatures at the

discretization nodes. For example, in Fig. 1.8(c) we illustrate local subsystems $\mathcal{S}_i$ defined by grouping the discretization nodes. The dynamics of each local subsystem is influenced by the states (temperatures) of its neighboring local subsystems. The strength of this dynamical interaction depends on the physical properties of the material (mainly it depends on the thermal conductivity constant).

The network structure of state-space models obtained by discretizing PDEs, is maybe best illustrated on an example of the 3D heat equation discretized using the FD method. For example, consider a plate shown in Fig. 1.9(a). The plate is heated by a heat flux acting on its top surface[11].

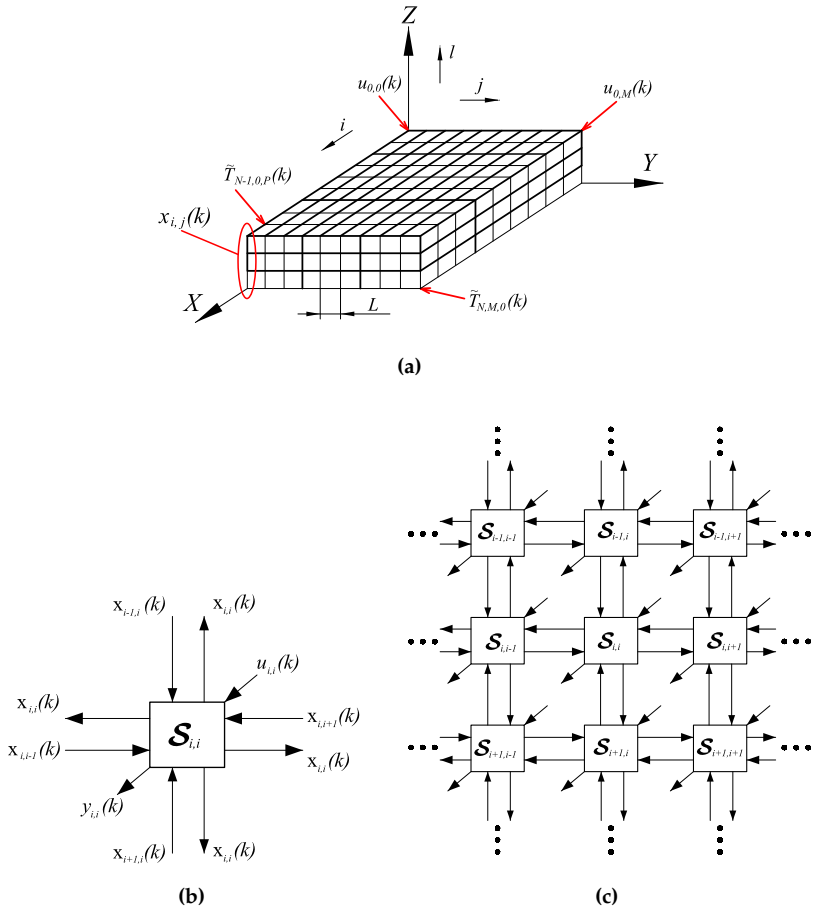


Figure 1.9: (a) The 3D spatial domain and uniform discretization mesh; (b) Inputs, outputs and interconnections of a local subsystem $\mathcal{S}_{i,i}$; (c) Interconnection pattern of local subsystems obtained by discretization of the heat equation on the 3D spatial domain.

[11] The plate can represent a segment of a mirror used in EUVL machines.

Let L denote the spatial discretization step and let the discrete coordinates be denoted by (i, j, l), see Fig. 1.9(a). The temperature at the node $x = iL, y = jL, z = lL$, and at the discrete time instant k, is denoted by $T_{i,j,l}(k)$. This temperature evolves in time according to the discretized 3D heat equation (for details see Chapter 2).

By lifting the temperatures over the z direction, we define *the local state*: $\mathbf{x}_{i,j}(k) = [T_{i,j,0}(k), T_{i,j,1}(k), \ldots, T_{i,j,P}(k)]^T$. With each local state $\mathbf{x}_{i,j}(k)$ we associate a *local subsystem* $\mathcal{S}_{i,j}$. *A local input* $u_{i,j}(k)$ of the local subsystem $\mathcal{S}_{i,j}$ is a heat flux that is acting on the top surface. *A local output* $y_{i,j}(k)$ is a temperature that is measured at the top surface. That is, we assume that it is not possible to measure the full local state $\mathbf{x}_{i,j}(k)$. The local inputs, outputs and interconnections of the local subsystem $\mathcal{S}_{i,i}$ are illustrated in Fig. 1.9(b). The interconnection structure of local subsystems is illustrated in Fig. 1.9(c). *The local state-space model* of the local subsystem $\mathcal{S}_{i,j}$ has the following form:

$$\mathcal{S}_{i,j} \begin{cases} \mathbf{x}_{i,j}(k+1) = A_{i,j}\mathbf{x}_{i,j}(k) + E_{i+1,j}\mathbf{x}_{i+1,j}(k) + E_{i-1,j}\mathbf{x}_{i-1,j}(k) \\ +E_{i,j+1}\mathbf{x}_{i,j+1}(k) + E_{i,j-1}\mathbf{x}_{i,j-1}(k) + B_{i,j}u_{i,j}(k) \\ y_{i,j}(k) = C_{i,j}\mathbf{x}_{i,j}(k) + n_{i,j}(k) \end{cases} \tag{1.3}$$

where $A_{i,j}$, $E_{i+1,j}$, $E_{i-1,j}$, $E_{i,j+1}$, $E_{i,j-1}$, $B_{i,j}$ and $C_{i,j}$ are *the local system matrices* of appropriate dimensions and $n_{i,j}(k)$ is the local measurement noise. By lifting the local subsystems $\mathcal{S}_{i,j}$, first over the x direction and then over the y direction, we obtain *the global state-space model*:

$$\mathcal{S} \begin{cases} \underline{\mathbf{x}}(k+1) & = \underline{A}\underline{\mathbf{x}}(k) + \underline{B}\underline{\mathbf{u}}(k) \\ \underline{\mathbf{y}}(k) & = \underline{C}\underline{\mathbf{x}}(k) + \underline{\mathbf{n}}(k) \end{cases} \tag{1.4}$$

where vector $\underline{\mathbf{x}}(k)$ is *the global state* consisting of the local states $\mathbf{x}_{i,j}(k)$. Similarly, the vectors $\underline{\mathbf{u}}(k)$, $\underline{\mathbf{y}}(k)$, $\underline{\mathbf{n}}(k)$, are *the global input*, *the global output* and *the global measurement noise*, respectively. The matrix $\underline{A}$ is a sparse multi-banded matrix and $\underline{B}$ and $\underline{C}$ are sparse, diagonal matrices. The sparsity patterns of these matrices are illustrated in Fig.1.10.

Two classes of optimal control and estimation problems can be formulated for large-scale interconnected systems. The first class of problems is to design controllers or estimators without any a priori assumptions on the structure of their matrices. In this thesis, these problems are called *the unstructured control and estimation problems*.

In the second class of control and estimation problems, we are interested in designing controllers and estimators that are described by (sparse) structured matrices. In this thesis, these problems are called *the distributed control and estimation problems*. For example, consider the problem of distributed estimation of the local states of the global state-space model (1.4). In its most simplest form, this problem consists of finding a structured gain matrix $\tilde{K}$, such that:

$$\hat{\underline{\mathbf{x}}} = \tilde{K} f(\underline{\mathbf{y}}, \underline{\mathbf{u}}) \tag{1.5}$$

where $\hat{\underline{\mathbf{x}}}$ is an estimate of the global state, $f(\cdot)$ is a known function of the global

output and input vectors and $\tilde{K}$ is a sparse matrix which structure is illustrated in Fig. 1.11.

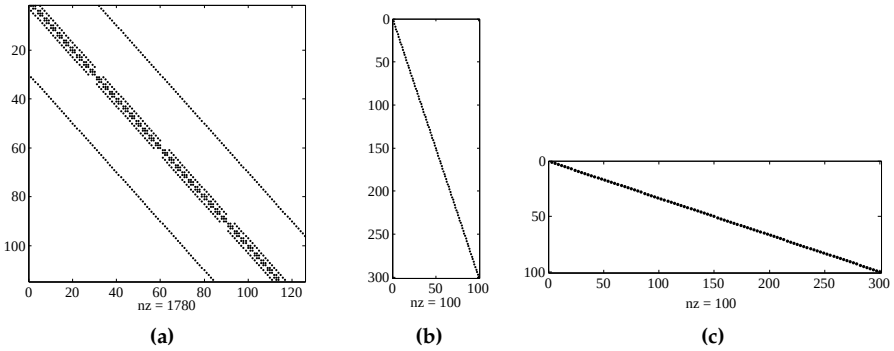


Figure 1.10: Structure of the matrices of the global state-space model (1.4): (a) segment of $\underline{A}$; (b) $\underline{B}$; (c) $\underline{C}$ ("nz" denotes the number of non-zero elements).

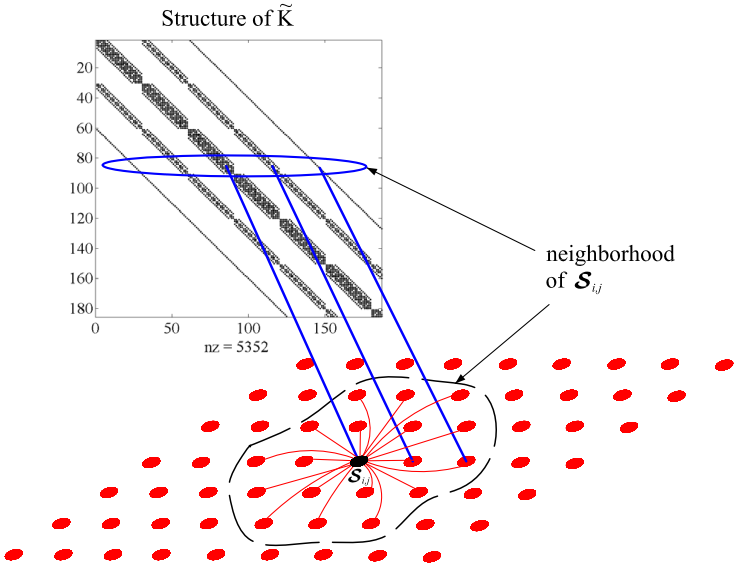


Figure 1.11: Physical interpretation of the sparsity structure of $\tilde{K}$.

The structure of the matrix $\tilde{K}$, illustrated in Fig. 1.11, implies that the state of a local subsystems $\mathcal{S}_{i,j}$ can be estimated from the inputs and outputs of the local subsystems that are in its neighborhood. Because of this, the estimator (1.5) can be implemented on a network of sensors and computing units that communicate locally.

For low dimensional state-space models, the unstructured control and estimation problems have been studied extensively in the past and a large variety of control and estimation methodologies have been proposed [108; 109; 110]. Some notable examples are: the Linear Quadratic Regulator (LQR) and the Kalman filter. In this thesis, these methods are called *the classical model based control and estimation methods*. These design methods heavily rely on numerical linear algebra algorithms [111; 112]. Although these algorithms have many good sides, such as numerical stability, they have one major drawback that makes them unsuitable for control and estimation of large-scale systems. Namely, their computational complexity scales at least cubically with the number of local subsystems N, and their memory complexity scales quadratically with N. Consequently, the classical model based control and estimation methods are not computationally feasible for large-scale interconnected systems.
In the literature two strategies have been used to decrease the computational complexity of the classical estimation and control techniques.

The first strategy, proposed by Rice and Verhaegen in [80; 85; 113], is based on the Sequentially Semi-Separable (SSS) matrix algebra [114; 115; 116]. The main idea of this approach is to construct the SSS matrices from the state-space matrices, and to use the linear computational complexity SSS algebra for designing optimal controllers and estimators. Further applications of the SSS matrix algebra to control and estimation problems can be found in [117; 118; 119; 120; 121; 122].

As it has been mentioned before, the second strategy is to exploit the sparsity of system matrices to decrease the computational complexity of the classical model based control and estimation methods. For example, methods summarized in [60; 96; 123] are following this approach. However, estimation and control matrices computed using these methods are dense. Consequently, the real time implementation of designed estimators and controllers might not be computationally feasible.

Distributed control and estimation problems have also received significant attention in the last few decades. In the case of infinite dimensional systems, a large variety of distributed control and estimation methods have been proposed [77; 78; 79; 84; 124; 125]. However, the real-time implementation of these distributed control and estimation methods always implies some form of discretization or decomposition. Consequently, in order to implement these methods the previously explained computational challenges need to be addressed.
Similarly, a large-variety of distributed control and estimation methods have been proposed for finite-dimensional distributed (interconnected) systems, see [83; 126; 127; 128; 129; 130; 131; 132; 133] and references therein. However, the methods proposed in the above cited papers are either restricted to special classes of large-scale interconnected systems, or the computational and memory complexities of these methods are very high.
For example, in [126; 127] the distributed control and identification methods are proposed for the class of decomposable systems. A decomposable system is basically an interconnected system that is composed of a number of identical local subsystems. In Chapter 2 we use the FD method to discretize the heat equation with constant coefficients. We show that if the boundary conditions of the heat

equation are ignored, then the discretized state-space model can be seen as a network of identical, local subsystems. Consequently, this state-space model belongs to the class of decomposable interconnected systems. However, if the boundary conditions are taken into account or if the coefficients of the heat equation depend on the spatial coordinates, then the derived state-space model cannot be seen anymore as an interconnection of identical local subsystems. This shows that the application of the methods proposed in [126; 127] is restricted to a relatively narrow class of interconnected systems.

The problem of designing sparse feedback gains for distributed control of finite dimensional interconnected systems has been studied in [134; 135]. In [135], the feedback gain matrix is determined by solving constrained H_2 optimal control problem. In the method described in [135], the structure of the gain matrix is fixed a priori and it is incorporated in the H_2 optimal control problem as a constraint. The approach proposed in [134] consists of the two steps. In the first step, the structure of the gain matrix is determined by minimizing a cost function that is composed of the H_2 cost function and a sparsity promoting penalty function. In the second step, the determined structure of the gain matrix is used as a constraint in the H_2 optimization problem. The methods proposed in [134; 135] have two drawbacks. First of all, gain matrices are found by solving non-convex optimization problems. This might imply that the derived gain matrices (determined as the local minima of the H_2 cost functions) do not guarantee a good performance of the closed loop system. More importantly, because these design methods heavily rely on optimization techniques, their computational and memory complexity scale at least with $O(N^3)$ and $O(N^2)$, respectively. This is the main reason why these methods are not suitable for distributed control of large-scale systems.
The approach presented in [136] determines the structured, H_∞ feedback controllers for interconnected systems, by using the ℓ_1 optimization framework. However, because this design methodology heavily relies on optimization techniques, it is not computationally feasible for large-scale interconnected systems.

In [137], a computationally efficient method has been developed for inversion of block banded matrices. Furthermore, it has been demonstrated that this method can be used to decrease the computational complexity of the Kalman filter for large-scale dynamical systems.
The problem of designing the distributed Kalman filter for large-scale systems, has been studied in [138]. The approach presented in [138] is developed on the basis of the Distributed Iterate-Collapse Inversion (DICI) algorithm [139]. The DICI algorithm combines the Jacobi algorithm for matrix inversion [140] and the theoretical framework for inversion of the block banded matrices proposed in [137].

As we will explain the sequel, first principles models of interconnected systems can be very inaccurate. This often implies that models of interconnected systems need to be identified from real data. However, the identification of large-scale interconnected systems is still an open problem. Namely, most of the above summarized estimation and control techniques are implicitly or explicitly exploiting the structure of interconnected systems. *From the identification point of view, this implies that the interconnection structure of a system has to be preserved in the identified state-space model.* The Subspace Identification Methods (SIMs) [58] are not able to

preserve the structure of the system in the identified model. On the other hand, the Prediction Error Methods (PEMs) [141] are able to incorporate some structural information about the system in the identified model, at the expense of using optimization techniques. Furthermore, the PEMs and SIMs are not suitable for identification of large-scale system because the computational and memory complexity of these methods scale with $O(N^3)$ and $O(N^2)$, respectively.

1.7 Scope and main contributions of the thesis

So far, we have explained the main problems in controlling TIWA in optical lithography machines and we have placed this problem in a more general, theoretical context of estimation and control of large-scale systems. Furthermore, we have pointed out some open problems in estimation and control of large-scale interconnected systems. We have now prepared the ground to explain the scope and the main contributions of this thesis.

Before developing predictive controllers for TIWA, it is first necessary to develop the thermoelastic model. In this thesis, we combine two approaches for developing thermoelastic models that are suitable for real time estimation and control of TIWA. The first approach is based on first principles models. Starting from the thermoelastic equations, we are using the FE method to obtain a sparse, descriptor state-space model that describes thermally induced deformations of optical elements. As an alternative to the FE method, we are also using the FD method to approximate the 3D heat equation describing temperature change in optical elements.
However, the first principles models have some limitations. First of all, because it is impossible to precisely know numerical values of model parameters and because it might not be possible to accurately model boundary conditions of the thermoelastic equations, the first principles models can be inaccurate. Secondly, some processes that affect the system's dynamics (disturbances) cannot be modeled a priori using first principles. To overcome the above explained drawbacks of the first principles models, we have developed a system identification framework for identifying large-scale, sparse, state-space models. *The first principles approach is still very useful because it can help us to understand the structure of the model that we want to identify*. For example, from discretization of the 3D heat equation, we know that the system matrices of the state-space model that we want to identify, are sparse, banded matrices. By exploiting this structure we are able to develop computationally efficient identification algorithms.

This thesis can be divided into two parts. In the first part we present theoretical framework on which basis we develop methods for identification and estimation of large-scale, interconnected systems. For brevity of this thesis, we did not develop predictive control algorithms for large-scale systems. The predictive control algorithms can be easily developed using the theoretical framework presented in Chapter 3. Because during the course of this thesis we were unable to test these methods on a real EUV system, we have validated these methods using numerical simulations.

However, in a cooperation with researchers from Optics Research Group, Delft University of Technology, we have built an experimental AO setup[12]. This experimental setup contains two DMs: one mirror is used to introduce wavefront aberrations and another mirror is used for correction. This way, we were able to simulate the AO system in a real EUV machine. Consequently, we were able to demonstrate the proof of concept for predictive control of wavefront aberrations. The predictive control method and its experimental validation is presented in the second part of the thesis. Because it is only used to demonstrate the proof of concept, the developed predictive controller is not based on the theoretical framework for large scale systems, that is presented in the first part of the thesis. Generalization of the predictive control framework for large-scale interconnected systems is briefly explained in Chapter 10 (see Section 10.5).
Beside being used to demonstrate the proof of concept for predictive control of wavefront aberrations, the experimental setup was used to develop and to test new identification and control methods for AO systems. These methods together with the predictive controller are not necessarily restricted to the AO systems for optical lithography machines and they can be used in a large variety of AO applications.

One of the main focuses of this thesis is to develop linear computational complexity algorithms for estimation and identification of the discretized heat equation and the discretized thermoelastic equations. For simplicity, in the thermoelastic equations we assumed that the Coefficient of Thermal Expansion (CTE) is constant. Consequently, the state-space representation of the discretized thermoelastic equations is linear and accordingly, the algorithms developed in this thesis mostly rely on the linear system theory. However, some of the materials used to make EUV mirrors might have CTE that depends on the temperature. In this case, the resulting state-space model has a linear state equation and a nonlinear output equation (for details see Chapter 2, Remark 2.2).
At the first glance, one might come to a conclusion that the computationally efficient algorithms presented in this thesis, are only applicable to linear state-space models and that they cannot be used in the case when the CTE depends on the temperature. To show that this conclusion is incorrect and that the presented algorithms can be generalized for the above mentioned case, in Chapter 4, Section 4.3, we demonstrate that the Newton observer for nonlinear systems can be implemented with linear computational complexity by using the algorithms developed in this thesis. Using the same principle, computationally efficient identification and control algorithms for nonlinear systems with an output nonlinearity, can be developed.

Readers who are only interested in the AO applications, can skip the first part of this thesis (Chapters 2-6) and they can focus on the second, experimental part (Chapters 7-10). Similarly, researchers who are solely interested in numerical aspects of estimating and controlling large-scale systems can skip the second part. However, we think that for engineers and scientists who are interested in developing real-time estimation and control algorithms for thermally induced wavefront aberrations, both parts are relevant. In the first part they can find the building blocks on which basis they can form computationally efficient algorithms, and

[12] A close collaboration has been established with the PhD student Alessandro Polo and his supervisors: professor Silvania Pereira and professor H. Paul Urbach.

in the second part they can get the main ideas of how these algorithms can be adapted to the problem of controlling wavefront aberrations. These ideas are briefly explained in Chapter 10 (see Section 10.5). In the sequel we provide a detailed explanation of the main contributions of the thesis.

1.7.1 Main theoretical contributions of the thesis

In this thesis we consider large-scale interconnected systems described by sparse banded or sparse multi-banded system matrices. As it has been already explained, these types of state-space models originate from discretization of PDEs using the FE or FD methods. *Throughout this thesis, when we refer to large-scale systems we mean large-scale interconnected systems with sparse banded or multi-banded state-space matrices.* Although we consider systems described by sparse (multi) banded state-space matrices, the methods proposed in this thesis can be generalized to interconnected systems with a more general interconnection patterns (see Chapter 3.4 and the conclusions in Chapter 11).

The main contributions to the theory of large-scale interconnected systems are:

1. **We propose a structure preserving lifting technique for state-space models of large-scale interconnected systems**. The newly proposed lifting technique, first lifts the local state-space models (1.3) over the time domain and then it lifts such lifted state-space models over the spatial domain. Consequently, the structure preserving lifting technique ensures that the distributed (sparse banded or multi-banded) structure of the state-space matrices of an interconnected system is preserved in the lifted state-space model. In this thesis we use the structure preserving lifting technique to prove some new, interesting properties of large-scale interconnected systems and to develop distributed estimation and identification algorithms.
   The importance of the new lifting technique mainly lies in the fact that the classical lifting technique that is widely used in the SIMs [142; 143; 144; 145; 146], moving horizon estimation [147; 148; 149; 150], Iterative Learning Control (ILC) [151; 152; 153; 154; 155; 156; 157], Model Predictive Control (MPC)[158; 159; 160], "destroys" the distributed structure of the large-scale interconnected systems. Consequently, the classical lifting technique cannot be the basis for the development of distributed control and estimation algorithms.

2. **The inverses of (finite-time) Gramians and lifted system matrices of large-scale interconnected systems are off-diagonally decaying matrices**.[13] This result is important because the off-diagonal decay of operators associated with interconnected (distributed) systems has only been studied in the infinite-dimensional case [84]. On the basis of the structure preserving lifting technique and using the theoretical framework presented in [161;

[13] Here, the off-diagonal decay does not necessarily mean the decay with respect to the main diagonal. We numerically show in Chapter 3.5 that inverses of multi banded matrices exhibit off-diagonal decay with respect to diagonals below or above the main diagonal.

162], we prove that inverses of lifted system matrices and (finite-time) Gramians of large-scale interconnected systems, belong to a class of off-diagonally decaying matrices.

3. **The inverses of Gramians and lifted system matrices of large scale interconnected systems can be approximated by sparse (multi) banded matrices with $O(N)$ computational and $O(N)$ memory complexity.** This directly follows from the fact that the inverses of lifted system matrices are off-diagonally decaying matrices. We have used two algorithms to perform this structure preserving inversion: the Chebyshev method [161] and the Newton iteration [163]. Using the spectral mapping theorem we have derived an upper bound on the error introduced by the Chebyshev approximation method. The Chebyshev method and the Newton iteration are used to develop computationally efficient identification and estimation methods for large-scale systems.

4. **The state estimate of a local subsystem can be computed as a linear combination of the input-output data of local subsystems that are in its neighborhood.** We prove that the size of this neighborhood depends on the condition number of the finite-time observability Gramian. In particular, if the condition number is larger then the size of this neighborhood is larger and vice-versa. *Thus, for systems with well-conditioned observability matrices and observability Gramians, to compute the local state estimate, a local system needs to communicate only with its neighboring local subsystems.* That is, there is no need for "all to all" communication between the local subsystems of a large-scale system.

On the basis of these new theoretical results, we have developed:

1. *Distributed and centralized Moving Horizon Estimation (MHE) methods for large scale interconnected systems in the standard, state-space form*. These methods are developed using the structure preserving lifting technique and by using the Chebyshev approximation method. The system matrices of the developed estimators are sparse (multi) banded matrices. Consequently, the computational and memory complexity of the developed MHE methods scale with $O(N)$. Furthermore, the distributed MHE method estimates the local state as a linear combination of the local input-output data. Thanks to this property, the developed MHE method is much faster than the distributed estimation methods that are based on the consensus-subgradient or diffusion based algorithms [164; 165; 166; 167; 168]. We have studied how the errors introduced by the Chebyshev approximation method influence the dynamics of the estimation error.
   Furthermore, using the Chebyshev method and the Newton iteration method, we have developed computationally efficient MHE methods for large-scale systems in the descriptor state-space form. These methods are used to estimate the state of the discretized thermoelastic equations.

   The direct consequence of the structure preserving lifting technique and the approximation algorithms is that the MHE estimators inherit the structure of large-scale,

interconnected systems. That is, the distributed MHE methods are derived without any a priori assumptions on the structure of the estimator gain matrices and without the need for formulating the distributed MHE problem as a non-convex optimization problem.

Beside the MHE methods, a variety of control strategies are relying on the lifted system representation. For example, the ILC and the MPC methods, explicitly or implicitly, lift state-space models over the time domain and they derive controllers by inverting the lifted system matrices[14]. Similarly to the development of the distributed MHE method, the structure preserving lifting technique and the approximation algorithms can be used to develop distributed ILC and MPC methods, that have sparse (multi) banded system matrices.
All this implies that the framework proposed in this thesis can be used as the basis for establishing distributed estimation and control methods that:

(a) Do not rely on non-convex optimization problems, and the structure of an estimator (or a controller) directly follows from the structure of interconnected systems.

(b) They are computationally feasible for systems with an extremely large number of local subsystems.

2. *A subspace identification algorithm for large scale interconnected systems.* By exploiting the fact that the state of a local subsystem can be estimated only from the local input-output data, the proposed subspace algorithm identifies state-space models of local subsystems in the decentralized manner. The computational complexity of estimating a local state-space model depends only on the dimension of the local state. Consequently, the developed subspace identification algorithm is computationally feasible for large-scale interconnected systems with a very large number of local subsystems. Even more importantly, the proposed identification algorithm is able to identify the global state-space model in which the interconnection structure of a system is preserved. Due to its decentralized nature, the subspace identification algorithm can be implemented on a network of local computing units and sensors that communicate locally.

3. *A parameter optimization algorithm for identification of large-scale interconnected systems.* The developed identification algorithm consists of two steps. In the first step, impulse response parameters of local subsystems are estimated in the decentralized manner. In the second identification step, the estimated impulse response parameters and input-output data are used to form a large-scale, separable least-squares problem. The local system matrices are identified by solving this optimization problem with $O(N)$ computational and $O(N)$ memory complexity. The linear computational complexity is achieved by approximating the inverses of lifted system matrices using the Chebyshev approximation method and Newton iteration. The local system matrices that are estimated using the proposed subspace identification

[14]For example, the lifted ILC algorithm inverts the impulse response matrix.

algorithm, can be used as initial guesses of decision variables of the separable least-squares optimization problem.

1.7.2 Contributions to the adaptive optics field

1. We developed an Iterative Learning Control (ILC) algorithm for controlling the shape of deformable mirrors. The developed algorithm is tested on an experimental AO setup consisting of a commercial Membrane DM (MDM), Shack-Hartmann WFS (S-H WFS) and a real-time controller. The developed algorithm can be used for accurate correction of both static and slowly-varying wavefront aberrations in optical systems. We studied the stability of the ILC method. Furthermore, we gave a physical interpretation of the parameters of the ILC algorithm and we gave guidelines for its tuning. The main advantage of the developed controller over the controllers for MDMs proposed in the literature, is that it can accurately produce the desired shape of a MDM in a small number of control iterations.

2. Using the subspace identification technique, we identified a low-dimensional, dynamical model of a prototype of Thermally Actuated Deformable Mirror (TADM). The prototype of TADM was manufactured to meet the requirements of AO systems for EUV lithography and it was designed by Eindhoven University of Technology[15]. Due to its low-dimensionality, the identified model can be easily used to design feed-forward or feedback model based controllers for a large variety of AO systems. The experimental results showed that the identified model can accurately predict the dynamical behavior of the TADM.
   The main contribution of the proposed identification methodology can be best explained by summarizing the commonly used approach for identifying models of TADMs. Because the dynamical behavior of TADMs is described by the thermoelastic equations[16], it is challenging to develop dynamical models of these devices. Consequently, in the AO literature, see for example [20; 169], steady-state models (influence functions) of these devices are usually identified. Controllers developed on the basis of these models cannot achieve fast and accurate correction of wavefront aberrations. In contrast to these approaches, the proposed method for identification of dynamical models of TADMs, can be used as the basis for the development of high performance AO controllers.

3. We experimentally demonstrated the proof of concept for predictive compensation of thermally induced wavefront aberrations in (EUV) optical lithography machines. The experiments were performed on an experimental setup consisting of the MDM and TADM. The TADM was used to generate wavefront aberrations, while the MDM was used as a correction element. In this

[15] The mirror was designed by a PhD student S.K. Ravensbergen and his supervisors: professor P.C.J.N. Rosielle and professor M. Steinbuch.

[16] In some cases, see for example [26], the thermoelastic equations need to be coupled with the biharmonic plate equation.

way, we simulated the AO system in a real EUVL system. Namely, the aberrations created by the TADM represent the TIWA, whereas the MDM represents an active optical element in a lithographic machine[17]. The predictive control algorithm was developed on the basis of the model of the TADM (that was identified using the subspace identification technique). We demonstrated that it is possible to predictively compensate wavefront aberrations by only using initial wavefront measurements. Furthermore, we showed that it is possible to compensate wavefront aberrations even when inputs of the prediction model are not known a priori (this case corresponds to the scenario in a lithographic machine when the inputs of the TIWA model are unknown or they are not modeled).
Apart from AO systems for optical lithographic machines, the proposed predictive controller can be used in other AO applications, such as compensation of wavefront aberrations induced by high-power lasers [20; 21; 22; 23; 24].

4. Using the developed MHE method for descriptor state-space models and the discretized thermoelastic equations, we demonstrated that it is possible to estimate the temperature distribution of an optical element from the measurements of surface deformations. In principle, these results imply that the state of the TIWA model can be estimated from wavefront measurements. Furthermore, this estimation framework can be easily generalized for simultaneous state and input estimation of the TIWA model.

1.8 Organization of the thesis and journal papers

This thesis can be divided into the following two parts.

- Part I: **Theoretical contributions**.
  - First, in Chapter 2 we start with the derivation of structured state-space models that describe dynamics of thermoelastic deformations of mirrors used in optical systems. To explain the basic lifting and partitioning procedures for obtaining large-scale interconnected systems from discretized PDEs, we first discretize the 2D heat equation using the FD method. After that, we use the FD method to discretize the 3D heat equation. The heat equation is defined on a rectangular prism (a plate), that can be seen as a segment of a mirror used in optical systems. Finally, we discretize the thermoelastic equations using the FE method. The discretized thermoelastic equations describe thermally induced deformations of circular mirrors used in optical systems. The models developed in this Chapter are used throughout this thesis to demonstrate

[17] Compared to AO systems for EUVL machines, in our experimental setup we reversed the role of the TADM. In AO systems for EUVL machines, TADMs are used to correct TIWA. However, in our experimental setup we used the TADM to induce wavefront aberrations. Namely, because of its slow dynamics, TADM is ideal for reproducing dynamical behavior of TIWA.

the main properties of large-scale systems and to demonstrate the performance of the developed algorithms.

– In Chapter 3, we present the structure preserving lifting technique for large-scale systems. Then, we prove that inverses of Gramians and lifted system matrices belong to a class of off-diagonally decaying matrices. Finally, we summarize two algorithms for structure preserving inversion of sparse lifted system matrices: the Chebyshev approximation method and Newton iteration. We derive an upper bound on the approximation errors introduced by the Chebyshev method. The material presented in this chapter is partly presented in:

  Moving horizon estimation for large-scale interconnected systems, A. Haber and M. Verhaegen, IEEE Transactions on Automatic Control, Vol. 58, Issue 11, 2013.

– In Chapter 4, we present centralized and distributed MHE methods for large-scale interconnected systems. Two types of MHE methods are developed. The first MHE method is developed for state-space models in the standard form. The second MHE method is developed for descriptor state-space models. The MHE method for state-space models in the standard form is presented in:

  Moving horizon estimation for large-scale interconnected systems, A. Haber and M. Verhaegen, IEEE Transactions on Automatic Control, Vol. 58, Issue 11, 2013.

  while the MHE method for descriptor state-space models is presented in:

  Moving horizon state estimation of thermoelastic deformations, A. Haber, I. Maj, R. Mustata and M. Verhaegen, preprint submitted to Journal of Computational Physics.

– In Chapter 5, we present the subspace identification algorithm for large scale systems. The method proposed in this chapter is presented in:

  Subspace identification of large-scale interconnected systems, A. Haber and M. Verhaegen, IEEE Transactions on Automatic Control, Note: Conditionally accepted, 2013 (arXiv:1309.5105v1).

– In Chapter 6, we present the parameter optimization method for identification of large-scale interconnected systems. The method presented in this chapter is based on the method proposed in:

  Identification of large-scale systems described by sparse banded matrices, A. Haber and M. Verhaegen, submitted to Automatica.

– In Chapter 7 we present numerical results of estimating the temperature distribution of a mirror used in optical systems. These results are

presented in:

Moving horizon state estimation of thermoelastic deformations, A. Haber, I. Maj, R. Mustata and M. Verhaegen, preprint submitted to Journal of Computational Physics.

- Part II: **Control and identification methods for adaptive optics**.
  - In Chapter 8, we present the ILC algorithm for optimal wavefront correction. This algorithm is presented in:

    Iterative learning control of a membrane deformable mirror for optimal wavefront correction, A. Haber, A. Polo, C.S. Smith, S.F. Pereira, H.P. Urbach and M. Verhaegen, Applied Optics, Vol. 52, Issue 11, 2013.

    The computationally efficient implementation of the ILC method is presented in:

    Linear computational complexity robust ILC for lifted systems, A. Haber, P.R. Fraanje and M. Verhaegen, Automatica, Vol. 48, Issue 6, 2012.

    Furthermore, the ILC algorithm is used to develop AO controllers in the following papers:

    Linear phase retrieval for real-time adaptive optics, A. Polo, A. Haber, S.F. Pereira, M. Verhaegen and H.P. Urbach, Journal of the European Optical Society-Rapid publications, Vol. 8, 2013.

    An innovative and efficient method to control the shape of push-pull membrane deformable mirror, A. Polo, A. Haber, S.F. Pereira, M. Verhaegen and H.P. Urbach, Optics Express, Vol. 20, Issue 25, 2012.

  - In Chapter 9, we present the identification method for the dynamical model of TADM. This chapter is based on:

    Identification of dynamical model of a thermally actuated deformable mirror, A. Haber, A. Polo, S.K. Ravensbergen, H.P. Urbach and M. Verhaegen, Optics Letters, Vol. 38, Issue 16, 2013.

  - In Chapter 10, we present the proof of concept for predictive correction of thermally induced wavefront aberrations. This chapter is based on:

    Predictive control of thermally induced wavefront aberrations, A. Haber, A. Polo, I. Maj, S.F. Pereira, H.P. Urbach and M. Verhaegen, Optics Express, Vol. 21, Issue 18, 2013.
  - In Chapter 11, we draw conclusions and we discuss the future research directions.

2 CHAPTER

Modeling

In this chapter we derive large-scale, state-space models that describe thermoelastic deformations of optical elements. We approximate the heat equation and the thermoelastic equations using the finite difference and finite element methods. The developed models are used in the remaining chapters to test and compare the performance of the developed estimation and identification algorithms.

2.1 Introduction

A large variety of physical phenomena and processes are described by Partial Differential Equations (PDEs). Some notable examples of physical processes that are modeled by PDEs are: compressible and incompressible flows [170; 171; 172; 173; 174; 175; 176], aero-optical aberrations [177; 178], heating of optical elements, elastic deformations of bodies under action of forces [179; 180] and electromagnetic fields [181].

From the form of a PDE and from numerical values of its coefficients, we can draw important conclusions about the stability and dynamical behavior of the underlying system. However, in order to confirm these conclusions or to fully understand the dynamical behavior of the system, a PDE needs to be discretized and its discretized form needs to be simulated.

In the systems and control literature, there are two approaches for designing estimators and controllers for systems modeled by PDEs. In the first approach, finite dimensional, approximate models of PDEs are obtained using discretization techniques, such as Finite Difference (FD) or Finite Element (FE) methods [97; 105; 182; 183]. After discretization of PDEs, the classical model based control and estimation techniques are applied [108; 110]. In the literature, this approach is referred to as *the early lumping approach* [184; 185; 186]. The main disadvantage of the early lumping approach is that the discretized models are usually large dimensional and consequently, the high computational complexity of the classical

estimation and control techniques becomes a serious problem. Furthermore, the early lumping approach introduces approximation errors in the early stage of the control design.

In the second approach, controllers or estimators are designed using models that are described by PDEs. During the control design process, PDEs are not discretized. Instead, the designed controllers are discretized in the final, implementation step. In the literature, this approach is referred to as *the late lumping approach* [184; 185; 186]. A large variety of control and estimation techniques are using this approach [124; 187; 188; 189]. Unfortunately, there is no universal late lumping approach. That is, there is no universal control and estimation strategy that can handle any type of PDE. Particular methods are tailored for particular types of PDEs and boundary conditions. Furthermore, the discretization and real time implementation of the designed controllers, are usually computationally expensive tasks.

In this thesis we will follow the early lumping approach. The FD and FE methods are used to develop structured state-space models of the heat equation and thermoelastic equations. In this chapter we explain in detail how to order discretized PDEs and boundary conditions such that the system matrices of the derived state-space models are sparse (multi) banded matrices. The models derived in this chapter will be used throughout this thesis to test and to compare the developed identification and estimation algorithms.

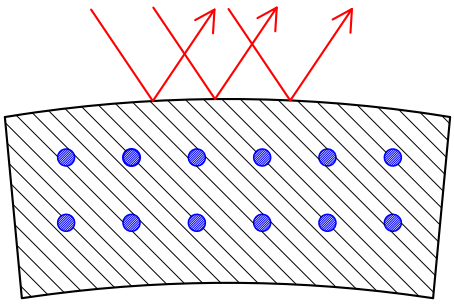
We start with the FD approximation of the heat equation defined on the 2D domain (2D heat equation). Then, we extend the discretization method to the 3D heat equation. Finally, we develop a FE state-space model describing thermoelastic deformations of mirrors used optical systems. This state-space model will be used in Chapter 7 to demonstrate the performance of the developed estimation methods.

2.2 Finite difference state-space model of the 2D heat equation

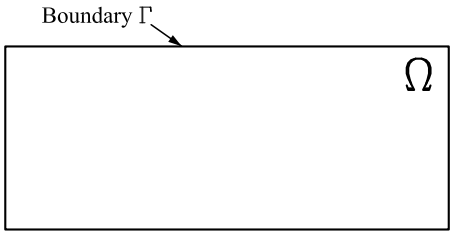
To explain the general idea of constructing structured state-space models of discretized PDEs, in this section we start with the Finite Difference (FD) approximation of the 2D heat equation. To make the story more interesting, discretization procedure will be placed in the context of developing a simplified model of a Thermally Actuated Deformable Mirror (TADM).

TADMs are an inexpensive option for accurate correction of slowly-varying wavefront aberrations in optical systems [20; 190; 191]. In Fig. 2.1 a simplified cross section of a TADM is illustrated. An array of heaters, denoted by an array of circles in Fig.2.1, heats the body of the mirror[1]. As the mirror body heats up, it starts to deform. The deformations of the top mirror surface influence the distortions of the reflected wavefront.

[1] In practice, an array of resistors can be used for generating the heat [21].



(a)



(b)

Figure 2.1: (a) Cross section of the thermally actuated deformable mirror with an array of heaters; (b) Interior region Ω and boundary domain Γ.

For the time being we will neglect the deformations and the heat diffusion in the direction perpendicular to the mirror's cross section. The temperature change of the cross-section of the mirror is governed by the 2D heat equation:

$$\frac{\partial T}{\partial t} = \alpha \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} \right) + ru, \quad (x, y) \in \Omega \tag{2.1}$$

where t is time, x, y are spatial coordinates, $T = T(x, y, t)$ is the temperature, u is heat generated by an actuator, r is a scaling factor of an input, α is the thermal diffusivity. For simplicity, the following boundary and initial conditions are introduced:

$$T(x, y, t) = 0, \quad (x, y) \in \Gamma, \quad t \geq 0 \tag{2.2}$$

$$T(x, y, 0) = 0, \quad (x, y) \in \Omega \tag{2.3}$$

For the discretization of the heat equation (2.1), we are using the FD method [97]. The FD method approximates the partial derivatives as follows:

$$\frac{\partial^2 T}{\partial x^2} \approx \frac{T_{i+1,j}(k) - 2T_{i,j}(k) + T_{i-1,j}(k)}{L^2}, \quad \frac{\partial^2 T}{\partial y^2} \approx \frac{T_{i,j+1}(k) - 2T_{i,j}(k) + T_{i,j-1}(k)}{L^2}$$

$$\frac{\partial T}{\partial t} \approx \frac{T_{i,j}(k+1) - T_{i,j}(k)}{h} \tag{2.4}$$

where L is a spatial discretization step, h is a time discretization step, $T_{i,j}(k)$ is the temperature at a node (Li, Lj), $i = 0, 1, \ldots, N+1$, $j = 0, 1, 2, 3$, and at the time instant kh, $k \geq 0$. The uniform discretization mesh is illustrated in Fig. 2.2.

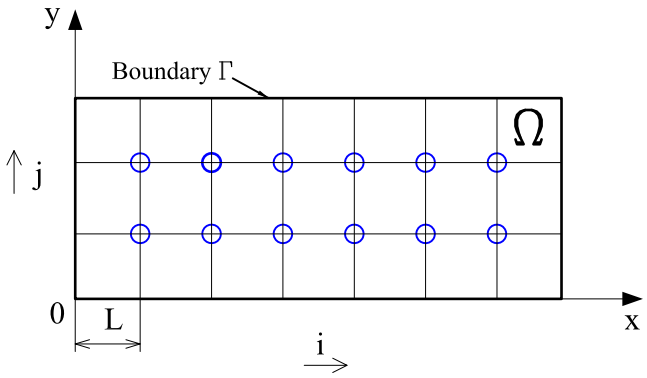


Figure 2.2: Mesh for finite difference discretization of the heat equation

For simplicity in (2.1), it is assumed that the input u does not change in the y direction. That is, we assume that the input u_i is acting at the nodes $(i, 1)$ and $(i, 2)$, $i = 1, \ldots, N$. Taking into account (2.4), the discretized heat equation takes the following form:

$$\begin{aligned} T_{i,j}(k+1) =& a_1 T_{i,j}(k) + a_2 T_{i+1,j}(k) + a_2 T_{i-1,j}(k) \\ &+ a_2 T_{i,j+1}(k) + a_2 T_{i,j-1}(k) + b u_i(k) \end{aligned} \tag{2.5}$$

where $a_1 = 1 - 4\frac{h\alpha}{L^2}$, $a_2 = \frac{h\alpha}{L^2}$ and $b = rh$. The discretized boundary conditions are:

$$\begin{aligned} T_{i,0}(k) = T_{i,3}(k) = 0, i = 0, \ldots, N+1, \ \forall k \geq 0 \\ T_{0,j}(k) = T_{N+1,j}(k) = 0, j = 0, 1, 2, 3, \ \forall k \geq 0 \end{aligned}$$

Taking these boundary conditions into account, from (2.5) we have:

- For $i = 1$, $j = 1, 2$:

$$\begin{bmatrix} T_{1,1}(k+1) \\ T_{1,2}(k+1) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ a_2 & a_1 \end{bmatrix} \begin{bmatrix} T_{1,1}(k) \\ T_{1,2}(k) \end{bmatrix} + \begin{bmatrix} a_2 & 0 \\ 0 & a_2 \end{bmatrix} \begin{bmatrix} T_{2,1}(k) \\ T_{2,2}(k) \end{bmatrix} + \begin{bmatrix} b \\ b \end{bmatrix} u_1(k) \tag{2.6}$$

- For $1 < i < N$, $j = 1, 2$:

$$\begin{bmatrix} T_{i,1}(k+1) \\ T_{i,2}(k+1) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ a_2 & a_1 \end{bmatrix} \begin{bmatrix} T_{i,1}(k) \\ T_{i,2}(k) \end{bmatrix} + \begin{bmatrix} a_2 & 0 \\ 0 & a_2 \end{bmatrix} \begin{bmatrix} T_{i-1,1}(k) \\ T_{i-1,2}(k) \end{bmatrix}$$

$$+\begin{bmatrix} a_2 & 0 \\ 0 & a_2 \end{bmatrix}\begin{bmatrix} T_{i+1,1}(k) \\ T_{i+1,2}(k) \end{bmatrix} + \begin{bmatrix} b \\ b \end{bmatrix} u_i(k) \tag{2.7}$$

- For $i = N$, $j = 1, 2$:

$$\begin{bmatrix} T_{N,1}(k+1) \\ T_{N,2}(k+1) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ a_2 & a_1 \end{bmatrix}\begin{bmatrix} T_{N,1}(k) \\ T_{N,2}(k) \end{bmatrix} + \begin{bmatrix} a_2 & 0 \\ 0 & a_2 \end{bmatrix}\begin{bmatrix} T_{N-1,1}(k) \\ T_{N-1,2}(k) \end{bmatrix} + \begin{bmatrix} b \\ b \end{bmatrix} u_N(k) \tag{2.8}$$

The local state $\mathbf{x}_i(k)$ and *the local system matrices* A, E and B are defined as follows:

$$\mathbf{x}_i(k) = \begin{bmatrix} T_{i,1}(k) \\ T_{i,2}(k) \end{bmatrix},\ A = \begin{bmatrix} a_1 & a_2 \\ a_2 & a_1 \end{bmatrix},\ E = \begin{bmatrix} a_2 & 0 \\ 0 & a_2 \end{bmatrix},\ B = \begin{bmatrix} b \\ b \end{bmatrix} \tag{2.9}$$

We assume that we can measure the temperatures at the grid points $(i, 2)$, $i = 1, \ldots, N$. In practice this can be achieved using the sensing principle explained in [192]. The local output equation is defined accordingly:

$$\mathbf{y}_i(k) = C\mathbf{x}_i(k) + \mathbf{n}_i(k) \tag{2.10}$$

where $C = \begin{bmatrix} 1 & 0 \end{bmatrix}$ and $\mathbf{n}_i(k)$ is the local measurement noise. By lifting the local states $\mathbf{x}_i$ over the i direction, we obtain *the global state-space model*:

$$\mathcal{S}\begin{cases} \underline{\mathbf{x}}(k+1) & = \underline{A}\underline{\mathbf{x}}(k) + \underline{B}\underline{\mathbf{u}}(k) \\ \underline{\mathbf{y}}(k) & = \underline{C}\underline{\mathbf{x}}(k) + \underline{\mathbf{n}}(k) \end{cases} \tag{2.11}$$

where *the global system matrices* $\underline{A}$, $\underline{B}$ and $\underline{C}$, and *the global vectors* $\underline{\mathbf{x}}(k)$, $\underline{\mathbf{u}}(k)$ and $\underline{\mathbf{n}}(k)$, are defined as follows:

$$\underline{A} = \begin{bmatrix} A & E & & \\ E & A & E & \\ & \ddots & \ddots & \ddots \\ & & E & A & E \\ & & & E & A \end{bmatrix}, \underline{B} = \begin{bmatrix} B & & \\ & \ddots & \\ & & B \end{bmatrix}, \underline{C} = \begin{bmatrix} C & & \\ & \ddots & \\ & & C \end{bmatrix},$$

$$\underline{\mathbf{y}}(k) = \begin{bmatrix} \mathbf{y}_1(k) \\ \vdots \\ \mathbf{y}_N(k) \end{bmatrix}, \underline{\mathbf{x}}(k) = \begin{bmatrix} \mathbf{x}_1(k) \\ \vdots \\ \mathbf{x}_N(k) \end{bmatrix}, \underline{\mathbf{u}}(k) = \begin{bmatrix} u_1(k) \\ \vdots \\ u_N(k) \end{bmatrix} \tag{2.12}$$

The global system $\mathcal{S}$, defined in (2.11), consists of the interconnection of N identical *local subsystems* $\mathcal{S}_i$. The interconnection pattern of local subsystems is illustrated in Fig. 2.3.

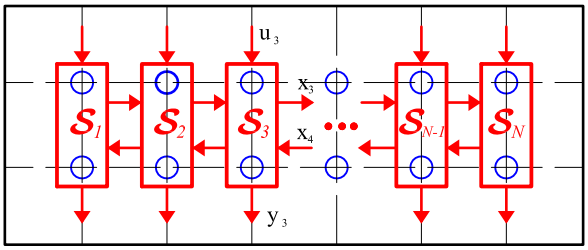


Figure 2.3: The discretized 2D heat equation can be seen as an interconnection of local subsystems S_i, $i = 1, \ldots, N$.

Because the thermal diffusivity is not depending on the spatial coordinates, the global system (2.11) belongs to the class of decomposable systems [127]. If in the 2D heat equation (2.1), the thermal diffusivity constant α depends on the spatial coordinates (x, y), then its discretized state-space model can be seen as an interconnection of nonidentical local subsystems. In this case the global system does not belong anymore to the class of decomposable systems.

The procedure for obtaining the global state-space model (2.67) *can be summarized as follows. First, for each* $i = 1, \ldots, N$ *the discretized heat equation* (2.5) *is lifted over the* j *direction. Then, starting from* $i = 1$*, such lifted equations are lifted over the* i *direction*. In the sequel we generalize this procedure for the 3D heat equation.

2.3 Finite difference state-space model of the 3D heat equation

Consider a rectangular prism (plate) shown in Fig. 2.4.

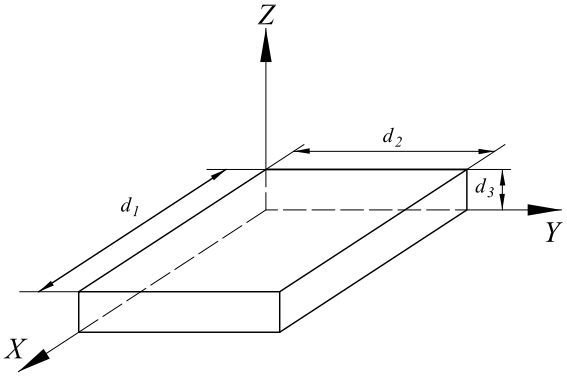


Figure 2.4: A plate with a coordinate system attached to it.

This plate can be seen as the part of a mirror used in the EUV lithography. On the top surface of the plate, the intensity of the diffracted light acts as a heat source that increases the temperature of the plate's body. The plate can be mathematically described by the following set of points:

$$\Omega = \{(x, y, z) \mid 0 \leq x \leq d_1,\ 0 \leq y \leq d_2,\ 0 \leq z \leq d_3\} \tag{2.13}$$

The boundary surfaces of the plate are mathematically described as follows:

$$\begin{aligned}
\Gamma_1 &= \{(x, y, z) \mid 0 \leq x \leq d_1,\ 0 \leq y \leq d_2,\ z = d_3\} \\
\Gamma_2 &= \{(x, y, z) \mid 0 \leq x \leq d_1,\ 0 \leq y \leq d_2,\ z = 0\} \\
\Gamma_3 &= \{(x, y, z) \mid x = d_1,\ 0 \leq y \leq d_2,\ 0 \leq z \leq d_3\} \\
\Gamma_4 &= \{(x, y, z) \mid 0 \leq x \leq d_1,\ y = d_2,\ 0 \leq z \leq d_3\} \\
\Gamma_5 &= \{(x, y, z) \mid x = 0,\ 0 \leq y \leq d_2,\ 0 \leq z \leq d_3\} \\
\Gamma_6 &= \{(x, y, z) \mid 0 \leq x \leq d_1,\ y = 0,\ 0 \leq z \leq d_3\}
\end{aligned} \tag{2.14}$$

The plate's temperatures are evolving according to the 3D heat equation:

$$\frac{\partial T}{\partial t} = \alpha \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} \right),\ (x, y, z) \in \Omega, t \in [0, \infty) \tag{2.15}$$

The energy exchange between the plate and the surrounding air occurs through the boundaries $\Gamma_1, \ldots, \Gamma_6$. More specifically, we assume that this energy exchange occurs through convection and radiation. Furthermore, the heat flux is acting on the top surface Γ_1, and we assume that the initial temperature of the plate is equal to the ambient temperature. These boundary and initial conditions can be mathematically described as follows:

$$k_1 \frac{\partial T}{\partial n_1} = k_2 u + \sigma\epsilon(T_0^4 - T^4) + h_1(T_0 - T),\ (x, y, z) \in \Gamma_1 \tag{2.16}$$

$$k_1 \frac{\partial T}{\partial n_i} = \sigma\epsilon(T_0^4 - T^4) + h_1(T_0 - T),\ (x, y, z) \in \Gamma_i, i = 2, \ldots, 6 \tag{2.17}$$

$$T(x, y, z, 0) = T_0,\ (x, y, z) \in \Omega \tag{2.18}$$

where k_1 is the thermal conductivity, $\partial / \partial n_i$ denotes the normal derivative, σ is the Stephan-Boltzmann constant, ϵ is the emissivity of the surface, T_0 is the constant ambient temperature, h_1 is the heat transfer coefficient, k_2 is the absorption coefficient and u is the heat flux acting on the top surface Γ_1.

The relative temperature is defined by:

$$\tilde{T} = T - T_0 \tag{2.19}$$

Using (2.19), we can rewrite the heat equation (2.15) as follows:

$$\frac{\partial \tilde{T}}{\partial t} = \alpha \left(\frac{\partial^2 \tilde{T}}{\partial x^2} + \frac{\partial^2 \tilde{T}}{\partial y^2} + \frac{\partial^2 \tilde{T}}{\partial z^2} \right), \tag{2.20}$$

After linearizing boundary conditions (2.16)-(2.17) around the ambient temperature T_0, we obtain:

$$k_1 \frac{\partial \tilde{T}}{\partial n_1} = k_2 u - (k_3 + h_1)\tilde{T}, \ (x, y, z) \in \Gamma_1 \tag{2.21}$$

$$k_1 \frac{\partial \tilde{T}}{\partial n_i} = -(k_3 + h_1)\tilde{T}, \ (x, y, z) \in \Gamma_i, i = 2, \ldots, 6 \tag{2.22}$$

$$\tilde{T}(x, y, z, 0) = 0, \ (x, y, z) \in \Omega \tag{2.23}$$

where $k_3 = 4\sigma\epsilon T_0^3$. Similarly to the discretization of the 2D heat equation, the spatial discretization step is denoted by L and the time discretization step by h. This means that we divide d_1, d_2 and d_3 into N, M and P segments, respectively, where the dimension of each segment is equal to L:

$$d_1 = NL, \ \ d_2 = ML, \ \ d_3 = PL \tag{2.24}$$

For convenience, we introduce the following notation:

$$\tilde{T}(iL, jL, lL, kh) = \tilde{T}_{i,j,l}(k), \quad u(iL, jL, kh) = u_{i,j}(k) \tag{2.25}$$

where $i = 0, 1, \ldots, N$, $j = 0, 1, \ldots, M$ and $l = 0, 1, \ldots, P$. The discretization mesh is illustrated in Fig. 2.5.

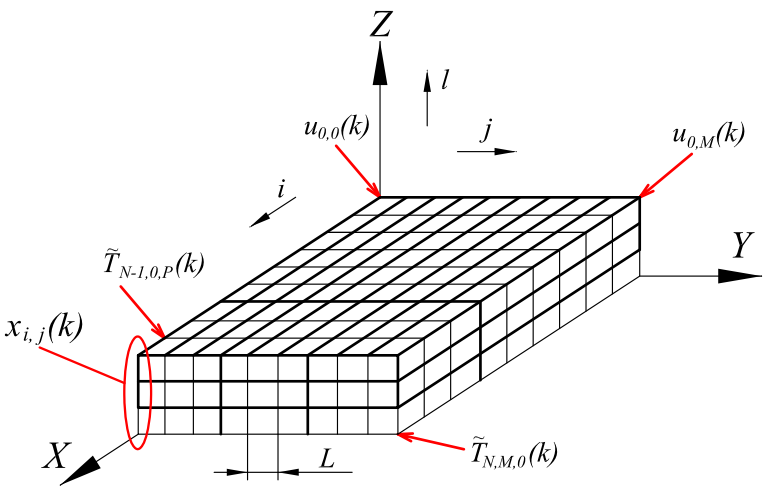


Figure 2.5: The discretization grid for the 3D heat equation (2.20).

We assume that it is possible to measure the temperatures at the top boundary surface Γ_1. That is, we assume that at each discrete time instant k we are able to measure the temperatures $T_{i,j,P}$, $i = 0, 1, \ldots, N$; $j = 0, 1, \ldots, M$. In reality this can

be achieved by using thermo-couples or using a thermal camera [193].

The measurement vector is defined as follows:

$$y_{i,j}(k) = \tilde{T}_{i,j,P}(k) + n_{i,j}(k) \tag{2.26}$$

where $n_{i,j}(k)$ is the measurement noise.

2.3.1 Discretized heat equation

The FD approximations of the partial derivatives are defined by [97]:

$$\frac{\partial \tilde{T}}{\partial t} \approx \frac{\tilde{T}_{i,j,l}(k+1) - \tilde{T}_{i,j,l}(k)}{h} \tag{2.27}$$

$$\frac{\partial^2 \tilde{T}}{\partial x^2} \approx \frac{\tilde{T}_{i+1,j,l}(k) - 2\tilde{T}_{i,j,l}(k) + \tilde{T}_{i-1,j,l}(k)}{L^2} \tag{2.28}$$

$$\frac{\partial^2 \tilde{T}}{\partial y^2} \approx \frac{\tilde{T}_{i,j+1,l}(k) - 2\tilde{T}_{i,j,l}(k) + \tilde{T}_{i,j-1,l}(k)}{L^2} \tag{2.29}$$

$$\frac{\partial^2 \tilde{T}}{\partial z^2} \approx \frac{\tilde{T}_{i,j,l+1}(k) - 2\tilde{T}_{i,j,l}(k) + \tilde{T}_{i,j,l-1}(k)}{L^2} \tag{2.30}$$

Using (2.27)-(2.30) we obtain the discretized heat equation:

$$\begin{aligned}\tilde{T}_{i,j,l}(k+1) =& k_4\tilde{T}_{i,j,l}(k) + k_5\tilde{T}_{i+1,j,l}(k) + k_5\tilde{T}_{i-1,j,l}(k) + k_5\tilde{T}_{i,j+1,l}(k) + k_5\tilde{T}_{i,j-1,l}(k) \\ &+ k_5\tilde{T}_{i,j,l+1}(k) + k_5\tilde{T}_{i,j,l-1}(k)\end{aligned} \tag{2.31}$$

where $k_4 = 1 - 6h\alpha/L^2$ and $k_5 = h\alpha/L^2$. To discretize the boundary condition on the surface Γ_1, the following approximation is introduced:

$$\frac{\partial \tilde{T}}{\partial n_1} \approx \frac{\tilde{T}_{i,j,P+1}(k) - \tilde{T}_{i,j,P-1}(k)}{2L} \tag{2.32}$$

From (2.21) and (2.32) we obtain:

$$\Gamma_1: \quad \tilde{T}_{i,j,P+1}(k) = \tilde{T}_{i,j,P-1}(k) - k_7\tilde{T}_{i,j,P}(k) + k_6 u_{i,j}(k) \tag{2.33}$$

where $k_6 = 2Lk_2/k_1$ and $k_7 = 2L(k_3+h_1)/k_1$. The remaining boundary conditions are discretized using the similar procedure. The discretized boundary conditions are:

$$\Gamma_2: \ \tilde{T}_{i,j,-1}(k) = \tilde{T}_{i,j,1}(k) - k_7\tilde{T}_{i,j,0}(k) \tag{2.34}$$

$$\Gamma_3: \ \tilde{T}_{N+1,j,l}(k) = \tilde{T}_{N-1,j,l}(k) - k_7\tilde{T}_{N,j,l}(k) \tag{2.35}$$

$$\Gamma_4: \ \tilde{T}_{i,M+1,l}(k) = \tilde{T}_{i,M-1,l}(k) - k_7\tilde{T}_{i,M,l}(k) \tag{2.36}$$

$$\Gamma_5: \ \tilde{T}_{-1,j,l}(k) = \tilde{T}_{1,j,l}(k) - k_7\tilde{T}_{0,j,l}(k) \tag{2.37}$$

$$\Gamma_6: \ \tilde{T}_{i,-1,l}(k) = \tilde{T}_{i,1,l}(k) - k_7\tilde{T}_{i,0,l}(k) \tag{2.38}$$

If we would write the discretized heat equation (2.31) for the mesh points belonging to the boundary surfaces, we would observe that these equations contain temperatures that are outside the domain Ω (body of the plate). We eliminate these additional terms by substituting (2.33)-(2.38) in (2.31). To illustrate this, consider the heat equation (2.31) written for the point $(0, 0, 0)$:

$$\begin{aligned}\tilde{T}_{0,0,0}(k+1) = k_4\tilde{T}_{0,0,0}(k) + k_5\tilde{T}_{-1,0,0}(k) + k_5\tilde{T}_{1,0,0}(k)\\ +k_5\tilde{T}_{0,1,0}(k) + k_5\tilde{T}_{0,-1,0}(k) + k_5\tilde{T}_{0,0,1}(k) + k_5\tilde{T}_{0,0,-1}(k)\end{aligned} \tag{2.39}$$

The boundary conditions corresponding to $(0, 0, 0)$ are:

$$\tilde{T}_{-1,0,0}(k) = \tilde{T}_{1,0,0}(k) - k_7\tilde{T}_{0,0,0}(k) \tag{2.40}$$

$$\tilde{T}_{0,-1,0}(k) = \tilde{T}_{0,1,0}(k) - k_7\tilde{T}_{0,0,0}(k) \tag{2.41}$$

$$\tilde{T}_{0,0,-1}(k) = \tilde{T}_{0,0,1}(k) - k_7\tilde{T}_{0,0,0}(k) \tag{2.42}$$

By substituting these boundary conditions in the discretized heat equation (2.39), we obtain:

$$\tilde{T}_{0,0,0}(k+1) = k_8\tilde{T}_{0,0,0}(k) + 2k_5\tilde{T}_{1,0,0}(k) + 2k_5\tilde{T}_{0,1,0}(k) + 2k_5\tilde{T}_{0,0,1}(k) \tag{2.43}$$

where $k_8 = k_4 - 3k_5k_7$. Using the same principle we can eliminate additional terms from the discretized equations formed for remaining boundary surfaces.

Lifting procedure

Here we will generalize the lifting procedure used to form the structured state-space model of the discretized 2D heat equation. The local states of the discretized 2D heat equation are defined by lifting the temperatures over the j direction. Then, these local states were lifted over the i direction to obtain the global state-space model (2.11).

Using the same lifting principle,

1. *We first lift the temperatures over the l direction (that is over the z domain, see Fig. 2.5). This way, we define the local state-vector $\mathbf{x}_{i,j}(k)$ of the local subsystem $\mathcal{S}_{i,j}$:*

$$\mathbf{x}_{i,j}(k) = \begin{bmatrix}\tilde{T}_{i,j,0}(k)\\ \tilde{T}_{i,j,1}(k)\\ \vdots\\ \tilde{T}_{i,j,P}(k)\end{bmatrix} \tag{2.44}$$

2. *Then, we lift these lifted vectors over the i direction.*

3. *Finally, we lift once more these "double lifted" vectors over the j direction. This way, we obtain the global state vector $\underline{\mathbf{x}}(k)$.*

Once we have eliminated from the discretized heat equation all the temperatures that are outside of the domain Ω, we can apply the above explained procedure to obtain the global state-space model. Namely, for $i = 0$ and $j = 0$ we can write:

$$\mathbf{x}_{0,0}(k+1) = A_c\mathbf{x}_{0,0}(k) + E_s\mathbf{x}_{0,1}(k) + E_s\mathbf{x}_{1,0}(k) + Bu_{0,0}(k) \tag{2.45}$$

where the vector $\mathbf{x}_{0,0}$ is defined in (2.44) for $(i,j) = (0,0)$, the matrices A_c, E_s and B are defined in (2.56) and the vector $u_{0,0}(k)$ is an input vector acting at the node (0,0,P) on the top surface Γ_1. The equation (2.45) is a state equation of the local subsystem $\mathcal{S}_{0,0}$. By grouping this state equation with the corresponding output equation (2.26), we obtain the state-space model of the local subsystem $\mathcal{S}_{0,0}$:

$$\mathcal{S}_{0,0}\begin{cases} \mathbf{x}_{0,0}(k+1) = A_c\mathbf{x}_{0,0}(k) + E_s\mathbf{x}_{0,1}(k) + E_s\mathbf{x}_{1,0}(k) + Bu_{0,0}(k) \\ y_{0,0}(k) = C\mathbf{x}_{0,0}(k) + n_{0,0}(k) \end{cases} \tag{2.46}$$

where $C = \begin{bmatrix} 0 & 0 & \dots & 1 \end{bmatrix}$ (the C matrix "tells us" that the temperature at the top surface Γ_1 is measured). Similarly, we define the state-space models of other local subsystems:

$$\mathcal{S}_{i,0}\begin{cases} \mathbf{x}_{i,0}(k+1) = A_s\mathbf{x}_{i,0}(k) + E_I\mathbf{x}_{i-1,0}(k) + E_I\mathbf{x}_{i+1,0}(k) + E_s\mathbf{x}_{i,1}(k) \\ +Bu_{i,0}(k) \\ y_{i,0}(k) = C\mathbf{x}_{i,0}(k) + n_{i,0}(k) \end{cases}$$
$$i = 1, \dots, N-1 \tag{2.47}$$

$$\mathcal{S}_{N,0}\begin{cases} \mathbf{x}_{N,0}(k+1) = A_c\mathbf{x}_{N,0}(k) + E_s\mathbf{x}_{N,1}(k) + E_s\mathbf{x}_{N-1,0}(k) + Bu_{N,0}(k) \\ y_{N,0}(k) = C\mathbf{x}_{N,0}(k) + n_{N,0}(k) \end{cases} \tag{2.48}$$

$$\mathcal{S}_{0,j}\begin{cases} \mathbf{x}_{0,j}(k+1) = A_s\mathbf{x}_{0,j}(k) + E_s\mathbf{x}_{1,j}(k) + E_I\mathbf{x}_{0,j-1}(k) + E_I\mathbf{x}_{0,j+1}(k) + Bu_{0,j}(k) \\ y_{0,j}(k) = C\mathbf{x}_{0,j}(k) + n_{0,j}(k) \end{cases}$$
$$j = 1, \dots, M-1 \tag{2.49}$$

$$\mathcal{S}_{i,j}\begin{cases} \mathbf{x}_{i,j}(k+1) = A_I\mathbf{x}_{i,j}(k) + E_I\mathbf{x}_{i-1,j}(k) + E_I\mathbf{x}_{i+1,j}(k) + E_I\mathbf{x}_{i,j-1}(k) \\ +E_I\mathbf{x}_{i,j+1}(k) + Bu_{i,j}(k) \\ y_{i,j}(k) = C\mathbf{x}_{i,j}(k) + n_{i,j}(k) \end{cases}$$
$$i = 1, \dots, N-1, j = 1, \dots, M-1, \tag{2.50}$$

$$\mathcal{S}_{N,j}\begin{cases} \mathbf{x}_{N,j}(k+1) = A_s\mathbf{x}_{N,j}(k) + E_s\mathbf{x}_{N-1,j}(k) + E_I\mathbf{x}_{N,j-1}(k) \\ +E_I\mathbf{x}_{N,j+1}(k) + Bu_{N,j}(k) \\ y_{N,j}(k) = C\mathbf{x}_{N,j}(k) + n_{N,j}(k) \end{cases}$$
$$j = 1, \dots, M-1, \tag{2.51}$$

$$\mathcal{S}_{0,M}\begin{cases}\mathbf{x}_{0,M}(k+1)=A_c\mathbf{x}_{0,M}(k)+E_s\mathbf{x}_{0,M-1}(k)+E_s\mathbf{x}_{1,M}(k)+Bu_{0,M}(k)\\ y_{0,M}(k)=C\mathbf{x}_{0,M}(k)+n_{0,M}(k)\end{cases} \tag{2.52}$$

$$\begin{aligned}&\mathcal{S}_{i,M}\begin{cases}\mathbf{x}_{i,M}(k+1)=A_s\mathbf{x}_{i,M}(k)+E_s\mathbf{x}_{i,M-1}(k)+E_I\mathbf{x}_{i-1,M}(k)\\ +E_I\mathbf{x}_{i+1,M}(k)+Bu_{i,M}(k)\\ y_{i,M}(k)=C\mathbf{x}_{i,M}(k)+n_{i,M}(k)\end{cases}\\ &i=1,\ldots,N-1\end{aligned} \tag{2.53}$$

$$\mathcal{S}_{N,M}\begin{cases}\mathbf{x}_{N,M}(k+1)=A_c\mathbf{x}_{N,M}(k)+E_s\mathbf{x}_{N,M-1}(k)+E_s\mathbf{x}_{N-1,M}(k)\\ +Bu_{N,M}(k)\\ y_{N,M}(k)=C\mathbf{x}_{N,M}(k)+n_{N,M}(k)\end{cases} \tag{2.54}$$

where

$$A_c=\begin{bmatrix}k_8 & 2k_5 & & & \\ k_5 & k_9 & k_5 & & \\ & \ddots & & & \\ & & k_5 & k_9 & k_5\\ & & & 2k_5 & k_8\end{bmatrix},\ A_s=\begin{bmatrix}k_9 & 2k_5 & & & \\ k_5 & k_{11} & k_5 & & \\ & \ddots & & & \\ & & k_5 & k_{11} & k_5\\ & & & 2k_5 & k_8\end{bmatrix},\ A_I=\begin{bmatrix}k_{11} & 2k_5 & & & \\ k_5 & k_4 & k_5 & & \\ & \ddots & & & \\ & & k_5 & k_4 & k_5\\ & & & 2k_5 & k_{11}\end{bmatrix}, \tag{2.55}$$

$$E_I=\begin{bmatrix}k_5 & & & \\ & k_5 & & \\ & & \ddots & \\ & & & k_5\\ & & & & k_5\end{bmatrix},\ E_s=2E_I,\ B=\begin{bmatrix}0\\0\\\vdots\\0\\k_6\end{bmatrix} \tag{2.56}$$

and where $k8=k_4-3k_5k_7$, $k_9=k_4-2k_5k_7$, $k_{10}=k_5k_6$ and $k_{11}=k_4-k_5k_7$. The interconnections of the local subsystem $\mathcal{S}_{i,i}$ are illustrated in Fig. 2.6.

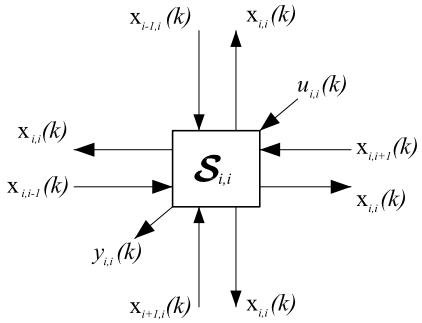


Figure 2.6: Inputs, outputs and interconnections of the local subsystem $\mathcal{S}_{i,i}$.

The interconnection pattern of the local subsystems (2.46)-(2.54) is illustrated in Fig.2.7.

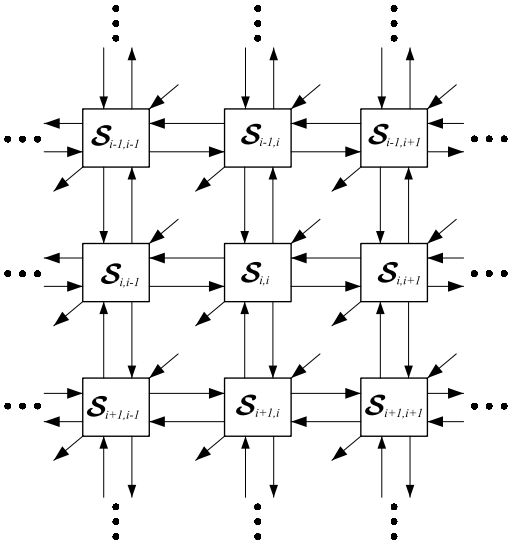


Figure 2.7: Interconnection pattern of local subsystems obtained by discretization of the heat equation on the 3D spatial domain.

By lifting the local states $\mathbf{x}_{i,j}(k)$ over the discretized x domain (the i direction) we define the following vector:

$$\mathbf{x}_j^x(k) = \begin{bmatrix} \mathbf{x}_{0,j}(k) \\ \mathbf{x}_{1,j}(k) \\ \vdots \\ \mathbf{x}_{N,j}(k) \end{bmatrix} \tag{2.57}$$

The following vectors are defined in the same manner:

$$\mathbf{u}_j^x(k) = \begin{bmatrix} u_{0,j}(k) \\ u_{1,j}(k) \\ \vdots \\ u_{N,j}(k) \end{bmatrix}, \ \mathbf{y}_j^x(k) = \begin{bmatrix} y_{0,j}(k) \\ y_{1,j}(k) \\ \vdots \\ y_{N,j}(k) \end{bmatrix}, \ \mathbf{n}_j^x(k) = \begin{bmatrix} n_{0,j}(k) \\ n_{1,j}(k) \\ \vdots \\ n_{N,j}(k) \end{bmatrix}, \tag{2.58}$$

The "double lifted" state and output equations have the following form:

$$\begin{aligned} \mathbf{x}_0^x(k+1) &= A_b^x \mathbf{x}_0^x(k) + E_b^x \mathbf{x}_1^x(k) + B^x \mathbf{u}_0^x(k) \\ \mathbf{y}_0^x(k) &= C_b^x \mathbf{x}_0^x(k) + \mathbf{n}_0^x(k) \end{aligned} \tag{2.59}$$

$$\mathbf{x}_j^x(k+1) = A^x \mathbf{x}_j^x(k) + E^x \mathbf{x}_{j-1}^x(k) + E^x \mathbf{x}_{j+1}^x(k) + B^x \mathbf{u}_j^x(k)$$

$$\begin{aligned} \mathbf{y}_j^x(k) &= C_b^x \mathbf{x}_j^x(k) + \mathbf{n}_j^x(k) \\ j &= 2, \ldots, M-1 \end{aligned} \tag{2.60}$$

$$\begin{aligned} \mathbf{x}_M^x(k+1) &= A_b^x \mathbf{x}_M^x(k) + E_b^x \mathbf{x}_{M-1}^x(k) + B^x \mathbf{u}_M^x(k) \\ \mathbf{y}_M^x(k) &= C_b^x \mathbf{x}_M^x(k) + \mathbf{n}_M^x(k) \end{aligned} \tag{2.61}$$

where

$$A_b^x = \begin{bmatrix} A_c & E_s & & & \\ E_I & A_s & E_I & & \\ & \ddots & & & \\ & & E_I & A_s & E_I \\ & & & E_s & A_c \end{bmatrix}, E_b^w = \begin{bmatrix} E_s & & & \\ & E_I & & \\ & & \ddots & \\ & & & E_I \\ & & & & E_s \end{bmatrix}, \; B^x = \begin{bmatrix} B & & & & \\ & B & & & \\ & & \ddots & & \\ & & & B & \\ & & & & B \end{bmatrix},$$

$$C^x = \begin{bmatrix} C & & & & \\ & C & & & \\ & & \ddots & & \\ & & & C & \\ & & & & C \end{bmatrix}, \; A^x = \begin{bmatrix} A_c & E_s & & & \\ E_I & A_I & E_I & & \\ & \ddots & & & \\ & & E_I & A_I & E_I \\ & & & E_s & A_s \end{bmatrix}, \; E^x = \begin{bmatrix} E_I & & & & \\ & E_I & & & \\ & & \ddots & & \\ & & & E_I & \\ & & & & E_I \end{bmatrix} \tag{2.62}$$

Finally, we lift the "double lifted" vectors $\mathbf{x}_j^x(k)$ over the discretized y domain (the j direction, see Fig. 2.5) and we define *the global state vector*:

$$\underline{\mathbf{x}}(k) = \begin{bmatrix} \mathbf{x}_0^x(k) \\ \mathbf{x}_1^x(k) \\ \vdots \\ \mathbf{x}_M^x(k) \end{bmatrix} \tag{2.63}$$

The global input, output and measurement noise vectors are defined as follows:

$$\underline{\mathbf{u}} = \begin{bmatrix} \mathbf{u}_0^x(k) \\ \mathbf{u}_1^x(k) \\ \vdots \\ \mathbf{u}_M^x(k) \end{bmatrix}, \; \underline{\mathbf{y}} = \begin{bmatrix} \mathbf{y}_0^x(k) \\ \mathbf{y}_1^x(k) \\ \vdots \\ \mathbf{y}_M^x(k) \end{bmatrix}, \; \underline{\mathbf{n}} = \begin{bmatrix} \mathbf{n}_0^x(k) \\ \mathbf{n}_1^x(k) \\ \vdots \\ \mathbf{n}_M^x(k) \end{bmatrix} \tag{2.64}$$

The global state-space model has the following form:

$$\mathcal{S} \begin{cases} \underline{\mathbf{x}}(k+1) & = \underline{A}\underline{\mathbf{x}}(k) + \underline{B}\underline{\mathbf{u}}(k) \\ \underline{\mathbf{y}}(k) & = \underline{C}\underline{\mathbf{x}}(k) + \underline{\mathbf{n}}(k) \end{cases} \tag{2.65}$$

$$\underline{A} = \begin{bmatrix} A_b^x & E_b^x & & & \\ E^x & A^x & E^x & & \\ & \ddots & & & \\ & & E^x & A^x & E^x \\ & & & E_b^x & A_b^x \end{bmatrix}, \underline{B} = \begin{bmatrix} B^x & & \\ & \ddots & \\ & & B^x \end{bmatrix}, \underline{C} = \begin{bmatrix} C^x & & \\ & \ddots & \\ & & C^x \end{bmatrix} \tag{2.66}$$

The sparsity patterns of the global system matrices $\underline{A}$, $\underline{B}$ and $\underline{C}$ are illustrated in Fig. 2.8. The matrix $\underline{A}$ is a sparse, multi-banded matrix, and $\underline{B}$ and $\underline{C}$ are diagonal matrices.

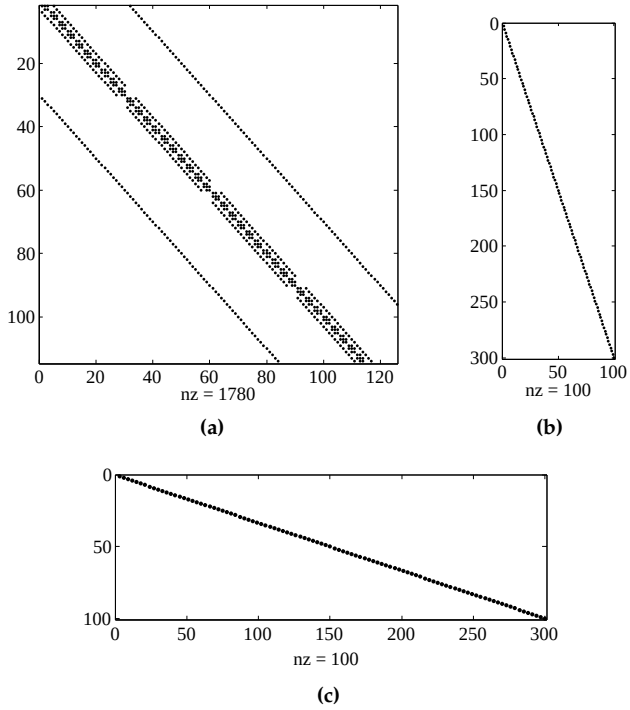


Figure 2.8: Structure of the system matrices of the global state-space model (1.4) obtained using finite differences approximation of the heat equation: (a) segment of $\underline{A}$; (b) $\underline{B}$; (c) $\underline{C}$ ("nz" denotes the number of non-zero elements).

2.4 Finite element discretization of the thermoelastic equations

In this section we present a Finite Element (FE) model that describes thermoelastic deformations of a mirror used in optical systems. The developed model is in a

descriptor state-space form [31, 55].

Consider a circular mirror illustrated in Fig. 2.9(a). A portion of the energy of the incoming beam is absorbed by a thin coating on the top surface. The absorbed energy heats up the mirror and it induces thermoelastic deformations [22; 57; 59]. Because the mirror surface is deformed, the reflected wavefront is distorted. These wavefront distortions (aberrations) can significantly degrade the imaging quality of an optical system.

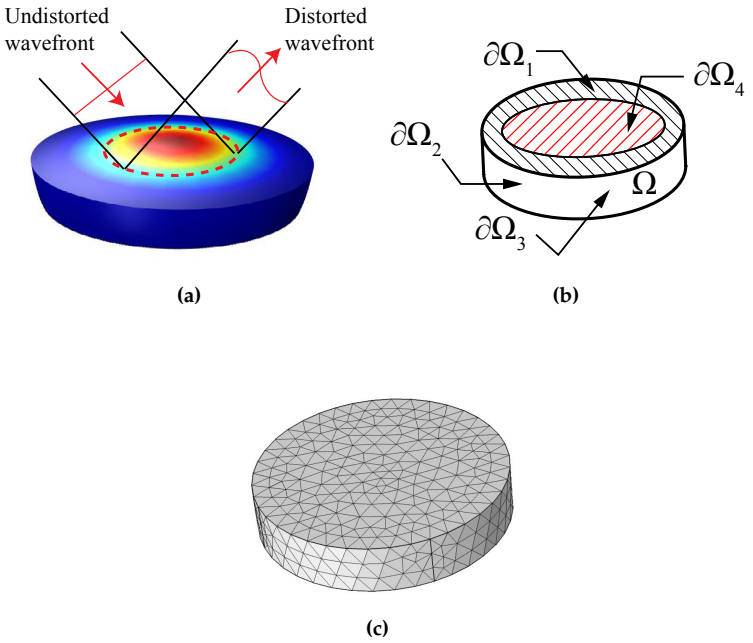


Figure 2.9: (a) Heating of the mirror by the absorption of the exposure energy and thermally induced wavefront aberrations; (b) Characteristics regions used to define the boundary conditions of the thermoelastic equations; (c) The FE mesh.

Without the loss of generality, we assume that the mirror is placed in a vacuum environment with a constant ambient temperature (that is, there is no convection between the mirror and the environment). Furthermore, it is assumed that mirror heat loses are only due to thermal radiation.
For presentation clarity, in Fig.2.9(b) we have denoted the characteristic regions of the mirror. The mirror is denoted by set of points Ω and its top surface is divided into two regions: $\partial\Omega_4$ region, denoting an area that reflects the light beam, and the remaining area of the top surface $\partial\Omega_1$. The side surface of the mirror and its bottom surface are denoted by $\partial\Omega_2$ and $\partial\Omega_3$, respectively.
The dynamical behavior of thermally induced deformations is governed by the

thermoelastic equations [57]:

$$\kappa\nabla^2 T = \rho c\frac{\partial T}{\partial t}, \quad (x,y,z)\in\Omega \tag{2.67}$$

$$c_1\nabla^2\mathbf{w} + (c_2+c_1)\nabla(\nabla\cdot\mathbf{w}) + c_3\nabla T = \rho\frac{\partial^2\mathbf{w}}{\partial t^2}, \quad (x,y,z)\in\Omega \tag{2.68}$$

where T is the temperature, $\mathbf{w}\in\mathbb{R}^3$ is the displacement vector, ρ is the density, c is the specific heat at constant deformation, κ is the thermal conductivity, (c_1, c_2) are Lamé constants, $c_3 = -\alpha(3c_2+2c_1)$ and α is the Coefficient of Thermal Expansion (CTE). We assume that the CTE does not depend on the temperature (see Remark 2.2). The boundary conditions are defined as follows:

$$-\kappa\frac{\partial T}{\partial z} = 4\sigma c_4 T_A^3(T-T_A), \quad (x,y,z)\in\partial\Omega_1 \tag{2.69}$$

$$-\kappa\frac{\partial T}{\partial \xi} = 4\sigma c_4 T_A^3(T-T_A), \quad (x,y,z)\in\partial\Omega_2 \tag{2.70}$$

$$-\kappa\frac{\partial T}{\partial z} = -4\sigma c_4 T_A^3(T-T_A), \quad (x,y,z)\in\partial\Omega_3 \tag{2.71}$$

$$-\kappa\frac{\partial T}{\partial z} = 4\sigma c_4 T_A^3(T-T_A) - c_4 f, \quad (x,y,z)\in\partial\Omega_4 \tag{2.72}$$

where $\partial/\partial z$ and $\partial/\partial\xi$ are normal derivatives (ξ is the radial coordinate), σ is the Stefan-Boltzmann constant, c_4 is the emissivity, T_A is a constant ambient temperature, f is the intensity distribution (a thermal load or an input) over $\partial\Omega_4$ and c_5 is the efficiency of conversion of light into the heat power. We assume that f is static. For simplicity, the nonlinear radiation boundary conditions are linearized. We assume that initial mirror temperatures are equal to ambient temperature.
Next, it is assumed that all six degrees of freedom of movement of the mirror are fixed. Finally, we assume that there are no external mechanical forces acting on the boundary surfaces.

The displacement vector $\mathbf{w}$ does not appear in the heat equation (2.67). Consequently, the solution of the heat equation, denoted by $T(x,y,z,t)$, can be found independently from the elastic equation (2.68). Once the solution $T(x,y,z,t)$ has been determined, it can be used to solve the elastic equation. However, in this thesis we will not directly solve the thermoelastic equations. Instead we will discretize them using the FE method.

For brevity, we will not explicitly derive the FE equations. The FE discretization of the thermoelastic equations (2.67)-(2.72) has been extensively studied in literature, see for example [194]. We have used COMSOL Multiphysics® software to model the mirror and to discretize the thermoelastic equations. The mesh that was used for FE discretization is illustrated in Fig. 2.9(c). Using LiveLink™, the FE matrices are exported from the COMSOL model to the MATLAB® workspace. The discretized elastic equation has the following form [194]:

$$M_{11}\ddot{\mathbf{s}} + K_{11}\mathbf{s} + K_{12}\mathbf{x} = \mathbf{l}_1 \tag{2.73}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the vector of temperatures at discretization nodes and $\mathbf{s} \in \mathbb{R}^m$ is the vector of displacements at discretization nodes, $\mathbf{l}_1 \in \mathbb{R}^m$ is a constant vector, and $K_{11} \in \mathbb{R}^{m \times m}$, $K_{12} \in \mathbb{R}^{m \times n}$ and $M_{11} \in \mathbb{R}^{m \times m}$ are the FE matrices.

The sparsity structure of the FE matrices is important for the development of computationally efficient moving horizon estimation method in Chapter 4.2. The structure of K_{11} and K_{12} is illustrated in Figs. 2.10(a)-(b) (the structure of $M_{11} \in \mathbb{R}^{m \times m}$ is similar to the structure of K_{11}). As we can see from these figures, the FE matrices are sparse banded matrices.

The discretized heat equation has the following form:

$$D_{22}\dot{\mathbf{x}} + K_{22}\mathbf{x} = \mathbf{l}_2 \tag{2.74}$$

where $\mathbf{l}_2 \in \mathbb{R}^n$ is a constant vector, $D_{22} \in \mathbb{R}^{n \times n}$ and $K_{22} \in \mathbb{R}^{n \times n}$ are the FE matrices. The vector $\mathbf{l}_2$ takes into account the intensity distribution f.

The sparsity structure of the matrices $D_{22} \in \mathbb{R}^{n \times n}$ and $K_{22} \in \mathbb{R}^{n \times n}$ is illustrated in Fig. 2.10(c)-(d).

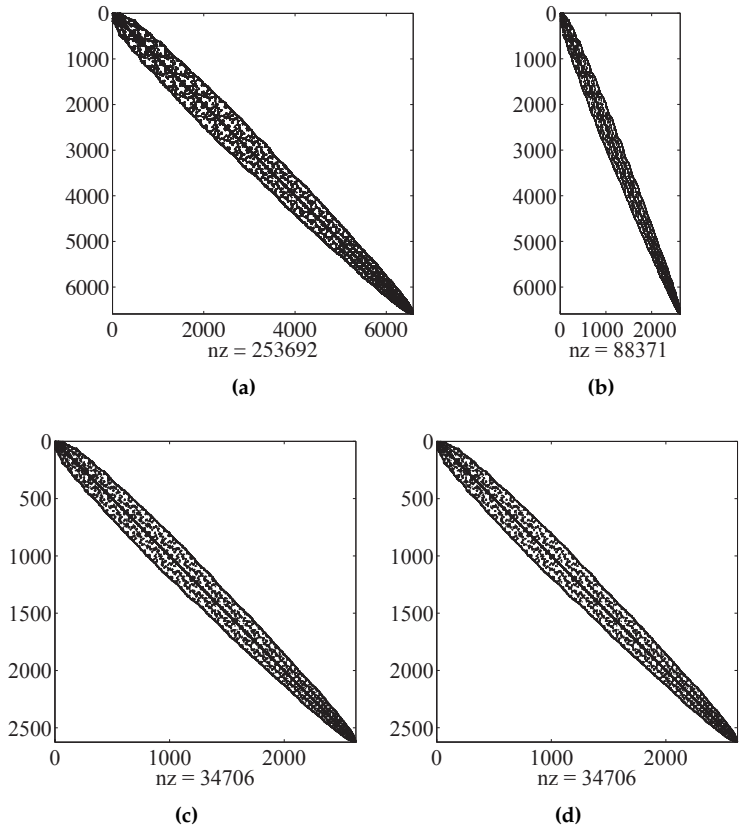


Figure 2.10: Sparsity pattern of the FE matrices; (a) K_{11} ; (b) K_{12}; (c) D_{22}; (d) K_{22}, ″nz″ stands for number of non-zero elements.

For studying the dynamical behavior of mirrors used in optics, it often justified to neglect the dynamics of the elastic equation. In engineering language, the justification for this model simplification follows from the fact that for the types of materials used to fabricate mirrors, there is a very short delay between the temperature change and induced deformations. Much larger delay occurs between the input f (intensity distribution acting on $\partial\Omega_4$) and the temperature change. Mathematically speaking, the dynamics of the heat equation is much slower than the dynamics of the elastic equation. On the basis of this physical insight, we introduce the following assumption:

Assumption 2.1 *The term* $\ddot{\mathbf{s}}$ *in* (2.73) *can be neglected.*

Taking into account Assumption 2.1, the discretized elastic equation takes the following form:

$$K_{11}\mathbf{s} + K_{12}\mathbf{x} = \mathbf{l}_1 \tag{2.75}$$

We assume that the wavefront aberrations induced by the deformations of the mirror's surface are measured using a WaveFront Sensor (WFS). Several wavefront sensing techniques have been developed for characterizing aberrations created by the surface irregularities and by deformations of optical elements. For example, in optical lithography machines interferometers are usually used to measure deformations of optical elements [51; 56]. In [52; 53] the possibility to use Hartmann sensor for measurement of wavefront aberrations in optical lithography machines has been investigated. In [54; 55], measurement techniques based on Shack-Hartmann WFS (S-H WFS) have been proposed for the measurement of thermally induced wavefront aberrations. From the measurements of wavefront aberrations, the deformations of the top mirror surface can be retrieved. The output equation has the following form:

$$\mathbf{y} = W\mathbf{s} + \mathbf{n} \tag{2.76}$$

where $W \in \mathbb{R}^{r\times m}$, $\mathbf{y} \in \mathbb{R}^r$ is the vector of surface deformations and $\mathbf{n} \in \mathbb{R}^r$ is the measurement noise. To make the measurement scenario more realistic, we assume that only surface deformations corresponding to the region that reflects the light beam (that is, corresponding to $\partial\Omega_4$) can be reconstructed. Stated in engineering language, from the total vector of displacements $\mathbf{s}$ the matrix W selects the displacements belonging to the region $\partial\Omega_4$ on the top mirror surface. The sparsity structure of W is illustrated in Fig. 2.11.

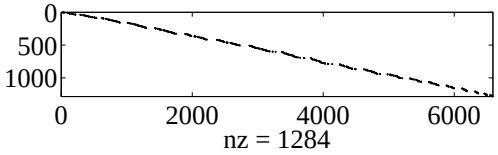


Figure 2.11: The structure of the matrix W for $m = 1284$ and $v = 6588$.

In reality, wavefront measurements are obtained at discrete time instants. Accordingly, time discretization of the FE equations needs to be performed. For the time discretization we use the Euler backward method [97; 105]. Let h denote the time discretization step and let k denote the discrete time instant (the total time is $t = kh$, $k = 0, 1, 2 \ldots$). The discretized elastic equation (2.75) has the following form:

$$K_{11}\mathbf{s}(k) + K_{12}\mathbf{x}(k) = \mathbf{l}_1 \tag{2.77}$$

From the last equation, we have:

$$\mathbf{s}(k) = -K_{11}^{-1}K_{12}\mathbf{x}(k) + K_{11}^{-1}\mathbf{l}_1 \tag{2.78}$$

Applying the backward Euler discretization method to (2.74) we obtain:

$$(D_{22} + hK_{22})\,\mathbf{x}(k) = D_{22}\mathbf{x}(k-1) + h\mathbf{l}_2 \tag{2.79}$$

Substituting (2.78) in the discretized version of (2.76), we obtain:

$$\mathbf{y}(k) = -WK_{11}^{-1}K_{12}\mathbf{x}(k) + WK_{11}^{-1}\mathbf{l}_1 + \mathbf{n}(k) \tag{2.80}$$

Combining (2.79) and (2.80), we arrive at the following state-space model:

$$Q_{22}\mathbf{x}(k) = D_{22}\mathbf{x}(k-1) + \mathbf{c}_1 \tag{2.81}$$

$$\mathbf{y}(k) = C\mathbf{x}(k) + \mathbf{d}_1 + \mathbf{n}(k) \tag{2.82}$$

where the matrices $Q_{22} \in \mathbb{R}^{n\times n}$ and $C \in \mathbb{R}^{r\times n}$ and the vectors $\mathbf{c}_1 \in \mathbb{R}^n$ and $\mathbf{d}_1 \in \mathbb{R}^r$ are defined as follows

$$Q_{22} = D_{22} + hK_{22},\;\; C = -WK_{11}^{-1}K_{12},\;\; \mathbf{c}_1 = h\mathbf{l}_2,\;\; \mathbf{d}_1 = WK_{11}^{-1}\mathbf{l}_1 \tag{2.83}$$

The state-space model (2.81)-(2.82) is in the descriptor form [106; 107]. On the basis of this model, in Chapter 4.2 a computationally efficient Moving Horizon Estimation (MHE) method is developed. In Chapter 7, the MHE method is used to estimate the mirror's temperature distribution from the measurements of surface deformations.

Remark 2.2 *Mirrors used in high power optical systems, such as EUVL machines, can be made of materials which have the CTE coefficient that (nonlinearly) depends on the temperature [62]. State space models of these mirrors have a linear state equation and a nonlinear output equation. That is, the state equation of these models is identical to* (2.81)*, whereas their output equation has the following form [62]:*

$$\mathbf{y}(k) = D_1\mathbf{x}(k) + D_2\mathbf{p}_2\left(\mathbf{x}(k)\right) + D_3\mathbf{p}_3\left(\mathbf{x}(k)\right) + \mathbf{n}(k)$$

where D_1 and D_2 are constant matrices and $\mathbf{p}_2(\cdot)$ and $\mathbf{p}_3(\cdot)$ are nonlinear (polynomial) functions of $\mathbf{x}(k)$. This nonlinearity might complicate the estimation of thermally induced wavefront aberrations. One of the possible solutions is to approximate the nonlinear output equation by a piecewise linear function and to use theoretical approaches proposed in [195; 196] to develop control and estimation methods. Another possibility for estimating

the state of the thermoelastic equations with temperature depending CTE, is to use the Newton observer for nonlinear systems that is briefly explained in Chapter 4, Section 4.3.

3 CHAPTER

Structure preserving lifting technique and inverses of Gramians and lifted system matrices

In this chapter a new, structure preserving lifting technique for large-scale interconnected systems is presented. This lifting technique ensures that the sparsity structure of the global system is preserved in its lifted state-space model. Next, it is proved that inverses of lifted system matrices and Gramians of interconnected systems belong to a class of off-diagonally decaying matrices. Consequently, these inverses can be approximated by sparse banded matrices. To compute the approximate inverses the Chebyshev approximation method and the Newton iteration are used. The accuracy of the Chebyshev approximation method is analyzed and a new upper bound on the approximation errors is derived. Several methods for decreasing the computational and memory complexity of these approximation methods are presented.
The approximation framework presented in this Chapter is used throughout the thesis to develop distributed and computationally efficient, centralized identification and estimation methods for large-scale interconnected systems.

3.1 Introduction

In a large variety of estimation and control problems, the observability (controllability) Gramians and the lifted system matrices need to be inverted. For example, in the norm optimal Iterative Learning Control (ILC) [197; 198; 199], a control action is derived by computing a regularized pseudo-inverse of the impulse response matrix. On the other hand, the Moving Horizon Estimation (MHE) method computes a state estimate by inverting the finite-time observability Gramian.

For the real-time implementation of control and estimation algorithms, these inverses need to be precomputed and stored in a controller memory[1]. However, the inverses of lifted system matrices and Gramians are in the general case dense matrices and consequently, n^2 memory locations are needed to store them (n is the number of states). All this implies that it might not be possible to implement the above mentioned control and estimation algorithms for large-scale systems.

This problem is illustrated by the following example. Our experience with large-scale computations in MATLAB, shows that on a standard desktop computer with 4 GB of Random Access Memory (RAM) it is not possible to perform basic operations on matrices which dimensions exceed 15,000. For example, the state-space model of the discretized 2D heat equation (see Chapter 2) can easily have more than 20,000 states. The system of 20,000 states corresponds to a small 2D grid of 100x200 nodes. The system matrices of this state-space model are sparse, and consequently, the finite-time observability Gramian is also sparse. Because the observability Gramian is a sparse matrix, it can be easily stored in a computer memory. However, the computer stops working when we try to invert this Gramian using standard MATLAB functions. This is because the capacity of the RAM is too small to perform this inversion[2].

To overcome these challenges, in this Chapter we develop a computationally efficient, sparsity preserving framework for inversion of lifted system matrices and Gramians. First, a structure preserving lifting technique for large-scale interconnected systems is introduced. This lifting technique ensures that the sparsity structure of the global system is preserved in its lifted state-space model. Then, it is proved that the (pseudo) inverses of lifted system matrices and Gramians are off-diagonally decaying matrices. Consequently, these inverses can be approximated by sparse banded matrices. To compute these approximate, sparse inverses of lifted system matrices and Gramians, the Chebyshev approximation method and the Newton iteration are used. The accuracy of the Chebyshev approximation method is analyzed and a new upper bound on the approximation errors is derived. Several methods for decreasing the computational and memory complexities of these approximation methods are presented.
The approximation framework presented in this Chapter is used throughout the thesis to develop distributed and computationally efficient centralized identification and estimation methods for large-scale interconnected systems.

This chapter is organized as follows. In Section 3.2 we present the structure preserving lifting technique. In Section 3.3 we study Gramians of large-scale interconnected systems. In Section 3.4 we present approximation algorithms. Finally, in Section 3.5 we present numerical results.

[1] Control and estimation algorithms can be also implemented without explicitly computing the inverses of lifted system matrices and Gramians. For example, a control action or a state estimate can be computed using iterative methods for solving linear system of equations, such as the conjugate gradient method [200]. However, because the computation time is limited by the sampling period of the control system, for real-time implementations it is generally not advisable to use iterative methods.

[2] Authors are aware that beside MATLAB there are other computer programs that are more suitable for large-scale computations on desktop computers. Although these programs can handle larger problems, they still cannot handle a system consisting of more than 100,000 states.

3.2 Structure preserving lifting technique

For presentation clarity the structure preserving lifting technique will be explained on the example of a large-scale system consisting of a string of local subsystems. Consider the following state-space model:

$$\mathcal{S}\left\{\begin{array}{rl} \underline{\mathbf{x}}(k+1) & = \underline{A}\underline{\mathbf{x}}(k)+\underline{B}\underline{\mathbf{u}}(k) \\ \underline{\mathbf{y}}(k) & = \underline{C}\underline{\mathbf{x}}(k)+\underline{\mathbf{n}}(k) \end{array}\right. \tag{3.1}$$

where the system matrices have the following structure:

$$\underline{A}=\begin{bmatrix} A_{1,1} & E_{1,2} & & & \\ E_{2,1} & A_{2,2} & E_{2,3} & & \\ & \ddots & & & \\ & E_{N-1,N-2} & A_{N-1,N-1} & E_{N-1,N} \\ & & E_{N,N-1} & A_{N,N} \end{bmatrix}, \underline{B}=\begin{bmatrix} B_1 & & \\ & \ddots & \\ & & B_N \end{bmatrix}$$
$$\underline{C}=\begin{bmatrix} C_1 & & \\ & \ddots & \\ & & C_N \end{bmatrix},\ \underline{\mathbf{y}}(k)=\begin{bmatrix} \mathbf{y}_1(k) \\ \vdots \\ \mathbf{y}_N(k) \end{bmatrix}, \underline{\mathbf{x}}(k)=\begin{bmatrix} \mathbf{x}_1(k) \\ \vdots \\ \mathbf{x}_N(k) \end{bmatrix},\ \underline{\mathbf{u}}(k)=\begin{bmatrix} \mathbf{u}_1(k) \\ \vdots \\ \mathbf{u}_N(k) \end{bmatrix} \tag{3.2}$$

and similarly we define $\underline{\mathbf{n}}(k)$. The system $\mathcal{S}$ is referred to as *the global system*, with *the global state* $\underline{\mathbf{x}}(k) \in \mathbb{R}^{Nn}$, *the global output* $\underline{\mathbf{y}}(k) \in \mathbb{R}^{Nr}$, *the global input* $\underline{\mathbf{u}}(k) \in \mathbb{R}^{Nm}$ and *the global measurement noise* $\underline{\mathbf{n}}(k) \in \mathbb{R}^{Nr}$. The system matrices $\underline{A} \in \mathbb{R}^{Nn\times Nn}$, $\underline{B} \in \mathbb{R}^{Nn\times Nm}$ and $\underline{C} \in \mathbb{R}^{Nr\times Nn}$ are referred to as *the global system matrices*. The global system $\mathcal{S}$ consists of the interconnection of N *local subsystems* $\mathcal{S}_i$:

$$\mathcal{S}_i\left\{\begin{array}{rl} \mathbf{x}_i(k+1) & = A_{i,i}\mathbf{x}_i(k)+E_{i,i-1}\mathbf{x}_{i-1}(k)+E_{i,i+1}\mathbf{x}_{i+1}(k)+B_i\mathbf{u}_i(k) \\ \mathbf{y}_i(k) & = C_i\mathbf{x}_i(k)+\mathbf{n}_i(k) \end{array}\right. \tag{3.3}$$

where $\mathbf{x}_i(k) \in \mathbb{R}^n$ is *the local state* of the local subsystem $\mathcal{S}_i$, $\mathbf{x}_{i-1}(k) \in \mathbb{R}^n$ and $\mathbf{x}_{i+1}(k) \in \mathbb{R}^n$ are the local states of *the neighboring local subsystems* $\mathcal{S}_{i-1}$ and $\mathcal{S}_{i+1}$ respectively, $\mathbf{y}_i(k) \in \mathbb{R}^r$ is *the local output*, $\mathbf{u}_i(k) \in \mathbb{R}^m$ is *the local input* and $\mathbf{n}_i(k) \in \mathbb{R}^r$ is *the local measurement noise*. The interconnection structure of local subsystems is illustrated in Fig. 3.1.

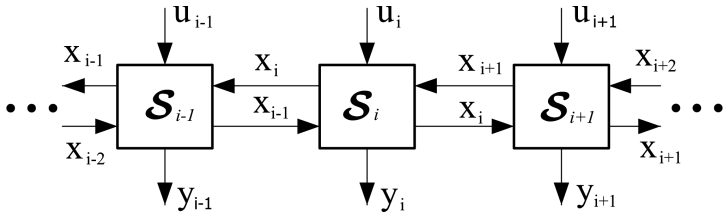


Figure 3.1: The interconnection structure of the local subsystems of the global system (3.1).

The matrices in (3.3) are constant matrices and are referred to as *the local system matrices*. Without loss of generality, we have assumed that all local subsystems have identical local order n, where $n \ll N$, and that all local system matrices are time-invariant. In (3.3), the index i is referred to as *the spatial index*. The spatial index takes the values from *the spatial domain* $\Pi = \{1, \ldots, N\}$.

Like it is explained in Chapter 2, the global state-space model (3.1) originates from the finite-difference discretization of the 2D heat equation with spatially dependent coefficients.

The first step in many classical control and estimation techniques, is to form the so-called *lifted data equations* from the global state-space model (3.1). The lifted data equations are formed by using the following strategy. Starting from $k-p$ and by lifting (3.1), p time steps we obtain:

$$\underline{\mathbf{x}}(k) = \underline{A}^p \underline{\mathbf{x}}(k-p) + R_{p-1}\mathbf{U}_{k-p}^{k-1} \tag{3.4}$$

$$\mathbf{Y}_{k-p}^{k} = O_p \underline{\mathbf{x}}(k-p) + \Gamma_{p-1}\mathbf{U}_{k-p}^{k-1} + \mathbf{N}_{k-p}^{k} \tag{3.5}$$

where

$$\mathbf{U}_{k-p}^{k-1} = \begin{bmatrix} \underline{\mathbf{u}}(k-p) \\ \underline{\mathbf{u}}(k-p+1) \\ \vdots \\ \underline{\mathbf{u}}(k-1) \end{bmatrix}, \quad \mathbf{Y}_{k-p}^{k} = \begin{bmatrix} \underline{\mathbf{y}}(k-p) \\ \underline{\mathbf{y}}(k-p+1) \\ \vdots \\ \underline{\mathbf{y}}(k) \end{bmatrix}, \quad \mathbf{N}_{k-p}^{k} = \begin{bmatrix} \underline{\mathbf{n}}(k-p) \\ \underline{\mathbf{n}}(k-p+1) \\ \vdots \\ \underline{\mathbf{n}}(k) \end{bmatrix}$$

$$O_p = \begin{bmatrix} \underline{C} \\ \underline{CA} \\ \vdots \\ \underline{CA}^p \end{bmatrix}, \quad R_{p-1} = \begin{bmatrix} \underline{A}^{p-1}\underline{B} & \underline{A}^{p-2}\underline{B} & \ldots & \underline{B} \end{bmatrix},$$

$$\Gamma_{p-1} = \begin{bmatrix} 0 & 0 & \ldots & 0 \\ \underline{CB} & 0 & \ldots & 0 \\ \underline{CAB} & \underline{CB} & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ \underline{CA}^{p-1}\underline{B} & \underline{CA}^{p-2}\underline{B} & \ldots & \underline{CB} \end{bmatrix} \tag{3.6}$$

In this this thesis, the matrix $O_p \in \mathbb{R}^{N(p+1)r \times Nn}$ will be called the *global observability matrix*. The matrices $\Gamma_{p-1} \in \mathbb{R}^{Nn \times Npm}$ and $R_{p-1} \in \mathbb{R}^{N(p+1)r \times Npm}$ will be called the *global impulse response* and *global controllability* matrices, respectively. For $p = 2$, the structure of these matrices is illustrated in Fig. 3.2. The lifting technique used to define the lifted system matrices O_p, Γ_p and G_p will be called *the classical lifting technique*.

Throughout this thesis we assume that $p \ll N$. Because of this assumption all the lifted system matrices are sparse. Without loss of generality, we assume that $p \geq n-1$.

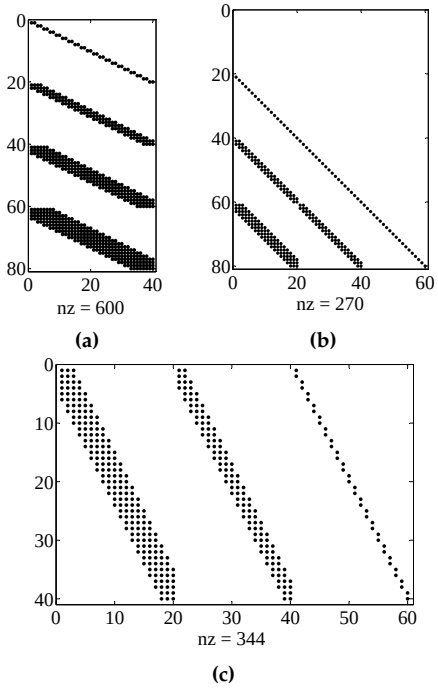


Figure 3.2: Sparsity structure of the lifted system matrices: (a) O_3; (b) Γ_2; (c) R_2 ("nz" denotes the number of non-zero elements).

From Fig. 3.2 we can see that the sparsity patterns of the global system matrices $\underline{A}$, $\underline{B}$ and $\underline{C}$ are not preserved in the lifted system matrices. This is because the classical lifting technique lifts the global inputs and the global outputs over the time domain.

In contrast to this type of lifting, we propose a lifting technique that first lifts the local outputs and inputs over the time domain and then, it lifts these lifted vectors over the spatial domain II. *This lifting method ensures that the sparsity structure of the global state-space model* (3.1) *is preserved in the lifted state-space model.*

To mathematically formulate this lifting technique, the following notation is introduced. The column vector $\mathcal{Y}_{i,k-p}^{k} \in \mathbb{R}^{(p+1)r}$ is defined by lifting local output of $\mathcal{S}_i$ over the discrete-time interval $[k-p, k]$:

$$\mathcal{Y}_{i,k-p}^{k} = \text{col}(\mathbf{y}_i(k-p), \mathbf{y}_i(k-p+1), \ldots, \mathbf{y}_i(k)) \tag{3.7}$$

where the notation $\text{col}(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N)$ denotes a column vector $[\mathbf{z}_1^T \ \mathbf{z}_2^T \ \ldots \ \mathbf{z}_N^T]^T$. In the same manner we define the lifted input vector $\mathcal{U}_{i,k-p}^{k} \in \mathbb{R}^{(p+1)m}$ and the lifted measurement noise vector $\mathcal{N}_{i,k-p}^{k} \in \mathbb{R}^{(p+1)r}$. Next, a column vector $\mathcal{Y}_{k-p}^{k} \in \mathbb{R}^{N(p+1)r}$ is defined as follows:

$$\mathcal{Y}_{k-p}^{k} = \text{col}(\mathcal{Y}_{1,k-p}^{k}, \ldots, \mathcal{Y}_{N,k-p}^{k}) \tag{3.8}$$

In the same manner we define vectors $\mathcal{U}_{k-p}^{k} \in \mathbb{R}^{N(p+1)m}$ and $\mathcal{N}_{k-p}^{k} \in \mathbb{R}^{N(p+1)r}$. It can be easily proved that:

$$\mathcal{Y}_{k-p}^{k} = P_Y \mathbf{Y}_{k-p}^{k}, \quad \mathcal{N}_{k-p}^{k} = P_Y \mathbf{N}_{k-p}^{k}, \quad \mathcal{U}_{k-p}^{k-1} = P_U \mathbf{U}_{k-p}^{k-1} \tag{3.9}$$

where P_Y and P_U are permutation matrices. By multiplying the lifted equation (3.5) from left with P_Y and keeping in mind that permutation matrices are orthogonal, we obtain:

$$\mathcal{Y}_{k-p}^{k} = \mathcal{O}_p \underline{\mathbf{x}}(k-p) + \mathcal{G}_{p-1} \mathcal{U}_{k-p}^{k-1} + \mathcal{N}_{k-p}^{k} \tag{3.10}$$

where the matrices $\mathcal{O}_p \in \mathbb{R}^{N(p+1)r \times Nn}$ and $\mathcal{G}_{p-1} \in \mathbb{R}^{N(p+1)r \times Npm}$ are defined as follows:

$$\mathcal{O}_p = P_Y O_p, \quad \mathcal{G}_{p-1} = P_Y \Gamma_{p-1} P_U^T \tag{3.11}$$

The equation (3.10) will be called *the global data equation*. On the other hand, using the orthogonality of the permutation matrix P_U from (3.4), we have:

$$\underline{\mathbf{x}}(k) = \underline{A}^p \underline{\mathbf{x}}(k-p) + \mathcal{R}_{p-1} \mathcal{U}_{k-p}^{k-1} \tag{3.12}$$

where the matrix $\mathcal{R}_{p-1} \in \mathbb{R}^{Nn \times Npm}$ is defined as follows:

$$\mathcal{R}_{p-1} = R_{p-1} P_U^T \tag{3.13}$$

The matrix $\mathcal{O}_p$ is a sparse, block banded matrix, with the block bandwidth equal to p. Similarly, the matrices $\mathcal{G}_{p-1}$ and $\mathcal{R}_{p-1}$ are sparse, block banded matrices with the block bandwidth equal to $p-1$.

In Fig. 3.3 we illustrate the sparsity patterns of $\mathcal{O}_p$ and $\mathcal{R}_{p-1}$, for $p = 3$ (compare with the structure of the lifted system matrices presented in Fig. 3.2(a) and Fig. 3.2(c)).

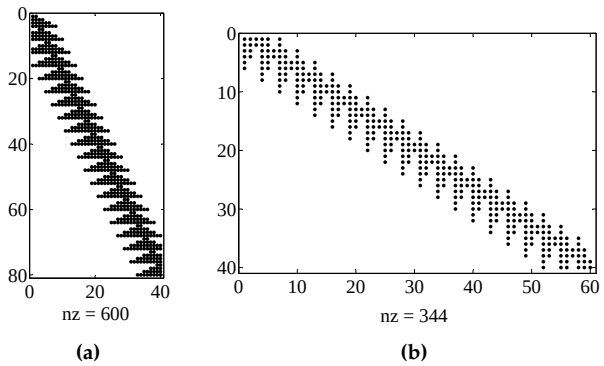


Figure 3.3: Sparsity patterns of lifted system matrices that are formed using the structure preserving lifting technique: (a) $\mathcal{O}_3$; (b) $\mathcal{R}_2$ (″nz″ denotes the number of non-zero elements).

Definition 3.1 *Definition of the structured lifted system matrices*

- *The matrix $\mathcal{O}_p$, defined in* (3.11)*, will be referred to as the structured observability matrix.*

- *The matrix $\mathcal{G}_{p-1}$, defined in* (3.11)*, will be referred to as the structured impulse response matrix.*

- *The matrix $\mathcal{R}_{p-1}$, defined in* (3.13)*, will be referred to as the structured controllability matrix.*

In the sequel, the structure of $\mathcal{O}_p$, $\mathcal{G}_{p-1}$ and $\mathcal{R}_{p-1}$ is explained in more details. The matrices $\mathcal{O}_p$ and $\mathcal{G}_{p-1}$ can be explicitly defined as follows:

$$\mathcal{O}_p = \begin{bmatrix} O_{1,1}^{(p)} \dots O_{1,1+p}^{(p)} \;\; 0 \dots \\ \ddots \\ \dots 0 \; O_{i,i-p}^{(p)} \dots O_{i,i}^{(p)} \dots O_{i,i+p}^{(p)} \; 0 \dots \\ \ddots \\ \dots 0 \;\; O_{N,N-p}^{(p)} \dots O_{N,N}^{(p)} \end{bmatrix}, \tag{3.14}$$

$$\mathcal{G}_{p-1} = \begin{bmatrix} G_{1,1}^{(p)} \dots G_{1,1+p-1}^{(p)} \;\; 0 \dots \\ \ddots \\ \dots 0 \; G_{i,i-p+1}^{(p)} \dots G_{i,i}^{(p)} \dots G_{i,i+p-1}^{(p)} \; 0 \dots \\ \ddots \\ \dots 0 \;\; G_{N,N-p+1}^{(p)} \dots G_{N,N}^{(p)} \end{bmatrix} \tag{3.15}$$

where the matrices $O_{i,j}^{(p)} \in \mathbb{R}^{(p+1)r \times n}$ and $G_{i,j}^{(p)} \in \mathbb{R}^{(p+1)r \times (p+1)m}$ are defined by:

$$\underbrace{\begin{bmatrix} 0 \\ \vdots \\ 0 \\ T_{i,i-p}^{(p)} \end{bmatrix}}_{O_{i,i-p}^{(p)}}, \dots, \underbrace{\begin{bmatrix} 0 \\ T_{i,i-1}^{(1)} \\ \vdots \\ T_{i,i-1}^{(p)} \end{bmatrix}}_{O_{i,i-1}^{(p)}}, \underbrace{\begin{bmatrix} T_{i,i}^{(0)} \\ T_{i,i}^{(1)} \\ \vdots \\ T_{i,i}^{(p)} \end{bmatrix}}_{O_{i,i}^{(p)}}, \underbrace{\begin{bmatrix} 0 \\ T_{i,i+1}^{(1)} \\ \vdots \\ T_{i,i+1}^{(p)} \end{bmatrix}}_{O_{i,i+1}^{(p)}}, \dots, \underbrace{\begin{bmatrix} 0 \\ \vdots \\ 0 \\ T_{i,i+p}^{(p)} \end{bmatrix}}_{O_{i,i+p}^{(p)}} \tag{3.16}$$

$$G_{i,i}^{(p)} = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ H_{i,i}^{(0)} & 0 & 0 & \dots & 0 \\ H_{i,i}^{(1)} & H_{i,i}^{(0)} & 0 & \dots & 0 \\ \ddots & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ H_{i,i}^{(p-1)} & H_{i,i}^{(p-2)} & \dots & H_{i,i}^{(1)} & H_{i,i}^{(0)} \end{bmatrix}, G_{i,i-1}^{(p)} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ H_{i,i-1}^{(1)} & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ H_{i,i-1}^{(p-1)} & \dots & H_{i,i-1}^{(1)} & 0 \end{bmatrix},$$

$$G_{i,i+1}^{(p)} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ H_{i,i+1}^{(1)} & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ H_{i,i+1}^{(p-1)} & \dots & H_{i,i+1}^{(1)} & 0 \end{bmatrix}, \dots, G_{i,i-p+1}^{(p)} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \\ H_{i,i-p+1}^{(p-1)} & 0 & \dots & 0 \end{bmatrix},$$

$$G_{i,i+p-1}^{(p)} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \\ H_{i,i+p-1}^{(p-1)} & 0 & \dots & 0 \end{bmatrix} \tag{3.17}$$

The block elements of $O_{i,j}^{(p)}$ and $G_{i,j}^{(p)}$ are the matrices $T_{i,j}^{(p)} \in \mathbb{R}^{r\times n}$ and $H_{i,j}^{(p)} \in \mathbb{R}^{r\times m}$ defined by:

$$T_{i,j}^{(p)} = CL_{i,j}^{(p)},\ H_{i,j}^{(p)} = CL_{i,j}^{(p)}B \tag{3.18}$$

and where the matrix $L_{i,j}^{(p)} \in \mathbb{R}^{n\times n}$ is (i,j) block of $\underline{A}^p$ (see Remark 3.1). The matrix $H_{i,j}^{(p)}$ will be referred to as *the local impulse response parameter* of the local subsystem $\mathcal{S}_i$.

Remark 3.1 *We have defined the matrices $L_{i,j}^{(p)} \in \mathbb{R}^{n\times n}$ by partitioning $\underline{A}^p$ as follows. Namely, we have divided the rows and columns of $\underline{A}^p$ into N block rows and N block columns, respectively, where each block row (block column) has n rows (columns). This partitioning of $\underline{A}^p$ can be visualized as follows:*

$$\underline{A}^p = \begin{bmatrix} L_{1,1}^{(p)} \dots L_{1,1+p}^{(p)} & 0 \dots & & \\ & \ddots & & \\ \dots 0\ L_{i,i-p}^{(p)} \dots L_{i,i}^{(p)} & \dots L_{i,i+p}^{(p)}\ 0 \dots & & \\ & & \ddots & \\ & \dots 0\ L_{N,N-p}^{(p)} \dots L_{N,N}^{(p)} & & \end{bmatrix}$$

On the other hand, the matrix $\mathcal{R}_{p-1}$ in the equation (3.12) has the following struc-

ture:

$$\mathcal{R}_{p-1} = \begin{bmatrix} R_{1,1}^{(p)} \dots R_{1,1+p-1}^{(p)} & 0 \dots & & \\ & \ddots & & \\ \dots 0 \; R_{i,i-p+1}^{(p)} \dots R_{i,i}^{(p)} & \dots R_{i,i+p-1}^{(p)} & 0 \dots & \\ & & \ddots & \\ & \dots 0 & R_{N,N-p+1}^{(p)} \dots R_{N,N}^{(p)} & \end{bmatrix} \tag{3.19}$$

$$R_{i,i}^{(p)} = \begin{bmatrix} V_{i,i}^{(p-1)} & V_{i,i}^{(p-2)} & \dots & V_{i,i}^{(0)} \end{bmatrix}$$

$$R_{i,i-1}^{(p)} = \begin{bmatrix} V_{i,i-1}^{(p-1)} & V_{i,i-1}^{(p-2)} & \dots & V_{i,i-1}^{(1)} & 0 \end{bmatrix}$$

$$R_{i,i+1}^{(p)} = \begin{bmatrix} V_{i,i+1}^{(p-1)} & V_{i,i+1}^{(p-2)} & \dots & V_{i,i-1}^{(1)} & 0 \end{bmatrix}$$

$$\vdots$$

$$R_{i,i-p+1}^{(p)} = \begin{bmatrix} V_{i,i-p+1}^{(p-1)} & 0 \dots 0 \end{bmatrix},$$

$$R_{i,i+p-1}^{(p)} = \begin{bmatrix} V_{i,i+p-1}^{(p-1)} & 0 \dots 0 \end{bmatrix}$$

$$V_{i,j}^{(p)} = L_{i,j}^{(p)} B_j$$

The i-th block row of the global data equation (3.10) has the following form:

$$\mathcal{Y}_{i,k-p}^{k} = \sum_{j=i-p}^{i+p} O_{i,j}^{(p)} \mathbf{x}_j(k-p) + \sum_{j=i-p+1}^{i+p-1} G_{i,j}^{(p)} \mathcal{U}_{j,k-p}^{k-1} + \mathcal{N}_{i,k-p}^{k-1} \tag{3.20}$$

where the matrices $O_{i,j}^{(p)}$ and $G_{i,j}^{(p)}$ are defined in (3.16) and (3.17), respectively.

The structure preserving lifting technique, that is introduced in this Chapter, can be easily generalized for state-space models with sparse multi-banded system matrices obtained by the finite-difference discretization of 3D Partial Differential Equations (PDEs). Furthermore, the structure preserving lifting technique can be generalized for descriptor state-space models, obtained by approximating PDEs using the finite element method.

3.3 Gramians of large-scale interconnected systems

The finite time observability Gramian of the global system (3.1) is defined by [61; 201]:

$$\mathcal{W} = \sum_{i=0}^{p} \left(\underline{A}^T\right)^i \underline{C}^T \underline{C} \underline{A}^i \tag{3.21}$$

It can be easily shown that:

$$\mathcal{W} = O_p^T O_p \tag{3.22}$$

where O_p is the global observability matrix defined in (3.6). *The finite time controllability Gramian* of the global system (3.1) is defined as follows [61; 201]:

$$\mathcal{Q} = \sum_{i=0}^{p-1} \underline{A}^i \underline{B}\underline{B}^T \left(\underline{A}^T\right)^i \tag{3.23}$$

It can be easily shown that:

$$\mathcal{Q} = R_{p-1} R_{p-1}^T \tag{3.24}$$

where R_{p-1} is the global controllability matrix defined in (3.6). Because the permutation matrices are orthogonal, from (3.11) and (3.22) we have:

$$\mathcal{W} = O_p^T O_p = \mathcal{O}_p^T P_Y P_Y^T \mathcal{O}_p = \mathcal{O}_p^T \mathcal{O}_p \tag{3.25}$$

That is, the finite-time observability Gramian is a sparse banded matrix with the block bandwidth equal to $2p$ (see Remark 3.2). Consequently, the finite-time observability Gramian is denoted as follows:

$$\mathcal{J}_{2p} = \mathcal{W} = \mathcal{O}_p^T \mathcal{O}_p \tag{3.26}$$

where the subscript $2p$ denotes the block bandwidth of $\mathcal{J}_{2p}$. Similarly, from (3.13) and (3.24) we have:

$$\mathcal{Q} = \mathcal{R}_{p-1} \mathcal{R}_{p-1}^T \tag{3.27}$$

That is, the finite-time controllability Gramian is a sparse banded matrix with the block bandwidth equal to 2(p-1).

Remark 3.2 *If a sparse banded matrix X has a (block) bandwidth equal to b, then the matrix $X \cdot X$ has a (block) bandwidth equal to $2b$. Similarly, the (block) bandwidth of $X \cdot X \cdot X$ is $3b$ and etc.*

3.4 Approximate sparse inverses of sparse matrices

The problem of computing sparse approximated inverses of sparse matrices originates from the problem of computing preconditioners for large-scale, sparse systems of linear equations [202]. For example, consider the problem of solving linear system of equations:

$$A\mathbf{x} = \mathbf{y} \tag{3.28}$$

where $A \in \mathbb{R}^{n \times n}$ is a large and sparse matrix. One of the standard methods for solving sparse linear systems is the Conjugate Gradient (CG) method [112; 200]. Starting from an initial guess, the CG method iteratively approximates the true solution of (3.28). By exploiting the sparsity of A, one iteration of the CG method can be computed efficiently. However, the convergence of the CG methods depends

on the spectral properties of A. In particular, if the matrix A is well-conditioned, then the CG method converges in a relatively small number of iterations.
In oder to improve the convergence rate of the CG method, the original system (3.28) is transformed into the following form:

$$AP\mathbf{z} = \mathbf{y}, \quad \mathbf{x} = P\mathbf{z} \tag{3.29}$$

where $P \in \mathbb{R}^{n\times n}$ is a preconditioner matrix. The matrix P should be constructed such that it is sparse and such that the product AP is a relatively good approximation of the identity matrix. This way, the matrix AP is well conditioned, and the CG method converges in a relatively small number of iterations.

There are several approaches for constructing sparse preconditioners [202]. The most natural approach is to chose P as an approximate sparse inverse of A. One of the first approaches that exploits this idea is presented in [203]. In the sequel we briefly summarize this approach. In [203], preconditioner P is constructed as the solution of the following optimization problem [203]:

$$\min_{P} \|AP - I\|_F^2 \tag{3.30}$$

where I is identify matrix. The cost function of (3.30) can be written in the following form:

$$\|AP - I\|_F = \sum_{i=1}^{n} \|A\mathbf{p}_i - \mathbf{l}_i\|_2^2 \tag{3.31}$$

where $\mathbf{p}_i$ is the i^{th} column of P and $\mathbf{l}_i$ is the i^{th} column of the identity matrix I. Because columns $\mathbf{p}_i$ are independent, the optimization problem (3.30) can be separated into n independent least-squares problems:

$$\min_{\mathbf{p}_i} \|A\mathbf{p}_i - \mathbf{l}_i\|_2^2, \quad i = 1, \ldots, n \tag{3.32}$$

The main advantage of solving (3.32) instead of (3.30), is that the optimization problems (3.32) can be solved in parallel. Furthermore, if the matrix P is sparse then each of the columns $\mathbf{p}_i$ has a small number of nonzero elements and consequently, (3.32) can be transformed into n small least-squares problems that can be solved in a computationally efficient manner.
However, the main challenge is how to determine a good sparsity pattern of the approximate inverse P. In [203], this is achieved automatically, by starting with an initial sparsity pattern of P (for example starting from a diagonal sparsity pattern) and augmenting P progressively until certain threshold on the residual norm has been satisfied or a prescribed maximum number of non-zero elements has been reached. Since this approach adaptively computes the "optimal" sparsity pattern of P, in literature it is often referred to as *the adaptive algorithm*. Beside the adaptive algorithm presented in [203], there are also other algorithms for computing sparse approximate inverse preconditioners, see for example [202; 204; 205; 206].

In [207] it has been demonstrated that for matrices originating from approximation of

PDE problems, good a priori sparsity pattern of P can be obtained by computing and sparsifying the powers of A. The advantage of this algorithm over the adaptive algorithms is that no additional calculations are needed to find the optimal sparsity pattern of P. That is, the structure of P in the optimization problem (3.31) is fixed a priori.
A similar idea is used in [208] to chose a priori sparsity pattern of P. It has been shown that by using the characteristic polynomial and by computing the Neumann series of A or A^TA a good a priori sparsity pattern of an approximate inverse can be obtained. Further discussion on how to chose the "best" a priori sparsity pattern of approximate inverse preconditioners is available in [209].
In this thesis we are not interested in developing new algorithms for computing sparse approximate inverse preconditioners. *However, the results of [207; 208] are important because they imply that the approximate sparse inverses of sparse matrices can be constructed by using matrix polynomials*. This idea also has its roots in the graph theory for predicting structure in sparse matrix computations [210].

In [161], the Chebyshev matrix polynomials have been used to compute approximate sparse inverses of sparse banded matrices. Furthermore, in [163] the Newton iteration (also known as Newton-Schultz iteration) has been used for the computation of approximate inverses. In this thesis, we will use these two algorithms to approximate inverses and pseudo-inverses of sparse matrices.*We will develop a rigorous analysis on the approximation accuracy of the Chebyshev matrix polynomials*. Furthermore, we will discuss the dropping strategy that helps us to additionally "sparsify" approximate inverses.

Before we present these algorithms, we illustrate one remarkable property of inverses of banded, positive definite matrices. In [161; 162] it has been proved that inverses of banded, positive definite matrices belong to a class of off-diagonally decaying matrices. Off-diagonal decaying matrices are characterized by a property that the absolute values of the off-diagonal elements decay as they are further away from the main diagonal.

3.4.1 Off-diagonally decaying matrices

In this Section we prove that the inverse of the (regularized) finite-time observability Gramian is an off-diagonally decaying matrix. Furthermore, we prove that the rate of the off-diagonal decay depends on the condition number of the finite-time observability Gramian. The theoretical results presented in this Section and in the remainder of this Chapter can be straightforwardly generalized for the inverses of the finite-time controllability Gramian and the impulse response matrix.

Consider the global state-space model (3.1)-(3.2). The matrix $\underline{A}$ is a block banded matrix and the global system matrices $\underline{C}$ and $\underline{B}$ are block diagonal. Like it has been illustrated in Fig. 3.2, the global observability matrix:

$$O_3 = \left[\underline{C}^T \quad (\underline{CA})^T \quad \left(\underline{CA}^2\right)^T \quad \left(\underline{CA}^3\right)^T\right]^T$$

is not a banded matrix. However, we have explained that there exists a permutation matrix P_Y that transforms O_3 into the structured observability matrix (for

more details see beginning of this chapter):

$$\mathcal{O}_3 = P_Y O_3 \tag{3.33}$$

that is a sparse banded matrix (sparsity pattern of $\mathcal{O}_3$ is illustrated in Fig. 3.3(a)). Because $\mathcal{O}_3$ is a banded matrix, the observability Gramian $\mathcal{O}_3^T\mathcal{O}_3$ is also a banded matrix. The structure of $\mathcal{O}_3^T\mathcal{O}_3$ is illustrated in Fig. 3.4(a).

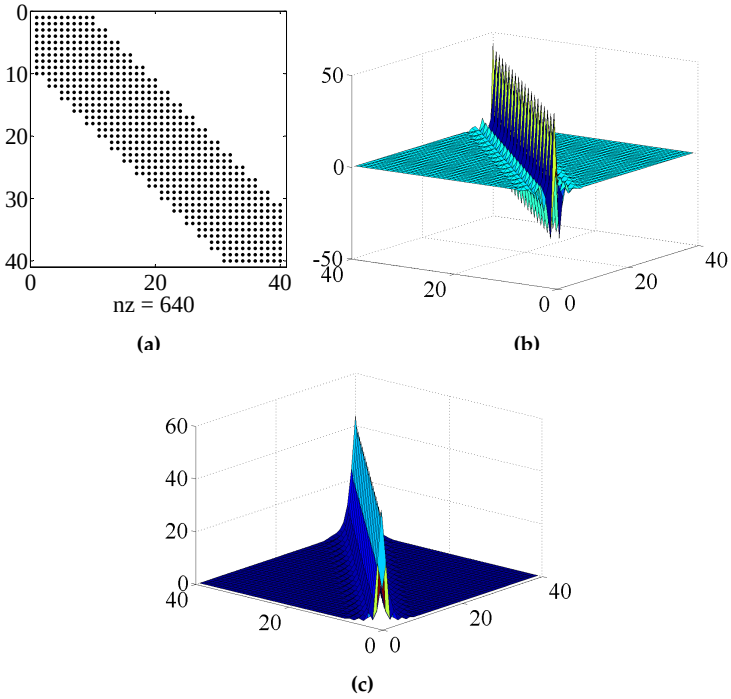


Figure 3.4: Off-diagonal decay of the inverse of the observability Gramian $\mathcal{O}_3^T\mathcal{O}_3$ ("nz" denotes the number of nonzero elements); (a) The sparsity pattern of $\mathcal{O}_3^T\mathcal{O}_3$; (b) The surface plot of $\left(\mathcal{O}_3^T\mathcal{O}_3\right)^{-1}$ generated by a MATLAB function surf$(\cdot)$; (c) The surface plot of absolute value of $\left(\mathcal{O}_3^T\mathcal{O}_3\right)^{-1}$.

If we assume that $\mathcal{O}_3$ has a full column rank, then the matrix $\mathcal{O}_3^T\mathcal{O}_3$ is a symmetric positive definite matrix. Consequently, its inverse is an off-diagonal decaying matrix [162]. The surface plot of $\left(\mathcal{O}_3^T\mathcal{O}_3\right)^{-1}$ is shown in Fig. 3.4 (b). In Fig. 3.4(c) we illustrate the surface plot of the absolute values of the elements of $\left(\mathcal{O}_3^T\mathcal{O}_3\right)^{-1}$. From Fig.3.4(c) we see that the absolute values of the off-diagonal elements of $\left(\mathcal{O}_3^T\mathcal{O}_3\right)^{-1}$ decay in an exponential manner as they are further away from the main diagonal.

In [162] it has been shown that the off-diagonal decay rate of $\left(\mathcal{O}_3^T\mathcal{O}_3\right)^{-1}$ is faster

if the condition number of $\mathcal{O}_3^T\mathcal{O}_3$ is smaller. Thus, inverses of well-conditioned matrices have the fast off-diagonal decay rate. In the sequel, this is explained in more details.
The observability index of the global system (3.1) is defined as follows.

Definition 3.2 *The observability index*[3]*, of the observable global system* (3.1)*, is the smallest integer ν, such that the global observability matrix*

$$O_\nu = \begin{bmatrix}\underline{C}^T & (\underline{CA})^T & \dots & (\underline{CA}^\nu)^T\end{bmatrix}^T \tag{3.34}$$

has rank equal to nN (that is, full column rank).

The following lemma tells us that if the global system is observable then the matrix $\mathcal{O}_p$ has full column rank.

Lemma 3.1 *Assume that $p \geq \nu$, where ν is the observability index of the global system. Then,*

$$rank(\mathcal{O}_p) = nN \tag{3.35}$$

Proof. We have shown that $\mathcal{O}_p = P_Y O_p$, where P_Y is a permutation matrix. Because this permutation does not change the rank of O_p, we have: rank $(\mathcal{O}_p)$ = rank (O_p). Now, let $p \geq \nu$, where ν is the observability index of the global system. This implies: rank(O_p) = rank$(\mathcal{O}_p) = nN$. □

The regularized observability Gramian $\mathcal{F}_{2p} \in \mathbb{R}^{Nn\times Nn}$ is defined by:

$$\mathcal{F}_{2p} = \mu I + \mathcal{J}_{2p} = \mu I + \mathcal{O}_p^T\mathcal{O}_p \tag{3.36}$$

where $\mu \geq 0$ is a regularization parameter.
The regularization parameter μ will play an important role in the algorithms that will be developed in the subsequent chapters. We will show that with a proper selection of μ it is possible to estimate the global state and moreover, to identify the global system with the computational and memory complexities that scale linearly with the number of local subsystems N.

In the next definition we define the class of matrices that are characterized by the off-diagonally decaying property.

Definition 3.3 *[161] We say that an $nN \times nN$ matrix $Z = [z_{i,j}]$ is an exponentially off-diagonally decaying matrix if there are constants $c, \lambda \in \mathbb{R}$, $c > 0$ and $\lambda \in (0, 1)$, such that:*

$$|z_{i,j}| \leq c\lambda^{|i-j|}, \;\; 0 < \lambda < 1 \tag{3.37}$$

[3]In the literature, see for example [211], the observability index is usually defined as the smallest integer ν for which the matrix $O_\nu = \begin{bmatrix}\underline{C}^T & (\underline{CA})^T & \dots & (\underline{CA}^{\nu-1})^T\end{bmatrix}^T$ has rank equal to nN. For convenience, in this chapter we slightly change this definition.

for all $i, j = 1, \ldots, nN$.

For the sequel, we introduce the following notation:

- The condition number of the matrix $\mathcal{J}_{2p}$ will be denoted by χ.
- The condition number of the matrix $\mathcal{F}_{2p}$ will be denoted by κ.

In the following theorem we prove that $\mathcal{F}_{2p}^{-1}$ is an exponentially off-diagonally decaying matrix and we will show how its decay rate depends on μ (similar results hold for $\mathcal{J}_{2p}^{-1}$).

Lemma 3.2 *Assume that* $\mu > 0$ *or* $p \geq \nu$*, where* ν *is the observability index of the global system. Furthermore, let the maximal and minimal singular values of* $\mathcal{O}_p$ *be denoted by* σ_1 *and* σ_{nN}*, respectively. Then,*

1. $\mathcal{F}_{2p}^{-1}$ *is an exponentially off-diagonally decaying matrix, with the exponential off-diagonal decay rate specified by:*

$$\lambda = \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{1/\theta}, \; c = \left\|\mathcal{F}_{2p}^{-1}\right\|_2 \max\left\{1, \frac{(1+\sqrt{\kappa})^2}{2\kappa}\right\} \tag{3.38}$$

where θ *is the bandwidth*[4] *of* $\mathcal{F}_{2p}$ *(*θ *is proportional to the product* pn*) and*

$$\kappa = \frac{\sigma_1^2 + \mu}{\sigma_{Nn}^2 + \mu} = 1 + \frac{\sigma_1^2 - \sigma_{nN}^2}{\sigma_{Nn}^2 + \mu} \tag{3.39}$$

2. *The parameters* c *and* λ*, defined in* (3.38)*, are decreasing functions of* μ.

Proof 1.) Because $p \geq n-1$, the matrix $\mathcal{O}_p \in \mathbb{R}^{N(p+1)r \times Nn}$ is square or "tall". The Singular Value Decomposition (SVD) of $\mathcal{O}_p$ is $\mathcal{O}_p = U\Sigma V^T$, where U and V are unitary matrices and Σ is a matrix defined as follows:

$$\Sigma = \begin{bmatrix}\Sigma_1 & 0\end{bmatrix}^T \tag{3.40}$$

where the matrix Σ_1 is a diagonal matrix of singular values $\sigma_1, \ldots, \sigma_{nN}$, $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{nN}$. The matrix $\mathcal{F}_{2p}$ can now be expressed as follows:

$$\mathcal{F}_{2p} = V(\Sigma_1^2 + \mu I)V^T \tag{3.41}$$

From (3.41) we have that the condition number of $\mathcal{F}_{2p}$ is given by (3.39). Since $\mu > 0$ or $p \geq \nu$, the matrix $\mathcal{F}_{2p}$ is a symmetric positive definite matrix. Because of this, from Theorem 2.4 stated in [162][5], we have that $\mathcal{F}_{2p}^{-1}$ is the exponentially

[4]The bandwidth θ is defined by $\theta = m/2$, where m is a constant in Eq. (2.6) in [162].

[5]The results of Theorem 2.4 in [162] are valid for both finite and infinite matrices. Similar results are stated in Proposition 2.2 of [162] and in [161].

off-diagonally decaying matrix with an exponential off-diagonal decay specified by (3.38).

2.) From (3.39), we conclude that κ is a monotonically decreasing function of μ. The partial derivative of λ with respect to κ is:

$$\frac{\partial \lambda}{\partial \kappa} = \frac{1}{\theta}\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{(1-\theta)/\theta} \frac{1}{\sqrt{\kappa}(\sqrt{\kappa}+1)^2} \tag{3.42}$$

Because $\kappa \geq 1$, the sign of (3.42) is positive or equal to zero, which implies that the function λ is an increasing function of κ. Because κ is a decreasing function of μ, we conclude that λ is a decreasing function of μ. In order to analyze the monotonicity of c, we first need to analyze the monotonicity when $c = \left\|\mathcal{F}_{2p}^{-1}\right\|_2$. Namely, it is easy to show that:

$$\left\|\mathcal{F}_{2p}^{-1}\right\|_2 = \frac{1}{\sigma_{Nn}^2 + \mu} \tag{3.43}$$

From (3.43), we conclude that c is a decreasing function of μ. Next, we analyze the monotonicity of c when:

$$c = \left\|\mathcal{F}_{2p}^{-1}\right\|_2 \frac{(1+\sqrt{\kappa})^2}{2\kappa} \tag{3.44}$$

Substituting (3.39) and (3.43) in (3.44), we obtain:

$$c = \frac{\left(\sqrt{\sigma_1^2+\mu} + \sqrt{\sigma_{Nn}^2+\mu}\right)^2}{2\left(\sigma_1^2+\mu\right)\left(\sigma_{Nn}^2+\mu\right)} \tag{3.45}$$

The derivative of c with respect to μ is:

$$\begin{aligned}\frac{dc}{d\mu} = &\left(\sqrt{\sigma_1^2+\mu} + \sqrt{\sigma_{Nn}^2+\mu}\right)^2 \times \\ &\times \frac{\left(\sqrt{\sigma_1^2+\mu}\sqrt{\sigma_{Nn}^2+\mu} - \left(\sigma_1^2+\sigma_{Nn}^2+2\mu\right)\right)}{2\left(\sigma_1^2+\mu\right)^2\left(\sigma_{Nn}^2+\mu\right)^2}\end{aligned} \tag{3.46}$$

Because

$$\sqrt{\sigma_1^2+\mu}\sqrt{\sigma_{Nn}^2+\mu} - \left(\sigma_1^2+\sigma_{Nn}^2+2\mu\right) \leq 0 \tag{3.47}$$

we conclude that (3.46) is negative or equal to zero. This implies that the function (3.45) is a decreasing function of μ. This completes the proof. □

3.4.2 Chebyshev method for computing approximate inverses of sparse matrices

We will first present the Chebyshev approximation method and then we will analyze the approximation errors. The Chebyshev approximation method will be explained on an example of approximating the inverse of regularized (finite-time) observability Gramian $\mathcal{F}_{2p}$. This method can be straightforwardly generalized for approximation of inverses of finite-time controllability Gramian and impulse response matrix. Throughout the remainder of this Chapter we will assume that the conditions of Lemma 3.2 are satisfied.

For the sequel, we will define the constants a and b as follows:

$$a = \lambda_{min}(\mathcal{F}_{2p}), \qquad b = \lambda_{max}(\mathcal{F}_{2p}) \tag{3.48}$$

where $\lambda_{min}(.)$ and $\lambda_{max}(.)$ denote the maximal and the minimal eigenvalues. Due to the fact that the matrix $\mathcal{F}_{2p}$ is a symmetric positive-definite matrix we have:

$$a = \sigma_{Nn}^2 + \mu, \qquad b = \sigma_1^2 + \mu \tag{3.49}$$

where σ_1 and σ_{Nn} are maximal and minimal singular values of $\mathcal{O}_p$. From (3.39) and (3.49) we have that the condition number of $\mathcal{F}_{2p}$ is:

$$\kappa = \frac{b}{a} \tag{3.50}$$

Because $\mathcal{F}_{2p}$ is a sparse banded matrix, the constants a and b can be computed with $O(N)$ computational complexity using the ARPACK software package [212; 213] or using MATLAB sparse matrix computations toolbox. The interval $[a, b]$ is the smallest interval on the real axis that contains the spectrum of $\mathcal{F}_{2p}$. In order to approximate $\mathcal{F}_{2p}^{-1}$ using the Chebyshev series approximation method, we define an inverse function $f(q)$:

$$f(q) = q^{-1} \tag{3.51}$$

where $q \in [a, b]$. For the approximation purpose, the argument of the function $f(q)$ has to belong to an interval $[-1, 1]$. In order to achieve this we shift and scale q as follows:

$$w = \frac{2}{b-a}q - \frac{a+b}{b-a} \tag{3.52}$$

Because $q \in [a, b]$, we have that $w \in [-1, 1]$. From (3.52) we have:

$$q = \frac{1}{2}\left((b-a)w + (a+b)\right) \tag{3.53}$$

By substituting (3.53) in (3.51) we define the function $g(w)$

$$g(w) = f(q(w)) = \frac{2}{(b-a)w + (a+b)} \tag{3.54}$$

For convenience we express the function $g(w)$ in the following form:

$$g(w) = \frac{2}{(a-b)(v-w)} \tag{3.55}$$

where

$$v = \frac{a+b}{a-b} = \frac{1+\kappa}{1-\kappa} \tag{3.56}$$

Because $w \in [-1, 1]$, we can approximate $g(w)$ using the Chebyshev series expansion. Due to the fact that $g(w) = f(q)$ the Chebyshev approximation of $g(w)$ is at the same time an approximation of $f(q)$. We will first approximate the function $\frac{1}{v-w}$. The approximation of $g(w)$ is obtained by multiplying the approximation of $\frac{1}{v-w}$ with the constant $\frac{2}{a-b}$. The Chebyshev series approximation of $\frac{1}{v-w}$ is defined as follows [214] :

$$\overline{g}_t(w) = \frac{c_0}{2} + \sum_{k=1}^{t} c_k T_k(w) \tag{3.57}$$

where $c_k \in \mathbb{R}$, $k = 0, \ldots, t$ and $T_k(w)$ is a Chebsyhev polynomial of the first kind. In [215] it has been shown that the coefficients c_k of the function $\frac{1}{v-w}$ can be computed explicitly (see Remark 3.3):

$$c_k = \frac{2}{-\sqrt{v^2-1}} \frac{1}{h^k}, \; h = v - \sqrt{(v^2-1)}, \; v \notin [-1, 1] \tag{3.58}$$

The Chebyshev polynomials $T_k(w)$ are defined recursively as follows [214]:

$$\begin{aligned} &T_0(w) = 1, \;\; T_1(w) = w, \\ &T_{k+1}(w) = 2wT_k(w) - T_{k-1}(w), \;\; k = 1, \ldots, t \end{aligned} \tag{3.59}$$

Remark 3.3 *The explicit formula for the coefficients of the Chebyshev series expansion of the inverse function has been derived in [215] (see equation 10.). This formula has been derived for a general case, when the argument of the inverse function is a complex polynomial. Like it has been stated in [215], in order to define the parameter h and coefficient c_k, the branch of $(v^2-1)^{1/2}$ (where v is in a general case a complex number) should be chosen such that $|h| > 1$. Because in our case v is negative and real, in order to ensure that $|h| > 1$, in* (3.58) *we have taken a negative branch of $(v^2-1)^{1/2}$. This is why in* (3.58) *the sign in front of $(v^2-1)^{1/2}$ differs from the sign of the corresponding quantity in equation 10. of [215].*

The polynomial $g_t(w)$, representing an approximation of $g(w)$, is defined by:

$$g_t(w) = \frac{2}{(a-b)} \overline{g}_t(w) = d_0 + d_1 T_1(w) + d_2 T_2(w) + \ldots + d_t T_t(w) \tag{3.60}$$

where $d_i \in \mathbb{R}$, $i = 0, \ldots, t$. From (3.57) and (3.60) it follows that:

$$d_0 = \frac{1}{(a-b)}c_0, \quad d_i = \frac{2}{(a-b)}c_i, i = 1, \ldots, t \tag{3.61}$$

By substituting (3.52) in (3.60), we obtain:

$$f_t(q) = g_t(w(q)) = e_0 + e_1 q + e_2 q^2 + \ldots + e_t q^t \tag{3.62}$$

where $e_i \in \mathbb{R}$, $i = 0, \ldots, t$. The polynomial $f_t(q)$ is an approximation of $f(q) = q^{-1}$. By substituting q with $\mathcal{F}_{2p}$ and w with $\mathcal{F}^*_{2p}$ in (3.51),(3.52), (3.53), (3.54) and (3.60) we obtain:

$$f(\mathcal{F}_{2p}) = \mathcal{F}^{-1}_{2p} \tag{3.63}$$

$$\mathcal{F}^*_{2p} = \frac{2}{b-a}\mathcal{F}_{2p} - \frac{a+b}{b-a}I \tag{3.64}$$

$$\mathcal{F}_{2p} = 0.5\left((b-a)\mathcal{F}^*_{2p} + (a+b)I\right) \tag{3.65}$$

$$g(\mathcal{F}^*_{2p}) = 2\left((b-a)\mathcal{F}^*_{2p} + (a+b)I\right)^{-1} \tag{3.66}$$

$$g_t(\mathcal{F}^*_{2p}) = \sum_{k=0}^{t} d_k T_k\left(\mathcal{F}^*_{2p}\right) \tag{3.67}$$

Because $f_t(\mathcal{F}_{2p}) = g_t(\mathcal{F}^*_{2p})$, for the Chebyshev approximation of $\mathcal{F}^{-1}_{2p}$ we can either use $f_t(\mathcal{F}_{2p})$ or $g_t(\mathcal{F}^*_{2p})$. However, since $f_t(\mathcal{F}_{2p})$ explicitly depends on $\mathcal{F}_{2p}$, throughout the reminder of the chapter we will use this matrix polynomial as the Chebyshev approximation of $\mathcal{F}^{-1}_{2p}$. Since the block bandwidth of $f_t(\mathcal{F}_{2p})$ is equal to $2tp$, we will denote this matrix polynomial by $\mathcal{E}_{2tp} \in \mathbb{R}^{Nn \times Nn}$, that is:

$$\begin{aligned} &\mathcal{F}^{-1}_{2p} \approx \mathcal{E}_{2tp}, \\ &\mathcal{E}_{2tp} = f_t(\mathcal{F}_{2p}) = e_0 I + e_1 \mathcal{F}_{2p} + e_2 \mathcal{F}^2_{2p} + \ldots + e_t \mathcal{F}^t_{2p} \end{aligned} \tag{3.68}$$

Because $p \ll N$, $\mathcal{F}_{2p}$ is a sparse banded matrix. This means that for $t \ll N$ the computational complexity of computing $\mathcal{E}_{2tp}$ is $O(N)$. Furthermore, for $t \ll N$ the matrix $\mathcal{E}_{2tp}$ can be stored using $O(N)$ memory locations. *So the key to the linear complexity computation of $\mathcal{E}_{2tp}$ is to choose $t \ll N$.* However, intuitively it should be clear that by decreasing t, the accuracy of the Chebyshev series approximation degrades. This is why in the following theorem we quantify the accuracy of approximating $\mathcal{F}^{-1}_{2p}$ by $\mathcal{E}_{2tp}$.

Theorem 3.4 *Consider the matrix function $\mathcal{F}^{-1}_{2p}$ and its Chebyshev series approximation $\mathcal{E}_{2tp}$, defined in* (3.68)*.Then,*

1. *The accuracy of the Chebyshev approximation is quantified by:*

$$\left\|\mathcal{F}^{-1}_{2p} - \mathcal{E}_{2tp}\right\|_2 \leq l(\mu, t) \tag{3.69}$$

$$l(\mu, t) = \frac{1}{a(\mu)}\left(1 - \frac{1}{\sqrt{\kappa(\mu)}}\right)\frac{1}{|h(\mu)|^t} \tag{3.70}$$

$$|h(\mu)| = \frac{\sqrt{\kappa(\mu)}+1}{\sqrt{\kappa(\mu)}-1} \tag{3.71}$$

where $\kappa(\mu)$ is defined in (3.39) *and $a(\mu)$ is defined in* (3.49).

2. *The function $l(\mu, t)$ is a decreasing function of μ and t.*

Proof. 1.) Using the Chebyshev truncation theorem (see [216], page 47, Theorem 6) we have:

$$\begin{aligned}\max_{w\in[-1,1]} |g_t(w) - g(w)| &= \max_{w\in[-1,1]} |\frac{2}{a-b}\overline{g}_t(w) - \frac{2}{(a-b)}\frac{1}{v-w}| \\ &= \frac{2}{b-a}\max_{w\in[-1,1]} |\overline{g}_t(w) - \frac{1}{v-w}| \leq \frac{2}{b-a}\sum_{k=t+1}^{\infty}|c_k|\end{aligned} \tag{3.72}$$

where the coefficients c_k are defined in (3.58). We can write:

$$\sum_{k=t+1}^{\infty}|c_k| = \sum_{k=0}^{\infty}|c_k| - \sum_{k=0}^{t}|c_k| \tag{3.73}$$

Next, we have:

$$\sum_{k=0}^{\infty}|c_k| = \frac{2}{\sqrt{v^2-1}}\left(1 + \frac{1}{|h|} + \frac{1}{|h|^2} + \dots\right) = \frac{2}{\sqrt{v^2-1}}\frac{|h|}{|h|-1} \tag{3.74}$$

Similarly, we have:

$$\sum_{k=0}^{t}|c_k| = \frac{2}{\sqrt{v^2-1}}\frac{|h|^{t+1}-1}{|h|^t(|h|-1)} \tag{3.75}$$

Substituting (3.74) and (3.75) in (3.73) we have:

$$\sum_{k=t+1}^{\infty}|c_k| = \frac{2}{\sqrt{v^2-1}}\frac{1}{|h|^t(|h|-1)} \tag{3.76}$$

It can be easily proved that:

$$\sqrt{v^2-1} = \frac{2\sqrt{ab}}{b-a} \tag{3.77}$$

$$|h| = \frac{(\sqrt{a}+\sqrt{b})^2}{b-a} = \frac{\sqrt{b}+\sqrt{a}}{\sqrt{b}-\sqrt{a}} = \frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1} \tag{3.78}$$

$$|h| - 1 = \frac{2a+2\sqrt{ab}}{b-a} \tag{3.79}$$

From (3.76)-(3.79) we have:

$$\frac{2}{b-a}\sum_{k=t+1}^{\infty}|c_k| = \frac{1}{a}\left(1-\frac{1}{\sqrt{\kappa}}\right)\frac{1}{|h|^t} = l(\mu,t) \tag{3.80}$$

Because $\mathcal{F}_{2p}$ is positive definite and because the polynomial $f_t(q)$ is real, we have [162]:

$$\left\|\mathcal{F}_{2p}^{-1} - f_t(\mathcal{F}_{2p})\right\|_2 = \max_{q\in\sigma(\mathcal{F}_{2p})}|\frac{1}{q} - f_t(q)| \tag{3.81}$$

where $f_t(q)$ is defined in (3.62) and $\sigma(\mathcal{F}_{2p})$ denotes the spectrum of $\mathcal{F}_{2p}$. Further, because

$$\begin{array}{l}\mathcal{F}_{2p}^{-1} = f(\mathcal{F}_{2p}) = g(\mathcal{F}_{2p}^*)\\ f_t(\mathcal{F}_{2p}) = g_t(\mathcal{F}_{2p}^*)\end{array} \tag{3.82}$$

we have:

$$\left\|\mathcal{F}_{2p}^{-1} - f_t(\mathcal{F}_{2p})\right\|_2 = \left\|g(\mathcal{F}_{2p}^*) - g_t(\mathcal{F}_{2p}^*)\right\|_2 \tag{3.83}$$

From (3.81) and (3.83) we have:

$$\left\|g(\mathcal{F}_{2p}^*) - g_t(\mathcal{F}_{2p}^*)\right\|_2 = \max_{q\in\sigma(\mathcal{F}_{2p})}|\frac{1}{q} - f_t(q)| \tag{3.84}$$

As it has been explained before, when $q \in \sigma(\mathcal{F}_{2p})$, w takes the values from $[-1,1]$. Because of this and because of the fact that $g(w) = f(q) = \frac{1}{q}$ and $f_t(q) = g_t(w)$, we have:

$$\max_{q\in\sigma(\mathcal{F}_{2p})}|\frac{1}{q} - f_t(q)| \leq \max_{w\in[-1,1]}|g(w) - g_t(w)| \tag{3.85}$$

From (3.83), (3.84) and (3.85) it can be concluded:

$$\left\|\mathcal{F}_{2p}^{-1} - f_t(\mathcal{F}_{2p})\right\|_2 = \left\|g(\mathcal{F}_{2p}^*) - g_t(\mathcal{F}_{2p}^*)\right\|_2 \leq \max_{w\in[-1,1]}|g(w) - g_t(w)| \tag{3.86}$$

From (3.72), (3.80) and (3.86) we obtain (3.69).

2.) Because $|h| > 1$ it is obvious that the function $l(\mu,t)$ is a decreasing function of t. Now in order to prove that the function $l(\mu,t)$ is a decreasing function of μ we will represent it as follows:

$$l(\mu,t) = l_1(\mu)l_2(\mu,t) \tag{3.87}$$

$$l_1(\mu) = \frac{1}{a}\left(1-\frac{1}{\sqrt{\kappa}}\right),\ l_2(\mu,t) = \frac{1}{|h|^t} \tag{3.88}$$

We have $\partial l/\partial\mu = (\partial l_1/\partial\mu)l_2 + l_1(\partial l_2/\partial\mu)$. Since l_1 and l_2 are positive functions, l is a decreasing function ($\frac{\partial l}{\partial\mu} \leq 0$) if both l_1 and l_2 are decreasing functions of μ

($\frac{\partial l_1}{\partial \mu} \leq 0$ and $\frac{\partial l_2}{\partial \mu} \leq 0$). It is easy to prove that $\frac{dl_1}{d\mu} < 0$, which means the function l_1 is a decreasing function of μ. On the other hand, we have:

$$l_2 = \frac{(\sqrt{\kappa}-1)^t}{(\sqrt{\kappa}+1)^t} \tag{3.89}$$

Using the same arguments that were used in proof of Theorem 3.4, we can prove that (3.89) is a decreasing function of μ. This completes the proof. □

On the basis of Theorem 3.4 we draw the following conclusions:

- Under the assumption that σ_1 and σ_{Nn} do not change significantly as N increases, from Theorem 3.4, it follows that the accuracy of the Chebyshev approximation is practically independent from N. This means that for extremely large N (for example in the order of 10^7), we can approximate $\mathcal{F}_{2p}^{-1}$ with a relatively good accuracy, using relatively small t (for example in the order of 10).
- If $\mathcal{J}_{2p}$ is well-conditioned ($\mathcal{J}_{2p}$ is defined in (3.36)) and σ_{Nn} is not close to 0, then for any μ we can always find $t \ll N$ for which the accuracy of the Chebyshev approximation of $\mathcal{F}_{2p}^{-1}$ is relatively good. To show this, let $\mu = 0$ and consider the parameter l defined in (3.70). From (3.39) we have that $\kappa = \chi$, where $\chi = \sigma_1^2/\sigma_{Nn}^2$ is a condition number of the matrix $\mathcal{J}_{2p}$. In this case, $|h|$ becomes:

  $$|h| = \frac{\sqrt{\chi}+1}{\sqrt{\chi}-1} \tag{3.90}$$

  When $\chi \to 1$, we have that $|h| \to \infty$, and $1/|h|^t$ approaches 0 for any t. If σ_{Nn} is not close to 0, then a is not close to 0 and $1/a$ is not large. This implies that when $\chi \to 1$ and σ_{Nn} is not close to 0, the parameter $l(\mu, t)$ approaches zero, for any t. If $\mu > 0$ the accuracy is even better, since $1/|h|^t$ and $l(\mu, t)$ are decreasing functions of μ (see the proof of Theorem 3.4). That is, any positive value of μ additionally increases accuracy of the Chebyshev approximation.
- If $\mathcal{J}_{2p}$ is badly conditioned, then for $t \ll N$ we can always find (not so large) μ for which the accuracy of the Chebyshev series approximation of $\mathcal{F}_{2p}^{-1}$ is relatively good. To show this, let us assume that $\mathcal{J}_{2p}$ is badly conditioned (χ is large). If $\mu = 0$, then $|h|$, given by (3.90), is close to 1 and $1/|h|^t$ is close to 1 for small t. This further implies that $l(\mu, t)$ is relatively large for small t. However, $l(\mu, t)$ is a decreasing function of μ (see the proof of Theorem 3.4). This means that we can always find (not so large) μ and $t \ll N$ for which $l(\mu, t)$ is small. This also follows from the results of Lemma 3.2.

Algorithm 3.1 *$O(N)$ computational complexity approximation of $\mathcal{F}_{2p}^{-1}$*
For $p \ll N$, and for a given μ, the approximation $\mathcal{E}_{2tp}$ of $\mathcal{F}_{2p}^{-1}$ is calculated by performing the following steps:

1. *Compute the constants a and b, defined in* (3.48) *using the ARPACK package [212; 213] or MATLAB sparse matrix computations toolbox.*

2. *Transform the matrix $\mathcal{F}_{2p}$ into the matrix $\mathcal{F}_{2p}^*$ using* (3.64).

3. *Choose $t \ll N$. Compute the coefficients $\{c_0, c_1, \ldots, c_t\}$ using* (3.58). *Using* (3.61), *compute the coefficients $\{d_0, d_1, \ldots, d_t\}$.*

4. *Compute $\mathcal{E}_{2tp} = g_t(\mathcal{F}_{2p}^*)$ using* (3.67).

3.4.3 Newton iteration

The Newton iteration (also known as the Newton-Schulz iteration) for approximating $\mathcal{F}_{2p}^{-1}$ is defined by [217; 218; 219]:

$$X_{k+1} = X_k \left(2I - \mathcal{F}_{2p} X_k\right) \tag{3.91}$$

The initial guess for the iteration (3.91) is usually chosen as follows [220]:

$$X_0 = \alpha \mathcal{F}_{2p} \tag{3.92}$$

where α is a sufficiently small number (see also Remark 3.5). There are several possibilities to chose α. In this thesis, we chose $\alpha = 2/(a^2 + b^2)$ [220].
Let the matrix E_k be defined as follows:

$$E_k = I - \mathcal{F}_{2p} X_k \tag{3.93}$$

At the k^{th} iteration, the residual can be computed by [218]:

$$\epsilon_k = \|E_k\|_2 = \|I - \mathcal{F}_{2p} X_k\|_2 \tag{3.94}$$

On the basis of this residual a stopping criteria for the Newton iteration can be defined. In general, the Newton iteration has a quadratic convergence rate. Namely, the matrix E_{k+1} can be expressed as follows:

$$\begin{aligned} E_{k+1} =& I - \mathcal{F}_{2p} X_{k+1} = I - \mathcal{F}_{2p} X_k \left(2I - \mathcal{F}_{2p} X_k\right) \\ &= \left(I - \mathcal{F}_{2p} X_k\right)\left(I - \mathcal{F}_{2p} X_k\right) = E_k E_k \end{aligned} \tag{3.95}$$

which implies

$$\epsilon_{k+1} \leq \epsilon_k^2 \tag{3.96}$$

If $\|E_0\|_2 < 1$, then from (3.96) it can be concluded that the Newton iteration has a quadratic convergence rate. An estimate of the number of operations to achieve the accuracy $\left\|\mathcal{F}_{2p}^{-1} - X_k\right\|_2 / \left\|\mathcal{F}_{2p}^{-1}\right\|_2 \leq \psi$ is given by [218]:

$$\log_2\left(\kappa^2 + 1\right) + \log_2 \ln \frac{1}{\psi} \tag{3.97}$$

where κ is the condition number of $\mathcal{F}_{2p}$. Similarly to the Chebyshev approximation method, the number of operations of the Newton iteration to achieve the

accuracy ψ depends on the condition number of $\mathcal{F}_{2p}$. If κ is smaller than the number of oeprations is smaller and vice-versa. Thus, if $\mathcal{F}_{2p}$ is well-conditioned, then its inverse can be approximated in a computationally efficient manner.

Remark 3.5 *The Newton iteration and Chebyshev approximation method can be elegantly combined to obtain very fast inversion algorithms. Namely, the initial guess for the Newton iteration can be obtained by computing the approximate inverse using Chebyshev approximation method.*

3.4.4 The dropping strategies

Because $p \ll N$, the matrix $\mathcal{F}_{2p}$ is a sparse banded matrix. This implies that for relatively small i the Chebyshev matrix polynomials $T_i(\mathcal{F}_{2p})$ are sparse, banded matrices. However, the bandwidth of $T_{i+1}(\mathcal{F}_{2p})$ is larger than the bandwidth of $T_i(\mathcal{F}_{2p})$ (see Remark 3.2). This means that for a large approximation order t, the fill-in (the number of non-zero elements of a matrix) of an approximate inverse will increase significantly. Consequently, both computation and memory complexity of computing $\mathcal{E}_{2tp}$ will significantly increase.

One way to overcome this problem is to use *the dropping strategies* [217; 221]. There are two types of dropping strategies. *The first one is the truncation strategy [217; 221].* The truncation strategy restricts the bandwidth of the Chebyshev matrix polynomials by setting to zero all the elements that are outside some prescribed bandwidth. This truncation operator, denoted by $\mathcal{L}$, can be defined entry-wise by:

$$\left(\mathcal{L}\left(Y\right)\right)_{i,j} = \begin{cases} Y_{i,j} & ,|i-j| \leq \beta \\ 0 & ,|i-j| > \beta \end{cases} \tag{3.98}$$

where $Y = [Y_{i,j}]$ is an arbitrary matrix, $\left(\mathcal{L}\left(Y\right)\right)_{i,j}$ is an (i,j) element of $\mathcal{L}\left(Y\right)$ and β is the prescribed truncation bandwidth. Applying the truncation operator to the Chebyshev matrix polynomials, we obtain:

$$\begin{aligned} &T_0 = I, \;\; T_1 = \mathcal{L}\left(\mathcal{F}_{2p}^*\right), \\ &T_{k+1} = \mathcal{L}\left(2\mathcal{F}_{2p}^* T_k\right) - T_{k-1}, \; k \geq 1 \end{aligned} \tag{3.99}$$

This way, we can significantly speed up the computation of an approximate inverse using the Chebyshev approximation method. Furthermore, we need less memory locations to store the approximate inverse.
The truncation bandwidth β can be determined using the results of Lemma 3.2. Namely, using (3.38) we can calculate the parameters λ and c. These parameters tell us how fast is the off-diagonal decay of $\mathcal{F}_{2p}^{-1}$. Consequently, we can determine the bandwidth outside which the elements of $\mathcal{F}_{2p}^{-1}$ can be neglected.

The truncation strategy can be also used to decrease the computational complexity the Newton iteration. Applying the truncation operator to the Newton iteration,

we define the truncated iteration:

$$\begin{aligned} X_0 &= \mathcal{L}\left(\alpha \mathcal{F}_{2p}\right) \\ X_{k+1} &= \mathcal{L}\left(X_k\left(2I - \mathcal{F}_{2p} X_k\right)\right), \quad k = 1, 2, \ldots \end{aligned} \tag{3.100}$$

In [217] it has been proved that the Newton iteration is robust with respect to the errors introduced by the truncation operator.

The second dropping strategy sparsifies matrices by replacing small elements with zeros. The sparsification operator, denoted by $\mathcal{B}$, can be defined entry-wise by:

$$\left(\mathcal{B}\left(Y\right)\right)_{i,j} = \begin{cases} Y_{i,j} & , |Y_{i,j}| \geq \phi \\ 0 & , |Y_{i,j}| < \phi \end{cases} \tag{3.101}$$

where ϕ is the sparsification tolerance. Applying the sparsification operator to the Chebyshev method and Newton iteration, we can obtain iterations that are similar to iterations (3.99) and (3.100). The fastest approximation algorithms can be obtained by combining the truncation and sparsification strategies.

In chapters 4, 6 and 7 we will use the Chebyshev approximation method and Newton iteration to approximate inverses of banded matrices. Furthermore, in these chapters we will demonstrate significant computational savings that can be obtained by using the dropping strategies.

In this chapter we proved that inverses of sparse, banded matrices can be approximated by sparse banded matrices. However, as we mentioned previously, the Chebyshev approximation method and Newton iteration can be applied to much broader class of sparse matrices. In the sequel this will be illustrated with numerical examples. Namely, we show that the finite-time observability Gramian of the global state-space model (2.65)-(2.66) *(the discretized 3D heat equation) is a sparse multi-banded matrix. Furthermore, we show that its inverse can be approximated by a sparse multi-banded matrix.*

3.5 Approximation of inverses of multi-banded matrices

Consider the state-space model (2.65)-(2.66) that is obtained by discretizing the 3D heat equation (2.15) using the finite difference method. The parameters of the model are: $L = 0.01$, $h = 5$, $N = 29$, $M = 29$, and $P = 3$. We assume that the plate is made of the BK7 material. We use the lifting window $p = 4$. The sparsity pattern of the structured observability matrix is shown in Fig. 3.5.

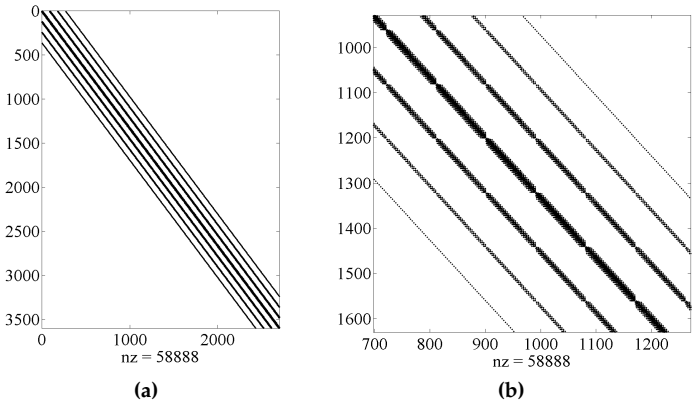


Figure 3.5: (a) Structured observability matrix of the state-space model (2.65)-(2.66); (b) A segment of the structured observability matrix.

In Fig. 3.6, the sparsity patterns of the structured controllability and impulse response matrices are illustrated.

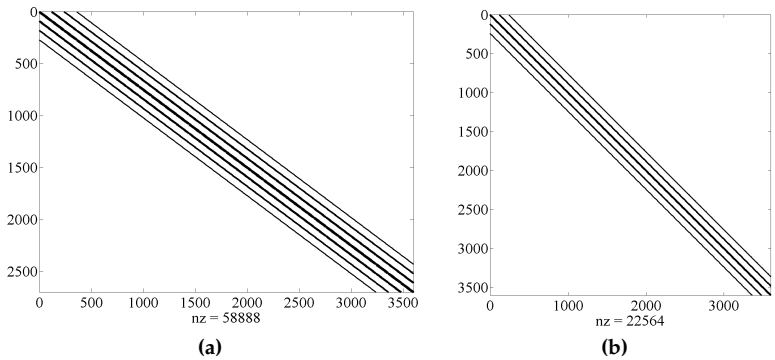


Figure 3.6: Sparsity patterns of the structured controllability and impulse response matrices of the state-space model (2.65)-(2.66) (a) Structured controllability matrix; (b) Structured impulse response matrix.

From Figs. 3.5 and 3.6 we clearly see that the lifted system matrices inherit the multi-banded structure of the global system matrix $\underline{A}$ (see Fig. 3.7(a)). The matrix $\underline{A}$ has a smaller number of nonzero elements than the lifted system matrices. This is because the lifted system matrices are constructed using the powers $\underline{A}^i$, $i = 1, \ldots, p$. The sparsity patterns of the matrix $\underline{A}$ and its powers are illustrated in Fig. 3.7.

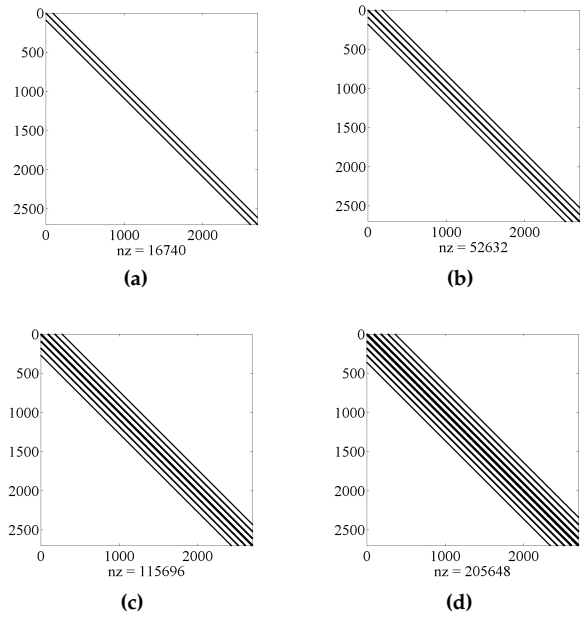


Figure 3.7: Sparsity patterns of the powers of the system matrix $\underline{A}$ of the state-space model (2.65)-(2.66) (a) $\underline{A}$; (b) $\underline{A}^2$; (c) $\underline{A}^3$; (d) $\underline{A}^4$.

The sparsity pattern of the finite-time observability Gramian, $\mathcal{J}_8 = \mathcal{O}_4^T \mathcal{O}_4$, is illustrated in Fig. 3.8.

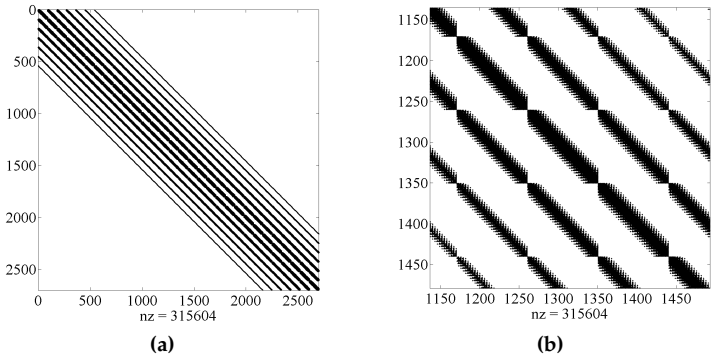
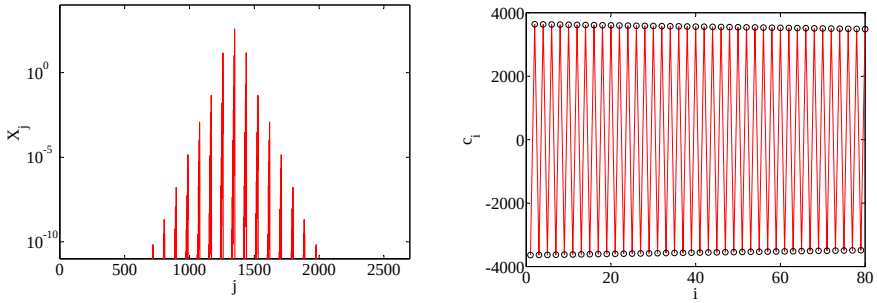


Figure 3.8: (a) Finite-time observability Gramian; (b) A segment of the finite-time observability Gramian.

The condition number of $\mathcal{J}_8$ is $\chi = 1.3245 \times 10^7$. Its minimal and maximal singular values are respectively: 2.6116×10^{-7} and 3.46. Let the inverse of $\mathcal{J}_8$ be denoted by $\mathcal{D}$. The absolute values of the elements of an arbitrary row of $\mathcal{D}$ are

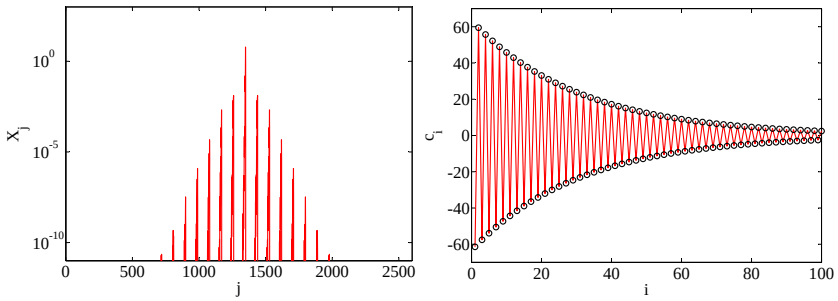
shown in Fig. 3.9(a). This figure suggests that $\mathcal{D}$ can be approximated by a sparse multi-banded matrix. The first 80 Chebyshev coefficients c_i (defined in (3.58)) are shown in Fig. 3.9(b). From Fig. 3.9(b) we see that the Chebyshev coefficients are slowly decaying. This is a numerical confirmation of the results of Theorem 3.4. Namely, because the condition number of $\mathcal{J}_8$ is large and the minimal singular value is small, the approximation order t that gives relatively good approximation accuracy is relatively large. This is why the Chebyshev coefficients in Fig. 3.9(b) are slowly decaying.



(a) **(b)**

Figure 3.9: (a) The absolute values of the elements of an arbitrary row of $\mathcal{D}$; (b) The Chebyshev coefficients c_i.

We chose a regularization parameter $\mu = 0.001$ and we compute a regularized, finite-time observability Gramian $\mathcal{F}_8 = \mathcal{J}_8 + \mu I$. The condition number of $\mathcal{F}_8$ is $\kappa = 3.78 \times 10^3$. Its minimal and maximal singular values are respectively: 9.2×10^{-4} and 3.46. The absolute elements of an arbitrary row of $\mathcal{F}_8^{-1}$ are shown in Fig. 3.10(a). The Chebyshev coefficients are shown in Fig. 3.10(b).



(a) **(b)**

Figure 3.10: (a) The absolute values of the elements of an arbitrary row of $\mathcal{F}_8^{-1}$; (b) The Chebyshev coefficients c_i.

The decay rate of the Chebyshev coefficients, calculated for the matrix $\mathcal{F}_8$, is faster

than the decay rate of the coefficients calculated for $\mathcal{J}_8$. This is because the condition number of $\mathcal{F}_8$ is smaller than the condition number of $\mathcal{J}_8$.

Next, we approximate $\mathcal{F}_8^{-1}$. The approximation error is measured by computing $e = \left\|\mathcal{F}_8^{-1} - \mathcal{E}\right\|_2$, where the matrix $\mathcal{E}$ is computed using the Chebyshev approximation method that is combined with the dropping strategies. More precisely, the approximation $\mathcal{E}$ is computed using the truncated iteration (3.99). Once the matrix $\mathcal{E}$ is computed, the sparsification operator is applied to $\mathcal{E}$ (that is, the sparsification operator is not used during the Chebyshev iteration). The sparsity patterns of the matrix $\mathcal{F}_8$ and its approximate inverses, calculated for several values of the truncation bandwidth β and the sparsification tolerance ϕ, are shown in Fig. 3.11

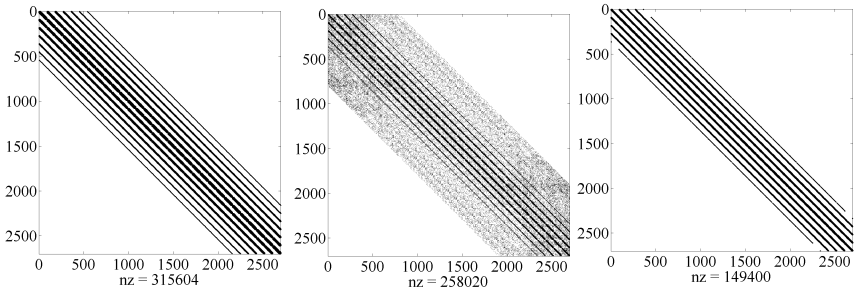


Figure 3.11: The sparsity patterns of $\mathcal{F}_8$ and its approximate inverse $\mathcal{E}$. (a) $\mathcal{F}_8$; (b) $\mathcal{E}$, $\beta = 800$, $\phi = 0.00001$, e=0.0099 ; (c) $\mathcal{E}$, $\beta = 800$, $\phi = 0.00005$, e=0.0103.

Finally, we approximate $\mathcal{F}_8^{-1}$ using the Newton iteration. Here we will be mainly interested how the truncation operator influence the convergence rate of the Newton iteration. In chapters 4, 6 and 7 we will illustrate the computational complexity of the Newton iteration. The convergence of the approximation error $\epsilon_k = \|E_k\|_2$ (defined in (3.94)), for several values of the truncation bandwidth β, is shown in Fig 3.12.

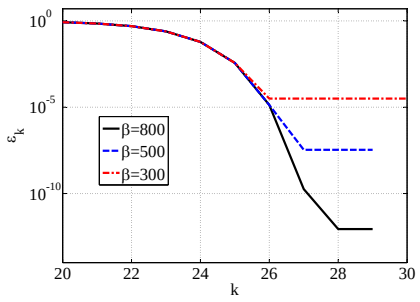


Figure 3.12: The convergence of the Newton iteration for several values of the truncation bandwidth β.

From Fig. 3.12 we see that the Newton iteration is robust with respect to the errors introduced by the truncation operator. We see that as the truncation bandwidth β decreases, the steady-state error increases. Furthermore, we see that the errors introduced by the truncation operator do not affect so much the convergence rate of the Newton iteration.

4 CHAPTER

Moving horizon estimation algorithms

Using the approximation methods summarized in Chapter 3, in this Chapter we develop Moving Horizon Estimation (MHE) methods for large-scale interconnected systems. First, we develop a linear computational complexity MHE method and a distributed MHE method for the state-space models in the standard form. The distributed method estimates the local state using only local input-output data. Furthermore, the distributed MHE method has a simple analytic form and it does not rely on consensus algorithms. Secondly, we develop a MHE method for large-scale systems in the descriptor state-space form.

4.1 Moving horizon estimation algorithms for state-space models in the standard form

The Moving Horizon Estimation (MHE) strategy determines the state of a dynamical system as the solution of an optimization problem that consists of input-output data over a moving time horizon [147; 148; 222; 223; 224]. Despite the fact that moving horizon state estimation of linear, low-dimensional systems has been extensively studied in literature, see for example [147; 148; 222; 223; 224], moving horizon state estimation of large-scale interconnected systems is still an open problem. Namely, there are two main problems that hinder the application of moving horizon estimation techniques for large-scale interconnected systems.

First of all, the design of the parameters (matrices) of the MHE methods proposed in [147; 148; 222; 223; 224] is a computationally challenging problem. Namely, in order to design the parameters of the above mentioned MHE methods, we need to perform sequences of basic matrix operations ($+-, \times, ^{-1}$) on lifted system matrices of a large-scale interconnected system. Furthermore, we need to compute

the spectral norm or the spectral radius of the matrices that result from these basic matrix operations. In the case of interconnected systems that are described by sparse system matrices, we can exploit the sparsity of the lifted system matrices to compute their inverses with $O(N^2)$ complexity [112], where N is the number of local subsystems. However, the problem lies in the fact that these inverse matrices are dense [60]. Namely, after we have inverted all matrices, the sparsity of the design problem is lost, and computational cost of the subsequent computations is larger than $O(N^2)$. Furthermore, for large N we cannot store dense matrices in a computer memory because we need $O(N^2)$ memory locations.

Secondly, the MHE methods proposed in [147; 148; 222; 223; 224], assume that the input-output data of all local subsystems are transmitted to one centralized computing unit, where calculations are performed. In the cases in which a large-number of local sensors collect measurement data of local subsystems, the transmission of all sensor data to one centralized computing unit requires a large amount of energy and communication [164; 225; 226]. In such situations, moving horizon state estimation should be performed in the decentralized/distributed manner on a network of local computing units and sensors that communicate locally.

In [131], the decentralized version of the MHE method (originally proposed in [148]) is developed. However, this decentralized MHE method requires "all-to-all" communication between the local computing units. In the case of a large number of sensors and local computing units, "all-to-all" communication might not be possible because it requires a large amount of energy and communication. Furthermore, in order to perform the stability analysis and to design the parameters of this decentralized MHE method, we need to compute the spectral norm or the spectral radius of large-dimensional, dense matrices. This problem has not been addressed in [131]. For all these reasons, the decentralized MHE method, proposed in [131], is restricted to the class of interconnected systems with a small or medium number of local subsystems.

In this section we present computationally efficient centralized and distributed moving horizon MHE methods for large-scale interconnected systems, that are described by sparse banded or sparse multi-banded system matrices. Like we have shown in Chapter 2, interconnected systems, with sparse (multi) banded system matrices originate from discretization of PDEs [60]. Both of the proposed MHE methods are developed by approximating a solution of the MHE problem using the Chebyshev approximation method. By exploiting the sparsity of this approximate solution we derive a centralized MHE method, which computational complexity and memory storage requirements scale linearly with the number of local subsystems of an interconnected system. Furthermore, on the basis of the approximate solution of the MHE problem, we develop a novel, distributed MHE method. This distributed MHE method estimates the state of a local subsystem using only local input-output data. In contrast to the existing distributed algorithms for the state estimation of large-scale systems, the proposed distributed MHE method is not relying on the consensus algorithms and has a simple analytic form. We have studied the stability of the proposed MHE methods and we have performed numerical simulations that confirm our theoretical results.

The proposed distributed MHE method can also be seen as a distributed param-

eter estimation or distributed optimization method. Unlike existing algorithms for distributed parameter estimation [164; 165] or for distributed optimization [166; 167; 168], that compute an estimate iteratively (using consensus-subgradient or diffusion based algorithms), the proposed distributed MHE method determines an estimate at the fixed time instant in a closed form. Namely, the proposed MHE method estimates a state of a local subsystem by computing a linear combination of the local input-output data of local subsystems that are in its neighborhood.

The remainder of this section is organized as follows: In Section 4.1.1 we present a problem formulation. In Section 4.1.2 we present the centralized MHE method. In Section 4.1.3 we present the distributed MHE method and in Section 4.1.4 we present simulation results. The conclusion is drawn in Section 4.1.5.

4.1.1 Problem formulation

For the sake of presentation clarity, we assume that the interconnected system whose state we want to estimate consists of a string of local subsystems [77; 78; 79]. The MHE strategies proposed in this thesis can be generalized for interconnected systems with a more general interconnection topologies (see Remark 4.4). The MHE strategy will be derived for the global system (3.1) that is not affected by the input vector:

$$\mathcal{S}\left\{\begin{array}{rl} \underline{\mathbf{x}}(k+1) & = \underline{A}\underline{\mathbf{x}}(k) \\ \underline{\mathbf{y}}(k) & = \underline{C}\underline{\mathbf{x}}(k)+\underline{\mathbf{n}}(k) \end{array}\right. \tag{4.1}$$

where the system matrices have the following structure:

$$\underline{A}=\begin{bmatrix} A_{1,1} & E_{1,2} & & \\ E_{2,1} & A_{2,2} & E_{2,3} & \\ & \ddots & & \\ & E_{N-1,N-2} & A_{N-1,N-1} & E_{N-1,N} \\ & & E_{N,N-1} & A_{N,N} \end{bmatrix},$$

$$\underline{C}=\begin{bmatrix} C_1 & & \\ & \ddots & \\ & & C_N \end{bmatrix},\ \underline{\mathbf{y}}(k)=\begin{bmatrix} \mathbf{y}_1(k) \\ \vdots \\ \mathbf{y}_N(k) \end{bmatrix},\underline{\mathbf{x}}(k)=\begin{bmatrix} \mathbf{x}_1(k) \\ \vdots \\ \mathbf{x}_N(k) \end{bmatrix} \tag{4.2}$$

Like it has been explained in Chapter 3, the global system $\mathcal{S}$ consists of the interconnection of N *local subsystems* $\mathcal{S}_i$:

$$\mathcal{S}_i\left\{\begin{array}{rl} \mathbf{x}_i(k+1) & = A_{i,i}\mathbf{x}_i(k)+E_{i,i-1}\mathbf{x}_{i-1}(k)+E_{i,i+1}\mathbf{x}_{i+1}(k) \\ \mathbf{y}_i(k) & = C_i\mathbf{x}_i(k)+\mathbf{n}_i(k) \end{array}\right. \tag{4.3}$$

The interconnection structure of local subsystems is illustrated in Fig.4.1.

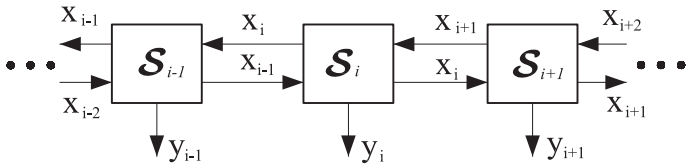


Figure 4.1: The interconnection structure of the local subsystems of the global system (4.1).

In this chapter we adopt the MHE problem formulation of [148]. *The state estimation horizon* will be denoted by a non-negative integer p, where $p \ll N$. Without the loss of generality we assume that $p \geq n-1$. For $k-p \geq 0$, we introduce the following notation.

- $\hat{\mathbf{x}}_i(k-p|k), \ldots, \hat{\mathbf{x}}_i(k|k)$ denote the estimates of the local states $\mathbf{x}_i(k-p), \ldots, \mathbf{x}_i(k)$, respectively, made at a time instant k. The estimates $\hat{\mathbf{x}}_i(k-p|k), \ldots, \hat{\mathbf{x}}_i(k|k)$ will be called *the local moving horizon estimates* and they should satisfy:

$$\begin{aligned} &\hat{\mathbf{x}}_i(j+1|k) = A_{i,i}\hat{\mathbf{x}}_i(j|k) + E_{i,i-1}\hat{\mathbf{x}}_{i-1}(j|k) + E_{i,i+1}\hat{\mathbf{x}}_{i+1}(j|k) \\ &j = k-p, \ldots, k-1 \end{aligned} \tag{4.4}$$

- *The global moving horizon estimates* are denoted by $\underline{\hat{\mathbf{x}}}(k-p|k), \ldots, \underline{\hat{\mathbf{x}}}(k|k)$. We have $\underline{\hat{\mathbf{x}}}(k-j|k) = \mathrm{col}(\hat{\mathbf{x}}_1(k-j|k), \ldots, \hat{\mathbf{x}}_N(k-j|k)),\ j = 0, 1, \ldots, p$. The global moving horizon estimates $\underline{\hat{\mathbf{x}}}(k-p+1|k), \ldots, \underline{\hat{\mathbf{x}}}(k|k)$, are determined from $\underline{\hat{\mathbf{x}}}(k-p|k)$ by propagating the equation:

$$\underline{\hat{\mathbf{x}}}(j+1|k) = \underline{A}\underline{\hat{\mathbf{x}}}(j|k), \quad j = k-p, \ldots, k-1 \tag{4.5}$$

- The local moving horizon estimate $\hat{\mathbf{x}}_i(k-p|k-1)$ is at the same time a prediction of the local state $\mathbf{x}_i(k-p)$ made at the time instant $k-1$. Consequently, $\hat{\mathbf{x}}_i(k-p|k-1)$ will be called *the local prediction* and similarly to (4.4) it is determined by:

$$\begin{aligned} &\hat{\mathbf{x}}_i(k-p|k-1) = A_{i,i}\hat{\mathbf{x}}_i(k-p-1|k-1) \\ &+ E_{i,i-1}\hat{\mathbf{x}}_{i-1}(k-p-1|k-1) + E_{i,i+1}\hat{\mathbf{x}}_{i+1}(k-p-1|k-1) \end{aligned} \tag{4.6}$$

- The vector $\underline{\hat{\mathbf{x}}}(k-p|k-1) = \mathrm{col}(\hat{\mathbf{x}}_1(k-p|k-1), \ldots, \hat{\mathbf{x}}_N(k-p|k-1))$ will be called *the global prediction* and similarly to (4.5) it is determined by $\underline{\hat{\mathbf{x}}}(k-p|k-1) = \underline{A}\underline{\hat{\mathbf{x}}}(k-p-1|k-1)$. The initial value of the global prediction $\underline{\hat{\mathbf{x}}}(0|p-1)$ is given a priori.

- A local lifted output vector $\mathcal{Y}^k_{i,k-p} \in \mathbb{R}^{(p+1)r}$ is defined in (3.7). In the same manner we define the lifted measurement noise vector $\mathcal{N}^k_{i,k-p} \in \mathbb{R}^{(p+1)r}$.

- A global lifted output vector $\mathcal{Y}_{k-p}^{k} \in \mathbb{R}^{N(p+1)r}$ is defined in (3.8). In the same manner we define $\mathcal{N}_{k-p}^{k} \in \mathbb{R}^{N(p+1)r}$.

With each local subsystem $\mathcal{S}_i$, we associate *the local cost function*:

$$J_i^k(\hat{\mathbf{x}}_{i-p}(k-p|k), \ldots, \hat{\mathbf{x}}_{i+p}(k-p|k)) = \\ \mu \left\| \hat{\mathbf{x}}_i(k-p|k) - \hat{\mathbf{x}}_i(k-p|k-1) \right\|_2^2 + \sum_{j=k-p}^{k} \left\| \mathbf{y}_i(j) - C_i \hat{\mathbf{x}}_i(j|k) \right\|_2^2 \tag{4.7}$$

where $\mu \geq 0$. The local cost function (4.7) consists of two parts. The first term on the right-hand side of (4.7) penalizes the difference between the local MHE estimate and the local prediction. The second term in (4.7) penalizes the difference between the measured local outputs and the local outputs computed on the basis of the local moving horizon estimates. By increasing μ we put more emphasis on the model and less emphasize on the data, and vice-versa. Because $p \ll N$ the local cost function (4.7) depends only on the few local states. *The global cost function* is defined as the sum of the local cost functions:

$$J^k(\hat{\underline{\mathbf{x}}}(k-p|k)) = \sum_{i=1}^{N} J_i^k \tag{4.8}$$

The global moving horizon estimate is determined by solving the following optimization problem:

$$\min_{\hat{\underline{\mathbf{x}}}(k-p|k)} J^k(\hat{\underline{\mathbf{x}}}(k-p|k)) \tag{4.9}$$

Problem Description 4.1 *The centralized MHE problem*

Let the vector $\mathcal{Y}_{k-p}^{k}$ and the global prediction $\hat{\underline{\mathbf{x}}}(k-p|k-1)$ be given at a time instant $k \geq p$. Then, at each time instant k approximate the solution $\hat{\underline{\mathbf{x}}}(k-p|k)$ of the optimization problem (4.9) *with $O(N)$ computational complexity and $O(N)$ memory storage requirements.*

In Section 4.1.2 we derive an approximate solution of the centralized MHE problem.

Remark 4.1 *Once $\hat{\underline{\mathbf{x}}}(k-p|k)$ has been computed, the global moving horizon estimates $\hat{\underline{\mathbf{x}}}(k-p+1|k), \ldots, \hat{\underline{\mathbf{x}}}(k|k)$ are computed by propagating the equation* (4.5). *The global moving horizon estimate $\hat{\underline{\mathbf{x}}}(k-p+1|k)$ is at the same time a global prediction of the estimate for the time instant $k+1$.*

In order to state the distributed MHE algorithm, we first define the architecture of the computational network. We assume that for each local subsystem $\mathcal{S}_i$ there exists a local computing unit, denoted by $\mathcal{T}_i$, that at each time instant $k \geq p$ memorizes $\mathcal{Y}_{i,k-p}^{k}$ and that computes and memorizes the local moving horizon estimates

$\hat{\mathbf{x}}_i(k-p|k), \ldots, \hat{\mathbf{x}}_i(k|k)$. At each time instant k, the local computing unit $\mathcal{T}_i$ is able to transfer these memorized quantities to other local computing units in its neighborhood.

Definition 4.1 *The set of computing units:*

$$R_s(\mathcal{T}_i) = \{\mathcal{T}_{i-s}, \ldots, \mathcal{T}_{i-1}, \mathcal{T}_{i+1}, \ldots, \mathcal{T}_{i+s}\} \tag{4.10}$$

where $s \ll N$, is called the neighborhood of the local computing unit $\mathcal{T}_i$.

A local computing unit, associated with a local subsystem, is illustrated in Fig. 4.2. Everything is now prepared to state the distributed MHE problem.

Problem Description 4.2 *The distributed MHE problem*

Let the parameter s be selected by the user. Let us assume that at each time instant $k \geq p$, each local computing unit $\mathcal{T}_i$ receives $\hat{\mathbf{x}}_j(k-p|k-1)$ and $\mathcal{Y}^k_{j,k-p}$ from every local computing unit $\mathcal{T}_j$ that belongs to $R_s(\mathcal{T}_i)$ (this communication is illustrated in Fig. 4.3). Using this data and using $\hat{\mathbf{x}}_i(k-p|k-1)$ and $\mathcal{Y}^k_{i,k-p}$, local computing unit $\mathcal{T}_i$ should approximate the local moving horizon estimate $\hat{\mathbf{x}}_i(k-p|k)$, that is a component of the solution $\underline{\hat{\mathbf{x}}}(k-p|k)$ of the optimization problem (4.9).

In Section 4.1.3 we derive an approximate solution of the Distributed MHE problem.

Remark 4.2 *Once all local moving horizon estimates $\hat{\mathbf{x}}_1(k-p|k), \ldots, \hat{\mathbf{x}}_N(k-p|k)$ are computed by the corresponding local computing units, the local computing unit $\mathcal{T}_i$ exchanges the moving horizon estimates with local computing units $\mathcal{T}_{i-1}$ and $\mathcal{T}_{i+1}$ in order to compute $\hat{\mathbf{x}}_i(k-p+1|k), \ldots, \hat{\mathbf{x}}_i(k|k)$ using equation* (4.4).

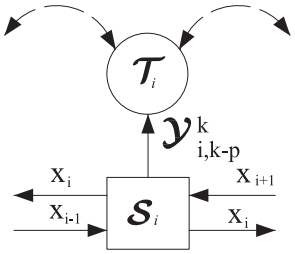


Figure 4.2: The local computing unit $\mathcal{T}_i$ associated with the local subsystem $\mathcal{S}_i$. The dashed lines indicate that the local computing unit can communicate with other local computing units.

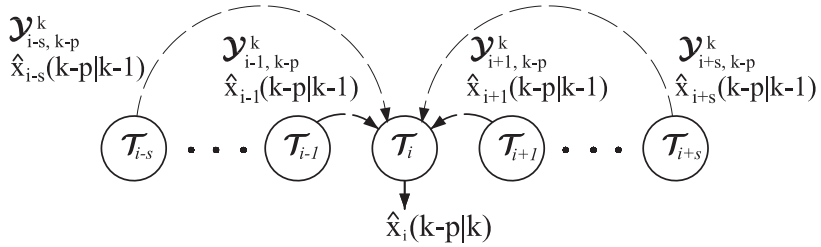


Figure 4.3: The communication that is necessary to compute the local MHE estimate $\hat{\mathbf{x}}_i(k-p|k)$ using the distributed MHE method.

4.1.2 Approximate sparse solution of the centralized MHE problem

In the first part of this section we present an exact solution of the centralized MHE problem (Problem 4.1). In the second part, using the Chebyshev approximation method, we develop its approximate, sparse solution. By exploiting its sparsity we implement this solution with $O(N)$ complexity. Finally, we study stability of the approximate MHE solution and we give some guidelines for its tuning.

Exact solution of the centralized MHE problem

Starting from a time instant $k-p$ and by lifting the output equation of the local subsystem $\mathcal{S}_i$ (4.3), p time steps, we obtain the local data equation (for more details see Chapter 3):

$$\mathcal{Y}^k_{i,k-p} = \sum_{j=i-p}^{i+p} O^{(p)}_{i,j}\mathbf{x}_j(k-p) + \mathcal{N}^k_{i,k-p} \tag{4.11}$$

where $\mathcal{Y}^k_{i,k-p}$ is defined in (3.7) and the matrices $O^{(p)}_{i,j}$, $j = i-p,\ldots,i+p$, are defined in (3.16). By lifting (4.11) from $i=1$ to $i=N$ we obtain the global data equation:

$$\mathcal{Y}^k_{k-p} = \mathcal{O}_p\underline{\mathbf{x}}(k-p) + \mathcal{N}^k_{k-p} \tag{4.12}$$

where $\mathcal{Y}^k_{k-p}$ is defined in (3.8) and the matrix $\mathcal{O}_p$ is defined in (3.14).

Using (3.16), (3.18) and (4.4), we express the local cost function (4.7) in the following form:

$$J^k_i(\hat{\mathbf{x}}_{i-p}(k-p|k),\ldots,\hat{\mathbf{x}}_{i+p}(k-p|k)) =$$
$$\mu||\hat{\mathbf{x}}_i(k-p|k) - \hat{\mathbf{x}}_i(k-p|k-1)||^2_2 + ||\mathcal{Y}^k_{i,k-p} - \sum_{j=i-p}^{i+p} O^{(p)}_{i,j}\hat{\mathbf{x}}_j(k-p|k)||^2_2 \tag{4.13}$$

By summing the local costs (4.13), for $i = 1, \ldots, N$, the global cost function becomes:

$$\begin{aligned} J^k(\hat{\underline{\mathbf{x}}}(k-p|k)) =& \mu \left\| \hat{\underline{\mathbf{x}}}(k-p|k) - \hat{\underline{\mathbf{x}}}(k-p|k-1) \right\|_2^2 \\ &+ \left\| \mathcal{Y}_{k-p}^k - \mathcal{O}_p \hat{\underline{\mathbf{x}}}(k-p|k) \right\|_2^2 \end{aligned} \tag{4.14}$$

Everything is prepared to state the solution of the optimization problem (4.9).

Theorem 4.3 *Suppose that $\mu > 0$ or $p \geq \nu$, where ν is the observability index of the global system. Then, the solution of the centralized MHE problem is given by:*

$$\hat{\underline{\mathbf{x}}}(k-p|k) = \mathcal{F}_{2p}^{-1}(\mu \hat{\underline{\mathbf{x}}}(k-p|k-1) + \mathcal{O}_p^T \mathcal{Y}_{k-p}^k) \tag{4.15}$$

where the matrix $\mathcal{F}_{2p}$ is defined in (3.36)*.*

Proof. Since $p \geq \nu$, from Lemma 3.1 we have that $\text{rank}(\mathcal{O}_p) = nN$. This implies that $\mathcal{F}_{2p}$ is positive definite. On the other hand, if $\mu > 0$, then independently from the value of p we have that $\mathcal{F}_{2p}$ is positive definite. Since $\mathcal{F}_{2p}$ is positive definite, by minimizing the cost function (4.14) with respect to $\hat{\underline{\mathbf{x}}}(k-p|k)$, we obtain (4.15). □

Linear computational-complexity centralized MHE

Throughout the remainder of the section we assume that the conditions of Lemma 3.2 are satisfied. From Lemma 3.2 it follows that if $\mathcal{F}_{2p}$ is well-conditioned, then the off-diagonal decay rate of $\mathcal{F}_{2p}^{-1}$ is rapid (the constants c and λ are small). The importance of this result lies in the fact that off-diagonally decaying matrices with a rapid decay rate can be approximated by sparse banded matrices. In Chapter, 3 we have presented the Chebyshev method for approximating $\mathcal{F}_{2p}^{-1}$. The Chebyshev approximation of $\mathcal{F}_{2p}^{-1}$ is denoted by $\mathcal{E}_{2tp}$, see (3.68).

By substituting $\mathcal{F}_{2p}^{-1}$ with $\mathcal{E}_{2tp}$ in (4.15), we define an approximate solution of the MHE problem:

$$\breve{\underline{\mathbf{x}}}(k-p|k) = \mathcal{E}_{2tp}(\mu \breve{\underline{\mathbf{x}}}(k-p|k-1) + \mathcal{O}_p^T \mathcal{Y}_{k-p}^k) \tag{4.16}$$

The prediction $\breve{\underline{\mathbf{x}}}(k-p|k-1)$, made at the time instant $k-1$, corresponding to (4.16), is given by:

$$\breve{\underline{\mathbf{x}}}(k-p|k-1) = \underline{A} \breve{\underline{\mathbf{x}}}(k-p-1|k-1) \tag{4.17}$$

Because $t \ll N$ and $p \ll N$, all matrices in (4.16) are sparse banded matrices, and consequently (4.16) can be computed with $O(N)$ computational complexity. *The equations* (4.16) *and* (4.17) *constitute the linear computational complexity MHE method*. The fundamental question that needs to be answered here is: "*How the approximation errors affect the stability and the performance of the MHE method*?'

Remark 4.4 *The MHE methods proposed in this thesis can be used for the state estimation of large-scale interconnected systems which system matrices have sparse banded or sparse multi-banded structure. Namely, if we would apply the lifting technique, that has been introduced in Chapter 3, to sparse multi-banded system matrices, the resulting matrix $\mathcal{O}_p$ and the matrix $\mathcal{F}_{2p}$ would inherit the sparse multi-banded structure. In [221], it has been shown that the Chebyshev approximation method can be generalized for approximating inverses of sparse multi-banded system matrices. State-space models, described by sparse (multi) banded system matrices, are important because they originate from discretization of 3D PDEs [60; 200; 227; 228].*

Stability and performance analysis

We define the estimation error as follows:

$$\breve{\underline{\mathbf{e}}}(k-p) = \underline{\mathbf{x}}(k-p) - \breve{\underline{\mathbf{x}}}(k-p|k) \tag{4.18}$$

In order to simplify the stability analysis of the MHE method, we introduce the following assumption.

Assumption 4.5 *The matrix $\mathcal{E}_{2tp}$ is positive definite.*

The justification for this assumption follows from the following facts. First of all, for $p \geq \nu$ or $\mu > 0$ the matrix $\mathcal{F}_{2p}$ is positive definite and invertible. Because $\mathcal{F}_{2p}$ is positive definite, so is its inverse $\mathcal{F}_{2p}^{-1}$. In Theorem 3.4 we have proved that the accuracy of the Chebyshev approximation can be made arbitrarily small. Because of this, if $\mathcal{E}_{2tp}$ is an accurate approximation of the matrix $\mathcal{F}_{2p}^{-1}$, then it is reasonable to assume that $\mathcal{E}_{2tp}$ is positive definite. We are now prepared to analyze the stability of the MHE method.

Theorem 4.6 *Suppose that $\|\underline{A}\|_2 < 1$. Then the system*

$$\begin{bmatrix} \breve{\underline{\mathbf{e}}}(k-p) \\ \underline{\mathbf{x}}(k-p+1) \end{bmatrix} = \begin{bmatrix} \mu\mathcal{E}_{2tp}\underline{A} & I - \mathcal{E}_{2tp}\mathcal{F}_{2p} \\ 0 & \underline{A} \end{bmatrix} \begin{bmatrix} \breve{\underline{\mathbf{e}}}(k-p-1) \\ \underline{\mathbf{x}}(k-p) \end{bmatrix} + \begin{bmatrix} -\mathcal{E}_{2tp}\mathcal{O}_p^T\mathcal{N}_{k-p}^k \\ 0 \end{bmatrix} \tag{4.19}$$

is asymptotically stable for any (non-negative) value of μ.

Proof In order to analyze the stability of the approximate MHE method, we define the following matrix:

$$\Delta = \mathcal{F}_{2p}^{-1} - \mathcal{E}_{2tp} \tag{4.20}$$

Due to Assumption 4.5, the matrix $\mathcal{E}_{2tp}$ is invertible. From (4.16) and (4.18) we have:

$$\breve{\underline{\mathbf{e}}}(k-p) = \mathcal{E}_{2tp}\left(\mathcal{E}_{2tp}^{-1}\underline{\mathbf{x}}(k-p) - \mu\breve{\underline{\mathbf{x}}}(k-p|k-1) - \mathcal{O}_p^T\mathcal{Y}_{k-p}^k\right) \tag{4.21}$$

Using the matrix inversion lemma (see [58], page 19, Lemma 2.2), from (4.20) we obtain:

$$\mathcal{E}_{2tp}^{-1} = \mathcal{F}_{2p} + \Delta_l \tag{4.22}$$

$$\Delta_l = \mathcal{F}_{2p}\Delta(I - \mathcal{F}_{2p}\Delta)^{-1}\mathcal{F}_{2p} \tag{4.23}$$

Taking into account (4.22) we can express (4.21) as follows:

$$\begin{aligned}\underline{\breve{\mathbf{e}}}(k-p) &= \mathcal{E}_{2tp}\left(\mathcal{F}_{2p}\underline{\mathbf{x}}(k-p) - \mu\underline{\breve{\mathbf{x}}}(k-p|k-1) - \mathcal{O}_p^T\mathcal{Y}_{k-p}^k\right)\\ &+ \mathcal{E}_{2tp}\Delta_l\underline{\mathbf{x}}(k-p)\end{aligned} \tag{4.24}$$

From (4.22) we have: $\mathcal{E}_{2tp}\Delta_l = I - \mathcal{E}_{2tp}\mathcal{F}_{2p}$. Taking this transformation into account, we can write (4.24) as follows:

$$\begin{aligned}\underline{\breve{\mathbf{e}}}(k-p) &= \mathcal{E}_{2tp}\left(\mathcal{F}_{2p}\underline{\mathbf{x}}(k-p) - \mu\underline{\breve{\mathbf{x}}}(k-p|k-1) - \mathcal{O}_p^T\mathcal{Y}_{k-p}^k\right)\\ &+ (I - \mathcal{E}_{2tp}\mathcal{F}_{2p})\underline{\mathbf{x}}(k-p)\end{aligned} \tag{4.25}$$

Multiplying (4.12) from left by $\mathcal{O}_p^T$ and adding to the both sides of the resulting equation $\mu\underline{\mathbf{x}}(k-p)$, we obtain:

$$\underbrace{(\mathcal{O}_p^T\mathcal{O}_p + \mu I)}_{\mathcal{F}_{2p}}\underline{\mathbf{x}}(k-p) = \mathcal{O}_p^T\mathcal{Y}_{k-p}^k - \mathcal{O}_p^T\mathcal{N}_{k-p}^k + \mu\underline{\mathbf{x}}(k-p) \tag{4.26}$$

Combining the state equation of the global system (4.1) with (4.26), we have:

$$\mathcal{F}_{2p}\underline{\mathbf{x}}(k-p) = \mathcal{O}_p^T\mathcal{Y}_{k-p}^k - \mathcal{O}_p^T\mathcal{N}_{k-p}^k + \mu\underline{A\mathbf{x}}(k-p-1) \tag{4.27}$$

Substituting (4.17) and (4.27) in (4.25), and keeping in mind that $\underline{\breve{\mathbf{e}}}(k-p-1) = \underline{\mathbf{x}}(k-p-1) - \underline{\breve{\mathbf{x}}}(k-p-1|k-1)$ we obtain the first equation of (4.19). The second equation of (4.19) is a state equation of the global system (4.1). In order to analyse stability of (4.19) we first determine an upper-bound on $\|\mathcal{E}_{2tp}\|_2$. From (3.57) we have:

$$\begin{aligned}&\frac{2}{b-a}|\overline{g}_t(w)|\\ \leq&\frac{2}{b-a}\left(|\frac{c_0}{2}| + |c_1||T_1(w)| + |c_2||T_2(w)| + \ldots + |c_t||T_t(w)|\right)\\ <&\frac{2}{b-a}(|\frac{c_0}{2}| + \sum_{i=1}^{\infty}|c_i||T_i(w)|)\end{aligned} \tag{4.28}$$

It is well known, see [214], that $\max_{w\in[-1,1]}|T_k(w)| = 1$, $\forall k$. Due to this, from (4.28) we have:

$$\frac{2}{b-a}\max_{w\in[-1,1]}|\overline{g}_t(w)| < \frac{2}{b-a}(|\frac{c_0}{2}| + \sum_{i=1}^{\infty}|c_i|) \tag{4.29}$$

Similarly to (3.76), we have:

$$\frac{2}{b-a}(|\frac{c_0}{2}|+\sum_{i=1}^{\infty}|c_i|)=\frac{2}{(b-a)\sqrt{v^2-1}}\left(\frac{|h|+1}{|h|-1}\right) \tag{4.30}$$

Using (3.77) and (3.78) we obtain:

$$\frac{2}{(b-a)\sqrt{v^2-1}}\left(\frac{|h|+1}{|h|-1}\right)=\frac{1}{a}=\frac{1}{\mu+\sigma_{Nn}^2} \tag{4.31}$$

From (4.28)-(4.31) we obtain:

$$\frac{2}{b-a}\max_{w\in[-1,1]}|\overline{g}_t(w)|<\frac{1}{\mu+\sigma_{Nn}^2} \tag{4.32}$$

In order to complete the proof, we will recall the Spectral mapping theorem [229; 230]. The spectral mapping theorem says that for every matrix M and every polynomial $p_t(x)$, the spectrum of $p_t(M)$ is $p_t(\sigma(M))$ (where $\sigma(M)$ denotes the spectrum of M). Applying the spectral mapping theorem to $\mathcal{F}_{2p}$ and $f_t(q)$, defined in (3.62), we conclude that the spectrum of $f_t(\mathcal{F}_{2p})$ is $f_t(\sigma(\mathcal{F}_{2p}))$. The matrix $\mathcal{E}_{2tp}$ is equal to $f_t(\mathcal{F}_{2p})$. Because $\mathcal{E}_{2tp}$ is symmetric positive definite (at least for sufficiently large t), we have that its eigenvalues are equal to its singular values. Using the spectral mapping theorem we conclude:

$$\left\|\mathcal{E}_{2tp}\right\|_2=\max_{q\in\sigma(\mathcal{F}_{2p})}|f_t(q)| \tag{4.33}$$

It is easy to see that:

$$\max_{q\in\sigma(\mathcal{F}_{2p})}|f_t(q)|\leq\max_{w\in[-1,1]}|g_t(w)|\leq\frac{2}{b-a}\max_{w\in[-1,1]}|\overline{g}_t(w)| \tag{4.34}$$

where $g_t(w)$ is defined in (3.60). From (4.32), (4.33) and (4.34), we have:

$$\left\|\mathcal{E}_{2tp}\right\|_2<\frac{1}{\mu+\sigma_{Nn}^2} \tag{4.35}$$

Because $\left\|\underline{A}\right\|_2<1$ (this is the main assumption of the theorem), from (4.19) we see that the system is asymptotically stable if $\rho(\mu\mathcal{E}_{2tp}\underline{A})<1$, where $\rho(.)$ denotes the matrix spectral radius. We have:

$$\rho(\mu\mathcal{E}_{2tp}\underline{A})\leq\left\|\mu\mathcal{E}_{2tp}\underline{A}\right\|_2\leq\mu\left\|\mathcal{E}_{2tp}\right\|_2\left\|\underline{A}\right\|_2 \tag{4.36}$$

From (4.35) and (4.36) we have:

$$\left\|\mu\mathcal{E}_{2tp}\underline{A}\right\|_2<\frac{\mu}{\mu+\sigma_{Nn}^2}\left\|\underline{A}\right\|_2 \tag{4.37}$$

It should be observed that $\frac{\mu}{\mu+\sigma_{Nn}^2}<1$ for any μ. Since $\left\|\underline{A}\right\|_2<1$, then we see that for any value of μ, the right-hand side of (4.37) is smaller than 1. This means that

if $\|\underline{A}\|_2 < 1$ then for any value of μ the augmented system (4.19) is asymptotically stable. This completes the proof. □

It should be noted that the results of Theorem 4.6 are consistent with the results of [148], see Remark 1 in [148].

Remark 4.7 *The effect of the Chebyshev approximation errors on the total estimation error is represented by the term* $(I - \mathcal{E}_{2tp}\mathcal{F}_{2p})\,\underline{\mathbf{x}}(k-p)$ *in* (4.19). *As* $t \to \infty$, *we have that* $\mathcal{E}_{2tp} \to \mathcal{F}_{2p}^{-1}$ *and consequently* $(I - \mathcal{E}_{2tp}\mathcal{F}_{2p}) \to 0$. *If* $(I - \mathcal{E}_{2tp}\mathcal{F}_{2p}) = 0$, *then the stability analysis of the MHE method is equivalent to the stability analysis presented in [148]. Furthermore, it should be observed that if there is no measurement noise, then the estimation error will converge to zero (this also holds for expectation of the estimation error, if the measurement noise is zero-mean). That is, the errors introduced by the Chebyshev approximation do not introduce bias. However, it can be easily shown that if an input is affecting the state equation of the global state-space model, then the combination of this input and Chebyshev approximation errors will produce bias of the estimation error.*

Corollary 4.8 *Suppose that* $\|\underline{A}\|_2 < 1$. *Furthermore, suppose that* $\mathbb{E}\left[\mathcal{N}_{k-p}^{k}\right] = 0$, *for all* k, *where* $\mathbb{E}[.]$ *denotes the expectation operator. Then,*

$$\|\mathbb{E}[\underline{\check{\mathbf{e}}}(k-p+m-1)]\|_2 < b_1 \|\mathbb{E}[\underline{\check{\mathbf{e}}}(k-p-1)]\|_2 + b_2 \|\underline{\mathbf{x}}(k-p-1)\|_2 \tag{4.38}$$

where

$$b_1 = \frac{\mu}{\mu + \sigma_{Nn}^2} \|\underline{A}\|_2^m, \qquad b_2 = \kappa\left(1 - \frac{1}{\sqrt{\kappa}}\right) \frac{1}{|h|^t} \|\underline{A}\|_2 \frac{1 - \|\underline{A}\|_2^m}{1 - \|\underline{A}\|_2} \tag{4.39}$$

and where the parameters κ *and* $|h|$ *are defined in* (3.39) *and* (3.71), *respectively.*

Proof From (4.19) we have:

$$\mathbb{E}[\underline{\check{\mathbf{e}}}(k-p+m-1)] = M_1 \mathbb{E}[\underline{\check{\mathbf{e}}}(k-p-1)] + M_2 \mathbb{E}[\underline{\mathbf{x}}(k-p-1)] \tag{4.40}$$

$$M_1 = (\mu \mathcal{E}_{2tp}\underline{A})^m, \; M_2 = \sum_{i=1}^{m} (\mu \mathcal{E}_{2tp}\underline{A})^{m-i} (I - \mathcal{E}_{2tp}\mathcal{F}_{2p})\,\underline{A}^i \tag{4.41}$$

From (4.40) we have:

$$\begin{aligned} \|\mathbb{E}[\underline{\check{\mathbf{e}}}(k-p+m-1)]\|_2 &\le \|M_1\|_2 \|\mathbb{E}[\underline{\check{\mathbf{e}}}(k-p-1)]\|_2 \\ &+ \|M_2\|_2 \|\underline{\mathbf{x}}(k-p-1)\|_2 \end{aligned} \tag{4.42}$$

From the proof of Theorem 4.6 we have: $\|\mathcal{E}_{2tp}\|_2 < \frac{1}{\mu+\sigma_{Nn}^2}$ and $\mu \|\mathcal{E}_{2tp}\|_2 < \frac{\mu}{\mu+\sigma_{Nn}^2} < 1$. This implies:

$$\|M_1\|_2 \le \mu^m \|\mathcal{E}_{2tp}\|_2^m \|\underline{A}\|_2^m < \left(\frac{\mu}{\mu + \sigma_{Nn}^2}\right)^m \|\underline{A}\|_2^m < b_1 \tag{4.43}$$

On the other hand

$$\|M_2\|_2 \leq \sum_{i=1}^{m} \|\mu\mathcal{E}_{2tp}\underline{A}\|_2^{m-i} \|I - \mathcal{E}_{2tp}\mathcal{F}_{2p}\|_2 \|\underline{A}\|_2^i \tag{4.44}$$

Because $\|\underline{A}\|_2 < 1$ and $\|\mu\mathcal{E}_{2tp}\|_2 < 1$, we have:

$$\|\mu\mathcal{E}_{2tp}\underline{A}\|_2^{m-i} \leq \|\mu\mathcal{E}_{2tp}\|_2^{m-i} \|\underline{A}\|_2^{m-i} < \|\mu\mathcal{E}_{2tp}\|_2 \|\underline{A}\|_2 < 1 \tag{4.45}$$

From (4.44) and (4.45) we have:

$$\|M_2\|_2 < \sum_{i=1}^{m} \|I - \mathcal{E}_{2tp}\mathcal{F}_{2p}\|_2 \|\underline{A}\|_2^i \tag{4.46}$$

It is easy to prove that

$$\sum_{i=1}^{m} \|I - \mathcal{E}_{2tp}\mathcal{F}_{2p}\|_2 \|\underline{A}\|_2^i = \|I - \mathcal{E}_{2tp}\mathcal{F}_{2p}\|_2 \|\underline{A}\|_2 \frac{1 - \|\underline{A}\|_2^m}{1 - \|\underline{A}\|_2} \tag{4.47}$$

Next, we will prove that:

$$\|I - \mathcal{E}_{2tp}\mathcal{F}_{2p}\|_2 \leq \kappa(1 - \frac{1}{\sqrt{\kappa}})\frac{1}{|h|^t} \tag{4.48}$$

First of all, it can be easily verified that the matrix $I - \mathcal{E}_{2tp}\mathcal{F}_{2p}$ is symmetric. Using the spectral mapping theorem (see proof of Theorem 4.6) we have:

$$\begin{aligned} \|I - \mathcal{E}_{2tp}\mathcal{F}_{2p}\|_2 &= \max_{q \in \sigma(\mathcal{F}_{2p})} |1 - f_t(q)q| \\ &\leq \max_{w \in [-1,1]} \left|1 - g_t(w)\frac{1}{2}\left((b-a)w + (a+b)\right)\right| \end{aligned} \tag{4.49}$$

We can write $g_t(w)$ as follows:

$$\begin{aligned} g_t(w) &= g(w) - z_{t+1}(w), \\ z_{t+1}(w) &= \frac{2}{a-b}\sum_{t+1}^{\infty} c_k T_k(w) \end{aligned} \tag{4.50}$$

On the other hand

$$g(w)\frac{1}{2}\left((b-a)w + (a+b)\right) = 1 \tag{4.51}$$

Using (4.50) and (4.51), we have:

$$\max_{w \in [-1,1]} \left|1 - g_t(w)\frac{1}{2}\left((b-a)w + (a+b)\right)\right|$$

$$= \max_{w\in[-1,1]} \left| z_{t+1}(w) \frac{1}{2} \left((b-a)w + (a+b) \right) \right| \leq b \max_{w\in[-1,1]} |z_{t+1}(w)| \tag{4.52}$$

On the other hand

$$\max_{w} |z_{t+1}(w)| \leq \frac{2}{b-a} \max_{w\in[-1,1]} \sum_{t+1}^{\infty} |c_k||T_k(w)| \leq \frac{2}{b-a} \sum_{t+1}^{\infty} |c_k| \tag{4.53}$$

From the proof of Theorem 3.4 (see (3.80)), we have:

$$\frac{2}{b-a} \sum_{t+1}^{\infty} |c_k| = l(\mu, t) = \frac{1}{a} \left(1 - \frac{1}{\sqrt{\kappa}} \right) \frac{1}{|h|^t} \tag{4.54}$$

From (4.52)-(4.54) we have:

$$b \max_{w\in[-1,1]} |z_{t+1}(w)| \leq \frac{b}{a} \left(1 - \frac{1}{\sqrt{\kappa}} \right) \frac{1}{|h|^t}, \;\; \frac{b}{a} \left(1 - \frac{1}{\sqrt{\kappa}} \right) \frac{1}{|h|^t} = \kappa \left(1 - \frac{1}{\sqrt{\kappa}} \right) \frac{1}{|h|^t} \tag{4.55}$$

From (4.49)-(4.55) we obtain (4.48). From (4.46), (4.47) and (4.48), we obtain:

$$\|M_2\|_2 < \kappa(1 - \frac{1}{\sqrt{\kappa}}) \frac{1}{|h|^t} \|\underline{A}\|_2 \frac{1 - \|\underline{A}\|_2^m}{1 - \|\underline{A}\|_2} = b_2 \tag{4.56}$$

From (4.42), (4.43) and (4.56), we obtain:

$$\begin{aligned} &\|\mathbb{E}\left[\underline{\breve{\mathbf{e}}}(k-p+m-1)\right]\|_2 < b_1 \|\mathbb{E}\left[\underline{\breve{\mathbf{e}}}(k-p-1)\right]\|_2 \\ &+ b_2 \|\underline{\mathbf{x}}(k-p-1)\|_2 \end{aligned} \tag{4.57}$$

This completes the proof. □

Some guidelines for selecting μ and t

In the sequel we will give some guidelines for selecting μ and t. These guidelines are valid for the noise-free case or for the case when $\mathbb{E}\left[\mathcal{N}_{k-p}^{k}\right] = 0$, for all k. Furthermore, these guidelines mainly address the computational issues of the developed MHE method. Additional guidelines for tuning the MHE method, that also take the effect of the measurement noise and disturbances into account, can be found in [148; 231].

We have proved that if $\mathcal{E}_{2tp}$ is positive definite and if $\|\underline{A}\|_2 < 1$, then any μ guarantees asymptotic stability of the approximate MHE method (stability of the system (4.19)). However, from Corollary 4.8 we see that μ influences convergence of the estimation error. Using the same arguments that are used in the proof of Lemma 3.2 and Theorem 3.4, it can be proved that b_1 is an increasing function of μ and that b_2 is a decreasing function of t and μ. Furthermore, we have that when $\mu \to 0$, then $b_1 \to 0$.

Since b_2 is a decreasing function of t, by increasing t we decrease the effect of the Chebyshev approximation errors on the total estimation error (see also Remark 4.7). However, for computational reasons we need to keep $t \ll N$.

As we have explained in Chapter 3, if $\mathcal{J}_{2p}$ is well-conditioned, κ is close to 1 and $|h|$ is relatively large. Consequently, independently from μ we can always find $t \ll N$ for which the accuracy of the Chebyshev approximation is good and for which b_2 is relatively small. If b_2 is relatively small, then the parameter b_1 dominantly describes the behavior of the estimation error. In this case, by decreasing μ we can decrease b_1, and consequently we can increase the convergence speed of the estimation error. However, by decreasing μ we increase b_2 and for very small μ and t, the parameter b_2 might become much larger than b_1. If b_2 is much larger than b_1, then b_2 dominantly describes the behavior of the estimation error. This is illustrated in Fig. 4.6, in Numerical experiments section.

If the matrix $\mathcal{J}_{2p}$ is badly conditioned, then in order to ensure that for $t \ll N$ the accuracy of the Chebyshev approximation is good, we need to select a sufficiently large μ. This value of μ can be found by choosing $t \ll N$ and by plotting the function $l(\mu, t)$ or b_2, for different values of μ. From this plot we can select μ for which $l(\mu, t)$ or b_2 are small. However, by increasing μ we increase b_1. This suggests that if $\mathcal{J}_{2p}$ is badly conditioned, then the convergence of the approximate MHE method is generally slower than in the case of well-conditioned $\mathcal{J}_{2p}$. *The degradation of the convergence speed (performance) is the price that we need to pay in order to have a computationally efficient MHE method. In the extreme case, when μ is much larger than σ^2_{Nn}, we have that $b_1 \approx \|A\|_2^m$.*

4.1.3 Distributed MHE method

On the basis of the approximate solution of the centralized MHE problem, in this section we derive an approximate solution of the distributed MHE problem (Problem 4.2 in Section 4.1.1). Because $\mathcal{E}_{2tp}$ is a sparse banded matrix, the i^{th} block row of (4.16) (corresponding to $\breve{\mathbf{x}}_i$) can be written as follows:

$$\breve{\mathbf{x}}_i(k-p|k) = \mu \sum_{j=i-2tp}^{j=i+2tp} P^1_{i,j} \breve{\mathbf{x}}_j(k-p|k-1) + \sum_{j=i-(2t+1)p}^{j=i+(2t+1)p} P^2_{i,j} \mathcal{Y}^k_{j,k-p} \tag{4.58}$$

where $P^l_{i,j}$, $l = 1, 2$ are block elements of $\mathcal{E}_{2tp}$ and $\mathcal{E}_{2tp}\mathcal{O}_p^T$ respectively. From (4.17), we can express the local prediction $\breve{\mathbf{x}}_i(k-p|k-1)$ as follows:

$$\begin{aligned}\breve{\mathbf{x}}_i(k-p|k-1) &= A_{i,i}\breve{\mathbf{x}}_i(k-p-1|k-1) \\ &+ E_{i,i-1}\breve{\mathbf{x}}_{i-1}(k-p-1|k-1) + E_{i,i+1}\breve{\mathbf{x}}_{i+1}(k-p-1|k-1)\end{aligned} \tag{4.59}$$

The equations (4.58) and (4.59) constitute the distributed MHE method. From (4.58) we see that in order to compute $\breve{\mathbf{x}}_i(k-p|k)$ the local computing unit $\mathcal{T}_i$ needs to receive local predictions and local output data from its neighbooring local computing units. That is, the parameter s (see the distributed MHE problem 4.2) should be chosen to be equal to $(2t+1)p$. From here we see that the amount

of communication that is necessary to compute $\breve{\mathbf{x}}_i(k-p|k)$ is proportional to the block bandwidth of $\mathcal{E}_{2tp}$ or equivalently, proportional to tp.

Like it has been proved in Chapter 3, if the matrix $\mathcal{J}_{2p}$ is well-conditioned, then for any value of μ we can find $t \ll N$ for which the accuracy of the Chebyshev approximation is good. *That is, if the matrix $\mathcal{J}_{2p}$ is well-conditioned, then we need small amount of communication to compute* (4.58). If the matrix $\mathcal{J}_{2p}$ is badly conditioned, then we can always determine the parameter μ such that there exists $t \ll N$ for which the accuracy of the Chebyshev approximation is relatively good.

In order to minimize the communication between the local computing units, the parameter t should be minimized. *How to select very small t without jeopardizing the accuracy of the Chebyshev approximation?* From Theorem 3.4 and Corollary 4.8, it follows that this can be achieved by increasing the parameter μ. In the case of well-conditioned $\mathcal{J}_{2p}$, this value of μ is not large at all. In the case of badly conditioned $\mathcal{J}_{2p}$, we need to select sufficiently large μ. However, by increasing μ we increase b_1. *This means that there exists a trade-off between the amount of communication between the local computing units and the convergence speed of the approximate MHE method*.

4.1.4 Numerical experiments

Numerical experiments are performed in MATLAB on a standard desktop personal computer. The data generating model is a global state-space model that consists of $N = 200$ identical local subsystems. The local system matrices of each local subsystem are

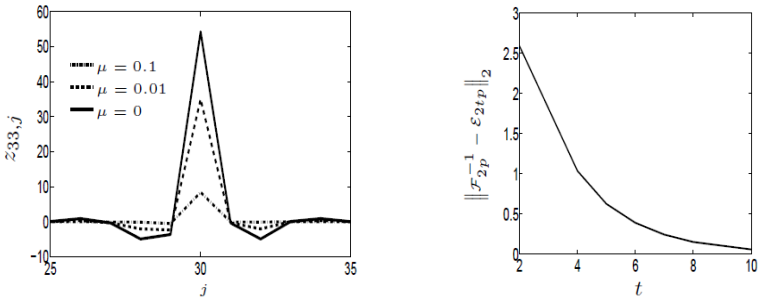
$$A = \begin{bmatrix} 0.7864 & 0.0534 \\ 0.0534 & 0.7864 \end{bmatrix}, E = \begin{bmatrix} 0.0534 & 0 \\ 0 & 0.0534 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{4.60}$$

This model is obtained using the finite difference approximation of the 2D heat equation (see Chapter 2). Heat diffusivity constant and discretization constants are: $\alpha = 0.6$, $h = 5$ and $L = 5.3$. Using the local system matrices (4.60), we construct the global system matrices $\{\underline{A}, \underline{B}, \underline{C}\}$ and we use the global state-space model to generate the sequence of the output data. For each local subsystem we assume that the signal to noise ratio of the noise corrupted output is $20[db]$. We choose $p = 2$. First, we numerically illustrate the results of Lemma 3.2. The parameters c and λ, defined in (3.38), for different values of μ are given in Table 4.1.

μ	0	0.01	0.1	1
λ	0.96	0.94	0.89	0.72
c	65.95	39.7	6.6	0.99

Table 4.1: The parameters c and λ for different values of μ

From Table 4.1 it follows that the parameters c and λ decrease as we increase μ. The exponential off-diagonal decay of $\mathcal{F}_2^{-1}$, for $\mu = 0, 0.1, 1$, is shown in Fig. 4.4(a). Furthermore, from Fig. 4.4(a), it follows that the off-diagonal decay of $\mathcal{F}_2^{-1}$ is more rapid as we increase μ.

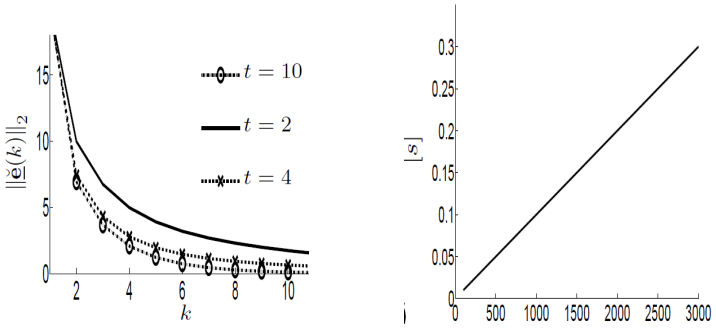


(a) (b)

Figure 4.4: a) The exponential off-diagonal decay of the elements of the 33^{rd} row of $\mathcal{F}_2^{-1}$. The elements are denoted by $z_{33,j}$; b) The Chebyshev approximation error for different orders t.

The error of approximating $\mathcal{F}_2^{-1}$ by $\mathcal{E}_{4t}$ ($p = 2$ and $\mu = 0.1$), for different orders t, is illustrated in Fig. 4.4(b). Fig. 4.4(b) shows that the Chebyshev approximation error decreases as t increases. Next, we illustrate how the estimation error (4.18) depends on t. We choose $\mu = 0.1$ and plot the estimation error for different orders $t = 2, 4, 7$. The results are presented in Fig. 4.5(a). As it is predicted by Corollary 4.8, the convergence rate of the MHE method increases as t is increased.

Next, we illustrate the linear computational complexity of approximating the matrix $\mathcal{F}_{2p}^{-1}$ for different number of local subsystems N. We fix the order of approximation (3.68) to $t = 7$ and choose $\mu = 0.1$. The computational times are presented in Fig. 4.5(b). From 4.5.b, we can clearly see that the proposed MHE method has the linear computational complexity.



(a) (b)

Figure 4.5: a) Convergence of the estimation error for different t. b) Computational times in seconds of approximating $\mathcal{F}_{2p}^{-1}$ by the matrix $\mathcal{E}_{2tp}$, for different numbers of local subsystems N.

In Fig. 4.6 we illustrate how the convergence rate of the estimation error depends

on μ. In Fig. 4.6(a), t is relatively small ($t = 4$). We see that by increasing μ, the estimation error converges more rapidly. This is because b_2 is much larger than b_1 and consequently, b_2 dominantly describes the behavior of the estimation error (b_2 is a decreasing function of μ). On the other hand, in Fig. 4.6(b), t is relatively large ($t = 15$). We see that when μ increases, the convergence rate of the estimation error decreases. In this case b_1 is larger than b_2, and consequently b_1 dominantly describes the behavior of the estimation error (b_1 is an increasing function of μ).

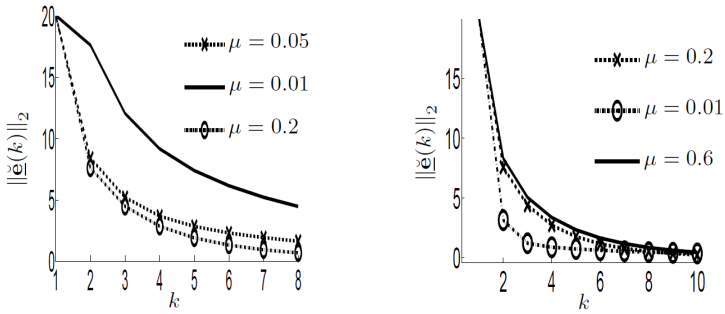


Figure 4.6: Convergence of the estimation error for: a) $t = 4$ b) $t = 15$

4.1.5 Conclusion

In this Section we presented a linear computational complexity MHE method for large-scale interconnected systems. Furthermore, we presented a novel, distributed MHE method, that computes the estimates of the local states using local input-output data. The proposed distributed algorithm is not relying on consensus algorithms and has a simple analytic form. We performed numerical simulations that confirm our theoretical results.

4.2 Moving horizon estimation for descriptor systems

In Chapter 2, we used the finite element method to approximate the thermoelastic system of PDEs (2.67)-(2.68). As a result, we obtained the descriptor state-space model (2.81)-(2.82):

$$Q_{22}\mathbf{x}(k) = D_{22}\mathbf{x}(k-1) + \mathbf{c}_1 \tag{4.61}$$

$$\mathbf{y}(k) = C\mathbf{x}(k) + \mathbf{d}_1 + \mathbf{n}(k) \tag{4.62}$$

where

$$Q_{22} = D_{22} + hK_{22}, \;\; \mathbf{c}_1 = h\mathbf{l}_2, \;\; C = -WK_{11}^{-1}K_{12}, \;\; \mathbf{d}_1 = WK_{11}^{-1}\mathbf{l}_1 \tag{4.63}$$

By inverting the matrix Q_{22}, the descriptor state-space model (4.61)-(4.62) can be transformed into the standard state-space form. However, Q_{22}^{-1} is a dense matrix and consequently, system matrices of the resulting state-space model are dense. That is, the sparsity of the model is "destroyed" (like it was explained previously, this fact makes the estimation problem computationally infeasible). This is the main reason why in this section we keep the state-space model in its original, sparse, descriptor form.
Because we do not want to invert Q_{22}, the problem of estimating the state of (4.61)-(4.62) in a moving horizon manner, is very similar to the MHE problems for singular descriptor systems[1]. The MHE method for singular descriptor systems has been proposed in [232]. However, this MHE method is computationally feasible only for low-dimensional descriptor systems.

Using the approximation framework developed in Chapter 3, in this section we develop a computationally efficient MHE method for descriptor systems. For presentation clarity, we will first formulate the state estimation problem as a least-squares problem.

4.2.1 Least-squares state estimation

The state sequence vector $\mathbf{x}_{k-p}^{k}$ is defined as follows:

$$\mathbf{x}_{k-p}^{k} = \operatorname{col}\left(\mathbf{x}(k-p), \mathbf{x}(k-p+1), \ldots, \mathbf{x}(k)\right) \tag{4.64}$$

The system (4.61)-(4.62) can be written compactly:

$$\underbrace{\begin{bmatrix} Q_{22} \\ C \end{bmatrix}}_{R} \mathbf{x}(k) + \underbrace{\begin{bmatrix} -D_{22} \\ 0 \end{bmatrix}}_{V} \mathbf{x}(k-1) + \underbrace{\begin{bmatrix} 0 \\ \mathbf{n}(k) \end{bmatrix}}_{\tilde{\mathbf{n}}(k)} = \underbrace{\begin{bmatrix} \mathbf{c}_1 \\ \mathbf{y}(k) - \mathbf{d}_1 \end{bmatrix}}_{\mathbf{g}(k)} \tag{4.65}$$

From (4.62) and (4.65) we have:

$$\Gamma \mathbf{x}_{k-p}^{k} + \mathbf{e}_{k-p}^{k} = \mathbf{q}_{k-p}^{k} \tag{4.66}$$

where

$$\Gamma = \overbrace{\begin{bmatrix} C & 0 & 0 & \ldots & \ldots & 0 \\ V & R & 0 & \ldots & \ldots & 0 \\ 0 & V & R & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & V & R & 0 \\ 0 & \ldots & 0 & 0 & V & R \end{bmatrix}}^{\text{p+1 blocks}}, \quad \mathbf{q}_{k-p}^{k} = \begin{bmatrix} \mathbf{y}(k-p) - \mathbf{d_1} \\ \mathbf{g}(k-p+1) \\ \mathbf{g}(k-p+2) \\ \vdots \\ \mathbf{g}(k-1) \\ \mathbf{g}(k) \end{bmatrix}, \quad \mathbf{e}_{k-p}^{k} = \begin{bmatrix} \mathbf{n}(k-p) \\ \tilde{\mathbf{n}}(k-p+1) \\ \vdots \\ \tilde{\mathbf{n}}(k-1) \\ \tilde{\mathbf{n}}(k) \end{bmatrix} \tag{4.67}$$

[1] Singular descriptor systems are characterized by the property that the matrix Q_{22} in (4.61) is not invertible.

The state sequence $\mathbf{x}_{k-p}^{k}$ can be estimated by solving the following least-squares problem:

$$\min_{\mathbf{x}_{k-p}^{k}} \left\| \mathbf{q}_{k-p}^{k} - \Gamma \mathbf{x}_{k-p}^{k} \right\|_2^2 \tag{4.68}$$

We assume that p is selected such that the matrix Γ is tall. Furthermore, we assume that Γ has full column tank. Under these assumptions the solution of (4.68) is:

$$\hat{\mathbf{x}}_{k-p}^{k} = \Gamma^{\dagger} \mathbf{q}_{k-p}^{k} \tag{4.69}$$

where $\Gamma^{\dagger} = \left(\Gamma^T \Gamma\right)^{-1} \Gamma^T$ is a pseudo-inverse of Γ. The sparsity pattern of $\Gamma^T \Gamma$ is illustrated in Fig. 4.7. Although this matrix is a sparse banded matrix, its inverse $\left(\Gamma^T \Gamma\right)^{-1}$ is a dense matrix. Consequently, it might not be possible to compute this inverse and to store it in a computer memory.

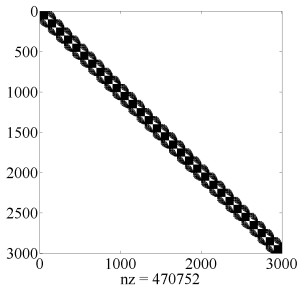


Figure 4.7: Sparsity pattern of the matrix $\Gamma^T \Gamma$ for an arbitrary p.

4.2.2 Moving horizon estimation

Similarly to the MHE problem formulation presented in Section 4.1, the vector $\hat{\mathbf{x}}(k-p|k)$ will denote an estimate of $\mathbf{x}(k-p)$ computed at the time instant k. However, for reasons that will become apparent later, the prediction of $\mathbf{x}(k-p)$ made at the time instant $k-1$ will be denoted by $\overline{\mathbf{x}}(k-p|k-1)$.

The MHE method described in Section 4.1 at the time step $k-1$ computes the prediction of $\mathbf{x}(k-p)$ by using $\hat{\mathbf{x}}(k-p-1|k-1)$ and by propagating the state equation. However, in the case of the descriptor state-space model (4.61)-(4.62), it might not be possible to compute the prediction using this strategy. In order to compute the prediction, we need to solve the following equation:

$$Q_{22} \overline{\mathbf{x}}(k-p|k-1) = D_{22} \hat{\mathbf{x}}(k-p-1|k-1) + \mathbf{c}_1 \tag{4.70}$$

where the unknown variable is $\overline{\mathbf{x}}(k-p|k-1)$. The linear system of equations (4.70) can be solved using the Conjugate Gradient (CG) method. However, in practice the linear system needs to be solved in the prescribed, short time interval (mainly determined by the sampling period of the system). Because the CG method is an

iterative method that does not explicitly compute Q_{22}^{-1}, it is not suitable for real-time implementation of the MHE method. Another option for solving (4.70) is to use the approximation strategy presented in Chapter 3 to compute an approximate, sparse inverse of Q_{22}^{-1}. However, at the prediction step it is not advisable to introduce approximation errors because they can cause divergence[2].

In practice we are mainly interested in the MHE estimate. The prediction is only an intermediate quantity that is used to compute the estimate. The above explained arguments motivate us to develop the MHE strategy that does not rely on the explicit computation of the state prediction.
Basically, the new MHE method at the time instant k estimates the state sequence $\mathbf{x}_{k-p}^{k}$ from the output data (in contrast to the MHE method presented in Section 4.1 that only directly estimates $\mathbf{x}(k-p)$ from the input-output data).

At the time instant $k \geq p$, the MHE cost function is defined by:

$$J_k\left(\hat{\mathbf{x}}(k-p|k),\ldots,\hat{\mathbf{x}}(k|k)\right) = \underbrace{\sum_{i=k-p}^{k} \mu \left\| Q_{22}\overline{\mathbf{x}}(i|k-1) - Q_{22}\hat{\mathbf{x}}(i|k)\right\|_2^2}_{\text{Part I}} + \underbrace{\sum_{i=k-p}^{k-1} \left\|\mathbf{g}(i+1) - R\hat{\mathbf{x}}(i+1|k) - V\hat{\mathbf{x}}(i|k)\right\|_2^2 + \left\|(\mathbf{y}(k-p) - \mathbf{d}_1) - C\hat{\mathbf{x}}(k-p|k)\right\|_2^2}_{\text{Part II}} \tag{4.71}$$

where $\mu \geq 0$ is a penalization parameter. The MHE cost function consists of the two parts. The first part penalize the difference between the prediction of the states made at the time instant $k-1$ and the MHE estimates that we want to determine. The second part is equivalent to the cost function of the least-squares problem (4.68). By increasing μ we put more emphasis on the model and less emphasis on the available data, and vice-versa. For $i = k-p,\ldots,k$ the prediction vectors $\overline{\mathbf{x}}(i|k-1)$ should satisfy:

$$Q_{22}\overline{\mathbf{x}}(i|k-1) = D_{22}\hat{\mathbf{x}}(i-1|k-1) + \mathbf{c}_1 \tag{4.72}$$

That is, the prediction $\overline{\mathbf{x}}(i|k-1)$ is not computed explicitly. Instead, the product $Q_{22}\overline{\mathbf{x}}(i|k-1)$ is computed using $\hat{\mathbf{x}}(i-1|k-1)$, $i = k-p,\ldots,k$.

For convenience and with a slight abuse of the notation, the MHE estimates $\hat{\mathbf{x}}(k-p|k),\ldots,\hat{\mathbf{x}}(k|k)$ will be grouped into the following vector:

$$\hat{\mathbf{x}}\left(k-p:k|k\right) = \text{col}\left(\hat{\mathbf{x}}(k-p|k), \hat{\mathbf{x}}(k-p+1|k),\ldots,\hat{\mathbf{x}}(k|k)\right) \tag{4.73}$$

It can be easily proved that the cost function (4.71) can be transformed into the following form:

$$J^k\left(\hat{\mathbf{x}}\left(k-p:k|k\right)\right) = \left(\mathbf{q}_3 - \Gamma_3\hat{\mathbf{x}}\left(k-p:k|k\right)\right)^T W \left(\mathbf{q}_3 - \Gamma_3\hat{\mathbf{x}}\left(k-p:k|k\right)\right) \tag{4.74}$$

[2]In the approximate MHE method presented in Section 4.1 approximation errors are only introduced at the estimation step, but not in the prediction step.

where

$$\mathbf{q}_3 = \begin{bmatrix} \mathbf{q}_{k-p}^k \\ Q_{22}\overline{\mathbf{x}}(k-p|k-1) \\ \vdots \\ Q_{22}\overline{\mathbf{x}}(k|k-1) \end{bmatrix}, \quad \Gamma_3 = \begin{bmatrix} \Gamma \\ \mathcal{Q} \end{bmatrix}, \quad W = \begin{bmatrix} \mathcal{I} & 0 \\ 0 & \mu\mathcal{I} \end{bmatrix}$$
$$\mathcal{Q} = \begin{bmatrix} Q_{22} & & \\ & \ddots & \\ & & Q_{22} \end{bmatrix}, \quad \mathcal{I} = \begin{bmatrix} I & & \\ & \ddots & \\ & & I \end{bmatrix} \tag{4.75}$$

where Γ and $\mathbf{q}_{k-p}^k$ are defined in (4.67). The MHE estimate is determined by solving the following optimization problem:

$$\min_{\hat{\mathbf{x}}(k-p:k|k)} J^k\left(\hat{\mathbf{x}}(k-p:k|k)\right) \tag{4.76}$$

Assuming that the matrix Γ has a full column rank, the solution of (4.76) is given as follows:

$$\hat{\mathbf{x}}(k-p:k|k) = M^{-1}\Gamma_3^T W \mathbf{q}_3 \tag{4.77}$$

where

$$M = \Gamma_3^T W \Gamma_3 = \Gamma^T\Gamma + \mu\mathcal{Q}^T\mathcal{Q} \tag{4.78}$$

The sparsity pattern of M is illustrated in Fig. 4.8.

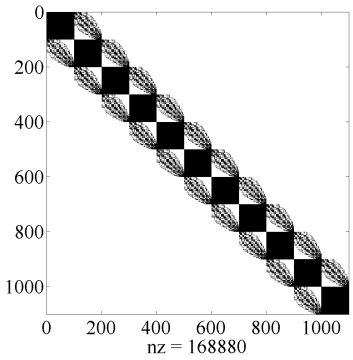


Figure 4.8: Sparsity pattern of the matrix M

In order to compute the estimate (4.77) it is necessary to invert the sparse banded matrix M. The matrix M is a symmetric positive definite matrix. If this matrix is well-conditioned, its inverse can be approximated by a sparse banded matrix, with the linear computational and memory complexity.

In the next proposition, we derive an upper bound on the condition number of M and we prove that this upper bound is a decreasing function of μ. This implies

that we can always find μ for which the off-diagonal decay of M^{-1} is rapid and consequently, the matrix M^{-1} can be approximated by a sparse banded matrix with linear computational complexity.

Proposition 4.9 *Consider the matrix M defined in* (4.78)*. Let the condition number of M be denoted by $\omega(M)$. Then,*

- *The condition number of M is bounded by:*

$$\omega(M) \leq \gamma(\mu), \tag{4.79}$$

$$\gamma(\mu) = \frac{\sigma_{max}\left(\Gamma^T\Gamma\right) + \mu\sigma_{max}\left(\mathcal{Q}^T\mathcal{Q}\right)}{\sigma_{min}\left(\Gamma^T\Gamma\right) + \mu\sigma_{min}\left(\mathcal{Q}^T\mathcal{Q}\right)} \tag{4.80}$$

  where $\sigma_{max}(X)$ and $\sigma_{min}(X)$ denote maximal and minimal singular values of an arbitrary matrix X.

- *The upper bound $\gamma(\mu)$ is a decreasing function of μ.*

Proof The condition number of M is defined by:

$$\omega(M) = \frac{\sigma_{max}(M)}{\sigma_{min}(M)} \tag{4.81}$$

Because M is a symmetric, positive definite matrix, its singular values are equal to its eigenvalues. This implies:

$$\omega(M) = \frac{\lambda_{max}(M)}{\lambda_{min}(M)} \tag{4.82}$$

where $\lambda_{max}(M)$ and $\lambda_{min}(M)$ denote maximal and minimal eigenvalues of M. On the other hand, using Weyl's inequalities [112], we have:

$$\lambda_{min}\left(\Gamma^T\Gamma\right) + \mu\lambda_{min}\left(\mathcal{Q}^T\mathcal{Q}\right) \leq \lambda_{min}(M) \tag{4.83}$$

$$\lambda_{max}(M) \leq \lambda_{max}\left(\Gamma^T\Gamma\right) + \mu\lambda_{max}\left(\mathcal{Q}^T\mathcal{Q}\right) \tag{4.84}$$

From (4.82),(4.83) and (4.84), we have:

$$\omega(M) \leq \frac{\lambda_{max}\left(\Gamma^T\Gamma\right) + \mu\lambda_{max}\left(\mathcal{Q}^T\mathcal{Q}\right)}{\lambda_{min}\left(\Gamma^T\Gamma\right) + \mu\lambda_{min}\left(\mathcal{Q}^T\mathcal{Q}\right)} \tag{4.85}$$

Furthermore, the matrices $\Gamma^T\Gamma$ are $\mathcal{Q}^T\mathcal{Q}$ are also symmetric positive definite and consequently, their singular values are equal to their eigenvalues. Taking these facts into account, we have:

$$\lambda_{min}\left(\Gamma^T\Gamma\right) + \mu\lambda_{min}\left(\mathcal{Q}^T\mathcal{Q}\right) = \sigma_{min}\left(\Gamma^T\Gamma\right) + \mu\sigma_{min}\left(\mathcal{Q}^T\mathcal{Q}\right) \tag{4.86}$$

$$\lambda_{max}\left(\Gamma^T\Gamma\right) + \mu\lambda_{max}\left(\mathcal{Q}^T\mathcal{Q}\right) = \sigma_{max}\left(\Gamma^T\Gamma\right) + \mu\sigma_{max}\left(\mathcal{Q}^T\mathcal{Q}\right) \tag{4.87}$$

From (4.85), (4.86), and (4.87) we obtain (4.79). It is obvious that the upper bound $\gamma(\mu)$ is a decreasing function of μ. □

In Chapter 7, we will use the developed MHE method to estimate the state (the temperature distribution) of the descriptor state-space model (4.61)-(4.62) that describes the dynamics of thermoelastic deformations of mirrors used in optical systems.

4.3 On the structure of the Newton observer for large-scale interconnected systems

Like it has been mentioned in Chapter 2, (see Remark 2.2), mirrors used in EUV lithographic machines can be made of materials that have Coefficient of Thermal Expansion (CTE) that depends on the temperature. As it has been shown in [233], in the case of materials with the temperature dependent CTE, the dynamical behavior of thermally induced wavefront aberrations is described by a linear state equation (the heat equation) and a nonlinear output equation.

Using the approximation framework presented in Chapter 3, in Section 4.1 of this chapter we have developed computationally efficient MHE algorithms for linear state-space models. In this section we will show that the approximation framework proposed in Chapter 3 can be used to develop a computationally efficient estimation algorithm for large-scale systems with a linear state equation and nonlinear output equation. The estimation algorithm is based on the Newton observer that has been proposed in [234; 235]. This algorithm can be used for state estimation of the thermoelastic equations that have the temperature depending CTE.

Without loss of generality, we will consider a state-space model obtained by discretizing the 1D heat equation. The 1D discretization domain is shown in Fig. 4.9.

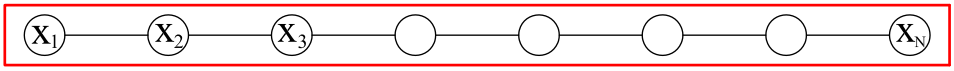


Figure 4.9: The discretization domain for the 1D heat equation. The temperatures at the discretization points are denoted by x_i, $i = 1, \ldots, N$.

For simplicity we will assume zero boundary and initial conditions. The discretized heat equation has the following form:

$$x_i(k+1) = ax_i(k) + ex_{i-1}(k) + ex_{i+1}(k) \tag{4.88}$$

where $a \in \mathbb{R}$ and $e \in \mathbb{R}$ are constants that depend on the material properties and on the discretization steps (see Chapter 2 for more details). From Fig. 4.9 we see that the discretized heat equation (4.88) can be interpreted as a state equation of a local subsystem $\mathcal{S}_i$. The local output equation is nonlinear and it has the following

form:

$$y_i(k) = bx_i^2(k) + cx_i(k) \tag{4.89}$$

where $b \in \mathbb{R}$ and $c \in \mathbb{R}$ are constants. The local output $y_i(k)$ is the temperature induced relative deformation at the i^{th} discretization point. The local output equation (4.89) is nonlinear because we assumed that the CTE depends on the temperature[3]. The global system has the following form:

$$\underline{\mathbf{x}}(k+1) = \underline{A}\underline{\mathbf{x}}(k) \tag{4.90}$$

$$\underline{\mathbf{y}}(k) = h\left(\underline{\mathbf{x}}(k)\right) \tag{4.91}$$

where

$$\underline{A} = \begin{bmatrix} a & e & & & & \\ e & a & e & & & \\ & \ddots & & & & \\ & & & e & a & e \\ & & & & e & a \end{bmatrix}, \ \underline{\mathbf{x}}(k) = \begin{bmatrix} x_1(k) \\ x_2(k) \\ \vdots \\ x_N(k) \end{bmatrix}, \ \underline{\mathbf{y}}(k) = \begin{bmatrix} y_1(k) \\ y_2(k) \\ \vdots \\ y_N(k) \end{bmatrix}$$

$$h\left(\underline{\mathbf{x}}(k)\right) = \begin{bmatrix} bx_1^2(k) + cx_1(k) \\ bx_2^2(k) + cx_2(k) \\ \vdots \\ bx_N^2(k) + cx_N(k) \end{bmatrix} \tag{4.92}$$

For simplicity, we write the global system (4.90)-(4.91) in a compact form:

$$\underline{\mathbf{x}}(k+1) = g\left(\underline{\mathbf{x}}(k)\right) \tag{4.93}$$

$$\underline{\mathbf{y}}(k) = h\left(\underline{\mathbf{x}}(k)\right) \tag{4.94}$$

where $g\left(\underline{\mathbf{x}}(k)\right) = \underline{A}\underline{\mathbf{x}}(k)$. By lifting the global system (4.93)-(4.94), p time steps, we obtain:

$$Y_k^{k+p} = H\left(\underline{\mathbf{x}}(k)\right) \tag{4.95}$$

where

$$H\left(\underline{\mathbf{x}}(k)\right) = \begin{bmatrix} h\left(\underline{\mathbf{x}}(k)\right) \\ h \circ g\left(\underline{\mathbf{x}}(k)\right) \\ h \circ g \circ g\left(\underline{\mathbf{x}}(k)\right) \\ \vdots \\ h \circ \underbrace{g \circ g \circ \ldots \circ g}_{\text{p-1 times}}\left(\underline{\mathbf{x}}(k)\right) \end{bmatrix}, \ Y_k^{k+p} = \begin{bmatrix} \underline{\mathbf{y}}(k) \\ \vdots \\ \underline{\mathbf{y}}(k+p) \end{bmatrix} \tag{4.96}$$

[3]For simplicity we assumed quadratic output nonlinearity. The estimation framework presented in this section can handle other types of differentiable, output nonlinearities.

where $g \circ g$ denotes function composition. That is,

$$g \circ g\left(\underline{\mathbf{x}}(k)\right) = g\left(g\left(\underline{\mathbf{x}}(k)\right)\right) = g\left(\mathbf{x}(k+1)\right) = \underline{A}\underline{\mathbf{x}}(k+1) = \underline{A}^2\underline{\mathbf{x}}(k)$$

Similarly, $h \circ g\left(\underline{\mathbf{x}}(k)\right)$ stands for:

$$h \circ g\left(\underline{\mathbf{x}}(k)\right) = h\left(g\left(\underline{\mathbf{x}}(k)\right)\right) = h\left(\mathbf{x}(k+1)\right)$$

If the output equation (4.91) is linear than the lifted equation (4.95) is very similar to the equation (3.5) (when the inputs are not affecting the state dynamics). The main idea of the Newton observer [234; 235] is to estimate $\mathbf{x}(k)$ by solving the system of nonlinear equations:

$$Y_k^{k+p} - H\left(\underline{\mathbf{x}}(k)\right) = 0 \tag{4.97}$$

using Newton's method. The solution is calculated iteratively :

$$\boldsymbol{\gamma}^{i+1} = \boldsymbol{\gamma}^i + \left[\frac{\partial H}{\partial \underline{\mathbf{x}}(k)}\left(\boldsymbol{\gamma}^i\right)\right]^{\dagger}\left(Y_k^{k+p} - H\left(\boldsymbol{\gamma}^i\right)\right) \tag{4.98}$$

where $\boldsymbol{\gamma}^i$ is the solution at the i^{th} iteration and $\frac{\partial H}{\partial \mathbf{x}(k)}$ is the Jacobian matrix (for more details see [234]). Because the matrix $\underline{A}$ is sparse banded matrix, the Jacobian is a sparse, structured matrix. Consequently, using the approximation framework presented in Chapter 3, the pseudo-inverse of the Jacobian matrix can be approximated by a sparse structured matrix. To show this, we define the functions F_i as follows:

$$F_1\left(\underline{\mathbf{x}}(k)\right) = h\left(\underline{\mathbf{x}}(k)\right), \quad F_2\left(\underline{\mathbf{x}}(k)\right) = h \circ g\left(\underline{\mathbf{x}}(k)\right), \quad F_3\left(\underline{\mathbf{x}}(k)\right) = h \circ g \circ g\left(\underline{\mathbf{x}}(k)\right), \ldots \tag{4.99}$$

The function $H\left(\underline{\mathbf{x}}(k)\right)$ can be written as follows:

$$H\left(\underline{\mathbf{x}}(k)\right) = \begin{bmatrix} F_1\left(\underline{\mathbf{x}}(k)\right) \\ F_2\left(\underline{\mathbf{x}}(k)\right) \\ \vdots \\ F_{p+1}\left(\underline{\mathbf{x}}(k)\right) \end{bmatrix} \tag{4.100}$$

The Jacobian is defined as follows:

$$\frac{\partial H}{\partial \underline{\mathbf{x}}(k)} = \begin{bmatrix} \frac{\partial F_1}{\partial x_1(k)} & \cdots & \frac{\partial F_1}{\partial x_N(k)} \\ \frac{\partial F_2}{\partial x_1(k)} & \cdots & \frac{\partial F_2}{\partial x_N(k)} \\ \vdots & & \vdots \\ \frac{\partial F_{p+1}}{\partial x_1(k)} & \cdots & \frac{\partial F_{p+1}}{\partial x_N(k)} \end{bmatrix} \tag{4.101}$$

It is easy to prove:

$$\frac{\partial F_1}{\partial x_1(k)} = \begin{bmatrix} 2bx_1(k)+c \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \frac{\partial F_1}{\partial x_2(k)} = \begin{bmatrix} 0 \\ 2bx_2(k)+c \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \dots \tag{4.102}$$

Now, because the matrix $\underline{A}^l$ has bandwidth equal to l and because $\underline{\mathbf{x}}(k+l) = \underline{A}^l \underline{\mathbf{x}}(k)$, we have:

$$x_i(k+l) = f_{i,l}(x_{i-l}(k), x_{i-l+1}(k), \dots, x_{i+l}(k)) \tag{4.103}$$

where $f_{i,l}$ is a function of local states $x_{i-l}(k), x_{i-l+1}(k), \dots, x_{i+l}(k)$. From (4.103) we have:

$$\frac{\partial x_i(k+l)}{\partial x_j(k)} = 0, j = 1, \dots, i-l-1, i+l+1, \dots, N \tag{4.104}$$

Using (4.104) we obtain:

$$\frac{\partial F_2}{\partial x_1(k)} = \begin{bmatrix} (2bx_1(k+1)+c)\frac{\partial x_1(k+1)}{\partial x_1(k)} \\ (2bx_2(k+1)+c)\frac{\partial x_2(k+1)}{\partial x_1(k)} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \frac{\partial F_2}{\partial x_2(k)} = \begin{bmatrix} (2bx_1(k+1)+c)\frac{\partial x_1(k+1)}{\partial x_2(k)} \\ (2bx_2(k+1)+c)\frac{\partial x_2(k+1)}{\partial x_2(k)} \\ (2bx_3(k+1)+c)\frac{\partial x_3(k+1)}{\partial x_2(k)} \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

$$\frac{\partial F_2}{\partial x_3(k)} = \begin{bmatrix} 0 \\ (2bx_2(k+1)+c)\frac{\partial x_2(k+1)}{\partial x_3(k)} \\ (2bx_3(k+1)+c)\frac{\partial x_3(k+1)}{\partial x_3(k)} \\ (2bx_4(k+1)+c)\frac{\partial x_4(k+1)}{\partial x_3(k)} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \dots \tag{4.105}$$

Using the same principle it can be proved that each of $\frac{\partial F_3}{\partial x_i(k)}$ is a sparse vector with no more than 5 nonzero elements and etc. From (4.102) we see that the matrix:

$$\begin{bmatrix} \frac{\partial F_1}{\partial x_1(k)} & \dots & \frac{\partial F_1}{\partial x_N(k)} \end{bmatrix}$$

that is the first block row of the Jacobian (4.101), is a diagonal matrix. From (4.105) we have that the matrix

$$\begin{bmatrix} \frac{\partial F_2}{\partial x_1(k)} & \dots & \frac{\partial F_2}{\partial x_N(k)} \end{bmatrix}$$

that is a second block row of the Jacobian, is a sparse, banded matrix, with the bandwidth equal to 1. Similarly, the third block row of the Jacobian is a sparse, banded matrix with the bandwidth equal to 2 and etc. That is, the Jacobian matrix has the sparsity structure of the observability matrix of the global system:

$$\begin{aligned}\underline{\mathbf{x}}(k+1) &= \underline{A}\underline{\mathbf{x}}(k)\\ \underline{\mathbf{y}}(k) &= \underline{C}\underline{\mathbf{x}}(k)\end{aligned}$$

where the matrix $\underline{A}$ is defined in (4.92) and $\underline{C}$ is a diagonal matrix. This sparsity structure is very similar to the sparsity structure of the observability matrix shown in Fig. 3.2(a).

The structure preserving lifting technique can be used to permute the rows of the equation (4.97). Starting from this equation, a permuted version of (4.98) can be obtained. The Jacobian matrix that corresponds to the permuted version of the equation (4.97) is a sparse, banded matrix (its sparsity pattern is very similar to the sparsity pattern of the matrix shown in Fig. 3.3(a)). This Jacobian can be obtained by directly permuting the rows of (4.101). If the permuted Jacobian is well-conditioned, then its pseudo-inverse can be approximated by a sparse banded matrix. If it is not, then the regularization technique can be used to decrease the condition number of the transformed Jacobian. All this implies that one iteration of Newton's observer, that is defined by (4.98), can be implemented with $O(N)$ complexity.

In this section we have shown that the approximation framework presented in Chapter 3 can be used to develop computationally efficient state estimators for interconnected systems with output nonlinearities. *In principle, this shows that the developed approximation methods can be used for efficient state estimation of the thermoelastic equations that have the temperature depending CTE.*

5 CHAPTER

Subspace identification of large-scale interconnected systems

In this chapter we propose a decentralized subspace algorithm for identifying large-scale, interconnected systems that are described by sparse banded or multi-banded system matrices. First, we prove that the state of a local subsystem can be approximated by a linear combination of inputs and outputs of the local subsystems that are in its neighborhood. Furthermore, we prove that for interconnected systems with well-conditioned observability matrices (or observability Gramians), the size of this neighborhood is relatively small (compared to the total number of local subsystems). On the basis of these results, we develop a subspace identification algorithm that identifies the state-space model of a local subsystem from the local input-output data. Consequently, the proposed algorithm is computationally feasible for interconnected systems with a large number of local subsystems. We numerically illustrate the effectiveness of the new identification algorithm.

5.1 Introduction

The problem of controlling large-scale interconnected systems has received a significant attention in the last few decades, see for example [76; 77; 78; 80; 236] and references therein. Unfortunately, the classical identification techniques, like the Subspace Identification Methods (SIMs) [58; 237] or the Prediction Error Methods (PEMs) [141], are not suitable for identification of large-scale interconnected systems because their computational and memory complexities scale with $O(N^3)$ and $O(N^2)$, respectively, where N is the number of local subsystems. Furthermore, the SIMs identify the state-space representation of an interconnected system, in which the interconnection structure is destroyed by unknown similarity transformation [58]. However, for efficient distributed controller synthesis we need a structured

state-space model of an interconnected system [76; 77; 78; 80; 236]. *From an identification point of view, this means that the interconnection structure of a large-scale system has to be preserved in the identified model.*

On the other hand, the SIMs and the PEMs are centralized identification methods that assume that input-output data of all local subsystems can be collected and processed in one computing unit. In the cases in which a large number of local sensors collect measurement data of local subsystems, the transfer of sensor measurements to one centralized computing requires a large amount of energy and communication [225; 226]. In such situations, identification should be performed in a decentralized/distributed manner on a network of local computing units that communicate locally.

In [238; 239; 240], identification strategies for ARX models of interconnected systems have been proposed. However, these methods cannot be used for the identification of state-space models of interconnected systems. Subspace identification algorithms for large-scale systems have been proposed in [127; 241]. Unfortunately, these algorithms are restricted to interconnected systems with identical local subsystems and they are computationally infeasible in the case of a large number of local subsystems.

In this chapter we propose a decentralized subspace algorithm for identifying large-scale, interconnected systems that have sparse banded or multi-banded system matrices. First, we prove that the state of a local subsystem can be approximated by a linear combination of inputs and outputs of the local subsystems that are in its neighborhood. Furthermore, we prove that for interconnected systems with well-conditioned observability matrices (or observability Gramians), the size of this neighborhood is relatively small (compared to the total number of local subsystems). On the basis of these results, we develop a subspace identification algorithm that identifies the state-space model of a local subsystem from the local input-output data. Consequently, the proposed algorithm is computationally feasible for interconnected systems with a large number of local subsystems. We numerically illustrate the effectiveness of the new identification algorithm.

The problem of identifying graph topologies of interconnected systems has been studied in [242; 243; 244; 245; 246]. For simplicity, in this chapter we assume that the graph topology, of an interconnected system that we want to identify, is known a priori. Integration of the proposed identification algorithm with the above mentioned graph topology identification algorithms is left for further research.

The chapter is organized as follows. In Section 5.2 we present the problem formulation. In Section 5.3, we postulate the main theorems that we use in Section 5.4 to develop the identification algorithm. In Section 5.5 we present the results of numerical simulations and in Section 5.6 we draw conclusions.

5.2 Problem formulation

For simplicity, the novel identification algorithm will be explained on the example of the global system (3.1). The identification algorithm can be generalized

to large-scale interconnected systems with more general interconnection patterns (see Remark 5.6). For presentation clarity, we rewrite the global state-space model (3.1):

$$\mathcal{S}\left\{\begin{array}{rl} \underline{\mathbf{x}}(k+1) & = \underline{A}\underline{\mathbf{x}}(k)+\underline{B}\underline{\mathbf{u}}(k) \\ \underline{\mathbf{y}}(k) & = \underline{C}\underline{\mathbf{x}}(k)+\underline{\mathbf{n}}(k) \end{array}\right. \tag{5.1}$$

where

$$\underline{A}=\left[\begin{array}{ccccccc} A_{1,1} & E_{1,2} & & & & & \\ E_{2,1} & A_{2,2} & E_{2,3} & & & & \\ & \ddots & & & & & \\ & E_{i,i-1} & A_{i,i} & E_{i,i+1} & & & \\ & & & \ddots & & & \\ & & E_{N-1,N-2} & A_{N-1,N-1} & E_{N-1,N} & & \\ & & & E_{N,N-1} & A_{N,N} & & \end{array}\right] \quad \begin{array}{l} \underline{B}=\text{diag}(B_1,\ldots,B_N) \\ \underline{C}=\text{diag}(C_1,\ldots,C_N) \\ \underline{\mathbf{y}}(k)=\text{col}(\mathbf{y}_1(k),\ldots,\mathbf{y}_N(k)) \\ \underline{\mathbf{x}}(k)=\text{col}(\mathbf{x}_1(k),\ldots,\mathbf{x}_N(k)) \\ \underline{\mathbf{u}}(k)=\text{col}(\mathbf{u}_1(k),\ldots,\mathbf{u}_N(k)) \\ \underline{\mathbf{n}}(k)=\text{col}(\mathbf{n}_1(k),\ldots,\mathbf{n}_N(k)) \end{array} \tag{5.2}$$

where $\underline{B}=\text{diag}(B_1,\ldots,B_N)$ stands for a block diagonal matrix with the matrices $B_1,\ldots,B_N$ on the main diagonal. Like it is explained in Chapter 3, the global system $\mathcal{S}$ is an interconnection of N *local subsystems* $\mathcal{S}_i$:

$$\mathcal{S}_i\left\{\begin{array}{rl} \mathbf{x}_i(k+1) & = A_{i,i}\mathbf{x}_i(k)+E_{i,i-1}\mathbf{x}_{i-1}(k)+E_{i,i+1}\mathbf{x}_{i+1}(k)+B_i\mathbf{u}_i(k) \\ \mathbf{y}_i(k) & = C_i\mathbf{x}_i(k)+\mathbf{n}_i(k) \end{array}\right. \tag{5.3}$$

The set of local subsystems:

$$V_h(S_i)=\{\mathcal{S}_{i-h},\ldots,\mathcal{S}_{i+h}\} \tag{5.4}$$

will be referred to as *the neighborhood of* $\mathcal{S}_i$. The subscript h quantifies the size of $V_h(S_i)$. In order to formulate the identification problem, the concept of *the structure preserving similarity transformation* is introduced.

Definition 5.1 *The structure preserving similarity transformation* $\underline{Q}=diag(Q_1,\ldots,Q_N)$*, transforms the global state-space model* (5.1)-(5.2) *into the following state-space model:*

$$\hat{\mathcal{S}}\left\{\begin{array}{rl} \underline{\hat{\mathbf{x}}}(k+1) & = \underline{\hat{A}}\underline{\hat{\mathbf{x}}}(k)+\underline{\hat{B}}\underline{\mathbf{u}}(k) \\ \underline{\mathbf{y}}(k) & = \underline{\hat{C}}\underline{\hat{\mathbf{x}}}(k)+\underline{\mathbf{n}}(k) \end{array}\right. \tag{5.5}$$

where $\underline{\hat{A}}$ *has block bandwidth equal to 1 (the same sparsity pattern like* $\underline{A}$*),* $\underline{\hat{B}}$ *and* $\underline{\hat{C}}$ *are block diagonal, and* $\underline{\mathbf{x}}(k)=\underline{Q}\underline{\hat{\mathbf{x}}}(k)$*,* $\underline{\hat{A}}=\underline{Q}^{-1}\underline{A}\underline{Q}$*,* $\underline{\hat{B}}=\underline{Q}^{-1}\underline{B}$ *and* $\underline{\hat{C}}=\underline{C}\underline{Q}$*.*

Problem Description 5.1 *Identification problem*

Consider the global system (5.1) *that consists of the interconnection of* N *local subsystems* (5.3)*. Then, using the sequence of the global input-output data* $\{\underline{\mathbf{y}}(k),\underline{\mathbf{u}}(k)\}$*,*

1. *Estimate the order of the local subsystems n.*
2. *Identify the global state-space model* (5.1) *up to a structure preserving similarity transformation.*

5.3 Main theorems

Using the results of Chapter 3, we postulate an approximate state-space model of $\mathcal{S}_i$ in which the local states $\mathbf{x}_{i-1}$ and $\mathbf{x}_{i+1}$ are replaced by the local inputs and outputs. This approximate state-space model is used in Section 5.4 to develop the identification algorithm.

Theorem 5.1 *Let $p \geq \nu$, where ν is the observability index of the global system. Then,*

$$\underline{\mathbf{x}}(k) = \underline{A}^p \mathcal{D} \left(\mathcal{O}_p^T \mathcal{Y}_{k-p}^k - \mathcal{O}_p^T \mathcal{G}_{p-1} \mathcal{U}_{k-p}^{k-1} - \mathcal{O}_p^T \mathcal{N}_{k-p}^k \right) + \mathcal{R}_{p-1} \mathcal{U}_{k-p}^{k-1} \tag{5.6}$$

where $\mathcal{D} = \mathcal{J}_{2p}^{-1}$, $\mathcal{D} \in \mathbb{R}^{Nn \times Nn}$, is the inverse of the finite-time observability Gramian $\mathcal{J}_{2p} = \mathcal{O}_p^T \mathcal{O}_p$.

Proof. From the global data equation (3.10), we have:

$$\mathcal{O}_p \underline{\mathbf{x}}(k-p) = \mathcal{Y}_{k-p}^k - \mathcal{G}_{p-1} \mathcal{U}_{k-p}^{k-1} - \mathcal{N}_{k-p}^k \tag{5.7}$$

Because $p \geq \nu$, from Lemma 3.1 we have: $\text{rank}(\mathcal{O}_p) = Nn$. This implies that $\mathcal{J}_{2p}$ is positive definite and invertible. Because of this, from (5.7) we have:

$$\underline{\mathbf{x}}(k-p) = \mathcal{D} \left(\mathcal{O}_p^T \mathcal{Y}_{k-p}^k - \mathcal{O}_p^T \mathcal{G}_{p-1} \mathcal{U}_{k-p}^{k-1} - \mathcal{O}_p^T \mathcal{N}_{k-p}^k \right) \tag{5.8}$$

Substituting (5.8) in (3.12) we arrive at (5.6). □

In the following theorem, we prove that $\mathcal{D}$ is an exponentially off-diagonally decaying matrix.

Theorem 5.2 *Let $p \geq \nu$, where ν is the observability index of the global system. Then, $\mathcal{D}$ is an exponentially off-diagonally decaying matrix with:*

$$\lambda = \left(\frac{\sqrt{\chi} - 1}{\sqrt{\chi} + 1} \right)^{1/\theta}, \; c = \|\mathcal{D}\|_2 \max \left\{ 1, \frac{(1 + \sqrt{\chi})^2}{2\chi} \right\} \tag{5.9}$$

where θ is the bandwidth[1] of $\mathcal{J}_{2p}$ (θ is proportional to the product pn) and χ is the condition number of $\mathcal{J}_{2p}$ (See Chapter 3).

Proof. Similar to the proof of Lemma 3.2 □

[1]The bandwidth θ is defined by $\theta = m/2$, where m is a constant in Eq. (2.6) in [162].

It should be clear that if the structured observability matrix $\mathcal{O}_p$ is well conditioned, then the finite-time observability Gramian $\mathcal{J}_{2p}$ is also well-conditioned. We introduce the following assumption:

Assumption 5.3 *The finite-time observability Gramian $\mathcal{J}_{2p}$ is well-conditioned.*

As we will prove later, this assumption ensures that the local state of $\mathcal{S}_i$ can be approximated by a linear combination of input and output data that are in a relatively small neighborhood of $\mathcal{S}_i$.

For small λ the off-diagonal decay of $\mathcal{D}$ is rapid. From (5.9) it follows that λ depends on the condition number χ and on the parameter pn (the bandwidth of $\mathcal{J}_{2p}$). Since $\mathcal{J}_{2p}$ is well-conditioned (see Assumption 5.3), $p \ll N$ and $n \ll N$, the parameter λ is small and consequently, the off-diagonal decay of $\mathcal{D}$ is rapid. The importance of this result lies in the fact that off-diagonally decaying matrices, with a rapid off-diagonal decay, can be approximated by sparse banded matrices. In that sense we introduce the following definition (see also Remark 5.5):

Definition 5.2 *[161] Let $\mathcal{D} = [d_{i,j}]$. The matrix $\breve{\mathcal{D}} = [\breve{d}_{i,j}]$ with its elements defined by:*

$$\breve{d}_{i,j} = \begin{cases} d_{i,j} & \text{if } |i-j| \leq s, \\ 0 & \text{if } |i-j| > s, \end{cases} \tag{5.10}$$

is a banded approximation of $\mathcal{D}$.

In the following proposition we give an upper bound on the approximation accuracy.

Proposition 5.4 *Consider the exponentially off-diagonally decaying matrix $\mathcal{D} \in \mathbb{R}^{nN \times nN}$ and its banded approximation $\breve{\mathcal{D}} \in \mathbb{R}^{nN \times nN}$. Then,*

$$\left\| \mathcal{D} - \breve{\mathcal{D}} \right\|_1 < ck_1, \quad k_1 = 2\lambda^{s+1} \frac{1 - \lambda^{Nn-s}}{1 - \lambda} \tag{5.11}$$

and the constants c and λ are defined in (5.9)*. Moreover, the parameter k_1 is an increasing function of χ.*

Proof Let $\mathcal{D} = [d_{i,j}]$ and $\breve{\mathcal{D}} = [\breve{d}_{i,j}]$. Then,

$$\left\| \mathcal{D} - \breve{\mathcal{D}} \right\|_1 = \max_{1 \leq j \leq Nn} \sum_{i=1}^{nN} |d_{i,j} - \breve{d}_{i,j}| < 2 \sum_{k=s+1}^{nN} c\lambda^k = 2c \left(\sum_{k=0}^{Nn} \lambda^k - \sum_{k=0}^{s} \lambda^k \right) = ck_1$$

By checking the sign of $\frac{\partial k_1}{\partial \chi}$, it can be easily proved that k_1 is an increasing function of χ. □

Let us assume that the bandwidth of $\breve{\mathcal{D}}$ is chosen such that $s = nt$, where t is a positive integer. Similarly to the partitioning of $\underline{A}^p$ (see Remark 3.1), we partition

$\breve{\mathcal{D}}$ into N^2 blocks, where each block is of dimension $n \times n$. After this partitioning, $\breve{\mathcal{D}}$ has the block bandwidth equal to t and in the spirit of the notation that is used in this chapter to denote block banded matrices, we will denote this matrix by $\breve{\mathcal{D}}_t$. In the sequel, $\breve{\mathcal{D}}_t$ will be referred to as *the block banded approximation* of $\mathcal{D}$.

From Proposition 5.4 we see that the accuracy of approximating $\mathcal{D}$ by $\breve{\mathcal{D}}$ increases as s increases or equivalently as t increases. Furthermore, we see that the approximation accuracy is better when χ is smaller. Because $\mathcal{J}_{2p}$ is well-conditioned, there exists $s \ll N$ or equivalently $t \ll N$ for which the accuracy of approximating $\mathcal{D}$ by $\breve{\mathcal{D}}_t$ is relatively good [161]. In the sequel it is assumed that $t \ll N$.

By substituting $\mathcal{D}$ with $\breve{\mathcal{D}}_t$ in (5.6), we define an approximation $\breve{\underline{\mathbf{x}}}(k)$ of the global state:

$$\breve{\underline{\mathbf{x}}}(k) = \underline{A}^p \breve{\mathcal{D}}_t \left(\mathcal{O}_p^T \mathcal{Y}_{k-p}^k - \mathcal{O}_p^T \mathcal{G}_{p-1} \mathcal{U}_{k-p}^{k-1} - \mathcal{O}_p^T \mathcal{N}_{k-p}^k \right) + \mathcal{R}_{p-1} \mathcal{U}_{k-p}^{k-1} \tag{5.12}$$

For the sequel we will partition $\breve{\underline{\mathbf{x}}}(k)$ as follows: $\breve{\underline{\mathbf{x}}}(k) = \text{col}(\breve{\mathbf{x}}_1(k), \ldots, \breve{\mathbf{x}}_N(k))$, where $\breve{\mathbf{x}}_i(k) \in \mathbb{R}^n$, $\forall i \in \Pi$. From (5.12) we have that $\breve{\mathbf{x}}_i(k)$ is a linear combination of the local lifted inputs, local lifted outputs and local lifted measurement noises of the local subsystems belonging to the neighborhoods $V_{3p+t-1}(\mathcal{S}_i)$, $V_{2p+t}(\mathcal{S}_i)$ and $V_{2p+t}(\mathcal{S}_i)$, respectively[2]. Because $t \ll N$, these neighborhoods are small. By substituting in (5.3) the local states $\mathbf{x}_{i-1}(k)$ and $\mathbf{x}_{i+1}(k)$ with their approximations, $\breve{\mathbf{x}}_{i-1}(k)$ and $\breve{\mathbf{x}}_{i+1}(k)$, we obtain the approximate, local state-space model:

$$\begin{aligned} \mathbf{x}_i(k+1) &\approx A_{i,i}\mathbf{x}_i(k) + \breve{Q}_i \breve{\Omega}_i + \breve{B}_i^{(3)} \breve{N}_i^{(1)} \\ \mathbf{y}_i(k) &= C_i \mathbf{x}_i(k) + \mathbf{n}_i(k) \end{aligned} \tag{5.13}$$

$$\begin{aligned} \breve{Q}_i &= \begin{bmatrix} \breve{B}_i^{(1)} & \breve{B}_i^{(2)} \end{bmatrix}, \; \breve{\Omega}_i = \begin{bmatrix} \breve{Y}_i^{(1)} \\ \breve{U}_i^{(2)} \end{bmatrix}, \\ \breve{Y}_i^{(1)} &= \text{col}\left(\mathcal{Y}_{i-1-2p-t,k-p}^k, \ldots, \mathcal{Y}_{i+1+2p+t,k-p}^k \right), \\ \breve{U}_i^{(2)} &= \text{col}\left(\mathcal{U}_{i-3p-t,k-p}^{k-1}, \ldots, \mathcal{U}_{i+3p+t,k-p}^{k-1} \right), \\ \breve{N}_i^{(1)} &= \text{col}\left(\mathcal{N}_{i-1-2p-t,k-p}^k, \ldots, \mathcal{N}_{i+1+2p+t,k-p}^k \right) \end{aligned} \tag{5.14}$$

Remark 5.5 *In Chapter 3 we have explicitly computed approximations of off-diagonally decaying matrices using the Chebyshev method or the Newton iteration. In this Chapter we are approximating the off-diagonally decaying matrix $\mathcal{D}$ by truncating its elements that are outside the prescribed bandwidth s. In practice, this approximation cannot be explicitly computed for large-scale systems, simply because to do so, we would first need to directly invert the finite-time observability Gramian $\mathcal{J}_{2p}$. The sole purpose of this approximation is to justify the existence of the state-space model* (5.13).

[2]This follows from the fact that the matrix $\underline{A}^p$ is a sparse block banded matrix with the block bandwidth equal to p. Each block of $\underline{A}^p$ is an $n \times n$ matrix, see Chapter 3.

5.4 Identification algorithm

The main idea of the identification algorithm can be described as follows. First, we use the approximate state-space model (5.13) to estimate the state sequence of the local subsystem $\mathcal{S}_i$. This identification step is repeated for all local subsystems. Because $t \ll N$, the input $\breve{\Omega}_i$ of the state-space model (5.13) contains input-output data of local subsystems that are in a relatively small neighborhood of $\mathcal{S}_i$. Consequently, using the SIMs [58] the state sequence of (5.13) can be efficiently estimated. Furthermore, the computational complexity of estimating the state of (5.13) is independent from the total number of local subsystems N. However, because we do not know the global system in advance, we do not know the precise value of t that determines the form of the input $\breve{\Omega}_i$. As it will be explained in Section 5.4.1, this problem can be solved by choosing several values of t and by computing the Variance Accounted For (VAF) [58] of the identified models.
Let the estimated state sequence of the approximate state-space model (5.13) be denoted by $\{\hat{\mathbf{x}}_i(k)\}$. The state sequence $\{\hat{\mathbf{x}}_i(k)\}$ is approximately related to the "true" state sequence of the local subsystem $\mathcal{S}_i$ via the following transformation:

$$\mathbf{x}_i(k) \approx Q_i \hat{\mathbf{x}}_i(k) \tag{5.15}$$

where Q_i is a square, invertible matrix. We will denote the estimated state-sequences of the local subsystems $\mathcal{S}_{i-1}$ and $\mathcal{S}_{i+1}$, that are estimated on the basis of (5.13), by $\{\hat{\mathbf{x}}_{i-1}(k)\}$ and $\{\hat{\mathbf{x}}_{i+1}(k)\}$, respectively. The state sequences $\{\hat{\mathbf{x}}_{i-1}(k)\}$ and $\{\hat{\mathbf{x}}_{i+1}(k)\}$ are approximately related to the "true" state-sequences of the local subsystems S_{i-1} and S_{i+1} via:

$$\mathbf{x}_{i-1}(k) \approx Q_{i-1}\hat{\mathbf{x}}_{i-1}(k),\ \mathbf{x}_{i+1}(k) \approx Q_{i+1}\hat{\mathbf{x}}_{i+1}(k) \tag{5.16}$$

where Q_{i-1} and Q_{i+1} are invertible matrices. By substituting (5.15) and (5.16) in (3.3), and transforming such state-space model, we obtain:

$$\hat{\mathcal{S}}_i \begin{cases} \hat{\mathbf{x}}_i(k+1) & \approx \underbrace{Q_i^{-1}A_{i,i}Q_i}_{\hat{A}_{i,i}}\hat{\mathbf{x}}_i(k) + \underbrace{Q_i^{-1}E_{i,i-1}Q_{i-1}}_{\hat{E}_{i,i-1}}\hat{\mathbf{x}}_{i-1}(k) + \underbrace{Q_i^{-1}E_{i,i+1}Q_{i+1}}_{\hat{E}_{i,i+1}}\hat{\mathbf{x}}_{i+1}(k) \\ & \quad + \underbrace{Q_i^{-1}B_i}_{\hat{B}_i}\mathbf{u}_i(k) \\ \mathbf{y}_i(k) & \approx \underbrace{C_iQ_i}_{\hat{C}_i}\hat{\mathbf{x}}_i(k) + \mathbf{n}_i(k) \end{cases} \tag{5.17}$$

From (5.17) we see that once the local state sequences are estimated, the local system matrices $\{\hat{A}_{i,i}, \hat{E}_{i,i-1}, \hat{E}_{i,i+1}, \hat{B}_i, \hat{C}_i\}$ can be estimated by solving a least-

squares problem formed on the basis of:

$$\left[\ \hat{\mathbf{x}}_i(k+1)\ \mathbf{y}_i(k)\ \right] \approx \underbrace{\left[\ \hat{A}_{i,i}\ \hat{E}_{i,i-1}\ \hat{E}_{i,i+1}\ \hat{B}_i\ \hat{C}_i\ \right]}_{\text{matrices to be estimated}} \begin{bmatrix} \hat{\mathbf{x}}_i(k) & 0 \\ \hat{\mathbf{x}}_{i-1}(k) & 0 \\ \hat{\mathbf{x}}_{i+1}(k) & 0 \\ \mathbf{u}_i(k) & 0 \\ 0 & \hat{\mathbf{x}}_i(k) \end{bmatrix} + \begin{bmatrix} 0 & \mathbf{n}_i(k) \end{bmatrix} \tag{5.18}$$

We can estimate the local system matrices of other local subsystems in a similar manner. Using the estimates of the local system matrices we can form the estimates $\{\underline{\hat{A}}, \underline{\hat{B}}, \underline{\hat{C}}\}$. Next, from (5.17) we have:

$$\begin{array}{l} \hat{A}_{i,i} \approx Q_i^{-1} A_{i,i} Q_i\ , \hat{E}_{i,i-1} \approx Q_i^{-1} E_{i,i-1} Q_{i-1},\ \hat{E}_{i,i+1} \approx Q_i^{-1} E_{i,i+1} Q_{i+1}, \\ \hat{B}_i \approx Q_i^{-1} B_i,\ \hat{C}_i \approx C_i Q_i \end{array} \tag{5.19}$$

Since (5.17) and (5.19) hold for all $i \in \Pi$, we conclude that $\underline{\mathbf{x}}(k) \approx \underline{Q}\underline{\hat{\mathbf{x}}}(k)$, $\underline{\hat{A}} \approx \underline{Q}^{-1}\underline{A}\underline{Q}$, $\underline{\hat{B}} \approx \underline{Q}^{-1}\underline{B}$, and $\underline{\hat{C}} \approx \underline{C}\underline{Q}$, where $\underline{Q} = \text{diag}(Q_1, \ldots, Q_N)$ is a structure preserving similarity transformation (see Definition 5.1). This shows that the identified model is (approximately) similar to the global state-space model (3.1). We are now ready to formally state the identification algorithm.

Algorithm 5.1 *Identification of the global state-space model* (5.1)

For $i = 1, \ldots, N$ perform steps 1 and 2:
1. Choose the parameters p and t and form the input vector $\breve{\Omega}_i$ of the state space model (5.13).
2. Estimate the local state sequence $\{\mathbf{x}_i(k)\}$ of state space model (5.13) *using the SIM.*
After the steps 1 and 2 are completed, the state sequences $\{\hat{\mathbf{x}}_i(k)\}$, $i = 1, \ldots, N$, are available. For $i = 1, \ldots, N$, perform the following steps:
3. On the basis of (5.18) *form a least-squares problem, and estimate the local system matrices*
$\{\hat{A}_{i,i}, \hat{E}_{i,i-1}, \hat{E}_{i,i+1}, \hat{B}_i, \hat{C}_i\}$.
4. Using the estimates $\{\hat{A}_{i,i}, \hat{E}_{i,i-1}, \hat{E}_{i,i+1}, \hat{B}_i, \hat{C}_i\}$, $i = 1, \ldots, N$, form the global system matrices $\{\underline{\hat{A}}, \underline{\hat{B}}, \underline{\hat{C}}\}$

5.4.1 Comments on the identification algorithm

The theory presented in this thesis predicts that for systems with well-conditioned, finite-time observability Gramians (or equivalently, for systems with well-conditioned structured observability matrices), there should exist relatively small t for which the matrix $\mathcal{D}$ can be approximated by the sparse banded matrix $\mathcal{D}_t$, with a relatively good accuracy. The parameter t needs to be selected in the first step of Algorithm 5.1. This problem can be solved by choosing any $t \ll N$ and by computing the VAF of the identified model. If the VAF value is not high enough, then a new value of t needs to be chosen and identification procedure needs to be re-

peated (usually the new value should be larger than the previous one). This has to be repeated until a relatively high value of VAF of the identified model is reached. By searching for "best" choice of t, we are actually determining the off-diagonal decaying properties of $\mathcal{D}$ directly from the input-output data.
As we show the next section, the form of the input $\breve{\Omega}_i$ can cause ill-conditioning of the data matrices used in the SIM. This is because $\breve{\Omega}_i$ consists of the delayed inputs and outputs of the local subsystems. Some of the outputs might be depending on the past local inputs and the local outputs. This problem can be resolved either by regularizing the data matrices used in the SIM or by eliminating certain outputs and inputs from $\breve{\Omega}_i$. In this thesis, we do not analyze the consistency of the identification algorithm. The consistency analysis left for future research.

Remark 5.6 *Algorithm 5.1 can be generalized for global systems described by sparse, multi-banded, state-space matrices. Like we have shown in Chapter 2, this class of interconnected systems arises in discretization of 3D PDEs using the finite difference method (for more details see for example [200; 228]). Using the lifting technique presented in Chapter 3, it can be easily shown that the finite-time observability Gramian $\mathcal{J}_{2p}$ of this class of systems, is a sparse, multi-banded matrix. In Chapter 3, Section 3.5, it has been shown that inverses of sparse multi-banded matrices can be approximated by sparse multi-banded matrices*[3]*. That is, the inverse of $\mathcal{J}_{2p}$ can be approximated by a sparse multi-banded matrix. From the identification point of view, this implies that the state of a local subsystem can be identified using the local input-output data of local subsystems that are in its 2D neighborhood.*

5.5 Numerical experiments

The data generating model, is a global state-space model that consists of $N = 500$ identical local subsystems. The local system matrices of each local subsystem are given by:

$$A = \begin{bmatrix} 0.5728 & 0.1068 \\ 0.1068 & 0.5728 \end{bmatrix}, E = \begin{bmatrix} 0.1068 & 0 \\ 0 & 0.1068 \end{bmatrix}, B = \begin{bmatrix} 0.2136 \\ 0.1068 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{5.20}$$

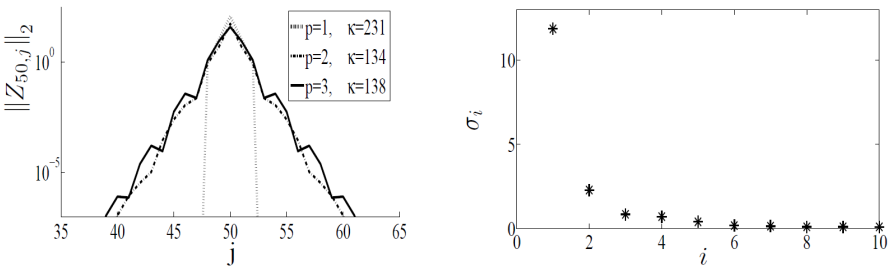
This model has been obtained using the finite difference approximation of the 2D heat equation (see Chapter 2). The local inputs are zero mean normally distributed random signals. The local inputs are mutually independent. We corrupt the outputs of local subsystems by a white noise (signal to noise ratio of each local output is 25 [dB]). In total we perform 100 identification experiments for different realizations of inputs and corresponding outputs of the local subsystems.

For identification of the local state of (5.13), we have used the SIM method summarized in [248]. This SIM is a modified version of the SIM presented in [249]. (we have also tested Predictor Based Subspace IDentification (PBSID) method

[3]For more details on computing sparse approximate inverses of sparse matrices, see for example [161; 205; 207; 221; 247] and references therein).

[250], and we have obtained similar identification results). The SIM presented in [248] has been applied with the past and future windows equal to 15 and 10, respectively. Because all local subsystems have identical local system matrices (5.20), to identify the global state-space model we only need to perform three identification experiments. Namely, on the basis of (5.13) we first estimate the state sequence of the local subsystem $\mathcal{S}_2$. Using the same methodology, we estimate the state sequence of the local subsystem $\mathcal{S}_1$. In the final identification step, we use the state-space model (5.18) (we set $i = 1$ in (5.18)) and the sequences $\{\mathbf{x}_1(k)\}$,$\{\mathbf{x}_2(k)\}$ and $\{\mathbf{y}_1(k), \mathbf{u}_1(k)\}$ to form a least-squares problem. By solving this least-squares problem we estimate the local system matrices $\{\hat{A}, \hat{E}, \hat{B}, \hat{C}\}$. On the basis of $\{\hat{A}, \hat{E}, \hat{B}, \hat{C}\}$ we form $\{\underline{A}, \underline{B}, \underline{C}\}$ and we compute the VAF of the identified model.

In Fig. 5.1(a), we illustrate how the off-diagonal decay of $\mathcal{D}$ depends on p and χ.



(a) (b)

Figure 5.1: (a) The norm of the block elements $Z_{50,j} \in \mathbb{R}^{2\times 2}$ of the 50th block row of $\mathcal{D} = [Z_{50,j}]$. (b) The singular values of the data matrix used to determine the order and to estimate the state-sequence of $\mathcal{S}_2$. The data matrix is formed on the basis of the input 3 in (5.21).

Results presented in Fig. 5.1(a) confirm that for well-conditioned $\mathcal{J}_{2p}$, the off-diagonal decay of $\mathcal{D}$ is rapid. This figure also suggests that the accuracy of approximating $\mathcal{D}$ by $\breve{\mathcal{D}}_t$ is relatively good for $t = 1$. To illustrate how the quality of the identified model depends on the selection of the input vector of the state-space model (5.13) we have estimated the state-sequence of $\mathcal{S}_2$ for 5 different forms of inputs:

$$
\begin{aligned}
&1.\ \breve{\Omega}_2 = \mathbf{u}_2(k)\\
&2.\ \breve{\Omega}_2 = \mathrm{col}(\mathbf{y}_1(k), \mathbf{y}_2(k), \mathbf{y}_3(k), \mathbf{u}_2(k))\\
&3.\ \breve{\Omega}_2 = \mathrm{col}(\mathcal{Y}^k_{1,k-1}, \ldots, \mathcal{Y}^k_{3,k-1}, \mathbf{u}_2(k), \mathbf{u}_1(k-1), \ldots, \mathbf{u}_3(k-1))\\
&4.\ \breve{\Omega}_2 = \mathrm{col}(\mathbf{y}_1(k-1), \ldots, \mathbf{y}_6(k-1), \mathbf{u}_2(k), \mathbf{u}_1(k-1), \ldots, \mathbf{u}_6(k-1))\\
&5.\ \breve{\Omega}_2 = \mathrm{col}(\mathcal{Y}^k_{1,k-1}, \ldots, \mathcal{Y}^k_{6,k-1}, \mathbf{u}_2(k), \mathbf{u}_1(k-1), \ldots, \mathbf{u}_6(k-1))
\end{aligned}
\tag{5.21}
$$

In the case of inputs $3, 4$ and 5, data matrices used to estimate the Markov parameters (impulse response parameters) of the state-space model (5.13) are ill-

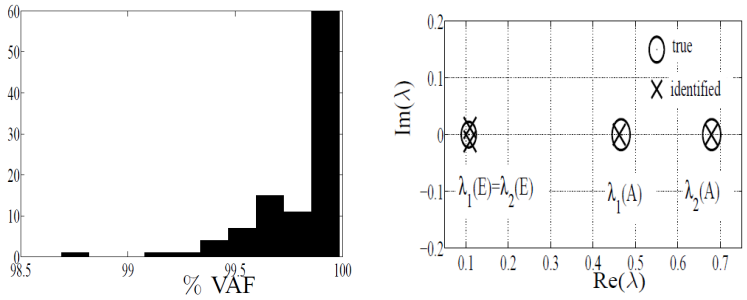
conditioned. This ill-conditioning is caused by the fact that the local outputs, that are the elements of $\breve{\Omega}_2$, are a linear combination of the delayed outputs and inputs. We have used regularization to improve condition number of the data matrix used in the identification of the Markov parameters of $\mathcal{S}_2$ (regularization parameter was 0.05). The order selection is performed by examining the singular values of the data matrix that is formed on the basis of $\{\breve{\Omega}_2, \mathbf{y}_2(k)\}$. For each input in (5.21), we form the data matrix and we select the local order $n = 2$. For illustration, in Fig. 5.1(b) we present the singular values of the data matrix formed on the basis of the input 3 (similar behavior of singular values can be observed for inputs 2, 4 and 5, while in the case of input 1 the state order could not be uniquely determined).

Using the same procedure we estimate the local order n and we estimate the state sequence of $\mathcal{S}_1$. Next, we estimate the local system matrices $\{\hat{A}, \hat{E}, \hat{B}, \hat{C}\}$. Using these local estimates we form $\{\underline{\hat{A}}, \underline{\hat{B}}, \underline{\hat{C}}\}$ and we compute the VAF of the global model. Average values of VAF of $\mathcal{S}_2$ are presented in Table 5.1.

input	1	2	3	4	5
VAF (without regularization)	40 %	99.7 %	5 %	30 %	20 %
VAF (with regularization)	-	99.6 %	97.7 %	99.2 %	98.5 %

Table 5.1: Average VAF of $\mathcal{S}_2$ for different inputs (5.21).

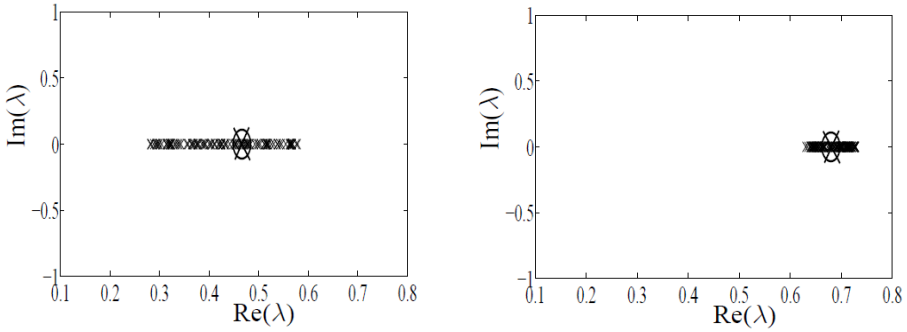
From Table 5.1 it can be concluded that the best VAF is obtained in the case of input 2. This input is formed for $t = 1$ and by eliminating the delayed inputs and outputs that cause ill-conditioning. In Fig. 5.2(a), we present the VAF values for the output of $\mathcal{S}_1$, when the input 2. is used for the identification (similar results are obtained for other local subsystems). The eigenvalues of $\hat{A}$ (when input 2 is used to perform the identification) are given in Fig. 5.3. Next, assuming that the local outputs are not corrupted by the noise, we have performed identification using the input 2 (with regularization). The results are given in Fig. 5.2(b).



(a) (b)

Figure 5.2: (a) Distribution of the VAF of $\mathcal{S}_1$. The identification of the state-sequence of $\mathcal{S}_2$ is performed using input 2, defined in (5.21). (b) Eigenvalues of the estimated matrix $\hat{A}$ and $\hat{E}$, when the input 2, defined in (5.21), is used for identification. The outputs are not corrupted by noise.

As it can be seen from Fig. 5.2(b), in the noise-free scenario we are able to obtain a relatively good identification results. Some of the eigenvalues are biased. This is because we are using an approximate state-space model (5.13) to estimate the local states.



(a) (b)

Figure 5.3: (a) and (b) The distribution of the eigenvalues of $\hat{A}$ for 100 identification experiments (with outputs corrupted by noise). The circle with the big "X" corresponds to the eigenvalue of A. The identification of the state-sequence of $\mathcal{S}_2$ is performed using input 2, defined in (5.21).

5.6 Conclusion

In this chapter we have proposed a decentralized subspace identification algorithm for identifying state-space models of large-scale interconnected systems. In

order to develop this novel identification algorithm, we have proved that the state of the local subsystem can be approximated by a linear combination of the inputs and outputs of the local subsystems that are in its neighborhood. The size of this neighborhood depends on the condition number of the finite-time observability Gramian of the global system. For systems with well-conditioned Gramians, the size of this neighborhood is small. Consequently, we are able to estimate the states and the system matrices of the local subsystems in the computationally efficient manner. We have performed numerical simulations that confirm the effectiveness of the proposed algorithm.

6 CHAPTER

Parameter optimization method for identification of large-scale interconnected systems

We propose a computationally efficient, parameter optimization method for identification of large-scale, interconnected systems described by sparse banded or multi-banded state-space matrices. The identification method consists of two steps. In the first step, impulse response parameters of local subsystems are estimated. In the second step, using the estimated impulse response parameters, the identification problem is formulated as a large-scale, structured, separable least-squares problem. We solve this optimization problem in a computationally efficient manner by approximating inverses of lifted system matrices by sparse banded matrices. In the case of interconnected systems with identical local subsystems, the computational and memory complexities of the proposed identification algorithm scale with $O(N)$, where N is the number of local subsystems. In the case of nonidentical local subsystems, the computational complexity of the identification algorithm is $O(N^2)$. Numerical results illustrate the effectiveness of the proposed identification method.

6.1 Introduction

In Chapter 5 we have explained the main problems of identifying large-scale interconnected systems. In this chapter we will present parameter optimization method for identification of large-scale interconnected systems. First, we briefly explain the need for such an identification method.

A widely used identification approach for low-dimensional systems consists of two steps. In the first identification step, the model of a system is identified using

the Subspace Identification Method (SIM). In the second step, the identified system matrices are used as initial guesses for the Prediction Error Method (PEM). Practice shows that this identification approach results in a relatively good quality of the identified model [58].

In Chapter 5, we have developed the decentralized SIM for the identification of large-scale interconnected systems. Now, the main question is: "*Using the theory presented in Chapter 3, can we develop a computationally efficient, parameter optimization method for improving the quality of the model that is identified using the decentralized SIM?*"

In this Chapter we propose such an algorithm. The identification algorithm consists of two steps. In the first step, the impulse response parameters of local subsystems are identified by solving low-dimensional, last-squares problems. Using the structure preserving lifting technique, the identified local impulse response parameters are used to form an estimate of the global structured impulse response matrix. In the second identification step, the input-output data of local subsystems and the estimate of the global impulse response matrix are used to form a large-scale, nonlinear, structured optimization problem. By approximating inverse functions of lifted system matrices by sparse (multi) banded matrices, we solve this optimization problem, and we identify the state-space model of an interconnected system. In the case of interconnected systems with identical local subsystem the computational complexity of the proposed identification algorithm is $O(N)$. In the case of nonidentical local subsystems, the computational complexity of the identification algorithm is $O(N^2)$. Numerical results illustrate the effectiveness of the proposed identification method.

The chapter is organized as follows. In Section 6.2, the problem formulation is presented. In Section 6.3, the identification algorithm is presented. The numerical simulations are performed in Section 6.4 and conclusions are drawn in Section 6.5.

6.2 Problem formulation

For the sake of presentation clarity, we will present a new identification method assuming that the local subsystems of an interconnected system are identical and that they are connected in a string. The identification algorithm proposed in this chapter can be generalized for large-scale interconnected systems with sparse banded or sparse multi-banded matrices. We consider an interconnected system obtained by finite difference approximation of the 2D heat equation (see Chapter 2):

$$\mathcal{S}\left\{\begin{array}{rl} \underline{\mathbf{x}}(k+1) & = \underline{A\mathbf{x}}(k)+\underline{B\mathbf{u}}(k) \\ \underline{\mathbf{y}}(k) & = \underline{C\mathbf{x}}(k)+\underline{\mathbf{n}}(k) \end{array}\right. \tag{6.1}$$

where the system matrices have the following structure:

$$\underline{A} = \begin{bmatrix} A & E & & \\ E & A & E & \\ & \ddots & \ddots & \ddots \\ & E & A & E \\ & & E & A \end{bmatrix}, \underline{B} = \begin{bmatrix} B & & \\ & \ddots & \\ & & B \end{bmatrix}, \underline{C} = \begin{bmatrix} C & & \\ & \ddots & \\ & & C \end{bmatrix},$$
$$\underline{\mathbf{y}}(k) = \begin{bmatrix} \mathbf{y}_1(k) \\ \vdots \\ \mathbf{y}_N(k) \end{bmatrix}, \underline{\mathbf{x}}(k) = \begin{bmatrix} \mathbf{x}_1(k) \\ \vdots \\ \mathbf{x}_N(k) \end{bmatrix}, \underline{\mathbf{u}}(k) = \begin{bmatrix} \mathbf{u}_1(k) \\ \vdots \\ \mathbf{u}_N(k) \end{bmatrix} \tag{6.2}$$

and similarly we define $\underline{\mathbf{n}}(k)$. The global system $\mathcal{S}$ consists of interconnection of N identical *local subsystems* $\mathcal{S}_i$:

$$\mathcal{S}_i \begin{cases} \mathbf{x}_i(k+1) &= A\mathbf{x}_i(k) + E\mathbf{x}_{i-1}(k) + E\mathbf{x}_{i+1}(k) + B\mathbf{u}_i(k) \\ \mathbf{y}_i(k) &= C\mathbf{x}_i(k) + \mathbf{n}_i(k) \end{cases} \tag{6.3}$$

For simplicity, we introduce the following assumption:

Assumption 6.1

- *Measurement noises* $\mathbf{n}_i(k)$ *of all local subsystems* (3.3)*, are white Gaussian and independent from each other.*
- *The global system* (6.1) *is asymptotically stable.*

In order to formulate the identification problem, we introduce the concept of *the structure preserving similarity transformation*.

Definition 6.1 *A structure preserving similarity transformation is a non-singular matrix* $T \in \mathbb{R}^{nN \times nN}$*, that by a transformation* $\underline{\mathbf{x}}(k) = \underline{T}\underline{\hat{\mathbf{x}}}(k)$*, transforms the global state-space model* (6.1)-(6.2) *into:*

$$\hat{\mathcal{S}} \begin{cases} \hat{\mathbf{x}}(k+1) &= \underline{\hat{A}}\hat{\mathbf{x}}(k) + \underline{\hat{B}}\underline{\mathbf{u}}(k) \\ \underline{\mathbf{y}}(k) &= \underline{\hat{C}}\hat{\mathbf{x}}(k) + \underline{\mathbf{n}}(k) \end{cases} \tag{6.4}$$

where

$$\underline{\hat{A}} = \begin{bmatrix} \hat{A} & \hat{E} & & \\ \hat{E} & \hat{A} & \hat{E} & \\ & \ddots & \ddots & \ddots \\ & \hat{E} & \hat{A} & \hat{E} \\ & & \hat{E} & \hat{A} \end{bmatrix}, \underline{\hat{B}} = \begin{bmatrix} \hat{B} & & \\ & \ddots & \\ & & \hat{B} \end{bmatrix}, \underline{\hat{C}} = \begin{bmatrix} \hat{C} & & \\ & \ddots & \\ & & \hat{C} \end{bmatrix}, \tag{6.5}$$

Problem Description 6.1 *Identification problem*

Consider the global system (6.1)-(6.2)*, and assume that the order* n *of the local system* $\mathcal{S}_i$ *is known a priori (see Remark 6.1). From the set of the global input-output data* $\{\underline{\mathbf{y}}(k), \underline{\mathbf{u}}(k)\}$*, identify the global state-space model* (6.1)-(6.2) *up to a structure preserving similarity transformation. The identification should be performed with* $O(N)$ *complexity and* $O(N)$ *memory requirements (see Remark 6.2).*

Remark 6.1 *For simplicity we have assumed that the order of the local subsystems* n *is known a priori. The order selection algorithms, that are used in the prediction error and output error identification methods [58; 141], can be easily incorporated in the identification algorithms proposed in this paper. Other option is to estimate the order of the local subsystems using the decentralized SIM proposed in Chapter 5.*

Remark 6.2 *It can be easily shown that by exploiting the sparsity of the global state-space model* (6.1)*, the computational complexity of the PEM and the output error methods [58; 141] can be reduced to* $O(N^3)$*. However,* $O(N^3)$ *computational complexity is still high, keeping in mind that the number of local subsystems* N *is very large.*

6.3 Identification algorithm

The identification algorithm consists of the two steps. In the first step, the structured impulse response matrix $\mathcal{G}_{p-1}$ (see Chapter 3) is identified by identifying the local impulse response parameters. In the second identification step, using the identified local impulse response parameters and the global data equation (3.10), a large-scale optimization problem is formed. The local system matrices are identified by solving this optimization problem in a computationally efficient manner.

6.3.1 Identification of local impulse response parameters

The theoretical framework developed in Chapter 3, can be straightforwardly applied to the global system (6.2) composed of identical local subsystems (6.3). Furthermore, the same notation can be used to denote the lifted system matrices.

From the output equation of the local subsystem (6.3) it follows:

$$\begin{aligned}\mathbf{y}_i(k) = &\sum_{j=-l}^{l} T_{i,i+j}^{(l)} \mathbf{x}_{i+j}(k-l) \\ &+ \sum_{g=-l+1}^{l-1} \sum_{s=0}^{l-1-|g|} H_{i,i+g}^{(l-1-s)} \mathbf{u}_{i+g}(k-l+s) + \mathbf{n}_i(k)\end{aligned} \tag{6.6}$$

where the matrices $T_{i,i+j}^{(l)}$ and $H_{i,i+g}^{(l-1-s)}$ are defined in (3.18). The matrix $H_{i,i+g}^{(l-1-s)}$ is called the local impulse response parameter of $\mathcal{S}_i$.

We assume that the parameter l in (6.6) is chosen such that[1] $l > p$ and $l \ll N$. Because by Assumption 6.1 the global system is asymptotically stable, for sufficiently large l the effect of initial local states in (6.6) can be neglected:

$$\mathbf{y}_i(k) \approx \sum_{g=-l+1}^{l-1} \sum_{s=0}^{l-1-|g|} H_{i,i+g}^{(l-1-s)} \mathbf{u}_{i+g}(k-l+s) + \mathbf{n}_i(k) \tag{6.7}$$

Furthermore, since the local subsystems are identical, for $l \geq 2$ and $i = l-1, \ldots, N-l+2$, we have:

$$H_{i,i+s}^{(f)} = H_{i,i-s}^{(f)}, s = 0, \ldots, l-1, f = s, \ldots, l-1 \tag{6.8}$$

On the basis of (6.7) and (6.8), a low-dimensional least-square problem can be formed and the local impulse response parameters of $\mathcal{S}_i$ can be estimated.

Because local subsystems are identical, we have that the local impulse response parameters of $\mathcal{S}_i$ are at the same time impulse response parameters of the local subsystems $\mathcal{S}_j$, $j \in \{l-1, \ldots, N-l+2\}$ (this property does not hold for nonidentical local subsystems, see Remark 6.2).
To complete the estimation, we need to estimate the impulse response parameters of the remaining local subsystems: $\mathcal{S}_j$, $j = 1, \ldots, l-2, N-l+3, \ldots, N$. This can be achieved by forming equations similar to (6.7) for $j = 1, \ldots, l-2, N-l+3, \ldots, N$, and by solving corresponding low-dimensional least-squares problems.

Since by assumption the local measurement noise of each local subsystem is a white noise (see Assumption 6.1), the consistency analysis of the linear regression problem, that is formed on the basis of (6.7) and (6.8), is simple and it has been extensively studied in the literature [141]. What needs to be emphasized here is that the estimates of the local impulse response parameters are biased. These estimates are biased because the local initial states are neglected in (6.7). The bias can be significantly reduced by choosing larger values of l.

Once we have estimated local impulse parameters of all local subsystems, using (3.15) and (3.17) we can form an estimate of the structured impulse response matrix $\mathcal{G}_{p-1}$. Since $l > p$, to form an estimate of $\mathcal{G}_{p-1}$ we only need local impulse response parameters that correspond to the lifting window p. The local impulse response parameters corresponding to $p+1, \ldots, l$ are not used for identification.

Remark 6.2 *In the case of nonidentical local subsystems, the impulse response parameters of all local subsystems are different. In this case, the impulse response parameters of all local subsystems have to be estimated separately.*

6.3.2 The global nonlinear optimization problem

Let an estimate of $\mathcal{G}_{p-1}$ be denoted by $\hat{\mathcal{G}}_{p-1}$. In the global data equation (3.10), we can substitute the unknown matrix $\mathcal{G}_{p-1}$ by its estimate $\hat{\mathcal{G}}_{p-1}$. On the basis of this

[1] In practice, the parameter l should be selected to be much larger than p. For example in numerical experiments section we chose $l = 10$ and $p = 2$.

substitution, we define:

$$\hat{\mathcal{D}}_{k-p}^{k} = \mathcal{Y}_{k-p}^{k} - \hat{\mathcal{G}}_{p-1}\mathcal{U}_{k-p}^{k-1} \tag{6.9}$$

Since $p < l$, the effect of the initial state in the global data equation (3.10) cannot be neglected. This means that from (3.10) and (6.9) we have:

$$\hat{\mathcal{D}}_{k-p}^{k} \approx \mathcal{O}_p \underline{\mathbf{x}}(k-p) + \mathcal{N}_{k-p}^{k} \tag{6.10}$$

From (6.10) *it can be concluded that the global system* (3.1) *can be identified by decomposing the vector* $\hat{\mathcal{D}}_{k-p}^{k}$ *into the product of a banded matrix and the global state vector. To the best of our knowledge, a computationally efficient algorithm for performing this structured decomposition has not been developed yet.*

To perform this decomposition, and consequently to identify the local subsystems, we will introduce a parametrization of the local system matrices. Throughout the remainder of this chapter, the parametrized local system matrices will be denoted by $A(\boldsymbol{\alpha})$, $E(\boldsymbol{\alpha})$ and $C(\boldsymbol{\alpha})$, where $\boldsymbol{\alpha}$ is a parametrization vector. We assume that $\boldsymbol{\alpha}$ is an element of an open set Ω. For simplicity, we assume that the local system matrices are fully parameterized by $\boldsymbol{\alpha}$. This means that $\Omega \subset \mathbb{R}^{n(2n+r)}$.
As it will be explained later, the estimate of the matrix B, denoted by $\hat{B}$, will be determined using the estimates of A, E and C. The family of the global systems, that are defined by the local system matrices $A(\boldsymbol{\alpha})$, $E(\boldsymbol{\alpha})$, $C(\boldsymbol{\alpha})$ and by a matrix $\hat{B}$, is denoted by $\mathcal{S}(\boldsymbol{\alpha})$. Using this parameterization and using (3.14), (3.16) and (3.18), we form the matrix $\mathcal{O}_p(\boldsymbol{\alpha})$. The local system matrices are identified by solving *the global optimization problem*:

$$\min_{\boldsymbol{\alpha}, \underline{\mathbf{x}}(k-p)} \left\{ \left\| \hat{\mathcal{D}}_{k-p}^{k} - \mathcal{O}_p(\boldsymbol{\alpha})\underline{\mathbf{x}}(k-p) \right\|_2^2 + \mu \left\| \underline{\mathbf{x}}(k-p) \right\|_2^2 \right\} \tag{6.11}$$

where $\mu \geq 0$ is a regularization parameter. As it will be explained later, the parameter μ will play crucial role in establishing computationally efficient identification algorithm. For convenience, we will write the optimization problem (6.11) in the following form:

$$\min_{\boldsymbol{\alpha}, \underline{\mathbf{x}}(k-p)} \left\| \mathcal{V} - \mathcal{Z}(\boldsymbol{\alpha})\underline{\mathbf{x}}(k-p) \right\|_2^2 \tag{6.12}$$

where

$$\mathcal{V} = \begin{bmatrix} \hat{\mathcal{D}}_{k-p}^{k} \\ 0 \end{bmatrix} \qquad \mathcal{Z}(\boldsymbol{\alpha}) = \begin{bmatrix} \mathcal{O}_p(\boldsymbol{\alpha}) \\ \sqrt{\mu} I \end{bmatrix} \tag{6.13}$$

In this thesis we are mainly focused on the computational aspects of solving the optimization problem (6.12). The global optimization problem is nonlinear and possibly non-convex. We do not analyze the bias and the variance of the solution of (6.12). This analysis is left for future research.

Remark 6.3 *In order to formulate the global optimization problem* (6.11)*, for the case of interconnected systems with nonidentical local subsystems, the local system matrices of*

all local subsystems need to be parameterized. Since we are considering interconnected systems described by sparse block banded matrices, in the case of nonidentical local subsystems the parameterization vector $\boldsymbol{\alpha}$ will have cN elements, where c is a constant and $c \ll N$.

6.3.3 The Separable-Least Squares form of the global optimization problem (6.11)

Using the the Separable Least Squares (SLS) technique [251], we can eliminate the global state from the global optimization problem (6.11). This way, we significantly reduce the number of its optimization variables.

Theorem 6.4 *Consider the optimization problem* (6.12) *and let ν be the observability index of the global system* (6.1)*. Assume that any of the following two conditions are fulfilled:*

1. *The observability index of $\mathcal{S}(\boldsymbol{\alpha})$ is equal to ν, $\forall \boldsymbol{\alpha} \in \Omega$, and the parameter p satisfies: $p \geq \nu$*
2. *The parameter μ satisfies: $\mu > 0$.*

Let $\hat{\boldsymbol{\alpha}}$ be a local minimizer of

$$\min_{\boldsymbol{\alpha}} \left\| \left(I - \mathcal{Z}(\boldsymbol{\alpha}) \mathcal{Z}^{\dagger}(\boldsymbol{\alpha}) \right) \mathcal{V} \right\|_2^2 \tag{6.14}$$

and let

$$\underline{\hat{\mathbf{x}}}(k-p) = \mathcal{Z}^{\dagger}(\hat{\boldsymbol{\alpha}}) \mathcal{V} \tag{6.15}$$

where $\dagger$ denotes the pseudo-inverse. Then $\hat{\boldsymbol{\alpha}}$ and $\underline{\hat{\mathbf{x}}}(k-p)$ are the local minimizers of (6.12)*, or equivalently of* (6.11)*.*

Proof. If the condition condition 1. is fulfilled, then from Lemma 3.1 we have that: $\text{rank}(\mathcal{O}_p(\boldsymbol{\alpha})) = nN, \forall \boldsymbol{\alpha} \in \Omega$. This implies that: $\text{rank}(\mathcal{Z}(\boldsymbol{\alpha})) = nN, \forall \boldsymbol{\alpha} \in \Omega$. On the other hand, if $\mu > 0$, then independently from p we have that $\text{rank}(\mathcal{Z}(\boldsymbol{\alpha})) = nN$. Because $\mathcal{Z}(\boldsymbol{\alpha})$ has full column rank $\forall \boldsymbol{\alpha} \in \Omega$, the conditions of Theorem 2.1, presented in [251], are fulfilled. This theorem states that the optimization problem (6.12) can be transformed into (6.14) using the following two steps. First, the optimization problem (6.12) is formally solved with respect to $\underline{\mathbf{x}}(k-p)$ (assuming that $\boldsymbol{\alpha}$ is constant). This formal solution is given by $\underline{\mathbf{x}}(k-p) = \mathcal{Z}^{\dagger}(\boldsymbol{\alpha}) \mathcal{V}$. Secondly, this solution is substituted back in (6.12) to obtain (6.14). To summarize, since $\text{rank}(\mathcal{Z}(\boldsymbol{\alpha})) = nN$, from Theorem 2.1 presented in [251], it follows that the pair $(\hat{\boldsymbol{\alpha}}, \underline{\hat{\mathbf{x}}}(k-p))$ is the local minimizer of (6.12), or equivalently of (6.11). □

Since we do not know the model of the global system is advance, we do not know the observability index of the global system ν. This means that for a chosen value of p we cannot check the condition (1) of Theorem 6.4. However, by choosing

$\mu > 0$, we can ensure that the condition (2) of Theorem 6.4 is satisfied, and that we can solve (6.14), instead of (6.12).

The advantage of solving (6.14) instead of (6.11) is evident. Instead of solving the optimization problem (6.11) involving $n(2n + r) + Nn$ parameters (where N is much larger than n and r), we solve the optimization problem (6.14) involving only $n(2n + r)$ parameters (see Remark 6.3). At a first glance, the disadvantage is that the sparsity of the original optimization problem (6.11) is destroyed by transforming it using the SLS strategy. That is, although $\mathcal{O}_p(\boldsymbol{\alpha})$ and $\mathcal{Z}(\boldsymbol{\alpha})$ are sparse block banded matrices, the matrices $\mathcal{O}_p^{\dagger}(\boldsymbol{\alpha})$ and $\mathcal{Z}^{\dagger}(\boldsymbol{\alpha})$ are fully populated for an arbitrary value of $\boldsymbol{\alpha}$. However, using the methods summarized in Chapter 3, we can approximate these pseudo-inverses in a computationally efficient manner.
To the best of our knowledge, the solution of the optimization problem (6.14) cannot be expressed in the analytic form. Consequently, the solution must be found using iterative numerical methods. A large variety of iterative methods for solving nonlinear optimization problems have been developed in the past, see for example [252]. In this thesis we use the steepest descent method. Other methods, like for example Gauss-Newton method, can also be used to solve (6.14). In order to solve (6.14) using the steepest descent method, we will parameterize matrices $\mathcal{J}_{2p} \in \mathbb{R}^{Nn \times Nn}$ and $\mathcal{F}_{2p} \in \mathbb{R}^{Nn \times Nn}$, defined in (3.26) and (3.36), respectively. The parametrized matrices are denoted by:

$$\mathcal{J}_{2p}(\boldsymbol{\alpha}) = \mathcal{O}_p^T(\boldsymbol{\alpha})\mathcal{O}_p(\boldsymbol{\alpha}), \qquad \mathcal{F}_{2p}(\boldsymbol{\alpha}) = \mu I + \mathcal{J}_{2p}(\boldsymbol{\alpha}) \tag{6.16}$$

In order to compute the Jacobian matrix of the optimization problem (6.14), we define [251]:

$$\mathcal{P}_Z(\boldsymbol{\alpha}) = \mathcal{Z}(\boldsymbol{\alpha})\mathcal{Z}^{\dagger}(\boldsymbol{\alpha}), \;\; \mathcal{P}_Z^{\perp}(\boldsymbol{\alpha}) = I - \mathcal{P}_Z(\boldsymbol{\alpha}), \;\; \mathbf{r}(\boldsymbol{\alpha}) = \mathcal{P}_Z^{\perp}(\boldsymbol{\alpha})\mathcal{V} \tag{6.17}$$

Let α_h be the h^{th} element of $\boldsymbol{\alpha}$ ($h = 1, \ldots, n(2n + r)$). Then, we can write the h^{th} block column of the Jacobian J, associated with the cost (6.14), as follows [251]:

$$\begin{aligned} J_h(\boldsymbol{\alpha}) &= -\frac{\partial \mathbf{r}(\boldsymbol{\alpha})}{\partial \alpha_h} = L_h(\boldsymbol{\alpha})\mathcal{V} \\ L_h(\boldsymbol{\alpha}) &= \left(\mathcal{P}_Z^{\perp}(\boldsymbol{\alpha}) \frac{\partial \mathcal{Z}(\boldsymbol{\alpha})}{\partial \alpha_h} \mathcal{Z}^{\dagger}(\boldsymbol{\alpha}) + (\mathcal{P}_Z^{\perp}(\boldsymbol{\alpha}) \frac{\partial \mathcal{Z}(\boldsymbol{\alpha})}{\partial \alpha_h} \mathcal{Z}^{\dagger}(\boldsymbol{\alpha}))^T \right) \end{aligned} \tag{6.18}$$

It can be easily shown that:

$$\mathcal{Z}^{\dagger}(\boldsymbol{\alpha}) = \mathcal{F}_{2p}^{-1}(\boldsymbol{\alpha}) \begin{bmatrix} \mathcal{O}_p^T(\boldsymbol{\alpha}) & \sqrt{\mu} I \end{bmatrix} \tag{6.19}$$

$$\mathcal{Z}(\boldsymbol{\alpha})\mathcal{Z}^{\dagger}(\boldsymbol{\alpha}) = \begin{bmatrix} \mathcal{O}_p \mathcal{F}_{2p}^{-1}(\boldsymbol{\alpha}) \mathcal{O}_p^T(\boldsymbol{\alpha}) & \sqrt{\mu} \mathcal{O}_p(\boldsymbol{\alpha}) \mathcal{F}_{2p}^{-1} \\ \sqrt{\mu} \mathcal{F}_{2p}^{-1}(\boldsymbol{\alpha}) \mathcal{O}_p^T(\boldsymbol{\alpha}) & \mu \mathcal{F}_{2p}^{-1}(\boldsymbol{\alpha}) \end{bmatrix} \tag{6.20}$$

$$\frac{\partial \mathcal{Z}(\boldsymbol{\alpha})}{\partial \alpha_h} = \begin{bmatrix} \frac{\partial \mathcal{O}_p(\boldsymbol{\alpha})}{\partial \alpha_h} \\ 0 \end{bmatrix} \tag{6.21}$$

The matrix $\frac{\partial \mathcal{O}_p(\boldsymbol{\alpha})}{\partial \alpha_h}$ is a block banded matrix, with the block bandwidth equal to p. The elements of $\frac{\partial \mathcal{O}_p(\boldsymbol{\alpha})}{\partial \alpha_h}$ can be computed recursively. Namely, because $\underline{A}^p =$

$\underline{AA}^{p-1}$ and using the product rule for differentiation, it can be easily shown that:

$$\begin{aligned}\frac{\partial L_{i,j}^{(p)}}{\partial \alpha_h} =& \frac{\partial A}{\partial \alpha_h} L_{i,j}^{(p-1)} + A\frac{\partial L_{i,j}^{(p-1)}}{\partial \alpha_h} + \frac{\partial E}{\partial \alpha_h} L_{i-1,j}^{(p-1)} + E\frac{\partial L_{i-1,j}^{(p-1)}}{\partial \alpha_h} \\ &+ \frac{\partial E}{\partial \alpha_h} L_{i+1,j}^{(p-1)} + E\frac{\partial L_{i+1,j}^{(p-1)}}{\partial \alpha_h}\end{aligned} \tag{6.22}$$

Using this recursion we can easily compute any element of $\frac{\partial \mathcal{O}_p(\boldsymbol{\alpha})}{\partial \alpha_h}$. The basic steepest descent method, for solving the optimization problem (6.14), consists of the following iteration:

$$\boldsymbol{\alpha}^{(k+1)} = \boldsymbol{\alpha}^{(k)} - aJ^T(\boldsymbol{\alpha}^{(k)}) \left(I - \mathcal{Z}(\boldsymbol{\alpha}^{(k)}) \mathcal{Z}^{\dagger}(\boldsymbol{\alpha}^{(k)}) \right) \mathcal{V} \tag{6.23}$$

where $a > 0$ is a positive constant. Throughout the remainder of the chapter, the converged value of $\boldsymbol{\alpha}^{(k)}$ will be denoted by $\hat{\boldsymbol{\alpha}}$.
Once we have obtained $\hat{\boldsymbol{\alpha}}$, we can determine the matrix $\hat{B}$ using two strategies. For example, from the identified local system matrices $C(\hat{\boldsymbol{\alpha}})$, $A(\hat{\boldsymbol{\alpha}})$, $E(\hat{\boldsymbol{\alpha}})$ and from the estimated impulse response parameters, we can determine $\hat{B}$ by solving a linear system of equations. Other strategy is to use (6.15) to determine the global state:

$$\hat{\underline{\mathbf{x}}}(k-p+s) = \mathcal{Z}^{\dagger}(\hat{\boldsymbol{\alpha}})\mathcal{V} \tag{6.24}$$

for $s = k, k+1$. From these global state estimates, we can extract the estimates of the local states. Then, on the basis of the local state-space models (6.3) we can form a linear regression problem in which the local states are substituted by the estimates of the local states, and the local system matrices A, C, E are substituted by $C(\hat{\boldsymbol{\alpha}}), A(\hat{\boldsymbol{\alpha}}), E(\hat{\boldsymbol{\alpha}})$. By solving this linear regression problem we can estimate $\hat{B}$.

From (6.17)-(6.24)*, we see that the main computational bottleneck of one iteration of the steepest descent method is the inversion of* $\mathcal{F}_{2p}^{-1}$. Namely, although $\mathcal{F}_{2p}(\boldsymbol{\alpha})$ is a sparse block banded matrix, its inverse $\mathcal{F}_{2p}^{-1}(\boldsymbol{\alpha})$ is a dense matrix. Since we need N^2 locations to store a dense matrix in a computer memory, for large N the matrix $\mathcal{F}_{2p}^{-1}(\boldsymbol{\alpha})$ cannot be memorized. Because the matrix $\mathcal{F}_{2p}(\boldsymbol{\alpha})$ is a sparse banded matrix, we can compute its inverse with $O(N^2)$ complexity [112]. However, since $\mathcal{F}_{2p}^{-1}(\boldsymbol{\alpha})$ is fully populated, the sparsity of the optimization problem is lost and the computational cost of the subsequent operations, that involve $\mathcal{F}_{2p}^{-1}(\boldsymbol{\alpha})$, are either $O(N^2)$ or $O(N^3)$. The above explained computational problems are illustrated in Figure 6.1 in numerical experiments section.
However, in Chapter 3 we have developed computationally, sparsity preserving methods for approximating $\mathcal{F}_{2p}^{-1}(\boldsymbol{\alpha})$. By using these methods we can compute one iteration the steepest descent method (6.23) with $O(N)$ complexity and $O(N)$ memory requirements. The identification algorithm that is based on the approximation of $\mathcal{F}_{2p}^{-1}$ using the Chebyshev method or Newton iteration is summarized below.

Algorithm 6.1 *Identification of the global state-space model* (6.1)

1. *Chose the parameter l and estimate the local impulse response parameters of all local subsystems S_j.*
2. *Choose $p < l$, and from the elements of the estimated local impulse response matrices (that correspond to p) and using* (3.17) *and* (3.15)*, form an estimate of the global impulse response matrix $\hat{\mathcal{G}}_{p-1}$. Using $\hat{\mathcal{G}}_{p-1}$ and using the input-output data, form the vector $\hat{\mathcal{D}}_{k-p}^{k}$ defined in* (6.9).
3. *Select an initial value of $\boldsymbol{\alpha}$ and parameter μ (see Section 6.3.4) and solve the optimization problem* (6.14). *In each iteration of the steepest descent method* (6.23) *approximate the matrix $\mathcal{F}_{2p}^{-1}(\boldsymbol{\alpha}^k)$ using the Chebyshev approximation method or the Newton iteration.*
4. *Use the sparse approximation of $\mathcal{F}_{2p}^{-1}(\boldsymbol{\alpha}^k)$ and* (6.19) *to approximate $\mathcal{Z}^{\dagger}(\hat{\boldsymbol{\alpha}})$ by a sparse matrix and estimate the global state $\hat{\underline{\mathbf{x}}}(k-p)$ using* (6.15). *Estimate the local matrix B.*

6.3.4 Some guidelines for selecting the parameter μ and initial guess $\alpha^{(0)}$

The following conclusions about the influence of the parameter μ on the identification algorithm can be easily drawn:

1. From the cost function of the optimization problem (6.11), it follows that by increasing μ we put more emphasis on the penalization of the global state-vector, and less emphasis on the penalization of the part of the cost function that involves input-output data. This indicates that for very large μ we only minimize the initial state, while we do not minimize the part of the cost function that involves the vector of parameters $\boldsymbol{\alpha}$. That is, very large μ has a negative effect on the identification quality.

2. The vector $\mathcal{V}$, defined in (6.13), contains the vector $\hat{\mathcal{D}}_{k-p}^{k}$, which is defined in (6.9). The vector $\hat{\mathcal{D}}_{k-p}^{k}$ contains the errors of estimating the structured impulse response matrix $\hat{\mathcal{G}}_{p-1}$ and measurement errors. Since μ is a regularization parameter of $\mathcal{F}_{2p}$, from (6.18) it follows that by selecting sufficiently large μ we can make the steepest descent method to be less sensitive to the errors that are present in $\hat{\mathcal{D}}_{k-p}^{k}$.

On the other hand, the parameter μ has an important role in establishing the computationally efficient method for approximating $\mathcal{F}_{2p}^{-1}$. In summary (see Chapter 3), with the parameter μ we can influence the condition number of $\mathcal{F}_{2p}$, and consequently, we can ensure that its inverse can be approximated with $O(N)$ complexity. Moreover, if the matrix $\mathcal{J}_{2p}$ is well-conditioned, then $\mathcal{F}_{2p}^{-1}$ can be approximated with a relatively good accuracy and $O(N)$ complexity, by a sparse banded matrix. If the matrix $\mathcal{J}_{2p}$ is ill-conditioned, then to approximate $\mathcal{F}_{2p}^{-1}$ with a good accuracy and $O(N)$ complexity, we need to choose an appropriate value of the parameter μ. If the approximation is performed using the Chebyshev method, this value should be chosen on the basis of Theorem 3.4 (choosing μ such that the upper bound (3.69) is small).

Before we start the steepest descent optimization algorithm we have to choose an initial guess $\boldsymbol{\alpha}^{(0)}$ of the optimization variable $\boldsymbol{\alpha}$, and we have to choose μ. In Numerical experiments section, we select $\boldsymbol{\alpha}^{(0)}$ by drawing samples from the normal distribution. Another option is to form $\boldsymbol{\alpha}^{(0)}$ from the elements of the local system matrices that are identified using the decentralized SIM proposed in Chapter 3.

Let us suppose that we have computed the condition number of $\mathcal{J}_{2p}(\boldsymbol{\alpha}^{(0)})$, and on the basis of this condition number we have chosen the parameter μ such that the matrix $\mathcal{F}_{2p}(\boldsymbol{\alpha}^{(0)})$ is well-conditioned. In each iteration k of the steepest descent method (6.23), $\boldsymbol{\alpha}^{(k)}$ changes. This means that in each iteration, the condition numbers of $\mathcal{J}_{2p}(\boldsymbol{\alpha}^{(k)})$ and $\mathcal{F}_{2p}(\boldsymbol{\alpha}^{(k)})$ change. It is possible that during convergence of the steepest descent method, $\boldsymbol{\alpha}^{k}$ becomes such that for selected μ the matrix $\mathcal{F}_{2p}(\boldsymbol{\alpha}^{(k)})$ becomes badly conditioned. In this case, a sparse banded matrix is no longer a good approximation of $\mathcal{F}_{2p}(\boldsymbol{\alpha}^{(k)})$ and consequently, the Jacobian matrix contains significant numerical errors. These errors might slow down the convergence of the steepest descent method, or more extremely, they might cause divergence (this situation is illustrated in Fig. 8.3(b)). This problem can be resolved using any of the following two strategies:

1. The first strategy is to select sufficiently large μ before we start the steepest descent method, and to keep this value constant during the optimization. Large μ will ensure that the condition number of $\mathcal{F}_{2p}\left(\boldsymbol{\alpha}^{(k)}\right)$ does not change significantly with the change of $\boldsymbol{\alpha}^{(k)}$. However, as we illustrate in Numerical experiments section (see Fig. 8.3(a)), by increasing μ we slow down the convergence of the steepest descent method. Furthermore, large μ might have a negative impact of the identification quality.

2. The second strategy is to compute the condition number of $\mathcal{J}_{2p}(\boldsymbol{\alpha}^{k})$ in each iteration of the steepest descent method. Then, on the basis of this condition number, we can select μ such that $\mathcal{F}_{2p}\left(\boldsymbol{\alpha}^{(k)}\right)$ is well-conditioned. This way, in each iteration of the steepest descent μ is adaptively changed.

6.4 Numerical experiments

Numerical experiments are performed in MATLAB on a standard desktop personal computer. The data generating model, is a global state-space model described by:

$$\begin{aligned} A &= \begin{bmatrix} 0.1440 & 0.089 \\ 0.089 & 0.1440 \end{bmatrix}, E = \begin{bmatrix} 0.0890 & 0 \\ 0 & 0.0890 \end{bmatrix}, \\ B &= \begin{bmatrix} 3.5600 \\ 1.7800 \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & 0.5 \end{bmatrix} \end{aligned} \tag{6.25}$$

The model (6.25) has been obtained using the finite-differences approximation of the heat equation (see Chapter 2). The local inputs are zero mean normally distributed random signals. Furthermore, the local inputs are mutually independent.

In order to make the identification problem more challenging, we corrupt the outputs of the local subsystems by a white noise (the signal to noise ratios of the outputs of the local subsystems are equal to 25 [dB]). In total we perform 100 identification experiments for different realizations of inputs and corresponding outputs of the local subsystems. For the identification we choose $p = 2$, $l = 10$. In each of 100 identification experiments, each element of $\boldsymbol{\alpha}^{(0)}$ is generated by drawing samples from a zero mean normal distribution, with variance equal to 0.1. Similar identification results are obtained when $\boldsymbol{\alpha}^{(0)}$ is formed from the elements of local system matrices identified using the decentralized SIM presented in Chapter 5.

First, we illustrate the $O(N)$ complexity of Algorithm 6.1. We vary the number of local subsystems N, and measure the time that is necessary to compute one iteration of the steepest descent method. We compare the computational times of the following three approaches. In the first approach, that we call *the direct implementation*, the steepest descent method is implemented using MATLAB's sparse matrix computations toolbox. In the second and in the third approach, we implement the steepest descent method by approximating $\mathcal{F}_{2p}^{-1}(\boldsymbol{\alpha}^{(k)})$ using the Chebyshev approximation method and Newton iteration, respectively. In these two approaches all matrix multiplications and additions were performed using MATLAB's sparse matrix computations toolbox. For computing $\mathcal{E}_{2tp}(\boldsymbol{\alpha}^{(k)})$, using the Chebyshev approximation method, we chose: $t = 10$ and $\mu = 0.3$. In the Newton iteration, we use the dropping strategy to restrict the block bandwidth of the approximate inverse ($\beta = 200$). The results are presented in Fig. 6.1.

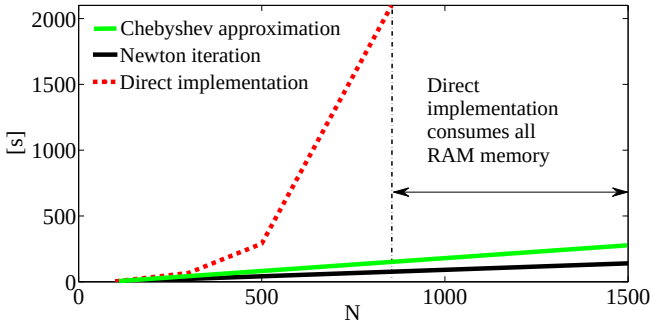


Figure 6.1: Computational complexity of the steepest descent method. In the direct implementation all matrix operations were performed using the MATLAB's sparse matrix computations toolbox.

From Fig. 6.1 it can be observed that using the direct implementation, we cannot identify interconnected systems with more than 800 local subsystems. This is because the direct implementation consumes all computer's RAM memory. Namely, the direct implementation computes dense matrix $\mathcal{F}_{2p}^{-1}(\boldsymbol{\alpha}^{(k)})$. The matrix $\mathcal{F}_{2p}^{-1}(\boldsymbol{\alpha}^{(k)})$, together with dense matrices that are obtained by multiplying $\mathcal{F}_{2p}^{-1}(\boldsymbol{\alpha}^{(k)})$ with $\mathcal{O}_p(\boldsymbol{\alpha}^{(k)})$ or $\mathcal{O}_p^T(\boldsymbol{\alpha}^{(k)})$ (see Section 6.11), need to be stored in the RAM memory. In order to perform this, we need $O(N^2)$ memory locations. On the other hand, the Chebyshev approximation method or the Newton iteration ap-

proximate $\mathcal{F}_{2p}^{-1}(\boldsymbol{\alpha}^{(k)})$ by a sparse banded matrix. Consequently all the matrices in (6.17)-(6.21) are sparse. Because of this, the sparsity of the steepest descent method is preserved. Furthermore, as it can be observed from Fig. 6.1, the approaches based on the Chebyshev approximation method and on Newton iteration have linear computational complexity.

Next, we show how μ influences the convergence of the steepest descent method. We construct the global system that consists of $N = 200$ local subsystems. The steepest descent method is implemented using the Chebyshev approximation method. The results are given in Fig.6.2(a). As it has been expected, larger values of μ slow down the convergence of the steepest descent method. In Fig.6.2(b), we illustrate the effect of the Chebyshev approximation errors on the convergence of the steepest descent method. The Chebyshev approximation of $\mathcal{F}_{2p}^{-1}$ is calculated for $t = 5$ and $t = 10$. For $t = 5$ the approximation error is 0.1, whereas for $t = 10$ this error is 0.005. As it can be observed from Fig. 6.2(b), in the case of $t = 5$ the steepest descent method diverges, whereas in the case of $t = 10$ it monotonically converges. This is because the Chebyshev approximation errors are larger for $t = 5$ than for $t = 10$. In the case of $t = 5$, the accumulation of the errors in each iteration of the steepest descent method is significant and after certain number of iterations, the steepest descent method diverges. This problem can be resolved by improving the accuracy of the Chebyshev approximation. For that purpose we increase t to 10. However, by increasing t, we increase the computational complexity of the Chebyshev approximation. From Theorem 3.4 it follows that the divergence problem can also be resolved by increasing μ. However, as it can be observed from Fig. 6.2(a), by increasing μ we slow down the convergence of the steepest descent method.

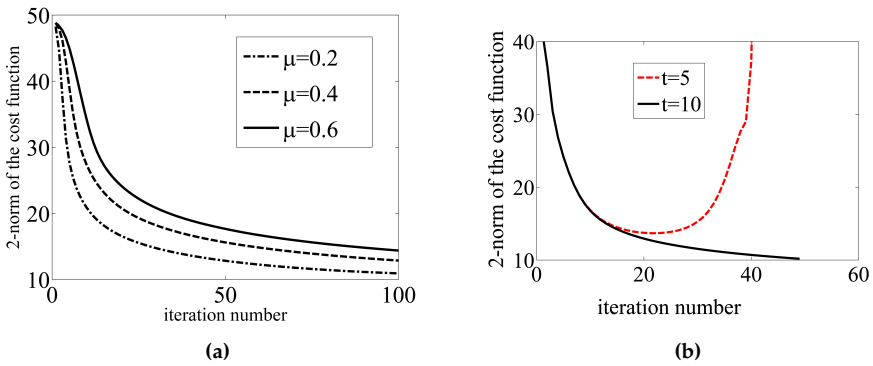


Figure 6.2: Convergence of the steepest descent method; a) influence of the parameter μ b) The divergence of the steepest descent method

Finally, we illustrate the quality of the identified model. For the approximation purpose, we chose $\mu = 0.3$ and $t = 10$. The distribution of the eigenvalues of the identified local matrix A is presented in Fig. 6.3. After we have identified the local system matrices, we reconstructed the global state-space model. Using the reconstructed global state-space model, we calculate the Variance Accounted For

(VAF) [58]. The distribution of the VAF of an arbitrary local subsystem, for 100 identification experiments is presented in Fig. 6.4.

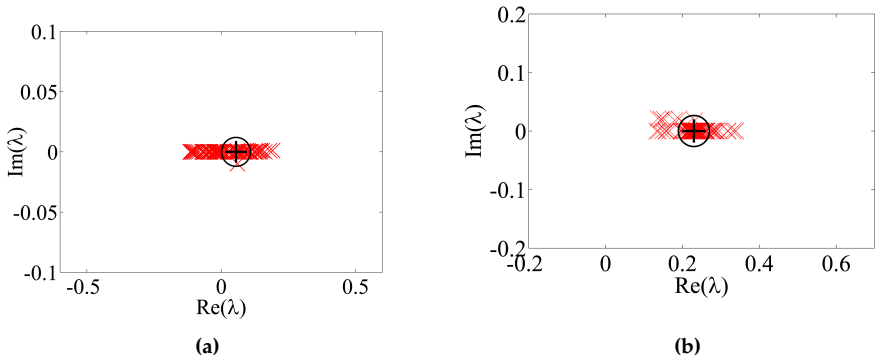


Figure 6.3: Distribution of the eigenvalues of the local matrix A for 100 identification experiments. The circled cross denotes the "true" eigenvalue.

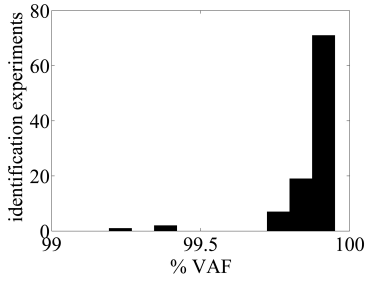


Figure 6.4: The distribution of the VAF for 100 identification experiments.

As it can be observed from Figs. 6.3 and 6.4, the identified models have relatively good quality.

6.5 Conclusion

In this Chapter we have developed a computationally efficient algorithm for identifying interconnected systems that are described by sparse banded or sparse multi-banded system matrices. The computational efficiency is achieved by approximating inverses of lifted system matrices by sparse banded matrices. We numerically demonstrate that it is possible to identify a global state-space model obtained by finite difference discretization of the 2D heat equation with $O(N)$ computational complexity and $O(N)$ memory requirements.

7 CHAPTER

State estimation of the discretized thermoelastic model

In this chapter we use the moving horizon estimation algorithm and the discretized thermoelastic equations to estimate the temperature distribution of the circular mirror illustrated in Fig. 2.8. We show that temperatures can be estimated with linear computational complexity and with relatively good accuracy. Furthermore, we illustrate the trade-off between the accuracy and computational efficiency of the estimation framework developed in this thesis.

7.1 Introduction

To make this thesis more accessible to engineers and scientists with various backgrounds, we first explain why state estimation is important for the development of predictive algorithms for thermally induced wavefront aberrations.

To fully predict the dynamical behavior of a system, that is, to predict the system's future state trajectory, we need to know the system's model, the system's inputs and the system's initial state.
For example, a simple way to predict the future state trajectory of the discretized 2D heat equation (2.11) is to start from an initial state $\underline{\mathbf{x}}(0)$ and by using the input sequence $\underline{\mathbf{u}}(0), \underline{\mathbf{u}}(1), \ldots$, to recursively compute the states $\underline{\mathbf{x}}(1), \underline{\mathbf{x}}(2), \ldots$. Once the state trajectory is known, the output sequence $\underline{\mathbf{y}}(0), \underline{\mathbf{y}}(1), \ldots$, can be predicted.
If the initial state $\underline{\mathbf{x}}(0)$ is zero, then we only need the input sequence and the model to predict the state trajectory. However, in practice systems are rarely initially at rest.

Consider the discretized thermoelastic system (2.81)-(2.82) that describes dynamical behavior of thermally induced deformations of the mirror illustrated in Fig. 2.8. The states of this system are mirror temperatures, the input is the heat flux

distribution (intensity distribution) on the top surface and the outputs are surface deformations. If we want to predict the output, then we need to know the initial temperature distribution and the input. If initially the mirror is in the thermal equilibrium then the initial state can be assumed to be zero. However, in optical lithography this is rarely the case. Namely, mirrors in EUV machines need tens of minutes to completely cool down. Because of the demand to continuously expose wafers, mirrors are always under thermal load and consequently, they do not have enough time to cool down.

All this implies that prediction of thermoelastic deformations can rarely be done in a completely feed-forward manner[1]. That is, we need to have some initial measurements of thermoelastic deformations from which we can estimate the initial state of the system (the initial temperature distribution). To perform this task we need to have a computationally efficient estimation strategy.

In this chapter we estimate the states of the discretized thermoelastic system (2.81)-(2.82) using the MHE strategy developed in Chapter 4.2. Furthermore, we show that using the approximation framework presented in Chapter 3, this estimation can be done with complexity that scales linearly with the state dimension. Due to this, the developed estimation framework is suitable for real time implementation. We also illustrate the trade-off between the accuracy and computational efficiency of the developed estimation framework.

We assume that the mirror is made of BK7 with a thin aluminum coating on its top surface. The mirror's dimensions and its material properties can be found in [253]. *Authors are aware that for high power optical systems, such as EUV lithography machines, mirrors made of materials with a low Coefficient of Thermal Expansion (CTE) are more suitable. However, we have chosen this mirror because real measurements of its thermoelastic deformations are reported in [253].* Consequently, by comparing our simulations results with the experimental data presented in [253], the model based on thermoelastic equations can be validated. The modeling and estimation strategies proposed in this chapter can be adapted such that they can be used for state estimation of mirrors made of various materials (for example, mirrors made of materials with a low CTE, such as Zerodur)

This chapter is organized as follows. In Section 7.2 the thermoelastic model is validated. In Section 7.3 least squares estimation results are presented. In Section 7.4 moving horizon estimation results are presented. Finally, in Section 7.5 the conclusions are drawn.

7.2 Validation

The mirror's dimensions and its material properties are summarized in Table 7.1.

[1] In feed-forward approach, prediction is performed without any measurements.

Parameter	Symbol	Units	Value
Density	ρ	$\left[\,\mathrm{kg}\cdot\mathrm{m}^{-3}\right]$	2.51×10^{3}
Specific heat	c	$\left[\,\mathrm{J}\cdot\mathrm{kg}^{-1}\cdot\mathrm{K}^{-1}\,\right]$	858
Thermal conductivity	κ	$\left[\,\mathrm{W}\cdot\mathrm{m}^{-1}\cdot\mathrm{K}^{-1}\,\right]$	1.114
CTE	α	$\left[\,\mathrm{K}^{-1}\,\right]$	7.10×10^{-6}
Young's modulus	E	$[\,\mathrm{GPa}\,]$	82
Poisson ratio	ν	$[\,-\,]$	0.206
Diameter of the mirror	ϕ	$[\,m\,]$	0.0245
Thickness of the mirror	δ	$[\,m\,]$	0.0065

Table 7.1: Material properties of BK7 and dimensions of the mirror. The data is taken from [253].

In order to assess the accuracy of the FE thermoelastic model (2.81)-(2.82), we compare the deformations predicted by the model with the experimental results reported in [253]. The thermoelastic deformations were simulated assuming the beam diameter of 250×10^{-6} [m] (the diameter of region Ω_4 in Fig. 2.9) and assuming the laser power of 100×10^{-3} [W]. The ambient temperature $T_A = 293.15$ and the emissivity of the aluminum coating $c_4 = 0.4$ were assumed. The comparison results are shown in Fig. 7.1.

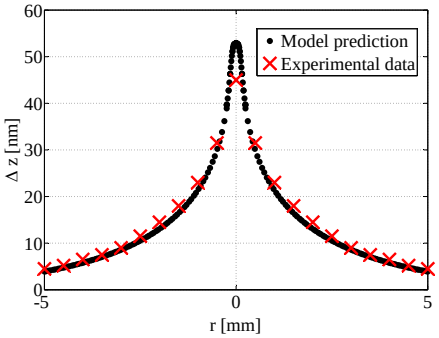


Figure 7.1: Comparison between the vertical surface deformations predicted by the model (2.81)-(2.82) and experimental data reported in [253]. The mirror geometry is illustrated in Fig. 2.9

As it can be seen from Fig. 7.1, the FE model obtained using COMSOL Multiphysics® software can relatively accurately predict the thermoelastic deformations of the mirror surface. The discrepancy between simulated deformations and experimental results partly originate from linearization of radiation boundary conditions. The linear boundary conditions imply that the heat losses due to radiation are underestimated. Furthermore, the errors partly originate from the noise of a sensor used in [253] to measure the displacements (the thermoelastic model is simulated without the measurement noise).

7.3 Least squares state estimation

Estimation is performed on a desktop personal computer with 12 GB of RAM. First, we show estimation results obtained by solving the least squares problem (4.68). The data generating model has 500 states, 150 outputs and 150 inputs (dimension of the vector $\mathbf{c}_1$ in (2.81)). The intensity distribution on the top surface is a Gaussian function. The measurement data (surface deformations) are corrupted by a white Gaussian noise with signal to noise ratio of 35 dB. The initial states are equal to the ambient temperature (assumed to be $297.15K$).

The estimates of the top surface temperatures are shown in Fig 7.2(b) (the top surface is defined by $\partial\Omega_1 \cup \partial\Omega_4$, see Fig. 2.9(b)). For comparison, in Fig. 7.2(a) we show the simulated temperatures of the same surface. In Fig. 7.2(c), the absolute errors between the simulated and the estimated temperatures are presented. In Fig. 7.3 we show the estimates of the middle surface temperatures (the surface parallel to the top surface). Finally, in Fig. 7.4, we show the estimated temperatures of the surface that passes through the center of the mirror and that is perpendicular to the top surface (the diameter cross section of the mirror).

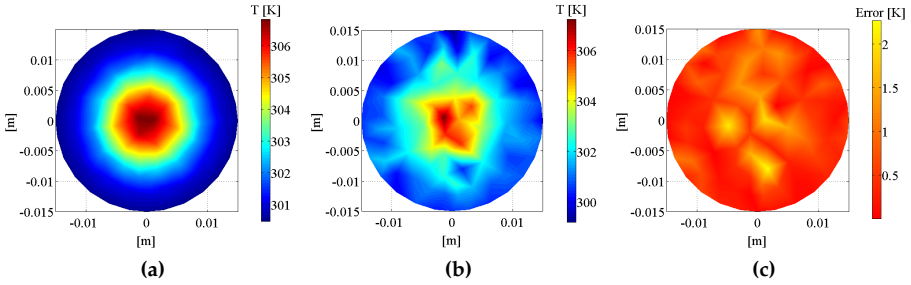


Figure 7.2: Least squares estimates of the top surface temperatures at an arbitrary time instant. (a) Simulated temperatures; (b) Estimated temperatures; (c) Relative error between simulated and estimated temperatures.

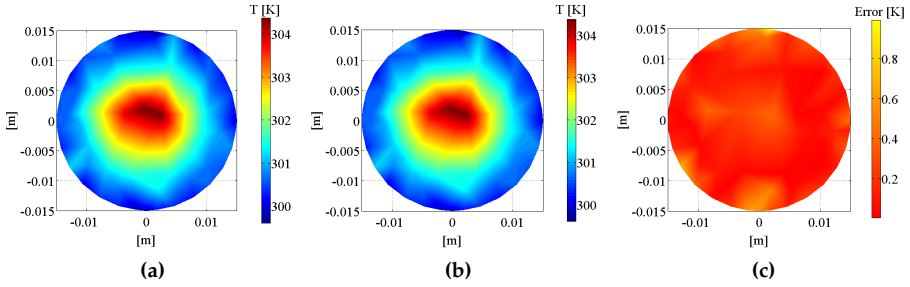


Figure 7.3: Least squares estimates of the middle surface temperatures. (a) Simulated temperatures; (b) Estimated temperatures; (c) Relative error between simulated and estimated temperatures.

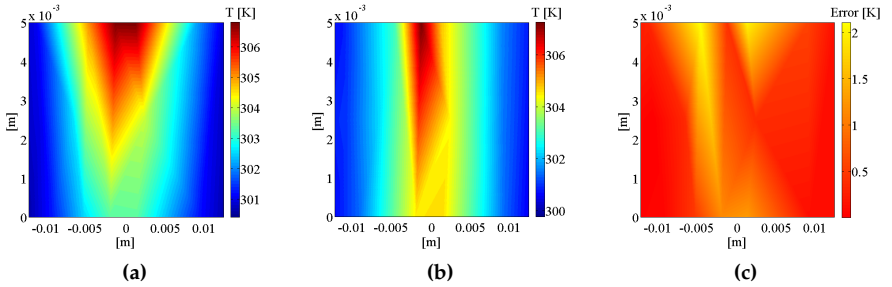


Figure 7.4: Least squares estimates of the temperatures of the diameter cross section of the mirror. (a) Simulated temperatures; (b) Estimated temperatures; (c) Relative error between simulated and estimated temperatures.

From Figs. 7.2-7.4 it follows that the mirror temperatures can be estimated with a maximal absolute error that is around 2K. These estimation errors partly originate from the amplification of the measurement noise. The amplification of the measurement noise is created by the ill-conditioning of the matrix Γ, defined in (4.67). The singular values of Γ are illustrated in Fig. 7.5.

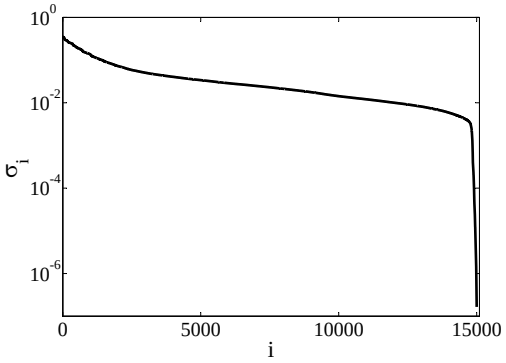


Figure 7.5: Singular values of the matrix Γ defined in (4.67)

From Fig. 7.5, it can be concluded the condition number of Γ is in the order of 10^6, which implies that the condition number of $\Gamma^T\Gamma$ is in order of 10^{12} (to compute the least squares estimate the matrix $\Gamma^T\Gamma$ is inverted). One of the ways to solve this problem and to improve the estimation quality is to employ regularization techniques [254; 255]. The MHE method is implicitly regularizing the pseudo-inverse of Γ. Furthermore, from Fig. 7.5 it can be concluded that some parts of the state sequence are poorly observable.

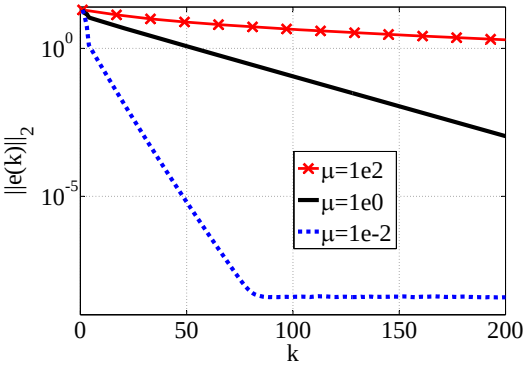
7.4 Moving horizon state estimation

First we will present the estimation results when the inverse of M (matrix M is defined in (4.78)) is computed exactly. The data generating model has 243 states (temperatures), 81 inputs and 81 outputs. We used the past window equal to $p = 60$.

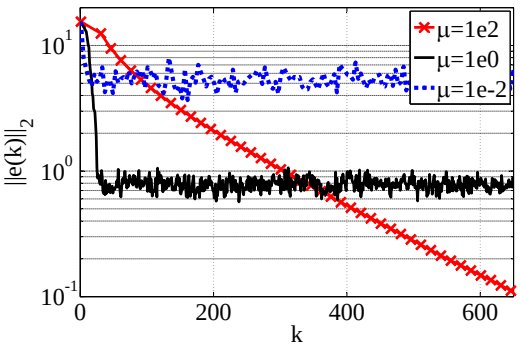
We quantify the estimation quality by computing the estimation error:

$$\mathbf{e}(k-p) = \mathbf{x}(k-p) - \hat{\mathbf{x}}(k-p|k) \tag{7.1}$$

Noise-free estimation results are presented in Fig. 7.6(a). These estimation results indicate that for larger values of μ the convergence of the estimation error is slower and vice-versa. This is expected, because for larger μ we put more emphasis on state predictions that are computed on the basis of the model (the first part of the MHE cost function (4.71)), and less emphasis on the measurement data (the second part of the MHE cost function).



(a)



(b)

Figure 7.6: Estimation error of the MHE method for several values of μ. (a) Noise free scenario; (b) Measurements corrupted by noise.

In Fig. 7.6(b) we present estimation results in the case of measurement noise. Figure 7.6(b) indicates that for larger values of μ the negative effect of the measurement noise on the estimation quality can be decreased. This is partly because for larger values of μ the condition number of M is smaller (see Proposition 4.9), and consequently, the estimation results are less-sensitive to measurement noise. However, for larger μ we have slower convergence of the MHE method. This indicates that there exists a trade-off between the convergence speed and the noise suppression of the proposed MHE method. The theoretical analysis of the effect of the parameter μ on the performance of the MHE method is left for future research.

In Fig. 7.7 we show the estimates of middle surface temperatures at the time instant $k = 170$. For estimation we have used $\mu = 10$.

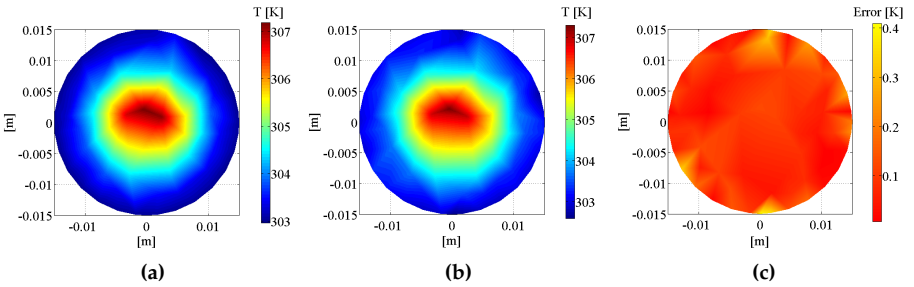


Figure 7.7: Estimates of middle surface temperatures. (a) Simulated temperatures; (b) Estimated temperatures; (c) Absolute error.

Figure 7.7 confirms that the MHE method is able to more accurately estimate the temperatures than the basic least squares method.

Next, we compute moving horizon estimates by approximating the inverse of M using the Newton iteration and the Chebyshev approximation method. At the same time, we are using the truncation strategy, explained in Section 3.4.4, to restrict the bandwidth of approximate inverses. We also show how the approximation errors influence the estimation accuracy.
Throughout the rest of this Chapter, the approximate inverse of M will be denoted by $\tilde{M}$. The final approximation accuracy is measured by computing:

$$\epsilon = \left\| I - M\tilde{M} \right\|_2 \tag{7.2}$$

If $\tilde{M}$ is a good approximation of M^{-1}, then $M\tilde{M}$ is approximately equal to the identity matrix I and consequently, ϵ is very small.
In Figs. 7.8-7.10 we show the estimates computed using the equation obtained by substituting the matrix M^{-1} by $\tilde{M}$ in the equation (4.77). These figures show the estimates of the middle surface temperatures. From Figs. 7.8-7.10 we can see how the truncation bandwidth β of the truncation strategy influences the estimation quality. Similar estimation results are obtained using the Chebyshev approximation method.

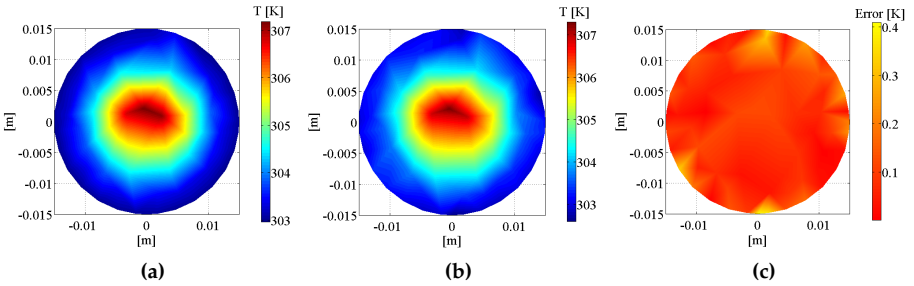


Figure 7.8: The moving horizon estimates computed using $\tilde{M}$, $\beta = 1500$, $\epsilon = 10^{-4}$. (a) Simulated states; (b) Estimated states; (c) Absolute error.

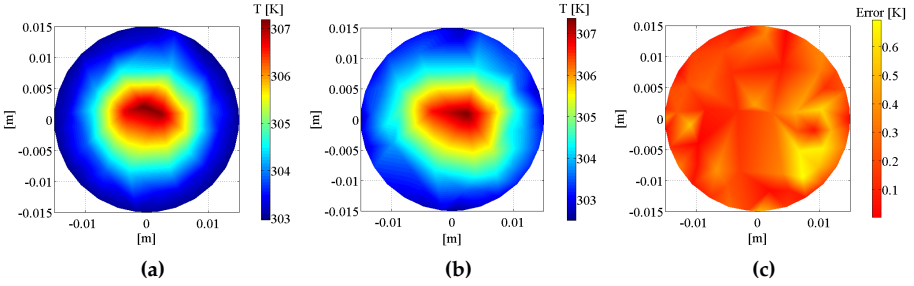


Figure 7.9: The moving horizon estimates computed using $\tilde{M}$, $\beta = 1000$, $\epsilon = 10^{-3}$. (a) Simulated states; (b) Estimated states; (c) Absolute error.

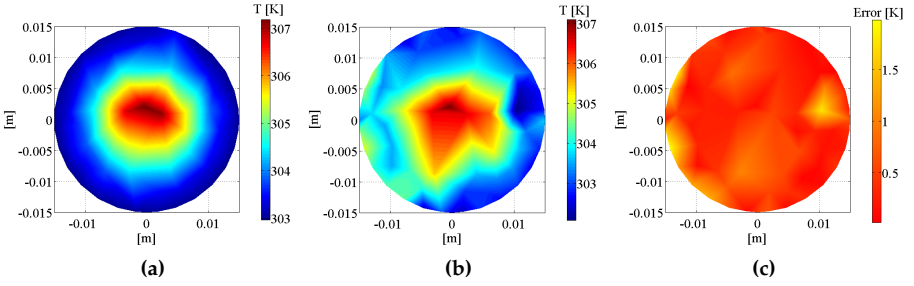


Figure 7.10: The moving horizon estimates computed using $\tilde{M}$, $\beta = 500$, $\epsilon = 5 \cdot 10^{-3}$. (a) Simulated states; (b) Estimated states; (c) Absolute error.

From Figs. 7.8-7.10 we conclude that as the truncation bandwidth β decreases, the estimation accuracy gets worse. This is expected because for smaller β the parameter ϵ, quantifying the error of the approximate inverse, is larger. On the other hand, for smaller β, the computational complexity of computing $\tilde{M}$ decreases.

That is, there is a trade-off between the accuracy and the computational complexity of the developed estimation method.
Next, we show how the approximation errors influence the convergence of the MHE method. The results are shown in Fig. 7.11. From Fig. 7.11 it follows that small approximation errors do not destabilize the convergence of the MHE method. For smaller β the steady-state approximation error is larger.

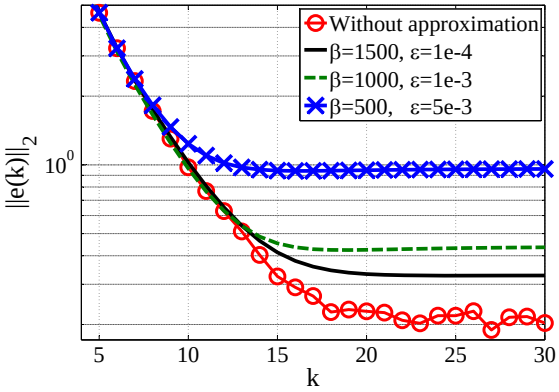


Figure 7.11: Influence of the approximation errors on the convergence of the MHE method.

7.4.1 Computational and memory complexity of the approximation methods

First, we illustrate the computational complexity of the Newton method for computing the approximate inverse $\tilde{M}$. The initial guess was computed by $X_0 = \alpha M$, where $\alpha = 2/(a^2+b^2)$ and a and b are minimal and maximal singular values of M. This initial guess is refined by computing 10 iterations of the Newton algorithm with a relatively small truncation bandwidth: $\beta = 400$. The Newton iteration was started again with this refined initial guess, and it is stopped when the accuracy of $\epsilon = 10^{-3}$ is reached. The computational and memory complexity of the Newton iteration are illustrated in Fig. 7.12. For comparison we also show the complexity of computing the "true inverse" of M using built-in MATLAB functions for performing operations on sparse matrices. Throughout the rest of Chapter, the computational approach that is completely based on the built-in MATLAB functions will be called "the direct approach".

As it can be seen from Fig. 7.12, the Newton method outperforms the direct approach both in terms of computational and memory complexity. Furthermore, we have observed that for matrix dimensions larger than 3×10^4 the direct implementation consumes all computer's RAM memory. In contrast, using the Newton method, we were easily able to invert matrices which dimensions exceed 3×10^4.

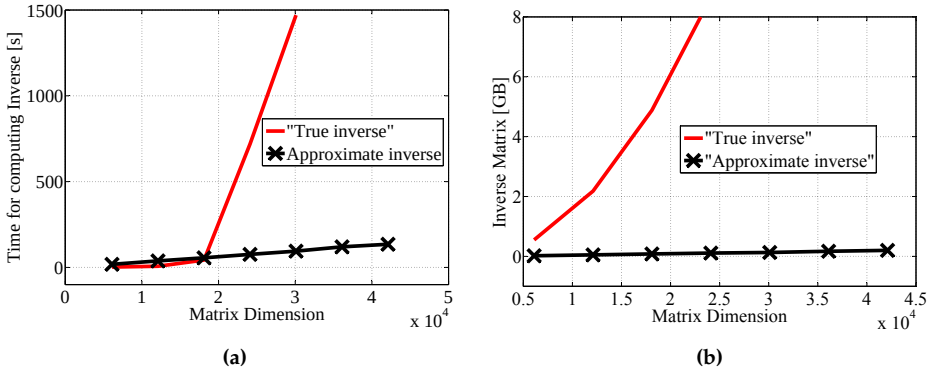


Figure 7.12: Newton iteration: (a) Computational complexity; (b) Memory complexity.

Finally, we illustrate the complexity of computing the moving horizon estimate (4.77) on the basis of the approximate, sparse inverse $\tilde{M}$. The approximate inverse is computed using the Newton method. For comparison, we illustrate the complexity of computing the moving horizon estimate using the direct approach. The results are shown in Fig. 7.13. As we can see from this figure, the moving horizon estimate can be computed with linear computational complexity using the proposed approximation framework. This is because the approximate inverse $\tilde{M}$ is a sparse banded matrix and consequently, the computational complexity of vector-matrix multiplications scales linearly with the size of the problem. On the other hand, the "true" inverse of M is a dense matrix, which implies quadratic computational complexity of vector-matrix multiplications. Because of its linear computational complexity, the developed MHE method is suitable for real-time applications.

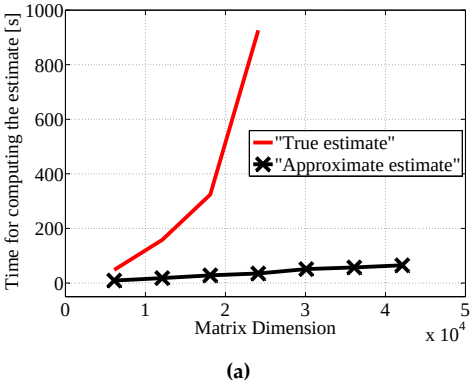


Figure 7.13: Time for computing the moving horizon estimate using the precomputed approximate inverse $\tilde{M}$.

7.5 Conclusion

In this chapter we have demonstrated that the MHE method, developed in Section 4.2, is able to relatively accurately reconstruct the temperature distribution of the mirror shown in Fig. 2.8. Furthermore, we have shown that using the structure preserving, approximate inversion framework developed in Chapter 3, the estimation can be performed with linear computational and memory complexity. Consequently, the proposed estimation framework can be used for real time estimation of the state of thermally induced wavefront aberrations in high power optical systems.

8 CHAPTER

Iterative learning control for optimal wavefront correction

In this chapter, we present an Iterative Learning Control (ILC) algorithm for controlling the shape of a membrane Deformable Mirror (DM). We furthermore give a physical interpretation of the design parameters of the ILC algorithm. On the basis of this insight, we derive a simple tuning procedure for the ILC algorithm that in practice guarantees stable and fast convergence of the membrane to the desired shape. In order to demonstrate the performance of the new algorithm we have built an experimental setup that consists of a commercial membrane DM, a wavefront sensor and a real-time controller. The experimental results show that by using the new ILC algorithm we are able to achieve a relatively small error between the real and desired shape of the DM while at the same time we are able to control the saturation of the actuators. Moreover, we show that the ILC algorithm outperforms other control algorithms available in literature.

8.1 Introduction

The main components of an Adaptive Optics (AO) system are WaveFront Sensor (WFS), an active optical element like a DM or a spatial light modulator and a control algorithm. There are different types of DMs. Most widely used are segmented, microelectromechanical systems (MEMS), bimorph piezoelectric and membrane deformable mirrors. Membrane deformable mirrors are relatively cheap, they have low hysteresis and they have low power consumption. However, the response of membrane mirrors is nonlinear and the coupling between control channels is relatively strong [256].

In literature various modeling and control strategies for membrane DMs have been proposed [256; 257; 258]. The control strategy presented in [256] is based

on the steepest-descent optimization algorithm. The control strategy presented in [258] is derived by inverting an influence function (matrix) of a DM. In [257], the problem of controlling a membrane DM is formulated as a non-negative least squares problem. However, the tuning of the parameters of the above mentioned control strategies is performed empirically, by trial and error. Because of this, these control methods might not guarantee optimal performance of an AO system. This non-optimality manifests itself in a slow convergence of the wavefront error and in a significant steady-state wavefront error. The non-optimality of the above mentioned control strategies also originates from a somewhat heuristic method used to identify a DM model (an influence matrix). Namely, in the above cited papers, the i^{th} column of an influence matrix is a measurement of a membrane response when the step voltage is applied to the i^{th} channel. Although simple, this identification method directly incorporates measurement errors into a DM model. Furthermore, this method does not take into account the "cross-talk" between the channels of a DM. Another drawback of the above cited control strategies is that they do not properly address the problem of saturation of the actuators of a DM. The saturation of the actuators is an undesired phenomena. On the one hand, it creates a strong mechanical stress on the membrane of a DM and thus it reduces its lifetime. On the other hand, saturation might prevent undamped control algorithms to converge.

In order to boost the performance of AO systems, in this chapter we present an iterative learning control (ILC) algorithm for controlling the shape of a membrane DM. The presented ILC algorithm is based on the linearized model of the DM, that is identified from the experimental data. We furthermore give a physical interpretation of the design parameters of the ILC algorithm. On the basis of this insight, we derive a simple tuning procedure for the ILC algorithm that in practice guarantees stable and fast convergence of the membrane to the desired shape. In order to demonstrate the performance of the new algorithm we have built an experimental setup that consists of a commercial membrane DM, a wavefront sensor and a real-time controller. The experimental results show that by using this new ILC algorithm we are able to achieve a relatively small error between the real and desired shape of the DM while at the same time we are able to control the saturation of the actuators. Moreover, we show that the ILC algorithm outperforms other control algorithms available in the literature. We also present a simple and effective statistical identification procedure for a linearized model of the DM. Using this model, and only one initial measurement of the DM shape, we can apply the ILC algorithm off-line. After the ILC algorithm has converged, we apply a "learned input" to a DM. Because it uses only one measurement, this is a fast method for generating the desired shape.

In [259], a method has been proposed for alignment of optical components in optical lithography systems. With minor modifications, the ILC method presented in this Chapter can be also used for precise alignment of optical components in lithography machines.

The benefits of the ILC algorithm for correcting wavefront aberrations were first demonstrated in [260]. However, in [260] the ILC algorithm only penalizes the differences between the voltages of two consecutive control iterations. Although this approach enables us to control the convergence rate of the ILC algorithm, in

some situations the steady-state voltages can still saturate. Furthermore, because the convergence and stability of the ILC algorithm have not been studied in [260], the tuning of the parameters of the ILC algorithm have to be performed on-line on a real setup which can be time consuming. In contrast to [260], in this chapter we penalize the values of the voltages for the next control iteration which enable us to directly control the voltage saturation. Beside this, we perform the stability and convergence analysis of the ILC algorithm and we propose a simple tuning procedure which can be performed off-line.

This chapter is organized as follows. In Section 8.2, we describe an experimental setup. In Section 8.3, we present the ILC algorithm. In section 8.4, we present an identification procedure for a model of the DM. In Section 8.5, we present experimental results and in Section 8.6, we draw conclusions.

8.2 Experimental setup

In this section we describe the optical test bench (AO system) that we use to validate the ILC algorithm for controlling the shape of the DM. The sketch of the experimental setup is shown in Fig. 8.1.

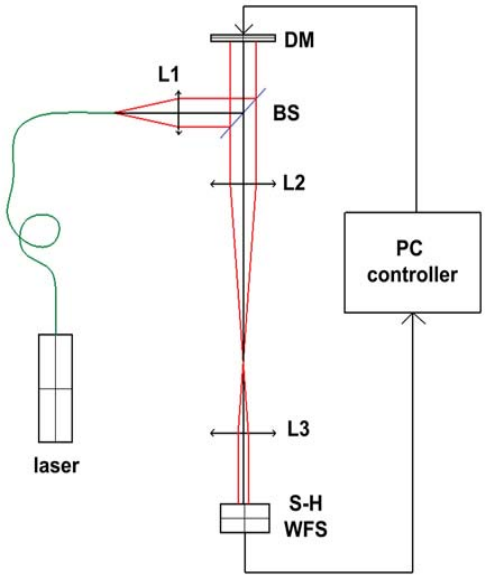


Figure 8.1: Simplified sketch of the experimental setup.

We use coherent light from a semiconductor laser working at $\lambda = 638$nm. The light from the laser is then coupled with a single mode fiber. The other end of the fiber is placed in the focal point of the spherical lens L1 (focal length f_1=100mm, diameter ϕ_1=1″). The resulting collimated beam is folded by 90° by the beam

splitter (BS). The central part (9 mm in diameter) of the beam illuminates the clear aperture (ϕ=11mm) of the deformable mirror (DM) uniformly. The DM is a commercial membrane mirror with 48 actuators, produced by Adaptica Srl. The specification of this mirror can be found in [261]. The reflected light goes again through the BS in a relay system consisting of the lens L2 (focal length f_2=250mm, diameter ϕ_2=2″) and L3 (focal length f_3=100mm, diameter ϕ_3=2″). The purpose of the optical system L2-L3 is twofold. The first purpose is to optically conjugate the surface of the DM with the Shack-Hartmann WaveFront Sensor (S-H WFS, Thorlabs WFS S300-14AR, 1.3 Mpixel, $\lambda/50$ rms accuracy). The optical system L2-L3 also has the function to decrease the beam diameter by a factor $M = f_3/f_2$ that is needed for the S-H WFS. The controller, that is implemented on a standard Personal Computer (PC), receives measurements from the S-H WFS and on the basis of these measurements sends a control signal to the DM. This feedback loop is established using a LabView interface. The control algorithm is implemented in MATLAB.

The wavefront is sampled with the maximal sampling rate of the S-H WFS, which is 15 Hz. The time constant of the membrane DM is in order of few milliseconds [262]. Because the sampling period of the S-H WFS is much larger than a time constant of the DM, we are not able to observe the transient response of the DM using the S-H WFS. Consequently, in Section 8.3 we model the DM as a static system. The control sampling rate is 1 Hz. Because the sampling rate of the S-H WFS is larger than the control sampling rate, in some degree we are able to reduce the effect of the measurement noise on the AO system. This is performed as follows. In between two consecutive control iterations, we take five measurement samples of the wavefront and we average these samples. This averaged wavefront is then used in the next control iteration.

8.3 Iterative learning control for membrane DM

In this section we present an Iterative Learning Control (ILC) algorithm for controlling the shape of a membrane DM. We study its stability and convergence rate. As a result of this study, we give a physical interpretation of the parameters of the ILC algorithm. This physical interpretation gives us guidelines for its (optimal) tuning.

In this chapter, the wavefront that is produced by the DM will be represented using the Zernike polynomials basis expansion (the Malacra notation is used in this Chapter [263]) :

$$\Phi(x,y) \approx \sum_{i=1}^{36} \alpha_i Z_i(x,y) \tag{8.1}$$

where $\Phi(x,y)$ is the spatial distribution of the wavefront, α_i is the i^{th} coefficient of the Zernike polynomials expansion and $Z_i(x,y)$ is the i^{th} Zernike polynomial.

The steady-state model of the DM is given by [257; 258]:

$$\mathbf{W} = F\mathbf{V} \tag{8.2}$$

where $\mathbf{W} \in \mathbb{R}^{36}$, $\mathbf{W} = \begin{bmatrix}\alpha_1 & \alpha_2 & \dots & \alpha_{36}\end{bmatrix}^T$ is the wavefront (membrane shape of the DM) expressed as Zernike coefficients, $F \in \mathbb{R}^{36\times 48}$ is the influence matrix, and the vector $\mathbf{V} \in \mathbb{R}^{48}$ is given as follows:

$$\mathbf{V} = \begin{bmatrix}u_1^2 & u_2^2 & \dots & u_{48}^2\end{bmatrix}^T \tag{8.3}$$

where u_i is the control voltage applied to the i^{th} channel of the DM. As it can be seen from (8.2)-(8.3), the model of the DM is a nonlinear (quadratic) function of the applied voltages. In contrast to [257; 258], in this chapter we propose a control algorithm that is based on the linearized model of the DM. We identify a linear model of the DM, directly from experimental data, using the identification procedure explained in Section 8.4. By neglecting higher-order terms of the Taylor expansion, a linearized model of the DM around the working point A can be written as (see Fig. 8.2):

$$\mathbf{w} = M\mathbf{u} \tag{8.4}$$

$$\mathbf{u} = \mathbf{U} - \mathbf{U}^A, \quad \mathbf{w} = \mathbf{W} - \mathbf{W}^A \tag{8.5}$$

where $\mathbf{U} = \begin{bmatrix}u_1 & u_2 & \dots & u_{48}\end{bmatrix}^T$ is the vector of (total) voltages, $\mathbf{U}^A$ is the vector of working point voltages, $\mathbf{u}$ is the vector of relative voltages, $\mathbf{W}^A$ is the wavefront (membrane shape of the DM) produced by $\mathbf{U}^A$, $\mathbf{W}$ is the wavefront produced by $\mathbf{U}$, $\mathbf{w}$ is the relative wavefront and M is the influence matrix of the linear model. In this thesis the working point voltages are chosen as follows:

$$\mathbf{U}^A = \begin{bmatrix}u_A & u_A & \dots & u_A\end{bmatrix}^T \tag{8.6}$$

where u_A is equal to 50% of the maximal voltages. It should be stressed that the model (8.4)-(8.5) depends on the working point voltages $\mathbf{U}^A$. From Fig. 8.2, we can conclude that for different working point voltages, we obtain a different linear model of the DM. Furthermore, if $\mathbf{U}$ is "close" to $\mathbf{U}^A$, then the linear model accurately describes the behavior of the DM. On the other hand, if $\mathbf{U}$ is "far away" from $\mathbf{U}^A$, then the linear model is less accurate. However, as we experimentally demonstrate in Section 5, the ILC algorithm can effectively handle these model uncertainties. For more details about robustness of the ILC algorithm with respect to model uncertainties, see [199; 264] and references therein.

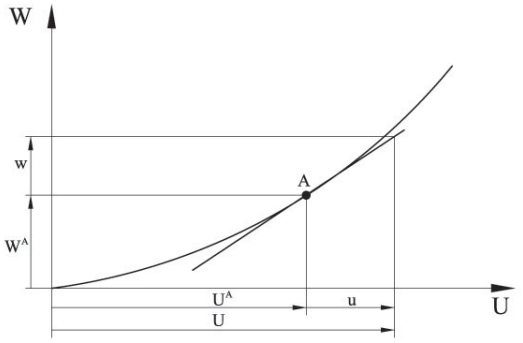


Figure 8.2: Linear and nonlinear model of the DM.

Let the control iteration be denoted by k (the control sampling rate is 1Hz). Further, let the wavefront (membrane shape of the DM), at the control iteration k, be denoted by $\mathbf{W}_k$. Similarly, we denote the relative membrane shape by $\mathbf{w}_k = \mathbf{W}_k - \mathbf{W}^A$. From (8.4)-(8.5) it follows:

$$\mathbf{w}_k = M\mathbf{u}_k \tag{8.7}$$

where $\mathbf{u}_k = \mathbf{U}_k - \mathbf{U}^A$ is the vector of relative voltages at the control iteration k, and $\mathbf{U}_k$ is vector of (total) voltages at control iteration k. The wavefront that is sensed by the S-H WFS, is the combination of the wavefront produced by actuating the DM and static wavefront aberrations initially present in the AO system. These static wavefront aberrations come from non-flatness of the DM (when the voltages are not applied to DM) and from imperfections and misalignments of the optical components. Because the wavefront $\mathbf{W}_k$ corresponds to the membrane shape, this wavefront is obtained by subtracting all static wavefront aberrations from the wavefront measured by S-H WFS. As we explain in Section 8.5.1, the ILC algorithm can be easily modified such that it takes into account static wavefront aberrations of the AO system.

For simplicity, we did not include the measurement noise in equation (8.7). In reality the measured wavefront is corrupted by the measurement noise of the S-H WFS. As we have explained in Section 8.2, the effect of the noise is reduced by averaging the wavefront between the two consecutive control iterations. Moreover, by adjusting the parameters of the ILC algorithm we are able to additionally increase noise immunity of the AO system.

The goal of the ILC algorithm is to produce a wavefront of the desired shape. Let such a wavefront be denoted by $\mathbf{W}_d$. The relative desired wavefront is denoted by $\mathbf{w}_d$ and it is defined as: $\mathbf{w}_d = \mathbf{W}_d - \mathbf{W}^A$. The wavefront error $\mathbf{e}_k$ at the control iteration k is defined as follows:

$$\mathbf{e}_k = \mathbf{w}_d - \mathbf{w}_k = \mathbf{w}_d - M\mathbf{u}_k \tag{8.8}$$

The error at the time instant $k+1$ is given by:

$$\mathbf{e}_{k+1} = \mathbf{w}_d - M\mathbf{u}_{k+1} \tag{8.9}$$

From (8.8) and (8.9) we have:

$$\mathbf{e}_{k+1} = \mathbf{e}_k - M\Delta\mathbf{u}_k \tag{8.10}$$

where

$$\Delta\mathbf{u}_k = \mathbf{u}_{k+1} - \mathbf{u}_k \tag{8.11}$$

For a given $\mathbf{u}_k$ and $\mathbf{e}_k$, the optimal ILC law [199] is obtained by solving the following optimization problem:

$$\min_{\mathbf{u}_{k+1}} \{\mathbf{e}_{k+1}^T Q_e \mathbf{e}_{k+1} + \mathbf{u}_{k+1}^T Q_u \mathbf{u}_{k+1} + \Delta\mathbf{u}_k^T Q_{\Delta u} \Delta\mathbf{u}_k\} \tag{8.12}$$

where Q_e, Q_u, and $Q_{\Delta u}$ are the weighting matrices. The first term in (8.12) penalizes the wavefront error at the control iteration $k+1$. The second term penalizes the voltages for the control iteration $k+1$. Finally, the last term of the cost function (8.12) penalizes the difference between the voltages between the two consecutive control iterations, k and $k+1$. We chose the weighting matrices of the cost function (8.12) as follows:

$$Q_e = I, \quad Q_u = \gamma I, \quad Q_{\Delta u} = \beta I \tag{8.13}$$

where β and γ are positive real numbers. By solving (8.12) for the selection of the weighting matrices (8.13) we obtain the control law in the following form [199]:

$$\mathbf{u}_{k+1} = Q(\mathbf{u}_k + L\mathbf{e}_k) \tag{8.14}$$

where

$$Q = \left(M^T M + \gamma I + \beta I\right)^{-1} \left(M^T M + \beta I\right) \tag{8.15}$$

$$L = \left(M^T M + \beta I\right)^{-1} M^T \tag{8.16}$$

At the initial iteration $k = 0$, the vector of relative voltages $\mathbf{u}_0$ needs to be initialized. We choose $\mathbf{u}_0 = 0$, which corresponds to the vector of total voltages $\mathbf{U} = \mathbf{U}^A$.

The experimental tuning of the parameters β and γ of the ILC control algorithm (8.14)-(8.16) might be time consuming and it might not guarantee optimal performance of the AO system. In order to simplify the tunning of the ILC algorithm, in the sequel we derive stability and monotonic convergence conditions for the ILC algorithm. Furthermore, we show how the steady-state tracking error and inputs depend on the parameters β and γ. Based on these insights we give application oriented guidelines for the selection of β and γ.

8.3.1 Stability and convergence rate of the ILC algorithm

By substituting (8.8) in (8.14), we obtain:

$$\mathbf{u}_{k+1} = Q(I - LM)\mathbf{u}_k + QL\mathbf{w}_d \tag{8.17}$$

The stability and convergence rate of the ILC algorithm is primarily determined by the spectral properties of the matrix $Q(I - LM)$. In order to perform the stability and convergence rate analysis, we introduce the Singular Value Decomposition (SVD) [112] of the full rank influence matrix M:

$$M = E_1 \begin{bmatrix} \Sigma & 0 \end{bmatrix} E_2^T \tag{8.18}$$

where $E_1 \in \mathbb{R}^{36\times 36}$ and $E_2 \in \mathbb{R}^{48\times 48}$ are unitary matrices, and the matrix $\Sigma \in \mathbb{R}^{36\times 36}$ is a diagonal matrix of singular values: $\Sigma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_{36})$, where $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{36} > 0$ are singular values. From (8.15) and (8.16) we have:

$$Q(I - LM) = \left[M^T M + (\gamma + \beta)\, I\right]^{-1} \beta \tag{8.19}$$

Using the SVD (8.18), we can write (8.19) as follows:

$$Q(I - LM) = E_2 \begin{bmatrix} \left[\Sigma^2 + (\gamma + \beta)\, I\right]^{-1} \beta & 0 \\ 0 & \frac{\beta}{\gamma+\beta} I \end{bmatrix} E_2^T \tag{8.20}$$

The ILC algorithm is stable if $\|Q(I - LM)\|_2 < 1$, where $\|\cdot\|_2$ denotes the 2-norm. From (8.20) we conclude:

$$\|Q(I - LM)\|_2 = \frac{\beta}{\gamma + \beta} \tag{8.21}$$

and consequently, we conclude that the ILC algorithm is stable if

$$\frac{\beta}{\gamma + \beta} < 1 \tag{8.22}$$

Because by definition $\beta > 0$ and $\gamma > 0$, from (8.22) we see that the ILC algorithm is always stable. However, a stable ILC algorithm does not necessarily imply a fast convergence of the control voltages. Due to this, in the sequel we study a convergence rate of the ILC algorithm.

We say that the ILC algorithm is monotonically convergent, if

$$\|\mathbf{u}_\infty - \mathbf{u}_{k+1}\|_2 \leq \nu \|\mathbf{u}_\infty - \mathbf{u}_k\|_2 \tag{8.23}$$

where $\mathbf{u}_\infty = \lim_{k\to\infty} \mathbf{u}_k$ and $0 \leq \nu < 1$ is the convergence rate of the ILC algorithm. The smaller ν is, the faster is the convergence of the ILC algorithm and vice-versa. From (8.17) we have:

$$\mathbf{u}_\infty = Q(I - LM)\mathbf{u}_\infty + QL\mathbf{w}_d \tag{8.24}$$

Because $\gamma > 0$, we have $\|Q(I - LM)\|_2 < 1$. This implies that the matrix $I - Q(I - LM)$ is invertible, and by solving (8.24) for $\mathbf{u}_\infty$ we obtain:

$$\mathbf{u}_\infty = [I - Q(I - LM)]^{-1} QL\mathbf{w}_d \tag{8.25}$$

By substituting (8.15) and (8.16) in (8.25) we obtain:

$$\mathbf{u}_\infty = \left(M^T M + \gamma I\right)^{-1} M^T \mathbf{w}_d \tag{8.26}$$

Equation (8.26) gives the value of the steady-state relative voltage vector. This equation helps us to estimate the steady state voltages for the desired wavefront $\mathbf{w}_d$, for the influence matrix M and for a chosen parameter γ. Next we will show how the convergence rate of the ILC algorithm depends on the parameters β and γ. From (8.25) we have:

$$QL\mathbf{w}_d = [I - Q(I - LM)]\,\mathbf{u}_\infty \tag{8.27}$$

Further we have:

$$\mathbf{u}_\infty - \mathbf{u}_{k+1} = \mathbf{u}_\infty - Q(I - LM)\mathbf{u}_k - QL\mathbf{w}_d \tag{8.28}$$

Substituting (8.27) in (8.28), we obtain:

$$\mathbf{u}_\infty - \mathbf{u}_{k+1} = Q(I - LM)(\mathbf{u}_\infty - \mathbf{u}_k) \tag{8.29}$$

From the last expression, we obtain:

$$\|\mathbf{u}_\infty - \mathbf{u}_{k+1}\|_2 \leq \|Q(I - LM)\|_2 \,\|(\mathbf{u}_\infty - \mathbf{u}_k)\|_2 \tag{8.30}$$

From (8.20) we have $\|Q(I - LM)\|_2 = \frac{\beta}{\beta+\gamma}$. This implies that the convergence rate of the ILC algorithm is given by:

$$\nu = \frac{\beta}{\beta + \gamma} \tag{8.31}$$

From the last equation we can observe that by increasing γ the convergence rate ν decreases. That is, by increasing γ we have a faster convergence of the ILC algorithm (since smaller ν gives faster convergence of the ILC algorithm). In the sequel, we will show that by increasing γ we decrease the steady-state relative voltages. This way, we can prevent actuator saturation. However, as we show in the sequel, the increase of γ has a negative effect on the steady-state tracking error. From (8.8) we obtain:

$$\mathbf{e}_\infty = \mathbf{w}_d - M\mathbf{u}_\infty \tag{8.32}$$

By substituting (8.26) in (8.32) we obtain:

$$\mathbf{e}_\infty = \left[I - M\left(M^T M + \gamma I\right)^{-1} M^T\right] \mathbf{w}_d \tag{8.33}$$

The equation (8.33) can help us to estimate the steady-state wavefront error for a desired wavefront, for the influence matrix M and for the chosen value of the parameter γ. As we prove in Appendix section, the steady-state relative input and the wavefront error are bounded as follows:

$$\|\mathbf{u}_\infty\|_2 \leq \frac{1}{2\sqrt{\gamma}} \|\mathbf{w}_d\|_2 \tag{8.34}$$

$$\|\mathbf{e}_\infty\|_2 \leq \frac{\gamma}{\gamma + \sigma_{36}^2} \|\mathbf{w}_d\|_2 \tag{8.35}$$

From the analysis presented in this section, we are able to give a physical interpretation of the design parameters β and γ, and to draw the following guidelines for the tuning of the ILC algorithm.

8.3.2 Physical interpretation of the parameters of the ILC algorithm and guidlines for its tuning

- From (8.34) we know that by increasing γ we decrease the upper bound on the steady-state voltages. This physically means that by increasing γ we can prevent the steady state voltages to saturate. At the same time, we know that by increasing γ the ILC algorithm converges more rapidly.

- On the other hand, from (8.35) we see that by increasing γ the upper bound on the steady state wavefront error increases. This physically means that by increasing γ we degrade the accuracy of the wavefront correction/generation.

- From (8.35) we can deduce that when $\gamma \to 0$, $\mathbf{e}_\infty \to 0$. However, this will never be achieved in practice since this condition only holds when there are no model uncertainties (due to linearization, we always have model uncertainties). Furthermore there is always measurement noise in the system.

- From (8.26) and (8.33) we see that the parameter β does not influence the steady-state input and the steady-state wavefront error. However, from (8.31) we have that by increasing β the convergence rate of the ILC algorithm is slower and vice-versa. The parameter β can also be used to regularize badly conditioned matrices M and consequently to improve the noise immunity of the system [199; 265].

In this section, we have derived the ILC algorithm by assuming that at each control iteration k we are able to measure $\mathbf{w}_k$. That is, we have assumed that the ILC algorithm is applied on-line. However, in Section 8.5.3. we will experimentally show that with only one initial measurement of the membrane shape, and by learning the control input off-line using the ILC algorithm, we are able to achieve a relatively good performance of the AO system. Although the accuracy of the wavefront generation using this off-line method is slightly worse than the on-line method, the off-line method can produce a relatively small wavefront error in only one control iteration.

8.4 Identification of the influence function

In order to identify the linearized model of the DM around the working point A (see Fig. 8.2), we introduce the following matrices:

$$P = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_N \end{bmatrix}, \quad D = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \dots & \mathbf{w}_N \end{bmatrix} \tag{8.36}$$

In (8.36), $\mathbf{u}_i \in \mathbb{R}^{48}$, $i = 1, \dots, N$, is a vector of random relative voltages varying between 0% and $\pm 30\%$ (30% of relative voltages correspond to 80% of total voltages, since the working point A is chosen at 50% of total voltages) and $\mathbf{w}_i$ is the membrane shape produced by $\mathbf{u}_i$. The influence matrix M is identified by solving the following least-squares optimization problem:

$$\min_{M} \|D - MP\|_F^2 \tag{8.37}$$

where $\| \bullet \|_F$ denotes the Frobenius norm [58] and N is a relatively large-number (in our case $N = 200$). The random voltages ensure that P has full row rank, so the solution of the optimization problem (8.37) is then given by $M = DP^{\dagger}$, where $P^{\dagger} = P^T(PP^T)^{-1}$ denotes the matrix pseudo-inverse.
The random voltages ensure that the mirror is persistently excited and that the cross-coupling between the channels is captured by the identified model. In order to determine $\mathbf{w}_k$ and consequently to define D, we need to know $\mathbf{W}^A$. That is, we need to know the membrane shape produced by the working point voltages. In theory, the wavefront $\mathbf{W}^A$ can be obtain by subtracting all static wavefront aberrations (originating from the experimental setup when the voltages are not applied to the DM) from the total measured wavefront. However, by applying the vector of working point voltages $\mathbf{U}^A$ at different time instants k, and by subtracting all static wavefront aberrations from the measured wavefront, we will observe different $\mathbf{W}_k^A$. This is due to the measurement noise of the S-H WFS. We solve this problem using the following strategy from [58]: we apply $\mathbf{U}^A$ 30 times to DM. Each time we measure the corresponding wavefront and subtract all static wavefront aberrations. After that, we determine $\mathbf{W}^A$ by averaging wavefronts from these $N = 30$ measurements (assuming that the effect of the hysteresis of the DM is not significant).

In Section 8.5.3 we will compare the dynamical behavior of the AO system with the dynamical behavior of the model of the AO system that is based on the identified influence function M.

8.5 Experimental results

In the first part of this section we present experimental results that illustrate the influence of β and γ on the dynamical behavior of the AO system. In the second part we illustrate the ability of the AO system to generate/compensate some typical Zernike polynomial wavefront aberrations. In the third part we compare the identified model of the DM with experimental results. Furthermore, we show that

by taking one initial wavefront measurement and by applying the ILC algorithm off-line, we are able to achieve a good performance of the AO system. Finally, in the fourth part we compare the ILC algorithm with other control algorithms.

8.5.1 Dynamical behavior

In Fig. 8.3 we present experimental convergence rates of the wavefront error and voltages for different values of β and γ. In this experiment, our goal is to make the wavefront sensed by S-H WFS to be equal to a flat wavefront. In this case, the desired total wavefront (desired membrane shape) for the ILC algorithm is: $\mathbf{W}_d = -\mathbf{d}$, where $\mathbf{d}$ is the measurement of all static wavefront aberrations in the AO system when the voltages are not applied to the DM. From Fig. 8.3 we see that by increasing γ we decrease the norm of the steady-state voltages. That is, by adjusting γ we can prevent actuator saturation. However, by increasing γ we increase the steady-state wavefront error, that is, we degrade the accuracy of the wavefront correction. We can also see that the convergence speed of the ILC algorithm decreases as β increases. These experimental results confirm the theoretical conclusions that we drew in section 8.3.

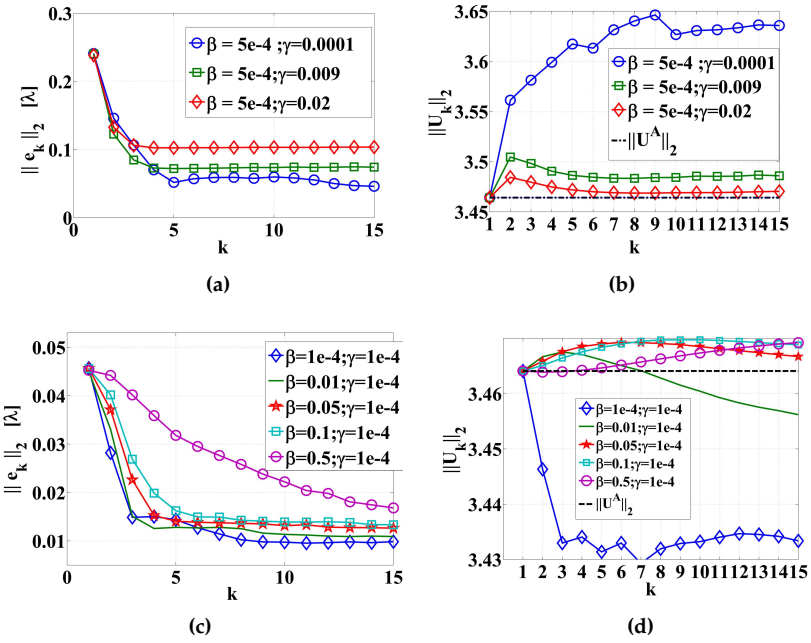


Figure 8.3: Analysis of the convergence of the wavefront from an arbitrary to a flat wavefront, for different values of β and γ. (a) Convergence of the norm of the tracking error for different γ; (b) Convergence of the norm of the total voltages for different γ; (c) Convergence of the norm of the tracking error for different β; (d) Convergence of the norm of the total voltages for different β.

8.5.2 Performance of the AO system

To demonstrate the performance of the AO system, we have generated several wavefronts described by different Zernike modes. The details of the desired wavefronts are listed in Table 8.1.

Desired aberration	Zernike Index	P-V [λ]	RMS [λ]
Astigmatism	α_4	0.34	0.1
Defocus	α_5	0.45	0.1
Trefoil	α_6	0.45	0.08
Coma	α_7	0.45	0.08

Table 8.1: desired wavefront details

The results of the wavefronts convergence are given in Fig. 8.4-8.8. We used the following parameters of the ILC algorithm:$\beta = 0.0005$ and $\gamma = 0.0001$. From these figures it can be concluded that the ILC algorithm guarantees relatively good wavefront generation performance with a final RMS wavefront error of about 0.01λ for all the considered cases.

Figures 8.4-8.8. (a) Convergence of the norm of the wavefront error $\|\mathbf{e}_k\|_2$; (b) Convergence of the total voltages $\|\mathbf{U}_k\|_2$; (c) Converged voltages of each channel of the DM; (d) Converged wavefront.

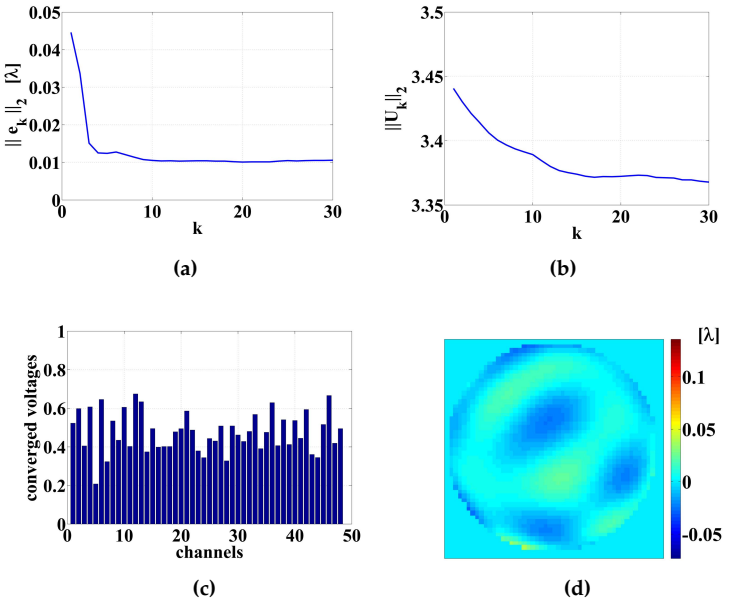


Figure 8.4: Flat wavefront generation.

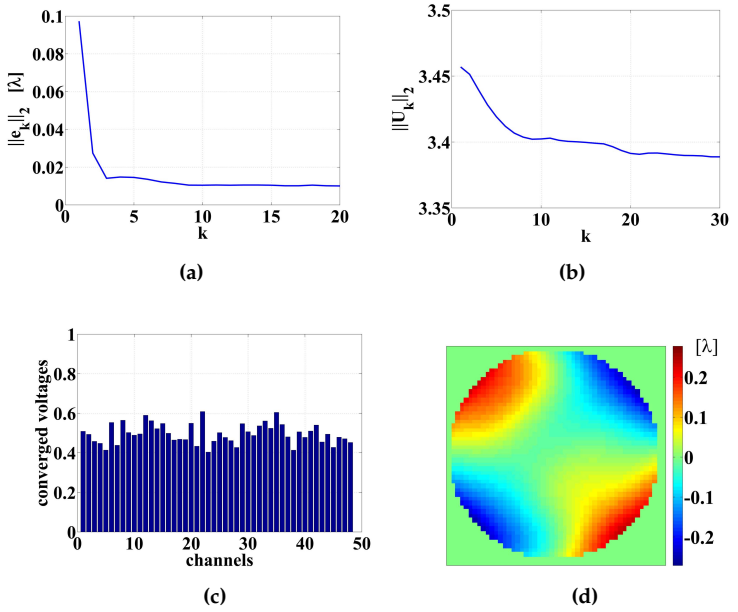


Figure 8.5: Astigmatism generation.

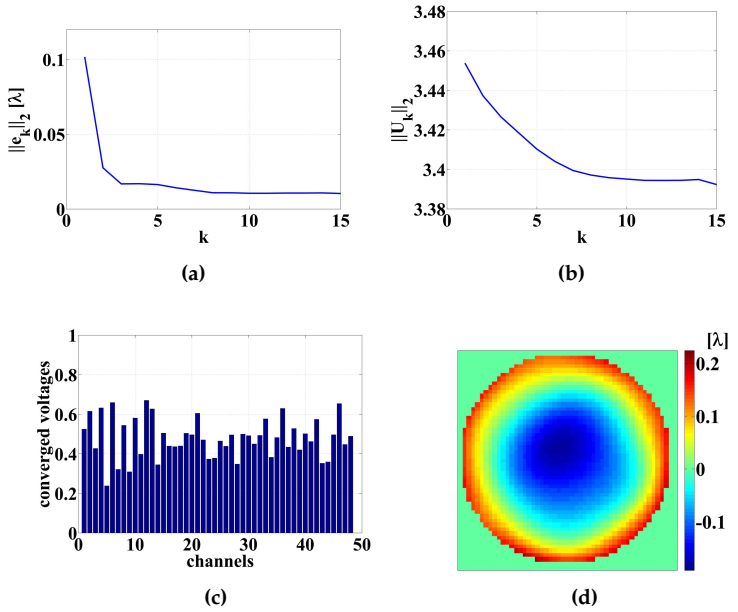


Figure 8.6: Defocus generation.

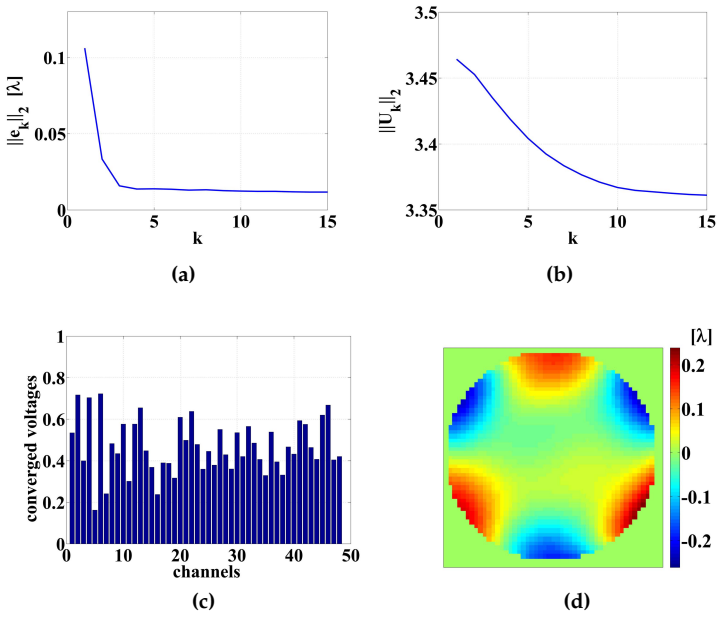


Figure 8.7: Trefoil generation.

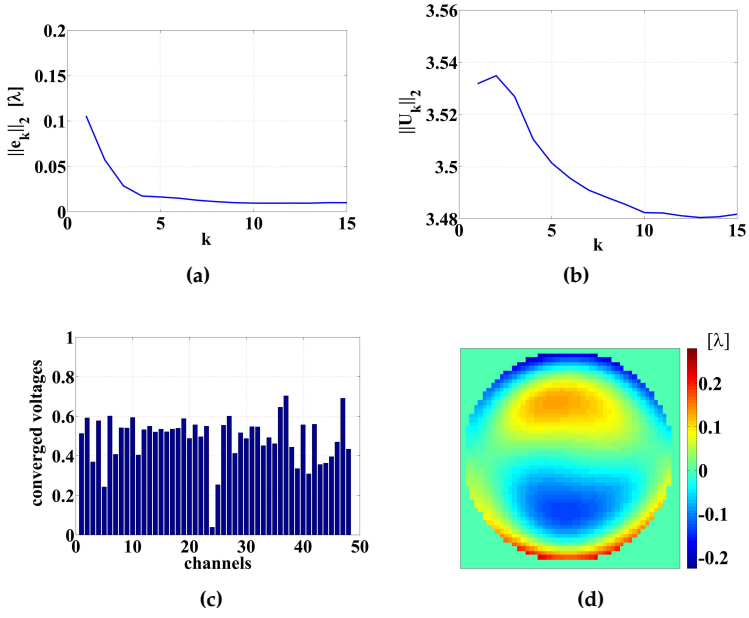


Figure 8.8: Coma generation.

8.5.3 Comparison between the model and the experimental setup

Using the identified influence function M as a model of the DM, we have simulated the dynamical behavior of the AO system. We assume that the desired wavefront is a flat wavefront. We compare such a simulated behavior with the experimental results. The comparison is presented in Fig.8.9.

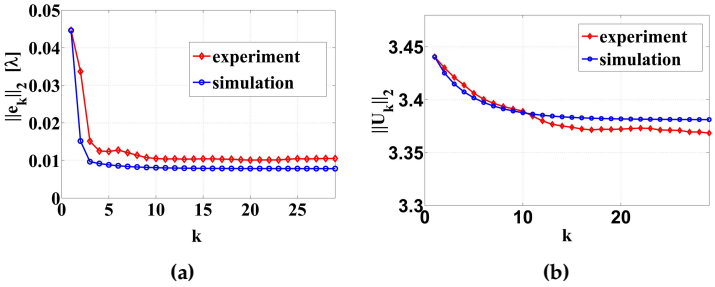


Figure 8.9: Comparison of the model with the experimental results. (a) Convergence of the norm of the wavefront error $\|\mathbf{e}_k\|_2$; (b) Convergence of the norm of the total voltages $\|\mathbf{U}_k\|_2$.

As it can be observed from Fig. 8.9, the identified model of the influence function gives a relatively good prediction of the dynamical behavior of the real AO system. This motivates us to perform the following experiment. For a desired wavefront equal to the flat wavefront and for one initial measurement of wavefront aberrations, we run the iterative learning control algorithm off-line. That is, we run it without taking any additional measurements except the initial one. After the ILC algorithm has converged, we apply the converged voltages to the DM. The converged wavefront (steady-state wavefront) is shown in Fig. 8.10. As a comparison, we show the converged wavefront in the case when the ILC algorithm was applied on-line.

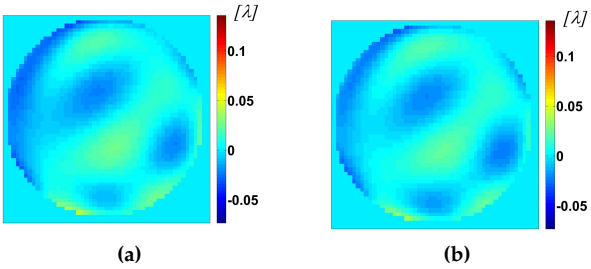


Figure 8.10: Comparison of the off-line with the on-line application of the ILC algorithm. (a) Converged wavefront when the ILC algorithm is applied off-line; (b) Converged wavefront when the ILC algorithm is applied on-line.

The RMS value of the steady state wavefront error for the case when the ILC algorithm is applied off-line is 0.015λ. On the other hand, the RMS value of the steady state wavefront error when the ILC algorithm is performed on-line is 0.011λ. These results show us that using the identified model and the ILC algorithm, we can generate/correct wavefront aberrations with only one initial measurement.

8.5.4 Comparison of ILC with other control algorithms

As a final demonstration of the advantages of the ILC algorithm, we compare it with the Non-Negative Least Squares (NNLS) control algorithm of [257]. Furthermore, we compare the proposed ILC algorithm with the Steepest Descent ILC algorithm [199]:

$$\mathbf{u}_{k+1} = \mathbf{u}_k + \alpha M^T \mathbf{e}_k \tag{8.38}$$

where $\alpha = 0.3$ is a step size. In this case the desired wavefront W_d is astigmatism of 0.1λ RMS. The results are presented in Fig. 8.11. In order to distinguish the proposed ILC algorithm from other algorithms, in this subsection we call it the optimal ILC algorithm. We have used $\beta = \gamma = 0.0001$.

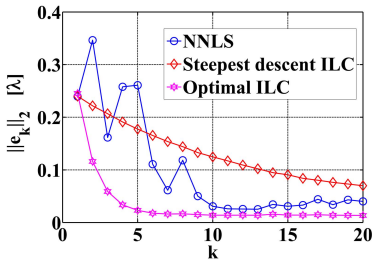


Figure 8.11: Comparison of different control algorithms

As it can be observed from Fig. 8.11, the optimal ILC algorithm outperforms other two control algorithms. First of all, the optimal ILC algorithm converges in only 5 control iterations, while the NNLS converges after 10 iterations and the steepest descent ILC algorithm does not converge in 20 iterations. Next, the optimal ILC algorithm reaches the smallest value of the steady-state tracking error.

8.6 Conclusion

In this Chapter we have proposed an Iterative Learning Control (ILC) algorithm for controlling the shape of a membrane DM. We have studied the stability and convergence rate of the novel algorithm and on the basis of this study we have given a physical interpretation of the controller parameters. This interpretation enabled us to derive a simple tuning procedure that in practice guarantees fast

and stable convergence of the wavefront error. The experimental results show that by using the ILC algorithm we are able to achieve a relatively small value of the residual wavefront, while at the same time we are able to effectively control the saturation of voltages. Furthermore, the experimental results show that the ILC algorithm produces a small residual wavefront when it is applied off-line with only one initial measurement.

8.6.1 Appendix

In this appendix we give a proof of (8.34) and (8.35). Using (8.18) we can express (8.26) and (8.33) as follows:

$$\mathbf{u}_\infty = E_2 K E_1^T \mathbf{w}_d \tag{8.39}$$

$$\mathbf{e}_\infty = E_1 S E_1^T \mathbf{w}_d \tag{8.40}$$

where

$$K = \begin{bmatrix} \Sigma(\Sigma^2 + \gamma I)^{-1} \\ 0 \end{bmatrix}, \; S = I - \Sigma^2 \left(\Sigma^2 + \gamma I\right)^{-1} \tag{8.41}$$

From (8.39), (8.40) and (8.41) we have:

$$\|\mathbf{u}_\infty\|_2 \leq \left\|E_2 K E_1^T\right\|_2 \|\mathbf{w}_d\|_2 \tag{8.42}$$

$$\|\mathbf{e}_\infty\|_2 \leq \left\|E_1 S E_1^T\right\|_2 \|\mathbf{w}_d\|_2 \tag{8.43}$$

From (8.41) we see that the singular values of $E_2 K E_1^T$ have the following form:

$$\frac{\sigma_i}{\sigma_i^2 + \gamma} \tag{8.44}$$

where σ_i is the ith singular value of M. The 2-norm of $E_2 K E_1^T$ is equal to its maximal singular value. Consider the function $f(x) = \frac{x}{x^2+\gamma}$, where x is a real number. The maximum of $f(x)$ is equal to $1/(2\sqrt{\gamma})$ and it is achieved for $x = \sqrt{\gamma}$. When $x = \sigma_i$, $f(\sigma_i)$ is equal to (8.44). From the above analysis, we have:

$$\max_{\sigma_i} \frac{\sigma_i}{\sigma_i^2 + \gamma} \leq \frac{1}{2\sqrt{\gamma}} \tag{8.45}$$

When a singular value σ_i is equal to $\sqrt{\gamma}$, then $\left\|E_2 K E_1^T\right\|_2 = \frac{1}{2\sqrt{\gamma}}$. From (8.42)-(8.45) we obtain (8.34). From (8.41) we see that the singular values of $E_1 S E_1^T$ are given by:

$$\frac{\gamma}{\sigma_i^2 + \gamma} \tag{8.46}$$

From (8.46) we see that the maximal singular value of $E_1 S E_1^T$ (that is the 2-norm) is given by:

$$\left\| E_1 S E_2^T \right\|_2 = \frac{\gamma}{\sigma_{36}^2 + \gamma} \tag{8.47}$$

where σ_{36} is the minimal singular value of M. Using (8.43) and (8.47) we obtain (8.35). This completes the proof.

9 CHAPTER

Identification of a dynamical model of a thermally actuated deformable mirror

Using the subspace identification technique, in this Chapter we identify a finite-dimensional, dynamical model of a recently developed prototype of a thermally actuated deformable mirror. The main advantage of the identified model, over the models that are described by partial differential equations, is its low complexity and low dimensionality. Consequently, the identified model can be easily used for high performance feedback or feed-forward control. The experimental results show a good agreement between the response of the model and the measured response of the thermally actuated deformable mirror.
The model identified in this Chapter is used in Chapter 10 to develop the predictive controller for the compensation of thermally induced wavefront aberrations.

9.1 Introduction

In a large variety of Adaptive Optics (AO) applications [2; 20; 21; 169; 190; 266; 267; 268; 269], slowly-varying or static wavefront aberrations must be corrected accurately. Thermally Actuated Deformable Mirrors (TADMs) are suitable for these AO applications because they have a high position resolution with high reproducibility [26]. Furthermore, they are less expensive than other types of Deformable Mirrors (DMs), such as membrane or piezoelectric DMs [21; 169].

In [20], a TADM has been used to correct static wavefront aberrations. In the above cited paper, a static (steady-state) model of the TADM has been identified, and on the basis of this model, a control action for the TADM has been derived as

the solution of a constrained least-squares problem. However, this control strategy requires that the time between two consecutive control iterations is approximately equal to the settling time (or the rise time) of a TADM. Consequently, this wavefront correction strategy is relatively slow and its performance might be additionally degraded in the case of time-varying wavefront aberrations.

To achieve fast correction of both static and time-varying wavefront aberrations, the time between control iterations has to be significantly smaller than the settling time of a TADM. In such cases, a dynamical model of a TADM has to be developed to accurately correct wavefront aberrations [245; 270]. Once this dynamical model has been obtained, model based control strategies [271; 272] can be employed to maximize the performance of the wavefront correction. Apart from the control perspective, a dynamical model of the TADM is important because it can be used to simulate the dynamical behavior of the AO system before the real system has been built.

A dynamical model of a TADM has to meet two requirements. First, it has to accurately capture the dynamics of TADM. Second, to be used for control, it has to be relatively simple [260; 273] and preferably low dimensional. However, the dynamics of TADMs are governed by the thermoelastic Partial Differential Equations (PDEs) [57]. Furthermore, in the case of the TADMs that have been proposed in [2; 169], the thermoelastic equations have to be coupled with the biharmonic plate equation [180]. The dynamical model that is based on these PDEs is infinite-dimensional and as such is too complex to be used for control. To apply model based control strategies of [271; 272], a more compact, finite dimensional model needs to be developed. One way to develop such a model, would be to discretize thermoelastic PDEs and corresponding boundary conditions using the Finite Element Method (FEM) [274]. However, the FEM can be applied only if all physical parameters of the TADM are known. Furthermore, the finite element model is usually high dimensional and thus is still relatively complex to be used for control.

In this Chapter, we follow another way of model building that is based on system identification techniques [58]. Accordingly, from experimental data we identify a low-order, state-space model of a recently developed prototype of a TADM. This device was developed by Eindhoven University of Technology [2; 26]. We have identified the dynamical model of the TADM using the subspace identification technique [58; 250]. Due to its low-dimensionality, the identified model can be easily used to design feed-forward or feedback model based controllers. As it will be shown later, the experimental results show a good match between the response of the identified state-space model and the response of the TADM. The model identified in this Chapter is used in Chapter 10 to develop the predictive controller for thermally induced wavefront aberrations.

This Chapter is organized as follows. In Section 9.2, on the basis of the experimental results we define the model of the TADM. In Section 9.3 we present the identification results. Finally, in Section 9.4 we present conclusions.

9.2 Model definition

The sketch of the TADM is shown in Fig. 9.1. The mirror B and the backplate A, are connected with 19 thermo-mechanical actuators C. The mirror diameter is 30 [mm] and the 19 actuators are placed inside of a circle with a diameter of 25 [mm]. The actuators consist of aluminum rods with heating coils warped around them. When applying a voltage to an actuator it heats up and it elongates. As it elongates, it exerts a mechanical force that deforms the mirror and a supporting back plate A. To identify the model of the TADM, we built an experimental setup where the surface of the TADM is illuminated by coherent light coming from a semiconductor laser of wavelength $\lambda = 638$ [nm]. The wavefront reflected by TADM is measured by a Shack-Hartmann Wavefront Sensor (S-H WFS) (Thorlabs WFS S300-14AR, 1.3 Mpixel, $\lambda/50$ rms accuracy). The mirror and the S-H WFS are optically conjugate through a relay system consisting of two spherical lenses.

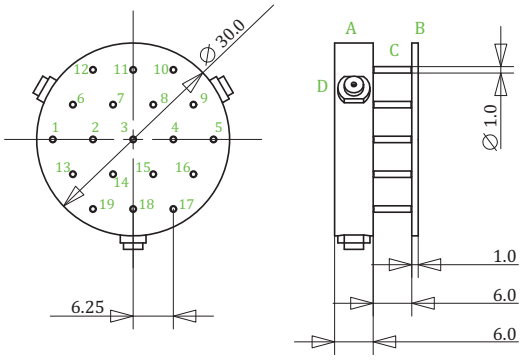


Figure 9.1: Sketch of the thermally actuated DM

In general, the sampling period should be 5-10 times shorter than the rise time of the TADM [58]. Our experimental results show that the rise time of the TADM is approximately 20 [s], which is in agreement with the results reported in [2]. Consequently, we chose a control and measurement sampling period of 2 [s] [58]. In this chapter, k will denote a discrete time instant corresponding to this sampling period. The wavefront that is produced by the TADM, at the time instant k, and that is sensed by the S-H WFS, is represented using a Zernike polynomial expansion (Noll [263]):

$$\mathbf{\Phi}(x, y, k) = \sum_{i=1}^{36} \alpha_i(k) Z_i(x, y)$$

where $\mathbf{\Phi}(x, y, k)$ is the wavefront, $\alpha_i(k)$ is the i^{th} coefficient of the Zernike polynomial expansion and $Z_i(x, y)$ is the i^{th} Zernike polynomial. All static wavefront aberrations that originate from the initial non-flatness of the TADM and from imperfections and misalignments of the optical components in the system, are subtracted from the measured wavefront.

The model of the TADM is identified using the following procedure [58]. First, we determined which Zernike coefficients were excited when we randomly actuated the TADM. It helped us to define the outputs of the model. Next, we investigated the linearity of the TADM. On the basis of this analysis, we postulated a state-space model of the TADM. In the final step, we identified the state-space model and assessed its quality.

To constrain the outputs of the model, we randomly actuated all 19 actuators over a certain time period and recorded the amplitude of all 36 measured Zernike coefficients. By analyzing this measurement data, we concluded that during the random actuation of the TADM, the first 10 Zernike coefficients were excited while others did not show a significant contribution to the measured wavefront. Therefore the output of the model should consists of the first 10 Zernike coefficients: $\mathbf{y}(k) = \begin{bmatrix}\alpha_2(k) & \alpha_3(k) & \dots & \alpha_{10}(k)\end{bmatrix}^T$ (where we neglect the first coefficient, i.e., piston). The voltage applied to the i^{th} actuator (input of the model) will be denoted by v_i (varies between 0% and 100% of the maximum voltage).

To analyze the step response, we applied the step functions of different magnitudes to the actuators and measure each response. Figure 9.2 shows the measured responses of the third actuator.

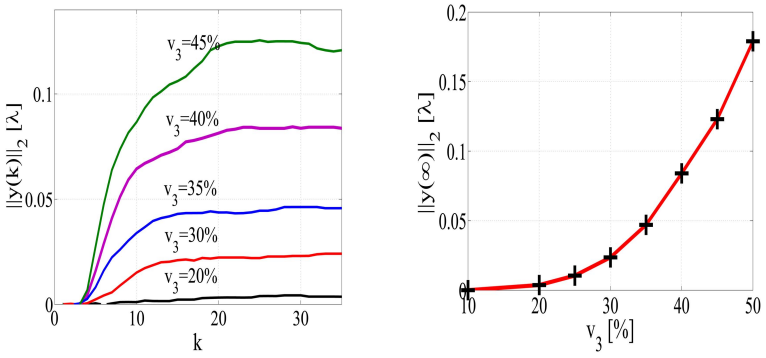


Figure 9.2: Step response of the TADM when the step functions are applied to the third actuator: (a) The RMS $\|\mathbf{y}(k)\|_2$; (b) Steady-state value of the RMS $\|\mathbf{y}(\infty)\|_2$

From Fig. 9.2(b) we see that the steady-state deformation of the TADM is a quadratic function of inputs. This is an experimental confirmation that the heat generated at the actuator is a quadratic function of the applied voltages [2]. As we show later, this nonlinearity can be eliminated from the model, by defining new inputs that are squares of the voltages applied to the actuators. Next, we tested the superposition principle between several actuators. The result of the superposition test between actuators 2 and 4 is shown in Fig. 9.3. We applied a 40% voltage step to actuator 2 and we measured the response (v_2 in Fig.9.3). Next, we applied a 40%

voltage step to actuator 4 and we measured the response. Finally, we actuated actuators 2 and 4 together ($v_2\&v_4$ in Fig.9.3) and measured the response.

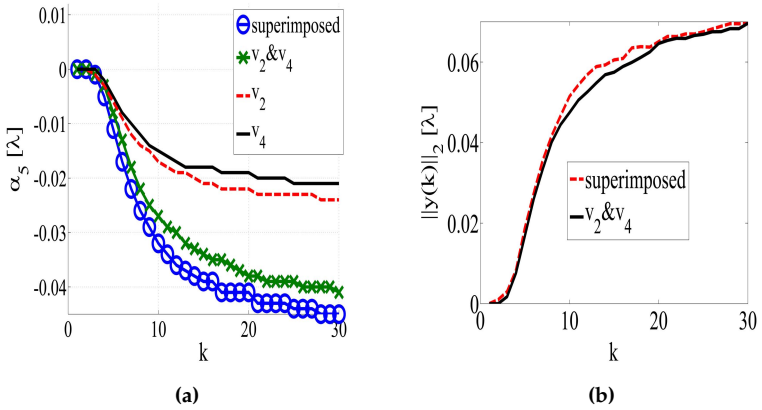


Figure 9.3: Test of the superposition principle: (a) 5th Zernike coefficient (b) The RMS of the total measured wavefront

The experimental results shown in Figs. 9.2 and 9.3, indicate that the TADM can be modeled as a linear state-space model in which the inputs are squares of the voltages:

$$\mathbf{x}(k+1) = A\mathbf{x}(k) + B\mathbf{u}(k) \tag{9.1}$$

$$\mathbf{y}(k) = C\mathbf{x}(k) + \mathbf{e}(k) \tag{9.2}$$

where $\mathbf{x}(k) \in \mathbb{R}^n$ is the state, n is the system order, $\mathbf{u}(k) = \begin{bmatrix}(v_1(k))^2 & (v_2(k))^2 & \dots & (v_{19}(k))^2\end{bmatrix}$ is the input vector of the squared voltages, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times 19}$ and $C \in \mathbb{R}^{9\times n}$ are the system matrices and $\mathbf{e}(k) \in \mathbb{R}^9$ is a S-H WFS measurement noise.

The identification problem of the state-space model (9.1)-(9.2), can be formulated as follows. *From the sequence of the input-output data:* $\{\mathbf{y}(k), \mathbf{u}(k)\}$, $k = 0, 1, \dots, N$, *estimate the system order* n *and identify the system matrices* A, B *and* C *up to a similarity transformation.*

The length N of the data sequence $\{\mathbf{y}(k), \mathbf{u}(k)\}$ should be chosen such that the total measurement time is at least ten times longer than the time constant (or roughly the rise time) of the TADM [58]. In general, the larger the length of the data sequence, the better the quality of the identified model. We chose $N = 450$. The estimates of A, B and C will be denoted by $\hat{A}$, $\hat{B}$ and $\hat{C}$, respectively. To identify the state-space model we used the predictor based subspace identification method [245; 250]. Similar identification results can be obtained using other subspace identification techniques [58].

We use the following strategy to assess the quality of the identified model. First, we generated an input sequence that is different from the input sequence that was

used for identification. Let such an input sequence be denoted by $\{\mathbf{u}_2(k)\}$. Then, using this input we actuated the TADM and generated the output data sequence $\{\mathbf{y}_2(k)\}$. From this input-output data, we estimate the initial state of (9.1)-(9.2) using the following methodology. By substituting the system matrices with their estimates, from (9.1)-(9.2) we obtain:

$$\underline{\mathbf{y}}_2 = \hat{O}\mathbf{x}(0) + \hat{D}\underline{\mathbf{u}}_2 + \underline{\mathbf{e}}_2 \tag{9.3}$$

where

$$\underline{\mathbf{y}}_2 = \begin{bmatrix} \mathbf{y}_2(0) \\ \mathbf{y}_2(1) \\ \vdots \\ \mathbf{y}_2(M) \end{bmatrix}, \underline{\mathbf{u}}_2 = \begin{bmatrix} \mathbf{u}_2(0) \\ \mathbf{u}_2(1) \\ \vdots \\ \mathbf{u}_2(M) \end{bmatrix}, \underline{\mathbf{e}}_2 = \begin{bmatrix} \mathbf{e}_2(0) \\ \mathbf{e}_2(1) \\ \vdots \\ \mathbf{e}_2(M) \end{bmatrix}, \hat{O} = \begin{bmatrix} \hat{C} \\ \hat{C}\hat{A} \\ \vdots \\ \hat{C}\hat{A}^M \end{bmatrix}$$

$$\hat{D} = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ \hat{C}\hat{B} & 0 & 0 & \dots & 0 \\ \hat{C}\hat{A}\hat{B} & \hat{C}\hat{B} & 0 & \dots & 0 \\ \vdots & & \ddots & \ddots & \\ \hat{C}\hat{A}^{M-1}\hat{B} & \hat{C}\hat{A}^{M-2}\hat{B} & \dots & \hat{C}\hat{B} & 0 \end{bmatrix} \tag{9.4}$$

and where M should be chosen such that the matrix $\hat{O}$ has full column rank. The initial state $\mathbf{x}(0)$ is estimated by solving:

$$\min_{\mathbf{x}(0)} \left\| \underline{\mathbf{z}} - \hat{O}\mathbf{x}(0) \right\|_2^2 \tag{9.5}$$

where $\underline{\mathbf{z}} = \underline{\mathbf{y}}_2 - \hat{D}\underline{\mathbf{u}}_2$. The solution of (9.5) is $\hat{\mathbf{x}}(0) = \hat{O}^\dagger \underline{\mathbf{z}}$, where $\hat{O}^\dagger = (\hat{O}^T\hat{O})^{-1}\hat{O}^T$ denotes the matrix pseudo-inverse. Starting from $\hat{\mathbf{x}}(0)$ and using $\{\mathbf{u}_2(k)\}$, we simulate the identified state-space model (9.1)-(9.2) (the system matrices are substituted by their estimates). This way, we generate the predicted output sequence $\{\hat{\mathbf{y}}_2(k)\}$. The quality of the identified model is assessed by comparing the predicted output sequence with the measured one. This quality is expressed using the Variance Accounted For (VAF) [58]:

$$VAF = \max\left\{0, \left(1 - \frac{\frac{1}{L}\sum_{k=0}^{L} \|\mathbf{y}_2(k) - \hat{\mathbf{y}}_2(k)\|_2^2}{\frac{1}{L}\sum_{k=0}^{L} \|\mathbf{y}_2(k)\|_2^2}\right) \times 100\%\right\} \tag{9.6}$$

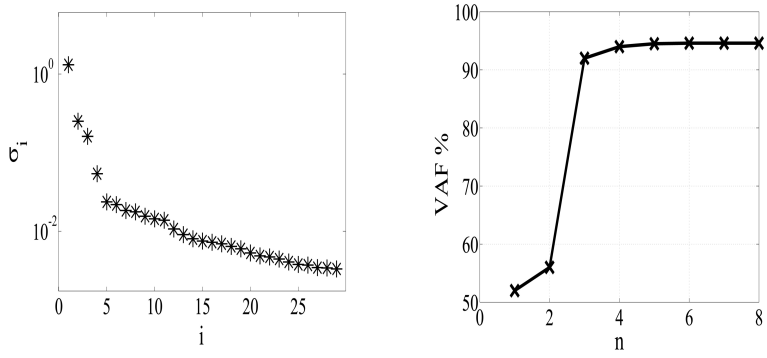
where L is the length of the data sequence. If VAF is equal to 100% it means a perfect match between the output that is predicted by the model and the TADM's actual response.

9.3 Identification results

As a first validation, we identified the state-space model when only the central actuator 3 is active (see Fig. 9.1). The input used for identification and the corre-

sponding output are given in Fig.9.5(a). The order of the state-space model (9.1)-(9.2) is estimated from the singular values and VAF plots given in Fig. 9.4. By identifying the gaps between the singular values that are shown in Fig. 9.4(a), we concluded that a relatively good state order estimate is $n = 4$ [58]. Figure 9.4(b) also confirms this conclusion, where we see that the value of the VAF for $n = 4$ is 93%. On the basis of the identified state-space model, we calculated the transfer function from v_3^2 to α_3. Figure 9.5(b) shows a Bode plot [271] of this transfer function. In Figs. 9.5(c) and 9.5(d), we compared the dynamical response of the TADM and the output of the identified state-space model. As it can be seen from these figures, the identified model of the 4^{th} order is able to predict the behavior of the DM with a relatively good accuracy.

Next, we identified the state-space model when all 19 actuators are active (the inputs used for identification were similar to the one that is shown in 9.5(a)). The order selection is performed on the basis of the singular values and VAF plots that are shown in Figures 9.6(a) and 9.6(b). From these figures we concluded that a relatively good state order estimate is $n = 40$. The value of VAF corresponding to this state order estimate is roughly 90%. The prediction performance of the identified state-space model (for $n = 40$) is illustrated in Fig. 9.6(c) and Fig. 9.6(d). As can be observed from these figures, the state-space model (9.1)-(9.2) of the 40^{th} order can predict relatively well the dynamic response of the TADM.



(a) **(b)**

Figure 9.4: (a) The singular values of the data matrix used in the identification; (b) The VAF values for different order n of the state-space model (9.1)-(9.2).

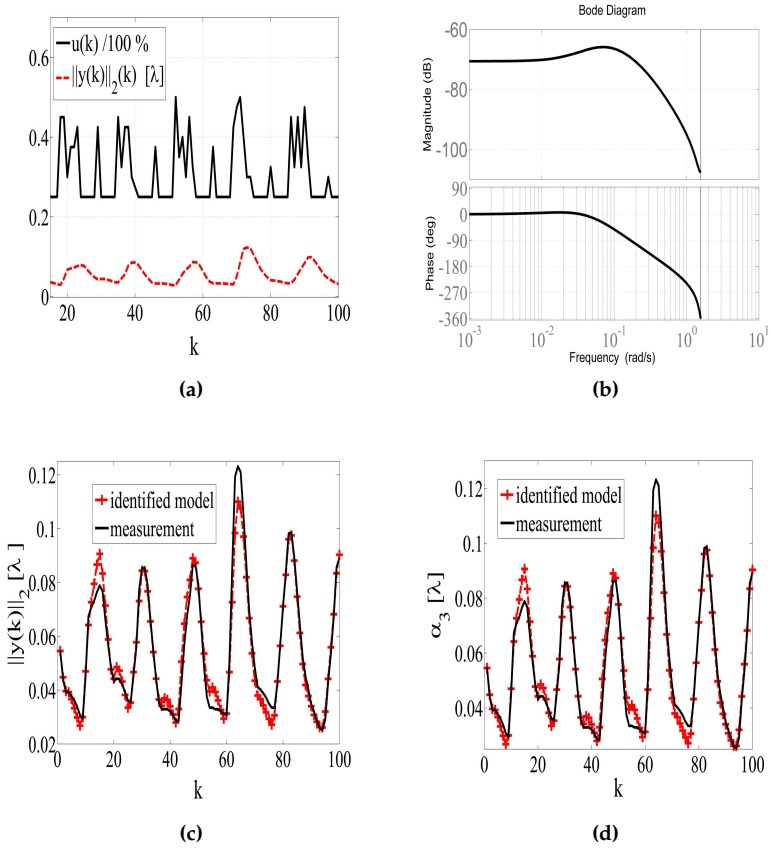


Figure 9.5: Identification results when only actuator 3 is active: (a) The input used for identification and the rms of the output; (b) Bode plot of the transfer function from v_3^2 to α_3; (c) and (d) Prediction and measurements of Zernike coefficients.

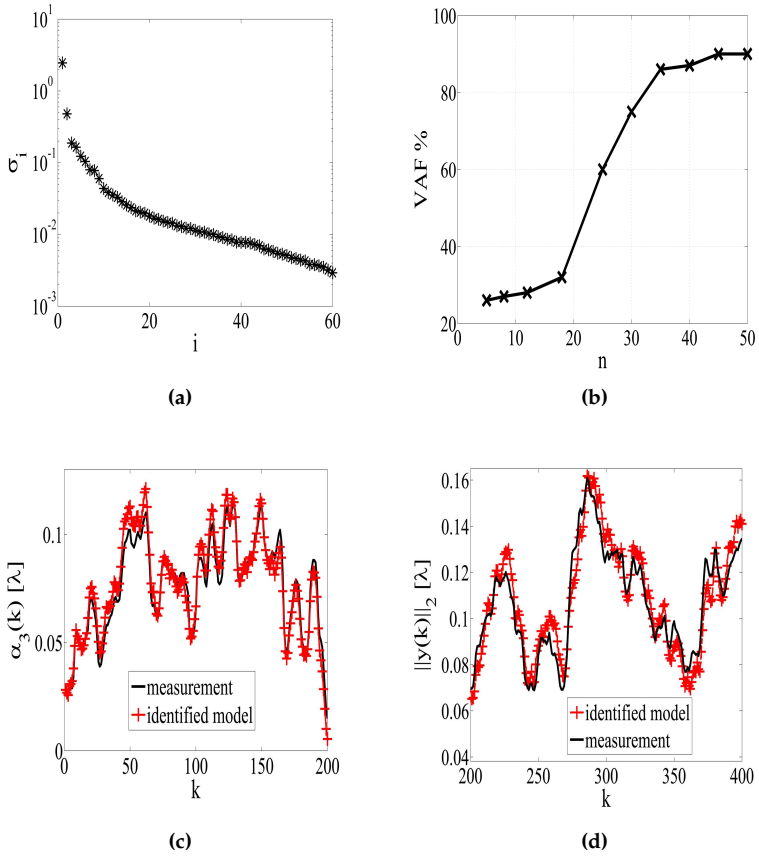


Figure 9.6: Identification results when all 19 actuators are active: (a) Singular values of the data matrix used for identification; (b) VAF for different model orders n; (c) and (d) the prediction performance of the identified model.

9.4 Conclusion

We have used the subspace identification technique to identify a low order dynamical model of a prototype of a thermally actuated deformable mirror (TADM). We have demonstrated that the identified state-space model of the 40^{th} order can accurately describe the dynamical behavior of the mirror (90% match). This contrasts to other modeling approaches that describe dynamics of TADMs using partial differential equations. Furthermore, the order of the identified model can be additionally decreased using model reduction techniques [245]. The proposed identification procedure is general and it can be used to identify dynamical models of other types of TADMs (for example, the types of TADMs proposed in [20; 21; 169]). Due to its low-dimensionality, the identified model can be used to

develop efficient and simple model based controllers.

The drawback of the proposed identification approach is that the structure of the real system is not preserved in the identified model. That is, the states of the identified model do not correspond directly to the physical states (temperatures) of the TADM.

10 CHAPTER

Predictive control of thermally induced wavefront aberrations

In this chapter we experimentally demonstrate the proof of concept for predictive control of thermally induced wavefront aberrations in high-power optical systems. On the basis of the model of thermally induced wavefront aberrations and using only past wavefront measurements, the proposed adaptive optics controller is able to predict and to compensate the future aberrations. Furthermore, the proposed controller is able to correct wavefront aberrations even when some parameters of the prediction model are unknown. The proposed control strategy can be used in high power optical systems, such as optical lithography machines, where the predictive correction of thermally induced wavefront aberrations is a crucial issue.

10.1 Introduction

In high power optical systems almost each element absorbs a portion of the beam's energy. The absorbed energy creates thermoelastic deformations and variation of refractive index of optical elements [22; 269; 275]. Consequently, it induces wavefront aberrations in the system. In this thesis, the aberrations caused by heating of optical elements are called the Thermally Induced Wavefront Aberrations (TIWA).

TIWA can limit the performance of a large variety of high power optical systems. For example, in gravitational wave interferometers high power lasers induce aberrations that can significantly decrease the sensitivity of the instruments [20; 21; 22]. TIWA can also degrade the beam quality of the lasers used in material processing [23; 24]. Furthermore, due to a constant demand for higher productivity and less production costs, the power transmitted through the projection optics of lithography machines constantly increases. Consequently, the energy absorbed by projection optics induces wavefront aberrations that can compromise the resolution of

the system [7; 9; 10; 11; 12]. In the next generation of optical lithography machines, that will use Extreme UltraViolet (EUV) sources, degradation of resolution due to the heating of optical elements will become even more severe [16; 17; 18].

Several types of active optical devices and Adaptive Optics (AO) concepts have been proposed for compensation of thermally induced aberrations in gravitational wave detectors [20; 21; 253; 269; 276]. In optical lithography machines, active optical elements have been introduced for correction of wavefront aberrations [7; 12]. Furthermore, one of the possible methods for correction of TIWA in the next generation of lithographic machines (EUV lithographic - EUVL machines), is to use Deformable Mirrors (DMs) and AO techniques [2; 26; 277]. However, correction of TIWA in lithographic machines might not be possible using the standard AO techniques [33; 34]. This is because the standard AO techniques require that a WaveFront Sensor (WFS) and a DM are connected using a real-time feedback. This condition is not fulfilled in lithography machines because wavefront aberrations can be measured only before and after the exposure of certain number of wafers [7]. Furthermore, because the measurement time decreases the wafer throughput, the total measurement time should be as small as possible. Ideally, the wavefront should be measured only at the beginning of the exposure process and when exposure conditions change. Obviously, the classical feedback control algorithms, on which most of the AO techniques rely upon, cannot be applied in this scenario. *Hence, a new type of adaptive optics control algorithms needs to be developed. These algorithms should be able to predict the future behavior of the wavefront aberrations and to compensate them using only past measurement data*.

In this Chapter we experimentally demonstrate the proof of concept for predictive control of thermally induced aberrations. On the basis of the model of TIWA and using only past wavefront measurements, the proposed controller is able to predict and to compensate the future wavefront aberrations. Furthermore, the proposed AO controller is able to correct wavefront aberrations even when some parameters of the prediction model are unknown. Beside optical lithography, the proposed predictive controller can be used in other high power optical systems where it is not possible to establish a real-time feedback between the controller and the wavefront sensor.

This Chapter is organized as follows. In Section 10.2, we present a problem formulation and we describe an AO experimental setup. In Section 10.3 and in Section 10.4, we develop the predictive control algorithm and present the experimental results, respectively. In Section 10.6, the conclusions are drawn.

10.2 Problem description and experimental setup

Optical lithography is a technology that uses electromagnetic radiation to project mask patterns onto a photo-resist on a semiconductor wafer. The main components of a lithographic machine are: source, illumination optics, reticle stage (with a mask), projection optics and the wafer stage [1].

Because it is not possible to establish a real-time feedback between a wavefront

sensor and a controller, the correction of TIWA in a lithography machine should be done in a predictive manner. Namely, the controller should be able to anticipate the future wavefront aberrations and to correct them. Obviously, the anticipation of the future wavefront aberrations must be done on the basis of the model of the TIWA. Furthermore, in order to do accurate prediction, the controller also needs to use past wavefront measurements (the main reason why the controller needs past wavefront measurements for prediction will be explained in Section 10.3).

The model of TIWA describes how exposure conditions influence the dynamical behavior of the wavefront aberrations. The exposure conditions that dominantly influence the wavefront aberrations are numerical aperture, source shape, reticle and mask pattern diffraction, exposure dose, throughput and resist stack [7]. In this thesis these exposure conditions will be called *the inputs of the TIWA model*. The TIWA model consists of two main parts [7]. The first part relates the inputs with the distribution of exposure energy (intensity distribution) over the surfaces of the optical elements. This relation is established by computing the full mask diffraction orders which are then convoluted with the illumination source to obtain the diffraction pattern [7]. The computed diffraction pattern determines the intensity distribution on the optical elements. The second part consists of thermoelastic Partial Differential Equations (PDEs), that relate the intensity distribution with the temperature change and deformations of the optical elements [22; 57].

The model of TIWA can be obtained using two approaches. The first approach relies on first principles modeling (for example, deriving the model by discretizing the thermoelastic equations using the finite element method). The second modeling approach is to identify the model directly from experimental data. During the testing and calibration of a lithographic machine, wavefront measurements can be collected and they can be used to estimate the model using system identification techniques [58]. However, during the exposure of the wafers, the exposure conditions are usually different than the ones that were used in machine calibration and testing (for example, a mask that is used during the exposure is usually different from the mask that was used during calibration). In mathematical terms, this means that the inputs of the TIWA model are different from the inputs that were used in model estimation. There are two ways to overcome this problem. If the new exposure conditions are precisely known then the model can be updated. For example, if the mask is changed, then the intensity distribution on optical elements changes. This new intensity distribution can be calculated using the knowledge of the mask's geometry [7]. However, this is a computationally challenging problem that needs to be solved in real-time [1]. Another approach, that we propose in this thesis, is to estimate unknown inputs or to estimate intensity distribution on optical elements from the measurements of wavefront aberrations.

To demonstrate the proof of concept of predictive wavefront correction, we built an experimental AO system that is illustrated in Figure 10.1.

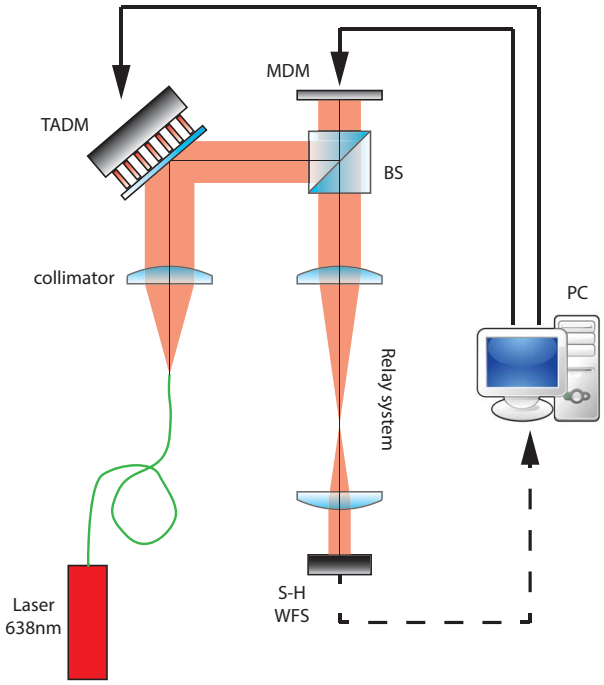


Figure 10.1: Experimental setup consisting of two DMs that was used to demonstrate the performance of the predictive control algorithm.

The AO system consists of two DMs. The first DM is a commercial Membrane DM (MDM) produced by Adaptica Srl [261]. The MDM has 48 actuators and it is used as a wavefront correction mirror. This mirror corresponds to an active optical element in a lithographic machine that is used for wavefront correction. The second mirror is a prototype of a Thermally Actuated DM (TADM) that was developed by Eindhoven University of Technology [2; 26]. The TADM consists of the two cylindrical plates that are connected by 19 actuators. The actuators consist of aluminum rods with heating coils warped around them. When voltage is applied to an actuator it heats up and elongates. As it elongates, it exerts a mechanical force that deforms the mirror and a supporting back plate (for more details see [26]). The TADM is used to simulate the effect of TIWA in a lithographic machine. Accordingly, its inputs (voltages applied to actuators) serve as exposure conditions that determine the wavefront aberrations.

The light of a semiconductor laser, working at the wavelength of $\lambda = 638$ [nm], is first collimated and then reflected by the TADM. The reflected light goes through the Beam Splitter (BS) and is again reflected by the MDM. The wavefront, at the surface of the MDM, is measured by a Shack-Hartmann Wavefront Sensor (S-H WFS) (Thorlabs WFS S300-14AR, 1.3 Mpixel, $\lambda/50$ rms accuracy). The MDM is optically conjugate to the S-H WFS (this is achieved using a relay system consisting of two spherical lenses).

A predictive controller is implemented on a standard Personal Computer (PC).

The PC sends two signals. One signal, that is calculated by the predictive controller, is sent to MDM. At the same time, another actuation signal is sent to TADM to introduce dynamical wavefront aberrations in the system. The predictive controller and the controller for TADM are implemented in MATLAB. The predictive controller cannot influence the signal that is sent to TADM. The PC receives measurements from the S-H WFS. However, only at certain discrete time instants, the predictive controller is able to receive wavefront measurements. In order to indicate this, the line connecting the WFS and PC is dashed. This way, we simulate the situation in a real lithographic machine, where the wavefront aberrations can be measured only before and after the exposure process. All the connections in the system are established using LabVIEW environment.

In the AO system, the control sampling period is 2 s. This value is chosen on the basis of the step response analysis of the TADM (previously characterized in Chapter 9). The measured wavefront at the discrete time instant[1] k, is represented using a Zernike polynomial expansion (Noll [263]):

$$\mathbf{\Phi}(x,y,k) = \sum_{i=1}^{36} \alpha_i(k) Z_i(x,y) \tag{10.1}$$

where $\mathbf{\Phi}(x,y,k)$ is the wavefront, $\alpha_i(k)$ is the i^{th} Zernike expansion coefficient and $Z_i(x,y)$ is the i^{th} Zernike polynomial. In Chapter 9, it has been experimentally demonstrated that TADM can produce variation of essentially the first 9 Zernike coefficients (omitting the piston). We therefore focus on correction of these 9 Zernike coefficients that will be grouped in a vector (see also Remark 10.1):

$$\mathbf{y}_T(k) = \begin{bmatrix} \alpha_2(k) & \dots & \alpha_{10}(k) \end{bmatrix}^T \tag{10.2}$$

The state-space model of the TADM has been identified in Chapter 9 and has the following form:

$$\mathbf{x}(k+1) = A\mathbf{x}(k) + Bv \tag{10.3}$$

$$\mathbf{y}_T(k) = C\mathbf{x}(k) \tag{10.4}$$

where $\mathbf{x}(k) \in \mathbb{R}^6$ is a state vector, $v \in \mathbb{R}$ is an input, and $A \in \mathbb{R}^{6\times 6}$, $B \in \mathbb{R}^{6\times 1}$ and $C \in \mathbb{R}^{9\times 6}$ are the identified system matrices.
The input of the TADM is $v = r_3^2$, where r_3 is a voltage applied to the 3rd actuator of the TADM. For simplicity, we actuate only one channel of the TADM (the generalization for all 19 actuators of the TADM is straightforward).
In Chapter 9 it has been demonstrated that the state-space model of the 4^{th} order (the dimension of the state $\mathbf{x}(k)$) is able to relatively accurately predict the dynamical behavior of the TADM. However, in order to achieve even better prediction accuracy, in this Chapter we are using the identified model of the 6^{th} order.

From the prediction point of view, the model of the TADM (10.3)-(10.4) represents the model of TIWA in a lithographic machine. Conceptually, the input v of the state-space model (10.3)-(10.4) represents the exposure conditions that are inputs

[1] The total time between k and $k+1$ is equal to the sampling period.

of the wavefront aberrations model (another modeling option is to consider v as an intensity distribution on optical elements, however, we will not develop this idea in this thesis). This input will be called *the disturbance input*. We assumed that the disturbance input is time independent because in a real system, exposure conditions do not change during the exposure of a relatively large number of wafers. Because during exposure the light source is turned on and off with high temporal frequency, the intensity distribution on optical elements oscillates in time. However, it can be easily shown that the dynamical response of the thermoelastic system of PDEs to a high frequency intensity distribution can be approximated by response to a static intensity distribution. That is, from the modeling point of view high frequency intensity distribution can be approximated by a static intensity distribution.
When however the exposure conditions change, the intensity distribution also changes. In the AO setup, this can be simulated by changing the magnitude of the voltage v that is applied to the TADM.
Physically speaking, the state vector $\mathbf{x}(k)$ corresponds to temperatures of the optical elements in a lithography machine. That is, it corresponds to the states of the TIWA model.

The model of the MDM has the following form [257; 258]:

$$\mathbf{y}_M(k) = M\mathbf{u}(k), \qquad \mathbf{u}(k) = \begin{bmatrix} q_1^2(k) & q_2^2(k) & \dots & q_{48}^2(k) \end{bmatrix}^T \tag{10.5}$$

where $q_i(k)$ is a voltage applied to the i^{th} actuator of the MDM, $\mathbf{y}_M(k) \in \mathbb{R}^9$ is a shape of the MDM described by the 9 Zernike coefficients and M is the influence matrix (see also Remark 10.1). The vector $\mathbf{u}(k) \in \mathbb{R}^{48}$ will be called *the control input*. We identified M using the identification procedure explained in [260; 273]. The total wavefront $\mathbf{y}(k)$, that is measured by the S-H WFS, is a sum of the wavefronts produced by the two DMs:

$$\mathbf{y}(k) = \mathbf{y}_T(k) + \mathbf{y}_M(k) \tag{10.6}$$

In reality the total wavefront $\mathbf{y}(k)$ is corrupted by a S-H WFS measurement noise. However, for simplicity in (10.6) we neglect the effect of the S-H WFS measurement noise. Combining (10.3), (10.5) and (10.6), we arrive at the model of the experimental setup:

$$\mathbf{x}(k+1) = A\mathbf{x}(k) + Bv \tag{10.7}$$

$$\mathbf{y}(k) = C\mathbf{x}(k) + M\mathbf{u}(k) \tag{10.8}$$

In order to define the prediction problem, we need to define the past and the future horizons. At the discrete time instant k, *the past horizon* is the set of discrete-time instants $\{k-p, k-p+1, \dots, k-1, k\}$ and *the future horizon* is the set $\{k+1, k+2, \dots, k+f\}$, for some appropriately chosen values of p and f. The past and the future horizons are illustrated in Figure 10.2.

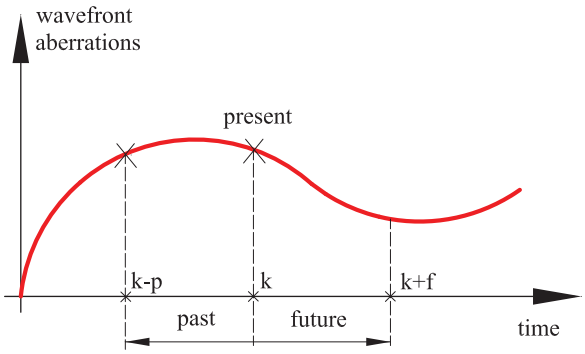


Figure 10.2: Past, present and future.

Next we define the following vectors:

$$\mathbf{v}_p = \underbrace{\begin{bmatrix} v \\ v \\ \vdots \\ v \end{bmatrix}}_{p \text{ entries of } v}, \quad \mathbf{u}_p = \begin{bmatrix} \mathbf{u}(k-p) \\ \mathbf{u}(k-p+1) \\ \vdots \\ \mathbf{u}(k) \end{bmatrix}, \quad \mathbf{y}_p = \begin{bmatrix} \mathbf{y}(k-p) \\ \mathbf{y}(k-p+1) \\ \vdots \\ \mathbf{y}(k) \end{bmatrix} \tag{10.9}$$

The vectors $\mathbf{v}_p \in \mathbb{R}^p$, $\mathbf{u}_p \in \mathbb{R}^{48(p+1)}$ and $\mathbf{y}_p \in \mathbb{R}^{9(p+1)}$ will be called *the past disturbance input*, *the past control input* and *the past wavefront measurement*, respectively. Similarly, we define the following vectors:

$$\mathbf{v}_f = \underbrace{\begin{bmatrix} v \\ v \\ \vdots \\ v \end{bmatrix}}_{f \text{ entries of } v}, \mathbf{u}_f = \begin{bmatrix} \mathbf{u}(k+1) \\ \mathbf{u}(k+2) \\ \vdots \\ \mathbf{u}(k+f) \end{bmatrix}, \quad \mathbf{y}_f = \begin{bmatrix} \mathbf{y}(k+1) \\ \mathbf{y}(k+2) \\ \vdots \\ \mathbf{y}(k+f) \end{bmatrix} \tag{10.10}$$

The vectors $\mathbf{v}_f \in \mathbb{R}^f$, $\mathbf{u}_f \in \mathbb{R}^{48f}$ and $\mathbf{y}_f \in \mathbb{R}^{9f}$ will be called *the future disturbance input*, *the future control input* and *the future wavefront*, respectively. We assumed that the same disturbance input v is an element of both $\mathbf{v}_p$ and $\mathbf{v}_f$ because in a lithography machine the exposure conditions do not change for several hours.
The first predictive control problem is formulated as follows:

Predictive control problem 1. Using the past wavefront measurements, past control inputs, past and future disturbance input, and the model (10.7)-(10.8), at the time instant k find the future control input that will correct future wavefront aberrations.

In the second predictive control problem, we will be interested in correction of

wavefront aberrations when the disturbance input is not known a priori. This case corresponds to the scenario when exposure conditions in a lithographic machine are either not modeled or they are not known a priori.

***Predictive control problem* 2.** Using the past wavefront measurement, past control input, and the model (10.7)-(10.8), at the time instant k estimate the disturbance input v and find the future control input that will correct future wavefront aberrations

10.3 Predictive control strategies

We solve the predictive control problems by deriving an output predictor from the state-space model (10.7)-(10.8) [58]. Similar predictive control strategy can be found in [278]. From (10.7)-(10.8) we have:

$$\underline{O}_p \mathbf{x}(k-p) = \mathbf{y}_p - \underline{I}_p \mathbf{v}_p - \underline{D}_p \mathbf{u}_p \tag{10.11}$$

where $\mathbf{y}_p$, $\mathbf{v}_p$ and $\mathbf{u}_p$ are defined in (10.9) and

$$\underline{O}_p = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^p \end{bmatrix}, \quad \underline{I}_p = \begin{bmatrix} 0 & 0 & \dots & \dots \\ CB & 0 & \dots & \dots \\ CAB & CB & 0 & \dots \\ \vdots & & \ddots & \\ CA^{p-1}B & CA^{p-2}B & \dots & CB \end{bmatrix},$$

$$\underline{D}_p = \underbrace{\begin{bmatrix} M & 0 & 0 & \dots & 0 \\ 0 & M & 0 & \dots & 0 \\ 0 & 0 & M & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & M \end{bmatrix}}_{p+1 \text{ blocks}} \tag{10.12}$$

where $\underline{O}_p \in \mathbb{R}^{9(p+1)\times 6}$, $\underline{I}_p \in \mathbb{R}^{9(p+1)\times p}$ and $\underline{D}_p \in \mathbb{R}^{9(p+1)\times 48(p+1)}$. In (10.12), the matrices A, B, C and M are the system matrices of the model (10.7)-(10.8). In order to ensure that equation (10.11) can be uniquely solved for $\mathbf{x}(k-p)$, the length p of the past horizon has to be larger than the observability index [279] of the system (10.7)-(10.8). That is, p has to be large enough such that $\underline{O}_p$ has full column rank. Assuming that this is the case, from (10.11) we have:

$$\mathbf{x}(k-p) = \underline{O}_p^{\dagger} \left(\mathbf{y}_p - \underline{I}_p \mathbf{v}_p - \underline{D}_p \mathbf{u}_p \right) \tag{10.13}$$

where $\underline{O}_p^{\dagger}$ denotes the pseudo-inverse of $\underline{O}_p$. From (10.3) we have:

$$\mathbf{x}(k) = A^p \mathbf{x}(k-p) + \underline{R}_{p-1} \mathbf{v}_p \tag{10.14}$$

where $\underline{R}_{p-1} = \begin{bmatrix} A^{p-1}B & A^{p-2}B & \dots & AB & B \end{bmatrix}$. Substituting (10.13) in (10.14), we obtain:

$$\mathbf{x}(k) = A^p \underline{O}_p^\dagger \mathbf{y}_p - A^p \underline{O}_p^\dagger \underline{D}_p \mathbf{u}_p + \left(\underline{R}_{p-1} - A^p \underline{O}_p^\dagger \underline{I}_p \right) \mathbf{v}_p \tag{10.15}$$

Similarly to (10.11), we have:

$$\mathbf{y}_f = \underline{O}_{f-1} A \mathbf{x}(k) + \overline{I}_f \mathbf{v}_f + \underline{D}_{f-1} \mathbf{u}_f \tag{10.16}$$

where

$$\underline{O}_{f-1} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{f-1} \end{bmatrix}, \overline{I}_f = \begin{bmatrix} CB & 0 & \dots & \dots \\ CAB & CB & 0 & \dots \\ \vdots & & \ddots & \ddots \\ CA^{f-1}B & CA^{f-2}B & \dots & CB \end{bmatrix},$$

$$\underline{D}_{f-1} = \underbrace{\begin{bmatrix} M & 0 & 0 & \dots & 0 \\ 0 & M & 0 & \dots & 0 \\ 0 & 0 & M & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & M \end{bmatrix}}_{f \text{ blocks}} \tag{10.17}$$

Substituting (10.15) in (10.16), we obtain:

$$\begin{aligned} \mathbf{y}_f = & \underbrace{\underline{O}_{f-1} A^{p+1} \underline{O}_p^\dagger \mathbf{y}_p - \underline{O}_{f-1} A^{p+1} \underline{O}_p^\dagger \underline{D}_p \mathbf{u}_p - \underline{O}_{f-1} A^{p+1} \underline{O}_p^\dagger \underline{I}_p \mathbf{v}_p}_{\text{The effect of the initial state } \mathbf{x}(k-p)} \\ & + \underline{O}_{f-1} A \underline{R}_{p-1} \mathbf{v}_p + \overline{I}_f \mathbf{v}_f + \underline{D}_{f-1} \mathbf{u}_f \end{aligned} \tag{10.18}$$

The equation (10.18) tells us how the future wavefront depends on the past wavefront, past and future disturbance inputs and past and future control inputs. This equation will be called *the prediction equation*. The first three therms in the prediction equation originate from the initial state $\mathbf{x}(k-p)$ in the system (10.7)-(10.8). That is, in the general case, the prediction cannot be done only on the basis of the model but the past wavefront measurements have to be taken into account. In the special case when $\mathbf{x}(k-p) = 0$ (that is, when initially the system is in the thermal equilibrium), the prediction equation can be simplified by neglecting the first three terms. In this special case, we do not need any past wavefront measurements to do the prediction.

The observability index of the identified model (10.7)-(10.8) is 2. That is, in theory, we only need 2 measurements samples of the past wavefront to predict the future behavior. However, in practice, the past measurements are corrupted by the S-H WFS measurement noise. The negative effect of the measurement noise on the prediction performance, can be minimized by selecting larger p. For convenience

we will write the prediction equation as follows:

$$\mathbf{y}_f = \mathbf{s} + \underline{D}_{f-1}\mathbf{u}_f \tag{10.19}$$

where

$$\mathbf{s} = \underline{O}_{f-1}A^{p+1}\underline{O}_p^{\dagger}\mathbf{y}_p - \underline{O}_{f-1}A^{p+1}\underline{O}_p^{\dagger}\underline{D}_p\mathbf{u}_p + \underline{O}_{f-1}A\left(\underline{R}_{p-1} - A^p\underline{O}_p^{\dagger}\underline{I}_p\right)\mathbf{v}_p + \overline{\underline{I}}_f\mathbf{v}_f \tag{10.20}$$

Solution of the prediction problem 1.

The first prediction problem will be solved using two approaches. In the first approach the future control input is determined by solving the following unconstrained least-squares problem:

$$\min_{\mathbf{u}_f}\{\mathbf{y}_f^T\mathbf{y}_f + \mathbf{u}_f^T W\mathbf{u}_f\} \tag{10.21}$$

where the weighting matrix $W \in \mathbb{R}^{48f\times 48f}$ is defined as follows:

$$W = \gamma QQ^T, \quad Q = \begin{bmatrix} I & -I & 0 & \dots & & \\ 0 & I & -I & 0 & \dots & \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & & \dots & 0 & I & -I \\ & & & \dots & 0 & I \end{bmatrix} \tag{10.22}$$

In (10.22), $I \in \mathbf{R}^{48\times 48}$ is an identity matrix and γ is a positive regularization parameter. The entries of the matrix W penalize the weighted 2-norm of the difference between two consecutive elements of $\mathbf{u}_f$. This way, we can control the convergence speed of the predictive control algorithm. The weighting matrix can also take another form, that penalizes the energy of the future input or that improves the conditioning of the optimization problem, for details see for example [273]. The solution of (10.21) is:

$$\hat{\mathbf{u}}_f = -\left(\underline{D}_{f-1}^T\underline{D}_{f-1} + W\right)^{-1}\underline{D}_{f-1}^T\mathbf{s} \tag{10.23}$$

The vector $\hat{\mathbf{u}}_f$ will be referred to as *the unconstrained predictive control input*. As it will be demonstrated experimentally (see Section 10.4), some of the elements of $\hat{\mathbf{u}}_f$ can be negative. Since the elements of $\hat{\mathbf{u}}_f$ are squares of the control voltages, negative entries of this vector are not feasible and we set them to zero. The consequence of this is a slower convergence of the predictive controller and increase of the control error (the predictive controller is suboptimal). In order to overcome this problem, we add "hard" constraints to the cost function (10.21):

$$\min_{\mathbf{u}_f}\{\mathbf{y}_f^T\mathbf{y}_f + \mathbf{u}_f^T W \mathbf{u}_f\}$$
$$\text{subject to } \mathbf{a}_1 \preceq \mathbf{u}_f \preceq \mathbf{a}_2 \tag{10.24}$$

where $\mathbf{a}_1 \in \mathbb{R}^{48f}$ is the vector of zeros, and $\mathbf{a}_2 \in \mathbb{R}^{48f}$ is a vector of ones. In (10.24), $\preceq$ denotes element-wise less-than-equal mathematical symbol. With the condition $\mathbf{u}_f \preceq \mathbf{a}_2$, we prevent saturation of the MDM. The solution of (10.24) will be referred to as *the constrained predictive control input*. Because the solution of (10.24) cannot be found in the closed form, we have determined it using MATLAB function lsqlin( ).

Solution of the prediction problem 2.

To solve the second prediction problem we need to estimate unknown disturbance input v from the past wavefront measurements. From (10.11) we have:

$$\mathbf{b} = L\mathbf{w}, \qquad \mathbf{w} = \begin{bmatrix}\mathbf{x}(k-p)\\ v\end{bmatrix} \tag{10.25}$$
$$\text{where} \quad \mathbf{b} = \mathbf{y}_p - \underline{D}_p\mathbf{u}_p, \qquad L = \begin{bmatrix}\underline{O}_p & \underline{I}_p\mathbf{q}\end{bmatrix}, \qquad \mathbf{q} = \underbrace{\begin{bmatrix}1 & 1 & \dots & 1\end{bmatrix}^T}_{\text{p entries}}$$

The unknown vector $\mathbf{w}$ consists of the initial state $\mathbf{x}(k-p)$ and the disturbance input. In reality the the past output vector $\mathbf{y}_p$ is corrupted by a measurement noise. That is, the vector $\mathbf{b}$ is only approximately equal to $L\mathbf{w}$. We want determine $\mathbf{w}$ such that the difference $\mathbf{b} - L\mathbf{w}$ is as small as possible. This can be done by solving the following least-squares problem:

$$\min_{\mathbf{w}}\{(\mathbf{b} - L\mathbf{w})^T(\mathbf{b} - L\mathbf{w})\} \tag{10.26}$$

Assuming that L has full column rank. The solution of (10.26) is: $\hat{\mathbf{w}} = L^{\dagger}\mathbf{b}$. By estimating $\mathbf{w}$ we have at the same time estimated the initial state and the disturbance input v. By substituting v in (10.21) or (10.24), we can easily compute the future control input.

Remark 10.1 *The predictive control algorithms are developed to correct for the first* 9 *Zernike coefficients that are produced by the TADM. On the other hand, the MDM can produce variation of more than* 9 *Zernike coefficients. This implies that when the predictive control inputs are applied to the MDM, the first* 9 *Zernike coefficients are corrected, while the higher order (higher than 9) Zernike coefficients are also changed. The variation of the higher order coefficient is not explicitly controlled by the predictive algorithms. However, we observed that when the predictive control inputs are applied to the MDM, the variation of the higher order coefficients is not significant. This is mainly because of the physical limits of the MDM. Due to this fact and for simplicity of presentation, in this chapter we assumed that the outputs of the MDM and TADM consist only of the first* 9 *Zernike*

coefficients.
The developed predictive control algorithms can be straightforwardly generalized such that they at the same time control the lower order Zernike modes, produced by the TADM and the higher order modes. This can be done by modifying the output equation of the state-space model (10.7)-(10.8)*, as follows:*

$$\mathbf{y}_1(k) = \begin{bmatrix} C \\ 0_1 \end{bmatrix} \mathbf{x}(k) + M_1 \mathbf{u}(k) \tag{10.27}$$

where $\mathbf{y}_1(k) \in \mathbb{R}^{Q_1}$ *is a vector of the first* Q_1 *Zernike coefficients that can be produced by the MDM,* $0_1 \in \mathbb{R}^{(Q_1-9)\times 6}$ *and* $M_1 \in \mathbb{R}^{Q_1 \times 48}$ *is a full influence matrix of the MDM.*

10.4 Experimental results

In this section we present the experimental results of validating the predictive control strategy on the AO setup described in Section 10.2. To create nonzero initial states in the system, and thus to make the prediction problem more challenging, before each experiment the disturbance input of 15% of the maximal voltage value is applied for 30 [s]. In order to clearly distinguish the controlled and uncontrolled wavefronts, we do not control the MDM during the past horizons. That is, in the prediction equation (10.18), the past control input $\mathbf{u}_p$ is zero.

In Fig. 10.3(a) we show the performance of the unconstrained predictive control strategy. The performance of the predictive controller was quantified by computing the 2-norm of the measured Zernike coefficients (denoted by $\|\mathbf{y}(k)\|_2$ in all the figures). We assume that the disturbance input is known a priori. Two experiments were performed. In both experiments, the disturbance input equal to the 35% of the maximal voltage value was applied to the TADM. In the first experiment, the MDM was not controlled. The uncontrolled wavefront is represented by a red dashed-dotted line in Fig 10.3(a). The wavefront $\|\mathbf{y}(k)\|_2$ starts from a value of $0.02\ \lambda$. This is because of the non-zero initial state in the system that was created by applying a 15% disturbance input before the beginning of the experiment. In about 20 discrete time samples (or 40 s in total), the wavefront approximately reaches its steady state $\|\mathbf{y}(\infty)\|_2 \approx 0.09\lambda$
In the second experiment, the unconstrained predictive control input was calculated using (10.23) and it was applied to the MDM. This was done at two time instants $k = 19$ and $k = 86$. The controlled wavefront is represented by a thick black line in Fig. 10.3(a). The future and past horizons were: $p = 17$ and $f = 30$. That is, 17 past measurements of the wavefront were taken, and the wavefront is predicted and controlled for 30 times instants in the future.
The elements of the unconstrained predictive control input, at an arbitrary time instant, are illustrated in Fig. 10.5(b). Some of the elements of the calculated predictive control input were negative and they were set to zero (marked by stars in Fig.10.5(b)).

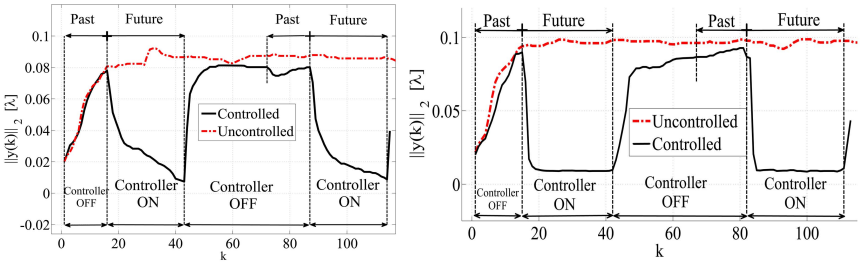


Figure 10.3: Performance of the predictive control strategy (a) Unconstrained controller; (b) Constrained controller

In Fig. 10.3(b) we show the performance of the constrained predictive control strategy. The predictive controller was calculated by solving (10.24). Similarly to the test of unconstrained predictive control strategy, two experiments were performed (the future and past horizons were the same like in the case of the unconstrained predictive control). From Fig. 10.3, it can be observed that both controllers are able to accurately correct wavefront aberrations ($\|\mathbf{y}(k)\|_2$ is below 0.015λ). By comparing Figs. 10.3(a) and 10.3(b), it can be observed that the constrained predictive controller corrects wavefront aberrations faster than the unconstrained one. Moreover, we have observed the constrained predictive controller achieves a smaller value of $\|\mathbf{y}(k)\|_2$ than the unconstrained predictive controller (this is difficult to see from Fig. 10.3). All this is because some of the elements of the unconstrained predictive control input were set to zero. That is, we sacrificed a part of the performance of the unconstrained controller to get physically realizable voltages.

In Fig. 10.4 we show how the convergence rate of the unconstrained predictive controller depends on the weighting parameter γ.

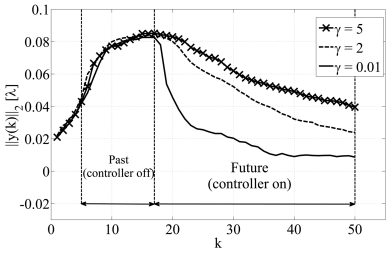


Figure 10.4: Convergence rate of the unconstrained predictive control strategy for different values of γ.

In Fig. 10.5(a), we show how the convergence rate of an arbitrary voltage of unconstrained predictive control input depends on γ. From Figs. 10.4 and 10.5(a),

we see that when γ is increased, the convergence of the predictive controller gets slower. This is because larger γ implies stronger penalization of the difference between two consecutive future control inputs. Similar results were observed for the case of the constrained controller (for brevity, we omit these results).

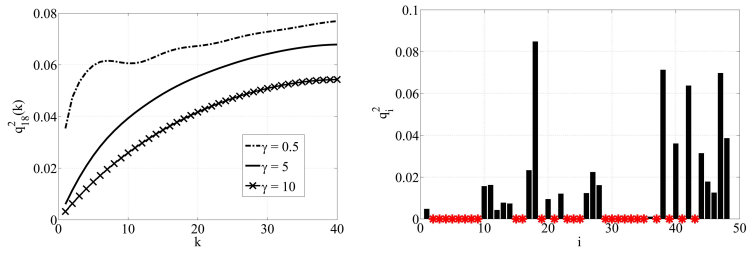


Figure 10.5: (a) Convergence of an arbitrary channel of the unconstrained predictive control input; (b) Calculated unconstrained control input. Negative values of the channels are set to zero.

Finally, in Fig. 10.6 we show the performance of the constrained predictive controller when unknown disturbance input is estimated (denoted by "estimated d. input").

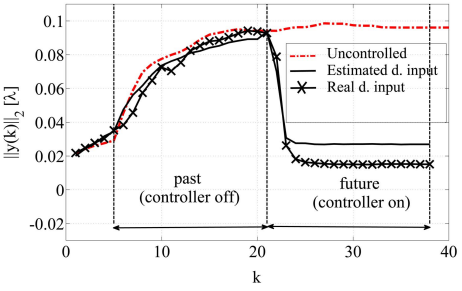


Figure 10.6: The performance of the predictive controller when unknown disturbance input is estimated.

For comparison in Fig. 10.6, we also show the performance of the constrained predictive controller when the disturbance input is known a priori (denoted by "real d. input"). The unknown disturbance input is estimated by solving the least-squares problem (10.26), and the estimate is substituted in (10.24) to derive the constrained predictive control input. It can be observed that the controller that is based on the estimated disturbance input has a slightly worse performance than the controller that is based on the "real" disturbance input (difference of 0.01λ between the converged future wavefronts). This is because the estimate

of the disturbance input is affected by the errors originating from the S-H WFS measurement noise and the model uncertainties of the state-space model (10.3)-(10.4). These model uncertainties originate from the stochastic nature of the heat convection that occurs between the actuators of the TADM and the surrounding air [169]. We observed that the estimation accuracy of the disturbance input v can be improved by increasing the length of the past horizon p. This is because more measurement data decrease the negative effect of the noise on the least-squares estimate.

10.5 Some comments on the generalization of the predictive control law for large-scale interconnected systems

In this section we will briefly explain how the theoretical framework presented in Chapter 3 can be used to develop distributed and centralized predictive control algorithms for large-scale interconnected systems. These control algorithms can be used for correction of TIWA in optical lithography machines.

To illustrate the main idea, let us assume that the disturbance model (10.3)-(10.4) is replaced by the global state-space model (2.65) (the discretized 3D heat equation). The global matrix $\underline{A}$ of this state-space model is a sparse multi-banded matrix and the matrices $\underline{B}$ and $\underline{C}$ are block diagonal. Using the structure preserving lifting technique proposed in Chapter 3 and following the derivation presented in Section 10.3, we can obtain a structured version of the prediction equation (10.18). If $p \ll N$ and $f \ll N$ then in the structured prediction equation all matrices except the pseudo-inverse of the structured observability matrix will be multi-banded or block diagonal matrices. Using the approximation framework proposed in Chapter 3, the pseudo-inverse of the structured observability matrix can be approximated by a sparse multi-banded matrix. This way we can derive a sparse prediction equation. On the other hand, using the the approximation algorithms presented in Chapter 3, we can approximate the matrix $\left(\underline{D}_{f-1}^T \underline{D}_{f-1} + W\right)^{-1}$ of the unconstrained predictive control input (10.23) by a sparse, multi banded matrix.
That is, the approximation framework proposed in Chapter 3 can be used to develop distributed (unconstrained) predictive control algorithms for large-scale interconnected systems. Moreover, the developed predictive controllers can be implemented in the centralized manner with complexity that scales linearly with the number of local subsystems.

The problem of solving the constrained prediction problem (10.24) for large-scale interconnected systems is more challenging. The main difficulty is that the solution of the optimization problem (10.24) cannot be found in the closed form. There are several iterative methods for solving this optimization problem. One of them is the interior point method [280]. In each iteration of the interior point method, a linearized system of equations needs to be solved. Because the global system matrices and matrices in the structured prediction equation are sparse multi-banded

matrices, the linearized system is also sparse. Using the approximation framework presented in Chapter 3, the inverse of the main matrix of the linearized system can be approximated by a sparse matrix. In this way, iterative predictive control algorithms can be developed.
However, this strategy introduces approximation errors in each iteration of the interior point method. The analysis of the effect of the approximation errors on the convergence of the interior point method might be difficult.

10.6 Conclusion

In this Chapter we have experimentally demonstrated the proof of concept for the predictive correction of TIWA. The experimental results show that the predictive controller is able to correct wavefront aberrations using a relatively small number of past wavefront measurements. Furthermore, we have demonstrated that the predictive controller is able to correct wavefront aberrations even when the inputs of the wavefront prediction model are not known a priori. The proposed controller can be used in optical lithography as well as in other high-power optical systems where it is not possible to establish real-time feedback between a wavefront sensor and a controller.

11 CHAPTER

Conclusions and recommendations

In this final chapter we present general conclusion about the algorithms and experimental results presented in this thesis. We also discuss some future research directions.

11.1 Conclusion about the theoretical part of the thesis

In this thesis we developed computationally efficient methods for estimation and identification of large-scale interconnected systems described by sparse banded or multi-banded state-space matrices. These systems originate from finite difference or finite element approximations of partial differential equations.

We proved that inverses of lifted system matrices and inverse of the finite-time observability (controllability) Gramian of interconnected systems, are off-diagonally decaying matrices. Furthermore, we demonstrated that these inverses can be approximated by sparse (multi) banded matrices with $O(N)$ computational complexity and $O(N)$ memory complexity. On the basis of these results we developed:

1. *The computationally efficient centralized and distributed Moving Horizon Estimation (MHE) methods for large-scale interconnected systems.* The distributed MHE method estimates the state of a local subsystem using only local input-output data. In contrast to the distributed estimation algorithms available in the literature, the developed distributed MHE method is not relying on the consensus algorithms and it has a simple, analytic form (closed-form).

   The computational complexity of the centralized MHE method scales linearly with the number of local subsystems. Consequently, it can be used for state estimation of interconnected systems that consist of an extremely large number of local subsystems. We also developed the computationally efficient MHE method for large-scale, descriptor state-space models.

2. *The decentralized subspace identification method for large scale interconnected systems.* First, we proved that the state of a local subsystem can be approximated by a linear combination of inputs and outputs of the local subsystems that are in its neighborhood. Furthermore, we proved that for interconnected systems with well-conditioned, finite-time observability Gramians (or observability matrices), the size of this neighborhood is relatively small (compared to the total number of local subsystems). On the basis of these results, we developed a subspace identification algorithm that identifies the state-space model of a local subsystem from the local input-output data. Consequently, the proposed algorithm is computationally feasible for interconnected systems that have a large number of local subsystems.

3. *The parameter optimization method for identification of large-scale interconnected systems.* This method can be used for identification of structured state-space models of interconnected systems. The computational and memory complexity of the proposed method scale linearly with the number of local subsystems. The initial estimate for this method can be obtained using the developed subspace identification method.

The above summarized methods were developed for large-scale systems that can be described by linear state-space models. Like we showed in Chapter 2.4, the discretized thermoelastic equations can be transformed into a linear state-space form, if the Coefficient of Thermal Expansion (CTE) is constant. If the CTE is depending on the temperature, then the state-space representation of the thermoelastic equations has a linear state equation and a nonlinear output equation (for details see Chapter 2, Remark 2.2).
In Chapter 4, Section 4.3, we showed that the approximation algorithms that are presented in Chapter 3, can be used to implement the Newton observer[1] with linear computational complexity. The computationally efficient Newton observer can be used for state estimation of thermoelastic equations with temperature depending CTE. Furthermore, the approximation framework presented in Chapter 3 can be used to develop computationally efficient identification and control algorithms for certain classes of nonlinear, large-scale systems.
Another option for estimating the state of thermoelastic equations with the temperature depending CTE, is to approximate the nonlinear output equation by a piecewise linear function. With some modifications, the MHE methods proposed in Chapter 4 can be applied to such approximate state-space models.

11.2 Recommendations for future theoretical research

- For presentation clarity, in this thesis we restricted our attention to large-scale systems described by sparse banded or multi-banded system matrices. The methods presented in this thesis exploit the fact that inverses of sparse, banded matrices can be approximated by sparse, banded matrices. On the

[1]The Newton observer is a state estimation method for nonlinear systems that was developed mainly by P. E. Moraal and J. W. Grizzle, see for example [234; 235].

other hand, in [202; 203; 204; 205; 207; 208] it has been shown that sparse approximate inverses can be computed for a large class of sparse matrices. This implies that the approximation algorithms presented in [202; 203; 204; 205; 207; 208], can be used to generalize the methods proposed in this thesis to interconnected systems with more general interconnection patterns. However, this generalization requires a graph theoretic approach to estimation and control problems for large-scale systems.

- The theory and methods developed in this thesis naturally give rise to the following questions. Does the solution of the algebraic Riccati equation for finite-dimensional, large-scale interconnected systems, exhibit some form of off-diagonal decay? This question is closely related with the following question. Can the solution of the algebraic Riccati equation for large-scale systems, be approximated in a computationally efficient manner by a sparse, structured matrix? *Undoubtedly, the solution of this problem would have a significant impact in the field of systems and control. Namely, if the solution of the algebraic Riccati equation can be approximated by a sparse matrix, then it it possible to design distributed Linear Quadratic Gaussian (LQG) controllers for large-scale interconnected systems, which performance is very close to the centralized LQG controllers. We think that the approximation methodology proposed in this thesis, can be a starting point for solving these fundamental problems.*

- In this thesis we used the dropping strategies to speed up the computations of approximate inverses (see Chapter 3.4.4). However, we did not analyze how the dropping strategies influence the convergence of the Newton iteration and the Chebyshev method. Furthermore, we did not analyze the errors introduced by the dropping strategies. These problems are left for future research.

- The problem of identifying graph topologies (interconnection patterns) of interconnected systems has been studied in [242; 243; 244; 245; 246]. For simplicity, in thesis we assumed that the graph topologies, of interconnected systems that we want to identify, are known a priori. Can the graph topology identification techniques be combined with the identification methods proposed in this thesis?

11.3 Conclusion about the experimental part of the thesis

Apart from the contributions to systems and control theory, in this thesis we developed and experimentally verified several new Adaptive Optics (AO) control algorithms.

1. We showed that the norm optimal Iterative Learning Control (ILC) algorithm can be used for accurate control of the shape of a membrane deformable mirror. Furthermore, we analyzed how the parameters of the ILC algorithm influence the wavefront correction performance.

2. Using the subspace identification technique, we identified a low order, dynamical model of a prototype of a thermally actuated deformable mirror. The identified model can be used for high performance correction of static or slowly-varying wavefront aberrations in high-power optical systems.

3. Finally, we demonstrated the proof of principle for predictive control of thermally induced wavefront aberrations in high-power optical systems. We demonstrated that on the basis of the model of Thermally Induced Wavefront Aberrations (TIWA) and using past wavefront measurements, it is possible to predictively correct wavefront aberrations. The experimental results give a good first indication that the developed predictive controller can be implemented in optical lithography machines.

11.4 Recommendations for future research on control of thermally induced wavefront aberrations

To verify the proposed strategy for predictive control of TIWA on a real lithographic machine, a complete model of TIWA needs to be developed.
Namely, in this thesis we developed the second part of the TIWA model, that relates the heat flux distribution on the mirror surface with thermoelastic deformations (see the beginning of Section 1.5). The first part, that relates exposure conditions with the heat flux distribution, needs to be developed. Furthermore, the mask and possibly optical elements in the illumination system, also absorb a portion of the light. Consequently, heating of these optical elements induces wavefront aberrations. To obtain an accurate model of thermally induced wavefront aberrations, thermoelastic deformation of these optical elements also needs to be modeled.
While developing a complete model of thermally induced wavefront aberrations, a special attention needs to be paid to the structure of the model. Like we demonstrated in this thesis, by exploiting the sparsity structure of the thermoelastic equations, we were able to develop the linear computational complexity estimation algorithms. Because of the need for real-time control of wavefront aberrations, similar philosophy should be adopted while developing the optical model of the complete TIWA model.
Finally, future research should be focused on the generalization of the developed methods for the case when the CTE in thermoelastic equations depends on the temperature (the temperature depending CTE gives rise to an output nonlinearity). In Chapter 4.3 we showed that the approximation algorithms presented in Chapter 3, can be used to implement the Newton observer for nonlinear systems with linear computational complexity. Similar approach should be followed in the development of computationally efficient predictive control algorithms for the case when the CTE depends on the temperature[2].

[2]In Chapter 10.5, we briefly explained how the approximation algorithms can be used to develop computationally efficient predictive control algorithms for large-scale linear systems.

Bibliography

[1] C. Mack. *Fundamental Principles of Optical Lithography: The Science of Microfabrication*. Wiley, 2008.

[2] S.K. Ravensbergen. *Adaptive optics for extreme ultraviolet lithography : actuator design and validation for deformable mirror concepts*. PhD thesis, Technische Universiteit Eindhoven, 2012.

[3] ASML. Technical specifications of TWINSCAN NXT:1950i. `http://www.asml.nl/asml/show.do?lang=EN&ctx=46772&dfp_product_id=822`.

[4] ASML. Technical specifications of TWINSCAN NXE:3300B Saturn. `http://www.asml.com/asml/show.do?lang=EN&ctx=46772&dfp_product_id=842`.

[5] J. E. Bjorkholm. Euv lithography - the successor to optical lithography? *Intel Technical Articles*, 1998.

[6] C. Wagner and N. Harned. EUV lithography: Lithography gets extreme. *Nat Photon*, 4(1):24–26, jan 2010.

[7] J. Bekaert, L. Van Look, G. Vandenberghe, P. Van Adrichem, M.J. Maslow, J.-W. Gemmink, H. Cao, S. Hunsche, J.T. Neumann, and A. Wolf. Characterization and control of dynamic lens heating effects under high volume manufacturing conditions. In *SPIE Advanced Lithography*, pages 79730V–79730V. International Society for Optics and Photonics, 2011.

[8] J. N. Hanson. Analysis of the thermoelastic deformation of optical surfaces. *J. Opt. Soc. Am.*, 55(8):916–921, Aug 1965.

[9] D. H. Beak, J. P. Choi, T. Park, Y. S. Nam, Y. S. Kang, C.-H. Park, K.-Y. Park, C.-H. Ryu, W. Huang, and K.-H. Baik. Lens heating impact analysis and controls for critical device layers by computational method. In *Proc. SPIE 8683, Optical Microlithography XXVI*, pages 86831Q–86831Q–8, 2013.

[10] S. Halle, M. Crouse, A. Jiang, Y. van Dommelen, T. Brunner, B. Minghetti, M. Colburn, and Y. Zhang. Lens heating challenges for negative tone develop layers with freeform illumination: a comparative study of experimental vs. simulated results. *Proc. SPIE*, 8326:832607–832607–19, 2012.

[11] C. Bikcora, M. Van Veelen, S. Weiland, and W.M.J. Coene. Lens heating induced aberration prediction via nonlinear kalman filters. *Semiconductor Manufacturing, IEEE Transactions on*, 25(3):384–393, 2012.

[12] F. Staals, A. Andryzhyieuskaya, H. Bakker, M. Beems, J. Finders, T. Hollink, J. Mulkens, A. Nachtwein, R. Willekers, P. Engblom, T. Grunner, and Y. Zhang. Advanced wavefront engineering for improved imaging and overlay applications on a 1.35 na immersion scanner. *Proc. SPIE*, 7973:79731G, 2011.

[13] H. Chen, H. Yang, X. Yu, and Z. Shi. Simulated and experimental study of laser-beam-induced thermal aberrations in precision optical systems. *Appl. Opt.*, 52(18):4370–4376, Jun 2013.

[14] Y Fujishima, S. Ishiyama, S. Isago, A. Fukui, T. Yamamoto, H. Hirayama, T. Matsuyama, and Y. Ohmura. Comprehensive thermal aberration and distortion control of lithographic lenses for accurate overlay. In *SPIE proc*, volume 86831I, 2013.

[15] Y. Ohmura, T. Ogata, T. Hirayama, H. Nishinaga, T. Shiota, S. Ishiyama, S. Isago, H. Kawahara, and T. Matsuyama. An aberration control of projection optics for multi-patterning lithography. In *Proc. SPIE*, volume 7973, pages 79730W–79730W–11, 2011.

[16] K. Liu, Y. Li, F. Zhang, and M. Fan. Transient thermal and structural deformation and its impact on optical performance of projection optics for extreme ultraviolet lithography. *Japanese Journal of Applied Physics*, 46(10A):6568–6572, 2007.

[17] Y Li, K. Ota, and K. Murakami. Thermal and structural deformation and its impact on optical performance of projection optics for extreme ultraviolet lithography. *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures*, 21(1):127–129, 2003.

[18] P. A. Spence, S. E. Gianoulakis, C. D. Moen, M.P. Kanouff, A. Fisher, and A. K. Ray-Chaudhuri. System performance modeling of extreme ultraviolet lithographic thermal issues. *Journal of Vacuum Science Technology B: Microelectronics and Nanometer Structures*, 17(6):3034–3038, 1999.

[19] A. K. Ray-Chaudhuri, S. E. Gianoulakis, P. A. Spence, M. P. Kanouff, and C. D. Moen. Impact of thermal and structural effects on euv lithographic performance. In *Proc. SPIE*, volume 3331, pages 124–132, 1998.

[20] M. Kasprzack, B. Canuel, F. Cavalier, R. Day, E. Genin, J. Marque, D. Sentenac, and G. Vajente. Performance of a thermally deformable mirror for correction of low-order aberrations in laser beams. *Appl. Opt.*, 52(12):2909–2916, Apr 2013.

[21] B. Canuel, R. Day, E. Genin, P. La Penna, and J. Marque. Wavefront aberration compensation with a thermally deformable mirror. *Classical and Quantum Gravity*, 29(8):085012, 2012.

[22] P. Hello and J.-Y. Vinet. Analytical models of transient thermoelastic deformations of mirrors heated by high power cw laser beams. *J. Phys. France*, 51(20):2243–2261, 1990.

[23] S. Piehler, C. Thiel, A. Voss, M. Abdou Ahmed, and T. Graf. Self-compensation of thermal lensing in optics for high-brightness solid-state lasers. *Proc. SPIE*, (8239):82390Z–82390Z–9, 2012.

[24] H. Haferkamp and D. Seebaum. Beam delivery by adaptive optics for material processing applications using high-power co2 lasers. *Proc. SPIE*, 2207:156–164, 1994.

[25] T. Stephens, D.C. Johnson, and M.T. Languirand. Beam Path Conditioning for High-Power Laser Systems. *The Lincoln Laboratory Journal*, 3(2):225–244, 1990.

[26] S.K. Ravensbergen, P.C.J.N. Rosielle, and M. Steinbuch. Deformable mirrors with thermo-mechanical actuators for extreme ultraviolet lithography: Design, realization and validation. *Precision Engineering*, 37:353–363, 2013.

[27] A. Polo, V. Kutchoukov, F. Bociort, S. F. Pereira, and H. P. Urbach. Determination of wavefront structure for a Hartmann wavefront sensor using a phase-retrieval method. *Optics express*, 20(7):7822–32, 2012.

[28] K. J. G. Hinnen. *Data-driven optimal control for adaptive optics*. PhD thesis, Delft University of Technology, Delft, The Netherland, 2007.

[29] H. Song, R. Fraanje, G. Schitter, Kroese, G. Vdovin, and M. Verhaegen. Model-based aberration correction in a closed-loop wavefront-sensor-less adaptive optics system. *Opt. Express*, 18(23):24070–24084, Nov 2010.

[30] R. Fraanje, P. Massioni, and M. Verhaegen. A decomposition approach to distributed control of dynamic deformable mirrors. *International Journal of Optomechatronics*, 4(3):269–284, 2010.

[31] P. Massioni, C. Kulcsár, H.-F. Raynaud, and J.-M. Conan. Fast computation of an optimal controller for large-scale adaptive optics. *JOSA A*, 28(11):2298–2309, 2011.

[32] C. Max. Introduction to adaptive optics and its history. In *American Astronomical Society 197th meeting*, 2001.

[33] F. Roddier. *Adaptive optics in astronomy*. Cambridge University Press, 1999.

[34] R.K. Tyson. *Principles of Adaptive Optics, Third Edition*. CRC PressINC, 2010.

[35] M. J. Booth. Adaptive optics in microscopy. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1861):2829–2843, 2007.

[36] J. Antonello, M. Verhaegen, R. Fraanje, T. van Werkhoven, H. C. Gerritsen, and C. U. Keller. Semidefinite programming for model-based sensorless adaptive optics. *J. Opt. Soc. Am. A*, 29(11):2428–2438, Nov 2012.

[37] A. Roorda, F. Romero-Borja, W. Donnelly III, H. Queener, T. Hebert, and M. Campbell. Adaptive optics scanning laser ophthalmoscopy. *Opt. Express*, 10(9):405–412, May 2002.

[38] R. Zawadzki, S. Jones, S. Olivier, M. Zhao, B. Bower, J. Izatt, S. Choi, S. Laut, and J. Werner. Adaptive-optics optical coherence tomography for high-resolution and high-speed 3d retinal in vivo imaging. *Opt. Express*, 13(21):8532–8546, Oct 2005.

[39] S.-W. Bahk, E. Fess, B. E. Kruschwitz, and J. D. Zuegel. A high-resolution, adaptive beam-shaping system for high-power lasers. *Opt. Express*, 18(9):9151–9163, Apr 2010.

[40] T. Weyrauch, M. A. Vorontsov, J. Gowens, and T. G. Bifano. Fiber coupling with adaptive optics for free-space optical communication. In *Proc. SPIE 4489, Free-Space Laser Communication and Laser Imaging*, volume 177, pages 177–184, 2002.

[41] R. Q. Fugate. Prospects for benefits to astronomical adaptive optics from u.s. military programs. In *Proc. SPIE*, volume 4007, pages 1012–1019, 2000.

[42] H. S. Rana. Towards generic military imaging adaptive optics. In *Proc. SPIE*, volume 7119, pages 711904–711904–12, 2008.

[43] J. Tesch, S. Gibson, and M. Verhaegen. Receding-horizon adaptive control of aero-optical wavefronts. *Optical Engineering*, 52(7):071406–071406, 2013.

[44] H. Song. *Model-based Control in Adaptive Optical Systems*. PhD thesis, Delft University of Technology, Delft, The Netherland, 20011.

[45] R. A. Zacharias, N. R. Beer, E. S. Bliss, Burkhart S.C., S. J. Cohen, S. B. Sutton, R. L. Van Atta, S. E. Winters, Salmon J. T., Stolz C. J., D. C. Pigg, and T. J. Arnold. National Ignition Facility alignment and wavefront control. *Proc. SPIE*, (5341), may 2004.

[46] L. Van Atta, M. Perez, R. Zacharias, and W. Rivera. The wavefront control system for the national ignition facility. *arXiv preprint physics/0111104*, 2001.

[47] J.C. Dainty. *Adaptive Optics for Industry and Medicine: Proceedings of the Sixth International Workshop, National University of Ireland, Ireland, 12-15 June 2007*. Imperial College Press, 2008.

[48] U. Wittrock. *Adaptive Optics for Industry and Medicine: Proceedings of the 4th International Workshop, Münster, Germany, Oct. 19-24, 2003*. Springer Proceedings in Physics. Springer, 2010.

[49] G. Chériaux, C.J. Hooker, M. Stupka, and Society of Photo-optical Instrumentation Engineers. *Adaptive optics for laser systems and other applications: 18-19 April 2007, Prague, Czech Republic*. Proceedings of SPIE–the International Society for Optical Engineering. SPIE, 2007.

[50] G.V. Vdovin. *Adaptive mirror micromachined in silicon*. Published and distributed by Delft University Press, 1996.

[51] M. A. van de Kerkhof, W. de Boeij, H. Kok, M. Silova, J. Baselmans, and M. Hemerik. Full optical column characterization of DUV lithographic projection tools. *Proc. SPIE*, 5377:1960–1970, 2004.

[52] A. Polo, N. van Marrewijk, S. F. Pereira, and H. P. Urbach. Subaperture phase reconstruction from a Hartmann wavefront sensor by phase retrieval method for application in EUV adaptive optics. *Proc. SPIE*, 8322:832219, 2012.

[53] P. Mercère, P. Zeitoun, M. Idir, S. Le Pape, D. Douillet, X. Levecq, G. Dovillaire, S. Bucourt, K. A. Goldberg, P. P. Naulleau, and S. Rekawa. Hartmann wave-front measurement at 13.4 nm with $\lambda_{EUV}/120$ accuracy. *Opt. Lett.*, 28(17):1534–1536, Sep 2003.

[54] J. Schwider and G. Leuchs. Multi-pass Shack-Hartmann planeness test: monitoring thermal stress. *Opt. Express*, 18(8):8094–8106, Apr 2010.

[55] B. Schäfer, M. Schöneck, A. Bayer, and K. Mann. Absolute measurement of surface and bulk absorption in DUV optics from temperature induced wavefront deformation. *Opt. Express*, 18(21):21534–21539, Oct 2010.

[56] Y. Zhu, Sugisaki. K., M. Okada, K. Otaki, Z. Liu, J. Kawakami, M. Ishii, J. Saito, K. Murakami, M. Hasegawa, C. Ouchi, S. Kato, T. Hasegawa, A. Suzuki, H. Yokota, and M. Niibe. Wavefront measurement interferometry at the operational wavelength of extreme-ultraviolet lithography. *Appl. Opt.*, 46(27):6783–6792, Sep 2007.

[57] W. Nowacki. *Dynamic Problems of Thermoelasticity*. Springer, 1975.

[58] M. Verhaegen and V. Verdult. Filtering and system identification: A least squares approach. *Cambridge University Press*, 2007.

[59] P. Hello and J.-Y. Vinet. Analytical models of thermal aberrations in massive mirrors heated by high power laser beams. *J. Phys. France*, 51(12):1267–1282, 1990.

[60] P. Benner. Solving large-scale control problems. *Control Systems, IEEE*, 24(1):44 – 59, 2004.

[61] A.C. Antoulas. *Approximation of Large-scale Dynamical Systems*. Advances in design and control. Society for Industrial and Applied Mathematics, 2005.

[62] C. Bikcora, S. Weiland, and W. M. J. Coene. Reduced-order modeling of thermally induced deformations on reticles for extreme ultraviolet lithography. In *American Control Conference (ACC), 2013*, pages 5542–5549, 2013.

[63] J. K. Rice and M. Verhaegen. Structure preserving model order reduction of heterogeneous 1-D distributed systems. In *American Control Conference*, pages 4109–4114, 2009.

[64] S. Lall, P. Krysl, and J. E. Mardsen. Structure-preserving model reduction for mechanical systems. *Physica D: Nonlinear Phenomena*, 184(14):304 – 318, 2003.

[65] R. W. Freund. Sprim: structure-preserving reduced-order interconnect macromodeling. In *Proceedings of the 2004 IEEE/ACM International conference on Computer-aided design*, pages 80–87. IEEE Computer Society, 2004.

[66] H. Yu, L. He, and S.X.D. Tar. Block structure preserving model order reduction. In *Behavioral Modeling and Simulation Workshop, 2005. BMAS 2005. Proceedings of the 2005 IEEE International*, pages 1–6, 2005.

[67] R. W. Freund. Structure-preserving model order reduction of rcl circuit equations. In *Model Order Reduction: Theory, Research Aspects and Applications*, pages 49–73. Springer, 2008.

[68] H. Yu, Y. Shi, and L. He. Fast analysis of structured power grid by triangularization based structure preserving model order reduction. In *Proceedings of the 43rd annual Design Automation Conference*, pages 205–210. ACM, 2006.

[69] H. Sandberg and R. M. Murray. Model reduction of interconnected linear systems using structured gramians. In *Proceedings of the 17th IFAC World Congress*, pages 8725–8730, 2008.

[70] J. Anderson, Y.-C. Chang, and A. Papachristodoulou. Model decomposition and reduction tools for large-scale networks in systems biology. *Automatica*, 47(6):1165 – 1174, 2011. Special Issue on Systems Biology.

[71] H. Sandberg and R. M. Murray. Model reduction of interconnected linear systems. *Optimal Control Applications and Methods*, 30(3):225–245, 2009.

[72] P. Mathai and B. Shapiro. Interconnection of subsystem reduced-order models in the electrothermal analysis of large systems. *Components and Packaging Technologies, IEEE Transactions on*, 30(2):317–329, 2007.

[73] T. Reis and T. Stykel. Balanced truncation model reduction of second-order systems. *Mathematical and Computer Modelling of Dynamical Systems*, 14(5):391–406, 2008.

[74] B. Salimbahrami and B. Lohmann. Structure preserving order reduction of large scale second order systems. In *Proceedings of the 10th IFAC Symposium on Large Scale Systems: Theory and Applications (Osaka, Japan*, pages 245–250, 2004.

[75] R.-C. Li and Z. Bai. Structure-preserving model reduction using a krylov subspace projection formulation. *Communications in Mathematical Sciences*, 3(2):179–199, 2005.

[76] D. Siljak. Decentralized control of complex systems. *Academic Press, Boston, MA*, 1991.

[77] B. Bamieh, O. Paganini, and M. A. Dahleh. Distributed control of spatially invariant systems. *IEEE Transactions on Automatic Control*, 47:1091–1118, 2002.

[78] R. D'Andrea and G. Dullerud. Distributed control design for spatially interconnected systems. *IEEE Transactions on Automatic Control*, 48:1478–1495, 2003.

[79] M. R. Jovanović. *Modeling, analysis, and control of spatially distributed systems*. PhD thesis, University of California, Santa Barbara, 2004.

[80] J. Rice and M. Verhaegen. Distributed control: A sequentially semi-separable approach for heterogeneous linear systems. *IEEE Transactions on Automatic Control*, 54(6):1270–1283, 2009.

[81] F. Lin, M. Fardad, and M. R. Jovanović. Optimal control of vehicular formations with nearest neighbor interactions. *IEEE Trans. Automat. Control*, 57(9):2203–2218, September 2012.

[82] M. R. Jovanović and B. Bamieh. On the ill-posedness of certain vehicular platoon control problems. *IEEE Trans. Automat. Control*, 50(9):1307–1321, September 2005.

[83] P. Massioni and M. Verhaegen. Distributed control for identical dynamically coupled systems: A decomposition approach. *IEEE Transactions on Automatic Control*, 54(1):124–135, 2009.

[84] N. Motee and A. Jadbabaie. Optimal control of spatially distributed systems. *IEEE Transactions on Automatic Control*, 53 (7):1616–1629, 2008.

[85] J. K. Rice. *Efficient Algorithms for Distributed Control: A Structured Matrix Approach*. PhD thesis, Delft University of Technology, 2010.

[86] G. E Stewart, D. M. Gorinevsky, and G. A. Dumont. Feedback controller design for a spatially distributed system: The paper machine problem. *Control Systems Technology, IEEE Transactions on*, 11(5):612–628, 2003.

[87] D. Gorinevsky, S. Boyd, and G. Stein. Design of low-bandwidth spatially distributed feedback. *Automatic Control, IEEE Transactions on*, 53(1):257–272, 2008.

[88] J. A. Jacquez and C. P. Simon. Qualitative theory of compartmental systems. *SIAM Review*, 35:43–79, 1993.

[89] H.M. Arneson and C. Langbort. Distributed control design for a class of compartmental systems and application to eulerian models of air traffic flows. In *46th IEEE Conference on Decision and Control*, pages 2876–2881, 2007.

[90] N.K. Bose, B. Buchberger, and J.P. Guiver. *Multidimensional systems theory and applications*. Kluwer Academic Publishers, 2003.

[91] K. Galkowski. State-space Realizations of Linear 2-D Systems with Extensions to the General nD ($n > 2$) Case. *Springer-Verlag, London, Great Britain*, Number 263, 2001.

[92] T. Kaczorek. Two-dimensional linear systems. *Springer-Verlag, Berlin, Germany*, volume 68, 1985.

[93] S.Y. Kung, B.C. Levy, M. Morf, and T. Kailath. New results in 2-d systems theory, part ii: 2-d state-space models - realization and the notions of controllability, observability,and minimality. *Proceedings of the IEEE*, 65(6):945–961, 1977.

[94] K. G. Gadkar, J. Varner, and F.J. Doyle III. Model identification of signal transduction networks from data using a state regulator problem. *Systems Biology, IEE*, 2(1):17–30, 2005.

[95] A. Kremling and Saez J. Rodriguez. Systems biology–an engineering perspective. *J Biotechnol*, 129(2):329–351, 2007.

[96] P. Benner and H. Faßbender. On the numerical solution of large-scale sparse discrete-time riccati equations. *Adv. Comput. Math.*, 35(2-4):119–147, 2011.

[97] G.D. Smith. *Numerical solution of partial differential equations: finite difference methods*. Oxford University Press, USA, 1985.

[98] Y.-Y. Liu, J.-J. Slotine, and Albert-László Barabási. Controllability of complex networks. *Nature*, 473(7346):167–173, 2011.

[99] W.-X. Wang, X. Ni, Y.-C. Lai, and C. Grebogi. Optimizing controllability of complex networks by minimum structural perturbations. *Physical Review E*, 85(2):026115, 2012.

[100] C.W. Wu and L.O. Chua. Synchronization in an array of linearly coupled dynamical systems. *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*, 42(8):430–447, 1995.

[101] C.W. Wu. On control of networks of dynamical systems. In *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, pages 3785–3788, 2010.

[102] L. Xiang, W. Xiaofan, and C. Guanrong. Pinning a complex dynamical network to its equilibrium. *Circuits and Systems I: Regular Papers, IEEE Transactions on*, 51(10):2074–2087, 2004.

[103] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási. Observability of complex systems. *Proceedings of the National Academy of Sciences*, 110(7):2460–2465, 2013.

[104] B. Bamieh, M. R. Jovanović, P. Mitra, and S. Patterson. Coherence in large-scale networks: dimension dependent limitations of local feedback. *IEEE Trans. Automat. Control*, 57(9):2235–2249, September 2012.

[105] M.G. Larson and F. Bengzon. *The Finite Element Method: Theory, Implementation, and Applications*. Texts in Computational Science and Engineering. Springer-Verlag New York Incorporated, 2013.

[106] M. Verhaegen and P.M. Van Dooren. A reduced order observer for descriptor systems. *Systems & Control Letters*, 8(1):29 – 37, 1986.

[107] D. G. Luenberger. Time-invariant descriptor systems. *Automatica*, 14(5):473–480, 1978.

[108] K. Zhou, J. C. Doyle, and K. Glover. *Robust and optimal control*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1996.

[109] F.L. Lewis and V.L. Syrmos. *Optimal Control*. A Wiley-Interscience publication. Wiley, 1995.

[110] D. Simon. *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. Wiley, 2006.

[111] P.M. Van Dooren. *Numerical Linear Algebra for Signals Systems and Control*. Lecture notes, University of Louvain, Belgium, 2003.

[112] G. H. Golub and C. F. Van Loan. *Matrix computations (3rd ed.)*. Johns Hopkins University Press, 1996.

[113] J.K. Rice and M. Verhaegen. A Structured Matrix Approach to Efficient Calculation of LQG Repetitive Learning Controllers in the Lifted Setting. *Accepted for publication in International Journal of Control*, 2010.

[114] S. Chandrasekaran, P. Dewilde, M. Gu, T. Pals, and A.J. van der Veen. Fast stable solver for sequentially semi-separable linear systems of equations. *Lecture Notes in Computer Science*, pages 545–554, 2002.

[115] P. Dewilde and A.J. Van der Veen. *Time-varying systems and computations*. Kluwer Academic Pub, 1998.

[116] S. Chandrasekaran, P. Dewilde, M. Gu, T. Pals, A.J. van der Veen, and D. White. Fast Stable Solvers for Sequentially Semi-Separable Linear Systems of Equations. *Report, Lawrence Livermore National Library*, 2003.

[117] J. K. Rice and J.-W. van Wingerden. A computational method for boundary estimation and control via the sequentially semi separable approach. In *American Control Conference (ACC), 2013*, pages 2600–2605. IEEE, 2013.

[118] E. van Solingen, J.-W. van Wingerden, P. Torres, J. Rice, R. de Breuker, and M. Verhaegen. Parameter estimation for spatially interconnected descriptor systems using sequentially semi-separable matrices. In *American Control Conference (ACC), 2013*, pages 1657–1662. IEEE, 2013.

[119] J. K Rice and J.-W. van Wingerden. Computational methods for distributed control of heterogeneous cyclic interconnection structures. 2013.

[120] J. K. Rice and M. Verhaegen. Robust and distributed control of a smart blade. *Wind Energy*, 13(2-3):103–116, 2010.

[121] J. K. Rice and M. Verhaegen. Efficient system identification of heterogeneous distributed systems via a structure exploiting extended kalman filter. *Automatic Control, IEEE Transactions on*, 56(7):1713–1718, 2011.

[122] J. K. Rice and J.-W. van Wingerden. Fast calculation of the ilc norm in iterative learning control. *International Journal of Control*, pages 1–5, 2013.

[123] P. Benner and J. Saak. Numerical solution of large and sparse continuous time algebraic matrix riccati and lyapunov equa-tions: A state of the art survey. Technical report, 2013.

[124] R.F. Curtain and H. Zwart. *An Introduction to Infinite-Dimensional Linear Systems Theory*. Springer, 1995.

[125] S. Dubljevic and D. C. Panagiotis. Predictive control of parabolic {PDEs} with boundary control actuation. *Chemical Engineering Science*, 61(18):6239 – 6248, 2006.

[126] P. Massioni, T. Keviczky, E. Gill, and M. Verhaegen. A decomposition-based approach to linear time-periodic distributed control of satellite formations. *IEEE Trans. Contr. Sys. Techn.*, 19(3):481–492, 2011.

[127] P. Massioni and M. Verhaegen. Subspace identification of distributed, decomposable systems. *Joint 48th Conference on Decision and Control and 28th Chinese Control Conference, Shahghai, P.R. China*, 2009.

[128] P. Massioni. *Decomposition Methods for Distributed Control and Identification*. PhD thesis, Delft University of Technology, 2010.

[129] P. Massioni, C. Kulcsár, H.-F. Raynaud, and J.-M. Conan. Approximating the riccati equation solution for optimal estimation in large-scale adaptive optics systems. *2012 American Control Conference*, 27 - 29 June 2012.

[130] M. Farina, G. Ferrari-Trecate, and R. Scattolini. Distributed Moving Horizon Estimation for Linear Constrained Systems. *IEEE Transactions on Automatic Control*, 55(11):2462–2475, 2010.

[131] M. Farina, G. Ferrari-Trecate, and R. Scattolini. Moving-horizon partition-based state estimation of large-scale systems. *Automatica*, 46:910–918, 2010.

[132] T. Keviczky, F. Borrelli, and G. J. Balas. Decentralized receding horizon control for large scale dynamically decoupled systems. *Automatica*, 42(12):2105–2115, 2006.

[133] F. Lin, M. Fardad, and M. R. Jovanović. Sparse feedback synthesis via the alternating direction method of multipliers. In *Proceedings of the 2012 American Control Conference*, pages 4765–4770, Montréal, Canada, 2012.

[134] F. Lin, M. Fardad, and M. R. Jovanović. Design of optimal sparse feedback gains via the alternating direction method of multipliers. *IEEE Trans. Automat. Control*, 58(9):2426–2431, September 2013.

[135] F. Lin, M. Fardad, and M. R. Jovanović. Augmented Lagrangian approach to design of structured optimal state feedback gains. *IEEE Trans. Automat. Control*, 56(12):2923–2929, December 2011.

[136] S. Schuler, P. Li, J. Lam, and F. Allgöwer. Design of structured dynamic output-feedback controllers for interconnected systems. *International Journal of Control*, 84(12):2081–2091, 2011.

[137] A. Asif and J. M. F. Moura. Block matrices with l-block-banded inverse: inversion algorithms. *Signal Processing, IEEE Transactions on*, 53(2):630–642, 2005.

[138] U. A. Khan and J. M. F. Moura. Distributing the kalman filter for large-scale systems. *Signal Processing, IEEE Transactions on*, 56(10):4919–4935, 2008.

[139] U. A. Khan and J. M. F. Moura. Distributed iterate-collapse inversion (dici) algorithm for l-banded matrices. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pages 2529–2532. IEEE, 2008.

[140] D. P. Bertsekas and J. N. Tsitsiklis. *Parallel and distributed computation*. Prentice Hall, Englewood Cliffs, 1989.

[141] L. Ljung. System identification: theory for the user. *PTR Prentice Hall, Upper Saddle River, NJ*, 1999.

[142] F. Felici, J.-W. van Wingerden, and M. Verhaegen. Subspace identification of mimo lpv systems using a periodic scheduling sequence. *Automatica*, 43(10):1684–1697, 2007.

[143] J.-W. van Wingerden and M. Verhaegen. Subspace identification of bilinear and lpv systems for open-and closed-loop data. *Automatica*, 45(2):372–381, 2009.

[144] G.J. van der Veen, J.W. van Wingerden, P.A. Fleming, A.K. Scholbrock, and M. Verhaegen. Global data-driven modeling of wind turbines in the presence of turbulence. *Control Engineering Practice*, 21(4):441–454, 2013.

[145] C. T. Chou and Michel Verhaegen. Subspace algorithms for the identification of multivarible dynamic errors-in-variables models. *Automatica*, 33:1857–1869, 1997.

[146] M. Verhaegen and P. Dewilde. Subspace model identification part 1. the output-error state-space model identification class of algorithms. *International Journal of Control*, 56(5):1187–1210, 1992.

[147] C. V. Rao, J. B. Rawlings, and J. H. Lee. Constrained Linear Estimation- a Moving Horizon Approach. *Automatica*, 37:1619–1628, 2001.

[148] A. Alessandri, M. Baglietto, and G. Battistelli. Receding-Horizon Estimation for Discrete-Time Linear Systems. *IEEE Transactions on Automatic Control*, 48(3):473–477, 2003.

[149] A. Alessandri, M. Baglietto, and G. Battistelli. Min-max moving-horizon estimation for uncertain discrete-time linear systems. *SIAM Journal on Control and Optimization*, 50(3):1439–1465, 2012.

[150] A. Alessandri, M. Baglietto, G. Battistelli, and V. Zavala. Computationally efficient, approximate moving horizon state estimation for nonlinear systems. In *Nonlinear Control Systems*, pages 759–764, 2010.

[151] D.A. Bristow, M. Tharayil, and A.G. Alleyne. A survey of iterative learning control. *IEEE control systems magazine*, 26(3):96–114, 2006.

[152] B.G. Dijkstra and O.H. Bosgra. Extrapolation of optimal lifted system ILC solution, with application to a waferstage. *American Control Conference*, pages 2595–2600, 2002.

[153] M.Q. Phan, R.W. Longman, and K.L. Moore. Unified formulation of linear iterative learning control. *Spaceflight mechanics 2000*, pages 93–111, 2000.

[154] M. Norrlöf. *Iterative Learning Control: Analysis, Design and Experiments*. Thesis no. 653, Linköpings universitet, 2000.

[155] W.B.J. Hakvoort, R.G.K.M. Aarts, J. van Dijk, and J.B. Jonker. A computationally efficient algorithm of iterative learning control for discrete-time linear time-varying systems. *Automatica*, 45:2925–2929, 2009.

[156] H.S. Ahn, Y.Q. Chen, and K.L. Moore. Iterative learning control: Brief survey and categorization. *IEEE Transactions On Systems, Man, And Cybernetics, Part C: Applications and Reviews*, 37(6):1099, 2007.

[157] M. Volckaert, M. Diehl, and J. Swevers. Generalization of norm optimal ilc for nonlinear systems with constraints. *Mechanical Systems and Signal Processing*, 2013.

[158] D. Q. Mayne, M. M. Seron, and S.V. Raković. Robust model predictive control of constrained linear systems with bounded disturbances. *Automatica*, 41(2):219–224, 2005.

[159] S.V. Raković, E. C. Kerrigan, K. I. Kouramas, and D. Q. Mayne. Invariant approximations of the minimal robust positively invariant set. *Automatic Control, IEEE Transactions on*, 50(3):406–410, 2005.

[160] H. J. Ferreau, H. G. Bock, and M. Diehl. An online active set strategy to overcome the limitations of explicit mpc. *International Journal of Robust and Nonlinear Control*, 18(8):816–830, 2008.

[161] M. Benzi and N. Razouk. Decay bounds and O(n) algorithms for approximating functions of sparse matrices. *Electronic Transactions on Numerical Analysis*, pages 16–39, 2007.

[162] S. Demko, W. F. Moss, and P. W. Smith. Decay rates for inverses of band matrices. *Mathematics of Computation*, 43(168):491–499, 1984.

[163] V. Y. Pan, Y. Rami, and X. Wang. Structured matrices and newton's iteration: unified approach. *Linear Algebra and its Applications*, 343344:233 – 265, 2002.

[164] F.S. Cattivelli and A.H. Sayed. Diffusion strategies for distributed kalman filtering and smoothing. *Automatic Control, IEEE Transactions on*, 55(9):2069 –2084, 2010.

[165] S. Kar and J. F. M. Moura. Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs. *J. Sel. Topics Signal Processing*, 5(4):674–690, 2011.

[166] K. Srivastava and A. Nedic. Distributed asynchronous constrained stochastic optimization. *J. Sel. Topics Signal Processing*, 5(4):772–790, 2011.

[167] A. Nedic and A. E. Ozdaglar. Distributed subgradient methods for multi-agent optimization. *IEEE Trans. Automat. Contr.*, 54(1):48–61, 2009.

[168] J. Chen and A.H. Sayed. Diffusion adaptation strategies for distributed optimization and learning over networks. *IEEE Transactions on Signal Processing*, 60(8):4289–4305, 2012.

[169] G. Vdovin and M. Loktev. Deformable mirror with thermal actuators. *Opt. Lett.*, 27(9):677–679, May 2002.

[170] M. R. Jovanović and B. Bamieh. Modeling flow statistics using the linearized navier-stokes equations. In *Decision and Control, 2001. Proceedings of the 40th IEEE Conference on*, volume 5, pages 4944–4949. IEEE, 2001.

[171] M. R. Jovanovic and B. Bamieh. Componentwise energy amplification in channel flows. *Journal of Fluid Mechanics*, 534:145–184, 2005.

[172] M. R. Jovanović. Turbulence suppression in channel flows by small amplitude transverse wall oscillations. *Physics of Fluids*, 20:014101, 2008.

[173] M. R. Jovanović and S. Kumar. Transient growth without inertia. *Phys. Fluids*, 22(2):023101 (19 pages), February 2010.

[174] B. K. Lieu, R. Moarref, and M. R. Jovanović. Controlling the onset of turbulence by streamwise traveling waves. Part 2: Direct numerical simulations. *J. Fluid Mech.*, 2010.

[175] R. Moarref and M. R. Jovanović. Controlling the onset of turbulence by streamwise traveling waves. Part 1: Receptivity analysis. *J. Fluid Mech.*, 663:70–99, November 2010.

[176] R. Moarref and M. R. Jovanović. Model-based design of transverse wall oscillations for turbulent drag reduction. *J. Fluid Mech.*, 707:205–240, September 2012.

[177] M. Wang, A. Mani, and S. Gordeyev. Physics and computation of aero-optics. *Annual Review of Fluid Mechanics*, 44:299–321, 2012.

[178] Q. Gao, Z. Jiang, S. Yi, W. Xie, and T. Liao. Correcting the aero-optical aberration of the supersonic mixing layer with adaptive optics: concept validation. *Appl. Opt.*, 51(17):3922–3929, Jun 2012.

[179] T. M. Atanackovic and A. Guran. *Theory of elasticity for scientists and engineers*. Springer, 2000.

[180] S. Timoshenko and S. Woinowsky-Krieger. *Theory of plates and shells*. Engineering societies monographs. McGraw-Hill, 1959.

[181] U. A. Bakshi and A.V. Bakshi. *Electromagnetic Field Theory*. Technical Publications, 2009.

[182] O. C. Zienkiewicz, R. L. Taylor, and J. Z. Zhu. *The finite element method: its basis and fundamentals*, volume 1. Butterworth-Heinemann, 2005.

[183] E. Zuazua. Propagation, observation, and control of waves approximated by finite difference methods. *SIAM review*, 47(2):197–243, 2005.

[184] A. V. Wouwer and M. Zeitz. State estimation in distributed parameter systems. *Control Systems, Robotics and Automation, Theme in Encyclopedia of Life Support Systems*, 2003.

[185] T. Meurer. On the extended luenberger-type observer for semilinear distributed-parameter systems. *Automatic Control, IEEE Transactions on*, 58(7):1732–1743, 2013.

[186] L. Jadachowski, T. Meurer, and A. Kugi. An efficient implementation of backstepping observers for time-varying parabolic pdes. In *Mathematical Modelling*, volume 7, pages 798–803, 2012.

[187] R. Vazquez, E. Schuster, and M. Krstic. Magnetohydrodynamic state estimation with boundary sensors. *Automatica*, 44(10):2517 – 2527, 2008.

[188] R. Vazquez and M. Krstic. Boundary observer for output-feedback stabilization of thermal-fluid convection loop. *Control Systems Technology, IEEE Transactions on*, 18(4):789–797, 2010.

[189] M. Kristic and A. Smyshlyaev. *Boundary Control of PDEs: A Course on Backstepping Designs*. Society for Industrial and Applied Mathematics, 2008.

[190] L. Huang, Q. Xue, P. Yan, M.L. Gong, T.H. Li, Z.X. Feng, and X.K. Ma. A thermo-field bimetal deformable mirror for wavefront correction in high power lasers. *Laser Physics Letters*, 11(1):015001, 2014.

[191] A. Haber, A. Polo, S. K. Ravensbergen, H. P. Urbach, and M. Verhaegen. Identification of a dynamical model of a thermally actuated deformable mirror. *Opt. Lett.*, 38(16):3061–3064, Aug 2013.

[192] M. Hauf, N. Baer, H. Walter, and J. Hartjes. Arrangement for mirror temperature measurement and/or thermal actuation of a mirror in a microlithographic projection exposure apparatus, March 2013. US Patent App. 13/784,979.

[193] R. Saathof, A. Haber, J.W. Spronck, M. Verhaegen, and R. H. Munnig Schmidt. Closed loop thermal control of a lithographic optical component. In *Proceedings of the ASPE Summer Topical Meeting, June 24-26, 2012, Berkeley, CA, USA*, pages 1–6.

[194] R. B Hetnarski and M. R. Eslami. *Thermal stresses: advanced theory and applications*. Springer, 2009.

[195] A. Hassibi and S. Boyd. Quadratic stabilization and control of piecewise-linear systems. In *American Control Conference, 1998. Proceedings of the 1998*, volume 6, pages 3659–3664. IEEE, 1998.

[196] E. Sontag. Nonlinear regulation: The piecewise linear approach. *Automatic Control, IEEE Transactions on*, 26(2):346–358, 1981.

[197] D. A. Bristow. Weighting matrix design for robust monotonic convergence in norm optimal iterative learning control. In *Proceedings of the American Control Conference*, pages 4554–4560, 2008.

[198] K.L. Barton, D.A. Bristow, and A. G. Alleyne. A Numerical Method for Determining Monotonicity and Convergence Rate In Iterative Learning Control. *International Journal of Control*, 2010.

[199] D. A. Bristow, M. Tharayil, and A. G. Alleyne. A survey of iterative learning control. *Control Systems, IEEE*, 26(3):96–114, june 2006.

[200] Y. Saad. *Iterative Methods for Sparse Linear Systems*. Society for Industrial and Applied Mathematics, 2003.

[201] K.B. Lim and W. Gawronski. Hankel singular values of flexible structures in discrete time. *Journal of guidance, control, and dynamics*, 19(6):1370–1377, 1996.

[202] M. Benzi. Preconditioning techniques for large linear systems: A survey. *Journal of Computational Physics*, 182(2):418 – 477, 2002.

[203] M. J. Grote and T. Huckle. Parallel preconditioning with sparse approximate inverses. *SIAM Journal on Scientific Computing*, 18(3):838–853, 1997.

[204] M. Benzi and M. Tuma. A sparse approximate inverse preconditioner for nonsymmetric linear systems. *SIAM Journal on Scientific Computing*, 19(3):968–994, 1998.

[205] E. Chow and Y. Saad. Approximate inverse preconditioners via sparse-sparse iterations. *SIAM Journal on Scientific Computing*, 19(3):995–1023, 1998.

[206] M. Benzi and M. Tuma. A comparative study of sparse approximate inverse preconditioners. *Applied Numerical Mathematics*, 30(2):305–340, 1999.

[207] E. Chow. A priori sparsity patterns for parallel sparse approximate inverse preconditioners. *SIAM Journal on Scientific Computing*, 21(5):1804–1822, 2000.

[208] T. Huckle. Approximate sparsity patterns for the inverse of a matrix and preconditioning. *Applied numerical mathematics*, (30):291–303, 1999.

[209] W.-P. Tang. Toward an effective sparse approximate inverse preconditioner. *SIAM journal on matrix analysis and applications*, 20(4):970–986, 1999.

[210] J.R. Gilbert. Predicting structure in sparse matrix computations. *SIAM J. Matrix Anal. Appl.*, 15(1):62–79, 1994.

[211] D. Luenberger. An introduction to observers. *Automatic Control, IEEE Transactions on*, 16(6):596–602, 1971.

[212] R. B. Lehoucq, D. C. Sorensen, and C. Yang. *Arpack User's Guide: Solution of Large-Scale Eigenvalue Problems With Implicitly Restorted Arnoldi Methods* . Society for Industrial and Applied Mathematics.

[213] D. C. Sorensen. Implicit application of polynomial filters in a k-step arnoldi method. *SIAM J. Matrix Anal. Appl.*, 13(1):357–385, January 1992.

[214] J. C. Mason and D. C. Handscomb. *Chebyshev Polynomials*. Chapman & Hall/CRC, 2003.

[215] R. J. Mathar. Chebyshev series expansion of inverse polynomials. *J. Comput. Appl. Math.*, 196(2):596–607, November 2006.

[216] J.P. Boyd. *Chebyshev and Fourier Spectral Methods*. Dover Publications, 2001.

[217] W. Hackbusch, B. N. Khoromskij, and E. E. Tyrtyshnikov. Approximate iterations for structured matrices. *Numerische Mathematik*, 109(3):365–383, 2008.

[218] V. Olshevsky, I. Oseledets, and E. Tyrtyshnikov. Superfast inversion of two-level toeplitz matrices using newton iteration and tensor-displacement structure. In *Recent Advances in Matrix and Operator Theory*, pages 229–240. Springer, 2008.

[219] V. Olshevsky, I. Oseledets, and E. Tyrtyshnikov. Tensor properties of multilevel toeplitz and related matrices. *Linear algebra and its applications*, 412(1):1–21, 2006.

[220] V. Pan and R. Schreiber. An improved newton iteration for the generalized inverse of a matrix with applications. *SIAM Journal on Scientific and Statistical Computing*, 12(5):1109–1130, 1991.

[221] N. Razouk. *Localization phenomena in matrix functions: theory and algorithms*. PhD thesis, Emory University, 2008.

[222] D. Sui, T. A. Johansen, and L. Feng. Linear Moving Horizon Estimation With Pre-Estimating Observer. *IEEE Transactions on Automatic Control*, 55(10):2363–2368, 2010.

[223] K.V. Ling and K.W. Lim. Receding horizon recursive state estimation. *IEEE Trans. Automat. Contr.*, 44(9):1750 –1753, 1999.

[224] W. H. Kwon, P. S. Kim, and P. Park. A receding horizon kalman fir filter for linear continuous-time systems. *IEEE Trans. Automat. Contr*, 44:2115–2120, 1999.

[225] M. Rabbat and R. Nowak. Distributed optimization in sensor networks. pages 20–27, 2004.

[226] F. S. Cattivelli, C.G. Lopes, and A.H. Sayed. Diffusion recursive least-squares for distributed estimation over adaptive network. *IEEE Transactions on Signal Processing*, 56:1865–1877, 2008.

[227] M. Gander, S. Loisel, and D. Szyld. An optimal block iterative method and preconditioner for banded matrices with applications to pdes on irregular domains. *SIAM Journal on Matrix Analysis and Applications*, 33(2):653–680, 2012.

[228] E. M. Ortigosa, L. F. Romero, and J. I. Ramos. Parallel scheduling of the pcg method for banded matrices rising from fdm/fem. *J. Parallel Distrib. Comput.*, 63(12):1243–1256, 2003.

[229] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, 1990.

[230] *Theory of Operator Algebras I*. Encyclopedia of Mathematical Sciences: Operator Algebras and Non-Commutative Geometry. Springer, 2001.

[231] A. Alessandri, M. Baglietto, and G. Battistelli. On estimation error bounds for receding-horizon filters using quadratic boundedness. *Automatic Control, IEEE Transactions on*, 49(8):1350–1355, 2004.

[232] B. Boulkroune, M. Darouach, and M. Zasadzinski. Moving horizon state estimation for linear discrete-time singular systems. *IET control theory & applications*, 4(3):339–350, 2010.

[233] C. Bikcora. *Modeling and prediction of the thermally induced imaging errors in deep- and extreme-ultraviolet lithography*. PhD thesis, Eindhoven University of Technology, 2013.

[234] P.E. Moraal and J.W. Grizzle. Observer design for nonlinear systems with discrete-time measurements. *Automatic Control, IEEE Transactions on*, 40(3):395–404, 1995.

[235] P.E. Moraal, J.W. Grizzle, and J.A. Cook. An observer design for single-sensor individual cylinder pressure control. In *Decision and Control, 1993., Proceedings of the 32nd IEEE Conference on*, pages 2955–2961. IEEE, 1993.

[236] P. Massioni and M. Verhaegen. Distributed control for identical dynamically coupled systems: A decomposition approach. *IEEE Trans. Automat. Contr.*, 54(1):124–135, 2009.

[237] P. Van Overschee and B. De Moor. *Subspace Identification for Linear Systems; Theory, Implementation, Applications*. Kluwer Academic Publishers, 1996.

[238] M. Ali, S. S. Chughtai, and H. Werner. Identification of spatially interconnected systems. In *Joint 48th Conference on Decision and Control and 28th Chinese Control Conference, Shahghai, P.R. China*, pages 7163–7168, 2009.

[239] A. Sarwar and P. G. Voulgaris. Spatially invariant systems: identification and adaptation. In *Proceedings of the 3rd International Conference on Applied Mathematics, Simulation, Modelling, Circuits, Systems and Signals*, pages 208–215, 2009.

[240] A. Sarwar, P. G. Voulgaris, and S. M. Salapaka. System identification of spatiotemporally invariant systems. In *American Control Conference (ACC)*, pages 2947 – 2952, 2010.

[241] P. Massioni and M. Verhaegen. Subspace identification of circulant systems. *Automatica*, 44 (11):2825–2833, 2008.

[242] D. Materassi and G. Innocenti. Topological identification in networks of dynamical systems. *IEEE Transactions on Automatic Control*, 55(8):1860 – 1871, 2010.

[243] M. Timme. Revealing network connectivity from response dynamics. *Phys. Rev. Lett.*, 98(22):224101, 2007.

[244] D. Napoletani and T. D. Sauer. Reconstructing the topology of sparsely connected dynamical networks. *Phys. Rev. E*, 77, 2008.

[245] A. Chiuso, R. Muradore, and E. Marchetti. Dynamic calibration of adaptive optics systems: A system identification approach. *Control Systems Technology, IEEE Transactions on*, 18(3):705–713, 2010.

[246] A. Chiuso and G. Pillonetto. Learning sparse dynamic linear systems using stable spline kernels and exponential hyperpriors. In *Proc. of NIPS 2010, accepted*, December 2010.

[247] M. Grote and T. Huckle. Parallel preconditioning with sparse approximate inverses. *SIAM Journal on Scientific Computing*, 18(3):838–853, 1997.

[248] A. Haber. Simple subspace method for identification of lti systems. *Delft Center For Systems For Control, Delft University of Technology*, 2012. `http://www.dcsc.tudelft.nl/~ahaber/sim_sub.pdf`.

[249] M. Jansson. Subspace identification and arx modelling. *Proceedings of the 13th IFAC SYSID Symposium, Rotterdam, NL*, 2003.

[250] A. Chiuso. The role of vector auto regressive modeling in predictor based subspace identification. *Automatica*, 43 (6):1034–1048, 2007.

[251] G. Golub and V Pereyra. The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM Journal on Numerical Analysis*, 10(2):413–432, 1973.

[252] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 2nd edition, 1999.

[253] H. Lück, K.-O. Müller, P. Aufmuth, and K. Danzmann. Correction of wavefront distortions by means of thermally adaptive optics. *Opt. Commun.*, 175(46):275 – 287, 2000.

[254] P. C. Hansen. *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*, volume 4. SIAM, 1998.

[255] P. C. Hansen. *Discrete inverse problems: insight and algorithms*, volume 7. SIAM, 2010.

[256] L. Zhu, P.C. Sun, D.U. Bartsch, W.R. Freeman, and Y. Fainman. Adaptive control of a micromachined continuous-membrane deformable mirror for aberration compensation. *Applied Optics*, 38(1):168–176, 1999.

[257] S. Bonora and L. Poletto. Push-pull membrane mirrors for adaptive optics. *Opt. Express*, 14(25):11935–11944, 2006.

[258] E. Fernandez and P. Artal. Membrane deformable mirror for adaptive optics: performance limits in visual optics. *Opt. Express*, 11(9):1056–1069, May 2003.

[259] H. N. Chapman and D. W. Sweeney. Rigorous method for compensation selection and alignment of microlithographic optical systems. In *Proc. SPIE*, volume 3331, pages 102–113, 1998.

[260] A. Polo, A. Haber, S. F. Pereira, M. Verhaegen, and H. P. Urbach. An innovative and efficient method to control the shape of push-pull membrane deformable mirror. *Opt. Express*, 20(25):27922–27932, Dec 2012.

[261] Adaptica Srl. Saturn user manual. *http://www.adaptica.com/site/en/pages/saturn*.

[262] G. Vdovin and P. M. Sarro. Flexible mirror micromachined in silicon. *Appl. Opt.*, 34(16):2968–2972, jun 1995.

[263] D. Malacara. *Optical shop testing*. Wiley-Interscience, 2007.

[264] A. Haber, R. Fraanje, and M. Verhaegen. Linear computational complexity robust ilc for lifted systems. *Automatica*, 48(6):1102–1110, 2012.

[265] A. Haber, P.R. Fraanje, and M. Verhaegen. Fast and robust iterative learning control for lifted systems. In *World Congress*, volume 18, pages 3617–3622, 2011.

[266] O. Albert, L. Sherman, G. Mourou, T. B. Norris, and G. Vdovin. Smart microscope: an adaptive optics learning system for aberration correction in multiphoton confocal microscopy. *Opt. Lett.*, 25(1):52–54, Jan 2000.

[267] C. Reinlein, Appelfelder. M., M. Goy, K. Ludewigt, and A. Tünnermann. Performance of a thermal-piezoelectric deformable mirror under 6.2kW continuous-wave operation. *Appl. Opt.*, 52(34):8363–8368, Dec 2013.

[268] S. Piehler, Weichelt. B., A. Voss, M. Abdou Ahmed, and T. Graf. Power scaling of fundamental-mode thin-disk lasers using intracavity deformable mirrors. *Opt. Lett.*, 37(24):5033–5035, Dec 2012.

[269] R. Lawrence, D. Ottaway, M. Zucker, and P. Fritschel. Active correction of thermal lensing through external radiative thermal actuation. *Opt. Lett.*, 29(22):2635–2637, Nov 2004.

[270] S. E. Winters, J. H. Chung, and S. A. Velinsky. Dynamic modeling and control of a deformable mirror. *Mechanics Based Design of Structures and Machines*, 32(2):195–213, 2004.

[271] K.J. Astrom and R.M. Murray. *Feedback Systems: An Introduction for Scientists and Engineers*. Princeton University Press, 2009.

[272] H. Song, R. Fraanje, G. Schitter, G. Vdovin, and M. Verhaegen. Controller design for a high-sampling-rate closed-loop adaptive optics system with piezo-driven deformable mirror. *European Journal of Control*, 17(3):290 – 301, 2011.

[273] A. Haber, A. Polo, C. S. Smith, S. F. Pereira, P. Urbach, and M. Verhaegen. Iterative learning control of a membrane deformable mirror for optimal wavefront correction. *Appl. Opt.*, 52(11):2363–2373, Apr 2013.

[274] M. Laslandes, E. Hugot, and M. Ferrari. Active optics: deformable mirrors with a minimum number of actuators. *Journal of the European Optical Society - Rapid publications*, 7(0), 2012.

[275] S. J. Sheldon, L. V. Knight, and J. M. Thorne. Laser-induced thermal lens effect: a new theoretical model. *Appl. Opt.*, 21(9):1663–1669, May 1982.

[276] M. A. Arain, W. Z. Korth, L. F. Williams, R. M. Martin, G. Mueller, D. B. Tanner, and D.H. Reitze. Adaptive control of modal properties of optical beams using photothermal effects. *Opt. Express*, 18(3):2767–2781, Feb 2010.

[277] R. Saathof. *Adaptive Optics to Counteract Thermal Aberrations-System Design for EUV lithography with sub-nm Precision*. PhD thesis, Delft University of Technology, 2013.

[278] M.Q. Phan, R.K. Lim, and R.W. Longman. Unifying Input-Output and State-Space Perspectives of Predictive Control. *Princeton University, Department of Mechanical and Aerospace Engineering Technical Report No. 3044*, 1998.

[279] A. Haber and M. Verhaegen. Moving horizon estimation for large-scale interconnected systems. *IEEE Transactions on Automatic Control*, 58(11), 2013.

[280] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

List of symbols

col$(\cdot)$	column vector
∞	infinity
diag	(block) diagonal matrix
λ_{min}	minimal eigenvalue
λ_{max}	maximal eigenvalue
σ_{min}	minimal singular value
σ_{max}	maximal singular value
$\mathbb{C}$	field of complex numbers
$\mathbb{R}$	field of real numbers
$\mathbb{N}$	field of natural numbers
$\dagger$	matrix pseudo-inverse

List of Abbreviations

AO	Adaptive Optics
BS	Beam Splitter
CG	Conjugate Gradient
CTE	Coefficient of Thermal Expansion
DM	Deformable Mirror
TIWA	Thermally Induced Wavefront Aberrations
TADM	Thermally Actuated Deformable Mirror
ILC	Iterative Learning Control
ICs	Integrated Circuits
LMI	Linear Matrix Inequality
LPV	Linear Parameter Varying
LQR	Linear Quadratic Regulator
LTI	Linear Time-Invariant
LTP	Linear Time-Periodic
LTV	Linear Time-Varying
MEMS	Micro Electro-Mechanical Systems
MIMO	Multiple-Input Multiple-Output
MDM	Membrane Deformable Mirror
MPC	Model Predictive Control
MOESP	Multivariable Output-Error State sPace Identification
PEM	Prediction Error Method
SISO	Single-Input Single-Output
SC	Science Camera
SVD	Singular Value Decomposition

S-H WFS	Shack-Hartmann WaveFrontSensor
SIM	Subspace Identification Method
SC	Science Camera
VAF	Variance Accounted For
FD	Finite Difference
FE	Finite Element

Summary

Estimation and control of large-scale systems with an application to adaptive optics for EUV lithography

Aleksandar Haber

To predictively control Thermally Induced Wavefront Aberrations (TIWA) in Extreme UltraViolet (EUV) lithographic machines, accurate models of aberrations need to be developed. However, the predictive control of TIWA is a challenging problem mainly because the dynamical behavior of wavefront aberrations is described by thermoelastic partial differential equations. By discretizing the thermoelastic equations using the finite difference or finite element methods, large-scale state-space models can be obtained. These state-space models can be seen as large-scale networks of local subsystems. Consequently, the problem of correcting wavefront aberrations in the EUV lithography can be placed in a much more general context of identifying, estimating and controlling large-scale interconnected (distributed) systems. However, currently used estimation and control algorithms are not computationally feasible for large-scale systems. For this reason, this thesis focuses on the development of computationally efficient identification, estimation and control algorithms for large-scale interconnected systems.

In this thesis we prove that the inverses of Gramians of large-scale systems described by sparse (multi) banded state-space matrices, belong to a class of off-diagonally decaying matrices. Consequently, these inverses can be approximated by sparse banded matrices with $O(N)$ complexity, where N is the number of local subsystems. To compute the approximate inverses, the Chebyshev approximation method and the Newton iteration are used. The accuracy of the Chebyshev approximation method is analyzed and a new upper bound on the approximation errors is derived. Several methods for decreasing the computational and memory complexity of the approximation methods are presented.

Using these theoretical results, we develop novel estimation and identification algorithms for large-scale interconnected systems. The computational and memory complexity of these methods scale with $O(N)$. This is the main advantage over the traditional estimation and identification methods, which complexity is $O(N^3)$.

Furthermore, in this thesis we present novel, Adaptive Optics (AO) algorithms for compensating wavefront aberrations. First of all, we develop an experimental AO

system that consist of two Deformable Mirrors (DMs). One mirror is used to introduce wavefront aberrations, while another one is used as a correction element. This way, we simulate the AO system in a real EUV lithographic machine. Using this experimental setup, we demonstrate the proof of concept for predictive control of TIWA. Next, we develop a new algorithm for identifying dynamical models of thermally actuated DMs. Finally, we develop and experimentally test an Iterative Learning Control (ILC) algorithm for high performance correction of wavefront aberrations. The ILC algorithm can be used in a real EUV lithographic machine for precise alignment of optical elements and for suppression of static wavefront aberrations.

Samenvatting

Het schatten en reguleren van grootschalige systemen, met een toepassing in EUV lithografie

Aleksandar Haber

Om warmte-ge induceerde golffront aberraties (WGGA) in extreem ultraviolet (EUV) op een voorspellende wijze te kunnen reguleren zijn accurate modellen voor deze abberaties nodig. Echter, het voorspellend reguleren van WGGA is een uitdagend probleem, met name doordat de dynamica van de aberraties beschreven worden door zogenaamde thermo-elastische parti ele differentiaal vergelijkingen. Desalniettemin kunnen grootschalige $state\text{-}space$ modellen geconstrueerd worden d.m.v. het discretizeren van de thermo-elastische vergelijkingen met behulp van eindige-elementen- en eindige-differentiemethodes. De resulterende state-space modellen kunnen gezien worden als grootschalige netwerken van lokale subsystemen. Een gevolg hiervan is dat het probleem van het corrigeren voor golffront aberraties in EUV lithografie in een ruimere context geplaatst kan worden, nl. die van het identificeren, schatten, en reguleren van grootschalige, aaneengesloten (gedistribueerde) systemen. Huidige schattings- en regulatie algoritmes zijn vanuit een rekentechnisch oogpunt te langzaam om toegepast te worden op grootschalige systemen. In dit proefschrift wordt derhalve nieuwe, effici entere algoritmes ge introduceerd voor de grootschalige, aaneengesloten systemen.

In dit proefschrift wordt vervolgens bewezen dat de inversen van de Gramianen van grootschalige systemen beschreven door schaars-gestrookte state-space matrices, onderdeel uitmaken van een klasse van matrices waarvan de niet-diagonale elementen vervallen met orde N. Hierdoor kunnen de inversen benaderd worden door schaars-gestrookte matrices van $O(N)$ complexiteit, waarbij N het aantal lokale subsystemen. Deze benaderde matrices worden uitgerekend met de Chebyshev benadering en de methode van Newton. De nauwkeurigheid van de Chebyshev benadering is geanalyseerd, waarbij een nieuwe bovengrens voor de benaderingsfout is afgeleid. Daarnaast zijn er verschillende methodes ontwikkeld om de benodigde rekentijd en het geheugengebruik in de computer te reduceren.

Met de bovengenoemde theoretische resultaten worden in dit proefschrift nieuwe schattings- en identificatie algoritmes voor grootschalige aaneengesloten systemen afgeleid. De complexiteit voor de benodige rekentijd en het geheugenge-

bruik schaalt met $O(N)$, wat in schril contrast staat met traditionele schattings- en identificatiemethodes, welke schalen met $O(N^3)$.

Daarnaast worden er nieuwe, adaptieve optica (AO) algoritmes beschreven met als doel om golffront aberraties te compenseren. Hiertoe hebben we allereerst een experimenteel AO systeem ontwikkeld dat bestaat uit twee vervormbare spiegels. Eén spiegel heeft als doel om golffront aberraties te cre eren, terwijl de andere gebruikt wordt als een correctie element. Op deze manier simuleren we een AO systeem in een echte EUV machine. Met deze experimentele opbouw tonen we het "proof-of-concept"voor voorspellende regulatie van WGGA. Verder is er een nieuw algoritme ontwikkeld voor het identificeren van dynamische systemen van warmte-aangedreven vervormbare spiegels. Tot slot hebben we een iteratief-lerend regulatie algoritme voor nauwkeurige en snelle correctie van golffront aberraties bedacht, en experimenteel getest. Dit algoritme kan gebruikt worden in echte EUV litografie machines voor het nauwkeurig uitlijnen van optische elementen en voor onderdrukking van statische golffront aberraties.

Curriculum Vitae

Aleksandar Haber was born on January 19, 1984, in Belgrade, Serbia. After finishing the high school, in 2003 he enrolled in the Faculty of Mechanical Engineering, University of Belgrade, Serbia, where he received his Dipl.-Ing. degree (5 years program equivalent to the MSc degree) in July of 2008. His average grade was 9.97, on a scale from 0-10 (where the passing grade was 6). He received the dean's award for the best graduate in Mechanical Engineering of the year. Since 2009 he has been a PhD candidate at the Delft Center for Systems and Control, Delft University of Technology, The Netherlands, under the supervision of Michel Verhaegen. The PhD research was supported by the Dutch Ministry of the Economic Affairs and the Provinces of Noord-Brabant and Limburg in the frame of the "Pieken in de Delta" program. The main focus of the PhD project was the development of estimation and control algorithms for large-scale interconnected systems with an application to adaptive optics for EUV lithography.

During the course of his PhD research, he closely collaborated with the scientists from the Optics Research Group, Delft University of Technology. Since July 2013, he is a post-doc researcher at the Delft Center for Systems and Control, Delft University of Technology.

List of journal publications

1. Moving horizon estimation for large-scale interconnected systems, **A. Haber** and M. Verhaegen, *IEEE Transactions on Automatic Control*, Vol. 58, Issue 11, 2013.

2. Predictive control of thermally induced wavefront aberrations, **A. Haber**, A. Polo, I. Maj, S.F. Pereira, H.P. Urbach and M. Verhaegen, *Optics Express*, Vol. 21, Issue 18, 2013.

3. Subspace identification of large-scale interconnected systems, **A. Haber** and M. Verhaegen, *IEEE Transactions on Automatic Control. Note: Conditionally accepted*, 2013 (arXiv:1309.5105v1).

4. Iterative learning control of a membrane deformable mirror for optimal wavefront correction, **A. Haber**, A. Polo, C.S. Smith, S.F. Pereira, H.P. Urbach and M. Verhaegen, *Applied Optics*, Vol. 52, Issue 11, 2013.

5. Linear computational complexity robust ILC for lifted systems, **A. Haber**, P.R. Fraanje and M. Verhaegen, *Automatica*, Vol. 48, Issue 6, 2012.

6. Identification of a dynamical model of a thermally actuated deformable mirror, **A. Haber**, A. Polo, S.K. Ravensbergen, H.P. Urbach and M. Verhaegen, *Optics Letters*, Vol. 38, Issue 16, 2013.

7. Linear phase retrieval for real-time adaptive optics, A. Polo, **A. Haber**, S.F. Pereira, M. Verhaegen and H.P. Urbach, *Journal of the European Optical Society-Rapid publications*, Vol. 8, 2013.

8. An innovative and efficient method to control the shape of push-pull membrane deformable mirror, A. Polo, **A. Haber**, S.F. Pereira, M. Verhaegen and H.P. Urbach, *Optics Express*, Vol. 20, Issue 25, 2012.

9. Identification of large-scale interconnected systems described by sparse banded system matrices. **A. Haber** and M. Verhaegen, *submitted to Automatica*.