Zhizheng Wu · Azhar Iqbal
Foued Ben Amara

Modeling and Control of Magnetic Fluid Deformable Mirrors for Adaptive Optics Systems

Springer

Modeling and Control of Magnetic Fluid Deformable Mirrors for Adaptive Optics Systems

Zhizheng Wu • Azhar Iqbal • Foued Ben Amara

Modeling and Control of Magnetic Fluid Deformable Mirrors for Adaptive Optics Systems

Springer

Zhizheng Wu
Shanghai University
Shanghai
China, People's Republic

Azhar Iqbal
University of Toronto
Toronto Ontario
Canada

Foued Ben Amara
University of Toronto
Toronto Ontario
Canada

ISBN 978-3-642-32228-0 ISBN 978-3-642-32229-7 (eBook)
DOI 10.1007/978-3-642-32229-7
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012951037

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Adaptive optics (AO) systems are used in various application areas to enhance the performance of optical systems, such as imaging systems, laser systems, and free space optical communication systems. Recent developments in optomechatronic systems targeted at AO applications cover two main areas, namely, the development of novel wavefront correctors and the development of advanced control algorithms. Adaptive optics (AO) systems make use of active optical elements, namely, wavefront correctors, to improve the resolution of imaging systems by compensating for complex optical aberrations. Recently, magnetic fluid deformable mirrors (MFDM) were proposed as a novel type of wavefront correctors that offer cost and performance advantages over existing wavefront correctors. These mirrors are developed by coating the free surface of a magnetic fluid with a thin reflective film of nanoparticles. The reflective surface of the mirrors can be deformed using a locally applied magnetic field and thus serves as a wavefront corrector. MFDMs have been found particularly suitable for ophthalmic imaging systems, where they can be used to compensate for the complex aberrations in the eye that blur the images of the internal parts of the eye, or astronomical imaging systems where an MFDM can be easily deployed and maintained. However, the practical implementation of MFDMs in real applications is hampered by the lack of effective methods to control the shape of their deformable surface.

The recent developments reported in this book present solutions to the surface shape control problem in an MFDM that will make it possible for such devices to become integral components of adaptive optics (AO) systems. The contents of this book address the following main objectives:

- *Model development.* This first objective involves the development of a comprehensive model of the MFDM surface shape response to the applied control magnetic field. The model should accurately capture the dynamics of the surface shape and should be simple enough to allow for the design of effective surface shape control algorithms.
- *Design and development of an MFDM and AO system.* This objective concerns the design and development of a prototype MFDM to be used in the validation of

the mirror model and in the evaluation of the designed controllers. It also includes the design and assembly of an experimental AO setup to test the performance of the MFDM. The AO setup could be further extended to an ophthalmic imaging system to compensate for the aberrations in the human eye.

- *Controller design.* The third objective involves the design of effective MFDM surface shape controllers such that the mirror could be used in a closed-loop AO system to cancel the optical aberrations of interest. The design of the proposed controllers is based on considering various closed-loop AO system performance specifications. As the aberrations in the human eye and in other applications are dynamic in nature and are not known a priori, the control problem is cast as a regulation problem where it is desired for the mirror surface shape to track a time-varying unknown reference shape.
- *Experimental validation and evaluation.* The final objective is to experimentally evaluate the effectiveness of the closed-loop AO system involving the prototype MFDM and the developed control algorithms. The closed-loop AO system performance evaluation is based on analyzing the mirror surface shape, tracking accuracy for given desired mirror surface shapes that are needed to cancel the optical aberrations.

Based on the aforementioned objectives, several research developments have been made with respect to the design, modeling, and control of magnetic fluid deformable mirrors. The first major contribution of this research is the development of an accurate analytical model of the dynamics of the mirror surface shape. The model is developed by analytically solving the coupled system of fluid-electromagnetic equations that govern the dynamics of the surface shape. The model is presented in state-space form and can be readily used in the development of surface shape control algorithms. The second major contribution of the research work is a novel and innovative design of the MFDM. The design change was prompted by the findings of the analytical work undertaken to develop the model mentioned above and is aimed at linearizing the response of the mirror surface. The proposed design also allows for mirror surface deflections that are much larger than those provided by the conventional MFDM designs. A third contribution of this book involves the development of control algorithms that allowed the first-ever use of an MFDM in a closed-loop adaptive optics system. A decentralized proportional-plus-integral (PI) control algorithm developed based on the DC model of the wavefront corrector is presented to deal mostly with static or slowly time-varying aberrations. To improve the stability robustness of the closed-loop AO system, a decoupled uncertain model of the plant is considered and a decentralized robust proportional-integral-derivative (PID) controller is developed. To improve the performance of the closed-loop AO system in dealing with complex dynamic aberrations, centralized control methods are developed and include a multivariable PID control approach where a fixed structure controller is designed based on solving linear matrix inequalities (LMIs) and an optimal control approach based on the mixed-sensitivity H_∞ design method. The proposed model prototype MFDM and control algorithms are experimentally tested and evaluated. The aforementioned developments are organized into nine chapters as summarized in the following.

Chapter 1 presents an introduction to the topics discussed in this book. The advantages of magnetic fluid deformable mirrors over the conventional solid mirrors are presented, and the major contributions of this book are summarized. Some preliminaries that help the readers better understand the contents in this book are also briefly introduced.

In Chap. 2, a brief review of the history on adaptive optics systems and their applications is presented. The basic concept of a wavefront and the wavefront-related optical aberrations is introduced. The basic setup of a typical adaptive optics system is illustrated, and a brief exposition of the ophthalmic technologies which have benefited from adaptive optics systems is introduced. The motivation for considering AO systems based on magnetic fluid deformable mirrors is discussed.

In Chap. 3, the brief history of the development of MFDMs is introduced. The working principle of MFDMs including the description of their main structure and composition is presented. The advantages of the MFDMs and significant challenges are discussed. An analysis of the problems impeding the potential utilization of these mirrors in practical applications is presented.

Chapter 4 introduces an analytical model of the MFDM surface shape dynamics. The model describes the dynamics of the mirror surface shape in terms of deflections of the mirror surface driven by an array of electromagnetic coils and is developed in both the Cartesian and the cylindrical coordinate systems. An innovative method of linearizing the equations governing the dynamics of the mirror surface is introduced. The linearization method is also utilized to propose an important modification in the MFDM design.

The design of the prototype MFDM and of an experimental AO setup is presented in Chap. 5. The novel conceptual design is prompted by the findings of the analytical work presented in Chap. 4. The details of the layout of the AO setup for experimental validation and the corresponding components used in the setup are described. The prototype mirror and the experimental setup are used to validate the developed analytical model of the mirror, including the static and dynamic response of the prototype MFDM, as well as validate the obtained new features of the MFDM compared with the conventional MFDMs, such as the linearization of the response, the bidirectional displacements of the mirror surface, and the amplified magnitude of the surface deflections.

In Chap. 6, a concise review of the different control algorithms that have been used for the shape control of deformable mirrors in AO systems is presented. The control problem for magnetic fluid deformable mirrors is then discussed, and the appropriate configuration of the proposed MFDM control system is presented.

In Chap. 7, two decentralized PID control algorithms are proposed to control the surface shape of an MFDM in the closed-loop adaptive optics system. First, a classical decentralized proportional-plus-integral (PI) controller is designed based on statically decoupling the plant model using the plant DC gain. To overcome stability robustness issues and minimize the effect of the model uncertainties, especially in the high frequency range, a decentralized robust proportional-integral-derivative (PID) controller is then proposed and verified experimentally.

To improve the control performance with respect to dynamic aberration signals, two centralized optimal control algorithms are proposed in Chap. 8 to address the MFDM surface shape control problem. The first control algorithm is associated with a multivariable PID controller designed based on properly formulated linear matrix inequalities. Controller designs that yield closed-loop systems with H_∞ and H_2 performance specifications are investigated. A robust mixed-sensitivity H_∞ controller design approach is also developed to provide the desired tracking performance, limit the input currents, and enhance the robustness of the closed-loop system. The performances of both control algorithms are experimentally evaluated using the closed-loop adaptive optics system, showing successful tracking of dynamic wavefront shapes that have a frequency content consistent with that of the human eye aberrations.

In Chap. 9, the concluding remarks and recommendations for future work are given. Important aspects of MFDMs that have to be addressed before the latter can be finally used practically in clinical or astronomical applications are discussed.

This book is primarily intended for researchers and engineers in the adaptive optics systems community. The contents of this book are presented in a self-contained manner. The nine chapters contained in this book exploit several independent yet related topics in detail. Instead of only giving the controller design procedure, some concepts behind the control approaches and corresponding details of the derivations are also presented selectively in this book to outline the design insights for the reader.

We are very fortunate to have had many helpful suggestions from our colleagues, friends, and coworkers through many stimulating discussions in relation to the developments presented in this book. First of all, we would like to express our sincere gratitude to numerous valuable suggestions and comments from Professor Ermanno Borra at Laval University. We would also like to express our sincere appreciation to a number of graduate and undergraduate students who have contributed to the collaborative studies of adaptive optics systems. They are Maurizio Ficocelli, Jason Li, Devina Dukhu, Geoff Fung, Steve Bristo, and Ryan Li. Special thanks go to Becky Zhao and Li Shen, Associate Editors of Springer-Verlag, for their consistent and reliable support and comments in the process of publishing this book. The research work presented in this book has been supported in part by the Natural Sciences and Engineering Research Council of Canada, the Canada Foundation for Innovation, the Provincial Government of Ontario, the National Natural Science Foundation of China, and Shanghai Pujiang Program.

We are greatly indebted to our families for their encouragement and sacrifices over the years it took to complete the research work presented in this book. It is our pleasure to dedicate this book to them for their invaluable support.

Zhizheng Wu
Azhar Iqbal
Foued Ben Amara

Contents

List of Abbreviations and Symbols

Throughout this book, vector quantities are referred to using boldface letters, whereas the corresponding magnitudes are referred to using italicized letters. In a given matrix, entries denoted by * are such that the overall matrix is symmetric. The following conventions and notations are adopted:

Abbreviations

AO	Adaptive optic(s)
ARE	Algebraic Riccati equation
BIBO	Bounded-input bounded-output
DM	Deformable mirror
DC	Direct current
FWHM	Full-width-at-half maximum
HEL	High-energy lasers
LFT	Linear fractional transformation
LMI	Linear matrix inequality
LMT	Liquid mirror telescope
LQG	Linear quadratic Gaussian
LQR	Linear quadratic regulator
LTI	Linear time invariant
MIMO	Multiple-input multiple-output
MFDM	Magnetic fluid deformable mirror
MELLF	Metal liquid-like film
MTF	Modulation transfer function
NA	Numerical aperture
OCT	Optical coherence tomography
OPL	Optical path length
OTF	Optical transfer function

PSF	Point spread function
PI	Proportional-plus-integral
PID	Proportional-integral-derivative
RMS	Root mean square
RS	Robust stability
WFC	Wavefront corrector
WFS	Wavefront sensor
SISO	Single-input single-output
SLD	Super-luminescent diode
SLO	Scanning laser ophthalmoscopy
SVD	Singular value decomposition

List of Symbols

$\mathbf{A}^{\mathrm{T}}$	Transpose of $\mathbf{A}$
$\mathbf{A}^{*}$	Complex conjugate transpose of $\mathbf{A}$
$\mathbf{A}^{-1}$	Inverse of $\mathbf{A}$
$\mathbf{A}^{\dagger}$	Pseudo-inverse of $\mathbf{A}$
Trace $(\mathbf{A})$	Trace of $\mathbf{A}$
$\lambda\,(\mathbf{A})$	Eigenvalue of $\mathbf{A}$
$\rho\,(\mathbf{A})$	Spectral radius of $\mathbf{A}$
$\bar{\sigma}\,(\mathbf{A})$	Largest singular value of $\mathbf{A}$
$\mathbb{C}$	Field of complex numbers
$\mathbf{I}_n$	$n \times n$ identity matrix
$\mathbb{R}$	Field of real numbers
$\mathbf{B}$	Magnetic flux density, $\mathbf{B} = B_r\hat{\mathbf{r}} + B_\theta\hat{\theta} + B_z\hat{\mathbf{z}} = B_x\hat{\mathbf{i}} + B_y\hat{\mathbf{j}} + B_z\hat{\mathbf{k}}$, $B = \lvert\mathbf{B}\rvert$ (T)
$\mathbf{H}$	Magnetic field, $\mathbf{H} = H_r\hat{\mathbf{r}} + H_\theta\hat{\theta} + H_z\hat{\mathbf{z}} = H_x\hat{\mathbf{i}} + H_y\hat{\mathbf{j}} + H_z\hat{\mathbf{k}}$, $H = \lvert\mathbf{H}\rvert$ (A/m)
$\mathbf{M}$	Magnetization, $\mathbf{M} = M_r\hat{\mathbf{r}} + M_\theta\hat{\theta} + M_z\hat{\mathbf{z}} = M_x\hat{\mathbf{i}} + M_y\hat{\mathbf{j}} + M_z\hat{\mathbf{k}}$, $M = \lvert\mathbf{M}\rvert$ (A/m)
$\mathbf{V}$	Fluid velocity, $\mathbf{V} = V_r\hat{\mathbf{r}} + V_\theta\hat{\theta} + V_z\hat{\mathbf{z}} = V_x\hat{\mathbf{i}} + V_y\hat{\mathbf{j}} + V_z\hat{\mathbf{k}}$, $V = \lvert\mathbf{V}\rvert$ (m/s)
$J_m\,(\cdot)$	Bessel function of the first kind of order m
$Y_m\,(\cdot)$	Bessel function of the second kind of order m
W	Wavefront function (m)
$\mathbf{b}$	Perturbed magnetic flux density, $\mathbf{b} = b_r\hat{\mathbf{r}} + b_\theta\hat{\theta} + b_z\hat{\mathbf{z}} = b_x\hat{\mathbf{i}} + b_y\hat{\mathbf{j}} + b_z\hat{\mathbf{k}}$, $b = \lvert\mathbf{b}\rvert$ (T)
$\mathbf{g}$	Gravitational acceleration $\left(\mathrm{m/s^2}\right)$
$\mathbf{h}$	Perturbed magnetic field, $\mathbf{h} = h_r\hat{\mathbf{r}} + h_\theta\hat{\theta} + h_z\hat{\mathbf{z}} = h_x\hat{\mathbf{i}} + h_y\hat{\mathbf{j}} + h_z\hat{\mathbf{k}}$, $h = \lvert\mathbf{h}\rvert$ (A/m)

$\mathbf{m}$	Perturbed magnetization, $\mathbf{m} = m_r\hat{\mathbf{r}} + m_\theta\hat{\theta} + m_z\hat{\mathbf{z}} = m_x\hat{\mathbf{i}} + m_y\hat{\mathbf{j}} + m_z\hat{\mathbf{k}}$, $m = \lvert\mathbf{m}\rvert$ (A/m)
$\hat{\mathbf{n}}$	Unit normal vector
k	Separation constant/wavenumber along $\hat{\mathbf{k}}$
k_x	Separation constant/wavenumber along $\hat{\mathbf{i}}$
k_y	Separation constant/wavenumber along $\hat{\mathbf{j}}$
m	Separation constant / wavenumber along $\hat{\theta}$
p	Thermodynamic pressure $\left(\mathrm{N/m^2}\right)$
p_m	Fluid-magnetic pressure, $\mu_0 \int_0^H M \,\mathrm{d}H$ $\left(\mathrm{N/m^2}\right)$
p_n	Magnetic normal pressure, $\frac{\mu_0}{2}(\mathbf{M}\cdot\hat{\mathbf{n}})^2$ $\left(\mathrm{N/m^2}\right)$
p_s	Magnetostrictive pressure, $\mu_0 \int_0^H \upsilon \left(\frac{\partial M}{\partial \upsilon}\right) \mathrm{d}H$ $\left(\mathrm{N/m^2}\right)$
r	Distance along $\hat{\mathbf{r}}$ from the origin of the cylindrical coordinate system
x	Distance along $\hat{\mathbf{i}}$ from the origin of the Cartesian coordinate system
y	Distance along $\hat{\mathbf{j}}$ from the origin of the Cartesian coordinate system
z	Distance along $\hat{\mathbf{k}}$ from the origin of the Cartesian coordinate system Distance along $\hat{\mathbf{z}}$ from the origin of the cylindrical coordinate system
∇	Gradient operator, $\nabla = \hat{\mathbf{i}}\frac{\partial}{\partial x} + \hat{\mathbf{j}}\frac{\partial}{\partial y} + \hat{\mathbf{k}}\frac{\partial}{\partial z} = \hat{\mathbf{r}}\frac{\partial}{\partial r} + \hat{\theta}\frac{1}{r}\frac{\partial}{\partial \theta} + \hat{\mathbf{z}}\frac{\partial}{\partial z}$
$\nabla\cdot$	Divergence operator, $\nabla\cdot\mathbf{v} = \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z} = \frac{1}{r}\frac{\partial}{\partial r}(r v_r) + \frac{1}{r}\frac{\partial v_\theta}{\partial \theta} + \frac{\partial v_z}{\partial z}$
∇^2	Laplacian operator, $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2} + \frac{\partial^2}{\partial z^2}$
Φ	Fluid velocity potential $\left(\mathrm{m^2/s}\right)$
$\Psi^{(i)}$	Magnetic potential (A)
χ	Magnetic susceptibility
ϕ	Perturbed fluid velocity potential $\left(\mathrm{m^2/s}\right)$
η	Dynamic viscosity $(\mathrm{kg/(m\cdot s)})$
κ	Surface curvature $\left(\mathrm{m^{-1}}\right)$
λ	Separation constant/wavenumber along $\hat{\mathbf{z}}\left(\mathrm{m^{-1}}\right)$
μ	Magnetic permeability of the magnetic fluid (H/m)
μ_0	Magnetic permeability of air (H/m)
θ	Angle from a reference line in a cylindrical coordination system (radians)
ρ	Density $\left(\mathrm{kg/m^3}\right)$
σ	Surface tension (N/m)
$\psi^{(i)}$	Perturbed magnetic potential (A)
φ	Phase of the electromagnetic wave (radians)
ζ	Fluid surface deflection (m)
$\in$	Belongs to
$:=$	Defined as
$\gg$	Much greater than
$\ll$	Much smaller than
$\left[\begin{array}{c\|c} A & B \\ \hline C & D \end{array}\right]$	Shorthand for state-space realization of $C(sI - A)^{-1}B + D$ or $C(zI - A)^{-1}B + D$

Chapter 1
Introduction and Preliminaries

Contents

1.1 Introduction

AO systems were initially developed for applications in astronomical imaging systems using ground-based telescopes for taking images of distant astral bodies (Babcock 1953). In these applications, AO systems are used to compensate for the aberrations caused by the atmospheric turbulence, which severely limits the resolution of the images provided by the telescopes. Today's astronomical AO systems typically work by deforming solid mirrors with hundreds of actuators to compensate for the turbulence in the atmosphere (Hart 2010). Adaptive optics systems work by sensing the wavefront incident on an imaging system and then correcting the aberrations using a deformable mirror to dynamically cancel the unknown aberrations, thus appreciably enhancing the image quality, sometimes approaching diffraction-limited performance. Adaptive optics has been synonymous with the correction of atmospheric aberrations, but many more applications have emerged in recent years and could benefit from the aberration corrections provided by adaptive optics systems. Examples of fairly novel applications include imaging of the retina, free-space optical communication, confocal microscopy, laser beam delivery, dispersion compensation in ultrafast lasers, and optical data storage. However, in order to fully commercialize these new applications, significant progress is needed especially in reducing the cost of adaptive optics components, such as deformable mirrors, driver electronics, or wavefront sensors, and in developing more accurate

Z. Wu et al., *Modeling and Control of Magnetic Fluid Deformable Mirrors for Adaptive Optics Systems*, DOI 10.1007/978-3-642-32229-7_1,

and effective control algorithms for these complex systems. Addressing these and other issues is expected to lead to much more widespread applications of adaptive optics in the near future. One of the most interesting applications of adaptive optics systems involves ophthalmic retinal imaging systems.

Ophthalmic imaging has evolved as one of the most effective tools available for the early detection and diagnosis of physiological disorders in the eye and is also playing a critical role in the treatment modalities such as retina surgery and drug delivery. The retina is a light-sensitive tissue lining the inner surface of the eye. It converts light into neural signals sent to the brain. Ophthalmologists utilize images of the retina in the human eye for the early detection of ocular diseases such as macular degeneration and retinopathy. The ability to visually observe the retina provides the opportunity to noninvasively monitor normal retinal function, the progression of retinal diseases, and the effectiveness of therapies for these diseases. Therefore, ophthalmology is one of the fields that will benefit from the opportunities provided by advanced adaptive optics imaging technologies (Godara et al. 2010; Bennett and Barry 2009).

The full potential of ophthalmic imaging has not yet been realized due to performance limitations of the conventional imaging systems, which result from the presence of aberrations in the eye. These aberrations are caused by naturally occurring defects in the eye. They affect the optical path length of the light waves that form the images of the retina, causing blurring or distortion of the resulting images. The aberrations that limit the resolution of retinal images are actually the same as those causing the usual loss of visual acuity of the eye. The simpler ones of these aberrations—commonly categorized as low-order aberrations—can be routinely compensated for using conventional optical components such as lenses and mirrors. Regular spectacles are an example of this remedy. However, conventional optical systems cannot rectify the more complex aberrations known as high-order aberrations. The high-order aberrations play a less significant role in the visual acuity of the eye. But they pose a formidable challenge in high-resolution retinal imaging where they limit the resolution of the images because of the inability of the conventional optics to compensate for these aberrations. Moreover, the aberrations in the human eye are dynamic in nature and cannot be corrected using conventional imaging systems, which provide static correction only.

Adaptive optics systems represent the technological remedy to the problem of canceling out the effects of high-order time-varying aberrations. Originally developed to enhance the quality of astronomical observations, adaptive optics systems make use of adaptive optical elements—wavefront correctors (WFC)—to compensate for these complex high-order aberrations (Tyson 2011; Hardy 1998). The wavefront correctors act on the wavefront of the light waves affected by the aberrations and cancel out the optical path difference caused by these aberrations. By far, the solid deformable mirrors are the most widely used type of wavefront correctors. These mirrors are characterized by a deformable reflective thin plate or membrane surface whose shape can be dynamically controlled by electrostatic or magnetic actuators underneath. The light waves affected by aberrations are directed to and then reflected back from this deformable surface. By precisely

controlling the shape of the surface, the optical path difference caused by the high-order time-varying aberrations is dynamically compensated. However, the common drawbacks of these solid thin plate or membrane-based mirrors are the high cost of per actuator channel and the relatively low stroke deflection, which may consequently prohibit the widespread AO applications. Notwithstanding the demonstrated potential of AO in ophthalmic imaging systems, the technology has been facing significant challenges in making its way to the clinical imaging systems. The high cost of wavefront correctors has until recently kept AO systems beyond the reach of clinical applications. The other major requirement of the wavefront corrector to be used in ophthalmic AO systems is the unusually large stroke (stroke of a deformable mirror is the magnitude of the maximum deflection of its deformable surface), as high as $\pm 12\ \mu$m or more (Doble and Miller 2006; Doble and Williams 2004), needed to compensate for the aberrations in the eye, which present a large peak-to-valley optical path difference and are beyond the stroke ability of many of the commercially available wavefront correctors.

A few years ago, magnetic fluid deformable mirrors (MFDMs) were proposed as a promising new type of wavefront correctors by Borra et al. (2009, 2008, 2006) and Iqbal et al. (2009, 2008, 2007). These mirrors are developed by coating the free surface of magnetic fluids with a thin film of the reflective materials called metal liquid-like fluids (MELLFs). Magnetic fluids are suspensions of small (about 10 nm in diameter) ferromagnetic nanoparticles dispersed in a liquid carrier. In the presence of an external magnetic field, the ferromagnetic particles align with the field, and the liquid becomes magnetized. The reflective surface can be deformed with the magnetic fluid using a locally applied magnetic field and thus serves as a deformable mirror. Magnetic fluid deformable mirrors have the major advantage over solid ones. These liquids have extremely smooth surfaces that naturally follow the equipotential surfaces created by magnetic fields. MFDMs can have smooth deviations from flatness that can be as small as a few nanometers to as large as several millimeters. The other advantage that MFDMs have with respect to solid deformable mirrors is their low cost per actuator. They are expected to cost orders of magnitude less than any of the known types of wavefront correctors (Borra et al. 2009; Laird et al. 2006). Moreover, the deflections of the reflective surface provided by these mirrors are far larger than those possible with any other known deformable mirrors (Brousseau et al. 2011, 2010; Laird et al. 2006). MFDMs have been found potentially suitable for the ophthalmic applications of AO systems, where they present a solution to the problem of high cost of the existing wavefront correctors as well to the major challenge requiring large stroke of the wavefront correctors.

The MFDM technology is still in the initial stages of its development. Besides the promising capabilities offered by these mirrors, initial studies have also identified some critical difficulties that need to be overcome before the technology is made available for practical applications in imaging systems (Borra et al. 2008; Brousseau et al. 2006, 2007; Laird et al. 2003, 2006; Thibault et al. 2006). Firstly, the amplitude of the deformations produced at the surface of the ferrofluid shows a nonlinear dependence on the applied magnetic field (or current) produced by the actuators. Secondly, vectorial behavior of the magnetic field prohibits the use of standard

control algorithms to predict the surfaces produced by the MFDM. Finally, since deformations are proportional to the square of the applied magnetic field, only positive deformations can be produced on the MFDM surface. These difficulties prohibit the usage of the appropriate standard control methods in the MFDM-based adaptive optics systems to control the shape of the deformable surface of these mirrors. Conventionally, the design of a controller for the wavefront corrector is performed using a DC model (i.e., static gain) of the WFC, which is referred to as the influence-function matrix. Influence-function-based controllers depend on the assumption that the wavefront corrector is a linear system. Since the response of the conventional MFDMs has been found to be nonlinear, the influence-function-based controllers become ineffective in controlling these mirrors (Seaman et al. 2007; Laird et al. 2006). Due to these difficulties, any future application of these mirrors is contingent upon the development of effective methods to control their surface shape in a closed-loop AO system.

The research work presented in this book aims at bridging this critical gap between the concept of a MFDM and its application in adaptive optics systems (Iqbal et al. 2010a, b, 2009; Iqbal and Ben Amara 2008, 2007; Wu et al. 2011, 2010a, b), for example, ophthalmic AO imaging, astronomical AO imaging, and laser beam shaping. The primary goal of the undertaken research work is to provide the necessary means to control the surface shape of the mirror such that the effects of the complex high-order aberrations can be canceled out. An accurate model of the mirror surface shape, which can be used in the development of effective controllers for the mirror surface shape, is sought. Secondly, to resolve the problem of nonlinearity in the response of the mirror surface, a novel change in the design of an MFDM is proposed by superposing a strong and uniform magnetic field to the magnetic field of the actuators, thereby linearizing the response of the MFDM. The major advantage of this linearization is that one can use the same proven control algorithms that are used with solid deformable mirrors. Third, novel control algorithms designed by explicitly accounting for the dynamics of the MFDM are used to optimally control the surface shape of the MFDM. The major contributions presented in this book are summarized as follows:

- *Analytical Model of an MFDM.* A comprehensive model of the dynamics of the surface shape of an MFDM is developed. The analytically developed model describes the dynamics of the surface shape in terms of time-varying displacements of the surface. The displacements are derived as a function of the magnetic field applied to control the mirror surface shape. The model is obtained by solving fundamental equations governing the coupled fluid-electromagnetic system representing the MFDM, is developed in both Cartesian and cylindrical coordinate systems, and is presented in its final form as a state-space model.
- *Modification of the Conceptual Design of an MFDM.* A novel modification in the conceptual design of an MFDM is presented. The design change was prompted by the findings of the analytical work undertaken to develop the model of the mirror. The proposed design involves placing the MFDM inside a Helmholtz coil where a uniform magnetic field is generated. As opposed to the nonlinear

character of the response of the existing mirror designs, the proposed design provides a linear change in the surface shape as a function of the magnetic field applied to control the surface shape. The proposed design accompanies other significant benefits such as bidirectional control and manyfold amplification of the maximum achievable displacements of the mirror surface.

- *Control Algorithms.* Control algorithms aimed at controlling the surface shape of the MFDM are developed and implemented. The above-mentioned model is utilized in the development of these control algorithms. The first control algorithm is a decentralized proportional-plus-integral (PI) controller developed based on the assumption that the plant can be approximated by its DC model. This type of control algorithms is commonly used in AO systems and can effectively handle static aberrations. To improve the stability robustness properties of the closed-loop AO system, a decentralized robust proportional–integral–derivative (PID) controller is then proposed. The resulting closed-loop system can effectively deal with static aberrations and has better stability guarantees than the closed-loop system based on the PI controller. The above-mentioned algorithms perform well in canceling static aberrations, but not as well when dynamic aberrations are present. To handle complex dynamic aberrations, centralized control algorithms are developed based on the analytic mathematical system model. The multivariable controllers consist mainly of a multivariable fixed structure (PID in this case) controller designed based on different desired optimal performance constraints and a dynamic output feedback controller structure designed using the mixed sensitivity H_∞ design approach. The resulting closed-loop system is capable of minimizing the effects of both static and dynamic aberrations while keeping the magnitude of the control signal relatively small.
- *Closed-Loop Operation of an MFDM.* The performance of MFDM is experimentally tested and evaluated in a closed-loop AO system for the first ever time using a designed prototype MFDM. It is practically demonstrated that wavefront aberrations can be corrected with a high level of spatial resolution. Moreover, the ability of the MFDM to compensate for the high-order dynamic aberrations is also demonstrated. The experimental results show the high potential offered by MFDM for applications in ophthalmic adaptive optics imaging systems and others.

The main advantages of MFDM with respect to solid mirrors are large stroke, low cost, and scalability. This technology could lead to extreme deformable optics capable of strokes of hundreds of micrometers and more (Borra et al. 2008, 2006). Early experiments indicated that MFDMs would be limited to dynamic uses at frequencies lower than 20 Hz. However, in recent experiments (Parent et al. 2009), it shows the problematic phase lag at high frequencies can be countered by increasing the viscosity of the liquid, and the bandwidth of MFDMs can be increased to 900 Hz by the approach of overdriving the actuators. Therefore, though the initial design of the prototype MFDM proposed in this book is relatively crude, it still shows the great advantages over solid optics. The developments described in this book pave

the way for further enhancements and improvements in the design and development of AO systems based on MFDMs and their applications in areas such as vision science, astronomical imaging, optical testing, and laser beam shaping.

1.2 Mathematical Preliminaries

This section introduces the reader to general mathematical tools used in the following chapters. These tools are essential in the construction of the analytical model and in the development of the mirror surface shape control algorithms. The first part of this section is related to the mathematical modeling of the MFDMs and includes the multiphysics laws governing the dynamics of the mirror and the corresponding analytic solutions. The underlying physical laws considered here are the basic conservation laws including conservation of fluid mass and magnetic field. Equations are generally limited to algebraic equations, ordinary differential equations, and partial differential equations, where time and distance usually enter as independent variables, and geometry as either a differential element or an entity of finite size. The sets of physical laws and expressions provide us with the tools for establishing a mathematical model.

The second part of this section contains the preliminaries of the system control theory. These preliminaries form the basic building blocks for the analysis and design techniques used to control the MFDMs. The treatment is limited to linear systems mainly because the model obtained for the MFDM has been effectively linearized and the theory for linear systems is relatively well developed and easier to implement. Experimental results presented in subsequent chapters will show that the linear controllers designed using the linear MFDM model provide satisfactory performance when applied to the MFDM-based adaptive optics system. The preliminaries also introduce some advanced modern multivariable control techniques including state-space realization, optimal control, and robust control. The powerful tool of linear matrix inequalities used in approaching optimal control problems is also covered briefly.

1.2.1 Basics of Fluid Mechanics and Electromagnetics Modeling

As mentioned above, an MFDM is based on the deformation of the free surface of a magnetic fluid under the influence of an externally applied magnetic field. The deformation of the fluid surface is governed by fundamental laws of fluid dynamics as well as magnetism. This section presents the basic properties of the fluid and magnetic fields as well as the laws governing the fields. The readers may refer to Rosensweig (1997) and Basmadjian and Farnood (2007) for more details in this area.

1.2.1.1 Magnetic Field Properties

Pole

When a magnet is dipped into a magnetic fluid, the fluid clings to the ends of the magnet. These ends are called *poles*. Like poles repel and unlike poles attract with a force that is proportional to the product of the pole strengths and inversely proportional to the square of the distance between them.

Magnetic Field (H)

The magnetic force acting on a unit pole is defined as *magnetic field*. If there is a unit north-seeking pole placed in the vicinity of a point pole p, the magnetic field $\mathbf{H}$ of the point pole p acting at the unit pole is

$$\mathbf{H} = \frac{p\hat{\mathbf{r}}}{4\pi\mu_0 r^2} = \frac{p\mathbf{r}}{4\pi\mu_0 r^3} \tag{1.1}$$

where $\mathbf{r}$ is the position vector directed from p to the location of the unit pole, $\hat{\mathbf{r}} \equiv \mathbf{r}/r$ is a unit vector having the orientation of $\mathbf{r}$, and μ_0 is called the *permeability* of free space.

Induction Field (B)

An *induction field* $\mathbf{B}$ is defined such that in vacuum, $\mathbf{B} = \mu_0\mathbf{H}$. The $\mathbf{B}$ field may be pictured as lines of induction and is also called *magnetic flux density*. In a uniform induction field $\mathbf{B}$ of unit intensity, one line (or Weber, denoted Wb) is said to cross each square meter of perpendicular surface. Thus, $\mathbf{B}$ has units of Weber per square meter, also known as Tesla (T). An electric current will be induced in any closed circuit when the magnetic flux through a surface bounded by the conductor changes. This applies whether the field itself changes in strength or the conductor is moved through it.

Intensity of Magnetization (M)

The *intensity of magnetization* $\mathbf{M}$ denotes the state of polarization of magnetized matter. For nonisothermal but sensibly constant-density materials, the magnetization thus depends only on the field $\mathbf{H}$ and the temperature T, so that $\mathbf{M} = \mathbf{M}(\mathbf{H},T)$. The induction field $\mathbf{B}$ is related to the intensity of magnetization $\mathbf{M}$ and the magnetic field $\mathbf{H}$ by the following constitutive relationship:

$$\mathbf{B} = \mu\mathbf{H} = \mu_o\,(\mathbf{H} + \mathbf{M}) \tag{1.2}$$

A linear relation between **M** and **H** exists in many materials, that is, diamagnets and paramagnets, and the relation is usually written as

$$\mathbf{M} = \chi \mathbf{H} \tag{1.3}$$

where χ is called the *magnetic susceptibility*.

1.2.1.2 Fluid Field Properties

Velocity

Consider the position coordinates x, y, z to locate a fixed point in the fluid volume comprising the magnetic fluid. The motion of a fluid particle located at this point is described by its *velocity* vector. The velocity components v_x, v_y, v_z measured in the x, y, and z directions, respectively, are functions of the coordinates x, y, z and, in general, the time t:

$$\begin{aligned} v_x &= v_x(x, y, z, t) \\ v_y &= v_y(x, y, z, t) \\ v_z &= v_z(x, y, z, t) \end{aligned} \tag{1.4}$$

If **V** is the velocity vector having components v_x, v_y, and v_z and if **r** is the position vector having components x, y, and z, then

$$\mathbf{V} = f(\mathbf{r}, t) \tag{1.5}$$

Because the velocity **V** is continuously distributed over space, **V**, like the magnetic vectors **B**, **H**, and **M**, is a field variable.

Description of Surface Curvature κ and Normal Vector $\hat{\mathbf{n}}$

The geometry of a surface can be conveniently described using *Monge representation,* which specifies the value of one of the coordinates in terms of the others. In the Cartesian coordinate system, for example, the surface can be represented using $z = \zeta(x, y)$. Accordingly, the expression $z - \zeta(x, y) = C$ represents contours having the same shape as the surface but displaced by magnitude C, and the gradient $\nabla[z - \zeta(x, y)]$ yields a vector normal to the surface. Therefore, the unit normal vector can be obtained as follows:

$$\begin{aligned} \hat{\mathbf{n}} &= \frac{\nabla[z - \zeta(x, y)]}{|\nabla[z - \zeta(x, y)]|} \\ &= \frac{-(\partial\zeta/\partial x)\,\hat{\mathbf{i}} - (\partial\zeta/\partial y)\,\hat{\mathbf{j}} + \hat{\mathbf{k}}}{\left[(\partial\zeta/\partial x)^2 + (\partial\zeta/\partial y)^2 + 1\right]^{\frac{1}{2}}} \end{aligned} \tag{1.6}$$

Ignoring the higher-order terms, (1.6) yields

$$\widehat{\mathbf{n}} = -\left(\frac{\partial \zeta}{\partial x}\right)\widehat{\mathbf{i}} - \left(\frac{\partial \zeta}{\partial y}\right)\widehat{\mathbf{j}} + \widehat{\mathbf{k}} \tag{1.7}$$

The mean curvature κ of a surface can be found using

$$\kappa = \tfrac{1}{2}\nabla \cdot \hat{\mathbf{n}} \tag{1.8}$$

Employing the linearized form (1.7) of the unit normal vector, (1.8) simplifies to

$$\kappa = -\frac{1}{2}\left(\frac{\partial^2 \zeta}{\partial x^2} + \frac{\partial^2 \zeta}{\partial y^2}\right) \tag{1.9}$$

1.2.1.3 Equations Governing the Dynamics of Magnetic Fluids

Continuity Equation

Application of the law of conservation of mass by requiring the sum of the input rates minus the output rates to equal the rate of accumulation yields

$$\frac{\partial \rho v_x}{\partial x} + \frac{\partial \rho v_y}{\partial y} + \frac{\partial \rho v_z}{\partial z} = -\frac{\partial \rho}{\partial t} \tag{1.10}$$

where ρ is the density of the fluid. Or, in vector notation,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{V}) = 0 \tag{1.11}$$

which is the continuity equation of fluid mechanics. For an incompressible liquid, ρ is constant, and the continuity equation reduces to

$$\nabla \cdot \mathbf{V} = 0 \tag{1.12}$$

Generalized Bernoulli Equation

The time rate of change of momentum for the constant mass contained in a deformable element having volume $\mathrm{d}x\mathrm{d}y\mathrm{d}z$ is

$$\frac{D}{Dt}(\rho \mathbf{V}\mathrm{d}x\mathrm{d}y\mathrm{d}z) = \rho \mathrm{d}x\mathrm{d}y\mathrm{d}z\frac{D\mathbf{V}}{Dt} + \mathbf{V}\frac{D\rho \mathrm{d}x\mathrm{d}y\mathrm{d}z}{Dt} \tag{1.13}$$

Because the mass $\rho \mathrm{d}x\mathrm{d}y\mathrm{d}z$ is constant, the last term vanishes and Newton's law normalized to unit volume can be written

$$\rho \frac{D\mathbf{V}}{Dt} = \underbrace{\mathbf{f}_p}_{\text{Pressure force}} + \underbrace{\mathbf{f}_v}_{\text{Viscous force}} + \underbrace{\mathbf{f}_g}_{\text{Gravity force}} + \underbrace{\mathbf{f}_m}_{\text{Magnetic force}} \tag{1.14}$$

where D/Dt is the substantial derivative defined as the following convention:

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{V} \cdot \nabla \tag{1.15}$$

The right side of (1.14) is the sum of the body forces normalized to a unit volume. The terms familiar from fluid mechanics are the pressure gradient

$$\mathbf{f}_p = -\nabla p\,(\rho, T)$$

the viscous force

$$\mathbf{f}_v = \nabla \cdot \mathbf{T}_v$$

the gravitational force

$$\mathbf{f}_g = \rho \mathbf{g}$$

and the magnetic force

$$\mathbf{f}_m = -\nabla \left[\mu_0 \int_0^H \left(\frac{\partial M \upsilon}{\partial \upsilon} \right)_{H,T} \mathrm{d}H \right] + \mu_0 M \nabla H$$

where $\mathbf{T}_v$ is the viscous stress tensor and $\mathbf{g}$ is the local acceleration due to gravity. $\upsilon = \rho^{-1}$ is the volume density. The viscous force density for an incompressible liquid is given by

$$\mathbf{f}_v = \eta \nabla^2 \mathbf{V}$$

where η is the viscosity coefficient of the fluid. Substituting the expressions for $\mathbf{f}_v$, $\mathbf{f}_p$, $\mathbf{f}_g$, and $\mathbf{f}_m$ into the equation of motion (1.14) gives the following definite form of the equation of motion for a magnetic fluid:

$$\begin{aligned} \rho \frac{D\mathbf{V}}{Dt} &= -\nabla p\,(\rho, T) + \eta \nabla^2 \mathbf{V} + \rho \mathbf{g} \\ &\quad - \nabla \left[\mu_0 \int_0^H \left(\frac{\partial M \upsilon}{\partial \upsilon} \right)_{H,T} \mathrm{d}H \right] + \mu_0 M \nabla H \end{aligned} \tag{1.16}$$

The ferrohydrodynamic Bernoulli equation can be obtained as a general integral of the equation of motion developed as

$$\rho\left(\frac{\partial \mathbf{V}}{\partial t}+\mathbf{V}\cdot\nabla\mathbf{V}\right)=-\nabla p^{*}+\mu_0 M\nabla H+\eta\nabla^2\mathbf{V}+\rho\mathbf{g} \tag{1.17}$$

This equation represents a generalization of the Navier–Stokes equation of conventional fluid mechanics such that a magnetic body force $\mu_0 M\nabla H$ appears on the right side and the composite pressure p^* is defined as follows:

$$\begin{aligned} p^{*} &= p(\rho,T)+\mu_0\int_0^H\left(\frac{\partial M\upsilon}{\partial\upsilon}\right)_{H,T}\mathrm{d}H \\ &= p(\rho,T)+\mu_0\int_0^H\left(\upsilon\frac{\partial M}{\partial\upsilon}\right)_{H,T}\mathrm{d}H+\mu_0\int_0^H M\,\mathrm{d}H \\ &= p+p_{\mathrm{s}}+p_{\mathrm{m}} \end{aligned} \tag{1.18}$$

where p is the fluid pressure, p_{s} is the magnetostrictive pressure defined as

$$p_{\mathrm{s}}\equiv\mu_0\int_0^H\upsilon\left(\frac{\partial M}{\partial\upsilon}\right)_{H,T}\mathrm{d}H \tag{1.19}$$

and p_{m} is the fluid-magnetic pressure given as

$$p_{\mathrm{m}}\equiv\mu_0\int_0^H M\,\mathrm{d}H=\mu_0\bar{M}H \tag{1.20}$$

The field-averaged magnetization $\bar{M}$ is defined as

$$\bar{M}=\frac{1}{H}\int_0^H M\,\mathrm{d}H \tag{1.21}$$

From a well-known vector identity, $\nabla^2\mathbf{V}=\nabla(\nabla\cdot\mathbf{V})-\nabla\times(\nabla\times\mathbf{V})$, where $\nabla\times\mathbf{V}\equiv\ \Omega$ is called the *vorticity*. For an incompressible liquid, using (1.12), $\nabla^2\mathbf{V}=-\nabla\times\Omega$. If the fluid is inviscid ($\eta=0$) or the flow is irrotational ($\Omega=0$), the viscous term in the equation of motion (1.17) is identically zero, and it can be written as

$$\rho\left(\frac{\partial \mathbf{V}}{\partial t}\right)+\rho\mathbf{V}\cdot\nabla\mathbf{V}=-\nabla p^{*}+\mu_0 M\nabla H+\rho\mathbf{g} \tag{1.22}$$

Using the following vector properties:

$$\mathbf{V}\cdot\nabla\mathbf{V}=\nabla\left(\frac{1}{2}v^2\right)-\mathbf{V}\times(\nabla\times\mathbf{V}) \tag{1.23}$$

$$\nabla \int_0^H M \, \mathrm{d}H = M \nabla H + \int_0^H \frac{\partial M}{\partial T} \nabla T \, \mathrm{d}H \tag{1.24}$$

equation (1.22) can be rewritten as

$$\rho \frac{\partial \mathbf{V}}{\partial t} - \rho \mathbf{V} \times \Omega = -\nabla \left(p^* + \rho \frac{v^2}{2} + \rho g h - \mu_0 \int_0^H M \, \mathrm{d}H \right) - \mu_0 \int_0^H \frac{\partial M}{\partial T} \nabla T \, \mathrm{d}H \tag{1.25}$$

In going from (1.22), (1.23), (1.24), (1.25), it was assumed that $g = |\mathbf{g}|$ is constant and h is the elevation in the direction opposite to gravity above some reference level. For irrotational flow, $\Omega = 0$, so there exists a velocity potential ϕ such that $\mathbf{V} = -\nabla \phi$. Then, if $\nabla T = 0$ or $\partial M / \partial T = 0$, (1.25) can be written as

$$\nabla \left(-\rho \frac{\partial \phi}{\partial t} + p^* + \frac{1}{2} \rho v^2 + \rho g h - \mu_0 \bar{M} H \right) = 0 \tag{1.26}$$

When (1.26) is integrated, the quantity in parentheses can at most equal a function of time $f(t)$:

$$-\rho \frac{\partial \phi}{\partial t} + p^* + \frac{1}{2} \rho v^2 + \rho g h - \mu_0 \bar{M} H = f(t) \tag{1.27}$$

Equation (1.27) is the time-dependent ferrohydrodynamic Bernoulli equation. For steady-state or time-invariant flow, $\partial \phi / \partial t = 0$ and $f(t) =$ constant, so the generalized Bernoulli equation reduces to

$$p^* + \frac{1}{2} \rho v^2 + \rho g h - \mu_0 \bar{M} H = \text{constant} \tag{1.28}$$

Or, equivalently, using the appropriate definition of p^*,

$$p + p_s + \frac{1}{2} \rho v^2 + \rho g h = \text{constant} \tag{1.29}$$

Maxwell's Equations

Maxwell's equations relate electric and magnetic fields to the charges and currents which produce them. They consist of four equations and are expressed as follows:

$$\nabla \cdot \mathbf{E} = 4 \pi \xi \tag{1.30}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{1.31}$$

$$\nabla \times \mathbf{E} = -\frac{1}{C}\frac{\partial \mathbf{B}}{\partial t} \tag{1.32}$$

$$\nabla \times \mathbf{B} = -\frac{1}{C}\frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{C}\mathbf{J} \tag{1.33}$$

where $\mathbf{E}$ denotes the electric field, $\mathbf{B}$ the magnetic field, ξ charge density, $\mathbf{J}$ the current density, and C the speed of the light. Maxwell's equations express the fact that time-varying magnetic fields produce electric fields, whereas time-varying electric fields, in turn, produce magnetic fields. In most work to date in magnetic fluid, the medium is ferromagnetically responsive, and both the free current density and Maxwell's displacement current are negligible. Hence, the field equations of magnetic fluid are usually employed in the magnetostatic limit of Maxwell's equations as

$$\nabla \times \mathbf{H} = \mathbf{0} \tag{1.34}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{1.35}$$

1.2.1.4 Boundary Conditions

The integral forms of Maxwell's equations describe the behavior of electromagnetic field quantities in all geometric configurations. The differential forms of Maxwell's equations are only valid in regions where the parameters of the media are constant or vary smoothly. In order for a differential form to exist, the partial derivatives must exist, and this requirement breaks down at the boundaries between different materials. For the special case of points along boundaries, we must derive the relationship between field quantities immediately on either side of the boundary from the integral forms. The same case also applies to the ferrohydrodynamic Bernoulli equation at the boundaries.

Magnetic Field Boundary Conditions

Normal Component of $\mathbf{B}$: The boundary condition for the normal component of the magnetic field can be obtained by applying Gauss's flux law:

$$\oint_S \mathbf{B} \cdot \mathrm{d}\mathbf{S} = 0$$

to a small pillbox. Figure 1.1a shows a small volume of the Gaussian pillbox whose upper and lower surface are parallel and located on either side of the interface. If we shrink the side wall of the pillbox to zero, all magnetic flux leaves/enters the pillbox through the top two surfaces, which results in

$$\oint_S \mathbf{B} \cdot \mathrm{d}\mathbf{S} = (B_{1n} - B_{2n})\,\mathrm{d}S = 0$$

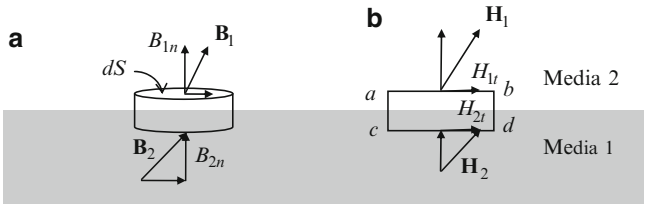


Fig. 1.1 The continuous conditions of the normal component of **B** and tangential component of **H** across the boundary between two different media

where B_{1n} and B_{2n} denote the normal component of $\mathbf{B}_1$ and $\mathbf{B}_2$, respectively. This result states that the normal component of **B** is continuous at boundaries. Hence, the boundary condition for the induced magnetic field **B** can be expressed as

$$B_{1n} - B_{2n} = 0 \tag{1.36}$$

or in vector form as

$$\hat{\mathbf{n}} \cdot (\mathbf{B}_1 - \mathbf{B}_2) = 0 \tag{1.37}$$

Tangential Component of **H**: The boundary condition of the tangential component of magnetic field **H** can be derived by applying Ampere's law:

$$\oint \mathbf{H} \cdot \mathrm{d}\mathbf{l} = 0$$

which is valid when there is no current flow. Ampere's law may be applied to the contour of a closed loop enclosing the interface as shown in Fig. 1.1b. Since the sides *ac* and *bd* approach zero length, they offer no contribution to the integral. The Ampere's law based on the remaining two sides *ab* and *cd* yields

$$\oint \mathbf{H} \cdot \mathrm{d}\mathbf{l} = (H_{1t} - H_{2t})\mathrm{d}l = 0$$

where H_{1t} and H_{2t} denote the tangential component of $\mathbf{H}_1$ and $\mathbf{H}_2$, respectively. Thus, the tangential component of magnetic field **H** is continuous at boundaries. This may be written

$$H_{1t} - H_{2t} = 0 \tag{1.38}$$

or in vector form as

$$\hat{\mathbf{n}} \times (\mathbf{H}_2 - \mathbf{H}_1) = 0 \tag{1.39}$$

Fluid Field Boundary Conditions

The governing equations for the velocity and pressure fields are partial differential equations applicable at every point in a fluid that is being modeled as a continuum. When they are integrated in discontinuous situation, the additional boundary conditions about the stress, velocity, and possibly its gradient at the natural boundaries of the flow domain are needed.

Stress continuity: When a fluid forms one of the boundaries of the flow, a boundary condition connecting the state of stress in each fluid at the interface needs to be considered. The condition can be written as two separate scalar boundary conditions as the tangential (shear) stress balance and the normal stress balance. The difference between the tangential stress vectors at a fluid–fluid interface depends on the gradient of interfacial tension, which normally is determined by the temperature and the composition of the interface. If we assume these to be uniform, the gradient of interfacial tension will vanish everywhere on the interface. This means that tangential stress is commonly continuous across the interface. For example, if two fluids meet and form a flow boundary, this boundary condition would require that the shear stress τ_{s1} in one fluid equal the shear stress τ_{s2} in the other at the boundary

$$\tau_{s1}|_{\text{at the boundary}} = \tau_{s2}|_{\text{at the boundary}} \tag{1.40}$$

Thus, for a viscous fluid in contact with an inviscid (zero or very low viscosity fluid), this means that at the boundary, the shear stress in the viscous fluid is the same as the shear stress in the inviscid fluid. Since the inviscid fluid can support no shear stress (zero viscosity), this means that the stress is zero at this interface. The boundary condition between a fluid such as a magnetic fluid and air, for example, would be that the shear stress τ_s in the magnetic fluid at the interface is zero:

$$\tau_s|_{\text{at the boundary}} = 0 \tag{1.41}$$

This condition is normally represented as that of vanishing shear stress at a free liquid surface. The normal stress jump boundary condition actually determines the curvature of the interface at the point in question and therefore the shape of the entire fluid–fluid interface. This shape is distorted by the flow. In fact, fluid mechanical problems involving the application of the normal stress balance at a boundary are complicated and must be solved numerically unless one assumes the shape distortion to be very small or of a particularly simple form. In such case, if κ denotes the arithmetic mean curvature of the surface, then the surface force density can be expressed as

$$p_c = 2\sigma\kappa \tag{1.42}$$

where σ is the interfacial tension. p_c is normally called the *capillary pressure*, which is produced on an interface having curvature and acted upon by interfacial tension or

forces of surface. Consider the interface between two immiscible fluids, where fluid 1 is taken to be magnetic ($\mathbf{M} > 0$) and fluid 2 is nonmagnetic ($\mathbf{M} = 0$), the magnetic fluid boundary condition in the absence of viscous forces can then be obtained as (Rosensweig 1997)

$$p_1^* + p_n = p_2 + p_c \tag{1.43}$$

where p_1^* is the composite pressure of fluid 1, p_2 is the thermodynamic pressure of fluid 2, and p_n is called the *magnetic normal traction*, which can be expressed as

$$p_n = \frac{1}{2}\mu_0(\mathbf{M} \cdot \hat{\mathbf{n}})^2 \tag{1.44}$$

In terms of the definition of p^* in (1.18), the boundary condition can be further expressed as

$$p_1 + p_s + p_m + p_n = p_2 + p_c \tag{1.45}$$

It is familiar that the pressure is continuous across a plane fluid boundary for ordinary fluids; however, this condition no longer holds for fluids possessing magnetization, and the magnetic stress will produce a traction force p_n shown in (1.44) at the interface. Note that the generalized Bernoulli equation (1.29) does not contain the fluid-magnetic pressure term p_m; however, p_m appears explicitly in the boundary condition (1.45).

Velocity continuity: It is convenient for the purpose of discussion to identify two types of velocity boundaries. One of the boundaries is at the interface between a fluid and a rigid surface. At such a surface, we shall require that the tangential component of the velocity of the fluid be the same as the tangential component of the velocity of the surface and similarly the normal component of the velocity of the fluid be the same as the normal component of the velocity of the surface. The former is known as the "no-slip" boundary condition and has been found to be successful in describing most practical situations. If we designate the velocity of the rigid surface as $\mathbf{V}_{\text{wall}}$ and that of the fluid as $\mathbf{V}_{\text{fluid}}$, the no-slip boundary condition can be stated as

$$\mathbf{V}_{\text{fluid}}|_{\text{at the boundary}} = \mathbf{V}_{\text{wall}} \tag{1.46}$$

Often, the walls are not moving, so the fluid velocity is zero, that is,

$$\mathbf{V}_{\text{fluid}}|_{\text{at the boundary}} = 0 \tag{1.47}$$

Consider the velocity boundary condition between two fluids; the velocity fields are continuous across the interface

$$\mathbf{V}_{\text{fluid1}}|_{\text{at the boundary}} = \mathbf{V}_{\text{fluid2}}|_{\text{at the boundary}} \tag{1.48}$$

This vector condition can be also viewed in two parts, the continuity of the tangential component of the two velocities and the continuity of the normal component of the two velocities, a kinematic consequence when there is no mass transfer across the interface. Notice that we have two unknown vector fields $\mathbf{V}_{\text{fluid1}}$ and $\mathbf{V}_{\text{fluid2}}$ now and therefore need twice as many boundary conditions compared with the case of a fluid and a rigid surface interface. Therefore, it is necessary to consider the stress boundary conditions in each fluid at the interface together.

The deflection of a free fluid surface can be related to the motion of the adjacent fluid using what is commonly known as the *kinematic condition*. The condition is written as follows:

$$v_z = \frac{\partial \zeta}{\partial t} \tag{1.49}$$

The condition (1.49) states that, at any point on the free surface of a fluid, the rate of deflection of the surface ζ is equal to the vertical component of the fluid velocity v_z at that point.

1.2.2 Control Systems Theory

In the following, an introduction to some basic control system concepts used in this book is presented. The proofs of all standard results presented here are omitted for simplicity. The interested reader could refer to Boyd et al. (1994), Zhou et al. (1995), Chen and Francis (1996), and Scherer and Weiland (2004) for a more complete exposure to these results.

1.2.2.1 Control System Presentation and Basic Properties

It is convenient to introduce some basic notations first. Let $\mathbb{R}$ denote the real scalar field and $\mathbb{C}$ the complex scalar field. $\mathbb{R}^n$ denotes the vector space over $\mathbb{R}$ and $\mathbb{C}^n$ the vector space over $\mathbb{C}$. A matrix $\mathbf{A}$ is Hermitian if it is square and $\mathbf{A} = \bar{\mathbf{A}}^{\mathrm{T}}$, where the bar denotes taking the complex conjugate of each entry in $\mathbf{A}$. If $\mathbf{A}$ is real, then this amounts to saying that $\mathbf{A} = \mathbf{A}^{\mathrm{T}}$ and we call $\mathbf{A}$ *symmetric*. For a given system Σ, $\Sigma(s)$ and $\Sigma(z)$ denote, respectively, the continuous and discrete time transfer functions of the system, and $\sum : \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$ refers to a state-space realization of the system Σ. In the block diagrams presented in this chapter, any given signal is represented using the same symbol in the time domain, Laplace domain, and the z domain.

A control system or plant or process is an interconnection of components to perform certain tasks and to yield a desired response, that is, to generate desired signal (the output) when it is driven by a manipulating signal (the input). In general, there are two categories of control systems, namely, open-loop systems

and closed-loop systems. In an open-loop system, a reference signal applied to the controller is used to generate the control signal that drives the plant. The output of the plant in this case is not fed back to the controller. In contrast, in a closed-loop control system, both the reference signal and the plant output measurement are used as inputs to the controller to generate the control signals for the plant, resulting in a feedback system.

In most cases, the evolution of physical control systems can be approximately modeled by real ordinary differential equations, that is, the state $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots, x_n(t)]^{\mathrm{T}}$ of the physical system at time t is the solution of the coupled first-order ordinary differential equations:

$$\dot{\mathbf{x}} = f_p\,(t, \mathbf{x}, \mathbf{u}) \tag{1.50}$$

where $\dot{\mathbf{x}}$ denotes the derivative of $\mathbf{x}$ with respect to the time variable t,

$$\mathbf{u}(t) = [u_1(t), u_2(t), \ldots, u_m(t)]^{\mathrm{T}}$$

are specified input variables, and the state $\mathbf{x}$ passes through the point $\mathbf{x}\,(t_0)$ at time $t = t_0$. Usually, we associate with (1.50) another equation

$$\mathbf{y} = h\,(t, \mathbf{x}, \mathbf{u}) \tag{1.51}$$

which defines a q-dimensional output vector $\mathbf{y}(t) = \left[y_1(t), y_2(t), \ldots, y_q(t)\right]^{\mathrm{T}}$ that comprises variables of particular interest in the analysis of the dynamical system, like variables which can be physically measured or variables which are required to behave in a specified manner. We call (1.51) the output equation and normally refer to Eqs. 1.50 and 1.51 together as the state-space presentation model. In general, the functions f_p and h are nonlinear functions of the state variable $\mathbf{x}$.

In closed-loop, a control law $\mathbf{u}(t) = g\,(\mathbf{x}(t), t)$ is selected. Thus, the closed-loop dynamics can be written as

$$\dot{\mathbf{x}} = f\,(\mathbf{x}, t) = f_p\,(t, \mathbf{x}, g\,(\mathbf{x}, t)) \tag{1.52}$$

A special case of (1.52) is when the function f does not depend explicitly on t, that is,

$$\dot{\mathbf{x}} = f\,(\mathbf{x}) \tag{1.53}$$

In these cases, the system is said to be *autonomous*.

A great majority of physical systems are linear within some range of the variables, though in general, systems ultimately become nonlinear as the variables are increased without limit. Furthermore, the nonlinearity of many physical systems can also be linearized approximately based on the assumption of small signal conditions (Dorf and Bishop 2008). For example, the linearization method of Taylor

series expansion around the system operating point is commonly utilized. Therefore, for the sake of simplicity in analyzing (1.50) and (1.51), f_p and h are frequently replaced or approximated by linear functions, hence resulting in a linearized system **G** given by

$$\mathbf{G}: \begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \mathbf{x}(t_0) = \mathbf{x}_0 \\ \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t) \end{cases} \tag{1.54}$$

where $\mathbf{x}(t) \in \mathbb{R}^n$ is called the *system state variable*, $\mathbf{x}(t_0)$ is called the initial condition of the system, $\mathbf{u}(t) \in \mathbb{R}^m$ is called the system input, and $\mathbf{y}(t) \in \mathbb{R}^q$ is the system output. The **A**, **B**, **C**, and **D** are appropriately dimensioned real constant matrices. In this case, we say the system (1.54) is *linear time-invariant* (LTI). A dynamical system with single input ($m = 1$) and single output ($q = 1$) is called a *single-input and single-output* (SISO) system; otherwise, it is called *multiple-input and multiple-output* (MIMO) system. The corresponding transfer matrix from **u** to **y** is defined as

$$\mathbf{Y}(s) = \mathbf{G}(s)\mathbf{U}(s) \tag{1.55}$$

where $\mathbf{U}(s)$ and $\mathbf{Y}(s)$ are the Laplace transform of $\mathbf{u}(t)$ and $\mathbf{y}(t)$ with zero initial conditions. The system transfer matrix can be obtained directly from (1.54) as

$$\mathbf{G}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D} \tag{1.56}$$

The state-space presentation (1.54) is sometimes written in a more compact matrix form as

$$\mathbf{G} := \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$$

The signals in control systems are usually continuous time signals, and the corresponding performance specifications are formulated in continuous time. However, with the application of the computer technology in the control systems, the controllers are frequently implemented using digital technology, which can take the form of a microcontroller to an ASIC to a standard desktop computer, etc. Normally, a digital controller samples and quantizes a continuous time signal, such as a tracking error, using an A/D converter to produce a digital signal, then it processes this digital signal according to the proper designed control algorithm and finally converts the resulting digital signal back into a continuous time control signal for the plant using a D/A converter. Such a control system thus involves both continuous time and discrete time signals and can be analyzed in a discrete time framework. Usually, two discretization techniques—zero-order hold transformation and bilinear transformation—are used to convert the continuous time system to a discrete time system. Zero-order hold transformation is suitable to discretize the physical plant

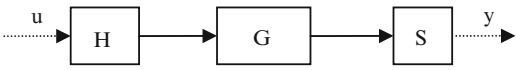


Fig. 1.2 Zero-order hold system

with a zero-order hold as shown in Fig. 1.2, where $\mathbf{G}$ is the continuous time plant, $\mathbf{H}$ and $\mathbf{S}$ are hold and sampling devices which are implemented using D/A and A/D converters, respectively. The transfer function of the zero-order hold is

$$\mathbf{H}(s) = \frac{1 - \mathrm{e}^{-sT}}{s} \tag{1.57}$$

where T is the sampling period. The Z transform can then be used directly to convert the continuous time model with the zero-order hold to the discrete time model as

$$\mathbf{G}_d(z) = Z\left(\mathbf{G}(s)\mathbf{H}(s)\right) \tag{1.58}$$

where Z denotes the Z transform. The Z transform maps the s plane in continuous time domain into a z plane in discrete time domain involving the following conformal relationship:

$$z = \mathrm{e}^{sT} \tag{1.59}$$

Let the state-space representation of $\mathbf{G}_d$ be given by

$$\begin{aligned} \mathbf{G}_d &:= \begin{pmatrix} A_d & B_d \\ C_d & D_d \end{pmatrix} \\ &:= \begin{cases} \mathbf{x}(k+1) = \mathbf{A}_d\mathbf{x}(k) + \mathbf{B}_d\mathbf{u}(k) \\ \mathbf{y}(k) = \mathbf{C}_d\mathbf{x}(k) + \mathbf{D}_d\mathbf{u}(k) \end{cases} \end{aligned} \tag{1.60}$$

The Z transfer function can be obtained by

$$\mathbf{G}_d(z) = \mathbf{C}_d(z\mathbf{I} - \mathbf{A}_d)^{-1}\mathbf{B}_d + \mathbf{D}_d \tag{1.61}$$

The above state-space presentation (1.60) can also be converted from (1.54) directly with

$$\mathbf{A}_d := \mathrm{e}^{T\mathbf{A}}, \quad \mathbf{B}_d := \int_0^T \mathrm{e}^{\tau\mathbf{A}}\mathrm{d}\tau\mathbf{B}, \quad \mathbf{C}_d := \mathbf{C}, \quad \mathbf{D}_d := \mathbf{D}$$

The bilinear transformation is frequently used to discretize an analog controller for the purpose of digital implementation. In this case, one might design an analog controller $\mathbf{K}$ based on a continuous time model, then convert from the analog

controller $\mathbf{K}$ to a digital controller $\mathbf{K}_d$. The bilinear transformation is also called *Tustin transformation* which converts the continuous time controller to digital controller with

$$s = \frac{2(z-1)}{T(z+1)} \tag{1.62}$$

Tustin transformation is actually the Pade approximation of the exponential function $z = \mathrm{e}^{sT}$:

$$\begin{aligned} z &= \mathrm{e}^{sT} \\ &= \frac{\mathrm{e}^{sT/2}}{\mathrm{e}^{-sT/2}} \\ &\approx \frac{1 + sT/2}{1 - sT/2} \end{aligned} \tag{1.63}$$

Based on (1.63), then it is easy to obtain (1.62). The digital controller will achieve an output which approaches the output of its respective analog controller as the sampling period is decreased.

A critical issue in control system design is that of system stability. An unstable system is of no practical value. This is because any control system is vulnerable to disturbances and noises in a real work environment, and the effect due to these signals would adversely affect the expected, normal system output in an unstable system. Feedback control techniques may reduce the influence generated by uncertainties and achieve desirable performance. However, an inadequate feedback controller may lead to an unstable closed-loop system though the original open-loop system is stable. When a dynamic system is just described by its input/output relationship such as a transfer function (matrix), the system is stable if it generates bounded outputs for any bounded inputs. This is called the bounded-input bounded-output (BIBO) stability. For a linear, time-invariant system modeled by a transfer function matrix ($\mathbf{G}(s)$ in (1.55)), the BIBO stability is guaranteed if and only if all the poles of $\mathbf{G}(s)$ are in the open-left-half complex plane, that is, with negative real parts. When a system is governed by a state-space model such as (1.54), a stability concept called *asymptotic stability* can be defined. A system is asymptotically stable if, for an identically zero input, the system state will converge to zero from any initial states. For a linear time-invariant continuous time system described by a model of (1.54), it is asymptotically stable if and only if all the eigenvalues of the state matrix $\mathbf{A}$ are in the open-left-half complex plane, that is, with negative real parts. While for a linear time-invariant discrete time system (1.60), the stability condition is that all the eigenvalues of the state matrix $\mathbf{A}_d$ are inside the unit circle on the Z plane, that is, the spectral radius of $\mathbf{A}_d$ satisfies $\rho(\mathbf{A}_d) < 1$. The above stabilities are defined both for open-loop systems and closed-loop systems. For a closed-loop system, a controller should be designed properly to make the system stable and satisfy the desired performance as well. In the following, some important concepts in linear system theory will be discussed.

Definition 1.1 The dynamical system described by Eq. 1.54 or the pair $(\mathbf{A}, \mathbf{B})$ is said to be controllable if, for any initial state $\mathbf{x}(0) = \mathbf{x}_0$ and final state $\mathbf{x}(t_1) = \mathbf{x}_1$, $t_1 > 0$, there exists an input $\mathbf{u}(\cdot)$ such that the solution of (1.54) satisfies $\mathbf{x}(t_1) = \mathbf{x}_1$. Otherwise, the system or the pair $(\mathbf{A}, \mathbf{B})$ is said to be uncontrollable. □

The definition of the controllability of the discrete time system is similar and will not be repeated here. The controllability of a system can be verified through the following theorem.

Theorem 1.1 *Consider the system (1.54) or (1.60), then* $(\mathbf{A}, \mathbf{B})$ *or* $(\mathbf{A}_d, \mathbf{B}_d)$ *is controllable if and only if*

(i) *The controllability matrix*

$$C = \left[\mathbf{B}\ \mathbf{AB}\ \mathbf{A}^2\mathbf{B}\ \ldots\ \mathbf{A}^{n-1}\mathbf{B}\right]$$

or

$$C_d = \left[\mathbf{B}_d\ \mathbf{A}_d\mathbf{B}_d\ \mathbf{A}_d^2\mathbf{B}_d\ \ldots\ \mathbf{A}_d^{n-1}\mathbf{B}_d\right]$$

has full row rank.

(ii) The matrix $[\mathbf{A} - \lambda\mathbf{I}, \mathbf{B}]$ *or* $[\mathbf{A}_d - \lambda\mathbf{I}, \mathbf{B}_d]$ *has full row rank for all* λ *in* $\mathbb{C}$.

(iii) The eigenvalues of $\mathbf{A} + \mathbf{BF}$ *or* $\mathbf{A}_d + \mathbf{B}_d\mathbf{F}$ *can be freely assigned (with the restriction that complex eigenvalues are in conjugate pairs) by a suitable choice of* $\mathbf{F}$. □

The dynamic system (1.54) or (1.60) is said to be stabilizable if there exists a state feedback $\mathbf{u} = \mathbf{Fx}$ such that the closed-loop system is stable, that is, $\mathrm{Re}\lambda(\mathbf{A} + \mathbf{BF}) < 0$ or $\rho(\mathbf{A}_d + \mathbf{B}_d\mathbf{F}) < 1$.

Definition 1.2 The dynamical system described by the Eq. 1.54 or by the pair $(\mathbf{C}, \mathbf{A})$ is said to be observable if, for any $t_1 > 0$, the initial state $\mathbf{x}(0) = \mathbf{x}_0$ can be determined from the time history of the input $\mathbf{u}(t)$ and the output $\mathbf{y}(t)$ in the interval of $[0, t_1]$. Otherwise, the system or $(\mathbf{C}, \mathbf{A})$ is said to be unobservable. □

The definition of observability of discrete time systems follows along the same lines. The observability of a system can be verified through the following theorem.

Theorem 1.2 *Consider the system (1.54) or (1.60), then* $(\mathbf{C}, \mathbf{A})$ *or* $(\mathbf{C}_d, \mathbf{A}_d)$ *is observable if and only if*

(i) The observability matrix

$$O = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \mathbf{CA^2} \\ \vdots \\ \mathbf{CA^{n-1}} \end{bmatrix} \text{ or } O_d = \begin{bmatrix} \mathbf{C}_d \\ \mathbf{C}_d\mathbf{A}_d \\ \mathbf{C}_d\mathbf{A}_d^2 \\ \vdots \\ \mathbf{C}_d\mathbf{A}_d^{n-1} \end{bmatrix}$$

has full column rank.

(ii) *The matrix* $\begin{bmatrix} \mathbf{A}-\lambda\mathbf{I} \\ \mathbf{C} \end{bmatrix}$ *or* $\begin{bmatrix} \mathbf{A}_d-\lambda\mathbf{I} \\ \mathbf{C}_d \end{bmatrix}$ *has full column rank for all* λ *in* $\mathbb{C}$.

(iii) *The eigenvalues of* $\mathbf{A}+\mathbf{LC}$ *or* $\mathbf{A}_d+\mathbf{LC}_d$ *can be freely assigned (with the restriction that complex eigenvalues are in conjugate pairs) by a suitable choice of* $\mathbf{L}$. □

Then, the dynamic system (1.54) or (1.60) is said to be detectable if there exists $\mathbf{L}$ such that the $\mathbf{A}+\mathbf{LC}$ is stable, that is, $\mathrm{Re}\lambda\,(\mathbf{A}+\mathbf{LC})<0$, or $\mathbf{A}_d+\mathbf{LC}_d$ is stable, that is, $\rho\,(\mathbf{A}_d+\mathbf{LC}_d)<1$.

Testing stability, controllability, and observability of a system is very important in linear system analysis and synthesis. However, these tests often have to be done indirectly. In this respect, the Lyapunov theory is sometimes useful. Consider the following Lyapunov equation for continuous time case:

$$\mathbf{A}^{\mathrm{T}}\mathbf{P}+\mathbf{PA}+\mathbf{Q}=0 \tag{1.64}$$

where $\mathbf{A}, \mathbf{P}, \mathbf{Q}$ are all square matrices, say, $n\times n$, with $\mathbf{Q}$ symmetric. The Lyapunov equation has a unique solution if and only if $\mathbf{A}$ has the property that no two of its eigenvalues add to zero, that is, $\lambda_i\,(\mathbf{A})+\bar{\lambda}_j\,(\mathbf{A})\neq 0,\ \forall i, j$. The following result presents the relationships between the stability of $\mathbf{A}$ and the solution of $\mathbf{P}$.

Theorem 1.3 *Consider the Lyapunov Eq. 1.64, we have*

(i) *Suppose* $\mathbf{A}$ *is stable, then we have the following statements:*

(a) *The unique solution is* $\mathbf{P}=\int_0^\infty \mathrm{e}^{\mathbf{A}^{\mathrm{T}}t}\mathbf{Q}\mathrm{e}^{\mathbf{A}t}\mathrm{d}t$.

(b) *If* $\mathbf{Q}\geq 0$, *then* $(\mathbf{Q},\mathbf{A})$ *is observable if and only if* $\mathbf{P}>0$.

(ii) *Suppose* $\mathbf{A}$, $\mathbf{Q}$, $\mathbf{P}$ *satisfy the Lyapunov equation,* $(\mathbf{Q},\mathbf{A})$ *is detectable, and* $\mathbf{Q}$ *and* $\mathbf{P}$ *are positive semi-definite, then* $\boldsymbol{A}$ *is stable.* □

Based on the above result, it is not difficult to obtain that, given a stable matrix $\mathbf{A}$, a pair $(\mathbf{C},\mathbf{A})$ is observable if and only if the solution to the following Lyapunov equation is positive definite:

$$\mathbf{A}^{\mathrm{T}}\mathbf{L}_0+\mathbf{L}_0\mathbf{A}+\mathbf{C}^{\mathrm{T}}\mathbf{C}=0 \tag{1.65}$$

The solution $\mathbf{L}_0$ is called the *observability Gramian*. Similarly, a pair $(\mathbf{A},\mathbf{B})$ is controllable if and only if the solution to the following Lyapunov equation is positive definite:

$$\mathbf{AL}_c+\mathbf{L}_c\mathbf{A}^{\mathrm{T}}+\mathbf{BB}^{\mathrm{T}}=0 \tag{1.66}$$

The solution $\mathbf{L}_\mathrm{c}$ is called the *controllability Gramian*.

Similarly, for the discrete time system, we define the following Lyapunov function:

$$\mathbf{A}_d\mathbf{X}\mathbf{A}_d^{\mathrm{T}}-\mathbf{X}+\mathbf{Q}=0 \tag{1.67}$$

where $\mathbf{A}_d$, $\mathbf{X}$, $\mathbf{Q}$ are all square matrices, say, $n \times n$, with $\mathbf{Q}$ symmetric. The Lyapunov equation has a unique solution if and only if $\mathbf{A}_d$ has the property that $\lambda_i(\mathbf{A}_d)\lambda_j\left(\mathbf{A}_d^{\mathrm{T}}\right) \neq 1, \forall i, j$. The following result presents the relationships between the stability of $\mathbf{A}_d$ and the solution of $\mathbf{X}$.

Theorem 1.4 *Consider the Lyapunov equation (1.67), we have*

(i) Suppose $\mathbf{A}_d$ is stable, then we have the following statements:

(a) The unique solution is $\mathbf{X} = \sum_{i=0}^{\infty} \mathbf{A}_d^i \mathbf{Q}\left(\mathbf{A}_d^{\mathrm{T}}\right)^i$, and $\mathbf{X} \geq 0$ if $\mathbf{Q} \geq 0$.
(b) If $\mathbf{Q} \geq 0$, then $(\mathbf{Q}, \mathbf{A}_d)$ is observable if and only if $\mathbf{X} > 0$.

(ii) Suppose $\mathbf{A}_d$, $\mathbf{X}$, $\mathbf{Q}$ satisfy the Lyapunov equation, $(\mathbf{Q}, \mathbf{A}_d)$ is detectable, and $\mathbf{Q}$ and $\mathbf{X}$ are positive semi-definite, then $\mathbf{A}_d$ is stable. □

Correspondingly, we denote $\mathbf{M}_{\mathrm{c}}$ and $\mathbf{M}_{\mathrm{o}}$ the controllability and observability Gramians with

$$\mathbf{A}_d \mathbf{M}_{\mathrm{c}} \mathbf{A}_d^{\mathrm{T}} - \mathbf{M}_{\mathrm{c}} + \mathbf{B}_d \mathbf{B}_d^{\mathrm{T}} = 0 \tag{1.68}$$

$$\mathbf{A}_d^{\mathrm{T}} \mathbf{M}_{\mathrm{o}} \mathbf{A}_d - \mathbf{M}_{\mathrm{o}} + \mathbf{C}_d^{\mathrm{T}} \mathbf{C}_d = 0 \tag{1.69}$$

Besides the aforementioned properties we need to consider for the control system, the optimal performance of the closed-loop system is another important consideration in the design of a control system. In optimal control, the performance of a control system is represented by a performance index. The controller is designed in such a way that the closed-loop system performance index is minimized, resulting in an optimal control system. Optimal control systems should possess good performance specifications, such as good regulation against disturbances, desirable responses to commands, actuator signals that stay within a prespecified range, and good robustness against uncertainties in the system model. One way to describe the performance specifications of a control system is in terms of the size of certain signals of interest. In the following, several norms for signals and induced norms of systems are presented.

1.2.2.2 Performance Specifications

Norms for Signals and Systems

First, consider the continuous or piecewise continuous time scalar-valued signals $\mathbf{x}(t)$, $t \in (-\infty, \infty)$, then l_2 and l_∞ norm of $\mathbf{x}(t)$ are defined as

$$\|\mathbf{x}\|_2 = \left(\int_{-\infty}^{\infty} |\mathbf{x}(t)|^2\right)^{1/2}$$

$$\|\mathbf{x}\|_\infty = \sup_{t \in (-\infty, \infty)} |\mathbf{x}(t)|$$

When $\mathbf{x}(t)$ are continuous or piecewise continuous vector-valued functions of the form $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots x_n(t)]^{\mathrm{T}}$, $t \in (-\infty, \infty)$, we have

$$\|\mathbf{x}\|_2 = \left(\int_{-\infty}^{\infty} \sum\nolimits_{i=1}^{n} |x_i(t)|^2 \right)^{1/2}$$

$$\|\mathbf{x}\|_\infty = \sup_{t \in (-\infty, \infty)} \left(\max_{1 \le i \le n} |x_i(t)| \right)$$

The norms for discrete time signals $\mathbf{v} = \{v(0), v(1), \ldots\}$ can be defined in a similar way

$$\|\mathbf{v}\|_2 = \left[v(0)^2 + v(1)^2 + \cdots \right]^{1/2}$$

$$\|\mathbf{v}\|_\infty = \sup_k |v(k)|$$

The normed spaces, consisting of signals with finite norm as defined correspondingly, are called $l_2(\mathbb{R})$ and $l_\infty(\mathbb{R})$, respectively, as

$$l_2(\mathbb{R}) := \{x(\cdot) : \|\mathbf{x}\|_2 < \infty\}$$

$$l_\infty(\mathbb{R}) := \{x(\cdot) : \|\mathbf{x}\|_\infty < \infty\}$$

From a signal point of view, the square of the 2-norm, $\|\mathbf{x}\|_2^2$, is often called the *energy* of the signal $\mathbf{x}(\cdot)$. The ∞-norm, $\|\mathbf{x}\|_\infty$, is the amplitude or peak value of the signal, and the signal is bounded in magnitude if $\mathbf{x}(\cdot) \in l_\infty(\mathbb{R})$.

System norms are actually the input–output gains of the system. Suppose that $\mathbf{G}$ is a linear and bounded system that maps the input signal $\mathbf{u}(\cdot)$ into the output signal $\mathbf{y}(\cdot)$, denote $\|\bullet\|_U$ and $\|\bullet\|_Y$ are the corresponding norms of the input signal and output signal, respectively. Then, the norm, maximum system gain, of $\mathbf{G}$ is defined as

$$\|\mathbf{G}\| := \sup_{\mathbf{u} \neq 0} \frac{\|\mathbf{G}\mathbf{u}\|_Y}{\|\mathbf{u}\|_U}$$

$\|\mathbf{G}\|$ is called the *induced norm* of $\mathbf{G}$ with regard to the signal norms $\|\bullet\|_U$ and $\|\bullet\|_Y$. In this book, we are particularly interested in the so-called H_2 norm and H_∞ norm of a system. The fact is that H_2 norm of a system is related to the average l_2 norm of the output when impulses are applied at the input channels and H_∞ of a system is related to the maximum l_2 norm of the output overall inputs of unit norm.

Computation of the Induced System Norms

For continuous time system $\mathbf{G}$, H_2 norm and H_∞ norm can be calculated in the frequency domain as

$$\|\mathbf{G}\|_2 := \sqrt{\frac{1}{2\pi}\int_{-\infty}^{\infty} Trace\left\{\mathbf{G}^*\left(j\omega\right)\mathbf{G}\left(j\omega\right)\right\}\mathrm{d}\omega}$$

and

$$\|\mathbf{G}\|_\infty := \operatorname*{ess\,sup}_{\omega} \bar{\sigma}\left\{\mathbf{G}\left(j\omega\right)\right\}$$

For discrete time system $\mathbf{G}_d$, H_2 norm and H_∞ norm can then be calculated as

$$\|\mathbf{G}_d\|_2 := \sqrt{\frac{1}{2\pi}\int_{0}^{2\pi} Trace\left\{\mathbf{G}_d^*\left(e^{j\theta}\right)\mathbf{G}_d\left(e^{j\theta}\right)\right\}\mathrm{d}\theta}$$

and

$$\|\mathbf{G}_d\|_\infty := \operatorname*{ess\,sup}_{\theta} \bar{\sigma}\left\{\mathbf{G}_d\left(\mathrm{e}^{j\theta}\right)\right\}$$

It is useful and convenient to compute the H_2 norm and H_∞ norm by state-space methods, which are presented in the following two lemmas.

Lemma 1.1 *Consider transfer matrices*

$$\mathbf{G}(s) = \left(\begin{array}{c|c} A & B \\ \hline C & 0 \end{array}\right), \quad \mathbf{G}_d(z) = \left(\begin{array}{c|c} A_d & B_d \\ \hline C_d & D_d \end{array}\right)$$

with $\mathbf{A}$ *and* $\mathbf{A}_d$ *stable. Then we have*

$$\|\mathbf{G}\|_2^2 = trace\left(\mathbf{B}^{\mathrm{T}}\mathbf{L}_{\mathrm{o}}\mathbf{B}\right) = trace\left(\mathbf{C}\mathbf{L}_{\mathrm{c}}\mathbf{C}^{\mathrm{T}}\right)$$
$$\|\mathbf{G}_d\|_2^2 = trace\left(\mathbf{D}_d^{\mathrm{T}}\mathbf{D}_d + \mathbf{B}_d^{\mathrm{T}}\mathbf{M}_0\mathbf{B}_d\right) = trace\left(\mathbf{D}_d\mathbf{D}_d^{\mathrm{T}} + \mathbf{C}_d\mathbf{M}_{\mathrm{c}}\mathbf{C}_d^{\mathrm{T}}\right)$$

where $\mathbf{L}_{\mathrm{c}}$, $\mathbf{M}_{\mathrm{c}}$ *and* $\mathbf{L}_{\mathrm{o}}$, $\mathbf{M}_{\mathrm{o}}$ *are controllability and observability Gramians which can be obtained from the following Lyapunov equation, respectively:*

$$\mathbf{A}\mathbf{L}_{\mathrm{c}} + \mathbf{L}_{\mathrm{c}}\mathbf{A} + \mathbf{B}\mathbf{B}^{\mathrm{T}} = 0$$
$$\mathbf{A}^{\mathrm{T}}\mathbf{L}_{\mathrm{o}} + \mathbf{L}_{\mathrm{o}}\mathbf{A} + \mathbf{C}^{\mathrm{T}}\mathbf{C} = 0$$
$$\mathbf{A}_d\mathbf{M}_{\mathrm{c}}\mathbf{A}_d^{\mathrm{T}} - \mathbf{M}_{\mathrm{c}} + \mathbf{B}_d\mathbf{B}_d^{\mathrm{T}} = 0$$
$$\mathbf{A}_d^{\mathrm{T}}\mathbf{M}_{\mathrm{o}}\mathbf{A}_d - \mathbf{M}_{\mathrm{o}} + \mathbf{C}_d^{\mathrm{T}}\mathbf{C}_d = 0$$

□

Lemma 1.2 *Consider a transfer matrix*

$$\mathbf{G}(s) = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right), \quad \mathbf{G}_d(z) = \left(\begin{array}{c|c} A_d & B_d \\ \hline C_d & D_d \end{array}\right)$$

with $\mathbf{A}$ *and* $\mathbf{A}_d$ *stable. Let* $\gamma > 0$*, then* $\|\mathbf{G}\|_\infty < \gamma$ *if and only if* $\bar{\sigma}(\mathbf{D}) < \gamma$ *and* $\mathbf{H}$ *has no eigenvalues on the imaginary axis where*

$$\mathbf{H} := \begin{bmatrix} \mathbf{A} + \mathbf{B}\left(\gamma^2\mathbf{I} - \mathbf{D}^{\mathrm{T}}\mathbf{D}\right)^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{C} & \mathbf{B}\left(\gamma^2\mathbf{I} - \mathbf{D}^{\mathrm{T}}\mathbf{D}\right)^{-1}\mathbf{B}^{\mathrm{T}} \\ -\mathbf{C}^{\mathrm{T}}\left(\mathbf{I} + \mathbf{D}\left(\gamma^2\mathbf{I} - \mathbf{D}^{\mathrm{T}}\mathbf{D}\right)^{-1}\mathbf{D}^{\mathrm{T}}\right)\mathbf{C} & -\left(\mathbf{A} + \mathbf{B}\left(\gamma^2\mathbf{I} - \mathbf{D}^{\mathrm{T}}\mathbf{D}\right)^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{C}\right)^{\mathrm{T}} \end{bmatrix} \tag{1.70}$$

Let $\upsilon > 0$*, then* $\|\mathbf{G}_d\|_\infty < \upsilon$ *if and only if* $\bar{\sigma}(\mathbf{D}) < \upsilon$ *and* $\boldsymbol{S}$ *has no eigenvalues on the unit circle, where* $\boldsymbol{S}$ *denotes the symplectic pair*

$$\mathbf{S} := \left(\begin{bmatrix} \mathbf{A}_d + \mathbf{B}_d\mathbf{D}_d^{\mathrm{T}}\left(\upsilon^2\mathbf{I} - \mathbf{D}_d\mathbf{D}_d^{\mathrm{T}}\right)^{-1}\mathbf{C}_d & \mathbf{0} \\ -\upsilon\mathbf{C}_d^{\mathrm{T}}\left(\upsilon^2\mathbf{I} - \mathbf{D}_d\mathbf{D}_d^{\mathrm{T}}\right)^{-1}\mathbf{C}_d & \mathbf{I} \end{bmatrix}, \begin{bmatrix} \mathbf{I} & -\upsilon\mathbf{B}_d\left(\upsilon^2\mathbf{I} - \mathbf{D}_d\mathbf{D}_d^{\mathrm{T}}\right)^{-1}\mathbf{B}_d^{\mathrm{T}} \\ \mathbf{0} & \left[\mathbf{A}_d + \mathbf{B}_d\mathbf{D}_d^{\mathrm{T}}\left(\upsilon^2\mathbf{I} - \mathbf{D}_d\mathbf{D}_d^{\mathrm{T}}\right)^{-1}\mathbf{C}_d\right]^{\mathrm{T}} \end{bmatrix}\right) \tag{1.71}$$

□

It should be noted that Lemma 1.2 does not provide a closed-form expression for the H_∞. Typically, a bisection algorithm based on Lemma 1.2 is used to search an approximation value of the H_∞ norm with the following steps (continuous time case):

(1) Select an upper bound γ_{u} and a lower bound γ_l such that $\gamma_l \leq \|\mathbf{G}\|_\infty \leq \gamma_{\mathrm{u}}$.
(2) If $(\gamma_{\mathrm{u}} - \gamma_{\mathrm{l}})/\gamma_{\mathrm{l}} \leq \varepsilon$, ε a predefined tolerance, stop, and we have $\|\mathbf{G}\|_\infty \approx (\gamma_{\mathrm{u}} + \gamma_{\mathrm{l}})/2$. Otherwise, go to next step:
(3) Set $\gamma = (\gamma_{\mathrm{l}} + \gamma_{\mathrm{u}})/2$.
(4) Test if $\|\mathbf{G}\|_\infty < \gamma$ by calculating the eigenvalues of $\mathbf{H}$ in (1.70) for a given γ.
(5) If $\mathbf{H}$ has an eigenvalue on the imaginary axis, then set $\gamma_{\mathrm{l}} = \gamma$; otherwise, set $\gamma_{\mathrm{u}} = \gamma$, then go back to step 2.

The same procedure can also be applied to discrete time systems.

Optimal Performance Consideration

The design of a controller for the closed-loop control system should in general satisfy three basic requirements:

(1) *Closed-loop Stability*
(2) *Performance:* The controller is typically designed so that the resulting closed-loop system meets certain performance specifications, such as an upper bound on the H_2 norm or H_∞ norm of the closed-loop system, the steady-state error.
(3) *Robustness:* The closed-loop controller has to remain stable and yield acceptable performance despite uncertainties in the description of the plant model.

The control design requirements are usually best expressed in the form of an optimization criterion subject to some constraints, for example, terminal constraints or practical constraints on plant variables. Control design optimization criteria were initially based on the idea of linear quadratic Gaussian (LQG) control which was later generalized to the idea of H_2 controller design. The development of small-gain theory laid the foundations of robust H_∞ control. For linear systems with suitable optimization criteria, such as LQG, H_2 and H_∞ performances, the solution to the optimization problem is readily found by solving Riccati equations.

Linear quadratic Gaussian (LQG) control problem is one of the most fundamental optimal control problems in the control theory. It concerns linear systems disturbed by additive white Gaussian noise, having incomplete state information (i.e., not all the state variables are measured and available for feedback) and undergoing control subject to quadratic costs. The LQG controller is simply the combination of a linear quadratic regulator (LQR) with a Kalman filter, that is, a linear quadratic estimator (LQE). The separation principle guarantees that these can be designed and computed independently. The LQG method has been successfully applied in astronomical adaptive optics systems to cancel the aberrations caused by stochastic atmosphere turbulence. However, the LQG problem can be cast into a standard H_2 optimal control problem presented in the next subsection. For simplicity, let's consider the LQR problem for continuous time systems. Given a plant model

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} + \tilde{\mathbf{w}} \tag{1.72}$$

where $\tilde{\mathbf{w}}$ is a zero mean white noise, find a control input $\mathbf{u}$ that minimizes the cost functional

$$J(\mathbf{u}) = \lim_{t\to\infty} \mathbf{E}\left(\mathbf{x}^{\mathrm{T}}(t)\mathbf{Q}\mathbf{x}(t) + \mathbf{u}^{\mathrm{T}}(t)\mathbf{R}\mathbf{u}(t)\right) \tag{1.73}$$

where $\mathbf{E}$ denoted the expected value. Assume $\mathbf{Q} \geq 0, \mathbf{R} \geq 0$, $(\mathbf{A}, \mathbf{B})$ stabilizable, and $(\mathbf{Q}, \mathbf{A})$ detectable. Then we have the well-known established result to solve (1.73) with the unique optimal control $\mathbf{u} = \mathbf{F}\mathbf{x}$, where $\mathbf{F} = -\mathbf{R}^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{X}$ and $\mathbf{X}$ satisfies the following Riccati equation:

$$\mathbf{A}^{\mathrm{T}}\mathbf{X} + \mathbf{X}\mathbf{A} - \mathbf{X}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{X} + \mathbf{Q} = 0 \tag{1.74}$$

The achieved minimization of the cost function J is $J_{\min} = trace\,(\mathbf{X})$.

The LQG problem assumes that the system state variable $\mathbf{x}$ is not available, then the overall optimal controller needs to combine the linear quadratic regulator (LQR) with a linear quadratic estimator (LQE). The LQR and LQE controller can be designed separately based on the well-known separation principle.

Standard Feedback System Configuration

The main objective behind the design of a controller for a feedback control system is to stabilize the closed-loop system and at the same time meet certain performance

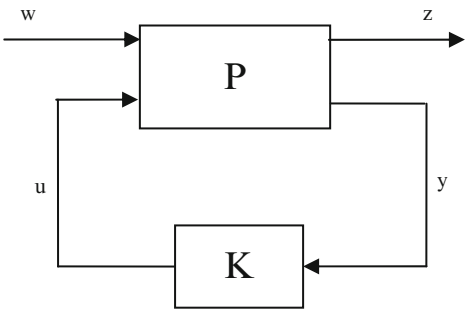


Fig. 1.3 General feedback system configuration

specification such as bounds on the H_2 or H_∞ norm of the closed-loop system. A standard feedback system configuration used in formulating controller synthesis problems is the linear-fractional transformation (LFT) configuration shown in Fig. 1.3, where **P** is the generalized plant with two sets of inputs: the exogenous inputs **w**, which include disturbances and commands, and control inputs **u**. The plant **P** also has two sets of outputs: the measured outputs **y** and the performance outputs **z**. A control problem in this setup is to design a feedback controller **K** such that the closed-loop system is stable in some appropriate sense and satisfies some performance requirements. For example, the overall control objective can be searching a controller **K** to minimize the H_∞ norm of the closed-loop transfer function from **w** to **z**. The plant transfer function matrix **P** in Fig. 1.3 can be partitioned as

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix}$$

The performance variable can then be expressed as

$$\mathbf{z} = \left[\mathbf{P}_{11} + \mathbf{P}_{12}\mathbf{K}(\mathbf{I} - \mathbf{P}_{22}\mathbf{K})^{-1}\mathbf{P}_{21}\right]\mathbf{w} \tag{1.75}$$

Let

$$F(\mathbf{P}, \mathbf{K}) = \mathbf{P}_{11} + \mathbf{P}_{12}\mathbf{K}(\mathbf{I} - \mathbf{P}_{22}\mathbf{K})^{-1}\mathbf{P}_{21} \tag{1.76}$$

F $(\mathbf{P}, \mathbf{K})$ is called a *linear-fractional transformation(LFT)* of **P** and **K**.

The example in Fig. 1.4 shows a standard single-loop feedback system with a plant **G**, controller **K**, and feedback sensor **F**. There are two exogenous inputs: a disturbance **d** and a noise signal **n** corrupting the plant output. Also shown are two filters, $\mathbf{W}_1$ and $\mathbf{W}_2$, generating a filtered control signal, $\tilde{\mathbf{u}}$, and a filtered plant output, $\tilde{\mathbf{y}}$. It is desired to set the closed-loop system performance specification in terms of the signals $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{y}}$. To convert the control system in Fig. 1.4 to the standard setup, define

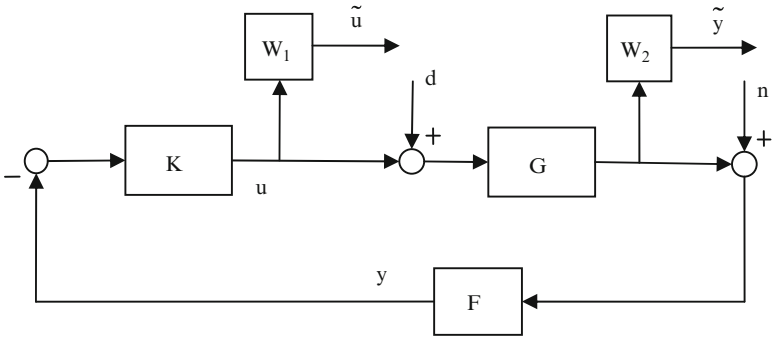


Fig. 1.4 A feedback system with weighted performance variable

$$\mathbf{w} = \begin{bmatrix} \mathbf{d} \\ \mathbf{n} \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} \tilde{\mathbf{y}} \\ \tilde{\mathbf{u}} \end{bmatrix}$$

Then we have

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix}$$

$$= \left[\begin{array}{cc|c} \begin{bmatrix} \mathbf{W_2G} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} & & \begin{bmatrix} \mathbf{W_2G} \\ \mathbf{W_1} \end{bmatrix} \\ \hline \begin{bmatrix} -\mathbf{FG} & -\mathbf{F} \end{bmatrix} & & -\mathbf{FG} \end{array}\right]$$

where the input and output of $\mathbf{P}$ are partitioned as $\begin{bmatrix} \mathbf{w} \\ \mathbf{u} \end{bmatrix}$ and $\begin{bmatrix} \mathbf{z} \\ \mathbf{y} \end{bmatrix}$. Let the state-space presentation of $\mathbf{P}$ be given as

$$\mathrm{P} := \left(\begin{array}{c|cc} \mathbf{A} & \mathbf{B}_1 & \mathbf{B}_2 \\ \hline \mathbf{C}_1 & \mathbf{D}_{11} & \mathbf{D}_{12} \\ \mathbf{C}_2 & \mathbf{D}_{21} & \mathbf{D}_{22} \end{array}\right)$$

If $\mathbf{D}_{22} = 0$, then the plant transfer matrix from $\mathbf{u}$ to $\mathbf{y}$ is strictly proper.

As an illustration of formulating a controller design problem using the standard feedback system configuration, consider the LQG controller design problem for the system Σ given by

$$\Sigma : \begin{cases} \dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} + \mathbf{w}_x \\ \mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{w}_y \end{cases} \tag{1.77}$$

where the process noise $\mathbf{w}_x$ and measurement noise $\mathbf{w}_y$ are independent and have constant power spectral density matrices $\tilde{\mathbf{W}}_x$ and $\tilde{\mathbf{W}}_y$, respectively. The LQG cost

function is the sum of the steady-state mean square weighted state $\mathbf{x}$ and the steady-state mean square weighted actuator signal $\mathbf{u}$ as

$$J = \lim_{t\to\infty} \mathbf{E}\left(\mathbf{x}(t)^{\mathrm{T}}\mathbf{Q}\mathbf{x}(t) + \mathbf{u}(t)^{\mathrm{T}}\mathbf{R}\mathbf{u}(t)\right) \tag{1.78}$$

where $\mathbf{Q}$ and $\mathbf{R}$ are positive semi-definite weight matrices. We can express this cost in the standard framework by forming the regulated output signal

$$\mathbf{z} = \begin{bmatrix} \mathbf{R}^{\frac{1}{2}}\mathbf{u} \\ \mathbf{Q}^{\frac{1}{2}}\mathbf{x} \end{bmatrix}$$

The exogenous input consists of the process and measurement noises, which can be represented as

$$\begin{bmatrix} \mathbf{w}_x \\ \mathbf{w}_y \end{bmatrix} = \begin{bmatrix} \left(\tilde{\mathbf{W}}_x\right)^{\frac{1}{2}} \\ \left(\tilde{\mathbf{W}}_y\right)^{\frac{1}{2}} \end{bmatrix} \tilde{\mathbf{w}}$$

where $\tilde{\mathbf{w}}$ a white noise signal. The state-space description of the augmented plant $\mathbf{P}$ for the LQG problem is thus

$$\mathbf{A} = \mathbf{A}, \quad \mathbf{B}_1 = \begin{bmatrix} \left(\tilde{\mathbf{W}}_x\right)^{\frac{1}{2}} & \mathbf{0} \end{bmatrix}, \quad \mathbf{B}_2 = \mathbf{B}$$

$$\mathbf{C}_1 = \begin{bmatrix} \mathbf{0} \\ \mathbf{Q}^{\frac{1}{2}} \end{bmatrix}, \quad \mathbf{C}_2 = \mathbf{C}, \quad \mathbf{D}_{11} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

$$\mathbf{D}_{12} = \begin{bmatrix} \mathbf{R}^{\frac{1}{2}} \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{D}_{21} = \begin{bmatrix} \mathbf{0} & \left(\tilde{\mathbf{W}}_y\right)^{\frac{1}{2}} \end{bmatrix}, \quad \mathbf{D}_{22} = \mathbf{0}$$

The optimal controller $\mathbf{K}$ takes the form of (Boyd and Barratt 1991)

$$\mathbf{K} : \begin{cases} \dot{\hat{\mathbf{x}}} = (\mathbf{A} - \mathbf{B}\mathbf{L}_{\mathrm{f}} - \mathbf{K}_{\mathrm{f}}\mathbf{C})\,\hat{\mathbf{x}} + \mathbf{K}_{\mathrm{f}}\mathbf{y} \\ \mathbf{u} = -\mathbf{L}_{\mathrm{f}}\hat{\mathbf{x}} \end{cases}$$

where $\mathbf{K}_{\mathrm{f}}$ and $\mathbf{L}_{\mathrm{f}}$ are the Kalman gain and state feedback gain to be designed, respectively.

Robustness Specifications

A control system is *robust* if it remains stable and meets certain performance objectives in the presence of uncertainties in the plant model. The H_∞ optimization approach and its related approaches have been shown to be effective and efficient robust design methods for linear, time-invariant control systems. In the following,

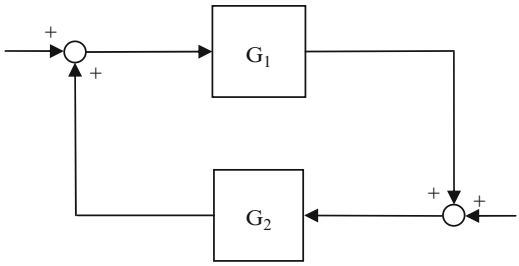


Fig. 1.5 Feedback configuration

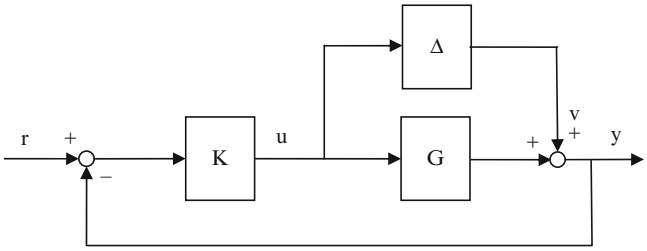


Fig. 1.6 Feedback system with additive perturbation in the plant

the *small-gain theorem*, which plays an important role in the H_∞ robust controller design methods, will be introduced. Consider the feedback configuration in Fig. 1.5, where $\mathbf{G}_1(s)$ and $\mathbf{G}_2(s)$ are the transfer function matrices of linear time-invariant systems. We then have the following theorem.

Theorem 1.5 *(Zhou et al. 1995) If* $\mathbf{G}_1(s)$ *and* $\mathbf{G}_2(s)$ *are stable, then the closed-loop system is stable if and only if*

$$\|\mathbf{G}_1\mathbf{G}_2\|_\infty < 1$$

and

$$\|\mathbf{G}_2\mathbf{G}_1\|_\infty < 1$$

□

A closed-loop system of the plant$\mathbf{G}$ and controller $\mathbf{K}$ is robustly stable if it remains stable for some level of uncertainty in the plant model. Consider, for example, the case of an additive perturbation as depicted in Fig. 1.6, where $\mathbf{\Delta}(s)$ is the perturbation, presented using a stable unknown matrix transfer function. Then the transfer function from the signal $\mathbf{v}$ to $\mathbf{u}$ is $\mathbf{T}_{uv} = -\mathbf{K}(\mathbf{I} + \mathbf{GK})^{-1}$. It is obvious that the controller $\mathbf{K}$ should stabilize the nominal plant $\mathbf{G}$ first, which means $\mathbf{T}_{uv}$ is stable. Then, based on the small-gain theorem, we have the following robustness result for the closed-loop system.

Theorem 1.6 *For stable* $\mathbf{\Delta}(s)$*, the closed-loop system is robustly stable if* $\mathbf{K}(s)$ *stabilizes the nominal plant and the following holds:*

$$\left\|\mathbf{\Delta}\boldsymbol{K}(\mathbf{I} + \mathbf{GK})^{-1}\right\|_\infty < 1 \tag{1.79}$$

and

$$\left\| \mathbf{K}(\mathbf{I} + \mathbf{GK})^{-1}\boldsymbol{\Delta} \right\|_{\infty} < 1 \tag{1.80}$$

□

Based on (1.79) and (1.80), the following sufficient condition guarantees the robust stability of the closed-loop system:

$$\left\| \mathbf{K}(\mathbf{I} + \mathbf{GK})^{-1} \right\|_{\infty} < \frac{1}{\|\boldsymbol{\Delta}\|_{\infty}} \tag{1.81}$$

In many cases, we may have a priori knowledge of an upper bound on the norm of the perturbation, such as, for example, an upper bound on the maximum singular value of the perturbation

$$\bar{\sigma}\left(\Delta\left(j\omega\right)\right) \leq \bar{\sigma}\left(\mathbf{W}_{\Delta}\left(j\omega\right)\right) \text{ for all } \omega \in \mathbb{R}$$

where $\bar{\sigma}(\bullet)$ denoted the maximum singular value and $\mathbf{W}_{\Delta}(s)$ is a known weight function. Consequently, the robust stabilization condition in this case is satisfied if

$$\left\| \mathbf{W}_{\Delta}\mathbf{K}(\mathbf{I} + \mathbf{GK})^{-1} \right\|_{\infty} < 1 \tag{1.82}$$

The reader can refer to Zhou et al. (1995) and Skogestad and Postlethwaite (2005) for more detailed discussions on robustness issues.

1.2.2.3 Linear Matrix Inequalities

Linear matrix inequalities (LMIs) and LMI techniques have emerged as powerful design tools in control engineering, where many optimal control problems can be stated in terms of linear matrix inequalities. Linear matrix inequalities are matrix inequalities which are linear (or affine) in a set of matrix variables. The equivalence between the Riccati equation and LMI formulations of the control problem was found at an early point in the 1970s. However, only during the past 10–15 years has the development of sophisticated numerical routines, that is, semi-definite programming, made it possible to solve LMIs in a reasonably efficient manner. From a control engineering perspective, one of the main attractions of LMIs is that they can be used to solve problems which involve several matrix variables, and, moreover, different structures can be imposed on these matrix variables. Another attractive feature of LMI methods is that they are flexible, so it is often relatively straightforward to pose a variety of problems as LMI problems amenable to LMI methods. Furthermore, in many cases the use of LMIs can remove restrictions associated with conventional methods and aid their extension to more general scenarios. While most problems with multiple constraints or objectives lack analytical solutions in terms of matrix equations, they often remain tractable in the LMI framework. Therefore, LMI methods can be applied in instances where

conventional methods either fail or struggle to find a solution. In actual fact, the flexibility of LMIs has created a much wider scope for controller design. They allow the efficient consideration of H_2 and H_∞ performance constraints and of robustness requirements. The control approaches developed in this book are heavily based on the LMI techniques; therefore, a brief LMI background is introduced in the following.

Fundamental LMI Properties

A notion central to the understanding of matrix inequalities is definiteness. In particular, a matrix $\mathbf{P}$ is defined to be positive definite if

$$\xi^{\mathrm{T}}\mathbf{P}\xi > 0,\ \forall \xi \neq 0$$

Likewise, $\mathbf{P}$ is said to be positive semi-definite if

$$\xi^{\mathrm{T}}\mathbf{P}\xi \geq 0,\ \forall \xi \neq 0$$

It is common practice to write $\mathbf{P} > 0$ ($\mathbf{P} \geq 0$) to indicate that it is positive (semi) definite. In particular, the positive-definite matrices $\mathbf{P}$ is square and symmetric, that is, $\mathbf{P} = \mathbf{P}^{\mathrm{T}}$. Because all eigenvalues of a real symmetric matrix are real, $\mathbf{P} > 0$ is equivalent to saying that all eigenvalues λ ($\mathbf{P}$) are positive real or that the minimum eigenvalue $\lambda_{\min}$ ($\mathbf{P}$) > 0. A positive semi-definite matrix shares the first attribute, but the last is relaxed to the requirement that all of its eigenvalues are positive real or zero. A matrix $\mathbf{Q} = -\mathbf{P}$ is said to be negative (semi) definite if $\mathbf{P}$ is positive (semi) definite. To indicate negative (semi) definiteness, we write $\mathbf{Q} < 0$ ($\mathbf{Q} \leq 0$). In fact, once the notation $\mathbf{P} > 0$ ($\mathbf{P} \geq 0$) or $\mathbf{Q} < 0$ ($\mathbf{Q} \leq 0$) is used, we implicitly require $\mathbf{P}$ and $\mathbf{Q}$ to be symmetric. A linear matrix inequality is an expression of the form

$$F(\mathbf{x}) := \mathbf{F}_0 + x_1\mathbf{F}_1 + \cdots + x_n\mathbf{F}_n < 0 \tag{1.83}$$

where $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ is a vector of n real numbers called the *decision variables*. $\mathbf{F}_0, \ldots, \mathbf{F}_n$ are real symmetric matrices, that is, $\mathbf{F}_j = \mathbf{F}_j^{\mathrm{T}}$, for $j = 0, \ldots n$. The basic LMI problem—the feasibility problem—is to find $\mathbf{x}$ such that inequality (1.83) holds.

Note that $F(\mathbf{x}) > 0$ describes an affine relationship in terms of the vector $\mathbf{x}$. Normally, the variable $\mathbf{x}$ to be determined is a vector obtained by stacking up the columns of a number of matrices. That is,

$$F(\mathbf{x}) = F(\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_m) \tag{1.84}$$

where $\mathbf{X}_i \in \mathbb{R}^{q_i \times p_i}$, $i = 1, \ldots, m$, are matrices to be determined, $\sum_{i=1}^{m} q_i \times p_i = n$, and the columns of all the $\mathbf{X}_i$ matrices are stacked up to form a single vector variable $\mathbf{x}$. When $F(\mathbf{x})$ is an affine function of the matrices $\mathbf{X}_i \in \mathbb{R}^{q_i \times p_i}$, $i = 1, ..., m$, (1.84) takes the form of

$$F(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m) = \mathbf{F}_0 + \mathbf{G}_1\mathbf{X}_1\mathbf{H}_1 + \cdots + \mathbf{G}_m\mathbf{X}_m\mathbf{H}_m$$
$$= \mathbf{F}_0 + \sum\nolimits_{i=1}^{m} \mathbf{G}_i\mathbf{X}_i\mathbf{H}_i$$

where $\mathbf{F}_0$, $\mathbf{G}_i$, $\mathbf{H}_i$ are given matrices and $\mathbf{X}_i$, $i = 1, \ldots, m$, are the matrix variables to be determined. It is worth noting that the linear matrix inequality (1.83) defines a convex constraint on $\mathbf{x}$, which means the set

$$\Im := \{\mathbf{x} | \, F(\mathbf{x}) < 0\}$$

of solutions of the LMI $F(\mathbf{x}) < 0$ is convex, that is, if $\mathbf{x}_1, \mathbf{x}_2 \in \Im$ and $\alpha \in (0, 1)$, then we have

$$F(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) = \alpha F(\mathbf{x}_1) + (1-\alpha) F(\mathbf{x}_2) < 0$$

The inequality follows from the fact that F is affine and that $\alpha > 0$ and $(1-\alpha) > 0$. The convex problem can be solved with efficient numerical tools. Some of the numerical tools are based on the ellipsoid algorithm which is simple, numerically robust, and easy to implement but may be slow for larger optimization problems. A major breakthrough in convex optimization lies in the introduction of interior point methods. Interior point methods (also referred to as barrier methods) are a certain class of algorithms to solve linear and nonlinear convex optimization problems. These algorithms have been inspired by Karmarkar's algorithm, developed by Narendra Karmarkar in 1984 for linear programming, and are significantly faster than classical convex optimization algorithms. The LMI solvers based on interior point algorithm can be found in many available commercial and noncommercial software products such as the Matlab LMI toolbox.

Many control problems can be solved using properly formulated LMIs. For example, consider the Lyapunov matrix inequality, that is,

$$\mathbf{A}^{\mathrm{T}}\mathbf{P} + \mathbf{P}\mathbf{A} < 0, \quad \mathbf{P} > 0$$

where $\mathbf{P} \in \mathbb{R}^{n\times n}$ is variable and $\mathbf{A} \in \mathbb{R}^{n\times n}$ is given. Lyapunov theory showed that this LMI is feasible if and only if the matrix $\mathbf{A}$ is stable, that is, all trajectories of the system $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ converge to zero as $t \to \infty$ or equivalently, all eigenvalues of $\mathbf{A}$ must have negative real part. This particular problem can be easily solved using numerical algorithms; however, it can also be solved in an analytic way. To solve this LMI problem, we pick any $\mathbf{Q} > 0$ and solve the Lyapunov equation $\mathbf{A}^{\mathrm{T}}\mathbf{P} + \mathbf{P}\mathbf{A} = -\mathbf{Q}$, which is nothing but a set of $n(n+1)/2$ linear equations for the $n(n+1)/2$ scalar variables in $\mathbf{P}$. This set of linear equations will be solvable and result in $\mathbf{P} > 0$ if and only if the LMI is feasible. In fact, this procedure not only finds a solution when the LMI is feasible but also parameterizes all solutions as $\mathbf{Q}$ varies over the positive-definite cone.

Systems of LMIs

In general, we are frequently faced with LMI constraints of the form

$$F_1(\mathbf{x}) < 0, \ldots, F_p(\mathbf{x}) < 0 \tag{1.85}$$

which is called a system of linear matrix inequalities. It is easily seen that the intersection of the feasible sets of each of the inequalities (1.85) is convex, and this set can be represented as the feasibility set of another large LMI. Indeed, $F_1(\mathbf{x}) < 0, \ldots, F_p(\mathbf{x}) < 0$ if and only if

$$F(\mathbf{x}) := \begin{bmatrix} F_1(\mathbf{x}) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & F_2(\mathbf{x}) & \mathbf{0} & \vdots \\ \vdots & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & F_p(\mathbf{x}) \end{bmatrix} < 0 \tag{1.86}$$

$F(\mathbf{x})$ is also symmetric for any $\mathbf{x}$, and the set of eigenvalues of $F(\mathbf{x})$ is simply the union of the eigenvalues of $F_1(\mathbf{x}), \ldots, F_p(\mathbf{x})$. Any $\mathbf{x}$ satisfies $F(\mathbf{x}) < 0$ also satisfies the system of LMIs (1.85) and vice versa. Therefore, multiple LMI constraints can always be converted to a single LMI constraint.

Congruence Transformation

For a given positive-definite matrix $\mathbf{P} > 0 \in \mathbb{R}^{n\times n}$, and given a real matrix $\mathbf{T} \in \mathbb{R}^{n\times n}$ such that $rank(\mathbf{T}) = n$, the following inequality holds (Herrmann et al. 2007):

$$\mathbf{TPT}^{\mathrm{T}} > 0 \tag{1.87}$$

In other words, definiteness of a matrix is invariant under pre- and post-multiplication by a full rank real matrix and its transpose, respectively. The process of transforming $\mathbf{P} > 0$ into Eq. 1.87 using a real full rank matrix is called a *congruence transformation*. The congruence transformation is very useful for removing bilinear terms in matrix inequalities and is often used, in conjunction with a change of variables, to transform a bilinear matrix inequality into an LMI. Often, $\mathbf{T}$ is chosen to have a diagonal structure. For example, consider a bilinear matrix inequality

$$\mathbf{Q} = \begin{bmatrix} \mathbf{A}^{\mathrm{T}}\mathbf{P} + \mathbf{PA} & \mathbf{PBF} + \mathbf{C}^{\mathrm{T}}\mathbf{V} \\ \mathbf{F}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{P} + \mathbf{VC} & -\mathbf{V} \end{bmatrix} < 0 \tag{1.88}$$

where the matrices $\mathbf{P} \in \mathbb{R}^{n\times n} > 0$, $\mathbf{V} \in \mathbb{R}^{m\times m} > 0$, and $\mathbf{F} \in \mathbb{R}^{p\times m}$ are the unknown matrix variables and the remaining matrices are the known constants with proper

dimensions. Notice that this inequality is bilinear in the variables $\mathbf{P}$ and $\mathbf{F}$ which occur in the (1,2) and (2,1) elements of the matrix $\mathbf{Q} \in \mathbb{R}^{(n+m)\times(n+m)}$. If we choose a nonsingular matrix

$$\mathbf{T} = \begin{bmatrix} \mathbf{P}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}^{-1} \end{bmatrix}$$

which is full rank with $rank\ (\mathbf{T}) = n + m$, then calculating $\mathbf{TQT}^{\mathrm{T}}$ gives

$$\mathbf{TQT}^{\mathrm{T}} = \begin{bmatrix} \mathbf{P}^{-1}\mathbf{A}^{\mathrm{T}} + \mathbf{AP}^{-1} & \mathbf{BFV}^{-1} + \mathbf{P}^{-1}\mathbf{C}^{\mathrm{T}} \\ \mathbf{V}^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}} + \mathbf{CP}^{-1} & -\mathbf{V}^{-1} \end{bmatrix} < 0$$

Define the new variables $\mathbf{X} = \mathbf{P}^{-1}$, $\mathbf{U} = \mathbf{V}^{-1}$, and $\mathbf{L} = \mathbf{FV}^{-1}$, then the bilinear matrix inequality (1.88) is transformed into a linear matrix inequality

$$\mathbf{TQT}^{\mathrm{T}} = \begin{bmatrix} \mathbf{XA}^{\mathrm{T}} + \mathbf{AX} & \mathbf{BL} + \mathbf{XC}^{\mathrm{T}} \\ \mathbf{L}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}} + \mathbf{CX} & -\mathbf{U} \end{bmatrix} < 0 \tag{1.89}$$

with unknown matrices $\mathbf{X} \in \mathbb{R}^{n\times n} > 0$, $\mathbf{U} \in \mathbb{R}^{m\times m} > 0$, and $\mathbf{L} \in \mathbb{R}^{p\times m}$. Notice that the original variables can be recovered by inverting $\mathbf{X}$ and $\mathbf{U}$ and computing $\mathbf{F} = \mathbf{LU}^{-1}$.

Schur Complement

The main use of the Schur complement is to transform quadratic matrix inequalities into linear matrix inequalities, as indicated in the following Theorem.

Theorem 1.7 *(Boyd et al. 1994) Let $F\ (\mathbf{x})$ be a symmetric and affine function in $\boldsymbol{x}$, which is partitioned according to*

$$F\ (\mathbf{x}) = \begin{bmatrix} F_{11}\ (\mathbf{x}) & F_{12}\ (\mathbf{x}) \\ F_{12}^{T}\ (\mathbf{x}) & F_{22}\ (\mathbf{x}) \end{bmatrix}$$

where $F_{11}\ (\mathbf{x})$ is square. Then $F\ (\mathbf{x}) < 0$ if and only if

$$\begin{cases} F_{11}\ (\mathbf{x}) < 0 \\ F_{22}\ (\mathbf{x}) - F_{12}^{T}\ (\mathbf{x})\ [F_{11}\ (\mathbf{x})]^{-1} F_{12}\ (\mathbf{x}) < 0 \end{cases} \tag{1.90}$$

if and only if

$$\begin{cases} F_{22}\ (\mathbf{x}) < 0 \\ F_{11}\ (\mathbf{x}) - F_{12}\ (\mathbf{x})\ [F_{22}\ (\mathbf{x})]^{-1} F_{12}^{\mathrm{T}}\ (\mathbf{x}) < 0 \end{cases} \tag{1.91}$$

□

The Schur complement result can also be generalized to nonstrict inequalities. For symmetric matrix $F(\mathbf{x})$, the condition $F(\mathbf{x}) \leq 0$ is equivalent to

$$F_{22}(\mathbf{x}) \leq 0$$

$$F_{11}(\mathbf{x}) - F_{12}(\mathbf{x}) F_{22}^{\dagger}(\mathbf{x}) F_{12}^{\mathrm{T}}(\mathbf{x}) \leq 0$$

$$F_{12}(\mathbf{x}) \left(\mathbf{I} - F_{22}(\mathbf{x}) F_{22}^{\dagger}(\mathbf{x})\right) = 0$$

where $F_{22}^{\dagger}(\mathbf{x})$ denotes the Moore–Penrose inverse of $F_{22}(\mathbf{x})$.

The second inequalities in (1.90) and (1.91) are nonlinear constraints in $\mathbf{x}$. Using this result, it follows that nonlinear matrix inequalities of the form (1.90) and (1.91) can be converted to linear matrix inequalities. For example, based on the Schur complement, the (maximum singular value) matrix norm constraint $\|Z(\mathbf{x})\| < 1$, where $Z(\mathbf{x}) \in \mathbb{R}^{p \times q}$ and depends affinely on $\mathbf{x}$, can be represented as the LMI

$$\begin{bmatrix} \mathbf{I} & Z(\mathbf{x}) \\ Z(\mathbf{x})^{\mathrm{T}} & \mathbf{I} \end{bmatrix} > 0$$

since $\|Z(\mathbf{x})\| < 1$ is equivalent to $\mathbf{I} - Z(\mathbf{x}) Z(\mathbf{x})^{\mathrm{T}} > 0$. Similarly, the constraint

$$c(\mathbf{x})^{\mathrm{T}} P(\mathbf{x})^{-1} c(\mathbf{x}) < 1, \quad P(\mathbf{x}) > 0$$

where $c(\mathbf{x}) \in \mathbb{R}^n$ and $P(\mathbf{x}) \in \mathbb{R}^{n \times n}$ depend affinely on $\mathbf{x}$, can be expressed as the LMI

$$\begin{bmatrix} P(\mathbf{x}) & c(\mathbf{x}) \\ c(\mathbf{x})^{\mathrm{T}} & 1 \end{bmatrix} > 0$$

Generalized Eigenvalue Problems

The generalized eigenvalue problem (GEVP) is to minimize the maximum generalized eigenvalue of a pair of matrices that depend affinely on a variable, subject to an LMI constraint. The general form of a GEVP is

$$\begin{aligned} &\text{minimize } \lambda \\ &\text{subject to } F_1(\mathbf{x}) + \lambda F_2(\mathbf{x}) > 0 \\ &\qquad\qquad F_2(\mathbf{x}) > 0 \\ &\qquad\qquad F_3(\mathbf{x}) > 0 \end{aligned}$$

where F_1, F_2, and F_3 are symmetric matrices that are affine functions of $\mathbf{x}$. We can express this as

$$\begin{aligned} \text{minimize} \quad & \lambda_{\max}\left(F_1(\mathbf{x}),\ F_2(\mathbf{x})\right) \\ \text{subject to} \quad & F_2(\mathbf{x}) > 0 \\ & F_3(\mathbf{x}) > 0 \end{aligned}$$

where $\lambda_{\max}(\mathbf{X}, \mathbf{Y})$ denotes the largest generalized eigenvalue of the pencil $\lambda\mathbf{Y} + \mathbf{X}$ with $\mathbf{Y} > 0$. This GEVP is a quasiconvex optimization problem since the constraint is convex and the objective, $\lambda_{\max}\left(F_1(\mathbf{x}),\ F_2(\mathbf{x})\right)$, is quasiconvex. Note that when the matrices are all diagonal and $F_1(\mathbf{x})$ and $F_2(\mathbf{x})$ are scalar, this problem reduces to the general linear-fractional programming problem, that is, minimizing a linear-fractional function subject to a set of linear inequalities. In addition, many nonlinear quasiconvex optimization problems can be represented in the form of GEVPs with appropriately defined F_1, F_2, and F_3. As an example of a GEVP, consider the problem

$$\begin{aligned} & \text{maximize} \quad \alpha \\ & \text{subject to} \quad \mathbf{A}^{\mathrm{T}}\mathbf{P} + \mathbf{A}\mathbf{P} + \alpha\mathbf{P} < 0, \mathbf{P} > 0 \end{aligned}$$

where the matrix $\mathbf{A}$ is given and the optimization variables are the symmetric matrix $\mathbf{P}$ and the scalar α. (This problem arises in Sect. 7.3.)

1.3 Summary

In this chapter, an introduction to adaptive optics systems and their applications is presented. Special focus is placed on the emerging area of ophthalmic imaging where advanced adaptive optics system technology can be used to cancel the high-order aberrations in the eye and obtain high-resolution retinal images of the retina. The latter allow the early detection of ocular diseases. The shortcomings of existing adaptive optics systems are discussed. Magnetic fluid deformable mirrors are introduced, and the advantages they offer over conventional solid mirrors are outlined. These advantages include low cost, large stroke of the wavefront corrector deformable surface, the extremely smooth surface of the mirror, and the ease of fabrication. The major contributions of this book, which is aimed at bridging the critical gap between the concept of an MFDM and its application in adaptive optics systems, for example, ophthalmic adaptive optics imaging systems, are summarized. Background information related to MFDM system modeling is presented and consists mainly of mathematical tools used in the derivation of MFDM analytic model. A review of elementary concepts in linear control systems theory as well as advanced modern multivariable control techniques used in this book, including

optimal control and robust control techniques, is briefly summarized. The powerful tool of linear matrix inequalities for approaching optimal control problems is also covered briefly.

References

Babcock HW (1953) The possibility of compensating astronomical seeing. Publ Astron Soc Pac 65:229–236

Basmadjian D, Farnood R (2007) The art of modeling in science and engineering with mathematica. Chapman and Hall/CRC, Boca Raton

Bennett TJ, Barry CJ (2009) Ophthalmic imaging today: an ophthalmic photographers viewpoint a review. Clin Experiment Ophthalmol 37:2–13

Borra EF (2009) Liquid mirrors in engineering, Optics & Photonics News, pp 14–17, September 2009

Borra EF, Brousseau D, Vincent A (2006) Large magnetic liquid mirrors. Astron Astrophys 446(1):389–393

Borra EF, Brousseau D, Cliche M, Parent J (2008) Aberration correction with a magnetic liquid active mirror. Mon Not R Astron Soc 391(4):1925–1930

Boyd S, Barratt C (1991) Linear controller design: limits of performance. Prentice Hall, Englewood Cliffs

Boyd S, Ghaoui LE, Feron E, Balakrishnan V (1994) Linear matrix inequalities in system and control theory. SIAM, Philadelphia

Brousseau D, Borra EF, Ruel HJ, Parent J (2006) A magnetic liquid deformable mirror for high stroke and low order axially symmetrical aberrations. Opt Express 14:11486–11493

Brousseau D, Borra EF, Thibault S (2007) Wavefront correction with a 37–actuator ferrofluid deformable mirror. Opt Express 15:18190–18199

Brousseau D, Borra EF, Rochette M, Landry DB (2010) Linearization of the response of a 91-actuator magnetic liquid deformable mirror. Opt Express 18(8):8239–8250

Brousseau D, Drapeau J, Piché M, Borra EF (2011) Generation of Bessel beams using a magnetic liquid deformable mirror. Appl Opt 50:4005–4010

Chen T, Francis B (1996) Optimal sampled data control systems. Springer, London

Doble N, Miller DT (2006) Wavefront correctors for vision science. In: Porter J, Queener H, Lin J, Thorn K, Awwal A (eds) Adaptive optics for vision science: principles, practices, design and applications. Wiley, Hoboken

Doble N, Williams DR (2004) The applications of MEMS technology for AO in vision science. IEEE J Sel Top Quantum Electron 10(3):629–635

Dorf RC, Bishop RH (2008) Modern control systems. Prentice Hall, Englewood Cliffs

Godara P, Dubis AM, Roorda A, Duncan JL, Carroll J (2010) Adaptive optics retinal imaging: emerging clinical applications. Optom Vis Sci 87(12):930–941

Hardy JW (1998) Adaptive optics for astronomical telescopes. Oxford University Press, New York

Hart M (2010) Recent advances in astronomical adaptive optics. Appl Opt 49(16):D17–D29

Herrmann G, Matthew C, Turner MC, Postlethwaite I (2007) Linear matrix inequalities in control. In: Mathematical methods for robust and nonlinear control, vol 367, LNCIS. Springer, Berlin, pp 123–142

Iqbal A, Ben Amara F (2007) Modeling of a magnetic fluid deformable mirror for retinal imaging adaptive optics systems. Int J Optomechatron 1(2):180–208

Iqbal A, Ben Amara F (2008) Modeling and experimental evaluation of a circular magnetic-fluid deformable mirror. Int J Optomechatron 2(2):126–143

Iqbal A, Wu Z, Ben Amara F (2009) Closed-loop control of magnetic fluid deformable mirrors. Opt Express 17(21):18957–18970

Iqbal A, Wu Z, Ben Amara F (2010a) A decentralized robust PID controller design for the shape control of a magnetic fluid deformable mirror. Int J Optomechatron 4(3):246–268

Iqbal A, Wu Z, Ben Amara F (2010b) Mixed sensitivity H_∞ control of magnetic fluid deformable mirrors. IEEE/ASME Trans Mechatron 15(4):548–556

Laird P, Bergamasco R, Berube V, Borra EF, Ritcey AM, Rioux M, Robitaille N, Thibault S, Lande Vieira da Silva Jr., Yockell-Lelivre H (2003) Ferrofluid-based deformable mirrors: a new approach to AO using liquid mirrors. In: Wizinowich PL, Bonaccini D (eds) Proceedings of the SPIE volume 4839, Adaptive Optical System Technologies II, SPIE – The International Society of Optical Engineering

Laird P, Caron N, Rioux M, Borra EF, Ritcey AM (2006) Ferrofluid adaptive mirrors. Appl Opt 45(15):3495–3500

Parent J, Borra EF, Brousseau D, Ritcey AM, Dery JP, Thibault S (2009) Dynamic response of ferrofluidic deformable mirrors. Appl Opt 48(1):1–6

Rosensweig RE (1997) Ferrohydrodynamics. Dover Publications, Mineola

Scherer C, Weiland S (2004) Linear matrix inequalities in control. Delft University of Technology, Delft

Seaman A, Macpherson JB, Borra EF, Ritcey AM, Asselin D, Jerominek H, Thibault S, Campbell MC (2007) Hartmann-Shack measurements of ferrofluidic mirror dynamics. In: Photonics North 2007, Proceedings of the SPIE 6796, no.679603

Skogestad S, Postlethwaite I (2005) Multivariable feedback control: analysis and design. Wiley, Chichester

Thibault S, Brousseau D, Rioux M, Senkow S, Dery JP, Borra EF, A Ritcey AM (2006) Nanoengineered ferrofluid deformable mirror: a progress report. In: Ellerbroek BL, Domenico BC (eds) Advances in adaptive optics II, Proceedings of SPIE 6272, no.627231

Tyson RK (2011) Principles of adaptive optics. CRC Press, Boca Raton

Wu Z, Azhar I, Ben Amara F (2010a) LMI-based decentralized PID controller design with application to the shape control of magnetic fluid deformable mirror. ASME international mechanical engineering congress and exposition, Vancouver, British Columbia, Canada, 12–18 Nov 2010

Wu Z, Azhar I, Ben Amara F (2010b) Magnetic fluid deformable mirror shape control with a multivariable PID controller. IEEE American control conference, Baltimore, Maryland, USA, June 30–July 2 2010

Wu Z, Azhar I, Ben Amara F (2011) LMI-based multivariable PID controller design and its application to the control of the surface shape of magnetic fluid deformable mirrors. IEEE Trans Control Syst Technol 19(4):717–729

Zhou K, Doyle J, Golver K (1995) Robust and optimal control. Prentice Hall, Upper Saddle River

Chapter 2
Adaptive Optics Systems

Contents

2.1 Introduction to Adaptive Optics Systems

The quality of images provided by an imaging system is affected by imperfections in the optical path of light traveling from the object being imaged to the location of the image. These imperfections could be represented by a nonuniformity in the properties of the medium through which light travels or by imperfections in the geometry of the optical components. The effects of these imperfections on the light rays or waves, and consequently on the resulting image, are referred to as *aberrations*. Some of these aberrations can be corrected using conventional optical components. For example, lenses and mirrors have been used for centuries to correct the static aberrations (Rubin 1986). However, more advanced solutions are needed for the complex types of aberrations. Adaptive optics is one of these advanced solutions, which utilizes adaptive optical elements called wavefront correctors to compensate for the complex and dynamic aberrations (Tyson 2011; Hardy 1998;

Z. Wu et al., *Modeling and Control of Magnetic Fluid Deformable Mirrors for Adaptive Optics Systems*, DOI 10.1007/978-3-642-32229-7_2,

Roggemann and Welsh 1996). Adaptive optics systems have been successfully used in astronomical seeing and free space laser communication systems to remove the effects of atmospheric distortions and in retinal imaging systems to reduce the impact of ocular optical aberrations.

AO was first envisioned by Horace W. Babcock in 1953 as a way to correct for the effects of atmospheric turbulence in astronomical imaging (Babcock 1953), but it did not come into common usage until advances in computer technology during the 1990s made the technique practical. In outer space, spherical waves emitted by the source (star) are essentially plane waves when they arrive at the Earth's atmosphere. The waves are then randomly distorted by turbulence in the atmosphere. The turbulence is mainly caused by spatial and temporal variations in the density of the atmosphere, aerosol scattering, and atmospheric temperature fluctuations which induce the wind velocity changes, eddies, etc. Images produced by any telescope larger than a few meters are blurred by these distortions. The idea of compensating these distortions for astronomical telescopes using adaptive optics systems was first introduced by Babcock, who proposed that the problem of astronomical seeing could be solved by incorporating a mirror that provided the possibility of canceling the distortions in the light waves collected for imaging distant astral bodies. Interestingly, the type of wavefront corrector proposed by Babcock was a liquid mirror formed by depositing a thin film of oil on the surface of a mirror. Though this wavefront corrector called Eidophor controlled the wavefront shape by changing the refractive properties of the film and hence is comparable to the modern day spatial light modulators, it is remarkable that the first proposed AO system was based on the concept of liquid mirrors. Babcock's idea was not directly put into practice, but it still remains the first published work on adaptive optics systems.

The works of Fried (1966) and Greenwood (1977) are considered to be seminal in defining the spatial and temporal requirements of AO systems. It was only in 1977 that the first AO system was realized for an astronomical application (Hardy et al. 1977). While the bulk of the research work published since then remains focused on applications in astronomy, a parallel stream of research has been going on in the field of high-energy lasers (HEL) intended for defense applications. The AO systems used in HEL were aimed at reducing the effects of atmospheric turbulence on the laser energy directed at long-range strategic targets. Due to the nature of the HEL applications, significant research effort in this area has gone into the concept of guide star (Happer et al. 1994; Feinleib 1982). A guide star is a source of light which can be used as a reference for measurement of the aberrations to be corrected using an adaptive optics system. In the early 1990s, a large part of the research work conducted in the defense sector was made public (Tyson 2000). The availability of this work triggered a surge in the astronomical applications of AO systems. Almost all major astronomical telescopes were either retrofitted with AO systems or provided with integrated AO systems in their design (Hart 2010; Roddier 1999; Paschall and Anderson 1993). As the technology matured, it found applications in other areas. Adaptive optics has been successfully applied to flood-illumination retinal imaging to produce images of single cones in the living

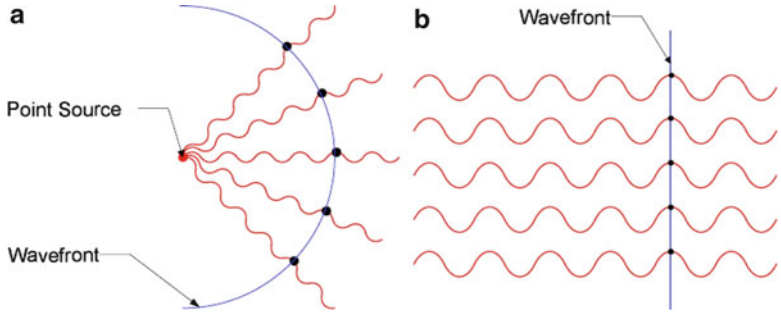


Fig. 2.1 Illustration of a wavefront. (**a**) Circular wavefront; (**b**) planar wavefront

human eye (Roorda and Williams 2002; Hofer et al. 2001b). Combined with optical coherence tomography and scanning laser ophthalmoscope systems, adaptive optics has enabled the three-dimensional imaging of living cone and rod photoreceptors (Zawadzki et al. 2011; Kocaoglu et al. 2011; Dubra et al. 2011; Ferguson et al. 2010; Torti et al. 2009; Fernandez et al. 2008). Besides its use in improving astronomical imaging and retinal imaging, adaptive optics technology has also been used in other applications. Adaptive optics is used for solar astronomy at observatories such as the German 1.5-m GREGOR solar telescope (Berkefeld et al. 2010) and the Swedish 1-m solar telescope (Scharmer et al. 2002). It is also expected to play a military role by allowing ground-based and airborne laser weapons to reach and destroy targets at a distance including satellites in orbit. Adaptive optics has been used to enhance the performance of free space optical communication systems (Vorontsov et al. 2010). Development of an Adaptive Scanning Optical Microscope (ASOM) using the adaptive optics technology was announced by Thorlabs in April 2007 (Potsaid and Wen 2008). In the future, it is believed that more and more new applications will be found, which could benefit the engineering science in a diversified perspective.

Before expanding on the details of these AO systems, a brief description of the concept of a wavefront and how the wavefront is related to optical aberrations is presented in the following.

2.1.1 The Basic Concept of a Wavefront

For light waves originating from a point source and having the same wavelength, a wavefront is defined as an imaginary surface that connects the points featuring the same phase. Figure 2.1 illustrates the concept of a wavefront. As depicted in Fig. 2.1a, the wavefront of the waves traveling unobstructed in a two-dimensional plane is circular in shape. When allowed to diverge in all three dimensions, the waves form perfect spherical wavefronts. If the point source of light is moved to infinity, the waves become collimated and present a planar wavefront, as shown in Fig. 2.1b. The waves with the planar wavefronts are called *plane waves* (Saleh and Teich 2007).

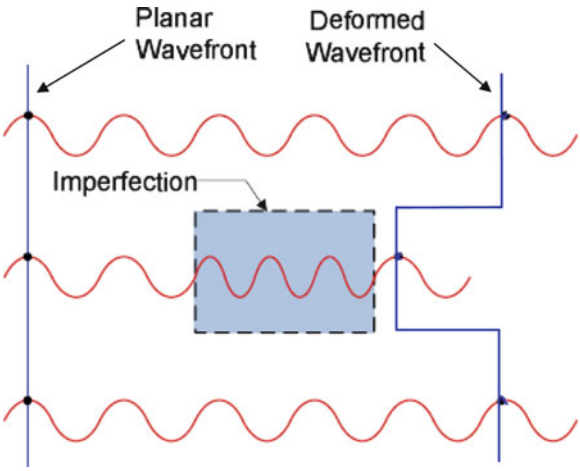


Fig. 2.2 Idealized illustration of a deformed wavefront

Ideally an imaging system should be able to project each point on the object being imaged to a corresponding point in the image plane. However, none of the real-life imaging systems are ideal. Even an optically perfect imaging system produces a dispersed image of a point object where the dispersion pattern is called *Airy disk*, and the system is termed as a diffraction-limited system (Born and Wolf 1997). As the name suggests, such a perfect system is limited only by the phenomenon of diffraction of light and is considered to have the theoretically best possible resolution.

Besides diffraction, the resolution of an imaging system is affected by optical aberrations introduced by the intervening medium between the object and the image plane. These aberrations produce a blurred image of a point object and result in a degradation of the resolution of images provided by the imaging system. The concept of a wavefront as illustrated below provides a convenient tool that can be used to know how the aberrations affect the resolution of the images provided by the system.

The optical aberrations are caused by imperfections in the optical path of light waves traveling between the object and the imaging plane. The imperfections affect the optical path length (OPL) of the waves and result in the deformation of the wavefront. The effect of aberrations on the wavefront of a plane wave is illustrated in Fig. 2.2. As shown, the optical path difference results in a wavefront shape that deviates from its planar shape.

The effect of optical aberrations on the resolution of the images provided by an imaging system is directly related to their effect on the wavefront shape of light waves traveling through the system. This phenomenon offers a black-box approach to the imaging problem: if we know how the wavefront of a plane wave is deformed by the imaging system, then we can fully predict how the image will be formed (Roorda 2002). The deformation of the wavefront accounts for the cumulative effect of all aberrations in the optical path. This approach not only

provides a comprehensive method of describing the optical aberrations but also offers an insight into the methods that can be used to rectify the effects of these aberrations.

Aberrations can be defined as the departure from the idealized conditions of Gaussian optics (Hetch 2002). Within this description aberrations fall into two classes: monochromatic aberrations and chromatic aberrations. Monochromatic aberrations are aberrations that occur because of the nature of the medium through which light travels. For example, in the human eye the main types of monochromatic aberrations include astigmatism, spherical, coma, and Petzval field curvature and distortion. Chromatic aberrations arise when different colors of light propagate at different speeds in a medium, a behavior that is due to the fact that the refractive index is wavelength dependent. The classical treatment of Gaussian optics is based on the assumption that rays of light proceeding from any object point unite in an image point and therefore an object space is reproduced in an image space. This permits the determination of the image of any object for any system. The Gaussian theory, however, is only true as long as the angles φ made by all rays with the optical axis (the axis of symmetry of the optical components such a lenses) are infinitely small, namely, that the approximation $\sin\varphi \approx \varphi$ is satisfied, which is usually called Gaussian theory or first-order theory. Similarly high-order theory uses the approximation

$$\sin\varphi = \varphi - \frac{\varphi^3}{3!} + \frac{\varphi^5}{5!} - \frac{\varphi^7}{7!} + \cdots \tag{2.1}$$

For example, if the first two terms in the expansion are retained as an approximation $\sin\varphi \approx \varphi - (\varphi^3/3!)$, we have then what is called third-order theory. In practice these conditions are seldom realized, since rays from the periphery of a lens are included in the formation of an image. The images then projected in uncorrected systems are ill defined and often blurred by these aberrations. Departures from the Gaussian theory (or first-order theory) result in the five primary aberrations (also known as Seidel aberrations). These aberrations are known as spherical aberration, coma, astigmatism, field curvature, and distortion. In addition to the first term, the series in (2.1) contains many more, smaller, terms. The inclusion of such higher-order terms results in additional higher-order aberrations, which are smaller than the primary aberrations but must still be dealt with. For the static and low-order aberrations, it is more direct and convenient to use geometric and paraxial theory, and the readers can refer to Hetch (2002) for the details. However, for dynamic and high-order aberrations, it is more convenient to describe them using wave optics theory with basis functions.

2.1.2 Representation of Aberrations with Basis Functions

The aberrations can be analyzed using physical optics or wave optics, which studies interference, diffraction, polarization, and other phenomena for which the ray

approximation of geometrical optics is difficult or not valid. In adaptive optics, it is convenient to study the wavefront aberrations, especially the high-order aberrations, using electromagnetic wave theory. An electromagnetic wave can be mathematically described by a complex function

$$u(\mathbf{r},t) = \Re\left\{A(\mathbf{r})\, e^{i\varphi(\mathbf{r})} e^{i2\pi f t}\right\} \tag{2.2}$$

where $u(\mathbf{r},t)$ represents each component of the electric as well as magnetic field vectors, $\Re\{\bullet\}$ denotes the real part of the complex function, $\mathbf{r}$ is a position vector and is represented by (r,θ), f is the frequency, $A(\mathbf{r})$ is the amplitude, and $\varphi(\mathbf{r})$ is the phase of the wave. Equation (2.2) may be written as

$$u(\mathbf{r},t) = \Re\left\{P(\mathbf{r})\, e^{i2\pi f t}\right\} \tag{2.3}$$

where $P(\mathbf{r}) = A(\mathbf{r})\, e^{i\varphi(\mathbf{r})}$ is called complex amplitude of the wave. The time dependence of the complex wave function (2.2) is related to the complex amplitude $P(\mathbf{r})$ by Helmholtz equation (see Saleh and Teich 2007 for details) and is therefore considered to be known a priori. Consequently, the complex amplitude $P(\mathbf{r})$ offers an adequate description of a wave. At any given position $\mathbf{r}$, the complex amplitude $P(\mathbf{r})$ is a complex variable whose magnitude is $|P(\mathbf{r})| = A(\mathbf{r})$, and the phase is the argument $\varphi(\mathbf{r})$. A wavefront is defined as a surface where all the points have the same phase $\varphi(\mathbf{r})$. The phase of the wave is related to its wavefront by

$$\varphi(\mathbf{r}) = \frac{2\pi}{\lambda} W(\mathbf{r}) \tag{2.4}$$

where $W(\mathbf{r})$ is a spatial function that expresses the shape of the wavefront and λ is the wavelength of the wave. For imaging systems, which transfer light waves from an object to an imaging plane, the wavefront function $W(\mathbf{r})$ is typically considered as a two-dimensional spatial function measured in the exit pupil plane, as illustrated in Fig. 2.3a. Exit pupil is the image of the aperture of an optical system formed in the image space by rays emanating from a point on the optical axis in the object space (Hecht 2002). Since most of the imaging systems have circular pupil, polar coordinates have been chosen for the illustration and will be used in all subsequent references to the wavefront function. Figure 2.3a shows the wavefront of an aberrated wave comparing it to the wavefront of an ideal wave. Function $W(r,\theta)$ represents the wavefront shape of the wave as measured with reference to the wavefront of the ideal wave. It is easier to visualize the wavefront function $W(r,\theta)$ in a pupil where the ideal wavefront shape is planar, as shown in Fig. 2.3b. Note that the lines drawn perpendicular to the wavefronts can be considered as rays, which determine the direction in which the segment of the wave is traveling and the position where the image will be formed. For the ideal wavefront shown in Fig. 2.3b, the image will be formed at infinity. The ideal spherical wavefront, as shown in Fig. 2.3a, forms a point image at its center of curvature. The aberrated

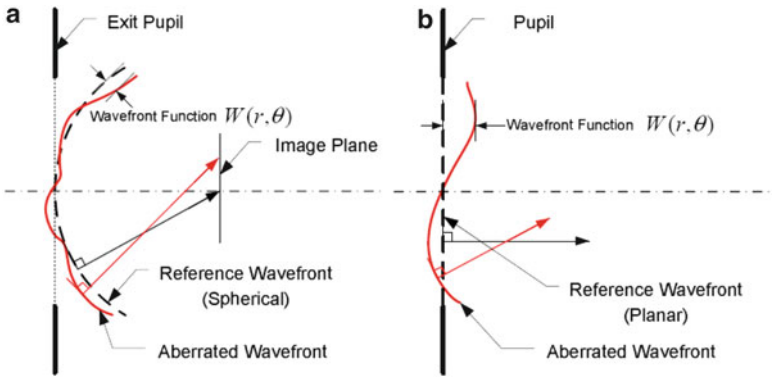


Fig. 2.3 Wavefront function of a distant point object

wavefront, on the other hand, results in the rays not converging to the ideal point in the image plane and hence results in an aberrated image. The wavefront function $W(r,\theta)$ as measured in the exit pupil represents the cumulative effect of all optical aberrations that may be present in the imaging system.

In vision science, it is customary to describe the optical aberrations in terms of simpler forms of aberrations such as defocus, astigmatism, and coma. These aberrations actually represent the various wavefront shapes, and the cumulative effect of all aberrations in an imaging system can be described as a linear combination of these simpler shapes. Mathematically, it amounts to writing the wavefront function $W(r,\theta)$ as a linear combination of simpler basis functions (also called shape functions) representing different types of optical aberrations. Various series of two-dimensional basis functions have been used to represent the wavefront shapes. Seidel series, Taylor series, and Zernike polynomials are some of the popular sets of basis functions. A generalized method of representing the wavefront shape and two of the specific series of basis functions are described in the following paragraphs.

2.1.2.1 Generalized Basis Functions

Analytically, the wavefront function $W(r,\theta)$ may be written as a linear combination of spatial basis functions F_i, $i = 0, 1, 2, \ldots$, as

$$W(r,\theta) = \sum_{i=0}^{\infty} c_i F_i(r,\theta) \tag{2.5}$$

where c_i is the expansion coefficient corresponding to the ith basis function. If the set of basis functions F_i is complete (Kreyszig 1993), any two-dimensional surface shape may be fully represented by an infinite series of these functions. However, it

is a common practice to use a linear combination of only a finite number of basis functions to represent a wavefront surface. For practical reasons, the chosen set of basis functions often consists of orthonormal functions.

2.1.2.2 Seidel Series

Classically, optical aberrations are described in terms of primary aberrations first studied systematically by Seidel (1856). The Seidel aberrations, as they are generally referred to, are described by the following functions:

$$S_i(r,\theta) = S_n^m(r,\theta) = r^n \cos^m\theta \tag{2.6}$$

where $i = 0, 1, 2, \ldots$, is the order of the Seidel function and $n = 0, 1, 2, \ldots$ and $m = 0, 1, 2, \ldots$ are termed the radial degree and azimuthal frequency, respectively. The indices n and m must satisfy $m \leqslant n$, and $n - m$ must be even. The single-index i and the double-index n and m are related by

$$i = \text{int}\left[\frac{\left(n^2 + 2n + 1\right)}{4}\right] + \text{int}\left[\frac{m}{2}\right] \tag{2.7}$$

where int $[x]$ stands for the largest integer smaller than x and, the conversion from the single-index to the double-index is given by

$$\begin{aligned} n &= \text{int}\left[\sqrt{4i+1} - 1\right] \\ m &= 2i - \text{int}\left[\frac{n(n+2)}{2+1}\right] + 1 \end{aligned} \tag{2.8}$$

Important Seidel aberrations are illustrated in Fig. 2.4 and are described as follows:

- *Distortion.* Distortion results from a variation of the magnification due to off-axis field positions, that is, different parts of the object have different magnification. This type of aberrations does not cause any blur in the image.
- *Field curvature or defocus.* This type of aberration causes the image to focus on a curved plane. While defocus exists for both on-axis and off-axis positions, field curvature is defined as an on-axis aberration. The aberration may be eliminated by using a curved imaging surface or, more practically, using a spherical lens.
- *Astigmatism.* Astigmatism occurs when the tangential and sagittal foci of the optical system do not coincide. Astigmatism is extremely important in vision science and eye care, since the human eye often exhibits this aberration due to imperfections in the shape of the cornea or the lens.
- *Coma.* When rays entering different off-axis parts of the pupil focus at different points, the result is called *coma*. Coma causes a tear-like image for a point object.

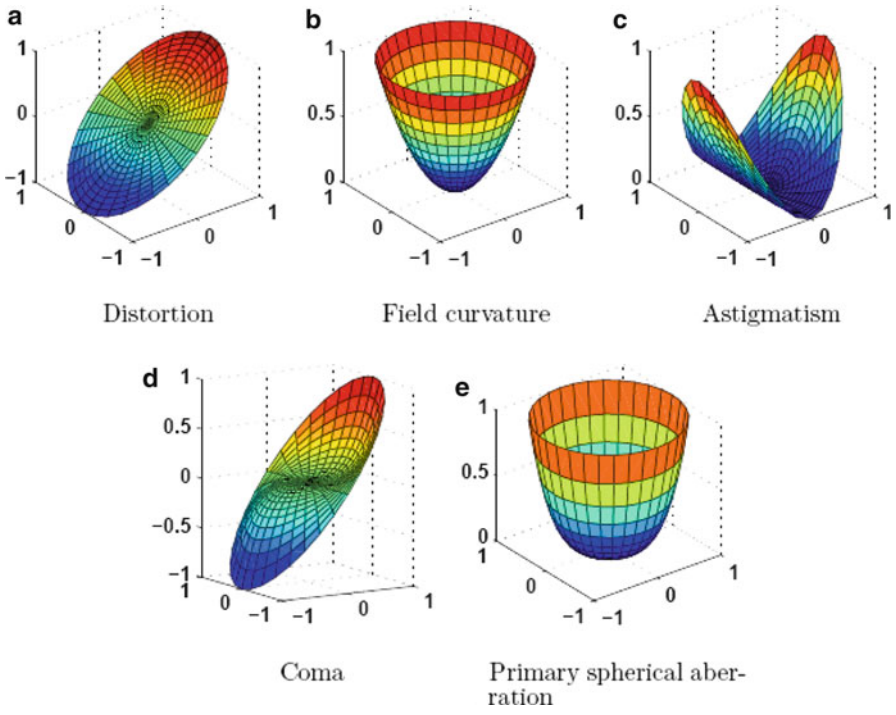


Fig. 2.4 Seidel aberrations. (**a**) Distortion; (**b**) Field curvature; (**c**) Astigmatism; (**d**) Coma; (**e**) Primary spherical aberrations

Coma can be an inherent property of telescopes using parabolic mirrors due to the off-center light source, where the different parts of the mirror do not reflect the light to the same point.

- *Spherical aberrations.* Spherical aberration occurs when rays from periphery of the pupil focus at a point different from the axis. In lens systems, the effect can be minimized using special combinations of convex and concave lenses, as well as using aspheric lenses.

2.1.2.3 Zernike Polynomials

The Zernike polynomials are a sequence of polynomials that are orthogonal on the unit disk. Named after Frits Zernike, they play an important role in beam optics and have been accepted in vision science as the standard for reporting of aberrations (Thibos et al. 2000). The wavefront function $W(r,\theta)$ of a wave with a pupil radius R can be expressed fully in terms of Zernike polynomials $Z_i\ (\rho, \theta)$ as

$$W(r,\theta) = W(R\rho,\theta) = \sum_{i=0}^{\infty} c_i Z_i\ (\rho,\theta) \tag{2.9}$$

where $\rho = (r/R)$ is the normalized pupil radius and c_i is the ith Zernike coefficient. The Zernike polynomials $Z_i(\rho, \theta)$ can be written as (Dai 2008)

$$Z_i(\rho, \theta) = Z_n^m(\rho, \theta) = \Re_n^{|m|}(\rho)\,\Theta^m(\theta) \tag{2.10}$$

where $n = 0, 1, 2, \ldots$, and $m = 0, \pm 1, \pm 2, \ldots$, are the radial degree and azimuthal frequency, respectively. The indices n and m must satisfy $m \leqslant n$, and $n - m$ must be even. The radial polynomials $\Re_n^{|m|}(\rho)$ are defined as

$$\Re_n^{|m|}(\rho) = \sum_{s=0}^{(n-|m|)/2} \frac{(-1)^s\,(n-s)!\rho^{n-2s}}{s!\,[(n+m)/2-s]!\,[(n-m)/2-s]!} \tag{2.11}$$

and the triangular functions are defined as

$$\Theta^m(\theta) = \begin{cases} \sqrt{2}\cos|m|\,\theta & (m > 0) \\ 1 & (m = 0) \\ \sqrt{2}\sin|m|\,\theta & (m < 0) \end{cases} \tag{2.12}$$

The first few nonzero radial polynomials are

$$\Re_0^0(\rho) = 1$$
$$\Re_1^1(\rho) = \rho$$
$$\Re_2^0(\rho) = 2\rho^2 - 1$$
$$\Re_2^2(\rho) = \rho^2$$
$$\Re_3^1(\rho) = 3\rho^3 - 2\rho$$
$$\Re_3^3(\rho) = \rho^3$$
$$\Re_4^0(\rho) = 6\rho^4 - 6\rho^2 + 1$$
$$\Re_4^2(\rho) = 4\rho^4 - 3\rho^2$$
$$\Re_4^4(\rho) = \rho^4$$

Figure 2.5 shows the wavefront shapes represented by the first ten Zernike polynomials.

2.1.3 Optical Metrics of Aberrations

In the following, the important question of how different wavefront aberrations affect the performance of an imaging system is addressed. There are various

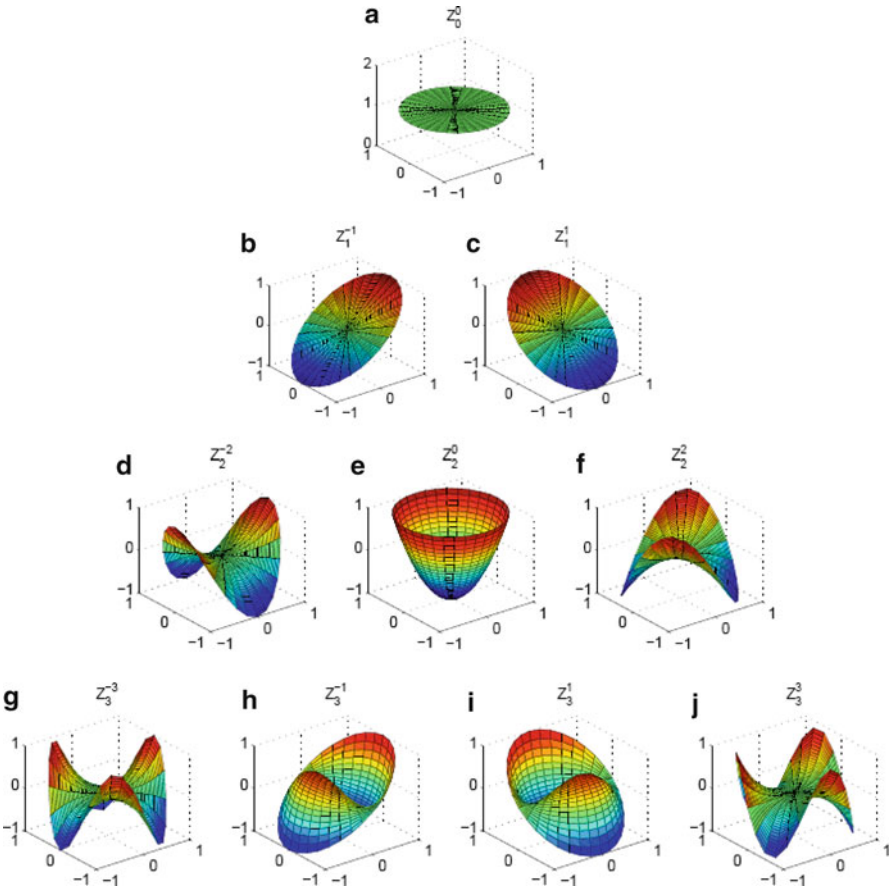


Fig. 2.5 Zernike polynomials. (**a**) Piston; (**b**) y-tilt; (**c**) x-tilt; (**d**) y-astigmatism; (**e**) defocus; (**f**) x-astigmatism; (**g**) y-trefoil; (**h**) y-coma; (**i**) x-coma; (**j**) x-trefoil

methods that can be used to express the effect of a known wavefront aberration on the quality of the resulting image. Some of these methods are based on the wavefront function $W(r,\theta)$ as measured in the pupil plane, while other methods are based on calculations done in the image plane.

2.1.3.1 Pupil Plane Metrics

The wavefront aberrations can be most conveniently measured in the exit pupil plane using a wavefront sensor or an aberrometer which samples the wavefront at discrete locations in the pupil. The discrete measurements can be used to interpolate the wavefront anywhere in the pupil by fitting the data using any of the series of basis

functions discussed earlier. When the wavefront function $W(r,\theta)$ is represented by one of these series of basis functions, the diameter of the pupil and the coefficients of the basis functions are the only quantities required to fully express the aberrations.

Root Mean Square Error

The most commonly used performance metric in the pupil plane is the root mean square (RMS) of the wavefront function $W(r,\theta)$. If the piston term (i.e., Z_0^0) is ignored, the RMS is the same as the standard deviation of the wavefront function $W(r,\theta)$. When the wavefront function $W(r,\theta)$ is decomposed into Zernike polynomials, the RMS of the function can be measured in terms of the coefficients c_i as

$$\sigma = \sqrt{\sum_{i=1}^{J} c_i^2} \tag{2.13}$$

where c_i is the coefficient of the ith Zernike polynomial and J is the total number of Zernike polynomials considered for the reconstruction of the wavefront. Similar expressions can be found for the other basis function series types mentioned above.

Wavefront Refraction

This is an approximate measure of wavefront error and considers low-order aberrations only. Wavefront refraction can be simply defined as the radial curvature of aberration, the details of which can be found in Dai (2008).

2.1.3.2 Image Plane Metrics

Though hard to measure directly, the image plane metrics offer a more detailed prediction of the performance of an imaging system. The following are the most commonly used metrics which are defined in the image plane:

Point Spread Function

The point spread function (PSF) describes the response of an imaging system to a point source or a point object. It can be expressed in terms of the distribution of the irradiance that results from a single point source in the object space. As shown in Fig. 2.6a, for a diffraction-limited optical system, the PSF can be visualized simply as the diameter of the Airy disk pattern which is the resulting image of a point source. The PSF of a typically aberrated system is shown in Fig. 2.6b. Sometimes, it is more convenient to represent the PSF as a two-dimensional cross-sectional plot

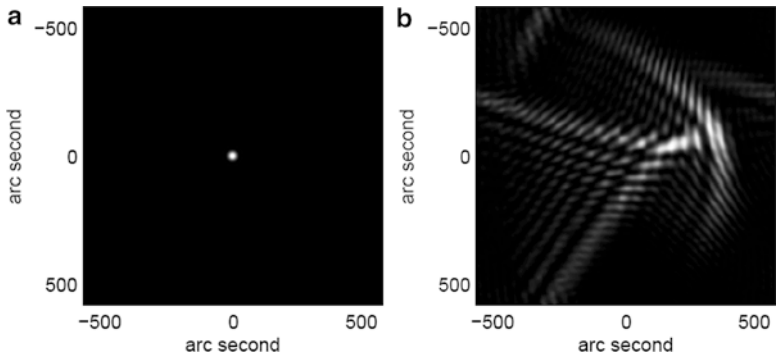


Fig. 2.6 Point spread function of (**a**) diffraction-limited wavefront versus (**b**) typically aberrated wavefront

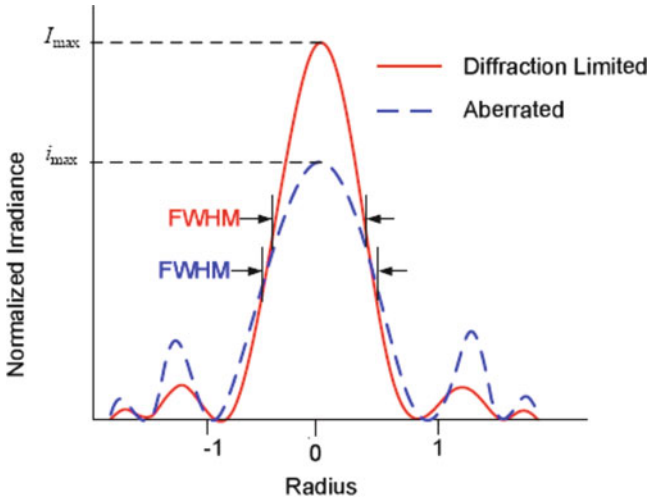


Fig. 2.7 Illustration of a 2D point spread function, Strehl ratio, and full-width-at-half maximum (FWHM)

of the irradiance distribution function. Figure 2.7 shows the cross-sectional view the PSF of a typically aberrated wavefront versus that of a diffraction-limited wavefront. Analytically, the PSF of an optical system can be computed using Fraunhofer approximation as follows (Goodman 2004):

$$\mathrm{PSF}\,(r,\theta) = K \cdot |\Im\,(P\,(r,\theta))|^2 \tag{2.14}$$

where $\Im\,(\bullet)$ represents the Fourier transform operator and K is a constant.

Strehl Ratio

It is defined as the ratio of the maximum irradiance of an optical system over that of a diffraction-limited optical system with the same pupil size and can be written as

$$S = \frac{i_{\max}}{I_{\max}} \tag{2.15}$$

where $i_{\max}$ is the maximum irradiance of the optical system and $I_{\max}$ is the maximum irradiance of a diffraction-limited optical system with the same pupil size. Strehl ratio can also be expressed as the ratio of maximum PSF value of an aberrated system to that of a diffraction-limited system, as shown in Fig. 2.7. However, this description makes sense only if the PSF of the aberrated system is not too badly distorted. The higher the Strehl ratio, the better is the quality of image. The best image quality is provided by a diffraction-limited system which has a Strehl ratio of unity.

Full-Width-At-Half Maximum

The full-width-at-half maximum (FWHM) intensity of the PSF of an optical system is another measure of its performance. FWHM is an expression of the extent of a function, given by the difference between the two extreme values of the independent variable at which the dependent variable is equal to half of its maximum value. The metric is illustrated in Fig. 2.7. Generally, the smaller the FWHM, the better is the quality of the resulting image.

Optical and Modulation Transfer Functions

These are measures of how much information is preserved, or modulated, from the object space into the image space. Analytically, the optical transfer function (OTF) of an imaging system can be computed using the Fourier transform of its PSF, the details of which can be found in Goodman (2004). The modulation transfer function (MTF) is the modulus of the OTF. For illustration purposes, MTF computed on the same point spread functions, as presented in Fig. 2.6, is shown in Fig. 2.8.

2.1.3.3 Order of Aberrations

The simple aberrations such as defocus and astigmatism are categorized as the low-order aberrations, while the more complex ones are known as high-order aberrations. The exact definition of the order of aberrations depends on the series of basis functions selected to describe the wavefront.

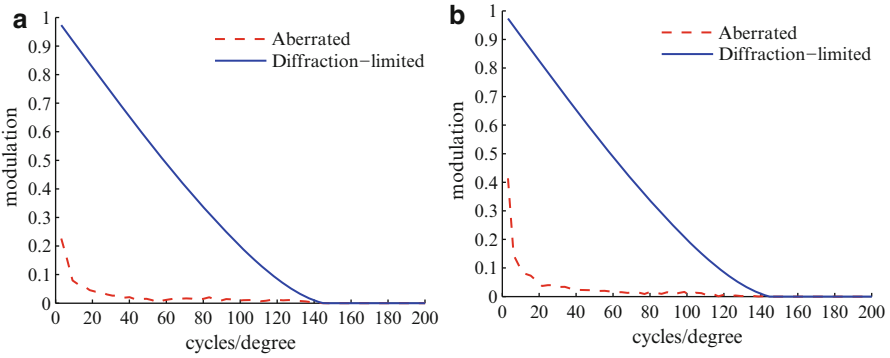


Fig. 2.8 Illustration of the modulation transfer function of a diffraction-limited wavefront compared to a typically aberrated wavefront. (**a**) Vertical. (**b**) Horizontal

The low-order aberrations are the main contributors to the overall loss of vision or, in the case of imaging systems, to the degradation of the quality of images provided by the systems. Though to a lesser extent, spherical aberrations, coma, and other higher-order aberrations also contribute significantly to the loss of vision or image quality.

In conventional optical systems, the low-order aberrations remain the major focus of the corrective actions taken to improve the vision or image quality. Due to relatively fewer performance benefits and higher complexity, correction of higher-order aberrations has been limited only to advanced applications, for example, in astronomy. However, with the significant developments in the imaging technology and the associated systems, the correction of higher-order aberrations has now become a viable—in many cases a necessary—feature.

2.1.4 Wavefront Aberration Correction

Having explained what a wavefront is, how it can be represented, and how it can be used to measure the quality of an imaging system, we now turn to how the wavefront aberrations can be corrected and how the correction enhances the quality of the aberrated images. Central to the idea of wavefront correction is the concept of phase conjugation as explained below (Dai 2008).

2.1.4.1 Phase Conjugation

The complex conjugate of the complex amplitude function $P(r,\theta) = A(r,\theta)\,e^{i\varphi(r,\theta)}$ is $A(r,\theta)\,e^{-i\varphi(r,\theta)}$, where (r,θ) represents position vector $\mathbf{r}$ as given in (2.2). Analytically, if we multiply the complex amplitude function $A(r,\theta)\,e^{i\varphi(r,\theta)}$

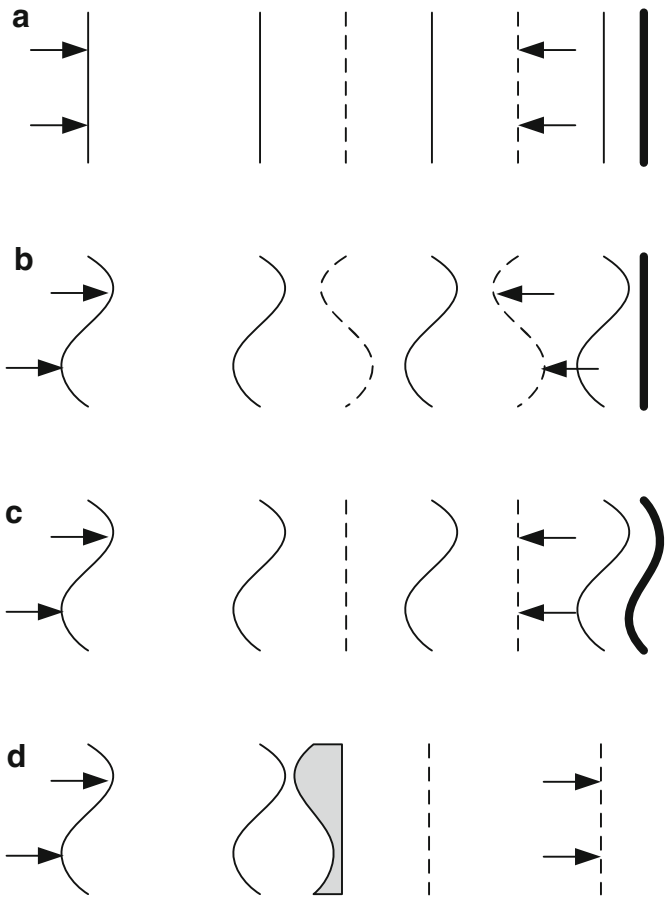


Fig. 2.9 Phase conjugation. Effect of (**a**) flat mirror on plane wave, (**b**) flat mirror on aberrated wave, (**c**) deformed mirror on aberrated wave, and (**d**) refractive medium on aberrated wave. The *dotted lines* in (**a**) to (**c**) represent the reflected wavefront

with the phase component of its complex conjugate (i.e., with $e^{-i\varphi(r,\theta)}$), the phase of the complex amplitude function is canceled out. Since the phase of the complex amplitude function represents the optical aberrations in the system, the multiplication operation amounts to the cancelation of the aberrations, which is the primary concern of the adaptive optics systems.

The concept is physically implemented by adding, to the original aberrated wave, an optical aberration which has an equal but opposite phase, as illustrated in Fig. 2.9. For a plane wave propagating toward a flat mirror, as shown in Fig. 2.9a, the reflected wavefront is the same as the incident wavefront. In this configuration, the mirror does not affect the phase of the wave but only reverses the direction of propagation. Similarly, an aberrated wavefront becomes inverted

due to the reflection from the mirror but otherwise maintains the same shape, as shown in Fig. 2.9b. However, when the flat mirror is replaced by a deformed mirror with a deformation half the magnitude of the incident wavefront, the reflected wavefront becomes flat (Fig. 2.9c). A similar effect can be obtained by allowing the aberrated wavefront to pass through a refractive medium which introduces a wavefront aberration with the same phase but opposite sign as that of the incident wavefront (Fig. 2.9d).

As described earlier, the aberrations in the wavefront result in an aberrated image. Therefore, the possibility of canceling out the aberrations in the wavefront using the principle of phase conjugation provides a direct means of improving the quality of images provided by an imaging system. The systems that utilize the concept of phase conjugation to compensate for the aberrations are called adaptive optics systems. The idea of phase conjugation is not new and has been used for centuries in the form of spectacles. The low-order aberrations in imaging systems are routinely compensated for using conventional optical devices such as lenses and mirrors—a manifestation of phase conjugation. However, the conventional optical components are static components that do not have the capability to correct for the higher-order aberrations. AO systems, on the other hand, make use of dynamic optical components to compensate for high-order time-varying aberrations.

2.2 Operating Principle of an Adaptive Optics System

The basic setup of a typical AO system is shown in Fig. 2.10. The main components of the system include a wavefront sensor, a wavefront corrector—usually a deformable mirror (DM)—and a controller. As shown in Fig. 2.10, the aberrated

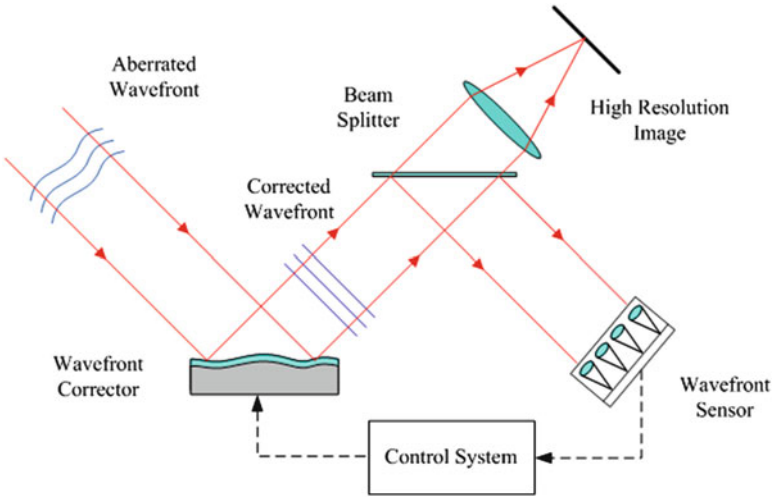


Fig. 2.10 A typical adaptive optics system

light from the source or the object to be imaged is collected and is directed to the wavefront sensor via the wavefront corrector. The wavefront sensor measures the deformations in the wavefront of the incident light. The wavefront deformation information is fed to a control system. The control system analyzes the information and computes the necessary actuator commands. When these commands are applied to the wavefront corrector, the wavefront corrector adjusts the shape of the mirror surface to cancel out the aberrations in the incident wavefront. Normally the process of measurement of wavefronts and application of commands to the wavefront corrector are iterative in nature. When a specified degree of correction has been achieved, the imaging system is activated to obtain images with enhanced quality.

In what follows, a brief description of the three basic components of an AO system is presented.

2.2.1 Wavefront Sensors

Over the years, various wavefront sensing technologies such as interferometric sensors, curvature sensors, and Shack–Hartman wavefront sensors have been used (Tyson 2011; Hardy 1998). These wavefront sensors are described in the following:

2.2.1.1 Interferometric Wavefront Sensors

This type of wavefront sensors works on the principle of interferometry. The wavefront measurement data is given in the form of an interferogram generated by the interference of two wavefronts: a reference wavefront and the wavefront to be measured. The shape and the magnitude of the latter are determined by reading the fringe pattern resulting from the interference of the two wavefronts. A description of the various types of interferometric wavefront sensors can be found in Tyson (2011).

2.2.1.2 Wavefront Curvature Sensors

A wavefront curvature sensor uses an array of small lenslets to focus the wavefront into an array of spots (Kellerer 2010). The local phase curvature of the wavefront is determined by measuring the relative intensities at two different places along the axis of the beam—one before the focal plane and the other after the focal plane. The curvature information is then utilized to obtain the wavefront shape.

2.2.1.3 Shack–Hartmann Wavefront Sensors

Recently, Shack–Hartmann wavefront sensor (SHWS) has emerged as the most commonly used type of wavefront sensors, particularly in vision science. This

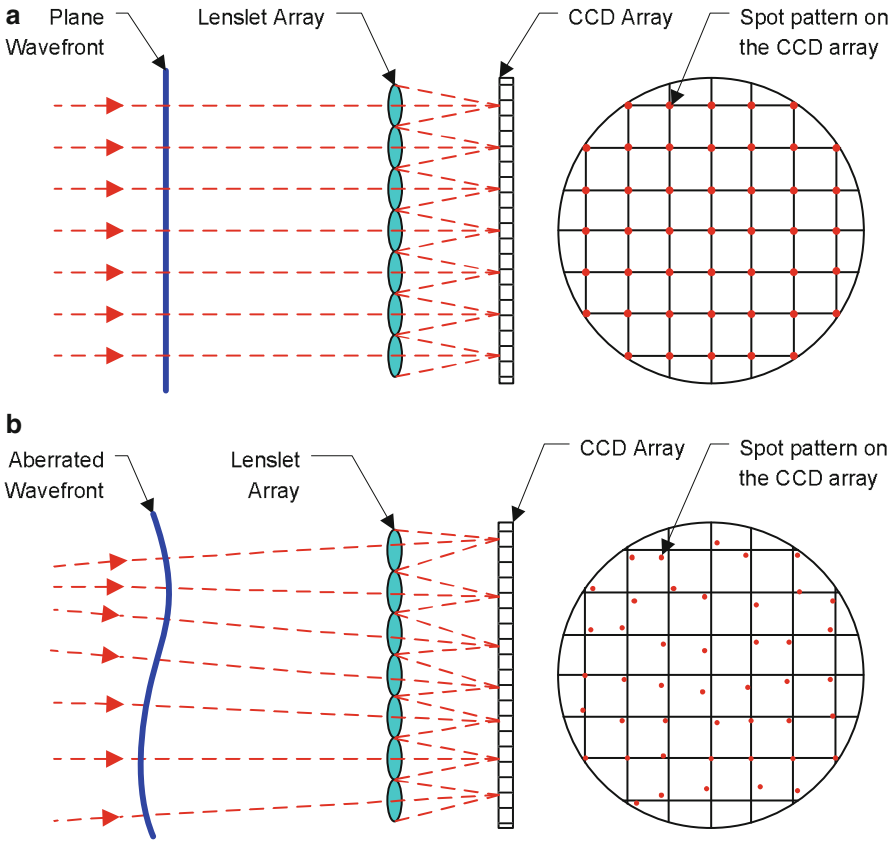


Fig. 2.11 Working principle of a Shack–Hartmann wavefront sensor. (**a**) Perfect wavefront; (**b**) aberrated wavefront

device samples the wavefront through an array of tiny lenslets. Each lenslet creates a focused spot, which is captured by a CCD camera. The array of spots, when sampling a planar wavefront, would form a uniform grid pattern, as shown in Fig. 2.11a. On the other hand, if the wavefront is aberrated, the spots will not focus on-axis in the focal plane of their corresponding lenslets but will deviate according to the local slope of the wavefront, as illustrated in Fig. 2.11b. The displacement of the spots from their on-axis position measures the local slope of the wave. Using this displacement data, the complete wavefront shape can be reconstructed. Generally, the data is fit on one of the series of two-dimensional spatial functions, as described in Sect. 2.1.2. A detailed description of how the SHWS data is used to reconstruct wavefront shape is given in Appendix C.

2.2.2 *Wavefront Correctors*

A wavefront corrector is the principal component of an AO system, which works on the incident aberrated wavefront and cancels out the aberrations. There are two basic types of wavefront correctors:

2.2.2.1 Spatial Light Modulators

This type of wavefront correctors uses arrays of liquid crystal micro-lenses (Huang et al. 2011; Love et al. 2010; Mu et al. 2010; Fernández et al. 2009; Prieto et al. 2004). The phase of the light passing through the array can be controlled by electronically or optically manipulating the refractive index of the individual micro-lenses. The phase modulation provides the necessary means to control the wavefront shape. The spatial light modulators (SLMs) are available in both reflective as well as transparent modes. This type of wavefront correctors has the advantage of very high spatial resolution provided by the extremely small liquid crystals. Since they are based on the existing LCD technology, they also have a significant cost advantage over the other types of wavefront correctors. SLMs are limited by the relatively small magnitude of correction that they can provide. Another limitation is the requirement of linearly polarized light since the liquid crystals can modulate only light polarized along their axis.

2.2.2.2 Deformable Mirrors

Deformable mirrors (DM) have evolved as the most widely used wavefront correction elements in adaptive optics systems. These mirrors are characterized by a reflective surface which can be locally deformed, hence providing a means to change the wavefront shape of the reflected light. The main advantage of these mirrors is their reflective nature that allows for a low loss of radiant energy and hence makes them particularly suitable for applications where the intensity of light is, or needs to be, low. Based on the type of reflective surface, DMs may be classified as segmented or continuous type.

Segmented mirrors feature an array of mirror segments, each of which can be individually controlled (Manzanera et al. 2011; Devaney et al. 2008; Helmbrecht and Juneau 2007). Figure 2.12 shows a typical segmented mirror. Some of these devices are piston only where each segment can be moved perpendicular to the mirror plane typically using a single actuator. Others can be tipped and tilted and are generally supported by three actuators per segment. Thanks to micromachining technology, these mirrors can be fabricated with a very high density of segments. For example, Boston Micromachines offers segmented mirrors named *Multi-DM* with 140 segments in 4.9-mm-diameter aperture and a stroke of 5.5 μm (Fig. 2.13). Most segmented mirrors have relatively small stroke lengths. However, the segmented

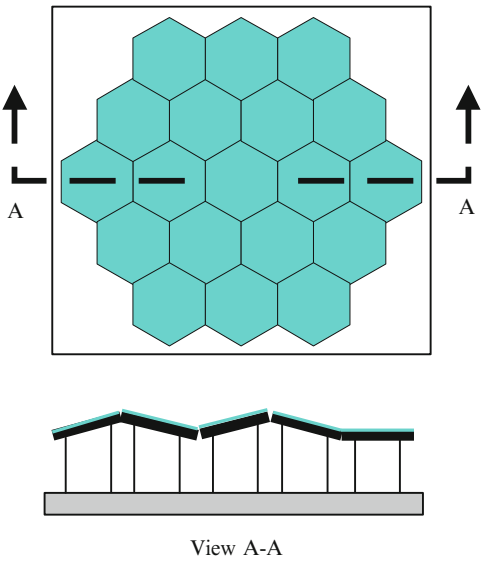


Fig. 2.12 Schematic of a segmented deformable mirror

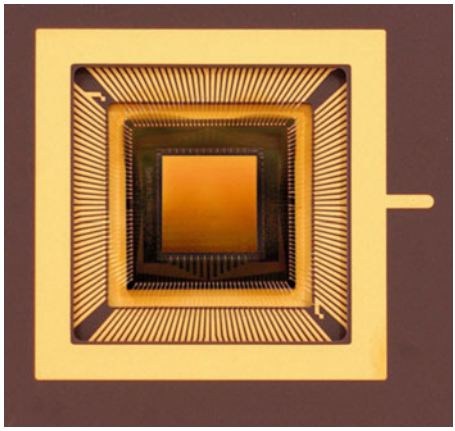


Fig. 2.13 140-actuator segmented mirror (Photo courtesy of Boston Micromachines)

mirrors are known to have better operating speeds. The percentage of the surface area of a segmented mirror covered by reflective segments is known as the fill factor. Segmented mirrors with low fill factor are not preferred due to the loss of energy at the discontinuities in their surface. The segments can be square or hexagonal in shape. As each segment can be adjusted independently, these mirrors are considered to be more suitable for correcting wavefront aberrations with higher spatial frequencies (Hampson 2008).

Continuous mirrors are characterized by a flexible, continuous reflecting surface which can be locally deformed using an appropriate actuation mechanism

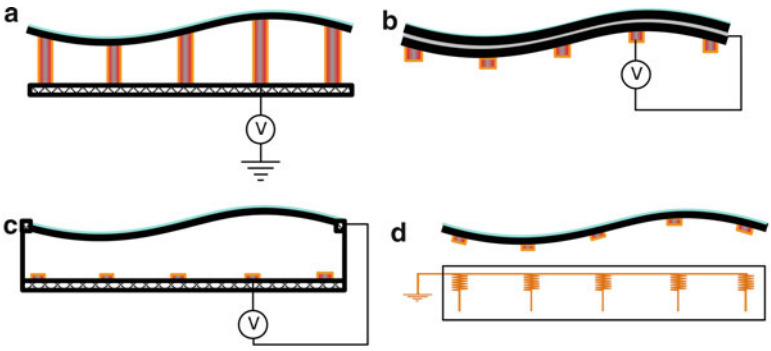


Fig. 2.14 Types of continuous surface mirrors

Fig. 2.15 Piezoelectric deformable mirror (Photo courtesy of OKO Technologies)

(Zawadzki et al. 2011; Ödlund et al. 2010; Guzmán et al. 2010; Correia et al. 2010b; Devaney et al. 2008; Vdovin et al. 2008). Typically when an actuator in a continuous DM is activated, the surface deflection is not restricted to the area directly above that actuator but extends to the whole mirror. Continuous mirrors remain the wavefront corrector of choice for many applications because of the low loss of energy. Based on the actuation mechanism, these mirrors may be further categorized into four types, as shown in Fig. 2.14.

– Unimorph deformable mirrors: Fig. 2.14a shows a type of continuous surface unimorph deformable mirrors which utilize piezoelectric actuators that expand when a voltage is applied (Verpoort and Wittrock 2010). Example of this type of mirrors is a 37-actuator mirror from OKO Technologies with 6-μm full surface stroke (Fig. 2.15).
– Bimorph deformable mirrors: They consist of layers of piezoelectric materials, as shown in Fig. 2.14b. By controlling the voltage applied to an electrode, the

Fig. 2.16 37-actuator membrane mirror (Photo courtesy of OKO Technologies)

curvature of the mirror surface can be changed, which serves the purpose of a deformable mirror (Verpoort and Wittrock 2010; Rodrigues et al. 2009). A 35-actuator mirror with 15-mm diameter and 16-μm total stroke produced by AOptix is the typical example of this category of deformable mirrors (Devaney et al. 2008).

– Electrostatically actuated deformable mirrors: Fig. 2.14c shows an example of membrane mirrors that utilize electrostatic forces to deform the mirror surface coated on a flexible membrane. Electrostatic-driven deformable mirrors usually do not have large strokes. OKO Technologies' mirror with 37-actuators and 15-mm-diameter pupil is an example of this type of mirrors (Fig. 2.16). The membrane clamped at its edges can provide up to 9-μm of peak-to-valley surface deflections.
– Electromagnetically actuated deformable mirrors: Fig. 2.14d shows another membrane-type mirror that utilizes electromagnetic force between small permanent magnets attached to the membrane and electromagnetic coils placed underneath the magnets. A 52-actuator mirror with 15-mm-diameter pupil offered by Imagine Eyes is an example of this type of mirrors (Fig. 2.17). The mirrors are reported to have up to 50-μm surface deflections (Ödlund et al. 2010).

The different types of mirrors mentioned above have their trade-offs, some of which have been investigated and reported in studies (Guzmán et al. 2010; Devaney et al. 2008; Dalimier and Dainty 2005). The common drawbacks of these solid thin plate or membrane-based mirrors are the high cost of per actuator channel and the relatively low stroke deflection. Recently, a new type of liquid deformable mirrors that uses magnetic liquids (ferrofluids) has been suggested by Borra et al. (2009, 2008), Brousseau et al. (2010, 2006), and Iqbal et al. (2010, 2009), Iqbal and Ben Amara (2008, 2007). Magnetic fluid deformable mirrors (MFDMs) have some major advantages over solid ones, such as their extremely smooth surfaces, low cost per actuator, large strokes, and ease of scalability compared to solid deformable mirrors.

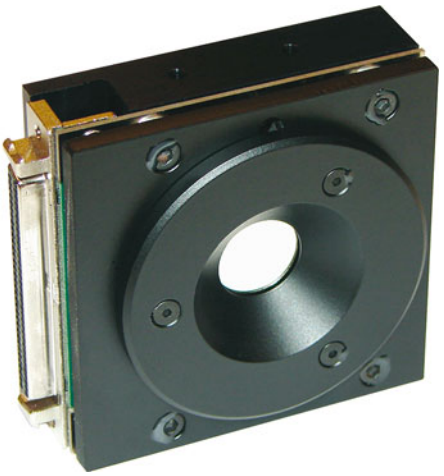


Fig. 2.17 52-actuator magnetic deformable mirror (Photo courtesy of Imagine Eyes)

2.2.3 Controllers

A controller is the vital link between the wavefront sensor and the wavefront corrector in an AO system. Its function is to generate, based on wavefront aberration measurements made by the wavefront sensor, a set of actuator commands that are applied to the wavefront corrector to cancel the wavefront aberrations. Since the wavefront correctors act on the wavefront shape of the incident wave, the wavefront corrector control problem faced in AO can be generalized as a shape control problem. Within the context of the DMs, the objective of the shape control is to obtain a shape for the mirror that best tracks the shape required to cancel the aberrations.

A wide range of controllers for AO systems have been reported in the literature. The controllers used in AO systems can be broadly grouped as proportional–integral–derivative (PID) type controllers, optimal controllers, and adaptive controllers. A brief review of these controllers is presented as follows:

2.2.3.1 PID Controllers

Most AO systems make use of proportional–integral–derivative (PID) feedback compensators (Stein and Gorinevsky 2005; Huang et al. 1995b; Downie and Goodman 1989). The objective of these systems is to drive the residual wavefront error to zero. Traditionally, single-input single-output (SISO) integral or PID controllers are developed for a single actuator. The controller is then duplicated for each actuator in the wavefront corrector. The design of these controllers used in AO systems is typically based on DC model (static gain) of the wavefront corrector. The DC model is experimentally obtained by measuring the wavefront shape resulting

from applying a constant input to each of the actuators. The resulting input–output map is usually referred to as the *influence function* of the wavefront corrector and is employed to determine the necessary control input for the wavefront corrector. Traditionally, either proportional or proportional-plus-integral structure is employed to provide updates of the inputs when using the influence-function-based controllers (Tyson 2011; Hardy 1998).

2.2.3.2 Optimal Controllers

Optimal controllers employ mathematical optimization methods to obtain the controller parameters that provide a specified performance objective such as the minimization of the wavefront shape error. A number of control algorithms used in AO systems employ linear quadratic Gaussian (LQG) criterion which seeks to minimize a quadratic cost function and, thus, determine the controller that minimizes the error between the actual shape and the desired shape of the wavefront corrector (Raynaud et al. 2011; Correia et al. 2010a; Petit et al. 2009; Looze 2009; Kulcsár et al. 2006; Wiberg et al. 2004a, b). H_2 and H_∞ based optimal controllers that seek to optimize the desired system norms of the overall system have also been reported in the literature (Hinnen et al. 2008, 2007; Baudouin et al. 2008; Voulgaris et al. 2003).

2.2.3.3 Adaptive Control

Due to the dynamic nature of the aberrations in most of the applications of AO systems, the shape control problem addressed in AO is the tracking of an unknown and time-varying shape for the wavefront corrector (i.e., desired shape of the mirror). Since the desired shape of the wavefront corrector is unknown and time varying, the controller must be tuned online to converge to the controller needed to achieve the desired control objective. This type of controllers whose parameters can be tuned online is called adaptive controllers. They are a relatively recent addition to the list of controllers used in AO systems and are expected to improve the performance and play an important role in the future AO applications (Ficocelli and Ben Amara 2012; Monirabbasi and Gibson 2010; Liu and Gibson 2007).

Using the above-mentioned methods, the controller design is usually implemented with one of the three control architectures: decentralized, centralized, and distributed.

2.2.3.4 Decentralized Control

In decentralized controllers, an array of single-input single-output controllers is used. Each controller is driven by the residual wavefront error corresponding to a single spatial location on the wavefront. The output of the controller is used to drive

only one actuator. Decentralized controllers are easy to implement. However, their performance is limited since each individual controller makes use of only partial residual wavefront error information.

2.2.3.5 Centralized Control

A centralized controller uses all the measurement data from the sensor to simultaneously compute the control inputs to all the actuators in the wavefront corrector. The controller is typically designed based on an optimal performance criterion such as LQG, H_2, or H_∞ design criteria. Centralized controllers achieve the best performance results but are computationally expensive. The implementation of the centralized controller may not be feasible for systems with a large number of actuators.

2.2.3.6 Distributed Control

Distributed control is sometimes referred to as localized control (Fraanje et al. 2010; Ellenbroek et al. 2006). With the increase in the number of actuator used in deformable mirrors, such as those used in the extremely large telescope system with a diameter more than 30 m, the computational requirements of centralized optimal control laws may be excessive. If the deformable mirror has a spatially invariant dynamics, then the system model can be decomposed into a number of identical subsystems. A distributed control strategy can then be used to control each subsystem with less computation cost. A localized controller only uses information from several neighboring sensors to generate commands to a given actuator, and the complexity of the localized control algorithm does not increase with the size of the mirror.

A more detailed overview of the control approaches for adaptive optics systems is presented in Chap. 6.

2.3 Retinal Imaging Adaptive Optics Systems

One of the most interesting applications of adaptive optics systems involves ophthalmic retinal imaging systems. The introduction of AO systems into ophthalmic imaging systems started more than two decades ago (Dreher et al. 1989). Research studies conducted since then have shown that the quality of images provided by ophthalmic imaging systems can be significantly improved by augmenting them with AO systems. Some of these initial studies have resulted in important breakthroughs in vision science (Zawadzki et al. 2011; Kocaoglu et al. 2011; Ferguson et al. 2010; Chen et al. 2007; Artal et al. 2004).

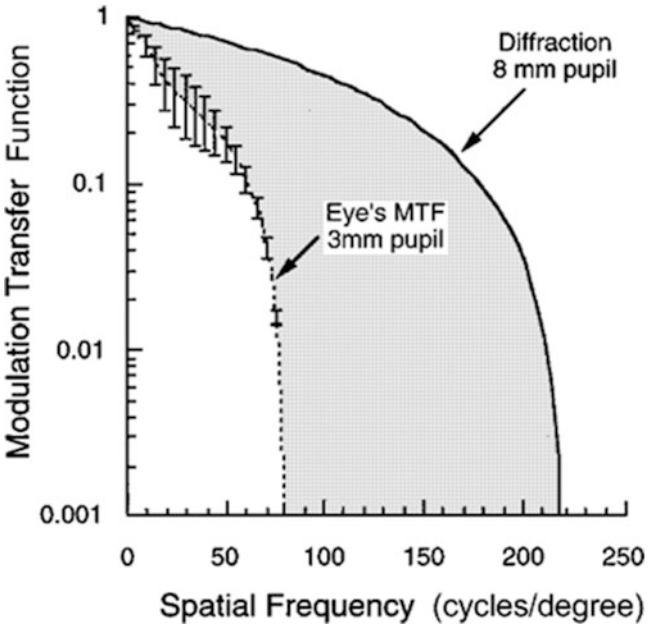


Fig. 2.18 Potential benefit of correcting the higher-order aberrations in the eye (Liang and Williams 1997)

The eye is an imaging system that forms one of the primary faculties of the human sensory system. It has long been known that the normal human eye suffers from many different aberrations that affect the vision of the individual. The most common aberrations in the human eye are defocus and astigmatism. The normal eyes are also known to suffer from additional aberrations such as spherical aberrations, coma-like aberrations, and a host of irregular aberrations. The uncorrected aberrations in the eye not only affect what it can see but also determine the smallest internal structures that can be observed when looking into the eye. The various imaging technologies, which are used to examine the internal parts of the eye, are affected by this phenomenon. Specifically, the naturally existing aberrations in the eye degrade the quality of the retinal images obtained by the various available technologies such as funduscopy, scanning laser ophthalmoscopy (SLO), and optical coherence tomography (OCT) (Godara et al. 2010; Hampson 2008).

Figure 2.18 shows the margin of improvement in the image quality that can be achieved by correcting higher-order aberrations (Liang and Williams 1997). Shown in the figure is the best MTF of the eye averaged over 14 eyes with 3-mm pupil with an optimal correction of defocus and astigmatism. This represents the best image quality that can be achieved by rectifying the low-order aberrations only. Now compare it to the ideal MTF of the eye with an 8-mm pupil blurred only by diffraction, which represents the best image quality with all the aberrations—including the high-order aberrations—compensated for. The figure shows that finer

spatial details can be observed in a diffraction-limited eye with a dilated pupil as compared to an aberrated eye with a regular pupil size. The shaded area between the two plots shows the range of image contrasts and spatial frequencies (resolution) that are available for imaging of the retina only if the high-order aberrations are corrected. This clearly shows the need of a system that can be used to compensate for the high-order aberrations in the eye and has been the driving force behind research efforts made over the years to incorporate AO systems in ophthalmic imaging systems.

2.3.1 Correction of the Aberrations in the Eye Using AO Systems

It is established that the correction of high-order aberrations in the eye carries significant visual and image quality benefits. This has been the driving force behind a number of research efforts undertaken during the last decade in order to integrate AO in systems ranging from funduscopes to laser surgery systems. A brief exposition of the ophthalmic technologies which have benefited from AO systems is presented in the following.

2.3.1.1 Retinal Imaging

The most extensively used application of AO in the eye is retinal imaging, which provides high-resolution images of the living retina. The ability to examine in vivo the minute details of the retina tissue allows for the early detection and regular monitoring of retinal diseases. There are three main imaging modalities which have thus far been utilizing AO systems (Godara et al. 2010; Hampson 2008).

Flood-Illumination Ophthalmoscopy

This is the most commonly used and straightforward imaging technique that combines AO with direct imaging of the retina. It employs a low-coherence source of light which illuminates a localized patch of retina. A CCD camera placed in conjugation with the retinal plane is used to collect the light reflected from the retina and take the image of the illuminated patch. Figure 2.19 shows the layout of a basic retinal imaging system augmented with AO. A light beam with a planar wavefront originating from a light source—usually a laser or a super-luminescent diode (SLD)—is directed to the retina of the eye. The light reflects from the retina and exits the eye as a beam having an aberrated wavefront—the aberrations being induced by the sources mentioned earlier. The beam is directed to the DM, and the reflection from the mirror is detected by the wavefront sensor. The wavefront sensor

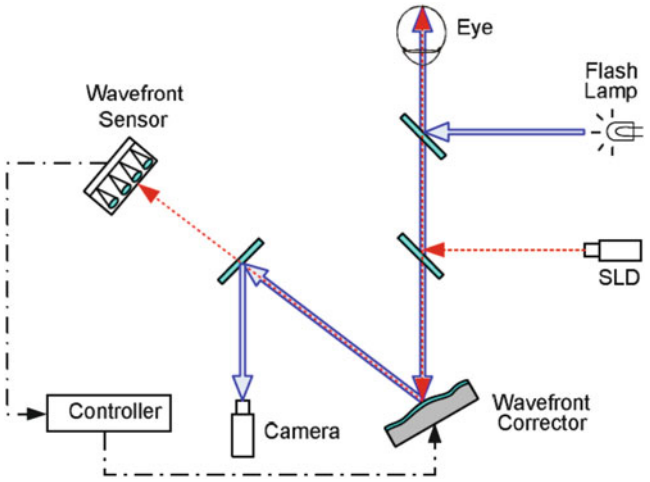


Fig. 2.19 A typical retinal imaging adaptive optics system

measures the magnitude of the aberrations over the whole pupil. The measurements are fed to a control system which computes the appropriate command signals for the DM. The signals, when applied to the DM, deform the shape of the mirror surface such that the aberrations in the wavefront of incident beam are iteratively canceled out. When a desired level of cancelation or correction is achieved, an illumination source such as a flash lamp is turned on. The light from this source floods the retina of the eye and is reflected back following the same path as that of the laser/SLD light. A dichroic mirror reflects the light to the imaging system which collects the image of the retina. Note that the flood light reflected from the retina still carries the aberration as it exits the eye. Before reaching the camera, however, the aberrations are canceled out by the DM which now is appropriately deformed to do so.

Scanning Laser Ophthalmoscopy

It may be noted that the ophthalmoscope described above provides no depth resolution on the retinal tissue and AO augmentation enhances only the transverse resolution of the images. Scanning laser ophthalmoscopy (SLO), on the other hand, is an imaging technique which also offers axial resolution, that is, the layers beneath the surface of the photoreceptor cells can be distinctly imaged. SLO utilizes a laser beam which is scanned across the retina, and the image is built up point by point. As shown in Fig. 2.20, the reflected light is made to pass through a small aperture called confocal pinhole, which is placed in conjugation with the retinal plane. This arrangement increases the contrast level of focused axial location by preventing light reflected from other layers of the retinal tissue from passing through and hence offers the desired depth resolution. An SLO can be augmented with an AO system

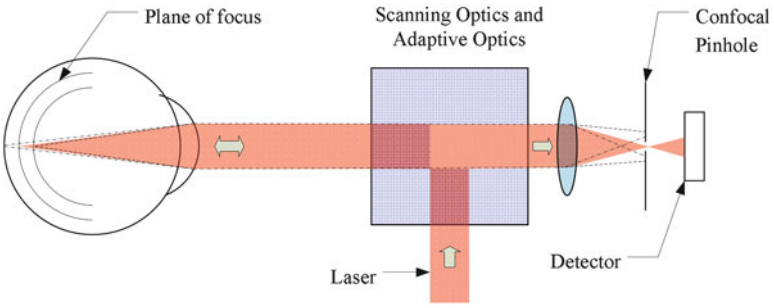


Fig. 2.20 The working principle of a scanning laser ophthalmoscope

by introducing a wavefront sensor and a wavefront corrector in the optical path. The wavefront sensor utilizes the same light as used for the basic SLO. The wavefront corrector is inserted into the optical system such that the wavefront aberrations introduced by the eye are canceled out before the reflected laser light passes through the pinhole and is detected by the image detector.

Optical Coherence Tomography

Optical coherence tomography (OCT) is a technique for obtaining subsurface images of translucent or opaque materials at a resolution equivalent to a low-power microscope. By effectively filtering the scattered light using optical coherence technology, cross-sectional images are obtained using the coherence light reflected from 1 to 2 mm below the surface in biological tissue. OCT systems offer submicron imaging resolution in both the transverse and axial directions of the retinal images, mostly performing better than SLO. An OCT system is based on interferometric principle and uses a Michelson interferometer in conjunction with a low-coherence light source (Zhang et al. 2005). OCT systems have been devised in two different types: time-domain OCT and spectral OCT.

2.3.2 History of Ophthalmic Adaptive Optics Systems

Shortly after Babcock (1953) introduced the concept of AO in astronomy, Smirnov (1961) suggested the idea of correcting the high-order aberrations in the eye, probably independently from the existing ideas in astronomy. However, it was not until late 1980s that this idea could be put into practice. Dreher et al. (1989) were the first to use a deformable mirror to improve the quality of images provided by scanning laser ophthalmoscope, though they corrected static astigmatism only. Artal and Navarro (1989) achieved measurement of inter-center distance between photoreceptor cones using speckle interferometry, a method well known

in astronomy. Miller et al. (1996) were the first to obtain direct images of the retina using a high-resolution fundus camera. Though they corrected defocus and astigmatism only, they were able to obtain up to 3.5-μm resolution. The real breakthrough came when Liang et al. (1994) demonstrated that the Shack–Hartman wavefront sensor could be used to measure the ocular aberrations. This prompted a number of studies that sought to provide wavefront correction promising diffraction-limited images of the retina.

Liang et al. (1997a, b) incorporated the Shack–Hartman wavefront sensor in an AO system and obtained images with unprecedented high resolution. By correcting the higher-order aberrations, they were able to obtain images with enough photometric accuracy to identify microscopic structures as small as the single cone cell. They reported up to 3.2-fold improvement in the transverse resolution over that of a 2.5-mm pupil. Significant enhancements in the contrast sensitivity of the eye were also reported. Hofer et al. (2001a, b) were the first to study and correct temporal variations in the ocular aberrations using closed-loop AO systems. They reported threefold improvements in Strehl ratio and up to 33% increase in contrast of the images of the cone photoreceptors. They reported dynamic compensation of up to fifth-order Zernike polynomials with frequency up to 0.8 Hz.

The newfound capability of measuring and correcting the high-order aberrations in the eye has facilitated developments such as the identification of the arrangement of the three classes of photoreceptor cells (Roorda and Williams 1999), psychological testing of the functioning of vision at the neural level (Artal et al. 2004), and finding alternate causes of color blindness (Carroll et al. 2004). Beside direct retinal imaging, OCT and SLO too have benefited from AO. Optical coherence tomography remains one of the most promising applications of AO systems (Kocaoglu et al. 2011; Mujat et al. 2010; Torti et al. 2009; Zhang et al. 2005). Moreover, when incorporated into a scanning laser ophthalmoscope (SLO), AO systems have significantly improved the lateral resolution of this type of ophthalmoscopy (Dubra et al. 2011; Ferguson et al. 2010; Mujat et al. 2009; Zhang et al. 2006).

2.3.3 Challenges to Ophthalmic Adaptive Optics Systems

Despite the demonstrated potential of the ophthalmic AO systems, the use of these systems has been limited only to research studies, and they are still not available for the widespread clinical imaging applications. Some of the relevant problems impeding the accessibility of these systems are as follows:

- High Stroke Requirement. A set of generalized requirements for ophthalmic applications of wavefront correctors is presented in Doble and Miller (2006). One of the most significant requirements of the wavefront correctors is the unusually high stroke needed to compensate for the large magnitude (peak-to-valley path length difference) of the wavefront aberrations in the eye. For a deformable mirror with the proposed size of 10-mm diameter, Doble and

Williams (2004) suggest a minimum of ± 12 μm of surface deflection as needed to compensate for the typical aberrations in the eye. For specific applications, Doble and Miller (2006) report a requirement of deformable mirror strokes as high as 53 μm. A survey of the wavefront correctors being used in ophthalmic AO systems is presented in Zawadzki et al. (2011) and Devaney et al. (2008). Few of the surveyed wavefront correctors meet the requirements of ophthalmic AO systems, as mentioned above. Only recently magnetic deformable mirrors with sufficient stroke have been reported (Zawadzki et al. 2011). Currently, a number of strategies are being used to work around the requirement of a large stroke. Some of these strategies include use of optical components, such as Badal optometer (Chen et al. 2006) and Alvarez lens pairs (Evans et al. 2009; Alvarez et al. 1967), to separately compensate the low-order aberrations, cascading of two deformable mirrors (Zou et al. 2011; Li et al. 2010; Zawadzki et al. 2009), and double incidence of the aberrated wave on the deformable mirror (Webb et al. 2004). All these strategies add to the complexity, size, and cost of the AO system. Therefore, a wavefront corrector that could provide a large peak-to-valley wavefront correction is needed.

- High cost of the current AO systems is another major impediment to the widespread accessibility of these systems in ophthalmic imaging systems. For wavefront correctors to be available for the routine ophthalmic imaging, Doble and Miller (2006) and Doble and Williams (2004) suggest an average cost of $10 or lower per actuator is expected for the deformable mirrors. With the relatively high cost of current wavefront correctors, only a significant cost reduction will render these systems viable for wide applications in ophthalmic imaging systems.

  Both of the above-mentioned problems that impede the widespread adoption of AO systems in ophthalmic imaging systems may be addressed using the recently developed magnetic fluid deformable mirrors that are discussed in detail in the following chapters.

2.4 Summary

A review of the literature on adaptive optics systems and their applications in ophthalmic imaging systems has been presented in this chapter. The fundamental motivation of adaptive optics systems is presented, and a brief history of adaptive optics systems is reviewed. In order to better understand the details of adaptive optics systems, the basic concept of a wavefront and how the wavefront is related to the optical aberrations are discussed in the framework of physical optics. The representation of wavefront aberrations using electromagnetic wave theory and the expression of the aberrations using orthonormal basis functions are introduced. The standard optical metrics that can be used to express the effect of a known wavefront aberration on the quality of the resulting image is presented. The operating principle of adaptive optics systems is briefly described. The basic setup of a typical adaptive optics system is illustrated, where the three main components of the systems,

namely, the wavefront sensor, the wavefront corrector, and the controller, are outlined in detail. A brief exposition of the ophthalmic technologies which have benefited from adaptive optics systems is also presented. The nature of optical aberrations in the eye is summarized, and the role and challenges associated with the use of AO in ophthalmic imaging systems are reviewed to provide one example of the motivations for considering AO systems based on magnetic fluid deformable mirrors.

References

Alvarez AW (1967) Two-element variable-power spherical lens, US Patent 3305294

Artal P, Navarro R (1989) High-resolution imaging of the living human fovea: measurement of the intercenter cone distance by speckle interferometry. Opt Lett 14:1098–1100

Artal P, Chen L, Fernandez EJ, Singer B, Manzanera S, Williams DR (2004) Neural compensation of the eye's optical aberrations. J Vis 4:281–287

Babcock HW (1953) The possibility of compensating astronomical seeing. Publ Astron Soc Pac 65:229–236

Baudouin L, Prieur C, Guignard F, Arzelier D (2008) Robust control of a bimorph mirror for adaptive optics systems. Appl Opt 47:3637–3645

Berkefeld T, Soltau D, Schmidt D, von der Lühe O (2010) Adaptive optics development at the German solar telescopes. Appl Opt 49(31):G155–G166

Born M, Wolf E (1997) Principles of optics. Cambridge University Press, Cambridge

Borra EF (2009) Liquid mirrors in engineering. Optics and Photonics News, pp 14–17, September 2009

Borra EF, Brousseau D, Cliche M, Parent J (2008) Aberration correction with a magnetic liquid active mirror. Mon Not R Astron Soc 391(4):1925–1930

Brousseau D, Borra EF, Ruel HJ, Parent J (2006) A magnetic liquid deformable mirror for high stroke and low order axially symmetrical aberrations. Opt Express 14:11486–11493

Brousseau D, Borra EF, Rochette M, Landry DB (2010) Linearization of the response of a 91-actuator magnetic liquid deformable mirror. Opt Express 18(8):8239–8250

Carroll J, Neitz M, Hofer H, Neitz J, Williams DR (2004) Functional photoreceptor loss revealed with adaptive optics: an alternate cause of *color blindness*. Proc Natl Acad Sci USA 101:8461–8466

Chen L, Kruger PB, Hofer H, Singer B, Williams DR (2006) Accommodation with higher-order monochromatic aberrations corrected with adaptive optics. J Opt Soc Am A 23:1–8

Chen DC, Jones SM, Silva DA, Olivier SS (2007) High-resolution adaptive optics scanning laser ophthalmoscope with dual deformable mirrors. J Opt Soc Am A 24:1305–1312

Correia C, Raynaud HF, Kulcsar C, Conan JM (2010a) Minimum variance control for woofer tweeter systems in adaptive optics. J Opt Soc Am A 27(11):133–144

Correia C, Raynaud HF, Kulcsar C, Conan JM (2010b) On the optimal reconstruction and control of adaptive optical systems with mirror dynamics. J Opt Soc Am A 27(2):333–349

Dai G (2008) Wavefront optics for vision correction, vol PM179, SPIE press monograph. SPIE Publications, Bellingham

Dalimier E, Dainty C (2005) Comparative analysis of deformable mirrors for ocular adaptive optics. Opt Express 13:4275–4285

Devaney N, Dalimier E, Farrell T, Coburn D, Mackey R, Mackey D, Laurent F, Daly E, Dainty C (2008) Correction of ocular and atmospheric wavefront: a comparison of the performance of various deformable mirrors. Appl Opt 47(35):6550–6562

Doble N, Miller DT (2006) Wavefront correctors for vision science. In: Porter J, Queener H, Lin J, Thorn K, Awwal A (eds) Adaptive optics for vision science: principles, practices, design and applications. Wiley, New York

Doble N, Williams DR (2004) The applications of MEMS technology for AO in vision science. IEEE J Sel Top Quantum Electron 10(3):629–635

Downie J, Goodman W (1989) Optimal wavefront control for adaptive segmented mirrors. Appl Opt 28:5326–5332

Dreher AW, Bille JF, Weinreb RN (1989) Active optical depth resolution improvement of the laser tomographic scanner. Appl Opt 28:804–808

Dubra A, Sulai Y, Norris JL, Cooper RF, Dubis AM, Williams DR, Carroll J (2011) Noninvasive imaging of the human rod photoreceptor mosaic using a confocal adaptive optics scanning ophthalmoscope. Biomed Opt Express 2(7):1864–1876

Ellenbroek R, Verhaegen M, Doelman N, Hamelinck R, Rosielle N, Steinbuch M (2006) Distributed control in adaptive optics: deformable mirror and turbulence modeling. Proc SPIE 6272:62723K

Evans JW, Zawadzki RJ, Jones SM, Olivier SS, Werner JS (2009) Error budget analysis for an adaptive optics optical coherence tomography system. Opt Express 17(16):13768–13784

Feinleib J (1982) Proposal 82–P4. Adaptive Optics Associates, Cambridge

Ferguson RD, Zhong Z, Hammer DX, Mujat M, Patel AH, Deng C, Zou W, Burns SS (2010) Adaptive optics scanning laser ophthalmoscope with integrated wide field retinal imaging and tracking. J Opt Soc Am A 27(11):265–277

Fernandez EJ, Hermann B, Povazay B, Unterhuber A, Sattmann H, Hofer B, Ahnelt PK, Drexler W (2008) Ultrahigh resolution optical coherence tomography and pancorrection for cellular imaging of the living human retina. Opt Express 16:11083–11094

Fernandez EJ, Prieto PM, Artal P (2009) Wave-aberration control with a liquid crystal on silicon (LCOS) spatial phase modulator. Opt Express 17(13):11013–11025

Ficocelli M, Ben Amara F (2012) Online tuning for retinal imaging adaptive optics systems. IEEE Trans Control Syst Technol 20(3):747–754

Fraanje R, Massioni P, Verhaegen M (2010) A decomposition approach to distributed control of dynamic deformable mirrors. Int J Optomechatron 4(3):269–284

Fried DL (1966) Limiting resolution looking down through the atmosphere. J Opt Soc Am 56:1380–1384

Godara P, Dubis AM, Roorda A, Duncan JL, Carroll J (2010) Adaptive optics retinal imaging: emerging clinical applications. Optom Vis Sci 87(12):930–941

Goodman JW (2004) Introduction to Fourier optics. Roberts & Company, Colorado

Greenwood DP (1977) Bandwidth specification for adaptive optics systems. J Opt Soc Am 67:390–393

Guzmán D, Juez F, Myers R, Guesalaga A, Lasheras FS (2010) Modeling a MEMS deformable mirror using non-parametric estimation techniques. Opt Express 18(20):21356–21369

Hampson KM (2008) Adaptive optics and vision. J Mod Opt 55(21):3425–3467

Happer W, MacDonald GJ, Max CE, Dyson FJ (1994) Atmospheric turbulence compensation by resonant optical backscattering from the sodium layer in the upper atmosphere. J Opt Soc Am 11(1):263–276

Hardy JW (1998) Adaptive optics for astronomical telescopes. Oxford University Press, New York

Hardy JW, Lefebvre JE, Koliopoulos CL (1977) Real-time atmospheric compensation. J Opt Soc Am 67:360–369

Hart M (2010) Recent advances in astronomical adaptive optics. Appl Opt 49(16):D17–D29

Hecht E (2002) Optics. Addison Wesley, San Francisco

Helmbrecht MA, Juneau T (2007) Piston-tip-tilt positioning of a segmented MEMS deformable mirror. Proc SPIE 6467:64670M

Hinnen K, Verhaegen M, Doelman N (2007) Exploiting the spatiotemporal correlation in adaptive optics using data-driven H_2-optimal control. J Opt Soc Am A 24:1714–1725

Hinnen K, Verhaegen M, Doelman N (2008) A data driven H_2 optimal control approach for adaptive optics. IEEE Trans Control Syst Technol 16(3):381–389

Hofer H, Artal P, Singer B, Aragon JL, Williams DR (2001a) Dynamics of the eye's aberration. J Opt Soc Am A 18(3):497–505
Hofer H, Chen L, Yoon GY, Singer B, Yamauchi Y, Williams DR (2001b) Improvement in retinal image quality with dynamic correction of the eye's aberrations. Opt Express 8:631–643
Huang J, Looze D, Denis N, Castanon D, Wirth A (1995a) Dynamic modeling and identification of an adaptive optics system. In: Proceedings of the 4th IEEE conference on control applications, Albany, NY, USA, pp 456–463
Huang J, Looze D, Denis N, Castanon D, Wirth A (1995b) Control design for an adaptive optics system. In: Proceedings of the IEEE conference on decision and control, New Orleans, LA, USA, pp 3753–3756
Huang H, Inoue T, Tanaka H (2011) Stabilized high-accuracy correction of ocular aberrations with liquid crystal on silicon spatial light modulator in adaptive optics retinal imaging system. Opt Express 19(16):15026–15040
Iqbal A, Ben Amara F (2007) Modeling of a magnetic fluid deformable mirror for retinal imaging adaptive optics systems. Int J Optomechatron 1(2):180–208
Iqbal A, Ben Amara F (2008) Modeling and experimental evaluation of a circular magnetic-fluid deformable mirror. Int J Optomechatron 2(2):126–143
Iqbal A, Wu Z, Ben Amara F (2009) Closed-loop control of magnetic fluid deformable mirrors. Opt Express 17(21):18957–18970
Iqbal A, Wu Z, Ben Amara F (2010) Mixed sensitivity H_∞ control of magnetic fluid deformable mirrors. IEEE/ASME Trans Mechatron 15(4):548–556
Kellerer A (2010) Curvature sensors: noise and its propagation. J Opt Soc Am A 27:A29–A36
Kocaoglu OP, Lee S, Jonnal RS, Wang Q, Herde AE, Derby JC, Gao W, Miller DT (2011) Imaging cone photoreceptors in three dimensions and in time using ultrahigh resolution optical coherence tomography with adaptive optics. Biomed Opt Express 2(4):748–763
Kreyszig E (1993) Advanced engineering mathematics. Wiley, New York
Kulcsár C, Raynaud HF, Petit C, Conan JM, de Lesegno Viaris P (2006) Optimal control, observers and integrators in adaptive optics. Opt Express 14:7464–7476
Li C, Sredar N, Ivers KM, Queener H, Porter J (2010) A correction algorithm to simultaneously control dual deformable mirrors in a woofer-tweeter adaptive optics system. Opt Express 18(16):16671–16684
Liang J, Williams DR (1997) Aberration and retinal image quality of the normal human eye. J Opt Soc Am A 14(11):2873–2883
Liang JZ, Grimm B, Goelz S, Bille JF (1994) Objective measurement of wave aberrations of the human eye with the use of a Hartmann-Shack wavefront sensor. J Opt Soc Am A 11:1949–1957
Liang J, Williams DR, Miller DT (1997) Supernormal vision and high-resolution retinal imaging through adaptive optics. J Opt Soc Am A 14(11):2884–2892
Liu YT, Gibson JS (2007) Adaptive control in adaptive optics for directed energy systems. Opt Eng 46(4):046601
Looze DP (2009) Linear quadratic Gaussian control for adaptive optics systems using a hybrid model. J Opt Soc Am 26(1):1–9
Love GD, Kirby AK, Ramsey RA (2010) Sub-millisecond, high stroke phase modulation using polymer network liquid crystals. Opt Express 18(7):7384–7389
Manzanera S, Helmbrecht MA, Kempf CJ, Roorda A (2011) MEMS segmented-based adaptive optics scanning laser ophthalmoscope. Biomed Opt Express 2(5):1204–1217
Miller DT, Williams DR, Morris GM, Liang JZ (1996) Images of cone photoreceptors in the living human eye. Vision Res 36:1067–1079
Monirabbasi S, Gibson S (2010) Adaptive control in an adaptive optics experiment. J Opt Soc Am A 27(11):A84–A96
Mu Q, Cao Z, Hu L, Liu Y, Peng Z, Xuan L (2010) Novel spectral range expansion method for liquid crystal adaptive optics. Opt Express 18(21):21687–21696
Mujat M, Ferguson RD, Iftimia N, Hammer DX (2009) Compact adaptive optics line scanning ophthalmoscope. Opt Express 17(12):10242–10258

Mujat M, Ferguson RD, Patel AH, Iftimia N, Lue N, Hammer DX (2010) High resolution multimodal clinical ophthalmic imaging system. Opt Express 18(11):11607–11621

Ödlund E, Raynaud HF, Kulcsár C, Harms F, Levecq X, Martins F, Chateau N, Podoleanu AG (2010) Control of an electromagnetic deformable mirror using high speed dynamics characterization and identification. Appl Opt 49(31):G120–G128

Paschall R, Anderson D (1993) Linear quadratic Gaussian control of a deformable mirror adaptive optics system with time-delayed measurements. Appl Opt 32:6347–6358

Petit C, Conan JM, Kulcsár C, Raynaud HF (2009) Linear quadratic Gaussian control for adaptive optics and multiconjugate adaptive optics: experimental and numerical analysis. J Opt Soc Am A 26(6):1307–1324

Potsaid B, Wen JT (2008) Adaptive scanning optical microscope: large field of view and high-resolution imaging using a MEMS deformable mirror. J Micro/Nanolithogr MEMS MOEMS 7(2):021009

Prieto PM, Fernandez EJ, Manzanera S, Artal P (2004) Adaptive optics with a programmable phase modulator: applications in the human eye. Opt Express 12:4059–4071

Raynaud HF, Correia C, Kulcsar C, Conan JM (2011) Minimum variance control of astronomical adaptive systems with actuator dynamics under synchronous and asynchronous sampling. Int J Robust Nonlinear Control 21(7):768–789

Roddier F (1999) Adaptive optics in astronomy. Cambridge University Press, Cambridge

Rodrigues G, Bastaits R, Roose S, Stockman Y, Gebhardt S, Schoenecker A (2009) Modular bimorph mirrors for adaptive optics. Opt Eng 48:034001

Roggemann MC, Welsh BM (1996) Imaging through turbulence. CRC-Press, Boca Raton

Roorda A (2002) Human visual system – image formation. In: Hornak JP (ed) Encyclopedia of imaging science and technology, vol 1. Wiley, New York, pp 539–557

Roorda A, Williams DR (1999) The arrangement of the three cone classes in the living human eye. Nature 397:520–522

Roorda A, Williams DR (2002) Optical fiber properties of individual human cones. J Vis 2:404–412

Rubin ML (1986) Spectacles: past, present and future. Surv Ophthalmol 30:321–327

Saleh BE, Teich MC (2007) Fundamentals of photonics. Wiley, New York

Scharmer GB, Dettori P, L̈ofdahl MG, Shand M (2002) Adaptive optics and correlation tracker systems for the new Swedish solar telescope. In: Keil S, Avakyan S (eds) Innovative telescopes and instrumentation for solar astrophysics, Proceedings of SPIE, vol 4853–52, SPIE press, Bellingham

Seidel L (1856) Zur Dioptrik Uber die Entwicklung der Glieder 3ter Ordnung. Astron Nachr 43:289

Smirnov MS (1961) Measurement of the wave aberration of the human eye. Biophysics 6:687–703

Stein G, Gorinevsky DM (2005) Design of surface shape control for large two-dimensional array. IEEE Trans Control Syst Technol 13:422–433

Thibos L, Applegate RA, Schweigerling JT, Webb R (2000) Standards for reporting the optical aberrations of eyes. In: Vision science and its applications. Optical Society of America, Washington, DC, pp 232–244

Torti C, Považay B, Hofer B, Unterhuber A, Carroll J, Ahnelt PK, Drexler W (2009) Adaptive optics optical coherence tomography at 120,000 depth scans/s for non-invasive cellular phenotyping of the living human retina. Opt Express 17(22):19382–19400

Tyson RK (2000) Adaptive optics engineering handbook. Marcel Dekker, New York

Tyson RK (2011) Principles of adaptive optics. CRC Press, Boca Raton

Vdovin G, Soloviev O, Samokhin A, Loktev M (2008) Correction of low order aberrations using continuous deformable mirrors. Opt Express 16:2859–2866

Verpoort S, Wittrock U (2010) Actuator patterns for unimorph and bimorph deformable mirrors. Appl Opt 49:G37–G46

Vorontsov M, Weyrauch T, Carhart G, Beresnev L (2010) Adaptive optics for free space laser communications. In: Applications of lasers for sensing and free space communications, vol LSMA1, OSA technical digest series (CD). Optical Society of America, New York

Voulgaris PG, Bianchini G, Bamieh B (2003) Optimal H_2 controllers for spatially invariant systems with delayed communication requirements. Syst Control Lett 50:347–361

Webb RH, Albanese MJ, Zhou Y, Bifano T, Burns SA (2004) Stroke amplifier for deformable mirrors. Appl Opt 43:5330–5333

Wiberg DM, Max CE, Gavel DT (2004a) A spatial non-dynamic LQG controller: part 1, application to adaptive optics. In: Proceedings of the IEEE conference on decision and control, Atlantis, Paradise Island, Bahamas, USA, pp 3326–3332

Wiberg DM, Max CE, Gavel DT (2004b) A spatial non-dynamic LQG controller: part 2, theory. In: Proceedings of the IEEE conference on decision and control, Atlantis, Paradise Island, Bahamas, USA, pp 3333–3338

Zawadzki RJ, Choi SS, Fuller AR, Evans JW, Hamann B, Werner JS (2009) Cellular resolution volumetric in vivo retinal imaging with adaptive optics optical coherence tomography. Opt Express 17(5):4084–4094

Zawadzki RJ, Jones SM, Pilli S, Balderas-Mata S, Kim DY, Olivier SS, Werner JS (2011) Integrated adaptive optics optical coherence tomography and adaptive optics scanning laser ophthalmoscope system for simultaneous cellular resolution in vivo retinal imaging. Biomed Opt Express 2(6):1674–1686

Zhang Y, Rha J, Jonnal R, Miller DT (2005) Adaptive optics parallel spectral domain optical coherence tomography for imaging the living retina. Opt Express 13:4792–4811

Zhang Y, Poonja S, Roorda A (2006) MEMS-based adaptive optics scanning laser ophthalmoscopy. Opt Lett 31:1268–1270

Zou W, Qi X, Burns SA (2011) Woofer-tweeter adaptive optics scanning laser ophthalmoscopic imaging based on Lagrange-multiplier damped least-squares algorithm. Biomed Opt Express 2(7):1986–2004

Chapter 3
Magnetic Fluid Deformable Mirrors

Contents

3.1 Brief History of the Development of MFDMs

The continuing improvements in optical mirror technology have resulted in significant progress in the astronomical imaging in the past few decades. Telescopes with mirrors up to 60 m in diameter are currently being planned. Most of the new generation of telescopes would use solid primary mirrors. However, the cost to produce such large solid mirrors is very high. It is therefore worthwhile to explore alternative technologies to produce different mirrors at a reduced cost. Liquid mirrors have been proposed as an alternative to conventional solid mirrors. The idea of using a liquid mirror is more than a century old (Wood 1909). However, it was only in the late 1980s that the concept was put into practical use. The exceptionally good surface qualities of liquids led NASA to use liquid-mirror telescopes (LMTs) for several astronomical applications (Cabanac and Borra 1998; Hickson and Mulrooney 1997; Wuerker 1997). Liquid mirrors take advantage of the fact that liquid surfaces are very smooth that need not to be polished, and that the support system is just a simple container, which results in inexpensive perfect optical quality surfaces. For example, a rotating liquid takes the parabolic shapes that could give inexpensive high-quality parabolic mirrors. Most liquid-mirror telescopes on Earth have used mercury. Mercury remains molten at room temperature, and it reflects

Z. Wu et al., *Modeling and Control of Magnetic Fluid Deformable Mirrors for Adaptive Optics Systems*, DOI 10.1007/978-3-642-32229-7_3,

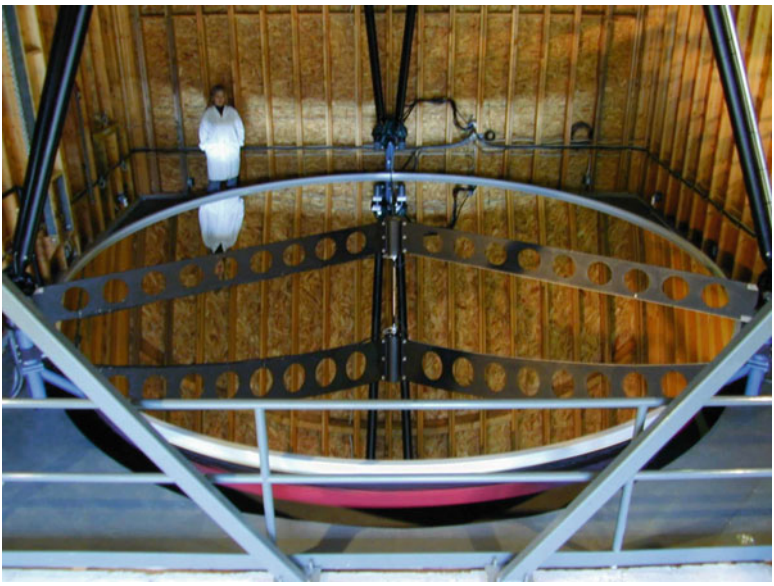


Fig. 3.1 A 6-m mercury-based liquid mirror for the Large Zenith Telescope at the University of British Columbia (P. Hickson, University of British Columbia)

about 75% of incoming light, almost as good as silver. The NASA Orbital Debris Observatory was the first to operate a liquid-mirror telescope with a 3-m diameter for an extended period of time to observe space debris. It operated continuously for 8 years and yielded useful astronomical research results. The biggest liquid-mirror telescope on Earth, the Large Zenith Telescope operated by the University of British Columbia in Canada (Fig. 3.1), is 6 m across—a diameter 20% larger than the famous 200-in. mirror of the Hale telescope at Palomar Observatory in California. When completed in 2005, the liquid-mirror telescope cost less than $1 million to build—only a few percent the cost of a solid-mirror telescope with the same diameter. However, the rotating liquid mirrors can only produce parabolic shapes and will also be limited by winds generated by the rotation of the mirror itself.

The other progress of LMTs led to the idea of deforming liquid mirrors to obtain mirror shapes other than a parabola. Shutter and Whitehead (1994) proposed to use magnets to reshape the classical mercury-based parabolic mirror into a sphere. They used an amalgam of iron and mercury to influence the liquid bulk by applying a magnetic field. Ragazzoni and Marchetti (1994) proposed the first adaptive liquid mirror that used an electrically conducting metallic reflective liquid that could carry a current and was shaped by a number of current-carrying coils. However, mercury suffers from serious limitations when subjected to a magnetic field to deform its surface shape. First, it is very difficult to obtain a stable mercury-based ferrofluid due to its chemistry properties (Rosensweig 1997). Second, the high density of mercury necessitates very large electromagnetic force to achieve even minimum required surface deflections. This requirement of large force means prohibitively

large currents running through the electromagnetic coils used to generate the magnetic field. Attempts to deform mercury by running currents through the liquid suffered from yet another drawback, that is, joule heating of the liquid (Borra 2009).

Confronted with these limitations of mercury, researchers came up with an innovative new technology that utilized the concept of ferrofluid deformation shaped with magnetic fields. Though the idea of these magnetic fluid deformable mirrors is quite recent, they have been appraised by a number of preliminary studies as a promising future technology (Brousseau et al. 2011, 2010; Parent and Thibault 2011; Parent et al. 2009; Iqbal et al. 2010a, b, 2009; Iqbal and Ben Amara 2008; Borra et al. 2008). Borra et al. (2008) have experimented with adaptive ferrofluidic mirrors coated with colloidal silver particles and shaped with magnetic actuators. Ferrofluids are ordinary liquids such as water or oil. Ferromagnetic nanoparticles coated with a chemical agent that prevents coalescence are then added into the liquid carry, which is thus rendered ferromagnetic so that its surface can be shaped by an externally applied magnetic field. The low density of carrier fluids such as water and oils offered a great advantage over mercury since the current requirements for the deformation of these liquids are orders of magnitude lower than those for mercury. Compared to mercury, a ferrofluid has a relatively poor reflective surface; however, this can be improved by spreading nanoengineered reflective colloidal films based on interfacial films of silver particles known as metal liquid-like films (MELLFs) on the surface of the liquid (Borra et al. 2008; Gingras et al. 2006). MELLFs combine the properties of metals and liquids, thus can be deformed and well adapted to applications in the field of liquid optics. These developments prompted a number of studies that explored the possibility of using the MFDMs in applications ranging from extremely large LMTs to ophthalmic AO systems. Their main advantages are that they are inexpensive and can produce very large deformations. The initial studies reported thus far show that these mirrors can easily provide tens of micrometers of surface deflection (Iqbal et al. 2009; Borra et al. 2008; Laird et al. 2006). It can be foreseen that ferrofluids could be used to make cost-competitive MFDMs with scalable dimensions and thousands of actuators capable of strokes ranging from nanometers to hundreds of microns. However, despite the initial success and promising future prospects, MFDMs have yet to see a practical application owing to some critical impediments as discussed in Sect. 3.3, which mainly motivated the research work presented in this book.

3.2 Principle of Operation

The free surface of a liquid follows an equipotential surface to a very high degree of precision. This characteristic of liquids can be exploited to deform their surface to any desired shape. The process of shaping the free surface may be controlled by varying one or more of the force fields that contribute toward the potential level. Magnetic fluids are a class of liquids which respond to variations in magnetic field, and hence their surface shape can be controlled by controlling the magnetic field

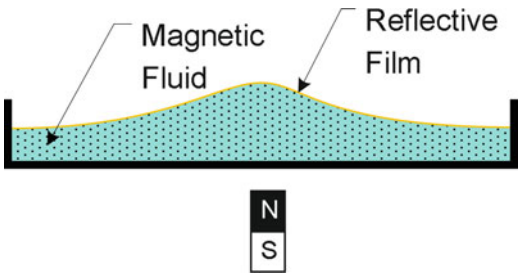


Fig. 3.2 The basic principle of a magnetic fluid deformable mirror

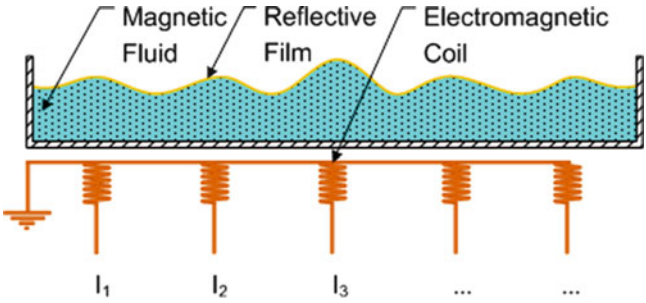


Fig. 3.3 Schematic diagram of a magnetic fluid deformable mirror

that they are exposed to. If a beam of light incident on the free surface of a magnetic fluid can be adequately reflected, the surface of the fluid can be utilized to control the shape of the wavefront of the reflected light, that is, the magnetic fluid surface can be used as a deformable mirror.

Figure 3.2 shows the working principle of an MFDM in its simplest form. The device comprises a thin layer of a magnetic fluid which, under the effect of gravity alone, features a planar free surface. When exposed to a magnetic field, such as the one created by a small magnet, the surface deforms to a new equipotential surface now supported by both gravity and the magnetic field. A more practical device is obtained by using electromagnetic coils to generate the magnetic field, which can be locally varied by controlling the currents applied to the individual coils. The varying magnetic field in turn affects the equipotential surface, thus providing a means to control the shape of the magnetic fluid surface. The schematic diagram of one such device is shown in Fig. 3.3. The deformation of a ferrofluid in a static magnetic field is determined by the equilibrium between gravity and the force on the liquid arising from the applied magnetic field **B**. The deformation amplitude can be approximated by (Brousseau et al. 2006)

$$h = \frac{(\mu_r - 1)}{2\mu_r \mu_0 \rho g} \left(|\mathbf{B} \cdot \hat{\mathbf{n}}|^2 + \mu_r |\mathbf{B} \times \hat{\mathbf{n}}|^2 \right)$$

where ρ and μ_r are the density and the relative permeability of the ferrofluid, $\hat{\mathbf{n}}$ is a unit vector normal to the surface of the liquid. The source of the external magnetic field **B** can come from permanent magnets or current-carrying wires and

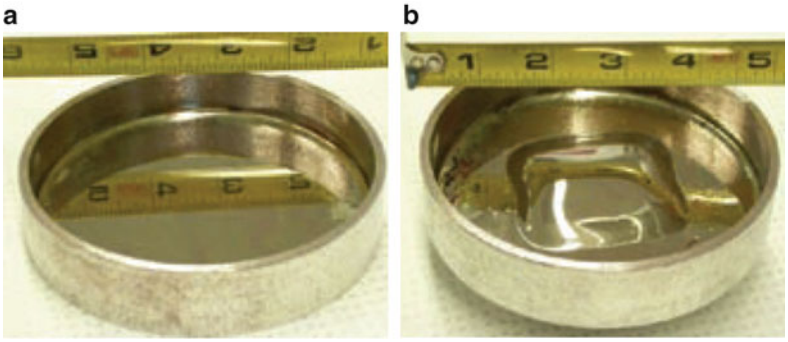


Fig. 3.4 MELLF-coated ferrofluid produced by Laval University. (**a**) Ferrofluid without a magnetic field. (**b**) Ferrofluid deformed by the magnetic field of a permanent magnet located underneath the container (Borra et al. 2009)

is the sum of all the individual magnetic fields produced by the actuators. It can be seen that the above equation is nonlinear with respect to the external magnetic field and depends on the vectorial addition of each magnetic field component produced by the corresponding actuators. This means that linear control algorithms used in conventional AO systems are not suitable to use with MFDMs due to the nonlinear response of the MFDM surface shape to the applied magnetic field. Furthermore, the equation shows that the surface deformations are proportional to the square of the applied magnetic field; therefore, only unidirectional deformations can be produced on the mirror surface, which means the actuators can only push not pull the mirror surface.

Typically, magnetic fluids have poor reflectivity and hence cannot be used as a functioning mirror. This limitation can be successfully overcome by using films of reflective particles thin enough that they pose no significant challenge to the deformation of the fluid but greatly enhance the reflectance of the deformable surface. Figure 3.4 shows an MELLF-coated magnetic fluid produced by Borra et al. (2009) in Laval University. The figure compares the magnetic fluid without a magnetic field to the same surface after it was deformed by several millimeters through the application of magnetic field from a permanent magnet located under the container. The relevant aspects of the magnetic fluids used in MFDMs and the reflective films are described in the following paragraphs.

3.2.1 Magnetic Fluids

3.2.1.1 Description

Magnetic fluids—also called *ferrofluids*—were originally discovered in the 1960s at the NASA Research Center, where scientists were investigating different possible methods of controlling liquids in space as the basis for a spacecraft's attitude control

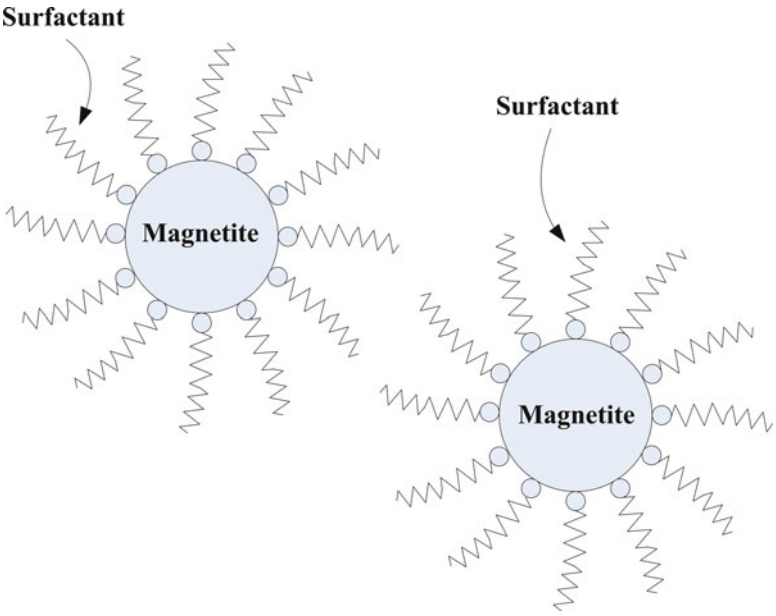


Fig. 3.5 Schematic representation of magnetic particles coated with surfactants

system. The benefits of using a magnetic fluid arise from the fact that the location of the fluid could be precisely controlled through the application of a magnetic field, and by varying the strength of the field, the fluid could be forced to flow, thus change the angular momentum and influence the rotation of the spacecraft. Magnetic fluids are stable colloidal suspensions of nano-sized, single-domain ferri-/ferromagnetic particles. When subjected to a magnetic field, the suspended particles affect the flow properties of the carrier fluid. As mentioned earlier, one of these effects is the variation of the fluid free surface shape, which has been exploited to develop deformable mirrors.

Depending on the type of the carrier fluid used, the magnetic fluids can be broadly categorized as water-based and oil-based systems. In most technological applications of magnetic fluids, the magnetic particles are obtained from one of a number of different ferrites. Researchers have prepared ferrofluids containing small particles of ferromagnetic metals, such as cobalt and iron, as well as magnetic compounds, such as manganese zinc ferrite. But by far, the most commonly used ferrites are magnetite (Fe_3O_4) and maghemite (γ-Fe_2O_3). The typical size of these particles is 10 nm in diameter. This is small enough for thermal agitation to disperse them evenly within a carrier fluid and for them to contribute to the overall magnetic response of the fluid.

Since the individual magnetic particles are in a permanent state of saturation magnetization, there exists a strong magnetostatic attraction between the particles and, as a result, the particles tend to agglomerate in the carrier fluid. To avoid the agglomeration, the magnetic particles are coated with surfactants (Fig. 3.5), which

introduce an entropic repulsion between the particles resulting in stable suspensions of the particles. The magnetic attraction of nanoparticles is weak enough that the surfactant's van der Waals force is sufficient to prevent magnetic clumping or agglomeration. Ferrofluids usually do not retain magnetization in the absence of an externally applied field and thus are often classified as superparamagnets. In some applications, the repulsive mechanism is achieved by charging the surface of the particles, which produces an electrostatic repulsion. Ideal ferrofluids are stable, which means that the magnetic particles do not agglomerate or could separate even in extremely strong magnetic fields. However, the surfactant tends to break down over years; therefore, the magnetic particles may eventually agglomerate and separate out, and the fluid will no longer respond to the applied magnetic field.

3.2.1.2 Properties

There exists a vast selection of fluids which may be utilized as carriers in the magnetic fluids. Similarly, a considerable degree of freedom may be exercised when choosing the magnetic particles and surfactant material properties. The choice of materials allows for a very wide range of properties that can be built into a desired magnetic fluid. Density, viscosity, surface tension, magnetic permeability, and saturation magnetization are some of the properties of a magnetic fluid that can be varied according to the requirements of an intended application. The response of an MFDM depends on the properties of the magnetic fluid selected for the application. The surface deflection resulting from applying a given magnetic field is a function of the fluid properties such as density, magnetic permeability, and surface tension.

Owing to the incompressible nature of the commonly used carrier fluids, most of the magnetic fluids can be characterized as incompressible. However, when exposed to a high enough magnetic field, they exhibit a unique behavior showing compression of the bulk fluid. This phenomenon is called *magnetostriction* which is a property of ferromagnetic materials that causes them to change their shape or dimensions during the process of magnetization.

When exposed to a uniform, vertical magnetic field, the magnetic fluids show no response at low magnetic flux densities. However, if the field strength exceeds a certain value (17 mT) (Gollwitzer et al. 2007), a well-known instability results in patterns of spontaneous protrusions of the free surface of the fluid and forms a regular pattern of peaks and valleys. This phenomenon is called *Rosensweig instability* (Cowley and Rosensweig 1967). The instability is driven by the magnetic field and can be explained by considering the potential energy of the system (Maxwell et al. 2008). It is well known that the shape of the fluid minimizes the total energy of the system. From the point of view of magnetic energy, the magnetic field is concentrated in the peaks; therefore, the magnetic field lines prefer to run through the fluid, and they try to ride the spikes of fluid out into space as far as possible. Meanwhile, the formation of peaks and valleys is resisted by gravity and surface tension. It costs energy to move fluid out of the valleys and up into the

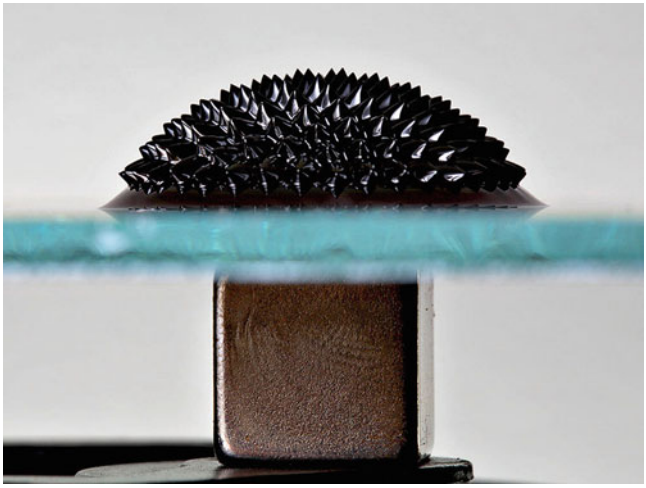


Fig. 3.6 A ferrofluid showing Rosensweig instability in a magnetic field produced by a magnet beneath the container (Maxwell et al. 2008)

spikes to increase the surface area of the fluid. When the magnetic field is strong enough, the protrusions of the fluid surface will form under the critical magnetic field strength, where the reduction in magnetic energy outweighs the increase in surface and gravitation energy terms. A more detailed mathematical analysis of the ferrohydrodynamic instability of magnetic fluids can be found in Rosensweig (1997). Figure 3.6 shows a magnetic fluid with spikes on the free surface caused by the Rosensweig instability in a strong magnetic field produced by a magnet beneath the container. The field strength at which the Rosensweig instability appears depends on the physical parameters of the ferrofluid and the geometry of the magnetic field. For typical magnetic fluids, the Rosensweig instability occurs around 80G (17 mT) when the applied magnetic vector is perpendicular to the fluid surface.

3.2.1.3 Synthesis

The magnetic fluids can be synthesized using a number of different methods. The earlier approach was to grind micro-sized magnetite particles in a ball mill in the presence of a surfactant material and a carrier fluid until the particle reaches the desired nanometer range size (Papal 1965). However, this process usually takes a very long time. Magnetic fluids are now produced more conveniently by chemical methods involving the coprecipitation of metal salts in aqueous solution using a base. A typical chemical process employed in the synthesis of magnetic fluids is as follows:

$$2Fe^{3+} + Fe^{2+} + 8OH^{-} = FeOFe_2O_3 + 4H_2O$$

For example, ferrofluids containing magnetite can be prepared by combining the appropriate amounts of an Fe(II) salt and an Fe(III) salt in basic solution, a combination that causes the mixed valence oxide, Fe_3O_4, to precipitate from solution:

$$2FeCl_3 + FeCl_2 + 8NH_3 + 4H_2O \rightarrow Fe_3O_4 + 8NH_4Cl$$

Besides the coprecipitation method, numerous other methods can be used to obtain both ferrite- and metal-based magnetic fluids (Charles 2002).

The particles of magnetite must remain small in order to remain evenly suspended in the liquid medium. However, for the small nanoparticles, magnetic and van der Waals interactions will cause them to agglomerate, and some approaches should be used to prevent the particles from agglomerating. Though thermal motion of magnetite particles smaller than 10 nm in diameter is sufficient to prevent agglomeration due to magnetic interactions, the van der Waals attraction between two particles becomes strongest when the particles approach each other at close distances. Therefore, the efficient method to prevent magnetite particles agglomerating due to van der Waals and magnetic forces is necessary by keeping the particles well separated. This separation can be accomplished by adding a surfactant to the liquid medium. The surfactants can generate either steric or electrostatic repulsions between the magnetic particles. The surfactants used to coat the nanoparticles include oleic acid, tetramethylammonium hydroxide, citric acid, and soy lecithin. These surfactants prevent the nanoparticles from clumping together, ensuring that the particles do not form aggregation that become too heavy to be held in suspension by Brownian motion. In this case, the magnetic particles in the ferrofluid do not settle out, even when exposed to a strong magnetic or gravitational field. For example, oleic acid can be used for oil-based ferrofluids as a surfactant that produces steric repulsions. The oleic acid surfactant is a long-chain hydrocarbon with a polar head that is attracted to the surface of the magnetite particle; thus, a surfactant coating is formed on the surface. The long chains of the tails act as a repellent cushion and prevent the close approach of other magnetite particles. Tetramethylammonium hydroxide can also be used as a surfactant that produces electrostatic repulsion in an aqueous medium. The hydroxide ions are attracted to the surface of each magnetite particle, forming a negatively charged layer at the magnetite surface. The tetramethylammonium cations are attracted to the negatively charged layer, forming a positive layer. When magnetite particles approach each other, the repulsions between their positively charged layers keep them from getting too close.

Surfactants are useful in prolonging the settling rate in ferrofluids; they also effect the fluid's magnetic properties, specifically the fluid's magnetic saturation. The addition of surfactants decreases the packing density of the ferroparticles while in its activated state, thus decreasing the fluids' on-state viscosity. Since the on-state viscosity is a primary fluid property for the majority of their commercial and industrial applications, a compromise must be met when considering on-state viscosity versus the settling rate of a ferrofluid by properly choosing and synthesizing the surfactants in the ferrofluid.

3.2.1.4 Applications of Magnetic Fluids

Since their introduction in the early 1960s, magnetic fluids have seen applications ranging from sink-float systems to drug delivery systems. Some of the well-known applications include liquid seals, dampers for audio speakers, and heat-transfer agents. A ferrofluid can behave as a liquid O-ring where a rotating shaft enters either a low- or high-pressure chamber. The ferrofluid is held in place by permanent magnets and forms a tight seal, eliminating most of the friction produced in a traditional mechanical seal. Ferrofluid seals are also used in high-speed computer disk drives to eliminate harmful dust particles or other impurities that can cause the data reading heads to crash into the disks. In a loudspeaker, the magnetic field induced by the electric energy causes the coil to vibrate and thus produces sound and heat. Bathing the electric coil in a ferrofluid, which is held in place by circular permanent magnets, dampens unwanted resonances and also provides a mechanism to dissipate heat from excess energy supplied to the coil, thus leads to an overall improved sound quality. Recently, ferrofluids are also being investigated for applications in biomedical applications, MEMS, and nanotechnology (Nakano et al. 2008; Zahn 2001). Their use as MFDMs, particularly in ophthalmic AO systems, remains one of the most advanced emerging optical applications. Besides ophthalmology, their potential use in extremely large telescopes is also widely anticipated (Borra et al. 2009, 2006).

3.2.2 MELLFs

Magnetic fluids typically show low reflectance to light. The reflectivity of EFH1 produced by Ferrotec Corporation, for instance, is only 4% (Brousseau et al. 2006). Due to this limitation, they cannot be used directly as mirrors. Compounds commonly known as metal liquid-like films (MELLFs) present a solution. MELLFs are homogeneous films of nanoparticles that exhibit reflective properties like liquid metals but are thin enough not to have any significant effect on the deformation of the substrate magnetic fluid. Typically, the nanoparticles used as MELLFs are silver, gold, and aluminum.

MELLFs were first reported by Yogev and Efrima (1988). They developed MELLFs by drop-wise addition of a reducing agent to a two-phase solution of a chlorinated organic solvent and an aqueous solution of ammonium silver nitrate, anisic acid, and an appropriate surfactant. The metallic silver particles are produced in the aqueous phase and eventually settle to the interface of the two phases forming a reflective film. These films are characterized as having an approximate thickness of 25 nm and show a reflectivity comparable to liquid mercury.

Since the inception of MELLFs in 1988, various other synthesis methods have been reported in the literature. Gorden et al. (1989) presented an alternative method which used an aqueous silver colloid to which an appropriate ligand is added

resulting in extraction of the reflective nanoparticles to the interface. Though having similar optical and physical properties, these films are chemically different from those presented in Yogev and Efrima (1988).

Over the years, several improvements to the MELLFs described above have been reported. However, these basic MELLFs also suffer from certain critical limitations. They are compatible with water-based magnetic fluids but not with the oil-based ones. Unfortunately, the water-based magnetic fluids have shortcomings of their own; they evaporate over time, and thus the properties of the fluid keep changing, eventually drying up altogether. Recently, magnetic fluids with glycol as carrier fluid have been developed by Dery et al. (2008), and this new type of magnetic fluids has been reported to be fully compatible with MELLFs they produced. The preparation of metal liquid-like films of silver nanoparticles is stabilized by thiolate surface ligands. These surface films, composed of particles with diameters of about 100 nm, are highly reflective and are employed in the fabrication of liquid mirrors. A number of different thiols could be considered as stabilizing ligands, including alkanethiols, aromatic thiols, and dithiols. Under identical preparation conditions, they lead to the spontaneous formation of reflective surface silver films which have a very high percentage of light reflectivity. Stable ferrofluids composed of positively charged magnetic iron oxide nanoparticles coated with 2-[2-(2-methoxyethoxy) ethoxy] acetic acid (MOEEAA) are prepared in ethylene glycol. Nanoparticles coated with MOEEAA and dispersed in ethylene glycol remained stable in the presence of a magnetic field. It is reported that these MOEEAA-stabilized new ferrofluids exhibit a magnetic response that is equivalent to that found for corresponding citrate-stabilized particles and is fully compatible with the deposition of surface films of silver nanoparticles, and thus allows for the practical preparation of magnetically deformable liquid mirrors.

3.3 Current State of Research and Challenges

Since the introduction of the idea of an MFDM as an alternative to the existing wavefront correctors (Laird et al. 2003), a number of studies characterizing various features of the mirrors have been reported. An overview of the literature reveals promising capabilities as well as significant challenges facing this new technology. An exposition describing these capabilities and challenges is presented below. An analysis of the problems impeding the potential utilization of these mirrors in practical applications leads to the identification of research goals presented in this book. Keeping in view the intended immediate application of these mirrors in ophthalmic AO systems, the application-specific requirements that these mirrors should meet are also stated in the following.

Important features of the MFDMs found by the studies conducted thus far are as follows:

- *Stroke.* The most promising aspect of the MFDMs is their ability to provide large deflections of their deformable surface. They are reported to have surface

deflections easily exceeding 20 μm (Brousseau et al. 2006). Recently, MFDM offering hundreds of microns of surface deflections have been discussed for their application in large astronomical telescopes (Borra et al. 2008). The feature—referred to as stroke in conventional deformable mirrors—shows an unparalleled advantage over the conventional micro-machined deformable mirrors which feature stroke lengths of a few microns.

- *Cost.* The simple design and expendable nature of MFDMs render them remarkably more economical than any other known type of wavefront correctors. Based on similar components used in the electronic industry, Laird et al. (2006) projected a cost only per channel of $2–$5. For large mirror telescope, it is estimated that for a given diameter, the cost of the magnetic liquid mirror is only about two orders of magnitude smaller than a solid glass mirror.
- *Speed.* Initial studies reported the magnetic fluid wavefront correction speeds between 10 and 50 Hz only and show the main disadvantage of their significantly lower bandwidth compared to solid deformable mirrors (Brousseau et al. 2007; Laird et al. 2006). However, recently Parent et al. (2009) have disclosed some unintuitive features, which allow the correction speed to be enhanced beyond 900 Hz. By increasing the viscosity of the magnetic fluids, along with an overdrive technique, the response time of magnetic fluid deformable mirror can be improved dramatically. Experiments have shown a bandwidth of the order of 1 kHz with a 494-cP-viscosity ferrofluid. However, the current results are limited to small amplitudes of the order of a few microns and need to be further improved for large amplitudes. On the other hand, the main application in ophthalmology may require a low bandwidth with respect to the property of dynamic wavefront aberrations in the eyes.
- *Vibrations.* The susceptibility of the magnetic fluid volume to mechanical vibrations is one of the most obvious concerns associated with the use of MFDMs. For applications that require the AO setup to be positioned on a moving platform, the mechanical vibrations as well as the strict requirement of keeping the MFDM surface horizontal remain the prohibiting factors. However, for applications such as ophthalmic AO systems which can afford static installation of the mirror, the effects of mechanical vibrations can be minimized. Firstly, like any other precision system, the optical setup should be mounted on vibration isolators. It has been found that the amplitudes of disturbances are less than 1/10 of a wave when a low-viscosity magnetic fluid layer (about 6 cP) with thickness of 2 mm is put on an optical table. Beyond that, Borra et al. (1992) assert that the effect of mechanical vibrations on a fluid mirror is minimal if the thickness of the fluid layer is limited to less than one millimeter. Increasing viscosity of the magnetic fluid is also reported to have minimizing impact on the mirror surface vibrations resulting from the ambient noise. Nonetheless, the most effective remedy is presented by the applied magnetic field itself. The magnetic field applied to control the surface profile of the mirror dampens the vibrations significantly (Brousseau et al. 2007).
- *Stability and Compatibility of Reflective Film.* A significant part of the research efforts expended thus far in the area of MFDMs has been directed at the

development of appropriate magnetic fluids and compatible MELLFs. As mentioned earlier, the previously known types of MELLFs were incompatible with the oil-based magnetic fluids. On the other hand, water-based magnetic fluids, which showed compatibility with the MELLFs, suffered from the drawback of evaporation over time (Thibault et al. 2006). The recent introduction of a new type of oil-based magnetic fluid (Dery et al. 2008), which are compatible with the silver MELLFs, is a significant advancement, which increased the likelihood of their practical application in the near future.

- *Control.* Difficulties in the control of the MFDM surface were first highlighted in Laird et al. (2006), which revealed that the deflections of the MFDM surface could not be obtained as a linear superposition of the influence of the individual actuator coils—a method routinely used in existing deformable mirrors. This unique feature of the MFDM has had serious implications in the control of these mirrors. Conventionally, control systems employed in AO systems have utilized static models of the response of their wavefront correctors. These models are referred to as the influence function matrix in the AO literature and the DC gain of the plant in the control community. The models are constructed by statically computing the mirror surface deflections—or to be exact, the wavefront shape—resulting from applying unit input to each of the actuators. The resulting wavefront shapes—also called *influence functions* of the individual actuators—are assembled into a matrix, the inverse of which is then utilized to control the deformable mirror in a closed-loop operation. Generally, cascading the inverse of the matrix thus obtained with a suitably tuned proportional or proportional-plus-integral controller gain provides actuator input updates for the closed-loop operation of the mirror.

  The assumption behind this method is that the wavefront corrector is a linear system, that is, the surface deflections and the resulting wavefront shapes at any point can be obtained using a linear combination of the influence functions of the individual actuators. As it was discovered that the assumption of linearity did not hold true in the case of MFDMs, the control methods based on the computation of influence function were rendered inapplicable to these mirrors. Notwithstanding these difficulties, some efforts have been made to control the MFDMs to generate desired wavefront shapes (Brousseau et al. 2007; Seaman et al. 2007). These efforts are based on a static model (Jones 1988) of the response of the MFDM surface, which describes the mirror surface deflections as a nonlinear function of the magnetic field applied to control the shape of the mirror. The use of this nonlinear model again precludes the possibility of directly using any linear control algorithms. In this backdrop, the closed-loop operation of the MFDMs is yet to be seen. A final look at the strengths and the challenges of MFDMs brings forward the control of the mirror as the most pressing requirement that needs to be met before they can be considered for a practical application. The research work presented in this book has been undertaken to facilitate the fulfillment of this requirement. In what follows, an analytical model that sufficiently represents

the dynamics of the MFDM will be developed. This model is then utilized to devise effective control algorithms, which will be implemented to control a prototype MFDM in a closed-loop system. Using an evocation prompted by the analytical work undertaken to develop a model of the mirror, we will devise a means to linearize the actual response of the mirror surface, thus easing the difficulties stemming from the nonlinearity in the MFDM model.

3.4 Performance Requirements for MFDMs in Ophthalmic AO Systems

Keeping in view the most immediate intended application of the MFDM, that is, an ophthalmic imaging AO system, the application-specific requirements that the proposed MFDM mirror should meet are stated as follows:

(1) The most stringent of the requirements of AO systems in ophthalmic systems is the depth of correction provided by the wavefront corrector, that is, the stroke length for the case of a deformable mirror. For a pupil large enough to provide the required spatial resolution, Doble and Williams (2004) argue that a least stroke length of ± 12 μm will be required to effectively compensate the ocular aberrations.
(2) Studies (Hofer et al. 2001; Fernandez et al. 2001) show that the significant dynamic component of the aberrations introduced by the human eye remains within a few Hertz. Using the rule of thumb that suggests a correction speed two times faster than the aberration dynamics, a wavefront corrector with more than 10 Hz bandwidth would be suffice for an ophthalmic AO system.
(3) Also, studies have shown that the contribution of the various spatial modes of aberrations in the eye decreases with the increasing order of the modes (Hofer et al. 2001). The improvements in image quality or vision acuity that can be attained by correction of the spatial modes higher than the tenth order become negligibly small. Although correction of modes as high as possible may be desirable, it also translates to a requirement of higher packing density of the actuators, which may not be viable due to increasing cost of development.
(4) Compactness is a highly desirable feature in the ophthalmic imaging systems intended mainly for clinical settings. As the dilated pupil size of the human eye ranges between 4 and 8 mm, an ideal deformable mirror to be used in an ophthalmic AO system should have a comparable size (Doble and Williams 2004). A deformable mirror larger than the pupil size requires optics to project the small pupil on the deformable mirror. The deformable mirrors larger than a few times the pupil size would require prohibitively large optical path lengths, thus compromising the compactness requirement.

In light of the above, the remaining chapters of this book address the following goals:

- *Model development.* Develop a comprehensive dynamic model of the response of an MFDM surface shape, which may be utilized in the design of a control system for the mirror.
- *Rectification of the nonlinearity in the MFDM response.* Devise means to rectify the problem of nonlinearity in the response of the MFMD surface.
- *Design of a Prototype MFDM and Closed-Loop AO Setup.* Design and develop a prototype MFDM keeping in view the requirements of ophthalmic AO system. Also, design and develop an experimental AO setup to use the prototype MFDM in a closed-loop system.
- *Controller design.* Design a controller for the surface shape of the mirror to be employed for the cancellation of the dynamically varying unknown aberrations.
- *Experimental validation and evaluation.* Experimentally validate the dynamic model of the mirror, and evaluate the performance of the closed-loop system using the prototype MFDM.

3.5 Summary

Liquid mirrors represent a promising technology. They offer an interesting alternative to the conventional deformable mirrors fabricated using thin plates or membranes. Large rotating mercury liquid mirrors have been used in observatories to yield perfect parabolic reflective surface and have delivered numerous scientific results. In order to generate more complex surface shapes other than the parabolic one, magnetic fluid deformable mirrors have been developed recently for a number of potential new optical applications. Compared with conventional solid mirrors, magnetic fluid deformable mirrors have the advantages of large strokes, low cost, and ease of scalability. The recent developments of MFDMs are very encouraging. In this chapter, the brief history of the development of MFDMs is introduced first. The working principle of MFDMs is then presented, including the description of their main structures and compositions. A review of the advantages of MFDMs and difficulties in controlling their surface shape due to the nonlinear behavior of the surface deformation as a function of the applied magnetic field is discussed. A number of studies characterizing various features of the mirrors have been reviewed, and the overview of the literature reveals promising capabilities as well as significant challenges facing this new technology. An analysis of the problems impeding the potential utilization of these mirrors in practical applications leads to the motivation of the outlined research topics discussed in this book.

References

Borra EF (2009) Liquid mirrors in engineering. Optics & Photonics News, pp 14–17, September 2009

Borra EF, Content R, Girard L, Szapiel S, Tremblay LM, Boily E (1992) Liquid mirrors: optical shop tests and contributions to the technology. Astrophys J 393:829–847

Borra EF, Brousseau D, Vincent A (2006) Large magnetic liquid mirrors. Astron Astrophys 446(1):389–393

Borra EF, Brousseau D, Cliche M, Parent J (2008) Aberration correction with a magnetic liquid active mirror. Mon Not R Astron Soc 391(4):1925–1930

Brousseau D, Borra EF, Ruel HJ, Parent J (2006) A magnetic liquid deformable mirror for high stroke and low order axially symmetrical aberrations. Opt Express 14:11486–11493

Brousseau D, Borra EF, Thibault S (2007) Wavefront correction with a 37–actuator ferrofluid deformable mirror. Opt Express 15:18190–18199

Brousseau D, Borra EF, Rochette M, Landry DB (2010) Linearization of the response of a 91-actuator magnetic liquid deformable mirror. Opt Express 18(8):8239–8250

Brousseau D, Drapeau J, Piché M, Borra EF (2011) Generation of Bessel beams using a magnetic liquid deformable mirror. Appl Opt 50:4005–4010

Cabanac R, Borra EF (1998) A search for peculiar objects with the NASA Orbital Debris Observatory 3-m Liquid Mirror Telescope. Astrophys J 509:309–323

Charles SW (2002) The preparation of magnetic fluids. In: Stefan Odenbach (ed) LNP 594, Springer-Verlag Berlin Heidelberg, pp 3–18

Cowley MD, Rosensweig RE (1967) The interfacial stability of a ferromagnetic fluid. J Fluid Mech 30(4):671–688

Dery JP, Borra EF, Ritcey AM (2008) Ethylene glycol based ferrofluid for the fabrication of magnetically deformable liquid mirrors. Chem Mater 20(20):6420–6426

Doble N, Williams DR (2004) The applications of MEMS technology for AO in vision science. IEEE J Sel Top Quantum Electron 10(3):629–635

Fernandez EJ, Iglesias I, Artal P (2001) Closed-loop adaptive optics in the human eye. Opt Lett 26:746–748

Gingras J, Dry JP, Yockell-Lelivre H, Borra E, Ritcey AM (2006) Surface films of silver nanoparticles for new liquid mirrors. Colloids Surf A Physicochem Eng Asp 279:79–86

Gollwitzer C, Matthies G, Richter R, Rehberg I, Tobiska L (2007) The surface topography of a magnetic fluid: a quantitative comparison between *experiment and numerical simulation*. J Fluid Mech 571:455–474

Gordon KC, McGarvey JJ, Taylor KP (1989) Enhanced Raman scattering from liquid metal films formed from silver solution. J Phys Chem 93:6814

Hickson P, Mulrooney MK (1997) University of British Columbia-NASA multi-narrowband survey. I. Description and photometric properties of the survey. Astrophys J Suppl 115:35–42

Hofer H, Artal P, Singer B, Aragon JL, Williams DR (2001) Dynamics of the eye's aberration. J Opt Soc Am A 18(3):497–505

Iqbal A, Ben Amara F (2008) Modeling and experimental evaluation of a circular magnetic-fluid deformable mirror. Int J Optomechatron 2(2):126–143

Iqbal A, Wu Z, Ben Amara F (2009) Closed-loop control of magnetic fluid deformable mirrors. Opt Express 17(21):18957–18970

Iqbal A, Wu Z, Ben Amara F (2010a) Mixed sensitivity H_∞ control of magnetic fluid deformable mirrors. IEEE/ASME Trans Mechatron 15(4):548–556

Iqbal A, Wu Z, Ben Amara F (2010b) A decentralized robust PID controller design for the shape control of a magnetic fluid deformable mirror. Int J Optomechatron 4(3):246–268

Jones TB (1988) Theory and application of ferrofluid seals. In: Berkovsky B (ed) Introduction to thermomechanics of magnetic fluids. Hemisphere, Washington, DC

Laird P, Bergamasco R, Berube V, Borra EF, Ritcey AM, Rioux M, Robitaille N, Thibault S, Lande Vieira da Silva Jr., Yockell-Lelivre H (2003) Ferrofluid-based deformable mirrors: a new approach to AO using liquid mirrors. In: Wizinowich PL, Bonaccini D (eds) Adaptive Optical System Technologies II, Proceedings of SPIE, vol. 4839, the International Society for Optical Engineering

Laird P, Caron N, Rioux M, Borra EF, Ritcey AM (2006) Ferrofluid adaptive mirrors. Appl Opt 45(15):3495–3500

Maxwell GF et al (2008) Ferrofluid. http://en.wikipedia.org/wiki/Ferrofluid, Wikipedia

Nakano M, Matsuura H, Dong-Ying J, Kumazawa T, Kimura S, Uozumi, Tonohata Y, Koide N, Noda K, Bian N, Akutsu P, Masuyama M, Makino K (2008) Drug delivery system using nano-magnetic fluid, ICICIC'08. In: 3rd international conference on innovative computing information and control, p 338, 18–20 June 2008

Papal SS (1965) Low viscosity magnetic fluid obtained by colloidal suspension of magnetic particles, US Patent, 3215572

Parent J, Thibault S (2011) Locally magnifying imager. Opt Express 19(6):5676–5689

Parent J, Borra EF, Brousseau D, Ritcey AM, Dery JP, Thibault S (2009) Dynamic response of ferrofluidic deformable mirrors. Appl Opt 48(1):1–6

Ragazzoni R, Marchetti E (1994) A liquid adaptive mirror. Astron Astrophys 283:L17–L19

Rosensweig RE (1997) Ferrohydrodynamics. Dover Publications, Mineola

Seaman A, Macpherson JB, Borra EF, Ritcey AM, Asselin D, Jerominek H, Thibault S, Campbell MC (2007) Hartmann-Shack measurements of ferrofluidic mirror dynamics. In: Photonics 575 North 2007, Proceedings of the SPIE 6796, no. 679603, Ottawa, Ontario, Canada

Shutter WH, Whitehead LA (1994) A wide sky coverage ferrofluid mercury telescope. Astrophys J 424:L139–L141

Thibault S, Brousseau D, Rioux M, Senkow S, Dery JP, Borra EF, Ritcey AM (2006) Nanoengineered ferrofluid deformable mirror: a progress report. In: Ellerbroek BL, Domenico BC (eds) Advances in adaptive optics II, Proceedings of SPIE 6272, no. 627231, Orlando, FL, USA

Wood RW (1909) The mercury paraboloid as a reflecting telescope. Astrophys J 29:164–176

Wuerker R (1997) Bistatic LMT lidar alignment. Opt Eng 36:1421–1424

Yogev D, Efrima S (1988) Novel silver metal liquid-like films. J Phys Chem 92:5754–5760

Zahn M (2001) Magnetic fluid and nanoparticle applications to nanotechnology. J Nanoparticle Res 3:73–78

Chapter 4
Analytical Model of a Magnetic Fluid Deformable Mirror

Contents

4.1 Analytical Model in Cartesian Geometry

Consider a MFDM diagram in Cartesian coordinates as shown in Fig. 4.1. Let the time-varying deflection of the mirror surface at any horizontal location (x, y) be denoted by $\zeta(x, y, t)$. The deflection is produced by the cumulative magnetic field generated by an array of electromagnetic microcoils located underneath the magnetic fluid layer. The magnetic field generated by any given coil j, $j = 1, 2, \ldots, J$, centered at the horizontal location (x_j, y_j), is idealized as that of a point source of magnetic potential $\psi_j(t)$. The idealization allows the electromagnetic field to be modeled as a current-free field and thus significantly simplifies the derivation of the model.

Z. Wu et al., *Modeling and Control of Magnetic Fluid Deformable Mirrors for Adaptive Optics Systems*, DOI 10.1007/978-3-642-32229-7_4,

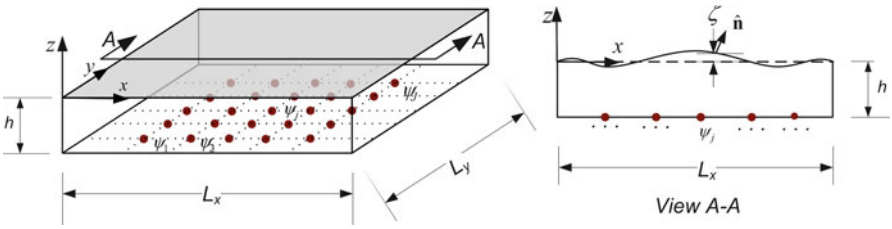


Fig. 4.1 Geometric representation of the magnetic fluid deformable mirror

4.1.1 Governing Equations

The deflection of the free surface of a magnetic fluid results from the fluid flow induced by the applied magnetic field. The fluid flow is governed by the fundamental principles of fluid dynamics appropriately modified to account for the effects of the magnetic field. The equations governing the fluid field, derived from the principles of conservation of mass and momentum, respectively, are as follows (Rosensweig 1997):

$$\nabla \cdot \mathbf{V} = 0 \tag{4.1}$$

$$\rho\left(\frac{\partial \mathbf{V}}{\partial t} + \mathbf{V} \cdot \nabla \mathbf{V}\right) = -\nabla\left(p + p_{\mathrm{s}} + p_{\mathrm{m}}\right) + \eta \nabla^2 \mathbf{V} + \rho \mathbf{g} + \mu_0 M \nabla H \tag{4.2}$$

where $\mathbf{V}$ is the velocity of the fluid; p, p_{s}, and p_{m} are, respectively, the thermodynamic, magnetostrictive, and fluid-magnetic pressures; ρ and η are the density and viscosity of the fluid; μ_0 is the magnetic permeability of free space; and M and H are the magnitudes of the magnetization vector $\mathbf{M}$ and the magnetic field vector $\mathbf{H}$, respectively.

The magnetic field itself is governed by Maxwell's equations. Since the magnetic field of the microcoils is idealized as that of point sources of magnetic potential located at the fluid domain boundary, a current-free electromagnetic field can be assumed. Using this assumption and further assuming that the displacement currents in the fluid are negligible, Maxwell's equations can be written as follows:

$$\nabla \times \mathbf{H} = 0 \tag{4.3}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{4.4}$$

where $\mathbf{B}$ is the magnetic flux density, which is related to the magnetic field $\mathbf{H}$ and the magnetization $\mathbf{M}$ by the following constitutive relationship:

$$\mathbf{B} = \mu \mathbf{H} = \mu_0 \left(\mathbf{H} + \mathbf{M}\right) \tag{4.5}$$

where μ is the magnetic permeability of the magnetic fluid. Assuming the magnetic fluid is linearly magnetized by the applied field, the magnetization vector $\mathbf{M}$ can be written as

$$\mathbf{M} = \chi \mathbf{H} \tag{4.6}$$

where $\chi = ((\mu / \mu_0) - 1)$ is considered to be a constant.

With the objective of obtaining the surface deflection $\zeta(x, y, t)$, the set of governing Eqs. (4.1), (4.2), (4.3), (4.4), (4.5), and (4.6) is augmented with two free-surface conditions. Firstly, the deflection ζ itself is related to the fluid motion by the following *kinematic condition*:

$$\frac{\partial \zeta}{\partial t} = V_z \text{ at } z = \zeta \tag{4.7}$$

where V_z is the vertical component of the fluid velocity $\mathbf{V}$ (see Sect. A.1). The other free-surface condition, referred to as the *surface dynamic equation,* is derived from the stress balance at the interface between the magnetic fluid and the air above it (see Sect. A.2). The condition describes a jump in the pressure states at the interface and can be written as follows:

$$p + \frac{\mu_0 \chi}{2}\left((\mathbf{H} \cdot \mathbf{H}) + \chi(\mathbf{H} \cdot \hat{\mathbf{n}})^2\right) = p^a + 2\sigma\kappa \text{ at } z = \zeta \tag{4.8}$$

where p is the fluid pressure immediately below the interface, p^a is the air pressure immediately above the interface, $2\sigma\kappa$ is the capillary pressure expressed as function of the coefficient of surface tension σ and the surface curvature κ, and $\hat{\mathbf{n}}$ is a unit vector directed normal to the surface. The second term on the left-hand side of (4.8) accounts for the magnetic effects—written in terms of magnetic field $\mathbf{H}$ immediately below the interface. In deriving (4.8), it has been assumed that the magnetization of the air is negligibly small.

4.1.2 Simplification of the Governing Equations

Keeping in view the actual flow conditions in the magnetic fluid layer, a number of approximations can be made in order to reduce the system of Eqs. (4.1), (4.2), (4.3), (4.4), (4.5), (4.6), (4.7), and (4.8) to a simpler form (see Sect. A.3). Firstly, if the fluid flow is considered to be incompressible as well as irrotational, then using a well-known vector identity, the velocity field $\mathbf{V}$ can be written in terms of a scalar potential $\Phi(x, y, z, t)$ as

$$\mathbf{V} = -\nabla\Phi \tag{4.9}$$

such that Φ obeys the Laplace equation

$$\nabla^2\Phi = 0 \text{ for} - h < z < \zeta \tag{4.10}$$

Since the applied magnetic field is not expected to induce any volume change in the fluid, the magnetostrictive pressure p_s can be ignored. Moreover, the assumption that the magnetization of the fluid depends on the magnetic field only allows the magnetic pressure term $-\nabla p_m$ and the magnetic body force $\mu_0 M \nabla H$ in (4.2) to cancel each other. If the magnetic fluid is considered to be nonviscous, the viscous force term $\eta\nabla^2\mathbf{V}$ also vanishes. Using these simplifications, the momentum Eq. (4.2) can be reduced to (Bashtovoi and Rosensweig 1993)

$$-\rho\frac{\partial\Phi}{\partial t} + p + \rho g z = 0 \tag{4.11}$$

Assuming a uniform magnetic permeability μ throughout the fluid domain and using the same vector identity as mentioned earlier, the magnetic field $\mathbf{H}$ can be written in terms of a scalar potential $\Psi(x, y, z, t)$ as

$$\mathbf{H} = -\nabla\Psi \tag{4.12}$$

such that Ψ satisfies the Laplace equation

$$\nabla^2\Psi = 0 \text{ for} - h < z < \zeta \tag{4.13}$$

Similarly, assuming a uniform magnetic permeability μ_0 in the air above the free surface of the fluid, the magnetic field in this domain too can be written as

$$\mathbf{H}^a = -\nabla\Psi^a \tag{4.14}$$

such that $\Psi^a(x, y, z, t)$ satisfies

$$\nabla^2\Psi^a = 0 \text{ for } z > \zeta \tag{4.15}$$

Using (4.11) (evaluated at $z = \zeta$) and (4.12), the surface dynamic Eq. (4.8) can be written as

$$-\rho\frac{\partial\Phi}{\partial t} + \rho g\zeta - \frac{\mu_0\chi}{2}\left(\nabla\Psi\cdot\nabla\Psi + \chi(\nabla\Psi\cdot\hat{\mathbf{n}})^2\right) + p^a + 2\sigma\kappa = 0 \text{ at } z = \zeta \tag{4.16}$$

The set of Eqs. (4.10), (4.13), (4.15), and (4.16) can be solved to obtain the four unknowns ζ, Φ, Ψ and Ψ^a. However, Eq. (4.16) above is nonlinear in Ψ, and, therefore, linear solution methods cannot be applied to this system of equations in its present form. This complication can be circumvented by introducing a large uniform magnetic field with a constant flux density B_0 superimposed on the input field generated by the array of microcoils. Without qualitatively affecting the resulting

surface shape, the presence of the large uniform field allows linearization of Eq. (4.16), thus greatly simplifying the solution process. It may be noted that the uniform vertical magnetic field itself does not cause any surface deflection unless it exceeds a critical value at which point a known instability takes place (Cowley and Rosensweig 1967).

In what follows, the magnetic fluid layer in an initial equilibrium state will be perturbed by the input magnetic field applied at the bottom of the layer. It is assumed that the applied field results in small perturbations of all the field quantities and hence small surface deflections. Making use of this assumption, the perturbation part of the reduced set of Eqs. (4.10), (4.13), (4.15), and (4.16) is extracted below and will be solved in the following section.

The magnetic fluid layer is initially exposed to the magnetic flux density B_0 and is in a state of equilibrium manifested by a perfectly flat surface. The initial state is characterized by

$$\begin{aligned}
\zeta &= 0 \\
\kappa &= 0 \\
\hat{\mathbf{n}} &= \hat{\mathbf{k}} \\
\mathbf{V} &= \mathbf{V}_0 = \mathbf{0} \\
\mathbf{H}_0 &= [0, 0, H_0]^{\mathrm{T}} = \left[0, 0, \frac{B_0}{\mu}\right]^{\mathrm{T}} \\
\mathbf{H}_0^a &= \left[0, 0, H_0^a\right]^{\mathrm{T}} = \left[0, 0, \frac{B_0}{\mu_0}\right]^{\mathrm{T}}
\end{aligned} \tag{4.17}$$

The fluid is then displaced by applying the input magnetic potential, which results in perturbations $\mathbf{v}$, $\mathbf{h}$, and $\mathbf{h}^a$ in the fluid velocity $\mathbf{V}_0$, the magnetic field $\mathbf{H}_0$ in the fluid, and the magnetic field $\mathbf{H}_0^a$ in the air, respectively. The resulting state is characterized by

$$\begin{aligned}
\hat{\mathbf{n}} &= -\frac{\partial \zeta}{\partial x}\hat{\mathbf{i}} - \frac{\partial \zeta}{\partial y}\hat{\mathbf{j}} + \hat{\mathbf{k}} \\
\kappa &= -\frac{1}{2}\left(\frac{\partial^2 \zeta}{\partial x^2} + \frac{\partial^2 \zeta}{\partial y^2}\right) \\
\mathbf{V} &= \mathbf{V}_0 + \mathbf{v} = \mathbf{v} \\
\mathbf{H} &= \mathbf{H}_0 + \mathbf{h} \\
\mathbf{H}^a &= \mathbf{H}_0^a + \mathbf{h}^a
\end{aligned} \tag{4.18}$$

where $\mathbf{h}$ and $\mathbf{h}^a$ are small in magnitude as compared to H_0 and H_0^a. The linearity of Laplace Eqs. (4.10), (4.13), and (4.15) allows the perturbations $\mathbf{v}$, $\mathbf{h}$, and $\mathbf{h}^a$ to be written as

$$\mathbf{v} = -\nabla\phi \tag{4.19}$$

$$\mathbf{h} = -\nabla\psi \tag{4.20}$$

$$\mathbf{h}^a = -\nabla\psi^a \tag{4.21}$$

such that

$$\nabla^2\phi = 0 \text{ for} - h < z < \zeta \tag{4.22}$$

$$\nabla^2\psi = 0 \text{ for} - h < z < \zeta \tag{4.23}$$

$$\nabla^2\psi^a = 0 \text{ for } z > \zeta \tag{4.24}$$

Using (4.17) and (4.18), the perturbation part of the surface dynamic Eq. (4.16) is extracted (see Sect. A.4) and is written as follows:

$$-\rho\frac{\partial\phi}{\partial t} + \rho g\zeta + \chi B_0\frac{\partial\psi}{\partial z} - \sigma\left(\frac{\partial^2\zeta}{\partial x^2} + \frac{\partial^2\zeta}{\partial y^2}\right) = 0 \text{ at } z = \zeta \tag{4.25}$$

4.1.3 Derivation of the Surface Response

The set of Eqs. (4.22), (4.23), (4.24), and (4.25), with the four unknowns ζ, ϕ, ψ, and ψ^a, describes the dynamics of the magnetic fluid mirror and will be solved to eventually find the mirror surface deflection ζ. We assume solutions of the following form:

$$\zeta\,(x, y, t) = \tilde{\zeta}(t)E\,(x, y) \tag{4.26}$$

$$\phi\,(x, y, z, t) = \tilde{\phi}\,(z, t)\,E\,(x, y) \tag{4.27}$$

$$\psi\,(x, y, z, t) = \tilde{\psi}\,(z, t)\,E\,(x, y) \tag{4.28}$$

$$\psi^a\,(x, y, z, t) = \tilde{\psi}^a\,(z, t)\,E\,(x, y) \tag{4.29}$$

where

$$E\,(x, y) = \mathrm{e}^{-i\left(k_x x + k_y y\right)} \tag{4.30}$$

k_x and k_y are the *mode numbers*, and it is understood that only the real part of the solution will be retained carriage.

The Laplace Eq. (4.22) can be solved using the assumed solution (4.27) and two boundary conditions on ϕ. The first boundary condition is derived from the free-surface kinematic condition (4.7), which can be written as

$$-\frac{\partial \phi}{\partial z} = \frac{\partial \zeta}{\partial t} \text{ at } z = \zeta \tag{4.31}$$

Considering that the surface deflection ζ is small in magnitude, condition (4.31) is written at $z = 0$ instead of $z = \zeta$. The second condition is deduced from the physical consideration that there cannot be any flow across the solid surface at the bottom of the layer, which implies

$$-\frac{\partial \phi}{\partial z} = 0 \text{ at } z = -h \tag{4.32}$$

The resulting solution is as follows (For the detailed solutions of the Laplace Eqs. (4.22), (4.23), and (4.24), see Sect. A.5):

$$\phi\left(x, y, z, t\right) = -\frac{1}{k}\frac{\cosh\left(k\left(z+h\right)\right)}{\sinh\left(kh\right)}\frac{\mathrm{d}\tilde{\zeta}(t)}{\mathrm{d}t}E\left(x, y\right) \tag{4.33}$$

where

$$k = \sqrt{k_x^2 + k_y^2} \tag{4.34}$$

and, depending on k_x and k_y, an infinite number of solutions can be obtained for ϕ. Equations (4.23) and (4.24) are simultaneously solved using the following magnetic field boundary conditions:

$$\hat{\mathbf{n}} \times \left(\mathbf{H} - \mathbf{H}^a\right) = 0 \quad \text{at } z = \zeta \tag{4.35}$$

$$\hat{\mathbf{n}} \cdot \left(\mathbf{B} - \mathbf{B}^a\right) = 0 \quad \text{at } z = \zeta \tag{4.36}$$

$$\lim_{z \to \infty} \psi^a < \infty \tag{4.37}$$

$$\psi\left(x, y, z, t\right) = \sum_{j=1}^{J} \psi_j(t)\delta^2\left(x - x_j\right)\left(y - y_j\right) \quad \text{at } z = -h \tag{4.38}$$

The standard boundary condition (4.35) states that, at the interface of the fluid and the air, the tangential components of the magnetic fields in the two media are equal. Similarly, condition (4.36) signifies that the normal components of the magnetic flux densities in the two media are also equal at the interface. Again,

keeping in view the small magnitude of ζ, these two conditions are written at $z = 0$ instead of $z = \zeta$. Condition (4.37) implies that the magnetic field in the region above the free surface remains bounded. Condition (4.38) specifies the magnetic field resulting from the input magnetic coils as point sources of magnetic potential applied at the bottom of the fluid layer. Using (4.35), (4.36), and (4.37), the following solutions are obtained for ψ^a and ψ, respectively:

$$\psi^a(x, y, z, t) = A(t)\mathrm{e}^{-kz}E(x, y) \tag{4.39}$$

$$\psi(x, y, z, t) = \left(A(t)\left(\cosh(kz) - \frac{\mu_0}{\mu}\sinh(kz)\right) - \frac{\chi}{\mu}B_0\tilde{\zeta}(t)\cosh(kz)\right)E(x, y) \tag{4.40}$$

where $A(t)$ is the integration constant that will be later determined using (4.38).

Application of the condition that the components of velocity and magnetic field normal to the container walls must be zero yields the following mode shapes:

$$E(x, y) = \cos(k_x x)\cos(k_y y) \tag{4.41}$$

and the characteristic equations

$$\sin k_x x = 0 \quad \text{at } x = L_x \tag{4.42}$$

$$\sin k_y y = 0 \quad \text{at } y = L_y \tag{4.43}$$

The conditions (4.42) and (4.43) can be satisfied by an infinite number of discrete values of k_x and k_y, which can be written in the series form as

$$k_x = \frac{(m-1)\pi}{L_x} \quad \text{for } m = 1, 2, 3, \ldots$$
$$k_y = \frac{(n-1)\pi}{L_y} \quad \text{for } n = 1, 2, 3, \ldots$$

Now, applying the input magnetic potential (4.38) as a boundary condition on ψ at $z = -h$ and making use of the orthogonality property of the mode shapes, the unknown constant $A(t)$ can be determined for each mode as

$$A_{mn}(t) = \frac{1}{\cosh(k_{mn}h) + (\mu_0/\mu)\sinh(k_{mn}h)} \times \left(\frac{\chi}{\mu}B_0\cosh(k_{mn}h)\,\tilde{\zeta}_{mn}(t) + \frac{c_m c_n}{L_x L_y}\sum_{j=1}^{J}\psi_j(t)\cos\left(k_m x_j\right)\cos\left(k_n y_j\right)\right) \tag{4.44}$$

where the subscripts m, n, and mn have been affixed to show that the mode numbers k_x, k_y, and k, as well as the variables dependent on them, can take only discrete values and

$$c_m = \begin{cases} 1 & \text{for } m = 1 \\ 2 & \text{for } m > 1 \end{cases}$$

$$c_n = \begin{cases} 1 & \text{for } n = 1 \\ 2 & \text{for } n > 1 \end{cases}$$

Now that the scalar potentials ϕ, ψ^a, and ψ have been determined as given in (4.33), (4.39), and (4.40), their substitution into the surface dynamic Eq. (4.25) gives

$$\rho \frac{1}{k_{mn}} \frac{1}{\tanh(k_{mn}h)} \frac{\mathrm{d}^2 \tilde{\zeta}_{mn}(t)}{\mathrm{d}t^2} + \rho g \tilde{\zeta}_{mn}(t) + \sigma k_{mn}^2 \tilde{\zeta}_{mn}(t) - \frac{\chi \mu_0}{\mu} B_0 k_{mn} A_{mn}(t) = 0 \tag{4.45}$$

Further substitution of (4.44) into (4.45), followed by its rearrangement, yields

$$\frac{\mathrm{d}^2 \tilde{\zeta}_{mn}(t)}{\mathrm{d}t^2} + \omega_{mn}^2 \tilde{\zeta}_{mn}(t) = f_{mn}(t) \tag{4.46}$$

where

$$\omega_{mn}^2 = g \tanh(k_{mn}h) k_{mn} + \frac{\sigma}{\rho} \tanh(k_{mn}h) k_{mn}^3 - \frac{\chi^2}{\mu} \frac{B_0^2}{\rho} \frac{\sinh(k_{mn}h)}{(\mu/\mu_0)\cosh(k_{mn}h) + \sinh(k_{mn}h)} k_{mn}^2$$

$$f_{mn}(t) = F_{mn} \sum_{j=1}^{J} \psi_j(t) E_{mn}^j$$

$$F_{mn} = \chi \frac{B_0}{\rho} \frac{\tanh(k_{mn}h)}{(\mu/\mu_0)\cosh(k_{mn}h) + \sinh(k_{mn}h)} k_{mn}^2 \frac{c_m c_n}{L_x L_y}$$

$$E_{mn}^j = \cos(k_m x_j) \cos(k_n y_j)$$

The second-order differential Eq. (4.46) can be solved to obtain the generalized displacements $\tilde{\zeta}_{mn}(t)$. The generalized displacements and the corresponding mode shapes E_{mn}, evaluated at any desired location (x_0, y_0), give the total surface displacement at the location as

$$\zeta(x_0, y_0, t) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \tilde{\zeta}_{mn}(t) E_{mn}(x_0, y_0) \tag{4.47}$$

The damping effect associated with the fluid viscosity may be introduced into the model by adding a factor of $\omega_{\mathrm{d}_{mn}} \left(\mathrm{d}\tilde{\zeta}/\mathrm{d}t\right)$ to the left-hand side of (4.46) (Blums et al. 1997), where

$$\omega_{d_{mn}} = 4\frac{\eta}{\rho}k_{mn}^2 \tag{4.48}$$

For convenience, the solution (4.47) is truncated to a finite number of modes such that $m = 1, 2, \ldots, M$ and $n = 1, 2, \ldots, N$. The dynamics of the system comprising (4.46) and (4.47), with the viscous damping term added to the former, is represented in state-space form as follows:

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{B}'\mathbf{u}' \\ y &= \zeta = \mathbf{C}\mathbf{x} \end{aligned} \tag{4.49}$$

where $\mathbf{x} = \left[\tilde{\xi}_{11}, \dot{\tilde{\xi}}_{11}, \tilde{\xi}_{12}, \dot{\tilde{\xi}}_{12} \ldots, \tilde{\xi}_{MN}, \dot{\tilde{\xi}}_{MN}\right]^{\mathrm{T}}$ is the vector of the generalized displacements and the corresponding velocities, $\mathbf{u}' = \left[\psi_1\ \psi_2 \cdots \psi_J\right]^{\mathrm{T}}_{(J\times 1)}$ is the vector of input magnetic potentials,

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ -\omega_{11}^2 & -\omega_{d_{11}} & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & -\omega_{12}^2 & -\omega_{d_{12}} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & -\omega_{MN}^2 & -\omega_{d_{MN}} \end{bmatrix}_{(2MN\times 2MN)}$$

$$\mathbf{B}' = \mathbf{F}\mathbf{E}$$

where

$$\mathbf{F} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ F_{11} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & F_{12} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & F_{MN} \end{bmatrix}_{(2MN\times MN)}$$

$$\mathbf{E} = \begin{bmatrix} E_{11}^{1} & E_{11}^{2} & \cdots & E_{11}^{J} \\ E_{12}^{1} & E_{12}^{2} & \cdots & E_{12}^{J} \\ \vdots & \vdots & \ddots & \vdots \\ E_{MN}^{1} & E_{MN}^{2} & \cdots & E_{MN}^{J} \end{bmatrix}_{(MN \times J)}$$

and

$$\mathbf{C} = \begin{bmatrix} E_{11} & 0 & E_{12} & 0 & \cdots & E_{MN} & 0 \end{bmatrix}_{1 \times 2MN}$$

The state-space model (4.49) describes the dynamics of the surface shape of a rectangular MFDM in terms of deflections ζ of the mirror surface. It provides the deflections of the mirror surface as a function of the magnetic field of the electromagnetic coils idealized as the magnetic field of the point sources of potential, ψ_j, $j = 1, 2, \ldots, J$. The model can be used to simulate the response of the MFDM surface as well as to develop controllers to control the surface shape.

4.2 Analytical Model in Circular Geometry

Almost all optical systems are designed and described in circular geometry. It is, therefore, pertinent to extend the model presented in the previous section to circular geometry. Accordingly, this section describes the model of a MFDM in a cylindrical coordinate system.

As shown in Fig. 4.2, the MFDM is represented by a cylindrical horizontal layer of a magnetic fluid. The top free surface of the fluid layer is coated with a reflective film and serves as the deformable surface of the mirror. The shape of the mirror is described by the deflection $\zeta(r, \theta, t)$ of the deformed surface as measured with respect to a point (r, θ) in the horizontal plane. The deflection is produced by the cumulative magnetic field generated by an array of miniature electromagnetic coils located underneath the magnetic fluid layer. The magnetic field generated by any given coil j, $j = 1, 2, \ldots, J$, centered at the horizontal location (r_j, θ_j), is idealized as that of a point source of magnetic potential $\psi_j(t)$.

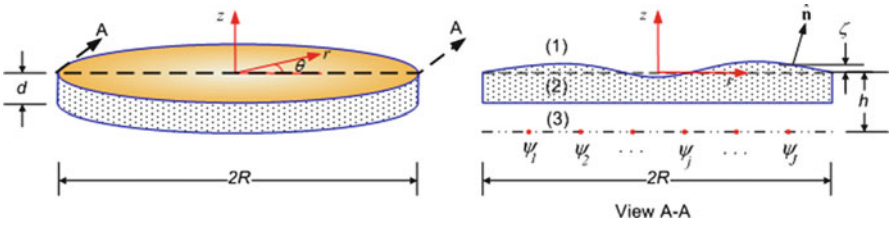


Fig. 4.2 Geometric representation of a circular magnetic fluid deformable mirror

4.2.1 Simplification of the Governing Equations

Since the system of governing Eqs. (4.1), (4.2), (4.3), (4.4), (4.5), (4.6), (4.7), and (4.8) is presented in operator notation, it is equally valid for the cylindrical coordinate system. Also, the simplifications given in Sect. 4.1.2 are independent of coordinate system and hence remain valid for the analysis presented in this section. Using these simplifications, the set of Eqs. (4.1), (4.2), (4.3), (4.4), (4.5), (4.6), (4.7), and (4.8) can be reduced to

$$\nabla^2 \Phi = 0, \quad -d \leq z \leq \zeta \tag{4.50}$$

$$\nabla^2 \Psi^{(i)} = 0, \quad i = 1, 2, 3 \tag{4.51}$$

$$-\rho \frac{\partial \Phi}{\partial t} + \rho g \zeta - \frac{\mu_0 \chi}{2} \left(\nabla \Psi^{(2)} \cdot \nabla \Psi^{(2)} + \chi \left(\nabla \Psi^{(2)} \cdot \hat{\mathbf{n}} \right)^2 \right) + p^{(1)} + 2\sigma\kappa = 0 \text{ at } z = \zeta \tag{4.52}$$

where Φ is a scalar potential that describes the fluid velocity vector $\mathbf{V}$ as follows:

$$\mathbf{V} = -\nabla \Phi \tag{4.53}$$

Note that in this case, the point sources of magnetic potential are placed underneath the magnetic fluid layer and are not restricted to the fluid domain boundary. Considering that the magnetic field extends into the space above and below the fluid layer, Maxwell's equations are applied to all three sub-domains marked in Fig. 4.2 as (1), (2), and (3). The scalar potentials $\Psi^{(i)}$, $i = 1, 2, 3$ describe the magnetic field vectors $\mathbf{H}^{(i)}$ in these sub-domains as follows:

$$\mathbf{H}^{(i)} = -\nabla \Psi^{(i)}, \quad i = 1, 2, 3 \tag{4.54}$$

The set of Eqs. (4.50), (4.51), and (4.52) can be solved to obtain the five unknowns ζ, Φ and $\Psi^{(i)}$, $i = 1, 2, 3$. However, the surface dynamic Eq. (4.52) is nonlinear in $\Psi^{(2)}$, which not only complicates the solution of the equations but also has implications in the eventual control of the fluid surface. To circumvent the nonlinearity in (4.52), we introduce a large uniform magnetic field with a constant flux density B_0 superimposed on the input field generated by the array of miniature coils. Below a critical level, at which a known instability occurs (Rosensweig 1997), the presence of the large uniform field by itself does not affect the surface shape.

The magnetic fluid mirror is considered to be in an initial equilibrium state characterized by a perfectly flat horizontal surface ($\zeta = 0$) and the uniform magnetic flux B_0 applied vertically across the whole fluid volume. The fluid is then perturbed by the magnetic field resulting from the input potentials $\psi_j(t)$, $j = 1, 2, \ldots, J$. Assuming small perturbations of all the field quantities and retaining only first-order

terms, the perturbation part of (4.50), (4.51), and (4.52) is extracted as given below:

$$\nabla^2\phi = 0, \quad -d \le z \le \zeta \tag{4.55}$$

$$\nabla^2\psi^{(i)} = 0, \quad i = 1,2,3 \tag{4.56}$$

$$-\rho\frac{\partial\phi}{\partial t} + \rho g\zeta + \chi B_0\frac{\partial\psi^{(2)}}{\partial z} - \sigma\left(\frac{\partial^2\zeta}{\partial r^2} + \frac{1}{r}\frac{\partial\zeta}{\partial r} + \frac{1}{r^2}\frac{\partial^2\zeta}{\partial\theta^2}\right) = 0 \text{ at } z = \zeta \tag{4.57}$$

where ϕ and $\psi^{(i)}$, $i = 1,2,3$ are the perturbation components of the scalar potentials Φ and $\Psi^{(i)}, i = 1,2,3$, respectively. Note that only the vertical component of the field generated by the miniature coils, that is, $\partial\psi^{(2)}/\partial z$, appears in the linearized surface dynamic Eq. (4.57) and that the component is being multiplied by the uniform vertical field B_0. These features will be referred to in the design of the prototype MFDM presented in Chap. 5.

4.2.2 *Derivation of the Surface Response*

Equations (4.55), (4.56), and (4.57), with the five unknowns ζ, ϕ, and $\psi^{(i)}$, $i = 1,2,3$, describe the dynamics of the magnetic fluid mirror and will be solved to eventually find the mirror surface deflection ζ. We assume the following separable solutions:

$$\zeta(r,\theta,t) = \tilde{\zeta}(t)R(r)\Theta(\theta) \tag{4.58}$$

$$\phi(r,\theta,z,t) = \tilde{\phi}(z,t)\,R(r)\Theta(\theta) \tag{4.59}$$

$$\psi^{(i)}(r,\theta,z,t) = \tilde{\psi}^{(i)}(z,t)\,R(r)\Theta(\theta), \quad i = 1,2,3 \tag{4.60}$$

The Laplace Eq. (4.55) is solved using the assumed solution (4.59) and the following two boundary conditions:

$$-\frac{\partial\phi}{\partial z} = \frac{\partial\zeta}{\partial t} \text{ at } z = \zeta \tag{4.61}$$

$$-\frac{\partial\phi}{\partial z} = 0 \text{ at } z = -d \tag{4.62}$$

The condition (4.61) is derived from the kinematic condition (4.7), and (4.62) has been deduced from the physical consideration that there cannot be any flow across

the solid surface at the bottom of the layer. Considering the surface deflection ζ to be small in magnitude, condition (4.61) is written at $z = 0$ instead of $z = \zeta$. The solution thus obtained is as follows:

$$\phi(r, \theta, z, t) = -\frac{1}{\lambda} \frac{\cosh(\lambda(z + d))}{\sinh(\lambda d)} \frac{\mathrm{d}\tilde{\zeta}(t)}{\mathrm{d}t} R(r)\Theta(\theta) \tag{4.63}$$

where λ is the separation constant, and where $\Theta(\theta)$ and $R(r)$ satisfy the following ordinary differential equations (Haberman 2003):

$$\frac{\mathrm{d}^2\Theta}{\mathrm{d}\theta^2} + m^2\Theta = 0 \tag{4.64}$$

$$\left(\frac{\mathrm{d}^2 R}{\mathrm{d}r^2} + \frac{1}{r}\frac{\mathrm{d}R}{\mathrm{d}r}\right) + \left(\lambda^2 - \frac{m^2}{r^2}\right) R = 0 \tag{4.65}$$

where m is yet another separation constant.

The magnetic field Eqs. (4.56) are simultaneously solved using the following boundary conditions:

$$\lim_{z \to \infty} \psi^{(1)} < \infty \tag{4.66}$$

$$\hat{\mathbf{n}} \times \left(\mathbf{H}^{(2)} - \mathbf{H}^{(1)}\right) = \mathbf{0} \quad \text{at } z = \zeta \tag{4.67}$$

$$\hat{\mathbf{n}} \cdot \left(\mathbf{B}^{(2)} - \mathbf{B}^{(1)}\right) = 0 \quad \text{at } z = \zeta \tag{4.68}$$

$$\hat{\mathbf{z}} \times \left(\mathbf{H}^{(3)} - \mathbf{H}^{(2)}\right) = \mathbf{0} \quad \text{at } z = -d \tag{4.69}$$

$$\hat{\mathbf{z}} \cdot \left(\mathbf{B}^{(3)} - \mathbf{B}^{(2)}\right) = 0 \quad \text{at } z = -d \tag{4.70}$$

$$\psi^{(3)}(r, \theta, z, t) = \sum_{j=1}^{J} \psi_j(t) \frac{1}{r} \delta\left(r - r_j\right) \delta\left(\theta - \theta_j\right) \quad \text{at } z = -h \tag{4.71}$$

Inequality (4.66) implies that the magnetic field in the region above the free surface remains bounded. Equations (4.67), (4.68), (4.69), and (4.70) are the standard boundary conditions of a magnetic field. Condition (4.71) specifies the magnetic field resulting from the input magnetic coils as point sources of magnetic potential applied in a plane located under the fluid layer at $z = -h$. Again, keeping in view the small magnitude of ζ, (4.67) and (4.68) are written at $z = 0$ instead of $z = \zeta$. Using (4.66), (4.67), (4.68), (4.69), and (4.70), the following solutions are obtained for $\psi^{(i)}$, $i = 1, 2, 3$:

$$\psi^{(1)}(r, \theta, z, t) = -A(t) \frac{\mu}{\mu_0} \mathrm{e}^{-\lambda z} R(r)\Theta(\theta) \tag{4.72}$$

$$\psi^{(2)}(r,\theta,z,t) = -\left(A(t)X(\lambda z) + \frac{\chi}{\mu}B_0\tilde{\zeta}(t)\cosh(\lambda z)\right)R(r)\Theta(\theta) \tag{4.73}$$

$$\psi^{(3)}(r,\theta,z,t) = \left(A(t)Y(\lambda z) - Z(\lambda z)\,B_0\tilde{\zeta}(t)\right)R(r)\Theta(\theta) \tag{4.74}$$

where $A(t)$ is the integration constant that will be later determined using (4.71) and

$$X(\lambda z) = \frac{\mu}{\mu_0}\cosh(\lambda z) - \sinh(\lambda z) \tag{4.75}$$

$$Y(\lambda z) = -\left(\frac{\mu}{\mu_0}\frac{\beta}{\alpha} + \frac{\chi}{\alpha}\right)\cosh(\lambda z) + \left(\frac{\mu}{\mu_0}\left(\frac{\alpha}{\beta} - \frac{\chi}{\alpha}\right) - \frac{\chi^2}{\alpha\beta}\right)\sinh(\lambda z) \tag{4.76}$$

$$Z(\lambda z) = (\beta\cosh(\lambda z) + \chi\sinh(\lambda z))\,\frac{\chi}{\mu}\frac{1}{\alpha} \tag{4.77}$$

$$\alpha = \tanh(\lambda d) - \coth(\lambda d) \tag{4.78}$$

$$\beta = \frac{\mu}{\mu_0}\tanh(\lambda d) - \coth(\lambda d) \tag{4.79}$$

Equation (4.64) has well-known periodic solutions $\cos m\theta$ and $\sin m\theta$ and results in the following eigenfunctions:

$$\Theta(\theta) = \begin{cases} \sin m\theta, & m = 1,2,3,\ldots \\ \cos m\theta, & m = 0,1,2,\ldots \end{cases} \tag{4.80}$$

Equation (4.65) is Bessel's equation, which can be solved for each m and has the general solution of the following form:

$$R(r) = C_1 J_m(\lambda r) + C_2 Y_m(\lambda r), \tag{4.81}$$

where C_1 and C_2 are constants of integration and $J_m(\cdot)$ and $Y_m(\cdot)$ are the Bessel functions of the first and the second kind, respectively, each of order m. The condition that both magnetic and fluid potentials are bounded throughout the domain can be satisfied only if $C_2 = 0$, which results in

$$R(r) = C_1 J_m(\lambda r) \tag{4.82}$$

Further considering that the miniaturized coils are located far from the walls of the fluid container, (4.82) yields the following characteristic equation:

$$J_m(\lambda r) = 0 \quad \text{at} \quad r = R \tag{4.83}$$

Equation (4.83) can be solved numerically and yields an infinite number of solutions $\varepsilon_{mn} = \lambda R$, $m = 0, 1, 2...$, $n = 1, 2, 3...$, providing the eigenvalue λ_{mn} for each mode as follows:

$$\lambda_{mn} = \frac{\varepsilon_{mn}}{R} \tag{4.84}$$

Equations (4.80) and (4.82) can be combined to obtain the following permissible mode shapes:

$$\begin{aligned} H_{mnc} &= J_m\left(\lambda_{mn} r\right) \cos\left(m\theta\right) \\ H_{mns} &= J_m\left(\lambda_{mn} r\right) \sin\left(m\theta\right) \end{aligned} \tag{4.85}$$

Applying the input magnetic potential (4.71) as a boundary condition on $\psi^{(3)}$ at $z = -h$ and making use of the orthogonality property of the mode shapes (4.85), the unknown constant $A(t)$ can be determined for each mode as

$$A_{mnc}(t) = \frac{1}{Y\left(-\lambda_{mn} h\right)} \times \left\{ Z\left(-\lambda_{mn} h\right) B_0 \tilde{\xi}_{mnc}(t) + \frac{k}{\pi R^2 \left[J_{m+1}\left(\varepsilon_{mn}\right)\right]^2} \sum_{j=1}^{J} \psi_j(t) J_m\left(\lambda_{mn} r_j\right) \cos\left(m\theta_j\right) \right\} \tag{4.86}$$

for $m = 0, 1, 2, \ldots,$ $n = 1, 2, 3, \ldots$ and

$$A_{mns}(t) = \frac{1}{Y\left(-\lambda_{mn} h\right)} \times \left\{ Z\left(-\lambda_{mn} h\right) B_0 \tilde{\xi}_{mns}(t) + \frac{k}{\pi R^2 \left[J_{m+1}\left(\varepsilon_{mn}\right)\right]^2} \sum_{j=1}^{J} \psi_j(t) J_m\left(\lambda_{mn} r_j\right) \sin\left(m\theta_j\right) \right\} \tag{4.87}$$

for $m, n = 1, 2, 3, \ldots$ where

$$k = \begin{cases} 1 & \text{for } m = 0 \\ 2 & \text{for } m \neq 0 \end{cases} \tag{4.88}$$

With the integration constant $A(t)$ obtained as $A_{mnc}(t)$ (4.86) and $A_{mns}(t)$ (4.87), the scalar potentials ϕ and $\psi^{(i)}$, $i = 1, 2, 3$ are fully determined as given in (4.63), (4.72), (4.73), and (4.74), and, for the case of symmetric modes ($\cos m\theta$), their substitution into the surface dynamic Eq. (4.57) gives

$$\rho \frac{1}{\lambda_{mn}} \frac{1}{\tanh\left(\lambda_{mn} d\right)} \frac{\mathrm{d}^2 \tilde{\xi}_{mnc}(t)}{\mathrm{d}t^2} + \rho g \tilde{\xi}_{mnc}(t) + \sigma \lambda_{mn}^2 \tilde{\xi}_{mnc}(t) + \chi B_0 \lambda_{mn} A_{mnc}(t) = 0 \tag{4.89}$$

Further substitution of (4.86) into (4.89), followed by its rearrangement, yields

$$\frac{\mathrm{d}^2 \tilde{\xi}_{mnc}(t)}{\mathrm{d}t^2} + \omega_{mn}^2 \tilde{\xi}_{mnc}(t) = f_{mnc}(t) \tag{4.90}$$

where

$$\omega_{mn}^2 = g \tanh(\lambda_{mn} d) \lambda_{mn} + \frac{\sigma}{\rho} \tanh(\lambda_{mn} d) \lambda_{mn}^3 + \frac{\chi}{\rho} B_0^2 \tanh(\lambda_{mn} d) \lambda_{mn}^2 \frac{Z(-\lambda_{mn} h)}{Y\left(-\lambda_{mn} h\right)} \tag{4.91}$$

$$f_{mnc}(t) = F_{mn} \sum_{j=1}^{J} \psi_j(t) \mathrm{H}_{mnc}^j \tag{4.92}$$

$$F_{mn} = -\frac{\chi}{\rho} B_0 \frac{\tanh(\lambda_{mn} d)}{Y(-\lambda_{mn} h)} \lambda_{mn}^2 \frac{k}{\pi R^2 [J_{m+1}\left(\varepsilon_{mn}\right)]^2} \tag{4.93}$$

$$\mathrm{H}_{mnc}^j = J_m\left(\lambda_{mn} r_j\right) \cos\left(m \theta_j\right) \tag{4.94}$$

$$m = 0, 1, 2, \ldots, \quad n = 1, 2, 3, \ldots, \quad j = 1, 2, 3, \ldots J$$

Using (4.87), a similar set of equations can be obtained for the antisymmetric modes ($\sin m\theta$) as follows:

$$\frac{\mathrm{d}^2 \tilde{\xi}_{mns}(t)}{\mathrm{d}t^2} + \omega_{mn}^2 \tilde{\xi}_{mns}(t) = f_{mns}(t) \tag{4.95}$$

$$f_{mns}(t) = F_{mn} \sum_{j=1}^{J} \psi_j(t) H_{mns}^j \tag{4.96}$$

$$H_{mns}^j = J_m\left(\lambda_{mn} r_j\right) \sin\left(m \theta_j\right) \tag{4.97}$$

$$m, n = 1, 2, 3, \ldots \quad j = 1, 2, 3, \ldots J$$

The generalized displacements $\tilde{\xi}_{mnc}(t)$ and $\tilde{\xi}_{mns}(t)$, obtained from the solution of the second-order differential Eqs. (4.90) and (4.95), respectively, and the corresponding mode shapes H_{mnc} and H_{mns} evaluated at any desired location (r_k, θ_k) give the total surface displacement at the location as

$$\zeta\left(r_k,\theta_k,t\right)=\sum_{m=0}^{\infty}\sum_{n=1}^{\infty}\tilde{\xi}_{mnc}(t)H_{mnc}\left(r_k,\theta_k\right)+\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\tilde{\xi}_{mns}(t)H_{mns}\left(r_k,\theta_k\right) \tag{4.98}$$

The damping effects associated with the fluid viscosity may be introduced into the model by adding the factors $\omega_{d_{mn}}\left(\mathrm{d}\tilde{\xi}_{mnc}/\mathrm{d}t\right)$ and $\omega_{d_{mn}}\left(\mathrm{d}\tilde{\xi}_{mns}/\mathrm{d}t\right)$ to the left-hand side of (4.90) and (4.95), respectively. The damping frequency $\omega_{d_{mn}}$ can be approximated as follows (Blums et al. 1997):

$$\omega_{d_{mn}}=4\frac{\eta}{\rho}\lambda_{mn}^2 \tag{4.99}$$

The solution (4.98) is truncated to a finite number of modes such that $m=0,1,2,\ldots,M$ and $n=1,2,\ldots,N$. Similarly, the surface deflection $\zeta\left(r_k,\theta_k,t\right)$ as provided by (4.98) is limited to a discrete number of surface locations k, $k=1,2,\ldots,K$. The dynamics of the system comprising (4.90), (4.95), and (4.98), with the viscous damping terms added to the first two, is represented in state-space form as follows:

$$\begin{aligned}\dot{\mathbf{x}}&=\mathbf{A}\mathbf{x}+\mathbf{B}'\mathbf{u}'\\ y&=\zeta=\mathbf{C}\mathbf{x}\end{aligned} \tag{4.100}$$

where

$$\mathbf{x}=\Big[\tilde{\xi}_{01}\ \dot{\tilde{\xi}}_{01}\ \tilde{\xi}_{02}\ \dot{\tilde{\xi}}_{02}\cdots\ \tilde{\xi}_{0N}\ \dot{\tilde{\xi}}_{0N}\ \ \tilde{\xi}_{11c}\ \dot{\tilde{\xi}}_{11c}\ \tilde{\xi}_{11s}\ \dot{\tilde{\xi}}_{11s}\ \tilde{\xi}_{12c}\ \dot{\tilde{\xi}}_{12c}\ \tilde{\xi}_{12s}\ \dot{\tilde{\xi}}_{12s}\cdots$$
$$\tilde{\xi}_{MNc}\ \dot{\tilde{\xi}}_{MNc}\ \tilde{\xi}_{MNs}\ \dot{\tilde{\xi}}_{MNs}\Big]^{\mathrm{T}}_{(1\times 2N(2M+1))}$$

is the vector of the generalized displacements and the corresponding velocities,

$$\mathbf{u}'=\left[\psi_1\ \ \psi_2\ \cdots\ \psi_J\right]^{\mathrm{T}}_{(1\times J)}$$

is the vector of input magnetic potentials, and

$$\mathbf{y}=\left[\zeta_1,\zeta_2,\ldots,\zeta_K\right]^{\mathrm{T}}_{(1\times K)}$$

is the vector of surface deflections defining the system output. With the system order determined as $N_d=2N(2M+1)$, the system matrices $\mathbf{A}\in\mathbb{R}^{N_d\times N_d}$, $\mathbf{B}\in\mathbb{R}^{N_d\times J}$, and $\mathbf{C}\in\mathbb{R}^{K\times N_d}$ are as follows:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ -\omega_{01}^2 & -\omega_{d_{01}} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & -\omega_{02}^2 & -\omega_{d_{02}} & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & -\omega_{MN}^2 & -\omega_{d_{MN}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & -\omega_{MN}^2 & -\omega_{d_{MN}} \end{bmatrix}_{(N_d \times N_d)}$$

$$\mathbf{B}' = \mathbf{F}\mathbf{H}$$

where

$$\mathbf{F} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ F_{01} & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & F_{02} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & F_{MN} & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & F_{MN} \end{bmatrix}_{(N_d \times N_d/2)}$$

$$\mathbf{H} = \begin{bmatrix} \mathrm{H}_{01c}^1 & \mathrm{H}_{01c}^2 & \cdots & \mathrm{H}_{01c}^J \\ \mathrm{H}_{02c}^1 & \mathrm{H}_{02c}^2 & \cdots & \mathrm{H}_{02c}^J \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{H}_{0Nc}^1 & \mathrm{H}_{0Nc}^2 & \cdots & \mathrm{H}_{0Nc}^J \\ \mathrm{H}_{11c}^1 & \mathrm{H}_{11c}^2 & \cdots & \mathrm{H}_{11c}^J \\ \mathrm{H}_{11s}^1 & \mathrm{H}_{11s}^2 & \cdots & \mathrm{H}_{11s}^J \\ \mathrm{H}_{12c}^1 & \mathrm{H}_{12c}^2 & \cdots & \mathrm{H}_{12c}^J \\ \mathrm{H}_{12s}^1 & \mathrm{H}_{12s}^2 & \cdots & \mathrm{H}_{12s}^J \\ \cdots & \cdots & \ddots & \vdots \\ \mathrm{H}_{MNc}^1 & \mathrm{H}_{MNc}^2 & \cdots & \mathrm{H}_{MNc}^J \\ \mathrm{H}_{MNs}^1 & \mathrm{H}_{MNs}^2 & \cdots & \mathrm{H}_{MNs}^J \end{bmatrix}_{(N_d/2 \times J)}$$

and

$$\mathbf{C} = \begin{bmatrix} \mathrm{H}^1_{01c}\ 0\ \mathrm{H}^1_{02c}\ 0\ \cdots\ \mathrm{H}^1_{0Nc}\ 0\ \mathrm{H}^1_{11c}\ 0\ \mathrm{H}^1_{11s}\ 0\ \mathrm{H}^1_{12c}\ 0\ \mathrm{H}^1_{12s}\ 0\ \cdots\ \mathrm{H}^1_{MNc}\ 0\ \mathrm{H}^1_{MNs}\ 0 \\ \mathrm{H}^2_{01c}\ 0\ \mathrm{H}^2_{02c}\ 0\ \cdots\ \mathrm{H}^2_{0Nc}\ 0\ \mathrm{H}^2_{11c}\ 0\ \mathrm{H}^2_{11s}\ 0\ \mathrm{H}^2_{12c}\ 0\ \mathrm{H}^2_{12s}\ 0\ \cdots\ \mathrm{H}^2_{MNc}\ 0\ \mathrm{H}^2_{MNs}\ 0 \\ \vdots\ \vdots\ \ddots\ \vdots\ \vdots\ \vdots\ \vdots\ \ddots\ \vdots\ \vdots \\ \mathrm{H}^K_{01c}\ 0\ \mathrm{H}^K_{02c}\ 0\ \cdots\ \mathrm{H}^K_{0Nc}\ 0\ \mathrm{H}^K_{11c}\ 0\ \mathrm{H}^K_{11s}\ 0\ \mathrm{H}^K_{12c}\ 0\ \mathrm{H}^K_{12s}\ 0\ \cdots\ \mathrm{H}^K_{MNc}\ 0\ \mathrm{H}^K_{MNs}\ 0 \end{bmatrix}$$

representing the matrix of shape functions evaluated at (r_k, θ_k), $k = 1, 2, \ldots, K$. The state-space model (4.100) can be used to measure the response of a MFDM surface to the inputs generated by an arbitrary array of miniature coils.

4.3 Current–Potential Relationship

For the development of the analytical model given above, electromagnetic coils were idealized as point sources of magnetic potential. The actual mirror, on the other hand, is controlled using currents applied to the coils. To generate corresponding surface responses using the analytical model, it is necessary to establish a relationship between the current input to a coil and the value of the magnetic potential of the corresponding point source. The current–potential relationship is determined as follows.

Firstly, the position of the point source with respect to the corresponding coil is fixed. Since the contribution of the top turns of the coil to the magnetic field observed at the surface of the fluid is much more significant than that of the lower turns, it is assumed that the point source of potential is located at the same position as that of the center of the top turns of the coil.

Secondly, using arbitrary values of current in the coil, the magnetic field of the actual coil is simulated using COMSOL Multiphysics^{TM} (finite element analysis software by COMSOL Inc. Stockholm, Sweden), and the values of the magnetic flux density at the location corresponding to a point on the liquid surface immediately above the center of the coil are determined. Then, using the derived analytical model, the values of the point source of potential that yield the same values for the magnetic flux density at the corresponding location are computed. The slope of the linear current–potential relationship is then determined based on the obtained data. Actually, only two points are sufficient since the relationship is found to be linear.

The current–potential relationship for the prototype MFDM, whose physical parameters are described in the Chap. 5, is given in Fig. 4.3. The relationship is based on the COMSOL simulation of the actual coils used in the prototype MFDM (see Table 5.1 for the coil specification used in the simulation). The relationship is determined for one electromagnetic coil and is then applied uniformly across all others in the array. It should be noted that this relationship was determined independently of the experimental data.

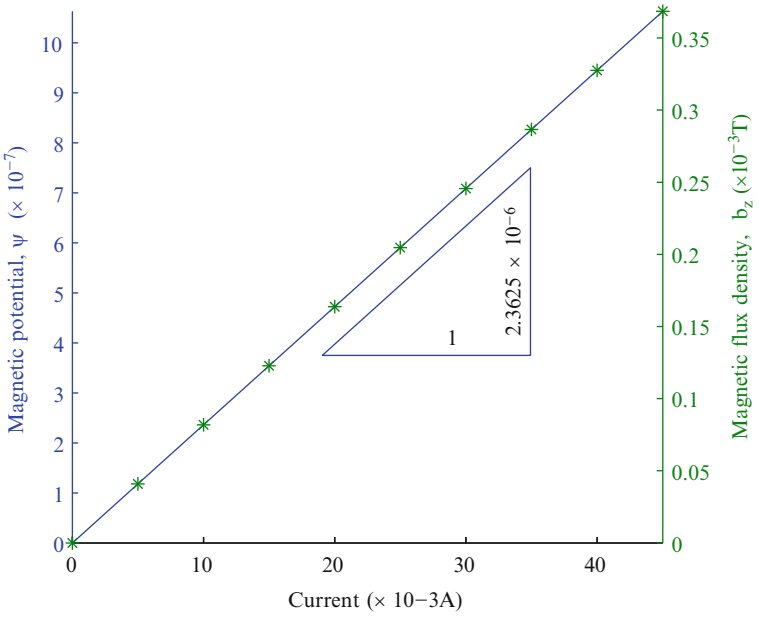


Fig. 4.3 Current–potential relationship

If α is the slope of the current–potential relationship, then

$$\mathbf{u}' = \alpha \mathbf{u} \tag{4.101}$$

where $\mathbf{u}$ is the vector of currents u_j, $j = 1, 2, 3, \ldots J$, applied to the electromagnetic coils. Using (4.101), the model (4.49) or (4.100) can be rewritten as follows:

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \\ \mathbf{y} &= \mathbf{C}\mathbf{x} \end{aligned} \tag{4.102}$$

where $\mathbf{B} = \alpha \mathbf{B}'$. A discrete-time equivalent representation of model (4.102) is given by

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{A_d}\mathbf{x}(k) + \mathbf{B_d}\mathbf{u}(k) \\ \mathbf{y}(k) &= \mathbf{C}\mathbf{x}(k) \end{aligned} \tag{4.103}$$

where $\mathbf{x}(k)$ is the vector of state variables, $\mathbf{u}(k)$ is the vector of control currents, $\mathbf{y}(k)$ is the vector of the mirror surface deflections, and $\mathbf{A_d}$ and $\mathbf{B_d}$ are the system matrices.

The DC gain of the system relating the steady-state response $\mathbf{y}_{ss}$ to a vector of static inputs $\mathbf{u}_{ss}$ can be obtained as (Skogestad and Postlethwaite 2005)

$$\mathbf{G_0} = -\mathbf{C}\mathbf{A}^{-1}\mathbf{B} \tag{4.104}$$

in the case of continuous time system and as (Chen and Francis 1996)

$$\mathbf{G_0} = \mathbf{C}(\mathbf{I} - \mathbf{A_d})^{-1}\mathbf{B_d} \tag{4.105}$$

in the case of discrete-time system.

4.4 Simulation of the MFDM Model

The analytical model presented above can be used to simulate the response of the MFDM surface to the magnetic field generated by an array of electromagnetic coils. The model provides the static as well as the dynamic surface shapes produced by the applied magnetic field. In the following, the analytical model representing the prototype MFDM presented in the next chapter is simulated. The same simulations will be utilized later to validate the analytical model by comparing the results predicted by the model to the experimental results obtained using the prototype mirror.

4.4.1 Model Parameters

The MFDM is modeled as a layer of EFH1 (ferrofluid for education—hydrocarbon type), which is a commercially available oil-based magnetic fluid (Ferrotec Corporation, NH, USA). The properties of the fluid used in the simulation are given in Table 4.1. The fluid layer has the following dimensions:

$$d = 1.0 \text{ mm}, \quad R = 30 \text{ mm}$$

The magnetic field is generated using an array of 19 electromagnetic coils arranged in a circular pattern and placed at distance $h = 2.0$ mm below the undisturbed surface of the mirror. To ensure accuracy, the actual location of the coils measured in the prototype mirror is used in the model. The location $\{(r_j \text{ (mm)}, \theta_j \text{ (rad)}), j = 1, 2, \ldots, 19\}$ of the coils is given as

Table 4.1 Properties of EFH1, the magnetic fluid used in the prototype MFDM

Density	1,210 kg/m^3
Viscosity	5.8 cP
Saturation magnetization	40 mT
Relative permeability	1.78
Coefficient of surface tension	29×10^{-3}N/m

$$\{(0,0),(4.1,4.7),(3.4,5.6),(3.9,0.8),(4.2,1.7),(5.2,2.7),(4.4,3.8),(7.6,4.8),\ldots$$
$$(7.4,5.3),(7.0,5.8),(7.5,0),(7.4,0.6),(7.1,1.3),(8.3,1.7),(7.8,2.2),(9.2,2.7),\ldots$$
$$(8.9,3.1),(8.0,3.7),(7.0,4.2)\}$$

The slope of the current–potential relationship, α, as determined from the COMSOL simulation of the actual electromagnetic coil, is set to 2.3625×10^{-6}/A. A vertical, uniform magnetic field with the flux density $B_0 = 2.5$ mT is superimposed on the magnetic field generated by the array of coils. The factors considered in the selection of the field strength are discussed in Sect. 5.1.3. The analytical model (4.102) used for simulation is based on the above mentioned parameters, and the model was obtained by truncating the radial and azimuthal modes to $N = 4$, $M = 4$, respectively.

4.4.2 Static Response

The static response of the mirror surface shape can be obtained by setting $\dot{\mathbf{x}} = 0$ with input $\mathbf{u}$ given as a vector of constant currents. Firstly, the response of the mirror surface to the current applied to only one coil is simulated. The simulated static shapes of the mirror surface obtained by applying constant currents to the coil located at (r_1, θ_1) are plotted in Fig. 4.4. Taking advantage of the symmetry of the surface shape generated by the chosen coil, that is, the one in the center of the mirror, only 2D shapes are shown in the figure. As can be observed from the figure, application of current to a single coil results in a Gaussian surface shape with its peak located immediately above the location of the energized coil. A 3D surface shape resulting from a specified set of input currents

$$\mathbf{u} = [0.1, 0.5, 0.4, 0.3, 0.1, -0.1, 0.2, 0.9, 0.7, 0.6, -0.5, 0.6, 0.1,$$
$$-0.4, -0.9, -0.3, 0.5, -0.3, 0.4]^{\mathrm{T}} \times \frac{1}{31.5}\mathrm{A}$$

applied to the complete array of 19 coils is presented in Fig. 4.5.

4.4.3 Dynamic Response

The analytical model can be used to observe the dynamics of the mirror surface shape. As an illustration of the ability of the model to capture the dynamic response of the mirror surface, the surface deflections resulting from a step input applied to the center coil are plotted in Fig. 4.6. The figure shows the time history of the surface deflections observed at a surface location immediately above the coil.

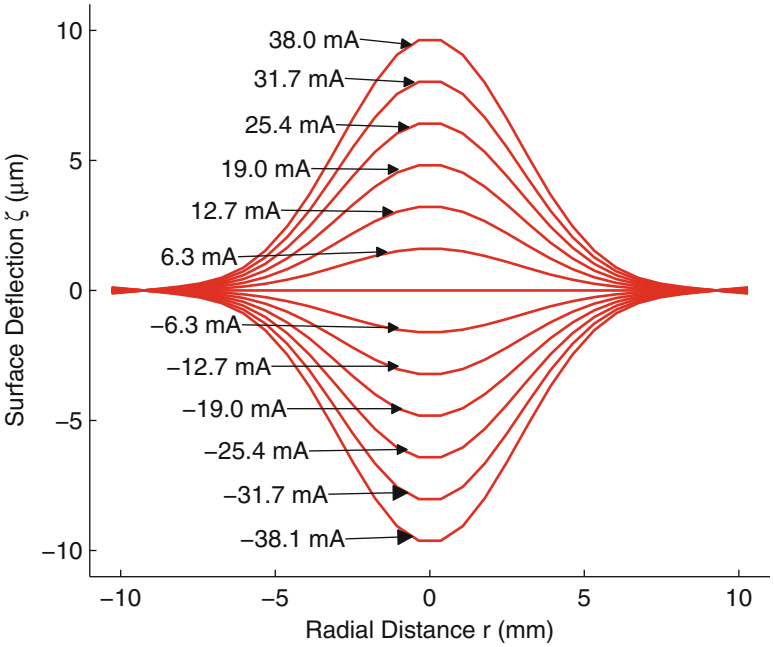


Fig. 4.4 The static surface shapes predicted by the analytical model

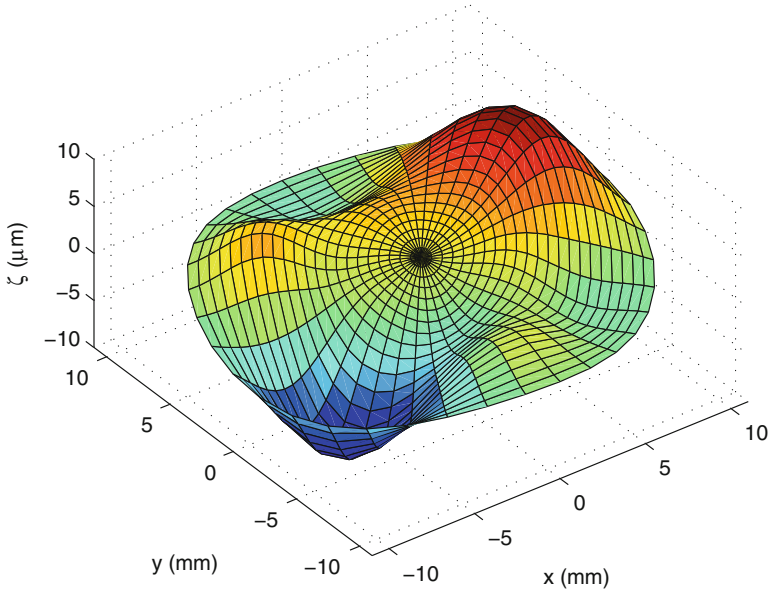


Fig. 4.5 3D static surface shapes predicted by the analytical model

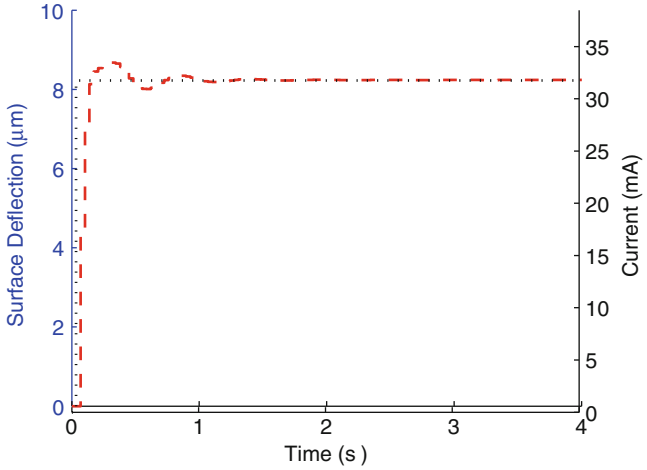


Fig. 4.6 Response to a step input. The input current is applied to the center coil only, and the surface deflection is observed at a point immediately above the same coil

A comprehensive picture of the dynamics of the MFDM surface can be seen by drawing the Bode plots of the MIMO model. Figure 4.7 shows the Bode plots for three selected channels where the input is applied at the coils located at (r_1, θ_1), (r_2, θ_2), and (r_{10}, θ_{10}) and the output surface deflections are observed at the surface points located immediately above the respective coils. While a detailed discussion on the dynamics of the mirror will follow in following chapters, a few observations may be noted here. Note that plots on the leading diagonal show the input–output relationship between an electromagnetic coil input and the surface displacement obtained immediately above the same coil. The off-diagonal plots, on the other hand, show the coupling effect between different input–output channels. It is obvious that for any input coil, the magnitude of the surface deflection immediately above the coil is significantly higher than the magnitude of the deflections obtained away from the coil. Also note that the coupling effect varies from channel to channel. While an input to the coil located at (r_1, θ_1) results in an output at (r_2, θ_2) with a significant magnitude, the magnitude of the resulting surface deflections at (r_{10}, θ_{10}) is negligibly small.

4.5 Summary

An analytical model of the dynamics of the surface shape of a MFDM has been presented in this chapter. The model is derived using a coupled system of equations obtained from the fundamental governing principles of conservation of mass and momentum, and Maxwell's equations. The model describes the dynamics of the

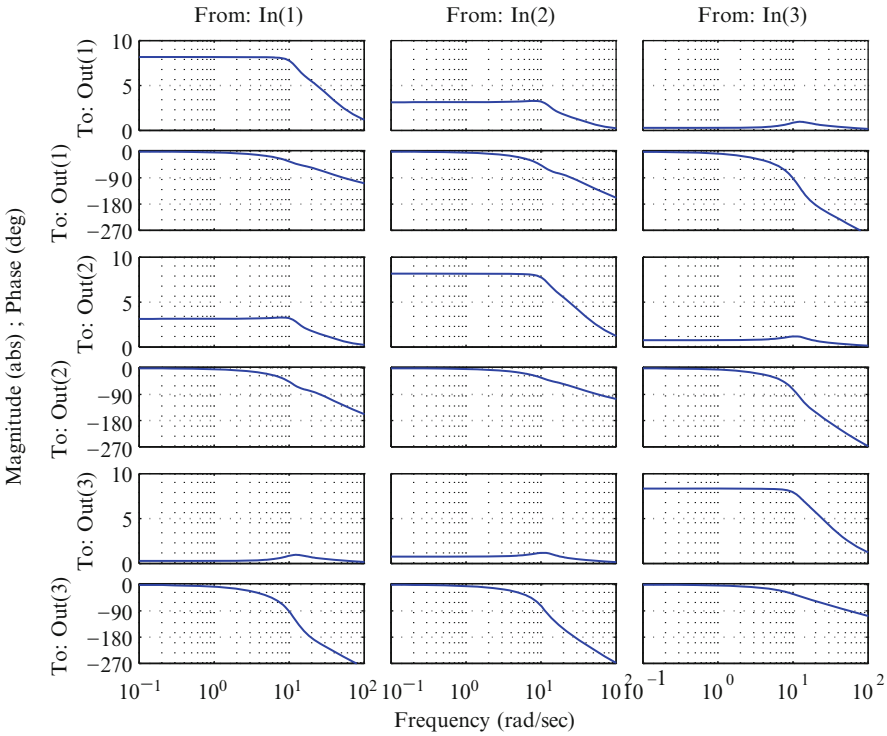


Fig. 4.7 Bode plot of the MFDM model. The array of Bode plots shows the system response for three selected input–output channels. The inputs In(1), In(2), and In(3) correspond to the electromagnetic coils $j = 1, 2$, and 10, respectively. The three plots in the *top row* are the magnitude plots of the output surface deflections predicted by the model for the location immediately above the center of coil #1. The *second row* shows the corresponding phase plots. The magnitude and phase plots for the predicted deflections above coils #2 and #10 are given alternatively in the following rows of plots

mirror surface shape in terms of deflections of the mirror surface produced by a magnetic field generated by an array of electromagnetic coils. The model has been developed in both the Cartesian and the cylindrical coordinate systems. An innovative method of linearizing the equations governing the dynamics of the mirror surface was introduced and will be utilized to propose an important modification in the design of a MFDM. The developed model was presented in state-space form and was used to simulate the static as well as the dynamic response of the mirror surface shape.

References

Bashtovoi VG, Rosensweig RE (1993) Excitation and study of sub-critical waves on a magnetic fluid surface. J Magn Magn Mater 122:234–240

Blums E, Cebers A, Maiorov MM (1997) Magnetic fluids. de Gruyter, Berlin

Chen T, Francis B (1996) Optimal sampled data control systems. Springer, Berlin

Cowley MD, Rosensweig RE (1967) The interfacial stability of a ferromagnetic fluid. J Fluid Mech 30(4):671–688

Haberman R (2003) Applied partial differential equations with Fourier series and boundary value problems. Prentice-Hall, Englewood Cliff

Rosensweig RE (1997) Ferrohydrodynamics. Dover Publications, New York

Skogestad S, Postlethwaite I (2005) Multivariable feedback control: analysis and design. Wiley, Chichester

Chapter 5
Design of a Magnetic Fluid Deformable Mirror and Experimental Model Validation

Contents

5.1 Design of Magnetic Fluid Deformable Mirror

The design of a practically deployable MFDM involves a wide range of considerations including—just to name a few—the application specific performance requirements, maintainability, and cost constraints. Covering all these aspects is beyond the scope of this book. However, in what follows, we will examine important design features of a MFDM that affect the control of the mirror and its potential application in ophthalmic imaging systems.

Z. Wu et al., *Modeling and Control of Magnetic Fluid Deformable Mirrors for Adaptive Optics Systems*, DOI 10.1007/978-3-642-32229-7_5,

5.1.1 Conceptual Design

5.1.1.1 Conventional Conceptual Design

The basic operating principle of a MFDM was briefly described in Sect. 3.2, and the schematic layout of the conventional design of the mirror was shown in Fig. 3.3. The primary elements of the mirror are a layer of a magnetic fluid, a thin film of a reflective material coated on the free surface of the fluid, and an array of electromagnetic coils placed beneath the fluid layer. Each of the electromagnetic coils is driven independently by applying a corresponding current input. The combined electromagnetic field of the coils causes the fluid surface and the reflective film to deform. The deformed reflective surface acts as a mirror whose shape can be controlled by adjusting the current applied to the individual coils. The material properties of the magnetic fluid and the reflective film, the shape and size of the fluid layer, the number and arrangement of the electromagnetic coils, and the dimensions of each coil are the variables that need to be established in the detailed design process.

5.1.1.2 Modified Design

The analytical work undertaken to develop a model of a MFDM, as given in Chap. 4, revealed that the deflection of the magnetic fluid surface is nonlinear in the applied magnetic field. This nonlinearity results in a fundamental limitation of the conventional MFDM design: since the surface displacement is a quadratic function of the applied magnetic field, only positive displacements can be achieved. It has already been known that the magnetic field of electromagnetic coils in conventional configuration can displace the magnetic fluid in the upward direction only, that is, the magnetic field can "push" the fluid surface but cannot "pull" (Laird et al. 2006). As it was briefly introduced in Sect. 4.1.2, one of the solutions to the problem of nonlinear response of the fluid surface is to linearize it by superimposing a large, uniform, vertical magnetic field on the field generated by the electromagnetic coils. The proposed modified design is the realization of this solution, where a large uniform magnetic field generated by an appropriate source is superimposed on the magnetic field of the conventionally used array of coils. In the proposed design, a Helmholtz coil is used to generate the uniform magnetic field. As shown in Fig. 5.1, the Helmholtz coil consists of two circular coils of equal diameter, which are separated by a distance equal to the radius of the coils. This distinctive design provides a magnetic field, which runs parallel to the central axis and is uniform over a significant volume extending around the axis.

The proposed modified design of a MFDM is a combination of a Helmholtz coil and the conventional design illustrated in Fig. 3.3. The modified design, as shown in Fig. 5.2, places the conventional MFDM inside the Helmholtz coil, thus allowing the uniform field of the Helmholtz coil to be superimposed on the field of the coils

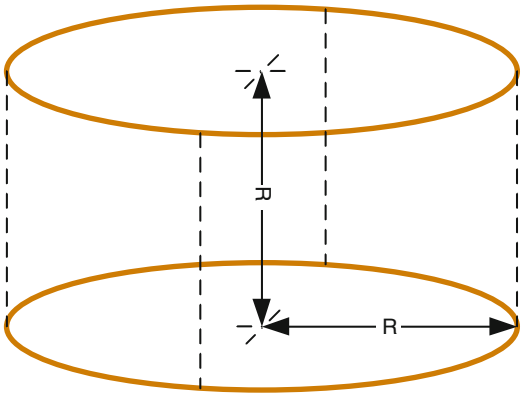


Fig. 5.1 Layout of a Helmholtz coil

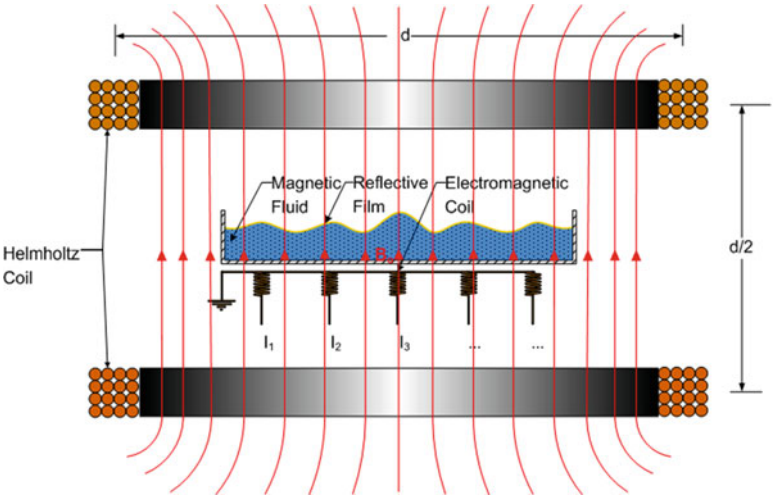


Fig. 5.2 Modified conceptual design of a magnetic fluid deformable mirror

placed underneath the fluid layer. The figure shows a more practical Helmholtz coil with each of the two coils comprising multiple turns. For clarity, the coils placed underneath the fluid layer will be referred to as miniature coils. If the magnetic field of the Helmholtz coil is significantly larger than the field of the miniature coils, the surface displacements become linearized in terms of the vertical component of the magnetic field of the miniature coils. As alluded to in the theoretical work presented in Sect. 4.1.2, the horizontal components of the magnetic field can be ignored when the magnetic field of the Helmholtz coil is present. The following are the main features of the proposed design:

- The conventional mirror is placed inside the Helmholtz coil such that the central axis of the coil is aligned with the gravity vector, that is, the uniform field of the coil runs perpendicular to the free static surface of the magnetic fluid.

- The strength of the uniform field is chosen to be significantly larger than the maximum magnetic field of each of the miniature coils.
- Contrary to the conventional configuration where the electromagnetic coils are augmented with a soft magnetic core for increasing the flux density, the miniature coils in the modified design are air-cored type. Theoretically, the coils with a soft magnetic core can still be used to serve the basic purpose of linearization of the response of the fluid surface. However, they cause the magnetic field of the Helmholtz coil to concentrate along the axis of the miniature coils submerged in the field of the Helmholtz coil, thus resulting in nonuniformities of the field. The resulting field gradients cause deflections of the fluid surface even when no currents are applied to the miniature coils.
- The diameter of the fluid layer is chosen to be significantly larger than the active area containing the coils, which implies that the diameter of the fluid layer is significantly larger than the optical pupil as measured on the fluid surface. This feature ensures that the effects of the magnetic field gradients at the edges of the fluid layer are excluded from the active mirror area. The gradients are caused by the unavoidable discontinuity of magnetic permeability at the interface of the fluid and the container. The exclusion of the edges is also desirable for the conventional design as it avoids the meniscus along the walls of the container.

5.1.2 Detailed Design

While the conceptual design provides only an outline of any product, the detailed design is an iterative process that determines the definite values of the various design features, that is, design variables, which meet a prescribed set of constraints and performance requirements. Depending on the nature of the application, different sets of constraints and performance requirements can be specified for the design of a MFDM. These requirements in turn dictate what features should be considered as design variables. We consider a generalized set of requirements and design variables as described below.

5.1.2.1 Design Requirements

Spatial resolution: Spatial resolution or the number of spatial modes to be corrected signifies the degree of accuracy of correction required. The finer the extent of defects that need to be compensated for, the higher the required order or the number of modes will be.

Bandwidth: The bandwidth is the required operating frequency range of the deformation of the mirror surface. The requirement is dictated by the frequency of the aberrations that need to be corrected.

Dynamic range: The dynamic range defines the maximum and the minimum achievable correction, that is, the magnitude of surface deflection that can be

generated using the MFDM. For example, a deformable mirror may be required to produce up to 10 μm deflection with a resolution of 10 nm. The stroke, that is, the maximum required deflection of the mirror surface, is dictated by the depth of the aberrations that need to be corrected. The requirement of the smallest applicable deflection is dictated by the precision of the correction.

5.1.2.2 Design Variables

Number of actuators and actuator spacing: The total number of actuators (electromagnetic coils) and spacing between them determines the spatial resolution or the number of modes that can be corrected. For a given shape of the array of actuators, it also determines the size of the active part of deformable mirror, which in turn determines the size of the AO system. A good approximation of the appropriate spacing can be obtained from the empirical observation that most of the wavefront correctors currently in use have 10–20% coupling between two adjacent actuators (Tyson 2011). The actuator spacing that provides the desired coupling can be directly observed by plotting the surface shape obtained by applying input to a single actuator.

Dimensions of the miniature coils: The strength of the magnetic field generated by each magnetic coil as measured at the surface of the magnetic fluid is the principle factor that controls the shape and magnitude of the resulting surface deflections. The strength of the magnetic field generated by the coil is a function of the number of turns in each coil, the length and diameter of the coil, and the maximum allowable current in the coil.

In the modified design featuring a vertical, uniform magnetic field superimposed on the field of the miniature coils, only the vertical component of the magnetic field of the miniature coils plays the significant role. It serves to provide the desired shape of deformation of the fluid surface, while the large uniform field provides the necessary amplification of the surface deflections.

Size of the fluid layer: The size of the fluid layer refers to the depth and diameter of the layer. The depth of the fluid layer affects the stroke of the mirror, that is, the maximum achievable surface deflections. As discussed in Sect. 4.2, the magnitude of the mirror surface deflection is a linear function of the vertical component, b_z, of the magnetic field of the miniature coils as observed at the mirror surface. The magnetic field component b_z itself can be varied by varying the vertical distance h between the top edge of the array of electromagnetic coils and the deformable surface of the mirror. Therefore, keeping all other factors constant, the maximum mirror surface deflection can be varied by varying the depth of the fluid layer. As shown in Fig. 5.3, the vertical component, b_z, of the magnetic field generated by an electromagnetic coil varies exponentially with the variation of the distance h. Consequently, any variation in the depth of the fluid layer also causes an exponential change in the stroke of the mirror.

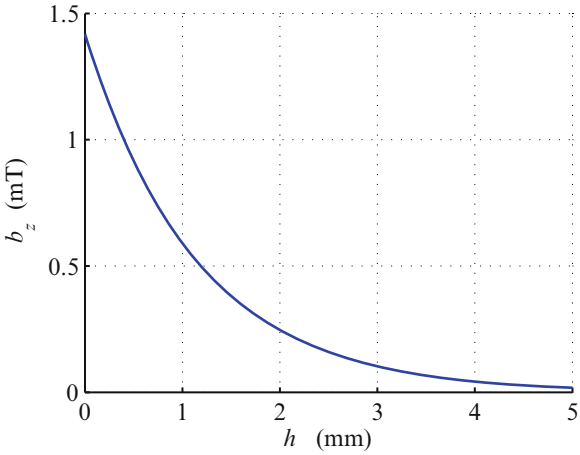


Fig. 5.3 The decay of the *vertical* component of the magnetic flux density (b_z) of a coil with an increasing distance, h, from the *top* edge of the coil

Magnetic fluid properties: Due to the synthetic nature of the magnetic fluids, the properties of a fluid intended for an application can always be tailored according to the requirements of the application. Therefore, the choice of appropriate fluid alone offers a great degree of freedom in the design of a MFDM. For example, the surface deflections obtained using a given magnetic field can be greatly varied by using magnetic fluids with different permeability or density. Similarly, the viscosity of the fluid affects the damping properties of the fluid motion and hence is an important determinant of the dynamic stability and controllability of the MFDM. Surface tension is another fluid property that affects the surface response and will be discussed at greater length in Sect. 5.3.2.

5.1.2.3 Design of the Helmholtz Coil

The nominal diameter (d in Fig. 5.2) of the two circular coils, the number of turns in each coil, and the maximum allowable current are the primary design variables pertaining to a Helmholtz coil. The Helmholtz coil should be able to provide a magnetic field, which is significantly (at least ten times) larger than the magnetic field of the miniature coils. Also, the uniformity of the magnetic field generated by the Helmholtz coil deteriorates with the increasing distance from the central axis. Therefore, the internal diameter should be large enough to ensure a minimum specified uniformity.

The nonuniformity of the magnetic field far from the central axis can produce a curvature on the initial surface of the liquid. This initial curvature in the liquid surface shape can be compensated using the miniature coil actuators; however, this compensation is done at the expense of using stroke that would otherwise be available for the correction of incoming wavefronts. An alternative to using a Helmholtz coil is to use a Maxwell coil (Brousseau et al. 2010). A Maxwell coil is

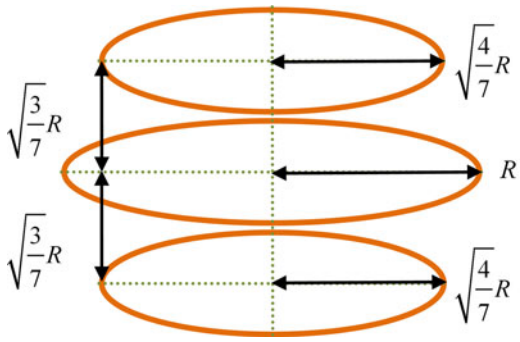


Fig. 5.4 Layout of a Maxwell coil

made up of three separate coils (see Fig. 5.4), which results in a uniform magnetic field up to the sixth order derivative with respect to the position near the center of the coil, compared with the 4th order derivative produced by Helmholtz coil (Caprari 1995). The radii of the coils and their vertical position must follow the ratios shown in Fig. 5.4. The number of ampere-turns (AT) of both the lower and upper coils must also be exactly in the ratio of 49/64 relative to the middle coil. It is usually more convenient to adjust the number of turns of the lower and upper coils than to adjust their current in the design of a Maxwell coil. By properly adjusting the number of turns of the lower and upper coils, the three coils can be arranged in a series configuration in the resulting electric circuit by supplying them with the same current. Using this structure, Brousseau et al. (2010) built a 91-actuator MFDM with a high linearization performance.

5.1.3 Description of the Prototype MFDM

A MFDM comprising 19 input channels has been developed for the experimental validation of the MFDM model and the controllers presented in this book. Though a practically deployable mirror should meet all of the requirements listed above, the mirror was built as a proof-of-concept prototype, which meets only as many requirements as could be accommodated within the limited time and resources available for the research work. Figure 5.5 shows the schematic layout of the mirror. The pictorial view of the assembly of the mirror is shown in Fig. 5.6. The salient features of the mirror are as follows:

- The fluid element of the mirror comprises a 60-mm-diameter layer of EFH1 (ferrofluid for education—hydrocarbon type) which is a commercially available oil-based magnetic fluid (Ferrotec Corporation, NH, USA). The thickness of the layer can be varied by adding appropriate volume of the fluid. Experiments were performed using 1.0-mm-thick layer. Important properties of the fluid are listed in Table 4.1.

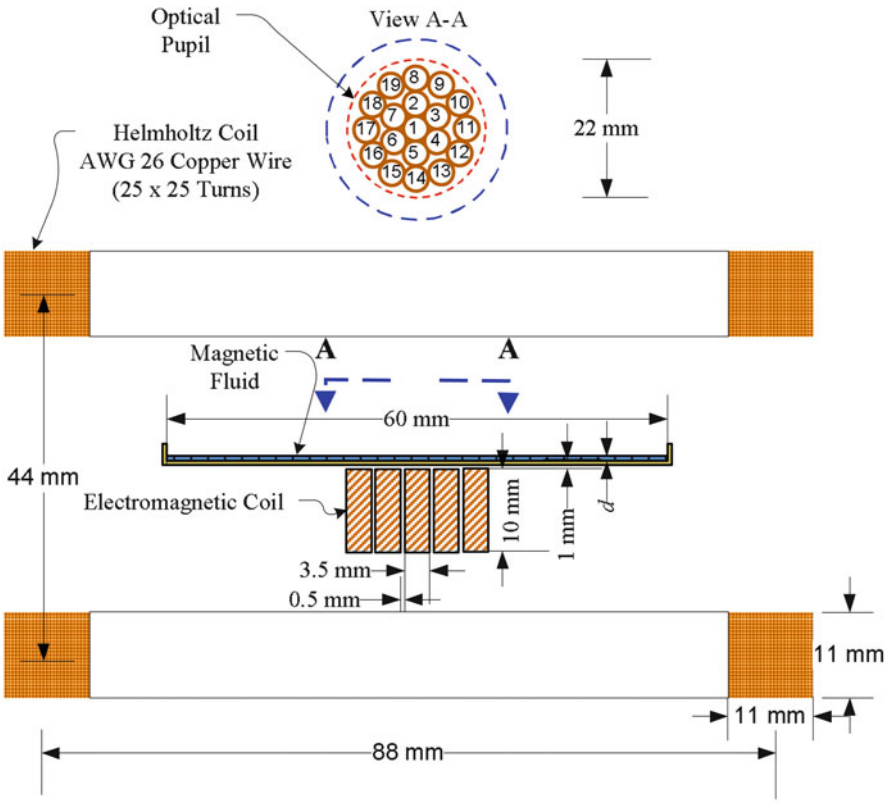


Fig. 5.5 Schematic diagram of the prototype MFDM

- The mirror is driven using an array of 19 electromagnetic coils arranged in a circular pattern as shown Fig. 5.7. The coils are radially spaced at 4 mm from center to center, thus presenting a total active footprint area of approximately 22-mm diameter.
- The electromagnetic coils are conventional circular coils wound on a cylindrical bobbin. The coils were obtained off-the-shelf from Dia-Netics Inc., CA, and have the properties/dimensions as given in Table 5.1.
- The Helmholtz coil was locally fabricated and has the properties/dimensions as given in Table 5.2. The strength of the magnetic field of the Helmholtz coil can be controlled by varying the current applied to the coil. The coil is normally operated at 194 mA, which produces a nominal magnetic flux density of 2.5 mT at the center of the coil. The strength of the magnetic field is chosen such that it is sufficiently larger than that of the miniature coils ($b_z \approx 0.25$) mT, as measured at the fluid surface, but is safely below the critical limit ($\sim$17 mT) at which the Rosensweig instability occurs (Gollwitzer et al. 2007).

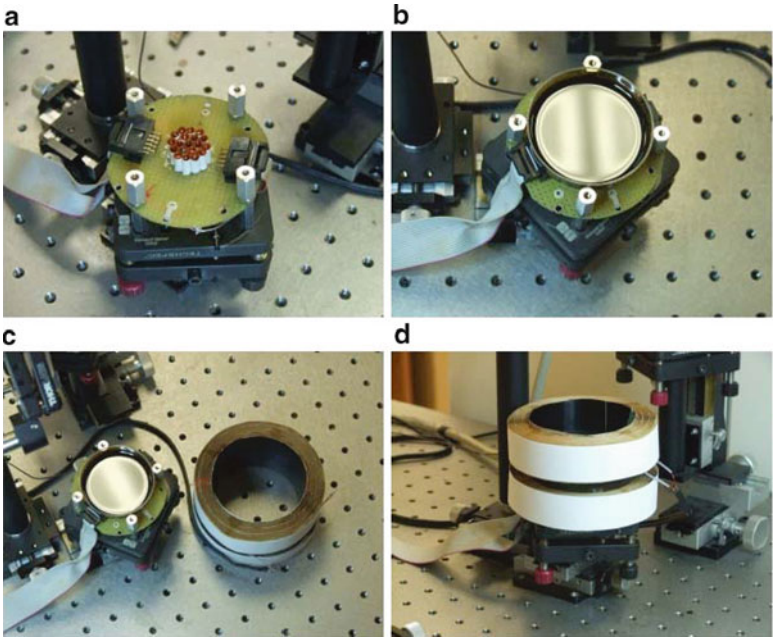
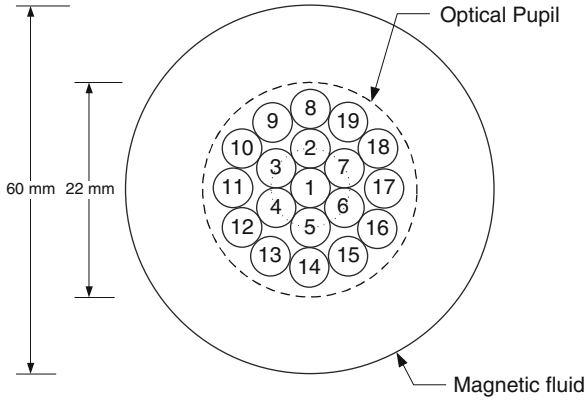


Fig. 5.6 Assembly of the prototype MFDM (**a**) Array of coils. (**b**) Fluid container added. (**c**) Mirror and Helmholtz coil. (**d**) Assembled mirror

Fig. 5.7 Layout of the array of electromagnetic coils



- The array of electromagnetic coils is driven by a PC-based system comprising a PCI-type analog voltage output card (National Instruments 6723) and a current amplification and protection unit. The analog voltage card features 32 output voltage channels with a maximum sampling speed of 45 kS/s. With the 13-bit resolution, it can provide voltages up to ± 10 V at a precision of 2.4 mV. Since the analog output voltage channels can supply only up to 5 mA current as against the

Table 5.1 Properties of the miniature electromagnetic coils

Manufacturer's part no.	PCRC 408-006
Length	0.4 in (10.1 mm)
Internal diameter	0.084 in (2.1 mm)
External diameter incl. bobbin flange	0.137 in (3.5 mm)
No. of turns	620
No. of layers	4
Core-type	Air-cored
Average resistance	31.5 Ω
Wire gauge	AWG41
Wire material	Copper
Wire diameter (bare)	0.0028 in (0.071 mm)
Wire thickness incl. insulation	0.0029 in (0.074 mm)
Bobbin internal diameter	0.074 in (1.9 mm)
Bobbin wall thickness	0.005 in (0.125 mm)
Bobbin material	Kapton®

Table 5.2 Properties of the Helmholtz coil

Nominal diameter	3.45 in (88 mm)
Internal diameter	3.02 in (77 mm)
No. of turns in each coil	625
No. of layers in each coil	25
Average resistance (combined)	49.5 Ω
Wire gauge	AWG26
Wire material	Copper
Wire diameter	0.0159 in (0.404 mm)
Bobbin internal diameter	3.0 in (76.2 mm)
Bobbin material	Aluminum

50 mA required by the electromagnetic coils to be driven, a current amplification circuit was developed. To avoid damage to the system, necessary protection measures using a voltage following circuit were incorporated into the circuit. The circuit diagram shown in Fig. 5.8 corresponds to the voltage following circuit and the voltage to current conversion circuit for one channel. The associated transfer function from the voltage input V_{in} to the resulting current in the coil I_0 is

$$I_0 = \left(\frac{1}{R_{\text{s}}} + \frac{R_{\text{s}} - R_{\text{L}}}{R_{\text{s}} \left(R_{\text{L}} + 2R_{\text{f}} \right)} \right) V_{\text{in}} \tag{5.1}$$

where $R_{\text{L}} = R_{\text{c}} + L_{\text{c}}s$, R_{c} is the resistance of the coil, and L_{c} is the inductance of the coil. For the case of $R_{\text{f}} \gg R_{\text{s}}$ and R_{c}, when the system works in the low-frequency range, the current I_0 of the coil in (5.1) can be approximated as

$$I_0 = \frac{V_{\text{in}}}{R_{\text{s}}} \tag{5.2}$$

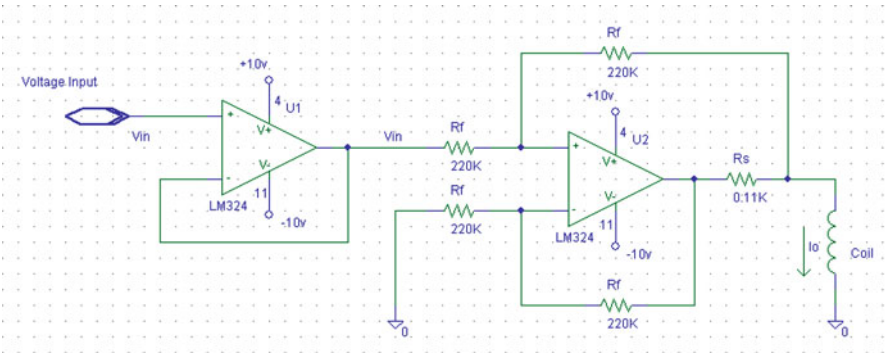


Fig. 5.8 Current amplification and protection circuits including a voltage following circuit and a voltage-controlled current source circuit. The diagram shows the circuit for one channel only. The circuit can be replicated to any desired number of channels

5.2 Experimental Setup

The prototype MFDM described above has been incorporated into an experimental adaptive optics setup. The system is designed to be used as platform for testing, characterization, and evaluation of the MFDM in the first stage and can be subsequently complemented with a science camera and auxiliary components that will render it a complete AO retinal imaging system.

5.2.1 Layout of the System

Figure 5.9 shows the snapshot of the experimental AO setup used in this study for validations. Important components of the system have been indicated. The schematic diagram of the system is given in Fig. 5.10. The components drawn as dotted lines in the figure represent the imaging subsystem that can be incorporated at a later stage.

In the final ophthalmic AO imaging system, the optical relay indicated as R_1 will be eliminated and a thin beam of laser or a super-luminescent diode (SLD) light will be directed into the eye using the beam splitter marked as BS_2. The light reflected from the eye, which exits the eye as a beam with a diameter equal to that of the pupil of the eye, will be directed to the MFDM using the optical relay R_2. Since the main purpose of the current system is only the characterization and control of the MFDM, the system is modified accordingly as described below.

A collimated, aberration-free beam of light from the laser source is magnified using the first optical relay R_1. The magnified beam is limited by an aperture stop

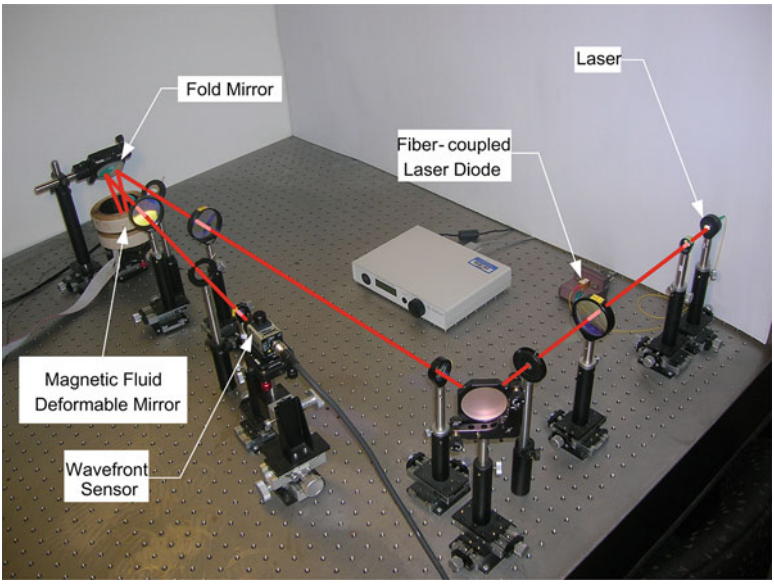


Fig. 5.9 Snapshot of the experimental setup

with 5.5-mm diameter, which is the average size of the pupil of a dilated human eye. The 5.5-mm beam is diverted and further magnified (×4) using relay R_2 to 22 mm. The 22-mm beam is directed on the horizontal fluid surface using the fold mirror, which also collects the reflected beam and folds it back to the Shack–Hartmann-type wavefront sensor at an angle of $\sim 8°$ from the incoming beam. The reflected beam, now carrying the aberrations induced by the mirror, is demagnified (×6) to 3.6 mm in order to be projected fully on the lenslet array of the Shack–Hartman-type wavefront sensor. The wavefront sensor measures the slope of the incoming beam using an array of 32×32 lenslets. The wavefront slope information is acquired by a PC-based reconstruction and control software, which provides necessary current inputs to the MFDM through an electronic control unit.

5.2.2 Description of the Main Components

Important features of the main components of the system are described below.

5.2.2.1 Wavefront Sensor

The Shack–Hartmann-type wavefront sensor (HASO 32 by Imagine Optic, Orsay, France) comprises a 32×32 array of lenslets. The sensor has an aperture dimension

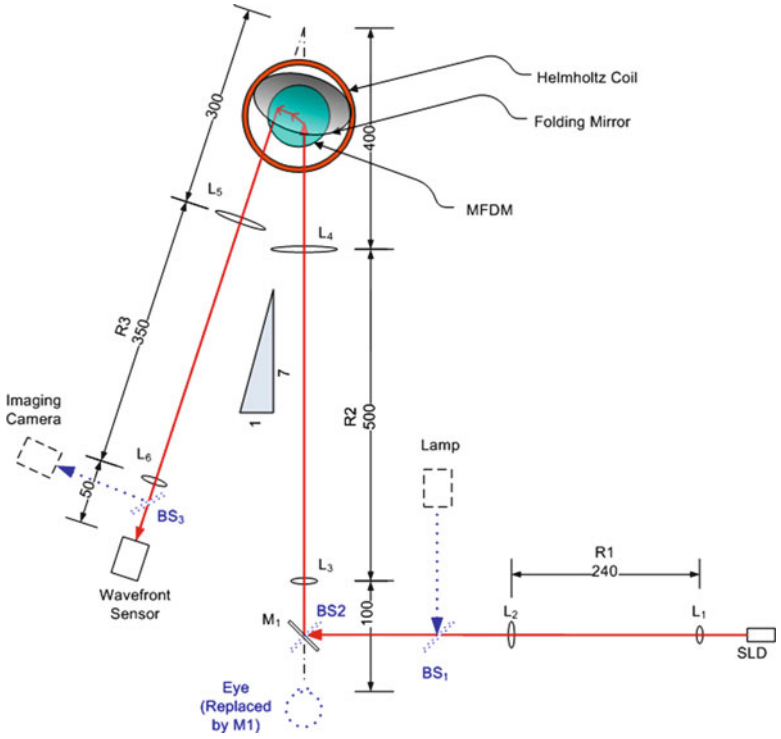


Fig. 5.10 Schematic layout of the experimental setup. The components shown in the *blue* color (*dashed lines*) are to be incorporated later. The subscripted letters L, M, R, and BS indicate a lens, mirror, relay, and beam splitter, respectively. All dimensions are in millimeter (mm)

of 3.6×3.6 mm backed by a CCD camera detector (Toshiba Teli CS8550i-01) and operates at a maximum acquisition frequency of 30 Hz. The wavefront slope data, measured discretely at the 32×32 subapertures of the wavefront sensor, is acquired by a PCI frame grabber (Euresys Domino Alpha2 acquisition board) which provides an interface to a host PC.

5.2.2.2 Laser Diode

The reference light used for the measurement of aberrations is provided by a 661-nm fiber-coupled diode laser source (QFLD-660-10S by Nolatech JSC, Moscow, Russia). The laser source can provide up to 10 mW of optical power, which is controlled by a current driver (QSDIL-300). The single-mode fiber-coupled laser diode with a 14-pin DIL packaging is mounted on a temperature controller DIL mount. The FC/APC terminated fiber is coupled to a dual aspheric lens collimator.

Table 5.3 Specifications of the optical components

	Focal length (mm)	Diameter (mm)
L_1	40	12.5
L_2	200	50.8
L_3	100	25.4
L_4	400	50.8
L_5	300	50.8
L_6	50	12.5

5.2.2.3 Optical Components

The optical system broadly comprises three optical relays (R_1, R_2, and R_3), which magnify and demagnify the light beam for proper projection on the MFDM and the wavefront sensor. The relays are constructed by plano-convex lenses coated with antireflection coatings. Table 5.3 provides the specifications of the lenses used in the setup. A manually operated fold mirror is used for directing the light beam to the MFDM whose surface lies essentially in the horizontal plane.

5.2.2.4 Software

The control software for the experimental AO system consists of three main components. The first component is the wavefront acquisition and reconstruction module. The acquisition and reconstruction functions used in this work are provided by the manufacturer of the wavefront sensor, Imagine Eyes Inc. The second component is the software implementation of the control algorithms. All controllers are written in discrete-time state-space form and are implemented as a C++ class. The third part is the driver software for the MFDM, which has been developed using the driver functions provided by National Instruments. The whole system is integrated using a C++ console application accessing the three components in a closed loop. A single PC, using a 4200+ AMD Athlon 64×2 Dual Core processor with two gigabytes of RAM, centrally controls the sensor acquisition, controller data processing, and MFDM actuation.

5.2.3 System Assembly

The experimental optical setup is composed of lenses, beam splitters, and mirrors. Essentially, the system optically conjugates the exit pupil of the eye, given by the virtual image of the retina through the cornea, with the plane of the DM and the wavefront sensor. This means that a point that is in focus at the exit pupil of the eye will be in focus on the plane of the DM and the wavefront sensor, thus minimizing aberrations caused by the hardware misalignment of AO system. Conjugation is

achieved by using relay telescopes formed using two positive lenses. All lenses are off-the-shelf plano-convex lenses. Figure 5.9 shows a picture of the actual system with its main components. Figure 5.10 shows a schematic of an AO system layout. In Fig. 5.10, conjugate planes are indicated with the letters P and R, and the direction of the light is shown with arrows. The image of the retina on the DM is magnified by a factor of 4. The image of the retina from the DM to the WFS's CCD detector is magnified by a factor of −6.

For the purposes of sensor alignment and testing, an artificial eye based on a dynamic eye model presented in Fernandez and Artal (2007) was used. The artificial eye consists of a doublet lens (acting as the artificial lens of the eye) and a rotating diffuser placed at a fixed distance from the lens (acting as the artificial retina). Different refractive errors can be simulated by rotating the diffuser. The artificial eye can be used for alignment and testing of the AO system due to its ability to create low-order aberrations.

In most AO systems, tip and tilt aberrations are compensated for using an additional (costly) mirror, known as a tip-tilt mirror. The tip-tilt mirror is needed because of the low achievable stroke of most of the electrostatic mirrors used in many AO systems for retinal imaging. Due to the high achievable stroke of the MFDM used in this AO systems, tip and tilt aberrations will be corrected for using the WFC, substantially reducing the overall cost of the system.

The DM and the WFS all lie in conjugate planes with the pupil of the eye, and since they all have different diameters, this requires optical relays, known as relay telescopes, to adjust the diameter of the beacon light as it propagates through the system. Before entering the eye, the beacon light shares a common path with the beacon light exiting the eye. In order to couple the beacon light path with the rest of the system, a pellicle beam splitter (PS) is used to direct light into the eye. Pellicle beam splitters are ideal because they minimize light dispersion and eliminate ghosting from the system. The pellicle beam splitter's membrane is extremely uniform exhibiting less than 1/2 wavelength of variation across a 25-mm beam splitter, using a 661-nm light source. This minimizes the wavefront error introduced into the reflected beam to less than half a wavelength, thus reducing the amount of the wavefront aberration introduced by the system itself. The entire system was designed to fit into a compact space, with dimensions of 1 m × 1 m.

The system control is performed using a single processor personal computer (PC) as shown in Fig. 5.11. The WFS data is collected by a frame grabber (Euresys Domino Alpha2 acquisition board) and sent to the PC, which uses the data to reconstruct the wavefront of the beacon light. The reconstructed wavefront is then used by a controller (also housed in the PC) to generate the required actuator signals, in the form of voltages, for each of the actuators in WFC. The actuator signals are sent to an analog voltage output card (National Instruments 6723), which transfers control signals to the current control circuit. Actuator current signals are then sent from the current drive box to the WFC. A detailed arrangement of the components used in this AO system is provided in Appendix D.

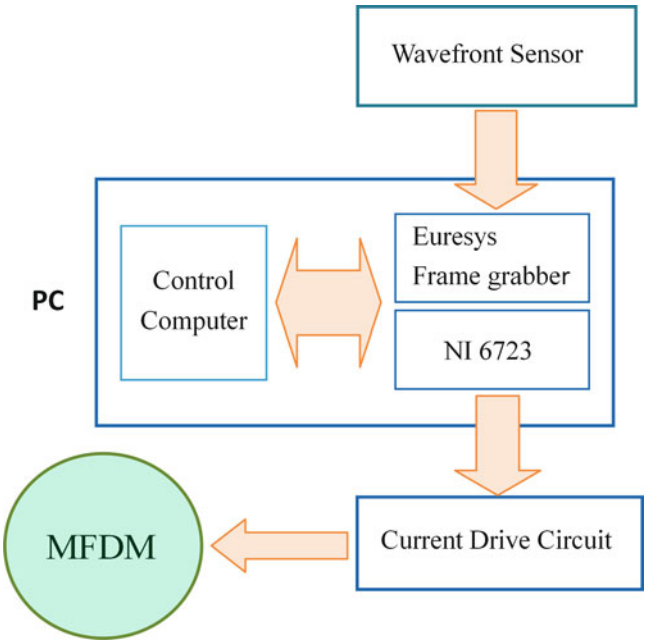


Fig. 5.11 Data flow for AO system with a MFDM

5.2.4 Alignment Procedure

Optical alignment refers to the process of adjusting the positions of various optical components in the system so that the AO system is set up optically as designed. A number of approaches can be taken to align the optics of an AO system. Many of these include the use of specialized equipment such as alignment telescopes and interferometers, which can add to the cost of an AO system. The optical alignment procedure adapted in this work involves four steps. The first step is to layout the main optics (the beacon light, DM, and WFS with respect to the position of the pupil of the eye) on an optical table according to design specifications. This will ensure that the components are placed close to the nominal design. It should be noted that the WFC and WFS are on X-Y-Z stages that allow precise movements in the *X*-*Y*-*Z* directions using adjustment dials on the stages.

The second step is to establish the optical axis of the system. In order to establish the optical axis, a line of sight (LOS) and set of targets are needed to define the intended optical axis. The targets will be permanent targets that can be continuously used throughout the life of the AO system to check for alignment. For this application, irises are used that can be closed when the targets are needed and opened, allowing the beacon light to pass through, when the targets are not needed. The LOS acts as the optical axis, making it easier to center the optics of the system, such as sources and detectors, with respect to the optical axis. In order to properly

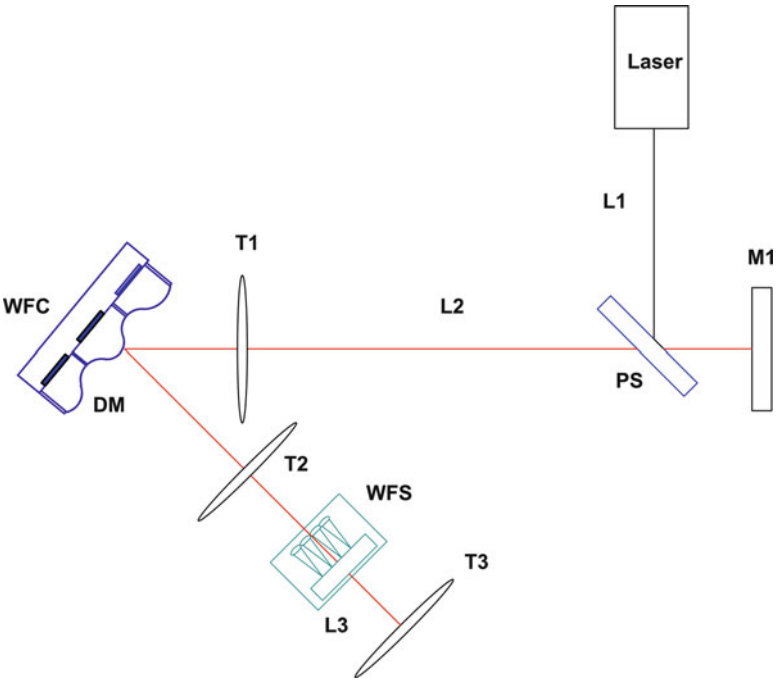


Fig. 5.12 Setting the LOS in the AO system

establish the LOS, permanent targets (T1, T2, and T3) are placed at prespecified positions. These targets are used to align the beacon light (or laser light) as it passes through different components in the system, shown in Fig. 5.12. Once the first leg (L1) of the system (the optical path from the alignment laser to PS) is aligned, the second leg (L2, the optical path from the eye to the DM) is aligned. The beacon light reflects from a flat mirror (M1, used to represent the pupil of the eye during alignment) and is directed toward the DM. The beacon light then reflects from the DM and is directed toward a second target, T2. The position of the DM can be adjusted using its X-Y-Z stage so that the laser hits T2 on target. T2 is then opened allowing the beacon light to pass through, hitting the third target, T3. The DM is adjusted until the light can hit both T2 and T3 at their centers. Next, the WFS is aligned by positioning the sensor using its alignment setting with the beacon light. The WFS is adjusted using its X-Y-Z stage so that the beacon light hits its lenslet array in the center.

The third step is to complete the alignment of the afocal relay telescopes in the system. It should be noted that a relay telescope is a lens arrangement, shown in Fig. 5.13, which accepts a collimated beam in and outputs a magnified collimated beam. The relay telescopes, shown in the Fig. 5.13, will each be aligned online using the laser. The lenses are centered using the predetermined LOS set up in step 2. If the LOS passes through the nodal point (or the center) of the lenses, then the LOS will

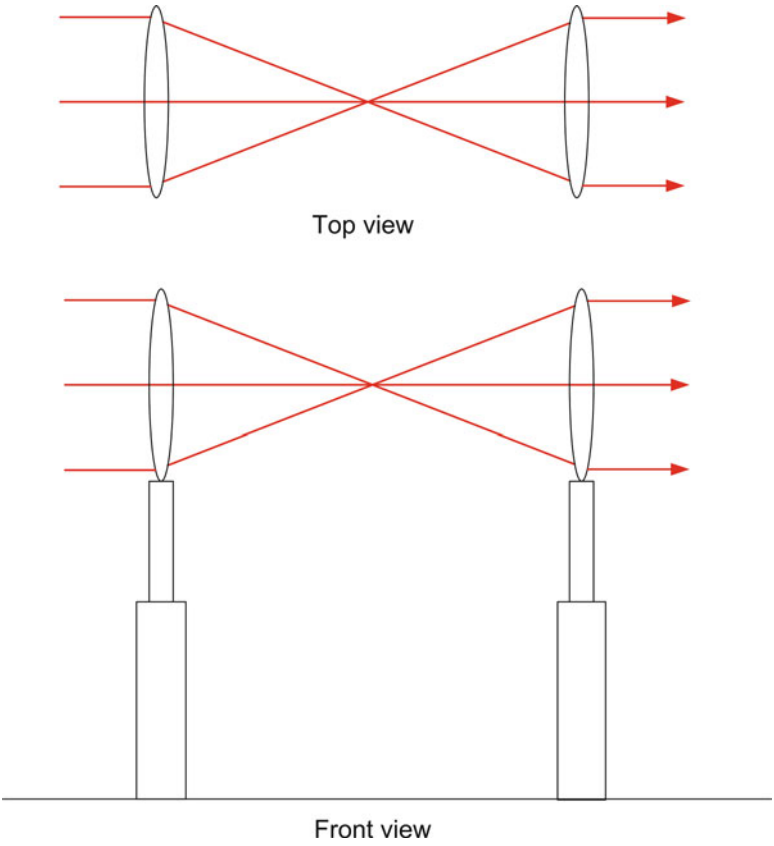


Fig. 5.13 Aligning the relay telescopes for the AO system

not deviate from its path. Therefore the height of each of the lenses is adjusted until the LOS hits the targets. Once the lenses are centered, the distance between the two lenses can be set using a shear plate. The two lenses can be adjusted with respect to each other until the output of the relay is a collimated beam. This will ensure that the two lenses are positioned correctly with respect to one another. The shear plate can be used to determine if the beam is collimated by analyzing the beam's wavefront. The shear plate is positioned in the optical path of the system immediately following the relay telescope that is being assembled. If the beam exiting the relay telescope is collimated, a fringe pattern is created, as shown in Fig. 5.14a. If the beam is not collimated, the fringe pattern will be misaligned, as shown in Fig. 5.14b. The relay telescope is then adjusted until the fringe pattern is similar to the pattern in Fig. 5.14a. The relay in L3 can be set in a similar manner. The relay telescopes will be known as RT2 and RT3, corresponding to L2 and L3.

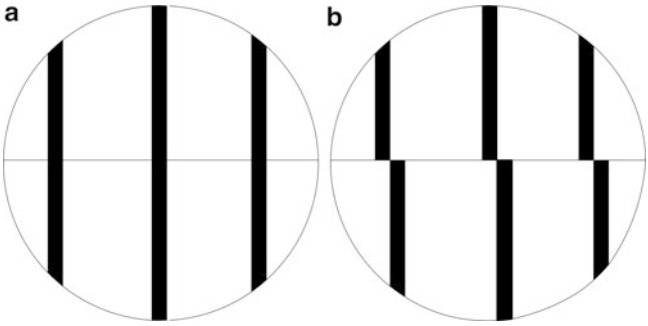


Fig. 5.14 Fringe pattern created by the shear plate: (**a**) collimated beam and (**b**) diverging beam

The fourth step is to align the conjugate planes of the system. In most AO systems, two sets of conjugate planes are important (Hecht 2002). The first set of planes is conjugate to the eye's retina and includes the aperture plane of the reference beacon and the CCD plane of the WFS. The second set of conjugate planes is conjugate to the eye's pupil and includes the DM and the lenslet array of the WFS. Since the field of view of the system is relatively small compared to its overall dimension, there is very little penalty in having the conjugate planes of the eye's pupil slightly out of position (viz., by a millimeter or two). Thus, the conjugate planes can easily be positioned using precise measurement. More detail regarding the alignment procedure and the design of the AO system are presented in Appendix D.

5.2.5 Miscellanea

5.2.5.1 Uniformity of the Initial Surface

The Helmholtz coil used in the prototype MFDM has a uniformity of less than 0.1% as measured at 20-mm diameter. Although small in magnitude, the nonuniformity of the magnetic field of the Helmholtz coil does produce a curvature in the initial surface of the fluid. Moreover, any misalignment in the vertical axis of the Helmholtz coil and the gravitational vector, as well as the imperfections in the optical components, result in a non-flat initial wavefront surface as measured by the wavefront sensor. Though the tilt and the curvature in the initial surface measurements can be minimized by properly adjusting the fold mirror (see Fig. 5.9) and the position of the wavefront sensor, some deformations of the initially measured wavefront surface remain uncorrected. Generally, these initial imperfections in the system are eliminated by presetting the actuators, which ensures a planar initial wavefront surface.

5.2.5.2 Effect of the Earth's Magnetic Field

The effects of the earth's magnetic field have been ignored. For our experimental setup with the miniature coil strength $b_z \approx 0.25$ mT and earth's magnetic field ≈ 0.055 mT in Toronto, Canada (Haines and Newitt 1997), the approximation is valid. However, with any smaller coils, the effects of the earth's magnetic field may be significant. Under such circumstances, the vertical component of the earth's magnetic field can be directly added to the uniform field of the Helmholtz coil without affecting the validity of the model.

5.3 Experimental Evaluation

This section is aimed at experimentally validating the analytical model developed in Chap. 4 using the prototype MFDM and the AO setup described in Sect. 5.2. After some preliminaries given in the following section, the analytical model is validated by comparing the predictions made by the model with the corresponding experimental results. The effects of the linearization of the MFDM response using a Helmholtz coil are investigated and verified. The experimental evaluation of the performance of the controllers will be presented and given later in Chaps. 7 and 8.

5.3.1 Preliminaries

5.3.1.1 Parameter Settings

Before presenting the experimental results, it is pertinent to specify the settings used in obtaining the results. The common settings are listed as follows:

- The magnetic fluid properties used in the analytical model of the MFDM are those of EFH1 as listed in Table 4.1.
- The MFDM model (4.102) is truncated to the following numbers of radial and azimuthal modes, respectively: $N = 4$ and $M = 4$.
- All results are obtained using a magnetic fluid layer which has a thickness of 1.0 mm.
- If not otherwise stated, the Helmholtz coil is set to run at 194 mA current, which corresponds to a nominal magnetic flux density of 2.5 mT.
- The optical pupil size, that is, the diameter of the laser beam, is set to 22 mm as measured on the MFDM surface. All optical metrics, for example, the PSF and the RMS of the wavefront shape error, are computed using this pupil size.

5.3.1.2 System Identification

System identification (SI) is a black-box approach to system modeling, which is used to estimate models of dynamic systems from experimentally measured input–output data. A supplementary model of the dynamics of the MFDM surface shape was obtained using this approach and has been used in the validation of the analytical model. The identified model of the dynamics of the MFDM surface shape is briefly described here. The identified model is obtained using the System Identification Toolbox of Matlab^{R} (Ljung 2009). Multiple sets of input–output data are generated by applying suitable currents to the electromagnetic coils and measuring the surface deflections at selected output positions using the wavefront sensor. The input–output data is then fit to a prespecified model structure. The prediction error minimization method (Ljung 1999) is used to estimate a discrete-time state-space model Σ_{SysID} of the MFDM response:

$$\Sigma_{\text{SysID}} : \begin{cases} \mathbf{x}(k+1) = \hat{\mathbf{A}}\mathbf{x}(k) + \hat{\mathbf{B}}\hat{\mathbf{u}}(k) \\ \hat{\mathbf{y}}(k) = \hat{\mathbf{C}}\mathbf{x}(k) + \hat{\mathbf{D}}\hat{\mathbf{u}}(k) \end{cases} \tag{5.3}$$

where $\hat{\mathbf{u}}$ and $\hat{\mathbf{y}}$ are the experimentally generated inputs and outputs, respectively, and $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$, $\hat{\mathbf{C}}$, $\hat{\mathbf{D}}$ are the estimated system matrices.

To get the best estimates of the system matrices, multiple sets of input–output data are generated by applying step, sinusoidal, and multifrequency inputs to each of the input channels and observing the outputs over time. The data is preprocessed by removing any spurious readings and trends. One set of data, obtained with the multifrequency inputs, is used as the primary data for identification, while the remaining sets are used for validation purpose. The identification process is done iteratively such that the prediction error is minimized, while the estimates of the system matrices are updated in each iteration. The order of the system is gradually increased until the simulated model output best fits the measured data. The identified model thus obtained for a single-input single-output system is given as:

$$\hat{\mathbf{A}} = \begin{bmatrix} 0.672 & -0.345 & -0.063 \\ 0.636 & -0.031 & 0.411 \\ -0.012 & -0.549 & -0.606 \end{bmatrix}$$

$$\hat{\mathbf{B}} = \begin{bmatrix} 0.079 \\ -0.137 \\ -0.266 \end{bmatrix}$$

$$\hat{\mathbf{C}} = \begin{bmatrix} 61.103 & 16.155 & 4.802 \end{bmatrix}$$

$$\hat{\mathbf{D}} = [0.000]$$

5.3.1.3 Jones' Model of Static Response

In the validation of the analytical model that will follow, in addition to comparing the predictions of the analytical model with the experimental results, a static model of the mirror response will also be used for comparison. The static model has been adapted from an existing model of the displacements of the free surface of a magnetic fluid presented by Jones (Jones 1988). Using Jones' model, the displacements of the free surface of a magnetic fluid produced by an applied magnetic field with the flux density b_z can be written as

$$\bar{\varsigma} = \frac{\chi b_z^2}{2\rho g \mu} \tag{5.4}$$

As presented here, the model is applicable to the special case where the magnetic fluid surface is exposed to a vertical magnetic field, that is, the flux density of the magnetic field has only a vertical component b_z and no horizontal component. Jones' model can be utilized to derive the displacement of the MFDM surface where a large magnetic field with the flux density B_0 is superimposed on a smaller magnetic field with the flux density b_z and where both of the fields are acting in the vertical direction. The derived surface displacements in this case can be found as

$$\hat{\varsigma} = \frac{\chi b_z B_0}{\rho g \mu} \tag{5.5}$$

This model can be used to obtain static displacements of the MFDM surface at the surface locations where the magnetic field of the miniature electromagnetic coils presents only a vertical component and no horizontal component. One such location is the peak of the surface shape produced by applying current input to only one electromagnetic coil and will be used in the following paragraphs for the comparison of the analytical model with Jones' model.

5.3.1.4 Initial Surface Map and Measurement Noise

The experimental results presented in this chapter are based on wavefront shape measurements obtained using a Shack–Hartmann wavefront sensor. Before presenting the results, a brief discussion on the accuracy of the measurements is presented in this section.

Theoretically, in absence of any input currents applied to the wavefront corrector and with no externally induced wavefront aberrations, the wavefront shape as measured by the wavefront sensor should be a perfect planar surface. However, practically achieving this is almost impossible because the optical and electromechanical components of system itself induce aberrations, which cannot be completely eliminated in the initial process of installation and alignment of these components. Therefore, the wavefront sensor reads an initial wavefront shape $\mathbf{y}_{s0}$ that needs to

be taken into consideration in all subsequent measurements. The initial wavefront shape is dynamic in nature, that is, it varies over time. However, the initial wavefront shape measurement $\mathbf{y}_{s0}$ has a relative large mean component that can be considered to be static and is called the initial surface map. It is a common practice in AO systems that the effect of the initial surface map is canceled out by giving a bias to the wavefront corrector using a set of initial inputs. On the other hand, the dynamic component of the initial wavefront shape measurement remains uncorrected. The static and the dynamic components of the initial wavefront shape measurement are discussed in the following.

Static Surface Map

The initial static surface map and the set of inputs used to cancel its effect were determined as follows. With no inputs applied to the MFDM, a series of wavefront measurements is taken such that at any instant k, $k = 1, 2, \ldots, K$, $\mathbf{y}_{s0}(k)$ is a vector of wavefront shape displacements measured at M discrete lenslet locations contained in the optical pupil. The initial static surface map is obtained as a mean surface defined by

$$\bar{\mathbf{y}}_{s0} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{y}_{s0}(k) \tag{5.6}$$

Once the initial static surface map $\bar{\mathbf{y}}_{s0}$ is known, the vector of input currents needed to cancel its effect can be obtained as follows:

$$\mathbf{u}_0 = \mathbf{G}_0^{-1} \bar{\mathbf{y}}_{s0} \tag{5.7}$$

where $\mathbf{G}_0$ is the DC gain of the system as given in (4.104) or (4.105) which can be obtained analytically. The application of this initial bias flattens the resulting wavefront shape, which becomes the reference for all subsequent wavefront shape measurements. Before applying any input currents computed in order to produce any desired wavefront shape, the initial input currents are added to these inputs. The surface plot of the initial surface map of the setup used to obtain the results presented in this chapter is given in Fig. 5.15. The set of input currents needed to flatten the surface is given as

$$\mathbf{u}_0 = [-0.4, 3.7, 23.6, 14.5, 28.0, 9.8, 17.3, 14.0, -9.9, 12.8, 11.4, 7.5, 8.4, \ldots$$
$$-2.9, 3.1, 2.2, -15.9, 19.8, 7.3] \times \frac{1}{31.5 \times 10^3} \text{A}$$

The surface plot of the resulting wavefront shape obtained after applying the bias currents is given in Fig. 5.16.

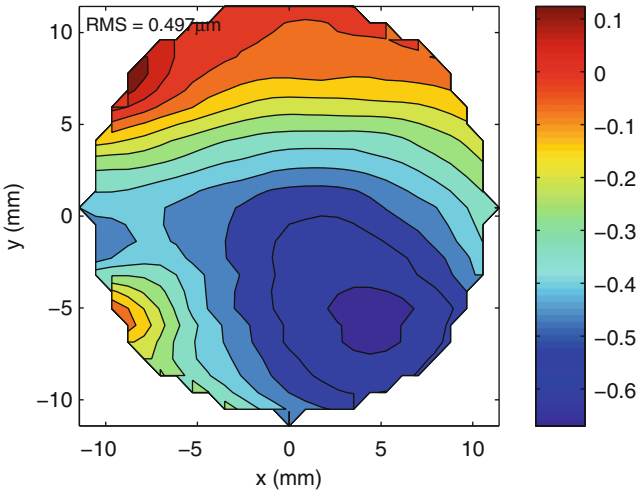


Fig. 5.15 Initial surface map

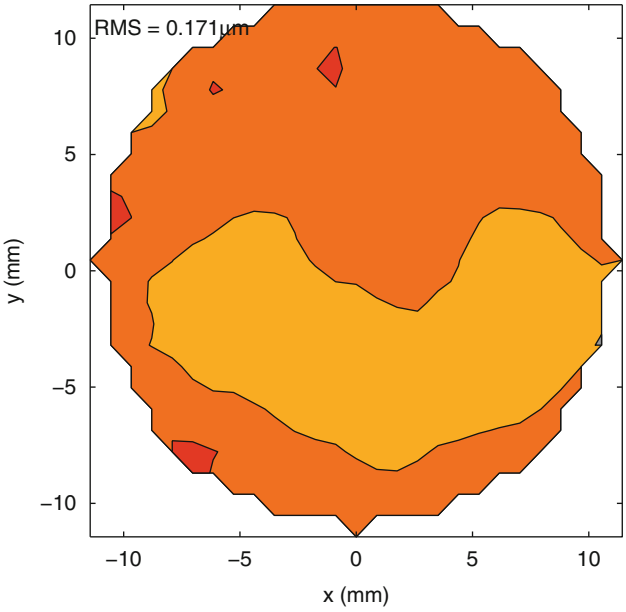


Fig. 5.16 Flattened surface

Dynamic Noise Component

The dynamic component of the initial wavefront shape remains uncorrected in the open-loop measurements and can be considered as measurement noise. Unless mitigated by closed-loop control system, the measurement noise poses a limit on the

performance of the AO system. An estimate of the magnitude of the measurement noise can be obtained by computing the standard deviation of the initial wavefront shape measurements $\mathbf{y}_{s0}(k)$ given as

$$\bar{\sigma} = \sqrt{\frac{1}{K}\sum_{k=1}^{K}(\mathbf{y}_{s0}(k) - \bar{\mathbf{y}}_{s0})^2} \tag{5.8}$$

The mean of the standard deviations computed over the 19 surface locations corresponding to the 19 actuator coils was found to be 0.181 μm. Another similar measure of the dynamic component of the initial surface shape is the standard deviation of the RMS values computed at instants k, $k = 1, 2, \ldots, K$. The computed standard deviation of the RMS values of the experimentally obtained initial surface shapes is 0.173 μm.

5.3.2 Model Validation

The results of an experimental investigation aimed at validating the analytical model are presented in this section. In the following subsection, the static response of the prototype MFDM is compared to the response predicted by the analytical model. The dynamic characteristics of the mirror as predicted by the model are also experimentally validated in Sect. 5.3.2.2.

5.3.2.1 Static Response

As the first step toward validating the analytical model, the static surface deflections predicted by the model are compared with those obtained experimentally. The steady-state response of the MFDM surface can be analytically derived from the state-space model given in (4.102) and is given as

$$\mathbf{y}_{ss} = -\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\mathbf{u}_{ss} \tag{5.9}$$

where $\mathbf{u}_{ss}$ is the vector of constant currents applied to the array of 19 electromagnetic coils and the system matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are 19 by 19 matrices, respectively. The steady-state response of the surface to a set of constant input currents is a static surface shape, which can be obtained from (5.9). The static response of the actual MFDM can be directly measured using the wavefront sensor.

Figure 5.17 presents the comparison of the experimentally obtained surface deflections with those predicted by the presented model. The figure shows the surface deflections resulting from applying various input currents to only the center coil (coil # 1 in Fig. 5.7). Since the resulting deflections are symmetric about the vertical axis, only 2D surface shapes are shown. The RMS of the error between the

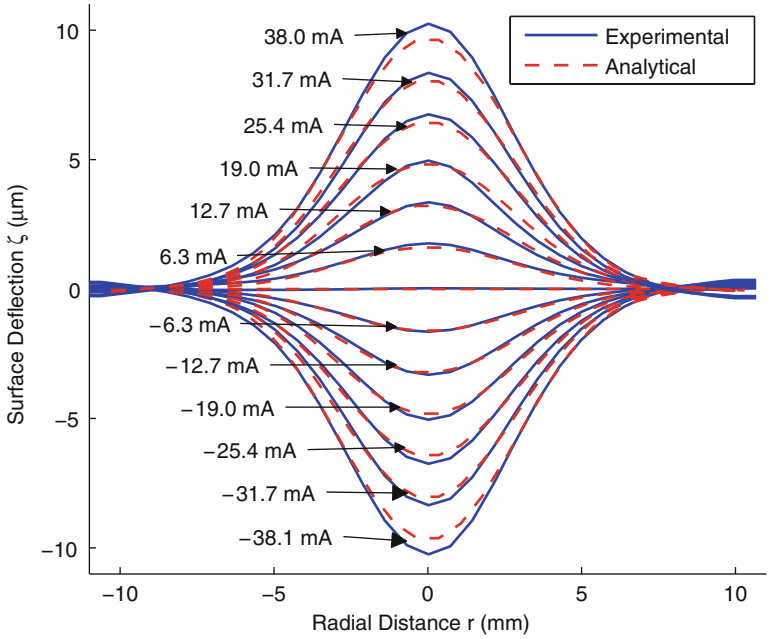


Fig. 5.17 Comparison of the static surface shapes obtained experimentally versus those predicted analytically

two corresponding surfaces computed over the 22-mm pupil is shown in Fig. 5.18. It is evident from the computed RMS errors that the steady-state surface deflections predicted by the presented model closely agree with those obtained experimentally. Besides the comparable deflections, the model also reveals the Gaussian profile of the deformed surface.

The peak surface displacement predicted by the analytical model can also be compared to those given by Jones' model as given in (5.5). Figure 5.19 shows the comparison, where it can be observed that the analytical model shows a better conformance to the experimental results than Jones' model. It is reiterated that the comparison of the Jones' model to both the experimental results and the analytical model presented in this section is valid only for the peak surface displacements where the magnetic field has only a vertical component and the horizontal component is zero. It may also be noted that, contrary to the analytical model presented in this book, Jones' model does not consider surface tension of the magnetic fluid.

The analytical model (4.102) represents a MIMO system. Experiments show that the surface deflections determined by the model are just as valid for the simultaneously applied multiple inputs as for the case of the single input. The 3D surface shapes resulting from a specified set of input currents

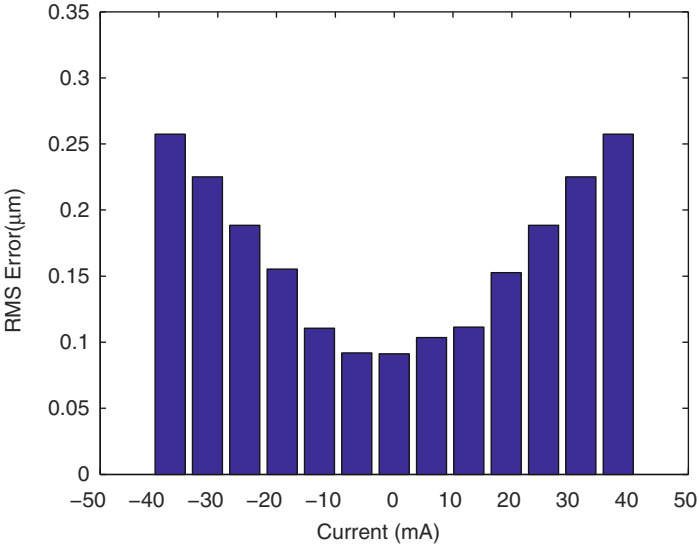


Fig. 5.18 RMS error between the analytical and experimental surface shapes of the MFDM measured at different input currents

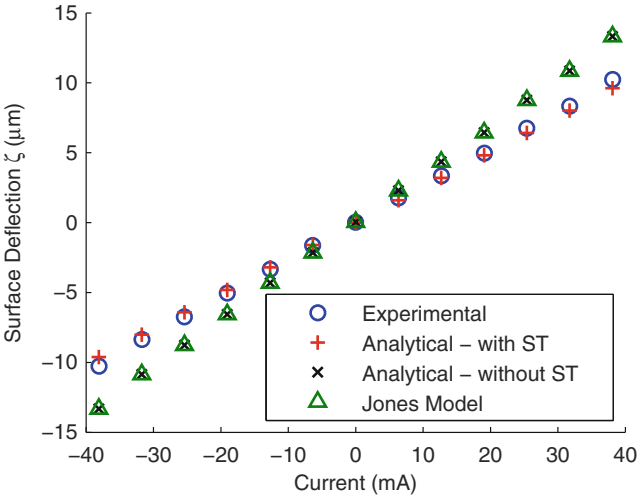


Fig. 5.19 Comparison of the peak surface displacements given by the analytical model, Jones' model, and the experimental results

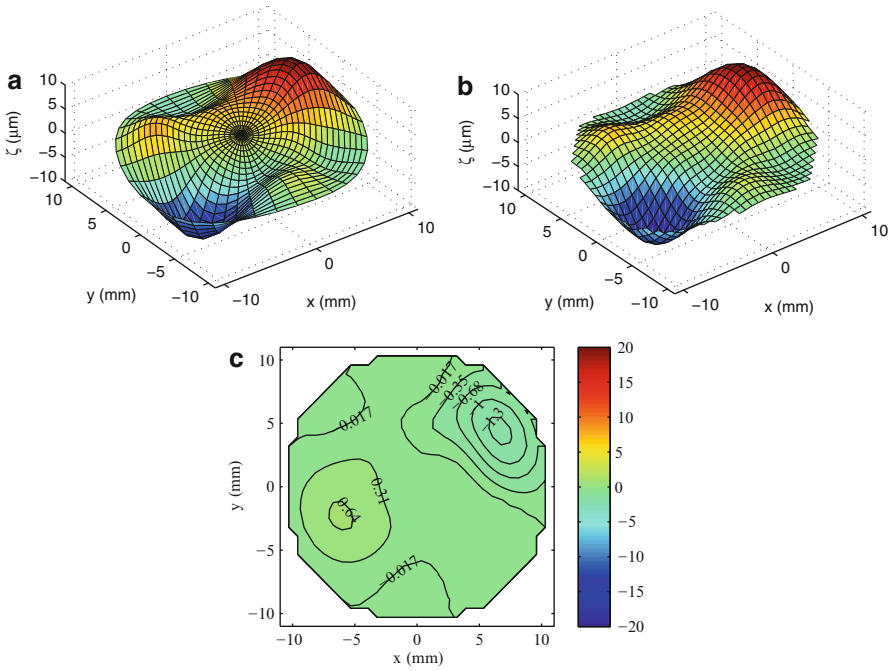


Fig. 5.20 Comparison of the 3D static surface shapes: (**a**) analytical, (**b**) experimental, and (**c**) error (The color bar and contour labels show the surface shapes in micrometers)

$$\mathbf{u} = [0.1, 0.5, 0.4, 0.3, 0.1, -0.1, 0.2, 0.9, 0.7, \ldots$$
$$0.6, -0.5, 0.6, 0.1, -0.4, -0.9, -0.3, 0.5, -0.3, 0.4] \times \frac{1}{31.5}\mathrm{A}$$

applied to the array of 19 coils are presented for comparison of the analytically predicted surface shapes with those obtained experimentally. Figure 5.20a shows the surface shape predicted by the analytical model. The corresponding shape obtained by applying the currents to the prototype MFDM is given in Fig. 5.20b. Figure 5.20c shows the contour plot of the shape error computed as the difference of the surfaces shapes shown in Figs. 5.20a, b. The contours are spaced at $\lambda/2$, where λ is the wavelength of the laser light used. The RMS of the shape error computed using the discrete displacements at the wavefront sensor lenslet locations contained in the pupil is 0.19 μm. Given the noise conditions in the laboratory as discussed in Sect. 5.3.1.4, the resulting RMS error shows that the surface shape predicted by the analytical model compares fairly well with the experimentally obtained shape.

5.3.2.2 Dynamic Response

In the following, important dynamic response parameters predicted by the model are compared with the corresponding experimental results. The dynamic response of the MFDM surface to a step, a sinusoidal, and a multitone input is presented. The Bode plots of the analytical model and the identified model are compared, and important features highlighted by the plots are discussed.

Response to a Step Input

Figure 5.21 presents a comparison between the step response of the actual MFDM and that of the analytical model of the mirror. A step input of 32 mA is applied to the center coil (coil # 1 in Fig. 5.7) of the MFDM. The time history of the mirror surface deflections immediately above the center of the coil is noted and plotted as shown in Fig. 5.21a. The response of the actual mirror surface is shown by the solid line (blue), while the analytically determined surface deflections are represented by the dashed line (red). The deflections predicted by the identified model are given as the dash-dotted line (magenta). For a direct visual comparison, the input current applied to the electromagnetic coil is scaled appropriately and is plotted against the right-hand side axis. The input current plot is shown as the dotted line (black). The error in the analytically predicted surface displacements as measured against the experimental results and the predictions of the identified model is presented in Fig. 5.21b. As can be observed from the figure, the response given by the analytical model agrees fairly well with the experimental results as well as with the identified model.

Response to a Sinusoidal Input

Figure 5.22 shows the response of the MFDM to a sinusoidal signal with 3 Hz frequency and an amplitude of 32 mA. Again, the input signal is applied to the center coil (coil #1 in Fig. 5.7). The output is measured at the location immediately above the center of the coil and is plotted in Fig. 5.22a. The experimentally obtained deflections of the mirror surface are represented by the solid line (blue); the response obtained using the analytical model is shown by the dashed line (red); and the dash-dotted line (magenta) represents the response given by the identified model. The sinusoidal input current is shown on the right-hand side axis with the plot drawn using a dotted line (black). Note that the scale of the right-hand side axis is the same as the one used in the response to the step input plotted in Fig. 5.21 above, thus highlighting the drop in the amplitude of the output surface deflections with the increased frequency of the input current signal. The error in the analytically

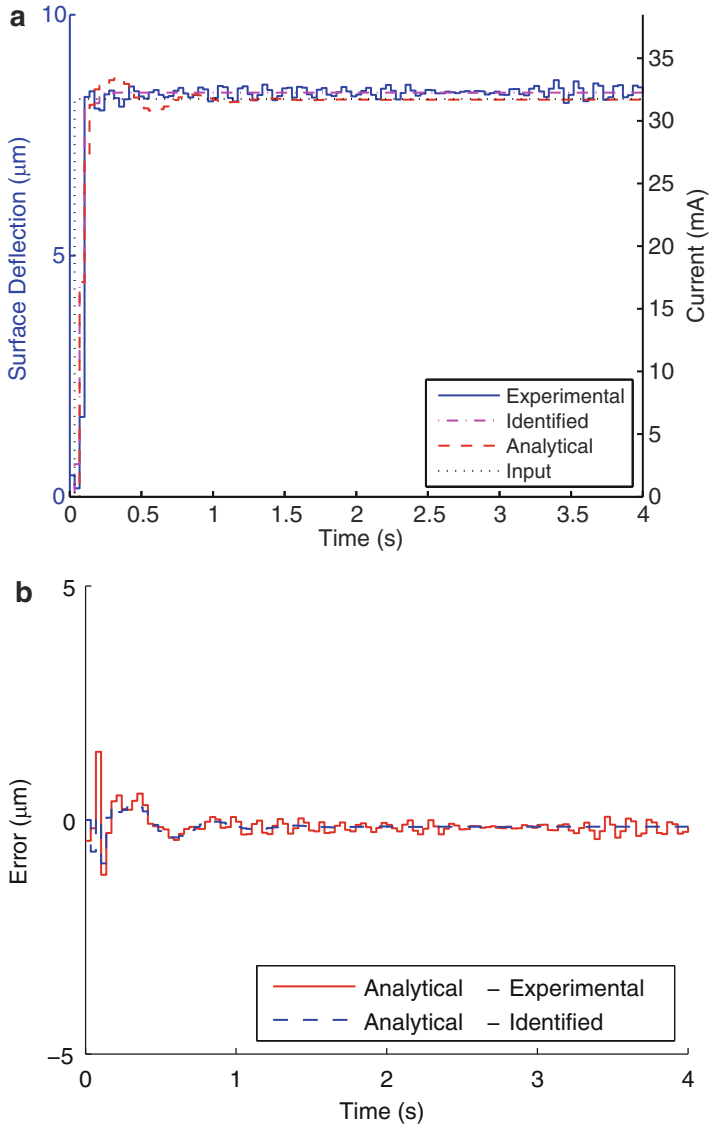


Fig. 5.21 Response to step input: (**a**) surface deflections and (**b**) error

predicted surface displacements as measured against the experimental results and the predictions of the identified model is presented in Fig. 5.22b. Although a significant error can be seen in the shown plot, it remains less than 10% of the peak-to-valley deflections of 15 μm as shown in Fig. 5.22a.

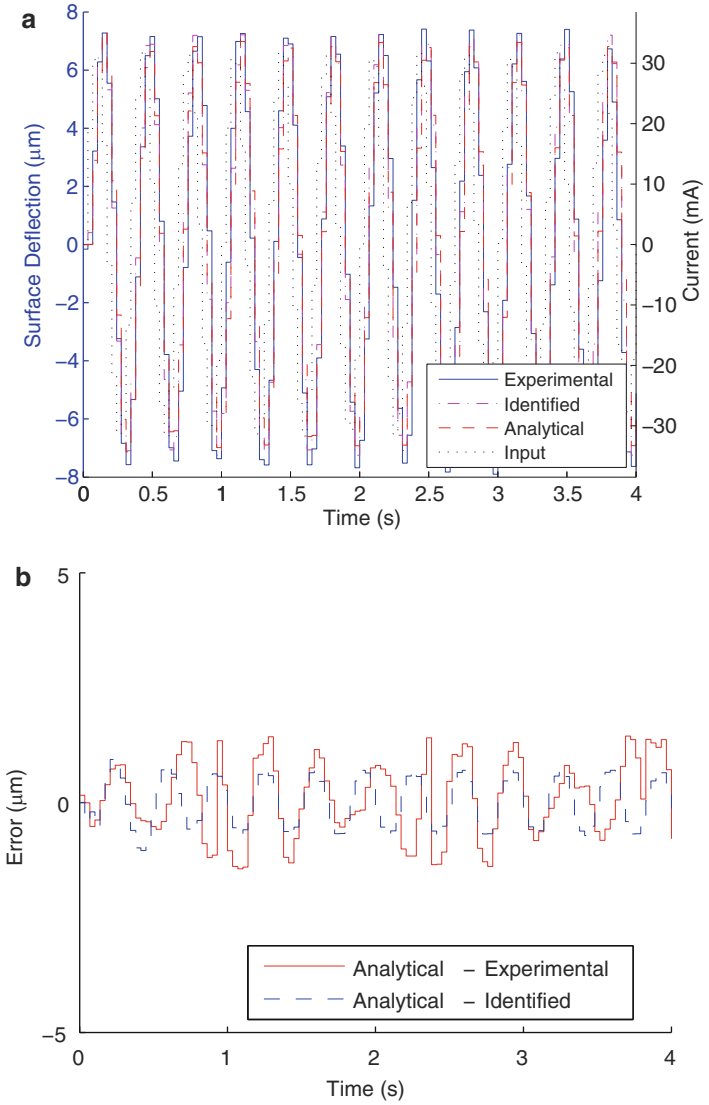


Fig. 5.22 Response to sinusoidal input: (**a**) surface deflections and (**b**) error

Response to a Multitone Input

Figure 5.23 shows the response of the MFDM surface to a multitone signal where the input current frequency varies over time. The input signal is given by

$$u_1 = I_0 \sin\left(2\pi\left(f_1 + \frac{f_2 - f_1}{T}t\right)t\right) \tag{5.10}$$

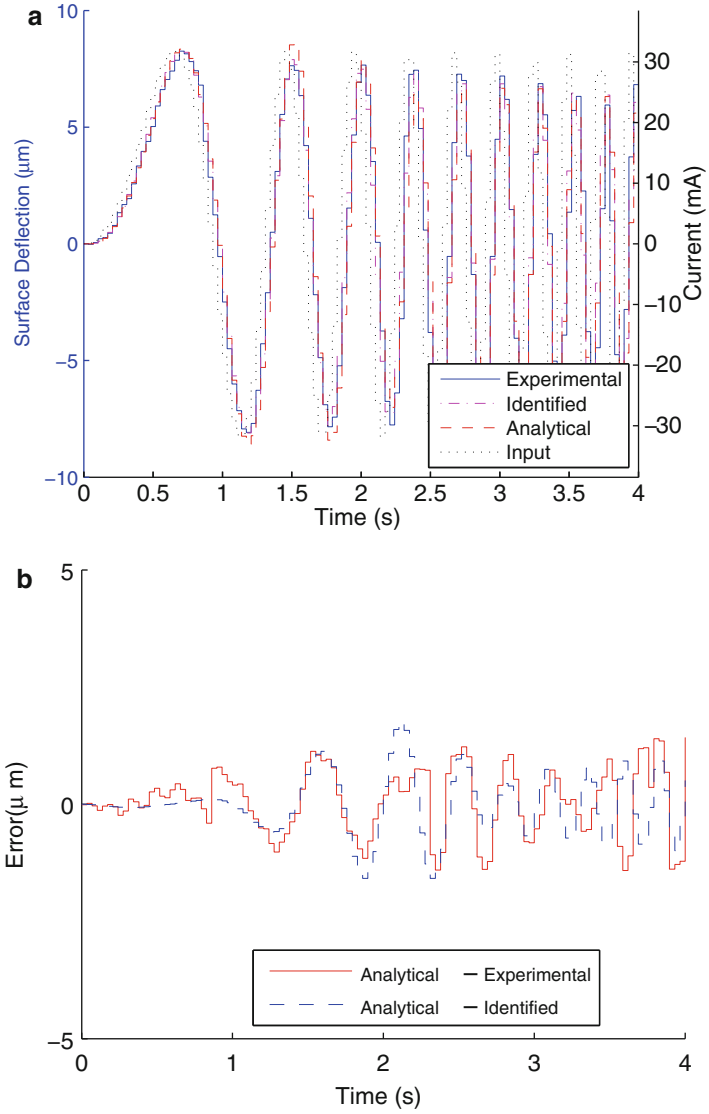


Fig. 5.23 Response to multitone input: (**a**) surface deflections and (**b**) error

where $I_0 = 32$ mA is the amplitude of the input current signal and the frequency of the signal linearly varies from $f_1 = 0$ to $f_2 = 5$ Hz in $T = 5$ s. Figure 5.23a shows a comparison of the analytically predicted surface deflection against the experimental results and surface displacements predicted by the identified model. The corresponding error is shown in Fig. 5.23b. It is evident from the comparison of the response given by the analytical model to the experimental results as well as to

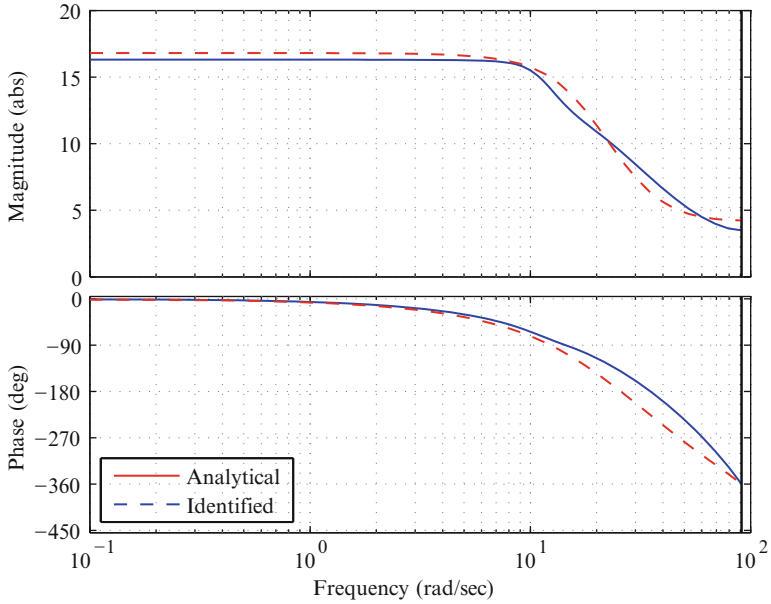


Fig. 5.24 Bode plot of the SISO system

the response given by the identified model that there exists good agreement among all three results. With the increasing frequency of the input current signal, certain trends in the experimental results can be observed. It can be seen that the output surface displacements follow the driving input closely in the low-frequency range. The agreement deteriorates as the frequency is increased. Firstly, the amplitude of the surface deflections decreases with the increasing frequency. Secondly, there exists a phase lag between the input current signal and the output displacements, which increases with the increasing frequency. It is evident from the presented plots that analytical model captures both these trends and closely follows the experimental results. On both trends, the analytical model agrees with the identified model as well.

Bode Plots

Bode plots present a comprehensive picture of the dynamics of a system. Figure 5.24 shows the Bode plots of the analytical model and the identified model of the plant. The plots presented are for the system where the input current is applied at the center coil (coil #1 in Fig. 5.7) of the MFDM, and the output wavefront displacement is measured at the location of the corresponding input coil. The analytical model agrees well with the identified model in the low-frequency range. Some mismatch between the two models appears for frequencies beyond 10 rad/s. Some important characteristics of the mirror response can also be observed from the Bode plots.

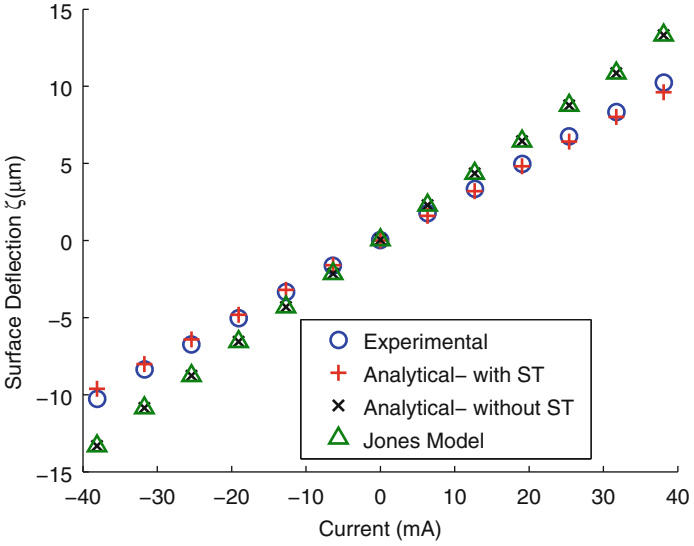


Fig. 5.25 Effect of surface tension on the peak surface displacements

The MFDM features a bandwidth of approximately 20 rad/s. Secondly, there exists a significant phase lag in the response of the mirror surface, which increases with the increasing frequency. This explains the lag between the input currents and the output surface deflections as plotted in Fig. 5.23.

The results presented above for the validation of the analytical model show that there exists a very close agreement between the response of the MFDM system as predicted by the analytical model and the corresponding results obtained experimentally. Similarly, these two results also agree fairly well with the response of the system as predicted by the identified model.

5.3.2.3 Effects of Surface Tension

The effects of surface tension in the magnetic fluid become evident when comparing the experimentally obtained peak displacements with those predicted by the analytical model, with and without including the surface tension. Such a comparison is made in Fig. 5.25 where the peak surface deflections already presented in Fig. 5.19 are reproduced with the predictions of the analytical model presented for both the cases, that is, with and without considering surface tension. It is obvious from the figure that the surface tension plays a significant role in determining the response of the mirror. It may be noted that the predictions of the analytical model and those provided by Jones' model (5.5) coincide when the effects of surface tension are not included in the analytical model.

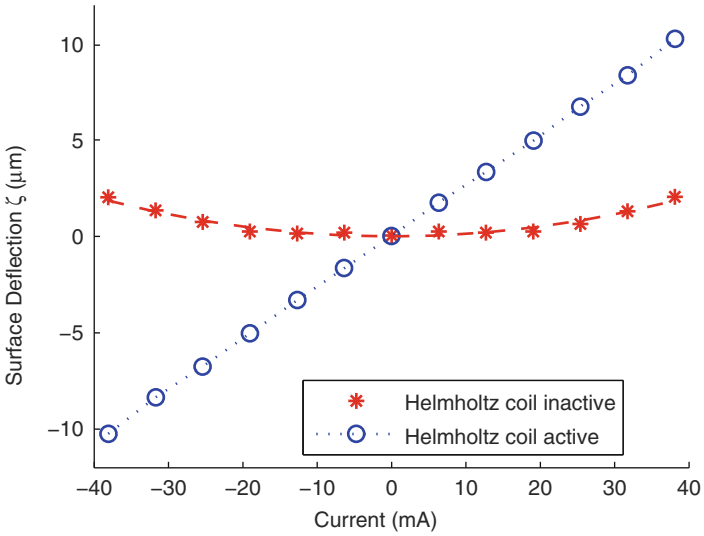


Fig. 5.26 Comparison of the response of the MFDM with conventional design versus proposed design

5.3.3 Linearity of the MFDM Response

A significant contribution of the work presented in this book is the linearization of the response of the MFDM, which is achieved by incorporating a Helmholtz coil in the design of the mirror. Experimental results which validate the proposed design modification and its effect on the response of the MFDM surface are presented here. First, the response of a MFDM, which uses the miniature coils only, is demonstrated by disabling the Helmholtz coil of the prototype mirror and recording the surface displacements as a function of the current applied to one of the miniature coils. Then, the linearizing effect of the Helmholtz coil is captured by observing the surface displacements while the Helmholtz coil is turned on. Finally, the proposition is substantiated by testing the principle of superposition of the response of the fluid surface.

Figure 5.26 shows the peak surface deflections as a function of currents applied to the miniature coil in the center of the mirror. The points marked as '*' signify the peak surface deflections of the MFDM when the Helmholtz coil is disabled. It can be seen that the surface deflections are positive for both positive and negative polarities of the input current applied to the miniature coil. Also, the surface deflections are nonlinear with respect to the applied current. A trend line fitted on the data reveals a quadratic approximation. This behavior can be explained using Jones' model of the static response of the fluid surface as given in (5.4). The model shows that the displacement $\bar{\zeta}$ of a magnetic fluid surface exposed to a magnetic field is a

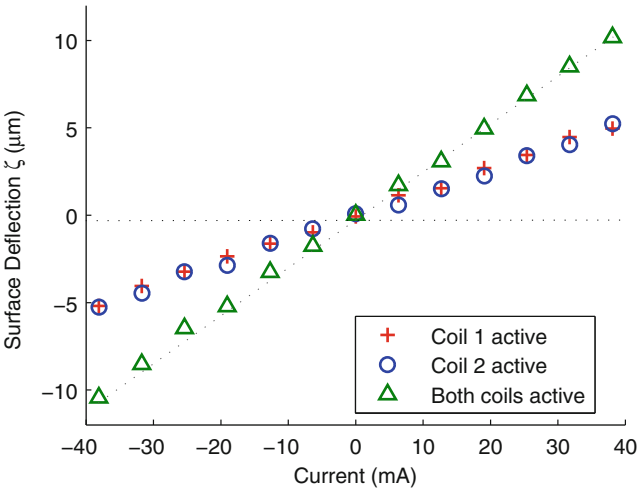


Fig. 5.27 Illustration of the principle of superposition of the MFDM surface displacements

quadratic function of the magnetic flux b_z of the field. The magnetic flux, in turn, is a linear function of the applied current implying that the surface deflection $\bar{\zeta}$ is a quadratic function of the applied current. The experimental peak surface deflections for the case when the Helmholtz coil is turn on are marked as '°'. It is obvious that (1) the surface deflections vary linearly with the increasing currents applied to the coil and (2) both negative and positive deflections are achieved.

The experimental results also show that when the Helmholtz coil is turned on, the principle of superposition holds distinctly. As illustrated in Fig. 5.27, when the two coils marked 1 and 2 (see Fig. 5.7) are energized, the surface deflection ('Δ') at a point midway between the two coils is the sum of the deflections ('+' and '°') at this point generated by each of the two coils energized separately. It is therefore evident that the proposed change in the design of the MFDM does produce the desired effect of linearizing the response of the mirror surface.

5.3.3.1 Bidirectional Displacements

The linearity of the response of the mirror also allows bidirectional displacements. Under the conventional design, the magnetic fluid can "push" only but not "pull." Due to this limitation of conventional MFDMs, the pulling effect is generated by biasing the electromagnetic coils to 50% of their maximum "push," which results in a reduction of the maximum achievable deflection by half. The design proposed in this book resolves this problem by generating both positive and negative displacements.

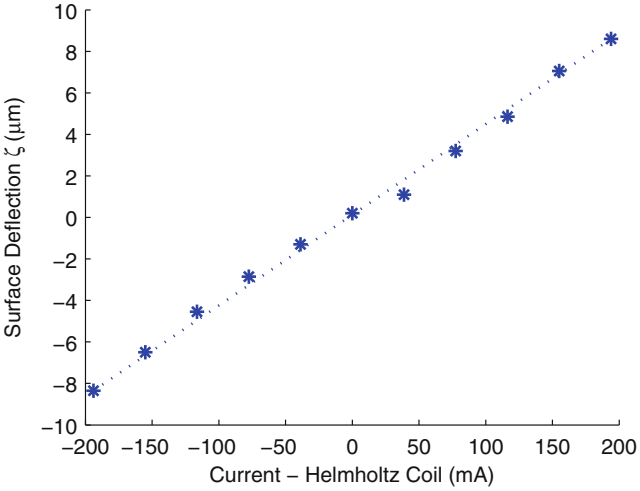


Fig. 5.28 Illustration of the amplifying effect of the large, uniform magnetic field generated using the Helmholtz coil. The *plot* shows the peak surface deflections as a function of the current applied to the Helmholtz coil, while the current applied to the miniature coil is kept constant at 32 mA

5.3.3.2 Amplification of the Surface Deflections

Another significant aspect of the proposed design modification is the amplification of the surface deflections achieved by applying the large uniform magnetic field. The amplification effect of this field provided by the Helmholtz coil can be seen in Fig. 5.28 where the current applied to the Helmholtz coil is gradually increased, while the current applied to the miniature coil (#1 in Fig. 5.7) is kept constant at 32 mA. Note that the polarity of the current applied to the Helmholtz coil also has the effect of reversing the "push" or "pull" to the mirror surface.

It may be noted that the magnetic field of the miniature coils generates the desired shape of the mirror surface. The large uniform magnetic field, on the other hand, amplifies the magnitude of the surface displacements. A very significant connotation of this effect is that very small electromagnetic coils can be used in conjunction with an appropriately large uniform magnetic field to produce far larger surface deflections than those possible with the small coils only. This will allow fabrication of very small mirrors—most likely developed using one of the MEMS microfabrication methods—with sufficient actuator density and surface deflections for possibly all applications of MFDMs.

5.4 Summary

In this chapter, the design of the prototype MFDM and the experimental AO setup has been described. The novel conceptual design change prompted by the findings of the analytical work presented in Chap. 4 is implemented in the design

of the prototype MFMD by using a Helmholtz coil. A description of the design requirements and the design variables of a MFDM are presented. The detailed design of the prototype MFDM is described. An adaptive optics setup is designed and assembled. The details of the layout of the experimental setup and the description of the components used in the setup are presented. The prototype mirror and the experimental setup are designed to be used to validate the analytical model of the mirror and to evaluate the performance of the controllers presented in the following chapters. In order to validate the analytical model developed in Chap. 4, the experimental static response of the prototype MFDM is first compared to the response predicted by the analytical model, and the dynamic characteristics of the mirror as predicted by the analytical model are validated using sinusoidal input signals. Then, the linearization of the response of the MFDM achieved by incorporating a Helmholtz coil in the design of the mirror is further verified. The experimental results show that when the Helmholtz coil is turned on, the principle of the superposition holds distinctly, which render the possibility of implementation of a closed-loop AO control system using linear system control method. The experiments also show that the bidirectional displacements of the mirror surface are achieved, and the magnitude of the surface deflections is significantly amplified by applying the large uniform magnetic field with the Helmholtz coil.

References

Brousseau D, Borra EF, Rochette M, Landry DB (2010) Linearization of the response of a 91-actuator magnetic liquid deformable mirror. Opt Express 18(8):8239–8250

Caprari RS (1995) Optimal current loop systems for producing uniform magnetic fields. Meas Sci Technol 6(5):593–597

Fernandez EJ, Artal P (2007) Dynamic eye model for adaptive optics testing. Appl Opt 46:6971–6977

Gollwitzer C, Matthies G, Richter R, Rehberg I, Tobiska L (2007) The surface topography of a magnetic fluid: a quantitative comparison between experiment and numerical simulation. J Fluid Mech 571:455–474

Haines GV, Newitt LR (1997) The Canadian geomagnetic reference field 1995. J Geomagn Geoelectr 49(2–3):317–336

Hecht E (2002) Optics. Addison Wesley, San Francisco

Jones TB (1988) Theory and application of ferrofluid seals. In: Berkovsky B (ed) Introduction to thermomechanics of magnetic fluids. Hemisphere, Washington, DC

Laird P, Caron N, Rioux M, Borra EF, Ritcey AM (2006) Ferrofluid adaptive mirrors. Appl Opt 45(15):3495–3500

Ljung L (1999) System identification: theory for the user. Prentice-Hall, Upper Saddle River

Ljung L (2009) System identification toolbox: users guide. The MathWorks, Natick

Tyson RK (2011) Principles of adaptive optics. CRC Press, Boca Raton

Chapter 6
Control System Design

Contents

6.1 Review of the Control Methods

In general, the imaging quality of an optical system is often expressed in terms of the Strehl ratio (Tyson 2011). The Strehl ratio is defined as the peak intensity of the image of a point source normalized to the diffraction limited peak intensity. The Strehl ratio is a useful and sensitive performance measure since the wavefront error is expected to diffract light away from the center of the image, thereby reducing the peak intensity. In the presence of any wavefront aberration, the Strehl ratio will be less than 1, and the general objective of AO control problem can be formulated as maximizing the Strehl ratio. However, from a practical point of view, the direct optimization of the Strehl ratio is not very convenient. Evaluating the Strehl ratio requires accurate knowledge of the wavefront over the entire aperture, which is generally not available. According to the Marechal approximation (Tyson 2000), the Strehl ratio is a decreasing function of the variance of the phase error over the aperture; therefore, the objective of maximizing the AO imaging quality can be replaced by that of minimizing the residual phase variance over the aperture using deformable mirrors.

The mirror control problem faced in AO systems can be generalized as a shape control problem. The objective of the AO system is to obtain a shape, for the fluid

Z. Wu et al., *Modeling and Control of Magnetic Fluid Deformable Mirrors for Adaptive Optics Systems*, DOI 10.1007/978-3-642-32229-7_6,

surface under consideration, which is as close as possible to the desired shape needed to cancel the wavefront aberrations present in the incoming wavefront. At the core of any AO system is a feedback loop, where an incoming uncorrected wavefront is reflected by the deformable mirror (DM), the reflected wavefront is measured by the WFS, and the DM is adjusted by a control algorithm, based on the WFS measurements. The objective of the control algorithms presented in this book is to select control inputs to reduce the error between the actual shape and the desired shape of the MFDM, thus obtaining the shape needed to cancel wavefront aberrations. Within the literature, a number of control design techniques have been proposed for the control of AO systems. In order to increase the image quality that can be achieved practically by an AO system, the designed closed-loop control system must be able to track and compensate for unknown and time-varying changes in the aberrations. Different controllers have been proposed in the literature to provide real-time correction of aberrations in AO systems, including proportional–integral–derivative (PID) controllers, optimal controllers, and adaptive controllers.

The nonlinear control problems of the conventional MFDM are simplified to linear control problems by superimposing a strong uniform magnetic field on top of the field created by the distributed coil actuators as shown in Chap. 5. This has the effect of linearizing the response of the liquid, which results in a significant simplification of the control system design in that instead of having to develop exotic nonlinear control algorithms, it is now possible to develop linear control algorithms to regulate the surface shape of the MFDM. In the following, different AO system control strategies proposed in the literature are reviewed.

6.1.1 Classical Decentralized AO Control Methods

Most AO control systems are based on a simple control law that involves a series of parallel single-input single-output compensators. Such an approach is usually referred to as a classical control approach (Hinnen et al. 2008, 2007). The classical control strategy is still the most widespread control approach used in AO systems. This approach implicitly assumes that the plant dynamics is fully decoupled. Moreover, the dynamics of both the DM and WFS and the temporal evolution of the wavefront are not explicitly accounted for in the controller design. A block diagram representation of a classical decentralized AO control system is shown in Fig. 6.1, where $\zeta_{\text{aberrated}}$ is the shape of the aberrated wavefront incident on the mirror surface, $\mathbf{u}$ is the vector of inputs to the DM actuators, ζ_s is the shape of the residual wavefront reflected from the mirror surface, and $\mathbf{s}$ is a vector of local slope measurements for the residual wavefront. The interaction that takes place at the level of the DM between the shape of the DM surface ζ and the aberrated wavefront shape $\zeta_{\text{aberrated}}$ and that yields the residual wavefront shape ζ_s is discussed later in the chapter.

The AO control system is designed to optimize the AO system imaging performance by minimizing the residual wavefront phase, that is, the phase of the wavefront reflected by the deformable mirror. Equivalently, the shape of the residual wavefront is driven as close as possible to a flat surface shape. The standard AO

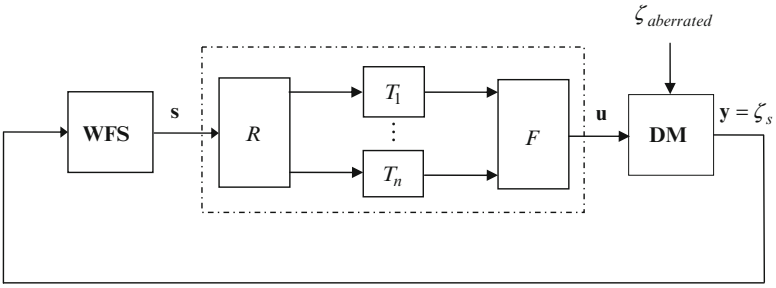


Fig. 6.1 Block diagram representation of a classical decentralized AO control system consisting of a wavefront sensor (*WFS*), wavefront reconstruction mapping *R*, a decentralized controller, a static mapping *F*, and a deformable mirror (*DM*)

system control loop mainly consists of a wavefront sensor (WFS), a wavefront corrector (WFC) that includes a deformable mirror (DM) and its corresponding actuators, and a controller. The WFS typically measures the slope or curvature of the wavefront using a curvature WFS, a pyramid WFS, a shearing interferometer, or a Shack–Hartmann sensor. The most widely used WFS is the Shack–Hartmann sensor which measures the subaperture gradients of the incoming wavefront shape. In this book, we will restrict our attention to the Shack–Hartmann sensor; however, the other types of WFSs can be handled in a similar way. The slope information provided by the WFS is used in the reconstruction of the residual wavefront shape and to express the latter as a linear combination of modal basis functions (Goodman 2004). The reconstruction can take place in any appropriate set of basis functions.

The classical AO control approach decomposes the controller into two static mappings and a series of parallel single-input single-output (SISO) feedback loops as shown in Fig. 6.1. The wavefront reconstruction problem is solved first, where the WFS measurements are multiplied by the matrix *R* to yield either an estimate of the residual wavefront shape or a set of coefficients related to the representation of the estimated residual wavefront shape as a linear combination of basis functions. The decentralized compensator is designed to achieve stability and closed-loop performance and to attenuate the residual wavefront. The outputs from the decentralized compensators are then processed by the second static mapping, denoted by F, to produce the control signals for the DM actuators so that the DM mirror achieves the desired surface shape. The DM dynamics maps the actuator commands $\mathbf{u}$ into a deformed mirror surface ζ as follows:

$$\zeta = M(\mathbf{u}) \tag{6.1}$$

where $M(\cdot)$ is a linear operator. The steady-state relationship between the mirror surface and the actuator commands is assumed to be a static linear mapping as

$$\zeta_{ss} = \mathbf{H}\mathbf{u}_{ss} \tag{6.2}$$

where ζ_{ss} and $\mathbf{u}_{ss}$ are the steady-state values of ζ and $\mathbf{u}$, respectively. $\mathbf{H}$ is the DC gain of the linear operator $M(\cdot)$, which is often defined as the influence function matrix of the DM.

The controller design approaches can be divided into two categories, namely, the modal compensation approach and the zonal compensation approach discussed below in more detail.

6.1.1.1 Modal Compensation

Let $\mathbf{c}_s$ be the vector of Zernike coefficients associated with the expression of the residual wavefront shape as a linear combination of Zernike polynomials. In the modal compensation-based AO control systems, the static mapping R represents the wavefront modal reconstructor between the measured local slopes $\mathbf{s}$ of the residual wavefront and the Zernike modal coefficients $\mathbf{c}_s$ as

$$\mathbf{c}_s = R\mathbf{s} \triangleq \tilde{\mathbf{Z}}^{\dagger}\mathbf{s} \tag{6.3}$$

where $\tilde{\mathbf{Z}}^{\dagger}$ is the pseudo-inverse of modal reconstruction matrix $\tilde{\mathbf{Z}}$ (see Appendix C). The vector of Zernike coefficients $\mathbf{c}_s$ is then processed by the decentralized controller to yield the vector of modified coefficients $\mathbf{c}_m$ as follows:

$$\mathbf{c}_m = T\mathbf{c}_s$$

where

$$T = \begin{bmatrix} T_1 & 0 & 0 & \cdots & 0 \\ 0 & T_2 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & T_n \end{bmatrix}$$

The vector of modified coefficients $\mathbf{c}_m$ represents Zernike coefficients associated with the desired DM shape when the latter is expressed as a linear combination of Zernike polynomials. It should be noted that for $i = 1, \ldots, n$, the input to the compensators T_i is the ith Zernike coefficient $\mathbf{c}_{s,i}$. Consequently, for each $i = 1, \ldots, n$, the compensator T_i aims at driving to zero the contribution of the ith Zernike mode to the total residual wavefront, that is, driving the Zernike coefficient $\mathbf{c}_{s,i}$ to zero.

The static mapping F represents the modal injection matrix between the modified coefficients $\mathbf{c}_m$ (the output of the compensators) and the DM actuator commands $\mathbf{u}$ as

$$\mathbf{u} = F\mathbf{c}_m \triangleq \mathbf{H}^{\dagger}\mathbf{Z}\mathbf{c}_m \tag{6.4}$$

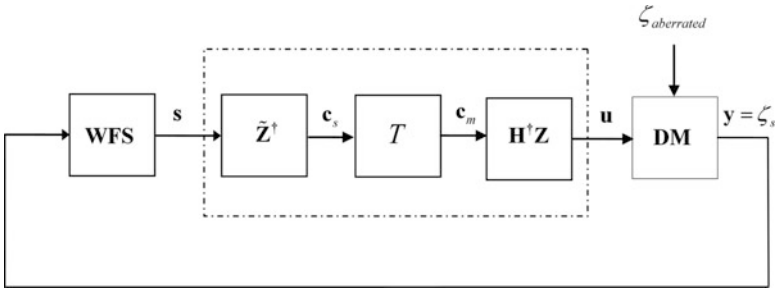


Fig. 6.2 Block diagram representation of an AO control system based on the modal compensation approach

where $\mathbf{Z}$ is the matrix of Zernike mode shapes computed at specified points (see Appendix C) and $\mathbf{H}^{\dagger}$ is the pseudo-inverse of the influence function matrix $\mathbf{H}$. A block diagram of an AO control system based on the modal compensation approach is presented in Fig. 6.2.

6.1.1.2 Zonal Compensation

For zonal compensation-based AO control systems, the static mapping R represents the wavefront reconstructor between the local slopes $\mathbf{s}$ of the residual wavefront and an estimate $\mathbf{z}_s$ of the residual wavefront shape $\mathbf{y}$ (or ζ_s) specified at a finite number of spatial locations as

$$\mathbf{z}_s = R\mathbf{s} \triangleq \mathbf{Z}\tilde{\mathbf{Z}}^{\dagger}\mathbf{s} \tag{6.5}$$

The vector $\mathbf{z}_s$ describing the estimate of the residual wavefront shape is then processed by the decentralized controller to yield the vector $\mathbf{z}_m$ representing the desired shape for the deformable mirror surface as follows:

$$\mathbf{z}_m = T\mathbf{z}_s$$

where the decentralized compensator T is as given above. In the zonal compensation approach, each compensator T_i processes the residual wavefront shape value at the ith spatial location on the wavefront. Consequently, for each $i = 1, \ldots, n$, the compensator T_i aims at driving the residual wavefront shape value at the ith spatial location to zero.

The static mapping F represents the zonal injection matrix between the modified zonal wavefront displacement $\mathbf{z}_m$ (the output of the compensator) and the DM actuator commands $\mathbf{u}$ as

$$\mathbf{u} = F\mathbf{z}_m \triangleq \mathbf{H}^{\dagger}\mathbf{z}_m \tag{6.6}$$

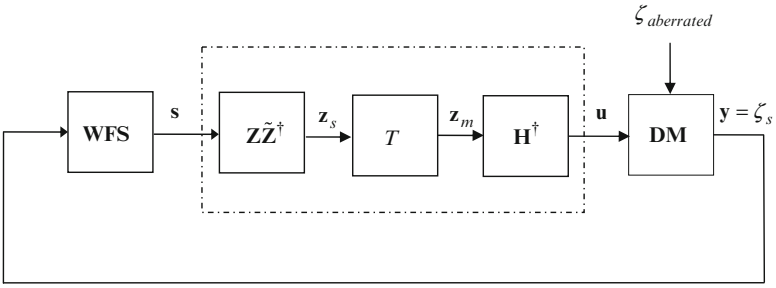


Fig. 6.3 Block diagram representation of an AO control system based on the zonal compensation approach

A block diagram of an AO control system based on the zonal compensation approach is presented in Fig. 6.3.

The task of the compensators is to update the required modal coefficient correction or zonal wavefront correction. It can be seen that in the wavefront reconstruction step and in the projection on the actuator space step (i.e., processing of the compensator outputs by F), typically all dynamics are neglected. The compensator design requires the consideration of the models of the significant dynamic components in the control loop. In any AO system, there are three sources of dynamic behavior in the control loop that should be considered. First, the time delay due to the computation time required for reading the WFS CCD, constructing the gradients, and computing the modal coefficients can be significant. Second, the inherent time discretization induced by the frame rate of the wavefront sensor and the dynamics of the digital to analog converter and sampling process are also important. Finally, the DM actuators always exhibit some dynamics. However, in most of the literature related to the classical control approach, the DM dynamics are normally represented using decoupled SISO linear systems and are usually neglected with the assumption that the dynamic response of the DM is much faster than the wavefront reconstruction. The compensator design should account for the dynamics of the system to achieve stability and the desired closed-loop system performance. As shown in Fig. 6.1, the compensators used in classical AO control systems consist of a series of parallel SISO feedback controllers T_i, $i = 1, \ldots, n$, which can be designed identically or separately. For zonal compensation approach, T_i, $i = 1, \ldots, n$, are usually identical, and the decomposition is based on the assumption that the actuator dynamics are identical for each actuator and the spatial and temporal dynamics can be decoupled. For modal compensation approach, T_i, $i = 1, \ldots, n$, can be identical or can be designed separately based on different modal properties to obtain the best closed-loop system performance.

The design of a compensator T_i typically uses the SISO classical control design approach. Commonly applied control structures include PID controllers, first-order lag filters, and Smith predictors. Given a control structure, the controller design problem finally boils down to determining the appropriate controller parameters that

yield a closed-loop system with the desired bandwidth and capable of effectively canceling the residual wavefront errors.

Most classical controllers for adaptive optics systems are either of the integral feedback type (Downie and Goodman 1989; Auburn et al. 1987) or proportional–integral–derivative (PID) feedback type (Stein and Gorinevsky 2005; Huang et al. 1995b). These controllers are driven by residual wavefront shape or phase errors. Each of the SISO controllers is to drive the wavefront error at a given spatial location to zero. Traditionally, a SISO integral or PID controller driving a single actuator has been developed based on a pure time delay model for the system. The controller is then duplicated for each channel in the control system. Because of the simplicity of such approach, it has been widely used to control segmented mirrors (which experience no inter-actuator coupling). This simple local feedback design approach works effectively since two assumptions are met. The first assumption is that there is very little or no coupling between the different actuator inputs and corresponding sensor measurements in segmented mirrors. The second assumption is that the bandwidth of the closed-loop system is significantly less than the dynamics of the mirror actuators.

The response of a continuous DM is typically described using the mirror static or DC gain. The continuity of the membrane always causes the deflection over an energized actuator to influence the deflection over neighboring actuators. Influence functions are static models that could be experimentally obtained by measuring the response of the DM to constant control inputs. The static models are usually developed by considering one actuator at a time and measuring the steady-state response of the mirror to a constant input from the actuator of interest. The response of the whole mirror is measured. By repeating the same experiment with all the actuators, a static map between the constant inputs and the resulting steady-state mirror response can be developed (Huang et al. 1995a). The resulting static input–output map is actually the DC gain of the dynamic system and is usually referred to in the adaptive optics literature as the influence function for the deformable mirror. The influence function can be further simplified by considering responses from sensors that are in the immediate vicinity of the actuator and ignoring the rest of the measured responses from the remaining sensors. By setting the mapping F to be the inverse of the influence function, the dynamics of combined mapping F and the DM is decoupled at DC (i.e., has a DC gain equal to the identity matrix). Classical SISO control design methods can be easily applied for each control channel.

In Roddier (1999) and Fried (1990), it was shown that computational time delay present in an AO system dramatically affects the systems performance. This is due to induced phase lag whose effect is to reduce the bandwidth of the AO system (i.e., the operating range where the AO system performance is acceptable). This occurs primarily when the system is modeled using a static model, since such models do not include any system dynamics. The time delay is represented in Roddier (1999) using the transfer function

$$\mathrm{WFS}(s) = \mathrm{e}^{-\tau s} \tag{6.7}$$

Equation (6.7) is the transfer function for the wavefront sensing (and reconstruction) system. The temporal characteristics of this device is a pure time delay, τ, due to the readout and computation of the sensing system.

In Roddier (1999), two examples of classical controllers that increase the AO system performance are given and compared, where the objective of the AO system is to reduce the error between the ideal shape of the DM and the actual shape. The first is the use of a PI controller to reduce the phase lag. The PI controller showed a 40% gain in AO system correction efficiency, as opposed to a pure integrator. The second controller examined was a Smith predictor, which is a controller especially dedicated to servos exhibiting time delay. Compared to the pure integrator, the Smith predictor showed a 44% gain in AO system correction efficiency.

The use of an estimation controller is a way to solve the time delay problem. The idea is to use theoretical and experimental knowledge of the temporal evolution of the wavefront phase fluctuations to compensate for the time delay (Voulgaris et al. 2003; Montera et al. 1997; Dessenne et al. 1997). In Voulgaris et al. (2003), a least mean square estimation algorithm is used to propagate the WFS outputs forward to the current time, for use by the controller. Analysis of the output data for the WFS revealed that the AO systems time delay corresponds to the sampling rate of the sensor. In order for the controller to provide accurate control commands to the DM, it requires up to date measurements of the wavefront (provided by the WFS).

The classical control approach has shown a reasonable performance under the moderate assumptions that the wavefront turbulence has a power spectrum mainly in the low frequency range. The controller with a built-in integrator function gives a high loop gain in the low frequency range, which implies a good disturbance rejection can be achieved for this slowly varying process. Looze (2006) has shown that if the temporal dynamics of the wavefront distortions can be modeled as independent first-order autoregressive processes, the classical AO control structure is optimal under some stringent conditions, such as that the only dynamics in the AO system is a one sample delay, the influence function matrix is invertible, and the uncorrected wavefront is an isotropic first-order temporal turbulence.

6.1.2 Centralized Optimal Control Methods

Even though the classical AO control strategy yields acceptable performance and offers a simple structure and a small computation burden, there is still a significant need for improvement in the performance of current systems through the use of modern control techniques. As pointed out by Roddier (1999), the compensation efficiency of large AO systems based on classical AO control approach is unduly low; therefore, it may be also necessary to search for optimal control approaches to improve the performance of AO systems. It is clear that the classical AO control approach does not explicitly account for the spatial–temporal evolution of the wavefront turbulence and the dynamics of the AO systems. A promising way to reduce the effect of the temporal error is to exploit the spatiotemporal correlation

in the turbulence to anticipate future wavefront distortions. On the other hand, in order to further improve the image quality in real time, the adaptive optics system needs to be designed to run sufficiently fast to attenuate the dynamic wavefront turbulence, and sampling times below 0.5 ms or even lower may be necessary. With these high sampling frequencies, the deformable mirror in adaptive optics systems cannot be considered as a static system, since the high-frequency DM dynamics can be easily excited in the high frequency range and a full-coupled dynamic model of the DM hence needs to be considered in the controller design. Therefore, it is necessary to consider the design of an AO system control loop from a modern control system perspective. In particular, this means first that the dynamics of the different elements involved in the control loop have to be taken into account. In particular, a specific dynamic model for the DM should be considered for the purpose of controller design. Second, a model of the turbulent wavefront, built from its spatial and temporal frequency domain characteristics, should also be accounted for in the controller design. Furthermore, the finite time delay between measurement and correction is known to be one of the main limitations on the performance of an AO system and needs to be minimized. Also, the measurement noise is known to constitute an error source of significant importance and should be tackled properly in the overall closed-loop AO system.

The residual wavefront represents the input to the controller in an AO system. As stated above, the objective of the control algorithm is to select control inputs to eliminate this residual wavefront error by minimizing a specified performance criterion. Many in the AO community have turned to a linear quadratic Gaussian (LQG) control formulation which seeks to minimize a quadratic cost function and, thus, determine the controller that minimizes the error between the actual shape and the desired shape of the deformable mirror. Under the LQG assumptions, the system to be controlled is assumed to be modeled by a set of static equations or linear differential equations driven by white Gaussian noise inputs. A relevant AO optimal performance criterion is to minimize the variance of residual wavefront shape ζ_s, defined as

$$J(u) \triangleq \lim_{T\to\infty} \frac{1}{T}\int_0^T \|\zeta_s(t)\|^2 \mathrm{d}t \tag{6.8}$$

For this class of minimum-variance problems, LQG techniques appear particularly attractive given the quadratic nature of the cost function J on the residual wavefront shape ζ_s.

An LQG design technique is first used by Paschall and Anderson (1993) to develop a controller for an adaptive optics system used in ground-based telescopes. Work in much of the literature pertaining to atmospheric aberrations suggests that, although both modulation of intensity and phase of the optical image contribute to the formation of aberrations, it is the modulation of phase that is the dominant distortive component. Therefore, the controller presented in Paschall and Anderson (1993) is directed toward phase distortion compensation only. Previous research had considered the wavefront sensor measurements to be available to the observer

instantaneously, namely, no consideration was given to the effect of the sample period time delay inherent in the sensor. The control law presented in Paschall and Anderson (1993) is an extension of the work of Von Bokern (1990), in which the measurements are delayed by one sample period. The LQG controller generates optimal inputs to the deformable mirror given (1) a time series of wavefront sensor measurements, (2) the statistical characteristics of those measurements (viz., second-order time and spatial correlations), and (3) the statistical characteristics of the measurement noise. The quadratic cost function J was given in discrete time as

$$J = E\left\{\sum_{k=0}^{N}\frac{1}{2}\left[\mathbf{x}^{\mathrm{T}}(k)\Gamma\mathbf{x}(k) + \mathbf{u}^{\mathrm{T}}(k)\Psi\mathbf{u}(k)\right]\right\} \tag{6.9}$$

where $E\{\cdot\}$ denotes the expectation operator and $\mathbf{x}(k)$ is the state vector of the state-space model that incorporates the dynamic behavior of the wavefront corrector (i.e., DM and actuators). The signal $\mathbf{u}(k)$ represents the control inputs, in the form of actuator voltages, applied to the DM, and Γ and Ψ are cost-weighting matrices selected such that quadratic penalties are assigned for state deviations from zero and for requiring excessive energy to control those states.

Design of the LQG regulator requires determination of the optimal vector of control inputs, $\mathbf{u}(k)$, to apply to the DM which minimizes the cost function (6.9). If it were possible to gain perfect knowledge of all the system states, the optimal feedback control law would be based on these states. Since this is not possible, a state observer is used to generate estimates of the states. Under the assumption of no measurement processing delay, the adaptive optics system was able to reduce the phase distortion in the reflected wavefront by more than 50%.

In Wiberg et al. (2004b), a static LQG controller design approach was presented. The feedback control law used to minimize the LQG criterion reduces to finding the finite-dimensional control vector as a function of the finite-dimensional sensor vector. Each control signal is associated with a specific actuator having a particular spatial influence function on the process; this influence function is called the controller mode shape. Next, it is shown that each sensor is similarly associated with a sensor mode shape. It is shown further that the best control is achieved when the controller mode shapes equal the sensor mode shapes. In Wiberg et al. (2004a), the theory of a spatial non-dynamic LQG controller developed in Wiberg et al. (2004b) is applied to astronomical adaptive optics telescopes. Simulation results suggest difficulties facing the practical design of a system with a large actuator/sensor array for adaptive optics applications.

Looze (2006) formulated the adaptive optics minimum-variance control problem as a linear quadratic Gaussian optimization problem using a discrete-time model that includes both the temporal dynamics of the DM and atmospheric aberrations, the computational loop delay, and the frame integration of the WFS. The objective is to select control inputs (in this case actuator voltages) to minimize the error between the actual and the desired shape of the mirror. The assumption of zero-mean Gaussian white noise is made to represent the aberrations. Given this assumption,

the controller objective is then met by minimizing the variance of the error signals. This is accomplished by minimizing the following functional:

$$J = \lim_{N\to\infty} \frac{1}{2N} \sum_{k=0}^{N} x_k^{\mathrm{T}} Q x_k \tag{6.10}$$

where Q is a symmetric, positive-semidefinite matrix and x_k is the state variable of the discrete-time state-space model representing for the DM, atmospheric disturbance, and WFS. The controller is first computed explicitly for an adaptive optics system in which the models for the DM, wavefront sensor, and the uncorrected wavefront are ideal: this includes an isotropic first-order temporal atmospheric aberrations, no computational loop delay, and no DM dynamics. Under these assumptions, when the time constant of the atmospheric model becomes increasingly large, the performance of the developed LQG controller approaches the performance of the integral controller that is commonly used in adaptive optics systems. If the time constant of the atmospheric model is finite, the dynamics of the LQG controller consist of identical first-order lags. The performance of the optimal LQG controller in the presence of deviations from the ideal conditions is further examined in Looze (2010), where a linear quadratic Gaussian (LQG) design approach is presented for the case that the WFS and zero-order hold (ZOH) need not be synchronized. The LQG controller is formulated based on an equivalent hybrid discrete-time model of the adaptive optics system presented in Looze (2009). The design model incorporates deformable mirror dynamics, an asynchronous wavefront sensor and zero-order hold operation, and a continuous time model of the incident wavefront. The hybrid LQG design does not assume that the beginning of the frame of the WFS is synchronized with the application of the DM commands via the zero-order hold. Using the structure of the discrete-time model, the dimensions of the Riccati equations to be solved are reduced. The LQG controller is shown to improve the AO system performance as compared to the LQG controller designed without accounting for intra-sample behavior of both the DM and the incident wavefront.

Le Roux et al. (2004) developed and simulated LQG designs for both classical and multiconjugate AO (MCAO) systems. The approach gives an optimal estimation of the turbulence using a Kalman filter in the closed loop. The optimal control law incorporates both spatial and temporal information on the turbulent phase, as well as information on the system noise statistics. Petit et al. (2009) further presented a comprehensive analysis of the linear quadratic Gaussian control approach applied to adaptive optics (AO) and multiconjugate AO (MCAO) based on both numerical and experimental validations. Various properties that make this control law a good option for different applications are discussed, such as open-loop control, aliasing reduction, or vibration filtering. The structure of an LQG controller and its relation to integral controllers were investigated by Kulcsár et al. (2006) under the assumption that the time response of the deformable mirror (DM) is fast compared with the sampling rate of the AO loop. The LQG control problem is broken down into an optimal deterministic control problem and an optimal estimation problem,

where the solution is linear quadratic (LQ) control with Kalman filter. Correia et al. (2010) presented a minimum-variance control approach based on the LQG formalism for adaptive optics systems with mirror dynamics. Such regulator is derived in a particularly instructive scalar case with a resonant second-order DM in an extremely large telescopes scenario, where large-amplitude wind-induced disturbances are considered. The simulation results show that neglecting DM dynamics gives rise to worse performance in terms of both residual phase and stability. The DM tends to oscillate with large amplitudes due to noise amplification at the resonant frequency, while the optimal regulator with DM dynamics grants stability for a broad range of natural resonant frequencies and damping coefficients. The result is further extended to the case of asynchronous DM/WFS sampling in adaptive optics systems in Raynaud et al. (2011).

Within the controls community, many researchers have turned to H_2- and H_∞-based optimal controllers for the control of spatially distributed systems. In Hinnen et al. (2008, 2007), a data-driven H_2 optimal control approach is proposed to exploit the spatiotemporal correlation in the wavefront without assuming any form of decoupling, and the minimum-phase-variance problem is solved in the H_2 optimal control design framework. This approach consists of a dedicated subspace identification algorithm that estimates the atmospheric disturbance model directly from open-loop WFS. The identified disturbance model should be able to capture the spatiotemporal correlation imposed by frozen turbulence satisfying the Taylor hypothesis. Given the identified atmospheric disturbance model and a model of the AO system dynamics, the AO control problem is then formulated in an H_2 optimal control framework. It has been shown that the general two Riccati equations can be simplified to one Riccati equation. By analyzing the dynamic behavior of the WFS camera, for any AO system with a DM that has a time constant that is short as compared to the WFS exposure time, an analytical expression for the H_2 optimal controller can be derived in the case that the AO system can be characterized by a scalar dynamic transfer function consisting of an integer number of samples delay and a two-tap impulse response. The data-driven optimal control approach obtained by combining the proposed subspace identification algorithm and the analytical expression for the H_2 optimal controller is also experimentally demonstrated on an AO laboratory setup. The validation study has shown that compared to the common control law consisting of a minimum-variance wavefront reconstructor and a first-order lag filter, the proposed optimal control helps improve the performance of current AO systems particularly in heavy turbulence conditions, including high wind speeds and small Fried parameters, as well as under low-level light conditions.

In Huang et al. (1995b), a sequence of AO control system designs of increasing complexity are presented for a continuous DM with piezoelectric actuators used in ground-based telescopes. The authors use models developed in Huang et al. (1995a) that use an influence function to capture the coupling between the different actuator inputs and sensor measurements. The authors develop SISO PI and PID control laws which are duplicated for each actuator. MIMO H_∞-based controller was also developed and compared to the SISO-based PI and PID controllers. The paper effectively compares the use of a SISO-based controller design to a MIMO design.

Both controllers met performance requirements with respect to error regulation when the closed-loop system has a high bandwidth. However, the MIMO H_∞ controller proved to be superior in performance in that the MIMO controller is able to perform well at low frequencies as well as high frequencies. In Baudouin et al. (2008), a robust control technique is proposed for an adaptive optics system using a bimorph deformable mirror as the wavefront corrector. The dynamic model of the mirror is modified from the usual plate equation, and the turbulence phase model is presented as the output of a linear shaping filter through the orthogonal basis of Zernike polynomials. By proper truncation of the number of eigenfunctions of Laplace operator and the number of Zernike model used to describe the phase turbulence, an H_∞ controller in an infinite-dimensional setting is designed to improve the robustness while optimizing the worst-case performance. Numerical results show a significant improvement in the performance level with respect to traditional single-input single-output methods due to the multivariable nature of the control problem involved in adaptive optics systems.

6.1.3 Distributed Control Methods

The distributed control techniques usually involve the use of distributed system models or spatially invariant distributed models for large-scale dynamic systems. The use of microactuators in the design of wavefront correctors has led to the development of large arrays of actuators. There is now a significant effort to develop deformable mirrors based on MEMS technology. Such mirrors will be relatively small, and they may have thousands of actuators to enable high-resolution images. In adaptive optics systems for astronomy, the deformable mirror is controlled to counteract the spatially distributed wavefront distortion with a spatial resolution which can be defined using the Fried parameter r_0. Fried parameter measures the optical quality of the atmosphere and corresponds to an area over which the RMS wavefront aberration is less than 1 rad, which is typically in the order of 20–60 cm. For extremely large telescopes with diameters $D > 30$ m, the number of actuators scales with $(D/r_0)^2$ and will be of the order of 10^4–10^5. This high-dimensional system complicates the modeling and controller design for DMs with regard to optimal objectives. The notion of spatial invariance addresses this issue by implying that the dynamics of a spatially distributed system is invariant with respect to translation along some spatial coordinates (Bamieh et al. 2002). The symmetric distribution of actuators and sensors in some systems allows spatial invariance to be exploited in their analysis and control. Standard controller techniques for finite-dimensional systems can then be applied. The ideas of spatial invariance can be implemented for both the tracking and stabilization of membranes with a large array of actuators and sensors as in AO systems.

In Bamieh (1997), distributed control of spatially invariant systems is explored, where a large array of sensors and actuators is considered. It was shown that as a general principle, LQG, H_2 and H_∞ optimal controllers, under a variety

of performance criteria inherit any spatial invariance of the plant. Namely, it was shown that if the plant structure is spatially invariant, the best achievable performance can be approached with a spatially invariant controller. The systems considered and the optimal controllers designed in Bamieh (1997) are typically infinite dimensional. However, by exploiting the properties of spatial invariance, it was shown in Castros and Paganini (2002) that optimal controllers, such as LQG, H_2, and H_∞ controllers, can be designed using finite-dimensional approximations of the infinite-dimensional problem. The finite-dimensional approximations of the infinite-dimensional controllers are developed using "spatial truncation" (developed by the authors) as opposed to the standard practice of modal truncation. The control algorithms are not implemented as centralized controllers, which would consider information from all the sensors in the array. Instead, the controllers are implemented as distributed controllers, where each actuator runs on an algorithm with information from its neighboring sensors. A similar controller was developed to control an AO telescope in Kulkarni et al. (2002).

In Castros and Paganini (2002), a method was proposed to synthesize localized and decentralized controllers for distributed systems with spatial invariance. However, unlike the method in Bamieh et al. (2002), this method directly imposes localization in the controller design. Conditions are provided for the existence of localized controllers that satisfy performance specifications, namely, H_∞ conditions. However, as opposed to Bamieh et al. (2002), these conditions are only sufficient. The work proposed in Castros and Paganini (2002) can be used in shape control problems for systems with spatial invariance. Voulgaris et al. (2003) considered the design of optimal H_2 controllers for stable distributed discrete-time spatially invariant systems with a temporal delay in the interaction of neighboring sites. By employing a standard parameterization approach, the constraints on the controller transform to convex constraints on the Youla parameter, given that the plant has the same cone-causality structure.

In Stein and Gorinevsky (2005), the control of a large space reflector structure, treated as a spatially invariant infinite-dimensional system, was achieved using a large distributed actuator array. The system was modeled using experimentally determined actuator influence functions. The influence functions were localized in the spatial neighborhoods of the respective actuators. The notion of spatial invariance was exploited to develop simple PI controllers as localized controllers. The use of localized constraints on the control law structure allows for scalability to very large arrays of actuators and sensors. Design examples resembling an actively controlled reflector whose deformations are controlled by an array of actuators and sensors were used to demonstrate the ability of the controller to satisfy closed-loop engineering specifications such as stability, robustness, and performance, where the performance criteria involve regulation of the error value as described above.

Fraanje et al. (2010) proposed a decomposition approach to distributed control for thin membrane or plate-based deformable mirrors. The actuators are assumed to be distributed in a hexagonal pattern, and each consists of a parallel connection of a spring, a damper, and a force actuator. The partial differential equation model of the thin plate is first discretized using the finite difference approximation of

the biharmonic operator over the hexagonal grid with unit distance between two neighboring actuators. Then, using a forward Euler discretization approach, the dynamic model of the deformable is written as a decomposable system consisting of an interconnection of a series of equal and small-size subsystems (Massioni and Verhaegen 2009), based on which efficient and structured controllers can be derived. This technique allows the design of a distributed H_2 controller for a deformable mirror of any size, with a computational cost that does not increase with the size of the mirror.

6.1.4 Adaptive Control Methods and Others

Due to the dynamic nature of the wavefront aberrations, the shape control problem of DM addressed in adaptive optics is the tracking of an unknown and time-varying shape for a distributed mirror surface (i.e., desired shape of the mirror). In the past, researchers in adaptive optics have introduced a variety of control approaches for wavefront prediction and reconstruction, but until recently the control loops in adaptive optics systems have been linear time-invariant. Classical AO control systems are based on classical feedback control methods so that the controllers are linear and time-invariant (LTI) and cannot adapt to the changes of the external disturbance. Optimal controllers are also LTI. Though they may involve minimum-variance prediction based on spatial and temporal wavefront statistics, the optimal controllers are only suitable to LTI plants with stationary disturbance and require either a priori modeling of wavefront statistics or off-line system identification of a Kalman predictor from sampled wavefront sequences. Control loops which have fixed gains that are based on known or assumed statistics of the system cannot be considered adaptive. Since the desired shape of the mirror in an AO system is unknown and time varying, it would be beneficial to have the controller tuned adaptively online to converge to the controller needed to achieve regulation. In recent years, adaptive control algorithms based on adaptive estimation of optimal reconstructor matrices have been proposed to improve the performance of AO systems in applications with strong, time-varying atmospheric turbulence. A number of adaptive control algorithms have been developed for adaptive optics based on different recursive algorithms (Ficocelli and Ben Amara 2012; Monirabbasi and Gibson 2010; Gibson et al. 2000; Zhu et al. 1999). Adaptive controllers have two important properties. The first is their capability to compensate for loop latency by predicting wavefront error, and the second is the capability to identify optimal gains in real time as the wavefront properties or statistics change.

In Zhu et al. (1999), an iterative control algorithm to track surface shapes based on experimentally measured mirror shape responses, that is, an influence function, is presented. The WFC contains a DM and 19 electrostatic actuators and is used primarily for astronomical applications. The authors use a centralized control scheme that is able to update all the control signals simultaneously. Following the approach of a standard steepest descent algorithm, the derivative of the measured

wavefront error (i.e., the error between the actual and ideal shapes of the DM) is evaluated with respect to the control signal vector. This gives the direction of the increasing wavefront error in the multidimensional control signal space at the point of the current control signal vector. An adjustment of the control signal vector in the opposite direction can reduce the wavefront error. This is done for all the control signals at once so that the total wavefront error moves toward a minimum. This approach is similar to that used in many adaptive signal processing applications with the difference that the method used in Zhu et al. (1999) uses a multivariable adaptive algorithm.

In Gibson et al. (2000), a feedback control loop containing a digital integrator and a feedback controller are used to choose control inputs for an AO system. The objective of the control system is to minimize a cost function J using a multichannel recursive-least-squares (RLS) lattice filter. The gains of the filter are adaptively chosen using an RLS algorithm to minimize the error between the desired shape of the mirror and the actual shape (i.e., minimize the cost function J). Multichannel lattice filters are used because they are among the few classes of algorithms that could yield the numerical stability and the speed required for real-time AO. The multichannel RLS lattice filter remained numerically stable for a large numbers of channels typical of AO systems.

Monirabbasi and Gibson (2010) presented an experimental implementation of a minimum-variance adaptive control scheme. The adaptive controllers are based on a multichannel recursive-least-squares (RLS) lattice filter, which identifies gains in real time for minimum-variance prediction of wavefront sequences. The coupled multichannel structure of this adaptive filter allows both spatial and temporal wavefront statistics to be incorporated implicitly in the prediction filter. The adaptive control approach is evaluated in an adaptive optics experiment setup, where closed-loop wavefront errors measured by a self-referencing interferometer are fed back to the control loops and drive a membrane deformable mirror with 31 actuators to correct the wavefront. The corrected laser beam is imaged by a diagnostic target camera. The experimental results show reduced closed-loop wavefront errors and correspondingly sharper diagnostic target images produced by the adaptive control loop as compared with the classical AO control loop.

In Ficocelli and Ben Amara (2012), an adaptive shape control algorithm is developed to compensate for unknown and time-varying aberrations in the eye. The proposed controller design approach relies on two steps. The first step is to construct a Q-parameterized set of stabilizing controllers for the system under consideration and to derive conditions on the Q parameter in the controller expression to achieve regulation. Partial decoupling of the closed-loop system dynamics is performed to facilitate the development of the adaptive regulator. Since the desired shape of the mirror is unknown and time varying, the second step is to derive an online tuning algorithm for the Q parameter in the expression for the parameterized stabilizing controller. The online tuning of the Q parameter allows the controller to converge to the controller needed to achieve regulation, hence compensating for the lack of information on the desired shape for the deformable mirror. The decoupling

introduced in the closed-loop system allows the tuning to be performed using decentralized adaptation algorithms.

Besides the control methods mentioned above, other modern control methods have also been proposed and applied to the AO systems. In Montera et al. (1997), the authors examine the use of neural networks for WFS slope measurement estimation with only the WFS measurements as inputs. They then compare the neural network solution to a linear minimum mean-square-error (LMMSE) estimator, which is based on prior knowledge of the second-order statistics of the wavefront slopes and measurement noise. The work shows that overall the neural network solution outperforms the LMMSE estimator. In (Dessenne et al. 1997), the argument is made that correction efficiency of AO systems is limited mainly by measurement noise and time delay. The authors describe a new optimized linear predictive controller whose parameters are optimized by means of a least-squares algorithm to minimize the residual optical phase variance. The least-squares algorithm is used to estimate parameters in the predictive controller according to the statistics of the noise and turbulence measurements of the WFS and was shown to work better than a classical optimized integral controller. Poyneer and Veran (2008) presented a predictive control approach for adaptive optics with arbitrary control loop delays based on a new form of the predictive Fourier control and Kalman filters. State-space model parameters are directly estimated from closed-loop telemetry and are used to solve the algebraic Riccati equation, producing the steady-state Kalman filter to predict the atmosphere. The fundamental structure of the predictive control law is independent of arbitrary computational delays. The simulation shows that the use of a predictive controller provides significantly better performance than that of an integral controller.

The above review is only a sample of the most important contributions in the literature pertaining to the control of adaptive optics systems. However, with the development of AO systems for emerging and more complex applications, it is anticipated that new control methods will continue to be developed in the future.

6.2 Control System for the Magnetic Fluid Deformable Mirrors

Magnetic fluid deformable mirrors (MFDM) have a major intrinsic advantage over solid optics. Their main advantages are that they are inexpensive and can produce very large deformations. Coated with MELLFs, the surface of MFDM can be shaped by the application of a magnetic field and generates reflecting surfaces with complicated shapes. However, earlier designs of MFDMs proposed by Laird et al. (2003, 2006) present a major inconvenience in that the mirror surface shape response to the applied control magnetic field is nonlinear and the induced deformations depend on the square of the external magnetic field, resulting in complicated control algorithms for the MFDMs. As shown in Chap. 5, these

problems can be overcome by superposing a strong and uniform magnetic field to the magnetic field of the actuators, thereby linearizing the response of the MFDMs. The major advantage of this linearization is that proven control algorithms used with solid mirrors or other modern control algorithms could be applied to the MFDM. Using the designed prototype mirror and the experimental setup, the linear static response of the prototype MFDM and the dynamic characteristics of the mirror have been verified by comparing them with those predicted by the analytical model in Chap. 5. The experimental results show that when the Helmholtz coil is turned on, the principle of superposition holds distinctly and that the bidirectional displacements of the mirror surface are achieved.

In the following two chapters, linear control algorithms will be developed to control the MFDM surface shape in a closed-loop AO system. In Chap. 7, decentralized control structures are used due to their simplicity and reduced computational burden. First, a decentralized proportional-plus-integral (PI) controller designed based on decoupling the plant model at DC is presented. This is actually the commonly used type of controllers in adaptive optics systems referred to as the classical decentralized control approach discussed in Sect. 6.1.1. The capabilities and limitations of this type of controllers are investigated. To overcome stability robustness issues in this first approach, a decentralized robust proportional–integral–derivative (PID) controller is then presented. The controller is designed based on a decoupled nominal dynamic model of the plant. To account for the modeling errors in the decoupled plant model, a robust PID controller design problem is formulated and solved using algorithms based on linear matrix inequalities (LMIs). The two control algorithms mentioned above provide good performance in removing static or very slowly changing aberrations but have limited capabilities in dealing with fast-changing aberrations.

To overcome this limitation, two centralized optimal control methods are developed in Chap. 8, including the multivariable PID controller and the mixed-sensitivity H_∞ multivariable controller, which are particularly suitable for applications in AO imaging systems with dynamic wavefront aberrations. For example, these controllers ensure that the mirror can cancel unknown dynamic wavefront aberrations that fall within the band of frequencies commonly seen in ocular wavefront aberrations. As opposed to the existing control algorithms, which are used to compensate for static high-order aberrations, the control algorithms proposed in Chap. 8 are capable of compensating dynamic wavefront aberrations.

Because of their simplicity, PID controllers are widely used and can specifically achieve asymptotic tracking of static reference inputs due to the presence of integrators in their structure. However, most of the existing PID design approaches are developed for SISO systems. New approaches have been recently developed for continuous time MIMO systems by transforming the PID controller design problem into an equivalent static output feedback controller design problem. However, these approaches are proposed based on solving bilinear matrix inequalities (BMIs) that should be solved iteratively. The search for the optimal controller is heavily dependent on the initial values fed into the synthesis procedure, and most likely the obtained solution may be unacceptable. Given that the control algorithms for an AO

system are typically implemented in a computer system, a new optimal multivariable PID controller design approach for discrete-time systems is proposed in Chap. 8. The multivariable PID controller parameters are designed directly by considering the actual closed-loop system involving the plant and the controller, without transforming the problem into an equivalent static output feedback controller design problem. The proposed controller synthesis procedures are based on solving properly formulated linear matrix inequalities (LMIs). The LMI-based formulation of the controller synthesis procedure results from introducing some assumptions on the state-space representation of the plant, from properly formulating the controller state-space representation, and from introducing slack variables with special structure in the controller synthesis results.

The centralized mixed-sensitivity H_∞ controller with integral function is also developed for the dynamic shape control problem for MFDMs. There are two main motivations for using the H_∞ controller design approach for MFDMs. First, the aberrations of the human eye can be modeled as a combination of static, harmonic, and random components, and the temporal frequency content of the aberrations in the human eye has a limited range. Second, the proposed control scheme minimizes the magnitude of high-frequency components in the controller output, that is, in the current input to the MFDM, which is beneficial in avoiding saturation of the control signals and in attenuating unmodeled high-frequency modes of the waveform corrector. In fact, an analytical model of the MFDM presented in Chap. 4 shows that the system model is infinite dimensional. For practical reasons, the controller is usually designed based on a truncated model where high-frequency dynamics are ignored. Consequently, the mixed-sensitivity H_∞ design method provides the means to deal with this model uncertainty in the high frequency range. The experimental results with the prototype 19-channel magnetic fluid deformable mirror demonstrate the efficiency of both proposed methods.

It should be noted that the methods proposed in the following two chapters are derived based on the analytic model developed in Chap. 4 and the dynamics of the MFDM model is fully considered in the design process. These proposed control methods are not limited to MFDM-based AO systems but can also be applied to other AO systems with different types of wavefront correctors. Some concepts behind the presented control approaches and corresponding details of the derivations are also included selectively in this book.

6.2.1 Description of the Closed-Loop System

In this section, the control architecture for the MFDM-based AO control system is presented. The system model is analytically obtained based on the prototype MFDM with 19 actuators as shown in Chap. 5. The Shack–Hartmann type wavefront sensor is used to measure the wavefront slope data at each subapertures of the wavefront sensor. The control algorithms are developed based on the zonal compensation framework. In the zonal approach, the wavefront reconstructor attempts to compute

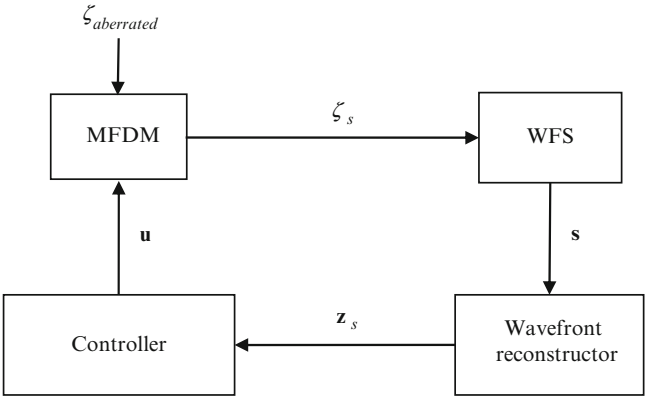


Fig. 6.4 Closed-loop diagram of the AO system

an estimate of the shape of the residual wavefront at a finite number of spatial locations corresponding to the actuator locations. In the closed-loop AO system, the function of the controller is to utilize the residual wavefront shape measurements to generate a set of actuator commands to effectively drive the residual wavefront shape measurements to zero.

A closed-loop diagram that includes the principal elements of the AO system, namely, the MFDM, the WFS with the wavefront reconstructor, and the controller, is shown in Fig. 6.4. $\zeta_{\text{aberrated}}$ is the shape of the aberrated wavefront incident on the mirror surface, and ζ_s is the shape of the wavefront reflected from the mirror and sensed by the wavefront sensor.

Physically, the aberrated wavefront shape $\zeta_{\text{aberrated}}$ and the residual wavefront shape ζ_s are distributed signals with continuous spatial and temporal variations. All of these signals are evaluated at a set of spatially distributed points. Thus, $\zeta_{\text{aberrated}}$ and ζ_s are represented as vectors. The Shack–Hartmann wavefront sensor measures local subaperature slopes of the residual wavefront $\zeta_{s,}$ and the slope measurements produced by the sensor are represented as a vector **s** and then sent to the wavefront reconstruction software. The wavefront reconstructor uses the slope measurements to produce an estimate of the residual wavefront shape in the form of a vector $\mathbf{z}_s$. The residual wavefront estimates are used by the controller to generate the actuator input signals. When the closed-loop system is operating properly, the shape of the MFDM should be such that at steady state, the residual wavefront shape ζ_s at the actuator locations should be zero in the case of static aberrations or, in the case of dynamic aberrations, the variance of the residual wavefront shape is minimized.

Figure 6.5 shows the control block diagram of the closed-loop AO system. The system makes use of a wavefront corrector where **u** is the vector of control currents applied to the MFDM and ζ is the resulting shape of the deformable mirror surface. The mirror surface shape and the incoming aberrated wavefront interact in such a way that the sensed residual wavefront shape ζ_s can be written as (Baudouin et al. 2008)

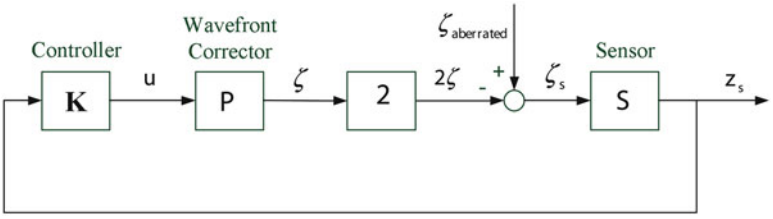


Fig. 6.5 Block diagram of a typical closed-loop adaptive optics system

$$\zeta_s = \zeta_{\text{aberrated}} - 2\zeta \tag{6.11}$$

The sensor block in Fig. 6.5 refers to both the wavefront sensor and the reconstruction algorithm that run inside the control computer. The sensor samples the residual wavefront ζ_s at discrete locations—using an array of lenslets in the case of the Shack–Hartmann wavefront sensor. Though the sampled data can be directly utilized in control computations, it is generally more helpful to reconstruct the shape of the sensed wavefront surface. The reconstructed shape can then be used to interpolate the wavefront surface displacements at any location in the pupil. The estimated wavefront shape, expressed as a vector $\mathbf{z}_s \in \mathbb{R}^M$ of M discrete displacements, is provided to the controller $\mathbf{K}$, which computes the actuator commands $\mathbf{u} \in \mathbb{R}^L$ to be applied to the array of L input actuators.

To simplify the implementation of the closed-loop system and the testing of the performance of the proposed control algorithms, an equivalent closed-loop system configuration is considered in Fig. 6.6a. In this configuration, a laser beam with a planar wavefront is directed toward the deformable mirror with surface shape ζ. The wavefront shape in the reflected beam is given by $\zeta_s = -2\zeta$. The residual wavefront shape slopes are measured by the wavefront sensor, and an estimate residual wavefront shape is reconstructed inside the control computer. The estimated residual wavefront shape vector $\mathbf{z}_s \in \mathbb{R}^M$ is sampled from the reconstructed shape. The estimated residual wavefront shape is then added, inside the control computer, to a hypothetical aberrated wavefront shape. The hypothetical aberrated wavefront shape is sampled at the same locations at which the estimated residual wavefront shape vector $\mathbf{z}_s$ is computed from the reconstructed shape. An equivalent block diagram representation that is consistent with the representation of negative feedback control systems is shown in Fig. 6.6b, where $\mathbf{y}_s = -\mathbf{z}_s$. Let $\mathbf{r} \in \mathbb{R}^M$ denote the sampled aberrated wavefront shape. The wavefront shape error fed to the controller $\mathbf{K}$ can be written as

$$\mathbf{e} = \mathbf{r} - \mathbf{y}_s \tag{6.12}$$

In this configuration, the controller $\mathbf{K}$ is required to provide input commands $\mathbf{u}$ such that the magnitude of the wavefront shape error $\mathbf{e}$ is minimized. A statement of the control problem addressed for the deformable mirror is then given as follows.

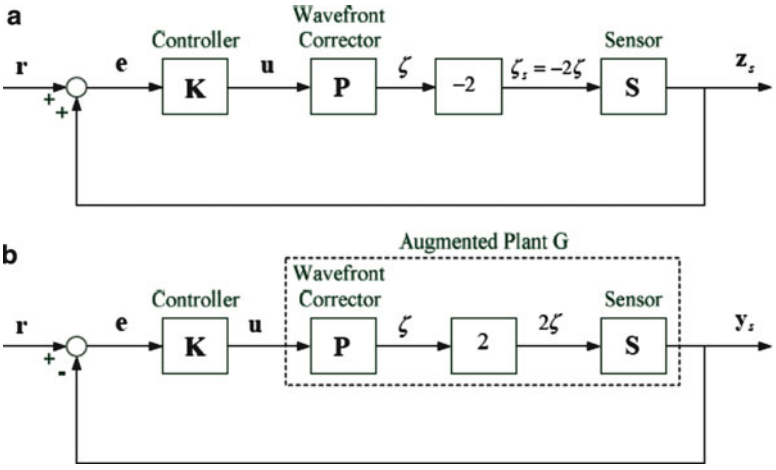


Fig. 6.6 (**a**) Block diagram of the modified closed-loop adaptive optics system. (**b**) Equivalent negative feedback block diagram representation

6.2.1.1 Control Problem Statement

> With respect to the feedback system shown in Fig. 6.6b, it is desired to design a controller **K** to provide input currents **u** such that the signal $\mathbf{y}_s$ (representing twice the shape of the MFDM surface) tracks the reference wavefront shape signal **r** (representing the aberrated wavefront shape $\zeta_{\text{aberrated}}$).

The main advantage in considering the feedback system shown in Fig. 6.6b is that any aberrated wavefront shape can be easily considered and analytically generated as opposed to physically generating it as in the closed-loop system shown in Fig. 6.5.

6.3 Summary

By superimposing a strong uniform magnetic field on top of the field created by the distributed coil actuators, the nonlinear response of the surface of MFDMs becomes linear, hence allowing for the effective control of the mirror surface shape in AO systems. Control algorithms that have been used extensively in AO systems are reviewed in this chapter. The classical decentralized control methods are simple and easy to implement. However, since full decoupling of the plant dynamics is assumed in the design of such controllers, the resulting AO closed-loop control systems cannot meet the high-performance requirements associated with advanced AO system applications. A promising way to improve the control system performance is to use centralized optimal control method by exploiting the spatiotemporal correlation both in the wavefront turbulence and the dynamics in DM as well. The recently developed distributed control methods are suitable for

AO systems involving a very large number of actuators and characterized by special invariance in their dynamics. In order to further deal with the time-varying nature of AO systems and of the aberrations, adaptive control methods and other approaches are developed where the controller parameters are adjusted online to achieve a better performance in the AO closed-loop control system. Control approaches designed for the shape control problem of MFDM are presented in the next two chapters. To simplify the implementation and the verification of the proposed control algorithms, an equivalent closed-loop system configuration is considered, where the arbitrary wavefront turbulence is analytically generated in the computer software instead of being physically generated.

References

Auburn J, Lorell K, Mast T, Nelson J (1987) Dynamic analysis of the actively controlled segmented mirror of the W.M. Keck ten-meter telescope. IEEE Contr Syst Mag 7(6):3–10

Bamieh B (1997) The structure of optimal controllers of spatially-invariant distributed parameter systems. Proc IEEE Conf Decis Contr 2:1056–1061

Bamieh B, Paganini F, Dahleh A (2002) Distributed control of spatially invariant systems. IEEE Trans Automat Contr 47:1091–1107

Baudouin L, Prieur C, Guignard F, Arzelier D (2008) Robust control of a bimorph mirror for adaptive optics systems. Appl Opt 47:3637–3645

Castros G, Paganini F (2002) Convex synthesis of localized controllers for spatially invariant systems. Automatica 38:445–456

Correia C, Raynaud HF, Kulcsar C, Conan JM (2010) On the optimal reconstruction and control of adaptive optical systems with mirror dynamics. J Opt Soc Am A 27(2):333–349

Dessenne C, Madec P, Rousset M (1997) Modal prediction for closed loop adaptive optics. Opt Lett 22:1535–1537

Downie J, Goodman W (1989) Optimal wavefront control for adaptive segmented mirrors. Appl Opt 28:5326–5332

Ficocelli M, Ben Amara F (2012) Online tuning for retinal imaging adaptive optics systems. IEEE Trans Contr Syst Technol 20(3):747–754

Fraanje R, Massioni P, Verhaegen M (2010) A decomposition approach to distributed control of dynamic deformable mirrors. Int J Optomechatronics 4(3):269–284

Fried D (1990) Time delay induced mean square error in adaptive optics. J Opt Soc Am A 7:1224–1227

Gibson JS, Chang CC, Ellerbroek BL (2000) Adaptive optics: wave-front correction by use of adaptive filtering and control. Appl Opt 39:2525–2538

Goodman JW (2004) Introduction to Fourier optics. Roberts & Company, Colorado

Hinnen K, Verhaegen M, Doelman N (2007) Exploiting the spatiotemporal correlation in adaptive optics using data-driven H_2-optimal control. J Opt Soc Am A 24:1714–1725

Hinnen K, Verhaegen M, Doelman N (2008) A data driven H_2 optimal control approach for adaptive optics. IEEE Trans Contr Syst Technol 16(3):381–389

Huang J, Looze D, Denis N, Castanon D, Wirth A (1995a) Dynamic modeling and identification of an adaptive optics system. In: Proceedings of the 4th IEEE conference on control applications, Albany, NY, USA, pp 456–463

Huang J, Looze D, Denis N, Castanon D, Wirth A (1995b) Control design for an adaptive optics system. In: Proceedings of the conference on decision & control, New Orleans, LA, USA, pp 3753–3756

Kulcsár C, Raynaud HF, Petit C, Conan JM, Viaris de Lesegno P (2006) Optimal control, observers and integrators in adaptive optics. Opt Express 14:7464–7476

Kulkarni J, D'Andrea R, Brandl B (2002) Application of distributed control techniques to the adaptive secondary mirror of Cornell's large Atacama telescope. In: SPIE astronomical telescopes and instrumentation conference, vol 4839, pp 750–756

Laird P, Bergamasco R, Berube V, Borra EF, Ritcey AM, Rioux M, Robitaille N, Thibault S, Lande Vieira da Silva Jr., Yockell-Lelivre H (2003) Ferrofluid-based deformable mirrors: a new approach to AO using liquid mirrors. In: Wizinowich PL, Bonaccini D (eds) Adaptive Optical System Technologies II, Proceedings of SPIE, vol. 4839, the International Society for Optical Engineering

Laird P, Caron N, Rioux M, Borra EF, Ritcey AM (2006) Ferrofluid adaptive mirrors. Appl Opt 45(15):3495–3500

Le Roux B, Conan JM, Kulcsar C, Raynaud HF, Mugnier LM, Fusco T (2004) Optimal control law for multiconjugate adaptive optics. J Opt Soc Am A 21:1261–1276

Looze DP (2006) Minimum variance control structure for adaptive optics systems. J Opt Soc Am A 23:603–612

Looze DP (2009) Linear quadratic Gaussian control for adaptive optics systems using a hybrid model. J Opt Soc Am 26(1):1–9

Looze DP (2010) Discrete time model for an adaptive optics system with input delay. Int J Contr 83(6):1217–1231

Massioni P, Verhaegen M (2009) Distributed control for identical dynamically coupled system: a decomposition approach. IEEE Trans Automat Contr 54(1):124–135

Monirabbasi S, Gibson S (2010) Adaptive control in an adaptive optics experiment. J Opt Soc Am A 27(11):A84–A96

Montera D, Welsh B, Roggemann C, Ruck D (1997) Prediction of wavefront sensor slope measurement with artificial neural networks. Appl Opt 36:675–681

Paschall R, Anderson D (1993) Linear quadratic Gaussian control of a deformable mirror adaptive optics system with time-delayed measurements. Appl Opt 32:6347–6358

Petit C, Conan JM, Kulcsár C, Raynaud HF (2009) Linear quadratic Gaussian control for adaptive optics and multiconjugate adaptive optics: experimental and numerical analysis. J Opt Soc Am A 26(6):1307–1324

Poyneer L, Veran JP (2008) Predictive wavefront control for adaptive optics with arbitrary control loop delays. J Opt Soc Am A 25(7):1486–1496

Raynaud HF, Correia C, Kulcsar C, Conan JM (2011) Minimum variance control of astronomical adaptive systems with actuator dynamics under synchronous and asynchronous sampling. Int J Robust Nonlin Contr 21(7):768–789

Roddier F (1999) Adaptive optics in astronomy. Cambridge University Press, Cambridge

Stein G, Gorinevsky DM (2005) Design of surface shape control for large two-dimensional array. IEEE Trans Contr Syst Technol 13:422–433

Tyson RK (2000) Adaptive optics engineering handbook. Marcel Dekker, New York

Tyson RK (2011) Principles of adaptive optics. CRC Press, Boca Raton

Von Bokern MA (1990) Design of a linear quadratic Gaussian control law for an adaptive optics system. MS Thesis, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson Air Force Base, Ohio

Voulgaris PG, Bianchini G, Bamieh B (2003) Optimal H_2 controllers for spatially invariant systems with delayed communication requirements. Syst Contr Lett 50:347–361

Wiberg DM, Max CE, Gavel DT (2004a) A spatial non-dynamic LQG controller: part 1, application to adaptive optics. In: Proceedings of the IEEE conference on decision and control, Atlantis, Paradise Island, Bahamas, USA, pp 3326–3332

Wiberg DM, Max CE, Gavel DT (2004b) A spatial non-dynamic LQG controller: part 2, theory. In: Proceedings of the IEEE conference on decision and control, Atlantis, Paradise Island, Bahamas, USA, pp 3333–3338

Zhu L, Sun P, Bartsch D, Freeman W, Fainman Y (1999) Adaptive control of a micromachines continuous-membrane deformable mirror for aberration compensation. Appl Opt 38:168–176

Chapter 7
Decentralized PID Controller Design

Contents

7.1 Plant Model Decoupling

A decentralized control system consists of a set of parallel decentralized single-input single-output compensators designed based on a properly decoupled model of the original MIMO system. This classical control strategy has the advantage of presenting a simple structure and a low computation burden and is still widely used in the AO community. The resulting closed-loop system is typically characterized by a reasonable performance under the moderate assumption that the wavefront aberrations have a power spectrum mainly in low frequency range. Through the integral action, the PID controller yields a high loop gain in the low frequency range, which implies effective rejection of slowly time-varying disturbances. In this chapter, decentralized PID controllers are designed based on decoupling the plant model in the AO system so that the controller design problem for the MIMO system is transformed into a set of controller design problems for SISO systems. The

Z. Wu et al., *Modeling and Control of Magnetic Fluid Deformable Mirrors for Adaptive Optics Systems*, DOI 10.1007/978-3-642-32229-7_7,

coupling among the different input–output channels of the MFDM is first minimized by appropriately choosing the location of the output points on the mirror surface, which ensures that each output point is influenced maximally by one input coil only, that is, the one corresponding to its own locations, while the effect of the other inputs is significantly small. Then, a decoupling technique using the DC gain (influence function) of the MIMO system is considered to further minimize the coupling effects between the different input–output channels.

7.1.1 The Plant Model

In the following, the analytical model of the MFDM presented in Chap. 4 is used in the development of controllers to control the mirror surface shape in the closed-loop system shown in Fig. 6.4b. As shown in Fig. 6.6b, an augmented plant $\mathbf{G}$ is defined and includes the WFC, a model of the optical interaction between the incident beam and the deformable mirror, and the wavefront sensor. Using the notation presented in Chap. 4, the WFC $\mathbf{P}$ will be represented by the transfer function $\mathbf{P}(s)$ and a corresponding state-space realization $\mathbf{P} : \left[\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{0} \end{array}\right]$, where the system matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are obtained from the mirror model (4.102). As mentioned earlier, the deformable mirror introduces a change in the wavefront shape, and the magnitude of this change is twice the magnitude of the mirror shape. The sensor also introduces a delay in the closed loop. Therefore, the augmented plant $\mathbf{G}$ is defined as having a transfer function

$$\mathbf{G}(s) = 2\mathrm{e}^{-\tau s}\mathbf{P}(s) \tag{7.1}$$

where τ is the sensor delay. By approximating the time delay transfer function $\mathrm{e}^{-\tau s}$ using a rational transfer function, that is, the Pade approximation (Chen and Francis 1996), an approximate model of the augmented plant $\mathbf{G}$ can be given in state-space form as

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{A_g}\mathbf{x} + \mathbf{B_g}\mathbf{u} \\ \mathbf{y_s} &= \mathbf{C_g}\mathbf{x} \end{aligned} \tag{7.2}$$

where $\mathbf{x} \in \mathbb{R}^{N_G}$ is the vector of state variables, $\mathbf{y_s} \in \mathbb{R}^{M}$ is the vector of sensor outputs, and $\mathbf{A_g} \in \mathbb{R}^{N_G \times N_G}$, $\mathbf{B_g} \in \mathbb{R}^{N_G \times L}$, and $\mathbf{C_g} \in \mathbb{R}^{M \times N_G}$ are the system matrices with appropriate dimensions.

7.1.2 Decoupling of the Input–Output Channels

Given the low frequency content of the aberrations of interest, it is desired in this section to investigate the decoupling of the plant model at DC. Decoupling the

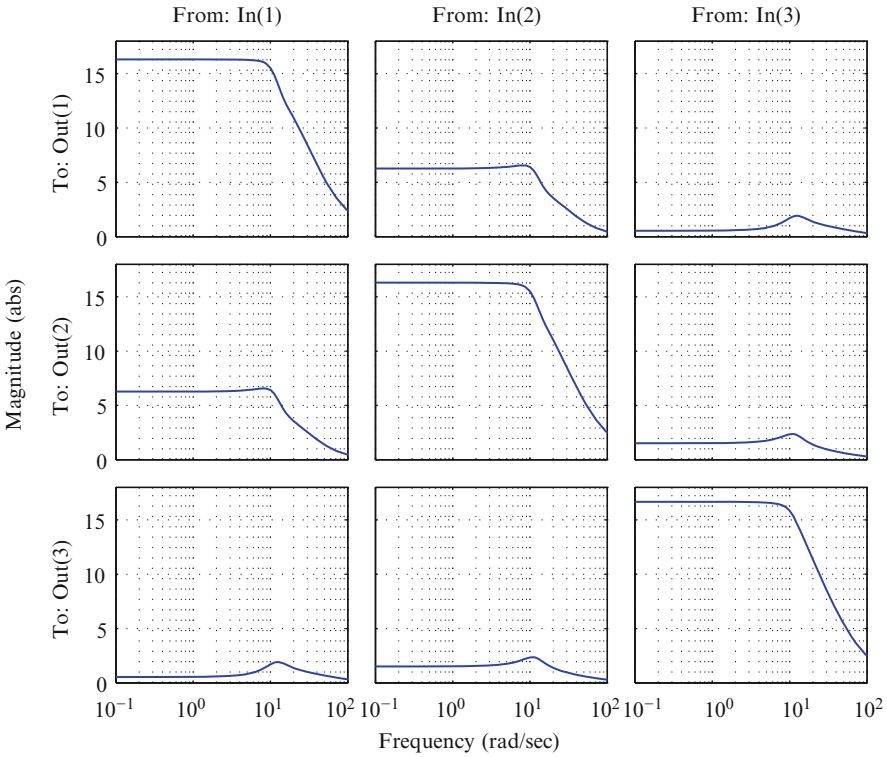


Fig. 7.1 Bode magnitude plots corresponding to three selected input–output channels in the augmented plant model **G**

plant model significantly facilitates the controller design for the AO system, as the problem of MIMO controller design can be reformulated as that of designing a set of SISO controllers. The developed DC-decoupled models are used in the design of decentralized PI and robust PID controllers presented later in this chapter.

As predicted by the analytical model (4.102), and experimentally validated in Chap. 5, the mirror surface shape generated by the magnetic field of a single electromagnetic coil features a Gaussian profile, with the peak surface deflections obtained immediately above the center of the coil. This feature of the MFDM allows the augmented plant **G** to be decoupled and represented as a collection of SISO systems since the coupling among the different input–output channels can be minimized by appropriately choosing the location of the output points. For the 19 input channel system, as described in Sect. 5.2, 19 output points in the reconstructed wavefront shape are chosen such that each of the output points corresponds to the location of an input coil as seen in the optical pupil. This arrangement ensures that each output point is influenced maximally by one input coil only, that is, the one corresponding to its own locations, while the effect of the other inputs is significantly small. The decoupling effect can be clearly observed from the Bode

magnitude plot of the augmented plant $\mathbf{G}$ as shown in Fig. 7.1, where magnitudes in the plots on the leading diagonal are significantly higher than those placed off-diagonal. The figure shows the Bode plots for three selected channels (#1, 2, and 10 as marked in Fig. 5.7) only.

Although the right choice of input–output points does produce the desired decoupling effect, the system is still not fully decoupled. For example, the coupling between the adjacent output points (e.g., 1 and 2) is of the order of 20 % (i.e., the magnitude of the wavefront measured at the output point 2 due to an input applied at coil 1 is approximately 0.2 times the magnitude of the wavefront measured at the output 1 due to the input applied at coil 1). Therefore, some means of further reducing the coupling effect are still needed.

A decoupling technique, which utilizes the DC gain of the MIMO system or the influence function matrix, is considered in the following to minimize the remaining coupling effects. This approach is motivated by the observation that the magnitude responses shown in Fig. 7.1 remain constant over a range of low frequencies. This implies that the augmented plant $\mathbf{G}$ can be considered as a static system in this range of frequencies. Since a static MIMO system can be fully decoupled using the DC gain of the system (Albertos et al. 2004), this property of the plant $\mathbf{G}$ is exploited to further improve the decoupling effect produced by the choice of output points as described above.

The DC gain of the augmented plant $\mathbf{G}(s)$ in (7.2) can be obtained as

$$\mathbf{G}_0 = \mathbf{G}(0) = -\mathbf{C}_g \mathbf{A}_g^{-1} \mathbf{B}_g \tag{7.3}$$

For frequencies $\omega < \omega_0$ over which the gains of all the channels of $\mathbf{G}(s)$ remain constant, the following relationship holds:

$$\left| \mathbf{G}\left(j\omega\right) \mathbf{G}_0^{-1} \right| \approx \mathbf{I} \quad \text{for } \omega < \omega_0 \tag{7.4}$$

where I is an identity matrix. It follows from (7.4) that for $\omega < \omega_0$ the system $\bar{G}$ given by

$$\bar{\mathbf{G}}(s) = \mathbf{G}(s)\mathbf{G}_0^{-1} \tag{7.5}$$

is a decoupled system with a unit DC gain associated with each of the decoupled channels. This decoupling effect can be incorporated in a closed-loop system by simply cascading the gain $\mathbf{G}_0^{-1}$ with the plant $\mathbf{G}$ as shown in Fig. 7.2. The decoupling effect can be more clearly observed in the magnitude response plots for the system $\bar{\mathbf{G}}$ given in Fig. 7.3. It is obvious that the system is significantly decoupled in the low frequency range $\omega < 12$ rad/s. Note that only the magnitude plots on the leading diagonal have significant values. Hence, we have that

$$\left| \bar{\mathbf{G}}\left(j\omega\right) \right| \approx \mathbf{I} \quad \text{for } \omega < 12 \text{ rad/s} \tag{7.6}$$

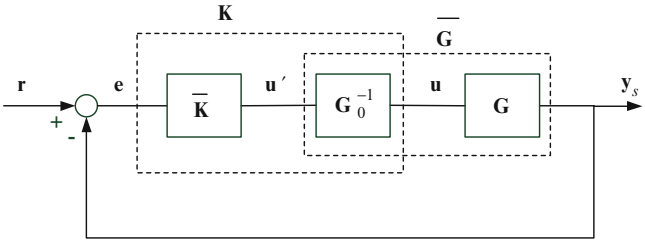


Fig. 7.2 Block diagram of the closed-loop system using DC gain matrix

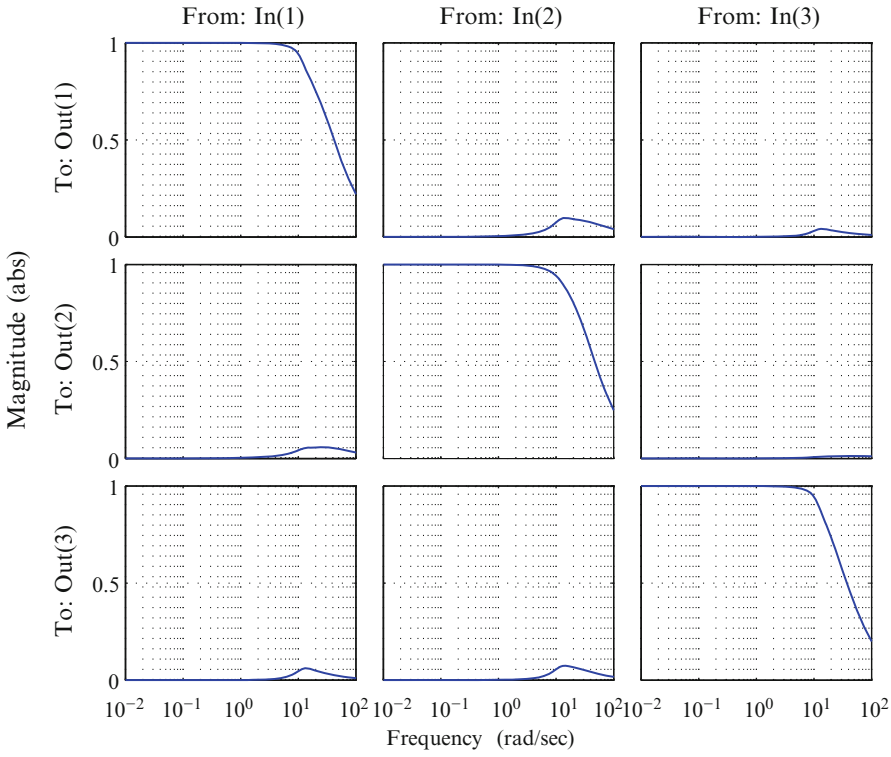


Fig. 7.3 Bode magnitude plots of the decoupled plant model

A decoupled system can be treated as a collection of SISO systems that can be controlled independently. For the system $\bar{\mathbf{G}}$, a decentralized type controller $\bar{\mathbf{K}}$ can be designed such that $\bar{\mathbf{K}} = k\mathbf{I}$, where k is a SISO controller. Given $\bar{\mathbf{K}}$, the overall controller $\mathbf{K}$ for the plant $\mathbf{G}$ can be written as

$$\mathbf{K} = \mathbf{G}_0^{-1}\bar{\mathbf{K}} \tag{7.7}$$

The decoupling approach presented in this section is used in the design of decentralized PI and PID controllers for the AO system as discussed in the following two sections.

7.2 Decentralized PI Controller

Most of the existing AO systems make use of controllers designed based on a DC model of the response of the reflected wavefront shape to the inputs applied to the wavefront corrector. In the AO community, the DC model is commonly referred to as the influence function matrix and is actually the DC gain of the augmented plant comprising the wavefront corrector and the wavefront sensor. For the purpose of controller design in this section, the closed-loop system block diagram is as shown in Fig. 7.2, except that the plant model $\mathbf{G}$ is replaced by its DC gain $\mathbf{G}_0$.

As mentioned in Sect. 3.3, the major difficulty in controlling the conventional MFDM arises from the nonlinear nature of the response of its deformable surface. A clear manifestation of this difficulty has been the failure of the influence-function-based control algorithms (Brousseau et al. 2007; Laird et al. 2006). Now that the nonlinearity of the response of the MFDM has been taken care of by modifying the design of the mirror, it is reasonable to expect that controllers designed based on the DC-decoupled plant model should succeed in removing the distortions in the aberrated wavefront, or at least do so with respect to the static aberrations. In the following, a decentralized PI controller is designed, and its performance is evaluated through simulation studies.

7.2.1 Controller Design

Assuming the plant model $\mathbf{G}(s)$ is approximated by its DC model $\mathbf{G}_0$, it follows that

$$\bar{\mathbf{G}}(s) = \mathbf{I} \tag{7.8}$$

The resulting system is a decoupled system with unit gains in each of the decoupled channels. For such a system, a decentralized type controller $\bar{\mathbf{K}}(s)$ can be designed such that $\bar{\mathbf{K}}(s) = k(s)\mathbf{I}$, where $k(s)$ is a scalar controller designed for the static plant (7.8). In this section, and with the aim of dealing mostly with static aberrations, a proportional-plus-integral (PI) type controller is considered. The controller $k(s)$ is given by

$$k(s) = k_\mathrm{p} + k_\mathrm{i}\frac{1}{s} \tag{7.9}$$

where k_p and k_i are, respectively, the proportional and integral gains of the controller. The corresponding discrete-time transfer function of the controller (7.9) is given by

$$k(z) = k_\mathrm{p} + k_\mathrm{i}\frac{z}{z-1} = \frac{\left(k_\mathrm{p} + k_\mathrm{i}\right)z - k_\mathrm{p}}{z-1} \tag{7.10}$$

The controller gains can be obtained using any of the standard control design methods (Phillips 2000). Using the classical pole placement method, the following parameters were obtained:

$$k_\mathrm{p} = 0.032, \quad k_\mathrm{i} = 0.288 \tag{7.11}$$

The overall controller $\mathbf{K}(s)$ is given by $\mathbf{G}_0^{-1}k(s)\mathbf{I}$ obtained as $k(s)\mathbf{G}_0^{-1}$. The controller thus obtained can be used to control the augmented plant $\mathbf{G}$ in the closed-loop system. Simulation results of the closed-loop system response are presented in the following section.

7.2.2 Simulation of the Closed-Loop System

To evaluate the performance of the decentralized PI controller, typical static and dynamic aberrated wavefront shapes are specified as reference signals $\mathbf{r}$ (see Fig. 7.2). Each entry in $\mathbf{r}$ is associated with a spatial location corresponding to the center of one of the 19 coils. The static reference wavefront shape considered in this simulation is given by

$$\begin{aligned}\mathbf{r}_0 = [&2.0, 4.0, 5.0, 1.0, 5.5, 0.5, 3.5, 8.5, 6.0, 3.0, 2.5, 2.0, 3.0, 4.0, -2.0,\\ &-4.0, -5.0, 1.0, 6.0]^\mathrm{T}(\mu\mathrm{m})\end{aligned} \tag{7.12}$$

whereas the dynamic reference wavefront shape is given by

$$\mathbf{r}(t) = \mathbf{r}_0 + [\sin(f_0 t + \phi_1), \ldots, \sin(f_0 t + \phi_{19})]\,(\mu\mathrm{m}) \tag{7.13}$$

where $\mathbf{r}_0$ is as in (7.12), f_0 is the frequency of the sinusoidally varying component of the aberrated wavefront shape, and ϕ_m, $m = 1, 2, \ldots, 19$, are the phases chosen randomly such that $0 < \phi_m < (\pi/2)$. All the entries in $\mathbf{r}_0$ and $\mathbf{r}(t)$ are expressed in micrometers. In (7.13), the 2-μm peak-to-valley dynamic reference wavefront shape signals are superimposed on the static offset $\mathbf{r}_0$ to represent the aberrations in the eye, which almost always have a large static component. These reference wavefront shapes will be utilized in the simulations of all the controllers presented in this chapter and in the experimental results presented in the following.

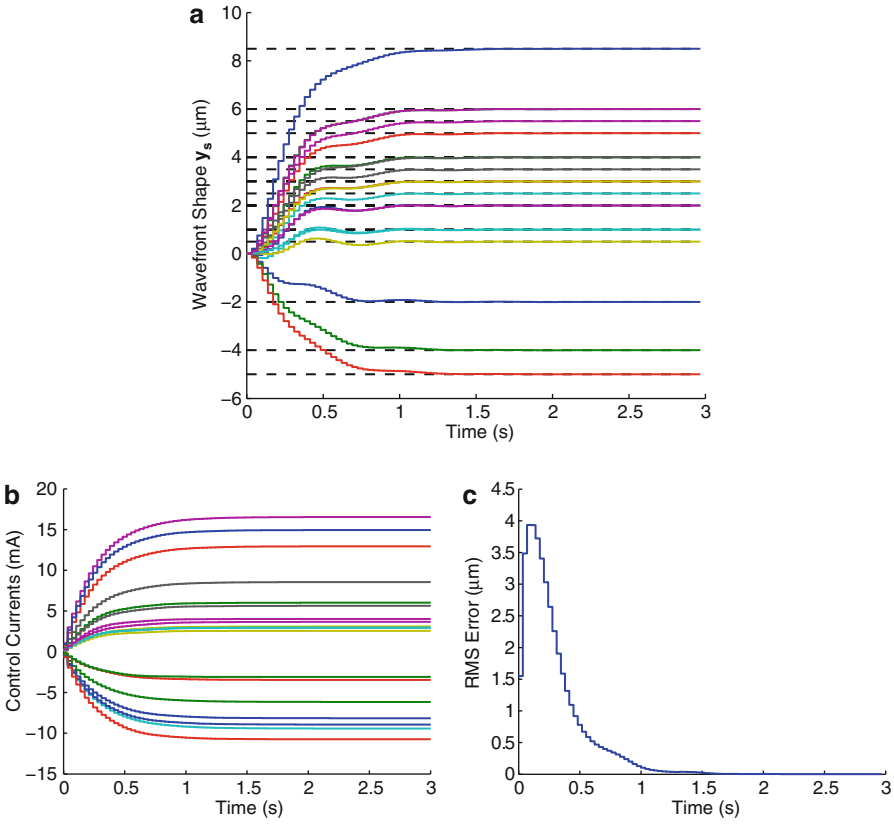


Fig. 7.4 Tracking of a static reference shape using the decentralized PI controller: (**a**) wavefront shape, (**b**) control inputs, (**c**) RMS error

Simulation results involving the static reference wavefront shape given in (7.12) are shown in Fig. 7.4. The time history of the wavefront shape measurements $\mathbf{y}_s$ at the 19 output locations is shown in Fig. 7.4a. The corresponding reference wavefront shape signals $\mathbf{r}_0$ are shown as dashed lines. The control inputs as determined by the designed controller are given in Fig. 7.4b. The root mean square (RMS) of the wavefront shape error $\mathbf{e}$ computed for the 19 channels is given in Fig. 7.4c. As evident from the simulated results, the output of the closed-loop system can successfully track the static reference shape.

Figure 7.5 shows simulation results involving the dynamic reference shape given in (7.13). The time history of the wavefront shape measurements $\mathbf{y}_s$ is shown in Fig. 7.5a. For clarity, the results for only selected channels (# 2, 4, 8, and 17) are shown in the figure. The corresponding reference wavefront shape signals are shown as dashed lines. The control inputs as determined by the designed controller are given in Fig. 7.5b. The RMS of the wavefront shape error $\mathbf{e}$ computed for the 19 channels is given in Fig. 7.5c. As expected, the tracking capability of the closed-loop system is very limited in this case.

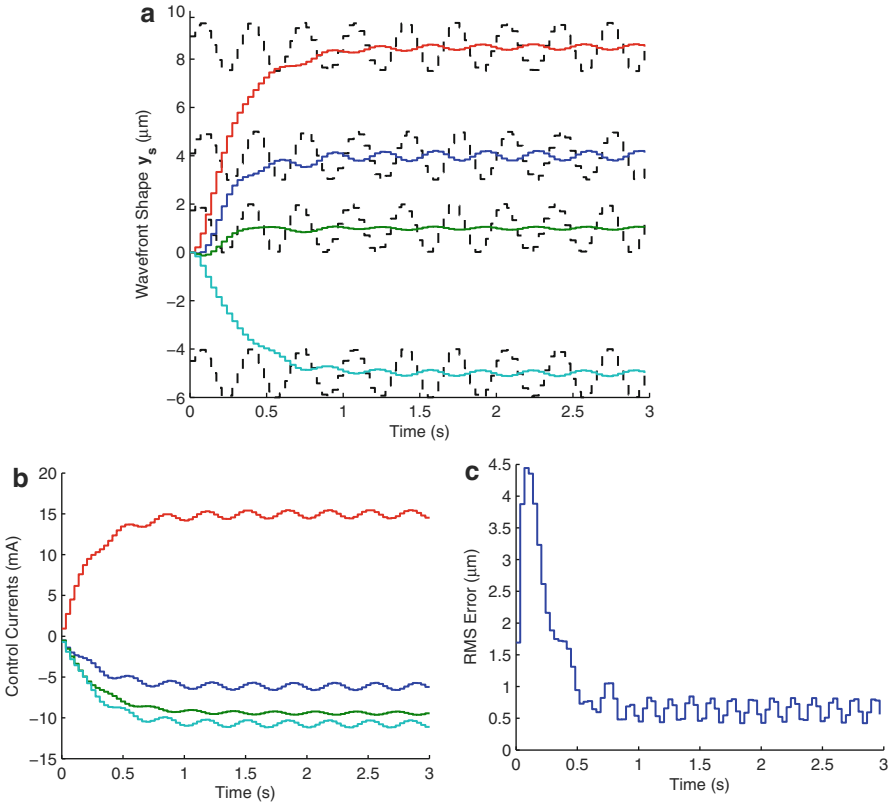


Fig. 7.5 Tracking of a dynamic reference shape using the decentralized PI controller: (**a**) wavefront shape, (**b**) control inputs, (**c**) RMS error

The following important observations can be made from the simulation results presented above:

- It was demonstrated that the PI controller designed based on a DC model (i.e., the influence function) of the response of the WFC can be successfully used to control the MFDMs developed according to the design modification proposed in the Chap. 5.
- The decentralized PI controller asymptotically drives the wavefront shape error to zero, when tracking the static reference wavefront shape signals. High-performance tracking of static reference signals remains a characteristic feature of the controllers that have an integrator in their structure. The same was verified by the simulations presented above.
- The performance of the decentralized PI controller in canceling dynamic wavefront aberrations was found to be inadequate. The controller provides the desired tracking of the dynamic reference wavefront shape signals when the reference

signals change very slowly with time. But the performance of the closed-loop system deteriorates rapidly with the increasing frequency of the reference signals as illustrated in Fig. 7.5a, which shows the results for the dynamic reference wavefront signals (7.13) with 3 Hz frequency. A more advanced PID controller design approach is presented in Chap. 8 and is shown to yield a better performance in dealing with dynamic wavefront aberrations.

- An important closed-loop system stability issue is related to the fact that the PI controller is designed based on the assumption that the augmented plant is a static system. The MFDM exhibits significant dynamics even in the low frequency range, which leads to stability robustness issues in the resulting closed-loop system. This latter issue is addressed in the following section.

7.3 Decentralized Robust PID Controller

A decentralize PI controller was presented in the previous section. The controller was designed based on the assumption that the augmented plant $\mathbf{G}$ can be considered to be a static system represented by its DC gain $\mathbf{G}_0$. Therefore, the resulting DC-decoupled plant model, which is strictly valid for $\omega = 0$ only, was assumed to be valid for all operating frequencies. However, the approximation of the plant model by a static model introduces modeling errors whose effects become more pronounced as the frequency content of the reference inputs increases. These modeling errors may eventually lead to instability in the resulting closed-loop system. In order to minimize the effect of these uncertainties, a decentralized robust PID controller design method for the MFDM shape control is presented in this section. The proposed controller maintains the computational and implementation simplicity of the PI controller considered in Sect. 7.2 but at the same time addresses the modeling uncertainties and provides the desired tracking performance in the closed-loop system for static or slowly varying reference signals. In the design of the decentralized controller, a decoupled approximate dynamic model of the plant is first obtained (as opposed to the static model considered in Sect. 7.2). Then, based on the resulting decoupled system model, a robust PID controller design approach is employed to ensure the system stability in the whole range of operating frequencies is maintained while also achieving the desired tracking performance in the closed-loop system with respect to static or very slowly time-varying shapes of the mirror surface. The design problem is first converted into a multi-objective H_2 / H_∞ static output feedback problem. In this new formulation, the controller to be designed is a static controller represented by the gains of the desired PID controller. The desired controller parameters are then obtained by iteratively solving some properly formulated linear matrix inequalities. The derivations presented in this section are based on a continuous time setting to facilitate the controller design. The resulting controller is then discretized and implemented in the actual experimental adaptive optics system.

7.3.1 Controller Design

A PID controller is considered since it has a very simple structure with only three gains to tune. To further simplify the design and implementation of the PID controller, a decentralized structure of the controller is considered, where the control signal applied to a given actuator is generated based only on one error signal associated with the spatial location of the actuator. To design a decentralized controller for the MFDM, an approximate decoupled dynamic model of the plant is considered and is obtained using three main steps. The first step involves choosing collocated actuation and sensing points on the mirror. Hence, each measurement is maximally influenced by the input from the actuator at the same spatial location, and the influence of the other actuator inputs is minimized. In the second step, the plant model is decoupled at DC by cascading the inverse of the DC gain of the plant model, as part of the controller, with the plant. The third step involves discarding the off-diagonal entries in the system, resulting from cascading the inverse of the DC gain of the plant with the plant model. Since the accuracy of the resulting model starts to deteriorate with the increasing operating frequency, a robust controller design approach is employed to ensure the closed-loop system stability in the whole operating frequency range. The PID controller structure ensures that the MFDM can successfully track desired static or slowly time-varying surface shapes while providing a fast transient response that is acceptable from a practical point of view. This latter goal is achieved by designing the controller to minimize the l_2 norm of the mirror shape tracking error with respect to static reference signals.

For the system under consideration, the use of the DC gain matrix serves well in decoupling the different channels in the low frequency range. As indicated by the selected Bode magnitude plots shown in Fig. 7.3, the diagonal entries represent the most significant entries in the augmented system $\bar{\mathbf{G}}(s)$, whereas the contribution of the other entries is much less significant. Moreover, the diagonal entries in $\bar{\mathbf{G}}(s)$ are all similar. Hence, these diagonal entries can all be represented using the same function $\bar{g}(s)$; therefore, we represent $\bar{\mathbf{G}}(s)$ as

$$\bar{\mathbf{G}}(s) = \begin{bmatrix} \bar{g}(s) & \bar{g}_{12}(s) & \cdots & \bar{g}_{1l}(s) \\ \bar{g}_{21}(s) & \bar{g}(s) & \cdots & \bar{g}_{2l}(s) \\ \vdots & \vdots & \ddots & \vdots \\ \bar{g}_{m1}(s) & \bar{g}_{m2}(s) & \cdots & \bar{g}(s) \end{bmatrix} \tag{7.14}$$

The function $\bar{g}(s)$ can be approximated using a dynamic model given by

$$\bar{g}(s) = p(s)d(s) \tag{7.15}$$

where $p(s) = \frac{7(s+194.3)}{(s+20)(s+68)}$ approximates each of the diagonal entries in $\bar{\mathbf{G}}(s)$ as a second-order system, assuming there is no time delay, and $\mathrm{d}(s) = \frac{12-6\tau s+(\tau s)^2}{12+6\tau s+(\tau s)^2}$ is the Pade approximation of the sensor delay transfer function $\mathrm{e}^{-\tau s}$, with τ=0.0345 s.

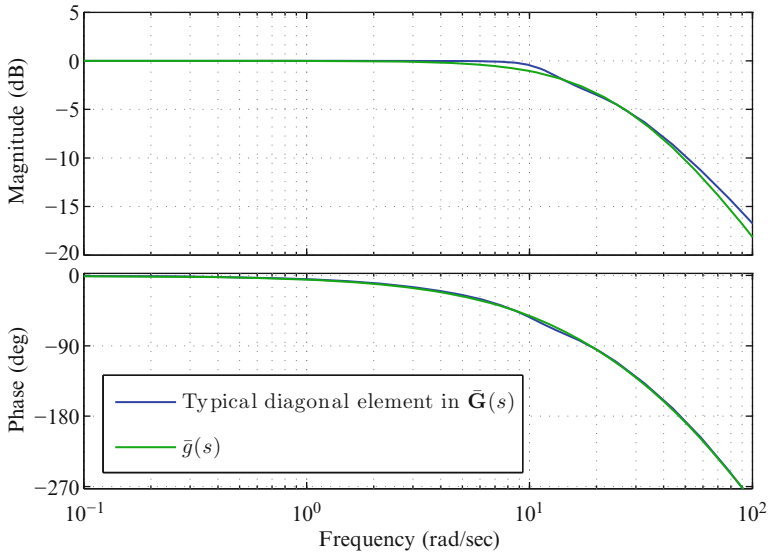


Fig. 7.6 Bode plots of $\bar{g}(s)$ and that of a typical diagonal entry in $\bar{\mathbf{G}}(s)$

Figure 7.6 shows a comparison of the Bode plots of $\bar{g}(s)$ and one of the diagonal entries in $\bar{\mathbf{G}}(s)$. The comparison indicates that $\bar{g}(s)$ provides an accurate representation of the diagonal entries in $\bar{\mathbf{G}}(s)$. The off-diagonal entries $\bar{\mathbf{G}}(s)$ in (7.14) denote the less significant uncertain dynamics in the decoupled system $\bar{\mathbf{G}}(s)$. Based on $\bar{g}(s)$ in (7.15), $\bar{\mathbf{G}}(s)$ then can be reformulated as

$$\begin{aligned}\bar{\mathbf{G}}(s) &= \bar{g}(s)\mathbf{I} + \left(\bar{\mathbf{G}}(s) - \bar{g}(s)\mathbf{I}\right) \\ &= \bar{g}(s)\mathbf{I} + \bar{\Delta}(s)\end{aligned} \tag{7.16}$$

where $\bar{\Delta}(s)$ represents the additive model uncertainty. In order to simplify the design of the PID controller, the uncertainty term $\bar{\Delta}(s)$ will be represented using a constant matrix Δ weighted by a band pass filter $w_\Delta(s)$ as

$$\bar{\Delta}(s) = w_\Delta(s)\Delta \tag{7.17}$$

where $\Delta \in \mathbb{R}^{m\times l}$ is such that all entries δ^{ij}, $i = 1,\ldots,m$, and $j = 1,\ldots,l$ in Δ satisfy $-\rho \leqslant \delta_{ij} \leqslant \rho$ and where ρ can be determined experimentally.

In the following, a robust PID controller will be developed for the augmented system (7.16). First, in order to eliminate the high-frequency noise in the real system, the filter given by $\mathbf{W}(s) = w(s)\mathbf{I}$, where $w(s)$ is a low-pass filter transfer function, will be applied to the error signal $\mathbf{e}$. By considering a decentralized PID controller of the form $\mathbf{K}_{\text{PID}}(s) = k_{\text{PID}}(s)\mathbf{I}$, it follows that the nominal closed-loop system is fully decoupled. The resulting closed-loop system is shown in Fig. 7.7.

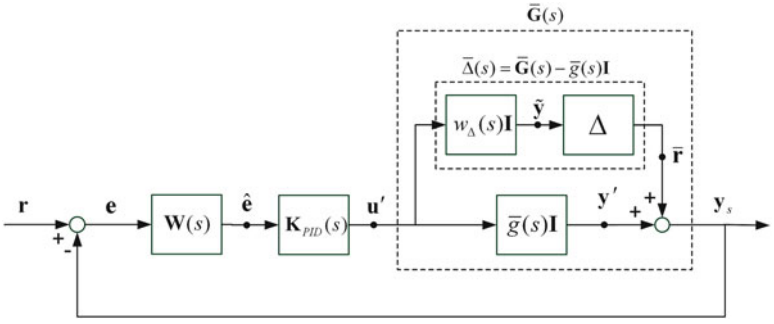


Fig. 7.7 Block diagram of the closed-loop system with the uncertain plant model

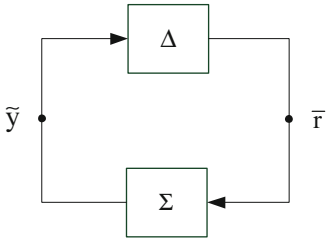


Fig. 7.8 Block diagram of the reconfigured closed-loop system showing the blocks Δ and Σ

For the purpose of stability robustness analysis, the signal $\mathbf{r}$ is set to zero first, and the closed-loop system block diagram is redrawn as shown in Fig. 7.8, where Σ is the system with input $\bar{\mathbf{r}}$ and output $\tilde{\mathbf{y}}$. The system Σ is given by

$$\Sigma(s) = -w_\Delta(s)\mathbf{K}_{\text{PID}}(s)\mathbf{W}(s)\mathbf{S}_0(s) \tag{7.18}$$

where $\mathbf{S}_0(s)$ is the sensitivity function given by

$$\begin{aligned}\mathbf{S}_0(s) &= (\mathbf{I} + \bar{g}(s)\mathbf{K}_{\text{PID}}(s)\mathbf{W}(s))^{-1} \\ &= (1 + \bar{g}(s)k_{\text{PID}}(s)w(s))^{-1}\mathbf{I}\end{aligned} \tag{7.19}$$

Let $s_0(s) = (1 + \bar{g}(s)k_{\text{PID}}(s)w(s))^{-1}$. Then, we have

$$\Sigma(s) = -w_\Delta(s)k_{\text{PID}}(s)w(s)s_0(s)\mathbf{I} \tag{7.20}$$

Assuming Σ is stable, then the robust stability of the closed-loop system shown in Fig. 7.8 is satisfied if (Zhou et al. 1995)

$$\|\Delta\|\,\|\Sigma\|_\infty < 1$$

which is equivalent to

$$\|\Sigma\|_\infty < \|\Delta\|^{-1} \tag{7.21}$$

Let $\mathbf{J} \in \mathbb{R}^{M\times L}$ be such that all its entries are equal to or less than one. Then, $\|\Delta\| \leqslant \rho\sigma_{\max}(\mathbf{J})$, where $\sigma_{\max}(\mathbf{J})$ denote the maximum singular value of $\mathbf{J}$. Based on (7.20), it follows that $\|\Sigma\|_\infty = \|w_\Delta k_{\text{PID}} w s_0\|_\infty$. Let $\gamma = 1/(\rho\sigma_{\max}(\mathbf{J}))$, then the robust stability condition of the closed-loop system can be written as

$$\|w_\Delta w k_{\text{PID}} s_0\|_\infty \leqslant \gamma \tag{7.22}$$

If (7.22) is satisfied, then the designed PID controller yields a stable closed-loop system. However, regulation performance of the closed-loop system is not considered yet if the design only involves the robustness constraint (7.22). In the following, the convergence rate of the closed-loop system with respect to the static inputs is performed by minimizing an H_2 performance specification. Consider the nominal closed-loop system subject to an input $r = v1_d$, where $v \in \mathbb{R}^m$ is a constant vector and 1_d is the scalar valued unit step function. An additional controller design objective is considered where it is desired to find a PID controller $\mathbf{K}_{\text{PID}}$ that minimizes the l_2 norm of e, that is, $\|e\|_2$, by considering only the nominal system $\bar{g}(s)I$ in the closed-loop system (i.e., without considering the uncertainty $\bar{\Delta}(s)$). Since the sensitivity function $\mathbf{S}_0$ relates the input $\mathbf{r}$ to the error $\mathbf{e}$, the solution of this design constraint can be obtained by considering a standard H_2 optimal control problem where it is desired to minimize the H_2 norm of the system $(1/s)\mathbf{S}_0$. In order to avoid the unstable pole introduced by $(1/s)$, the system $(1/s)\mathbf{S}_0$ can be approximated by the system $(1/(s+\alpha))\mathbf{S}_0$, where α is a small positive constant. Therefore, the decentralized robust PID controller design problem is converted into a mixed H_2/H_∞ multi-objective design problem as

$$\min_{k_{\text{PID}}} \left\| \frac{1}{s+\alpha} S_0 \right\|_2 \text{ subject to (7.22)} \tag{7.23}$$

In the following, the above optimization problem will be first expressed as a mixed H_2/H_∞ static output feedback problem, then a synthesis algorithm based on solving iterative linear matrix inequalities (ILMIs) is developed to solve this nonconvex problem by obtaining a suboptimal solution.

Consider the ith channel of the closed-loop system in Fig. 7.7. Write the following state-space representation for $\bar{g}(s)$

$$\sum_{\bar{g}} : \begin{cases} \dot{\mathbf{x}}_{\bar{g}} = \mathbf{A}_{\bar{g}}\mathbf{x}_{\bar{g}} + \mathbf{B}_{\bar{g}}u'_i \\ y_i = \mathbf{C}_{\bar{g}}\mathbf{x}_{\bar{g}} \end{cases} \tag{7.24}$$

where $\mathbf{x}_{\bar{g}}$ is the vector of state variables, $u'_i \in \mathbb{R}$ is the ith control input, $y_i \in \mathbb{R}$ is the ith output displacement, and $\mathbf{A}_{\bar{g}}$, $\mathbf{B}_{\bar{g}}$, and $\mathbf{C}_{\bar{g}}$ are the corresponding system matrices. A state-space representation for $w(s)$ is given as follows:

$$\sum_{w} : \begin{cases} \dot{\mathbf{x}}_w = \mathbf{A}_w\mathbf{x}_w + \mathbf{B}_w e_i \\ \hat{e}_i = \mathbf{C}_w\mathbf{x}_w \end{cases} \tag{7.25}$$

where $\mathbf{x}_w$ is the vector of state variables, $e_i = r_i - y_i \in \mathbb{R}$ is the ith performance error, $\hat{e}_i \in \mathbb{R}$ is the ith output of the filter, and $\mathbf{A}_w$, $\mathbf{B}_w$, and $\mathbf{C}_w$ are the corresponding system matrices. Similarly, the weight function $w_\Delta(s)$ and $1/(s+\alpha)$ are represented, respectively, as

$$\sum_{\Delta}: \begin{cases} \dot{\mathbf{x}}_\Delta = \mathbf{A}_\Delta \mathbf{x}_\Delta + \mathbf{B}_\Delta u'_i \\ \tilde{y}_i = \mathbf{C}_\Delta \mathbf{x}_\Delta \end{cases} \tag{7.26}$$

and

$$\sum_{s}: \begin{cases} \dot{\mathbf{x}}_s = -\alpha \mathbf{x}_s + e_i \\ \tilde{e}_i = \mathbf{x}_s \end{cases} \tag{7.27}$$

where $\mathbf{x}_\Delta$ and $\mathbf{x}_s$ are the vector of state variables and $\mathbf{A}_\Delta$, $\mathbf{B}_\Delta$, $\mathbf{C}_\Delta$, and $\mathbf{D}_\Delta$ are the corresponding system matrices. The control signal generated by the PID controller k_{PID} is given by

$$u'_i(t) = k_p \hat{e}_i(t) + k_i \int_0^t \hat{e}_i(\tau)\mathrm{d}\tau + k_d \dot{\hat{e}}_i(t)$$

where k_p, k_i, and $k_d \in \mathbb{R}$ are the parameters to be designed. Let $\boldsymbol{\chi} = \begin{bmatrix} \chi_1 & \chi_2 & \chi_3 & \chi_4 & \chi_5 \end{bmatrix}^T$, where $\chi_1 = \mathbf{x}_g$, $\chi_2 = \mathbf{x}_w$, $\chi_3 = \mathbf{x}_\Delta$, $\chi_4 = \mathbf{x}_s$, and $\chi_5(t) = \int_0^t \hat{e}_i \mathrm{d}t$. It follows that

$$\begin{aligned} \hat{e}_i &= \begin{bmatrix} \mathbf{0} & \mathbf{C}_w & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \boldsymbol{\chi} \\ \int_0^t \hat{e}_i \mathrm{d}t &= \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} \boldsymbol{\chi} \\ \dot{\hat{e}}_i &= \begin{bmatrix} -\mathbf{C}_w \mathbf{B}_w \mathbf{C}_g & \mathbf{C}_w \mathbf{A}_w & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \boldsymbol{\chi} + \mathbf{C}_w \mathbf{B}_w r_i \end{aligned}$$

Define the new output variables $\hat{\mathbf{y}} = [\hat{y}_1, \hat{y}_2, \hat{y}_3]^{\mathrm{T}}$, where $\hat{y}_1 = \hat{e}_i$, $\hat{y}_2(t) = \int_0^t \hat{e}_i \mathrm{d}\tau$, and $\hat{y}_3 = \dot{\hat{e}}_i$, and the new performance variable $\begin{bmatrix} z_2 \\ z_\infty \end{bmatrix} = \begin{bmatrix} \tilde{e}_i \\ \tilde{y}_i \end{bmatrix}$, then the state-space representation from u'_i and r_i to $\begin{bmatrix} z_2 \\ z_\infty \end{bmatrix}$ and $\hat{\mathbf{y}}$ can be written as

$$\sum_{ag}: \begin{cases} \dot{\boldsymbol{\chi}} = \boldsymbol{A}\boldsymbol{\chi} + \mathbf{B}_1 r_i + \mathbf{B}_2 u'_i \\ \begin{bmatrix} z_2 \\ z_\infty \end{bmatrix} = \begin{bmatrix} \mathbf{C}_2 \\ \mathbf{C}_\infty \end{bmatrix} \boldsymbol{\chi} \\ \hat{\mathbf{y}} = \mathbf{C}_{\hat{y}} \boldsymbol{\chi} + \mathbf{D}_{\hat{y}} r_i \end{cases} \tag{7.28}$$

where $\mathbf{A} = \begin{bmatrix} \mathbf{A}_g & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{B}_w\mathbf{C}_g & \mathbf{A}_w & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_\Delta & \mathbf{0} & \mathbf{0} \\ -\mathbf{C}_g & \mathbf{0} & \mathbf{0} & -\alpha & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_w & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$, $\mathbf{B}_1 = \begin{bmatrix} \mathbf{0} \\ \mathbf{B}_w \\ \mathbf{0} \\ \mathbf{1} \\ \mathbf{0} \end{bmatrix}$, $\mathbf{B}_2 = \begin{bmatrix} \mathbf{B}_g \\ \mathbf{0} \\ \mathbf{B}_\Delta \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}$,

$\mathbf{C}_2 = \begin{bmatrix} \mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{1}\,\mathbf{0} \end{bmatrix}$, $\mathbf{C}_\infty = \begin{bmatrix} \mathbf{0}\,\mathbf{0}\,\mathbf{C}_\Delta\,\mathbf{0}\,\mathbf{0} \end{bmatrix}$, $\mathbf{C}_{\hat{y}} = \begin{bmatrix} \mathbf{0} & \mathbf{C}_w & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} \\ -\mathbf{C}_w\mathbf{B}_w\mathbf{C}_g & \mathbf{C}_w\mathbf{A}_w & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$,

and $\mathbf{D}_{\hat{y}} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{C}_w\mathbf{B}_w \end{bmatrix}$. Therefore, with respect to the system (7.28), the design of the PID controller is converted into a static output feedback controller design problem where the control signal is given by $u_i' = \mathbf{F}\hat{\mathbf{y}}$ and where $\mathbf{F} = \begin{bmatrix} k_p & k_i & k_d \end{bmatrix}$. Therefore, the controller design problem is cast as a mixed H_2/H_∞ static output feedback controller design problem for the system (7.28) based on solving the optimization problem (7.23). Using the static output feedback controller $\mathbf{F}$, the state-space representation of the closed-loop system involving the modified plant (7.28) becomes

$$\sum_{cl}: \begin{cases} \dot{\chi} = \left(\mathbf{A} + \mathbf{B}_2\mathbf{F}\mathbf{C}_{\hat{y}}\right)\chi + \left(\mathbf{B}_1 + \mathbf{B}_2\mathbf{F}\mathbf{D}_{\hat{y}}\right) r_i \\ \begin{bmatrix} z_2 \\ z_\infty \end{bmatrix} = \begin{bmatrix} \mathbf{C}_2 \\ \mathbf{C}_\infty \end{bmatrix} \chi \end{cases} \tag{7.29}$$

It should be noted that the H_∞ norm of the closed-loop system relating $\bar{r}_i$ to z_∞ (i.e., to $\tilde{y}_i$) is the same as that of the closed-loop system Σ_{r_i,z_∞} in (7.29) relating r_i to z_∞. Also, $\left\| \frac{1}{s+\alpha} s_0 \right\|_2$ is the same as the H_2 norm of the closed-loop system Σ_{r_i,z_2} in (7.29) relating r_i to z_2 (i.e., to $\tilde{e}_i$). Consequently, the design of the PID controller presented in the following is based on considering $\|\Sigma_{r_i,z_\infty}\|_\infty$ and $\|\Sigma_{r_i,z_2}\|_2$. Hence, the mixed H_2/H_∞ static output feedback problem can then be cast in the form of an optimization problem involving matrix inequalities as

$$\text{OP1}: \quad \min_{k_{\text{PID}}} trace(\mathbf{C}_2\tilde{\mathbf{P}}\mathbf{C}_2^{\mathrm{T}})$$

$$\text{subject to} \begin{bmatrix} \left(\mathbf{A} + \mathbf{B}_2\mathbf{F}\mathbf{C}_{\hat{y}}\right)\tilde{\mathbf{P}} + \tilde{\mathbf{P}}\left(\mathbf{A} + \mathbf{B}_2\mathbf{F}\mathbf{C}_{\hat{y}}\right)^{\mathrm{T}} & \left(\mathbf{B}_1 + \mathbf{B}_2\mathbf{F}\mathbf{D}_{\hat{y}}\right) \\ \left(\mathbf{B}_1 + \mathbf{B}_2\mathbf{F}\mathbf{D}_{\hat{y}}\right)^{\mathrm{T}} & -\mathbf{I} \end{bmatrix} < 0 \tag{7.30}$$

$$\begin{bmatrix} \left(\mathbf{A} + \mathbf{B}_2\mathbf{F}\mathbf{C}_{\hat{y}}\right)\mathbf{P} + \mathbf{P}\left(\mathbf{A} + \mathbf{B}_2\mathbf{F}\mathbf{C}_{\hat{y}}\right)^{\mathrm{T}} & \left(\mathbf{B}_1 + \mathbf{B}_2\mathbf{F}\mathbf{D}_{\hat{y}}\right) & \mathbf{P}\mathbf{C}_\infty^{\mathrm{T}} \\ \left(\mathbf{B}_1 + \mathbf{B}_2\mathbf{F}\mathbf{D}_{\hat{y}}\right)^{\mathrm{T}} & -\mathbf{I} & \mathbf{0} \\ \mathbf{C}_\infty\mathbf{P} & \mathbf{0} & -\gamma^2\mathbf{I} \end{bmatrix} < 0 \tag{7.31}$$

where $\tilde{\mathbf{P}} > 0$ and $\mathbf{P} > 0$. Since the above matrix inequalities are nonlinear in the unknown parameters $\tilde{\mathbf{P}}$, $\mathbf{P}$ and $\mathbf{F}$, they cannot be solved directly using LMI solvers. In the following, an algorithm is proposed to obtain a solution for **OP1** by iteratively solving a set of properly formulated linear matrix inequalities. First, using Schur complement, inequality (7.30) can be written as (Boyd et al. 1994)

$$\left(\mathbf{A}+\mathbf{B}_2\mathbf{F}\mathbf{C}_{\hat{y}}\right)\tilde{\mathbf{P}}+\tilde{\mathbf{P}}\left(\mathbf{A}+\mathbf{B}_2\mathbf{F}\mathbf{C}_{\hat{y}}\right)^{\mathrm{T}}+\left(\mathbf{B}_1+\mathbf{B}_2\mathbf{F}\mathbf{D}_{\hat{y}}\right)\left(\mathbf{B}_1+\mathbf{B}_2\mathbf{F}\mathbf{D}_{\hat{y}}\right)^{\mathrm{T}}<0 \tag{7.32}$$

which is equivalent to

$$\begin{aligned}\mathbf{A}\tilde{\mathbf{P}}+\tilde{\mathbf{P}}\mathbf{A}^{\mathrm{T}}-\tilde{\mathbf{P}}\mathbf{C}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\tilde{\mathbf{P}}+\left(\mathbf{B}_2\mathbf{F}+\tilde{\mathbf{P}}\mathbf{C}_{\hat{y}}^{\mathrm{T}}\right)\left(\mathbf{B}_2\mathbf{F}+\tilde{\mathbf{P}}\mathbf{C}_{\hat{y}}^{\mathrm{T}}\right)^{\mathrm{T}}\\ +\left(\mathbf{B}_1+\mathbf{B}_2\mathbf{F}\mathbf{D}_{\hat{y}}\right)\left(\mathbf{B}_1+\mathbf{B}_2\mathbf{F}\mathbf{D}_{\hat{y}}\right)^{\mathrm{T}}<0\end{aligned} \tag{7.33}$$

Since $\left(\tilde{\mathbf{X}}-\tilde{\mathbf{P}}\right)^{\mathrm{T}}\mathbf{C}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\left(\tilde{\mathbf{X}}-\tilde{\mathbf{P}}\right)\leqslant 0$, then we have

$$\mathbf{P}\mathbf{C}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\mathbf{P}\geqslant\tilde{\mathbf{X}}\mathbf{C}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\mathbf{P}+\mathbf{P}\mathbf{C}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\tilde{\mathbf{X}}-\tilde{\mathbf{X}}\mathbf{C}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\tilde{\mathbf{X}}$$

Then, a sufficient condition for (7.33) to be satisfied can be obtained with $\alpha\leqslant 0$ as

$$\begin{aligned}&\mathbf{A}\tilde{\mathbf{P}}+\tilde{\mathbf{P}}\mathbf{A}^{\mathrm{T}}-\alpha\tilde{\mathbf{P}}-\tilde{\mathbf{X}}\mathbf{C}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\tilde{\mathbf{P}}-\tilde{\mathbf{P}}\mathbf{C}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\tilde{\mathbf{X}}+\tilde{\mathbf{X}}\mathbf{C}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\tilde{\mathbf{X}}\\ &+\left(\mathbf{B}_2\mathbf{F}+\tilde{\mathbf{P}}\mathbf{C}_{\hat{y}}^{\mathrm{T}}\right)\left(\mathbf{B}_2\mathbf{F}+\tilde{\mathbf{P}}\mathbf{C}_{\hat{y}}^{\mathrm{T}}\right)^{\mathrm{T}}+\left(\mathbf{B}_1+\mathbf{B}_2\mathbf{F}\mathbf{D}_{\hat{y}}\right)\left(\mathbf{B}_1+\mathbf{B}_2\mathbf{F}\mathbf{D}_{\hat{y}}\right)^{\mathrm{T}}<0\end{aligned} \tag{7.34}$$

Let $\Psi=\mathbf{A}\tilde{\mathbf{P}}+\tilde{\mathbf{P}}\mathbf{A}^{\mathrm{T}}-\alpha\tilde{\mathbf{P}}-\tilde{\mathbf{X}}\mathbf{C}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\tilde{\mathbf{P}}-\tilde{\mathbf{P}}\mathbf{C}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\tilde{\mathbf{X}}+\tilde{\mathbf{X}}\mathbf{C}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\tilde{\mathbf{X}}$. Then, (7.34) is equivalent to

$$\begin{bmatrix}\Psi & \left(\mathbf{B}_1+\mathbf{B}_2\mathbf{F}\mathbf{D}_{\hat{y}}\right) & \left(\mathbf{B}_2\mathbf{F}+\tilde{\mathbf{P}}\mathbf{C}_{\hat{y}}^{\mathrm{T}}\right)\\ \left(\mathbf{B}_1+\mathbf{B}_2\mathbf{F}\mathbf{D}_{\hat{y}}\right)^{\mathrm{T}} & -\mathbf{I} & \mathbf{0}\\ \left(\mathbf{B}_2\mathbf{F}+\tilde{\mathbf{P}}\mathbf{C}_{\hat{y}}^{\mathrm{T}}\right)^{\mathrm{T}} & \mathbf{0} & -\mathbf{I}\end{bmatrix}<0 \tag{7.35}$$

Similarly, a sufficient condition for (7.31) can be obtained as

$$\begin{bmatrix}\Omega & \left(\mathbf{B}_1+\mathbf{B}_2\mathbf{F}\mathbf{D}_{\hat{y}}\right) & \mathbf{B}_2\mathbf{F}+\mathbf{P}\mathbf{C}_{\hat{y}}^{\mathrm{T}} & \mathbf{P}\mathbf{C}_{\infty}^{\mathrm{T}}\\ \left(\mathbf{B}_1+\mathbf{B}_2\mathbf{F}\mathbf{D}_{\hat{y}}\right)^{\mathrm{T}} & -\mathbf{I} & \mathbf{0} & \mathbf{0}\\ \left(\mathbf{B}_2\mathbf{F}+\mathbf{P}\mathbf{C}_{\hat{y}}^{\mathrm{T}}\right)^{\mathrm{T}} & \mathbf{0} & -\mathbf{I} & \mathbf{0}\\ \mathbf{C}_{\infty}\mathbf{P} & \mathbf{0} & \mathbf{0} & -\gamma^2\mathbf{I}\end{bmatrix}<0$$

where $\Omega = \mathbf{AP} + \mathbf{PA}^{\mathrm{T}} - \alpha\mathbf{P} - \mathbf{XC}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\mathbf{P} - \mathbf{PC}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\mathbf{X} + \mathbf{XC}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\mathbf{X}$. In the following, an iterative LMI algorithm will be used to obtain a suboptimal solution for the optimization problem **OP1**. A similar algorithm has been used in (Cao et al. 1998) to solve a static output feedback stability problem. The main idea behind the following algorithm is to first find a set of gains $\mathbf{F}$ such that the robust stability condition (7.31) is satisfied, and then to further search within this set of gains for a solution that minimizes the H_2 performance.

Algorithm

Step 1: Select $\tilde{\mathbf{Q}} > 0$, $\mathbf{Q} > 0$, and solve for $\tilde{\mathbf{P}} > 0$ and $\mathbf{P} > 0$ from the following algebraic Riccati equation:

$$\tilde{\mathbf{P}}\mathbf{A}^{\mathrm{T}} + \mathbf{A}\tilde{\mathbf{P}} - \tilde{\mathbf{P}}\mathbf{C}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\tilde{\mathbf{P}} + \tilde{\mathbf{Q}} = 0$$

$$\mathbf{PA}^{\mathrm{T}} + \mathbf{AP} - \mathbf{PC}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\mathbf{P} + \mathbf{Q} = 0$$

Set $i = 1$, $\tilde{\mathbf{X}}_i = \tilde{\mathbf{P}}$, $\mathbf{X}_i = \mathbf{P}$, and an initial H_2 performance value γ_2. Let

$$\Psi_i = \mathbf{A}\tilde{\mathbf{P}}_i + \tilde{\mathbf{P}}_i\mathbf{A}^{\mathrm{T}} - \alpha\tilde{\mathbf{P}}_i - \tilde{\mathbf{X}}_i\mathbf{C}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\tilde{\mathbf{P}}_i - \tilde{\mathbf{P}}_i\mathbf{C}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\tilde{\mathbf{X}}_i + \tilde{\mathbf{X}}_i\mathbf{C}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\tilde{\mathbf{X}}_i$$

and

$$\Omega_i = \mathbf{AP}_i + \mathbf{P}_i\mathbf{A}^{\mathrm{T}} - \alpha\mathbf{P}_i - \mathbf{X}_i\mathbf{C}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\mathbf{P}_i - \mathbf{P}_i\mathbf{C}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\mathbf{X}_i + \mathbf{X}_i\mathbf{C}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\mathbf{X}_i$$

Step 2: Minimize α_i subject to the following LMI constraints for unknown $\tilde{\mathbf{P}}_i > 0$, $\mathbf{P}_i > 0$, $\mathbf{F}$, and α_i

$$\begin{bmatrix} \Psi_i & (\mathbf{B}_1 + \mathbf{B}_2\mathbf{F}\mathbf{D}_{\hat{y}}) & \left(\mathbf{B}_2\mathbf{F} + \tilde{\mathbf{P}}_i\mathbf{C}_{\hat{y}}^{\mathrm{T}}\right) \\ (\mathbf{B}_1 + \mathbf{B}_2\mathbf{F}\mathbf{D}_{\hat{y}})^{\mathrm{T}} & -\mathbf{I} & \mathbf{0} \\ \left(\mathbf{B}_2\mathbf{F} + \tilde{\mathbf{P}}_i\mathbf{C}_{\hat{y}}^{\mathrm{T}}\right)^{\mathrm{T}} & \mathbf{0} & -\mathbf{I} \end{bmatrix} < 0 \tag{7.36}$$

$$\begin{bmatrix} \Omega_i & (\mathbf{B}_1 + \mathbf{B}_2\mathbf{F}\mathbf{D}_{\hat{y}}) & \mathbf{B}_2\mathbf{F} + \mathbf{P}_i\mathbf{C}_{\hat{y}}^{\mathrm{T}} & \mathbf{P}_i\mathbf{C}_{\infty}^{\mathrm{T}} \\ (\mathbf{B}_1 + \mathbf{B}_2\mathbf{F}\mathbf{D}_{\hat{y}})^{\mathrm{T}} & -\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \left(\mathbf{B}_2\mathbf{F} + \mathbf{P}_i\mathbf{C}_{\hat{y}}^{\mathrm{T}}\right)^{\mathrm{T}} & \mathbf{0} & -\mathbf{I} & \mathbf{0} \\ \mathbf{C}_{\infty}\mathbf{P}_i & \mathbf{0} & \mathbf{0} & -\gamma^2\mathbf{I} \end{bmatrix} < 0 \tag{7.37}$$

$$\mathbf{C}_2\tilde{\mathbf{P}}_i\mathbf{C}_2^{\mathrm{T}} < \gamma_2 \tag{7.38}$$

Step 3: If $\alpha_i \leqslant 0$, let

$$\Psi_i = \mathbf{A}\tilde{\mathbf{P}}_i + \tilde{\mathbf{P}}_i\mathbf{A}^{\mathrm{T}} - \tilde{\mathbf{X}}_i\mathbf{C}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\tilde{\mathbf{P}}_i - \tilde{\mathbf{P}}_i\mathbf{C}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\tilde{\mathbf{X}}_i + \tilde{\mathbf{X}}_i\mathbf{C}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\tilde{\mathbf{X}}_i$$

$$\Omega_i = \mathbf{AP}_i + \mathbf{P}_i\mathbf{A}^{\mathrm{T}} - \mathbf{X}_i\mathbf{C}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\mathbf{P}_i - \mathbf{P}_i\mathbf{C}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\mathbf{X}_i + \mathbf{X}_i\mathbf{C}_{\hat{y}}^{\mathrm{T}}\mathbf{C}_{\hat{y}}\mathbf{X}_i$$

and $\beta_i = trace(\mathbf{C}_2\tilde{\mathbf{P}}_i\mathbf{C}_2^{\mathrm{T}})$, then go to step 5.

Step 4: If $|a_{i-1} - a_i| < \delta$, a prescribed tolerance, the optimal problem **OP1** may not be solved by $\mathbf{F}$, or try different $\tilde{\mathbf{Q}}$ and $\mathbf{Q}$ in step 1 and run the algorithm again. Otherwise, minimize $trace\left(\begin{bmatrix} \tilde{\mathbf{P}}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_i \end{bmatrix}\right)$ subject to (7.36), (7.37), and (7.38) for the unknowns $\tilde{\mathbf{P}}_i$, $\mathbf{P}_i$, and $\mathbf{F}$. Set $i = i + 1$, $\tilde{\mathbf{X}}_i = \tilde{\mathbf{P}}_{i-1}$, and $\mathbf{X}_i = \mathbf{P}_{i-1}$, then go to step 2.

Step 5: Solve min β_i subject to (7.36), (7.37), and (7.38) for unknown $\tilde{\mathbf{P}}_i$, $\mathbf{P}_i$, and $\mathbf{F}$. If $|\beta_{i-1} - \beta_i| < \varepsilon$, a prescribed tolerance, stop.

Step 6: Minimize $trace(\mathbf{P}_i)$ subject to (7.36), (7.37), and (7.38) for the unknowns $\mathbf{P}_i$ and $\mathbf{F}$. Set $i = i + 1$, $\tilde{\mathbf{X}}_i = \tilde{\mathbf{P}}_{i-1}$, and $\mathbf{X}_i = \mathbf{P}_{i-1}$, then go to step 5.

7.3.2 *Simulation of the Closed-Loop System*

Using the algorithm described above, a controller is designed for the system $\bar{\mathbf{G}}$ in (7.14). In the controller design process, the low-pass filter $w(s)$ is selected as

$$w(s) = \frac{1}{6.944 \times 10^{-3} s^2 + 1.179 \times 10^{-1} s + 1} \tag{7.39}$$

For the 19-input–19-output system, based on Bode plot analysis of the decoupled system, we have the uncertainty bound $\rho \leqslant 0.2$, and each output is mostly affected by the neighboring actuators. Based on the data, we obtain $\|\Delta\|_\infty \leqslant 3.8$ and the uncertain weight function $w_\Delta(s)$ as

$$w_\Delta(s) = \frac{s}{s + 10} \tag{7.40}$$

Using these parameters, the proposed iterative algorithm provides the following PID controller:

$$k_p = 0.30, \quad k_i = 0.26, \quad k_d = 0.65 \tag{7.41}$$

Using the above parameters, the controller k_{PID} for each channel can be obtained as

$$k_{\text{PID}}(s) = k_p + k_i \frac{1}{s} + k_d s \tag{7.42}$$

which gives

$$\bar{\mathbf{K}}(s) = \mathbf{K}_{\text{PID}}(s) = k_{\text{PID}}(s) I \tag{7.43}$$

Using (7.43) in (7.7), the overall controller $\mathbf{K}(s)$ can be obtained as

$$\mathbf{K}(s) = k_{\text{PID}}(s) \mathbf{G}_0^{-1} \tag{7.44}$$

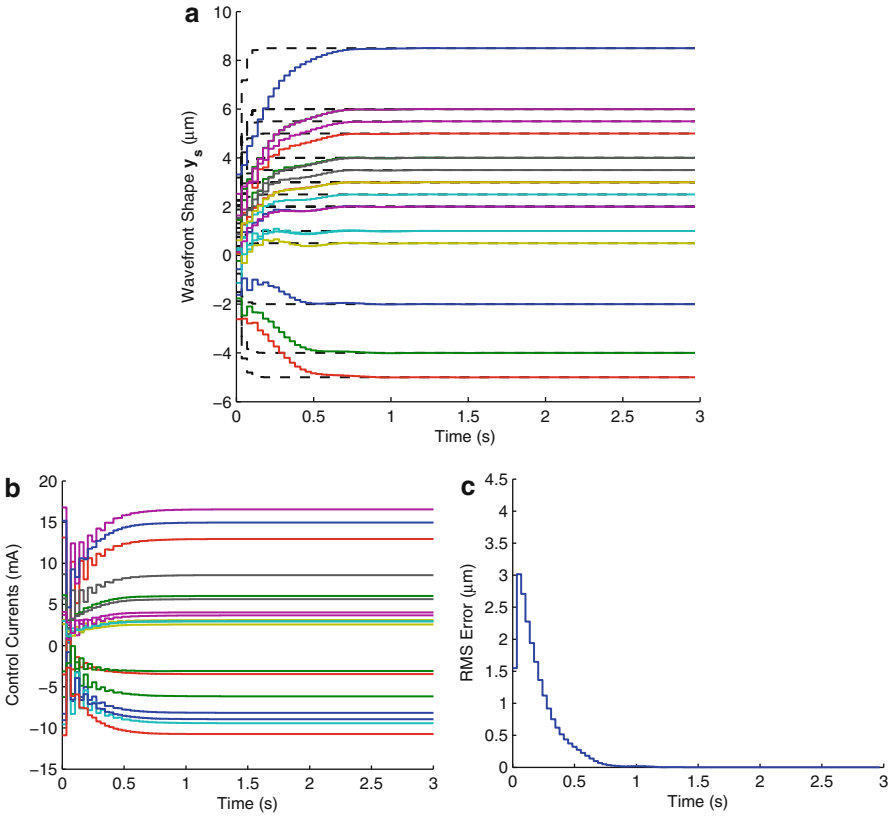


Fig. 7.9 Tracking of a static reference shape using decentralized robust PID controller: (**a**) wavefront shape, (**b**) control inputs, (**c**) RMS error

A closed-loop system comprising the augmented plant model **G** and the decentralized PID controller (7.44) was simulated in order to determine the ability of the closed-loop system to track the reference wavefront shapes signals. The results of the simulations are as follows:

Figure 7.9 shows the results of simulations where the closed-loop system is required to track the static shape defined by (7.12). Figure 7.9a shows the time history of the wavefront shape measurements $\mathbf{y}_s$ at the 19 output locations. The corresponding reference wavefront shape signals are shown as dashed lines. The control inputs as determined by the designed controller are given in Fig. 7.9b. The RMS of the wavefront shape error **e** computed for the 19 channels is given in Figure 7.9c. The output of the closed-loop system was found to converge successfully to the desired reference shape in less than 1s. Note that the RMS error asymptotically approaches zero, which is the characteristic feature of controllers with an integrator.

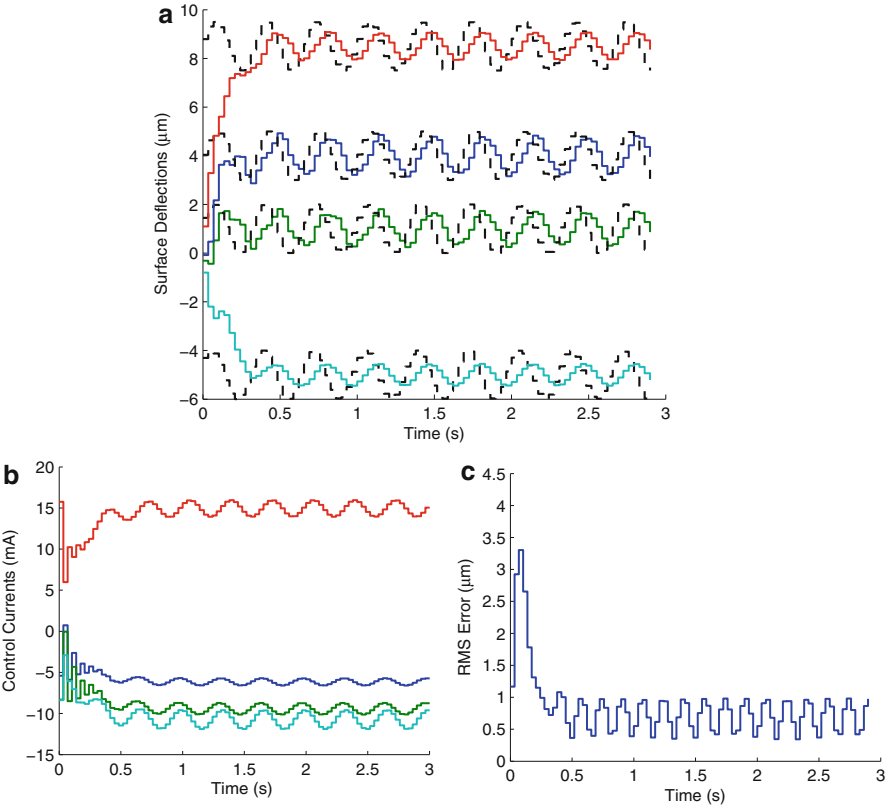


Fig. 7.10 Tracking of a dynamic reference shape using decentralized robust PID controller: (**a**) wavefront shape, (**b**) control inputs, (**c**) RMS error

Figure 7.10 shows the results of simulations where the closed-loop system is required to track the dynamic reference wavefront shape signals specified in (7.13). Figure 7.10a shows the time history of the wavefront shape measurements $\mathbf{y}_s$ at selected locations (# 2, 4, 8, and 17). The control inputs are given in Figure 7.10b. As can be observed from the figure, the controller manages to drive the wavefront shape $\mathbf{y}_s$ to the static offset values of the reference signals. However, there exists a phase difference between the reference wavefront shape signals and the corresponding wavefront shape measurements $\mathbf{y}_s$, which results in a relatively large RMS error as shown in Fig. 7.10c.

The closed-loop system based on the PID controller described above performs satisfactorily in tracking slowly changing reference shapes. The integrator in the structure of this type of controllers ensures that the error in tracking a constant reference shape is driven to zero asymptotically. When tracking a dynamic reference wavefront shape, the output of the closed-loop system lags the reference wavefront shape signal, resulting in significant tracking errors.

7.4 Experimental Evaluation

This section presents the results of the experimental work carried out to evaluate the performance of the controllers presented in Sects. 7.2 and 7.3. Where possible, a comparison to the numerical simulations is also presented.

7.4.1 Decentralized PI Controller

In order to implement the decentralized PI controller as presented in Sect. 7.2, the closed-loop system shown in Fig. 7.2 is constructed such that the augmented plant **G** is the actual system comprising the MFDM and the sensor. The sampling time in the closed-loop system is of 0.0345 s. In order to cancel the sensor noise, an extra low-pass filter f is chosen as follows:

$$f(z) = \frac{0.3757(z+1)}{z-0.2487} \tag{7.45}$$

which is the transfer function for a Butterworth filter with 5 Hz cutoff frequency. The controller is given by $\mathbf{K} = k\mathbf{G}_0^{-1}$, where $\mathbf{G}_0^{-1}$ is obtained from (7.3) and where the parameters in k are obtained using the classical pole placement method. To evaluate the performance of the decentralized PI controller, aberrated wavefront shapes are specified as reference signals $\mathbf{r}$ where each entry in $\mathbf{r}$ is associated with a spatial location corresponding to the center of one of the 19 coils in the prototype MFDM described in the preceding section.

7.4.1.1 Tracking of a Generalized Reference Wavefront Shape

Firstly, a generalized static reference wavefront shape defined by the following desired displacements is considered:

$$\begin{aligned}\mathbf{r} = [&2.0, 4.0, 5.0, 1.0, 5.5, 0.5, 3.5, 8.5, 6.0, 3.0, 2.5, 2.0, 3.0, 4.0, -2.0,\\ &-4.0, -5.0, 1.0, 6.0]^{\mathrm{T}}(\mu\mathrm{m})\end{aligned} \tag{7.46}$$

With the closed-loop system setup as described above, the static reference wavefront shape signals specified in (7.46) are tracked using the developed closed-loop AO system. Figure 7.11a shows the time history of the experimentally obtained wavefront shape measurements $\mathbf{y}_s$, where only selected channels (# 2, 4, 8, and 17) are shown. The corresponding reference signals are shown as dotted lines. The RMS of the wavefront shape error $\mathbf{e}$ is given in Fig. 7.11b where the corresponding simulation results are also shown for comparison. While the simulations showed that the controller can drive the RMS error to zero, the experimental results show a residual steady-state error of 0.21 μm.

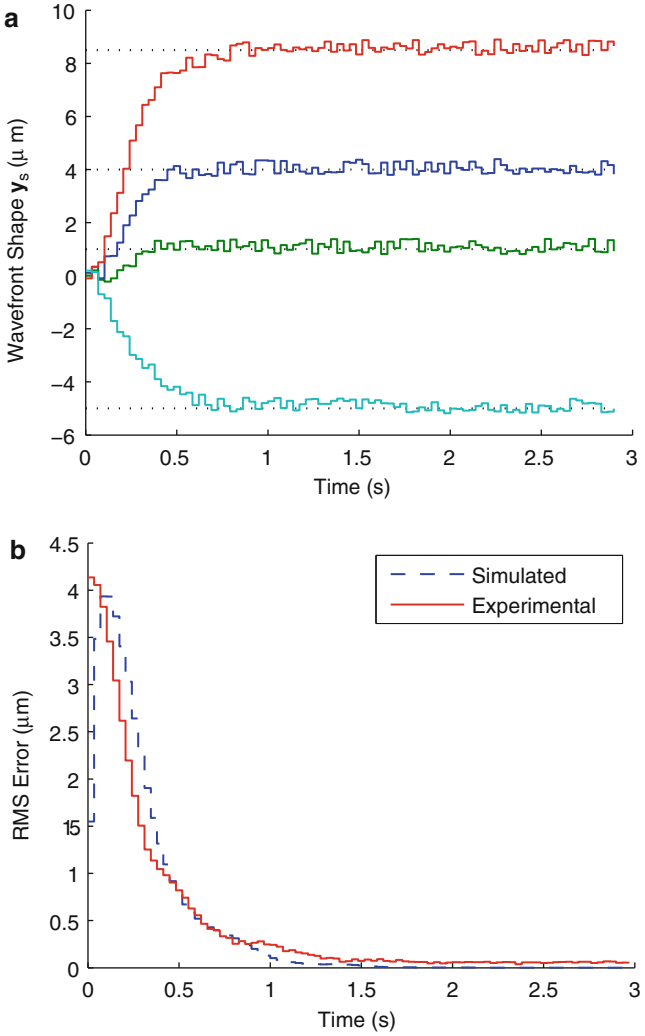


Fig. 7.11 Experimental tracking of a static reference shape using the PI controller: (**a**) wavefront shape, (**b**) RMS error

The optical benefit of compensating for a wavefront error can be visualized by observing its effect on the point spread function (PSF) of the system of the wavefront computed before and after the correction has been applied. The comparison for the above-mentioned case of correction of the static reference shape is shown in Fig. 7.12. Figure 7.12a represents the image of a point source of light formed by an imaging system, which carries wavefront aberrations equivalent to the wavefront shape error defined by this static reference shape. Figure 7.12b represents the

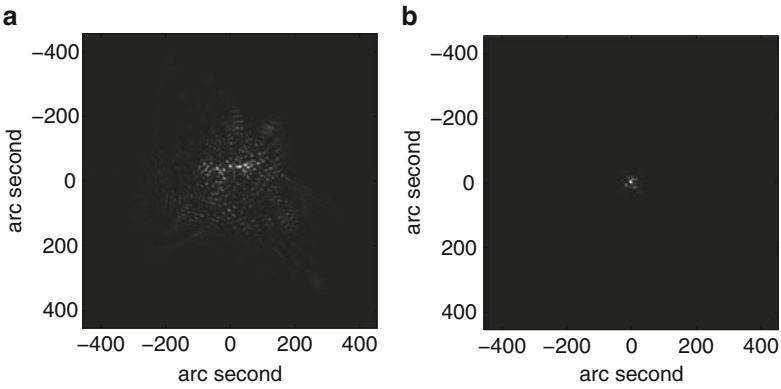


Fig. 7.12 The point spread function of the static wavefront error before and after applying the AO correction: (**a**) before correction, (**b**) after correction

image of a point source of light formed by an imaging system, which carries wavefront aberrations equivalent to the residual wavefront error measured after the AO correction had been applied. The results show that a significant improvement in the image quality can be achieved using the closed-loop system comprising the presented control algorithm and a prototype MFDM with only 19 actuators.

7.4.1.2 Tracking of Zernike Mode Shapes

In the following, it will be shown that, using the decentralized PI controllers presented earlier, a wide range of static shapes of the MFDM can be achieved. It follows from the linearity of the MFDM response that any combination of these surface shapes can also be generated. The Zernike modes present a complete set of basis functions that can be used to represent any desired shape using a linear combination of these shapes. The number of Zernike modes that the mirror surface can produce is a good estimate of the spatial resolution of the wavefront correction offered by the mirror. Therefore, the MFDM was tested for its ability to generate the series of Zernike mode shapes. A simulated signal generator as shown in Fig. 7.13 was used to provide the reference mode shapes or a combination of these shapes to be tracked by the mirror. The reference shape vector $\mathbf{r}$ can be written as

$$\mathbf{r} = \mathbf{Z}\mathbf{c} \tag{7.47}$$

where $\mathbf{c}$ is the vector of Zernike coefficients and $\mathbf{Z}$ is the matrix of Zernike mode shapes computed at the 19 output points coinciding with the location of electromagnetic coils in the optical pupil. The mode shape matrix $\mathbf{Z}$ is written as

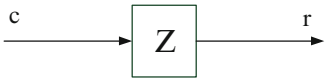


Fig. 7.13 Reference signal generator with Zernike mode shapes

$$\mathbf{Z} = \begin{bmatrix} Z_{11} & Z_{12} & \cdots & Z_{1j} & \cdots & Z_{1J} \\ Z_{21} & Z_{22} & \cdots & Z_{2j} & \cdots & Z_{2J} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ Z_{m1} & Z_{m2} & \cdots & Z_{mj} & \cdots & Z_{mJ} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ Z_{M1} & Z_{M2} & \cdots & Z_{Mj} & \cdots & Z_{MJ} \end{bmatrix}$$

where Z_{mj}, $j = 1, 2, \ldots, J$, $m = 1, 2, \ldots, M$ is the jth Zernike mode shape computed at (r_m, θ_m) output location. The Zernike coefficients determine the magnitude of the resulting reference shape. Note that individual Zernike mode shapes can be set as the reference wavefront shape signals by setting all coefficients, other than that of the desired mode, to zero. Also, by introducing time-varying coefficients, dynamic wavefront shapes can be provided as reference signals.

The closed-loop system with the decentralized PI controller presented in (7.10) is used to track the desired mode shapes. The magnitude of the mode shape to be tracked is specified by a 10-μm coefficient (±10 μm peak-to-valley). Figure 7.14 shows some sampled results of the total 32 Zernike mode shapes that were attained by the MFDM. For each Zernike mode Z_j, $j = 1, 2, \ldots, 35$, the contour plot on the LHS is the experimentally obtained wavefront shape. The one in the middle is the exact Zernike mode shape provided as the reference signal. The shape on the RHS shows the wavefront shape error. Note that the contours are spaced according to the wavelength ($\lambda = 0.661$ μm) of the laser light used. While the graphical illustration does show effectiveness of the MFDM in tracking the specified mode shapes, a more comprehensive measure is the RMS of the wavefront shape error as shown by the bar chart in Fig. 7.15. The initial RMS error defined by a 10-μm modal coefficient of the reference shape is reduced to the magnitudes shown in the figure. The average RMS error is 0.11 μm $\left(\approx \frac{\lambda}{6}\right)$. The successful tracking of up to 32 Zernike mode shapes using only a 19-channel MFDM demonstrates that higher-order wavefront aberrations can be effectively canceled out using this WFC.

7.4.1.3 Tracking of Dynamic Shapes

The experiments conducted with the dynamic reference wavefront shape signals (7.13) show that the PI controller can track very slowly changing signals only. With the increasing frequency of the reference signals, the error increases rapidly and results in the saturation of the actuator electromagnetic coils.

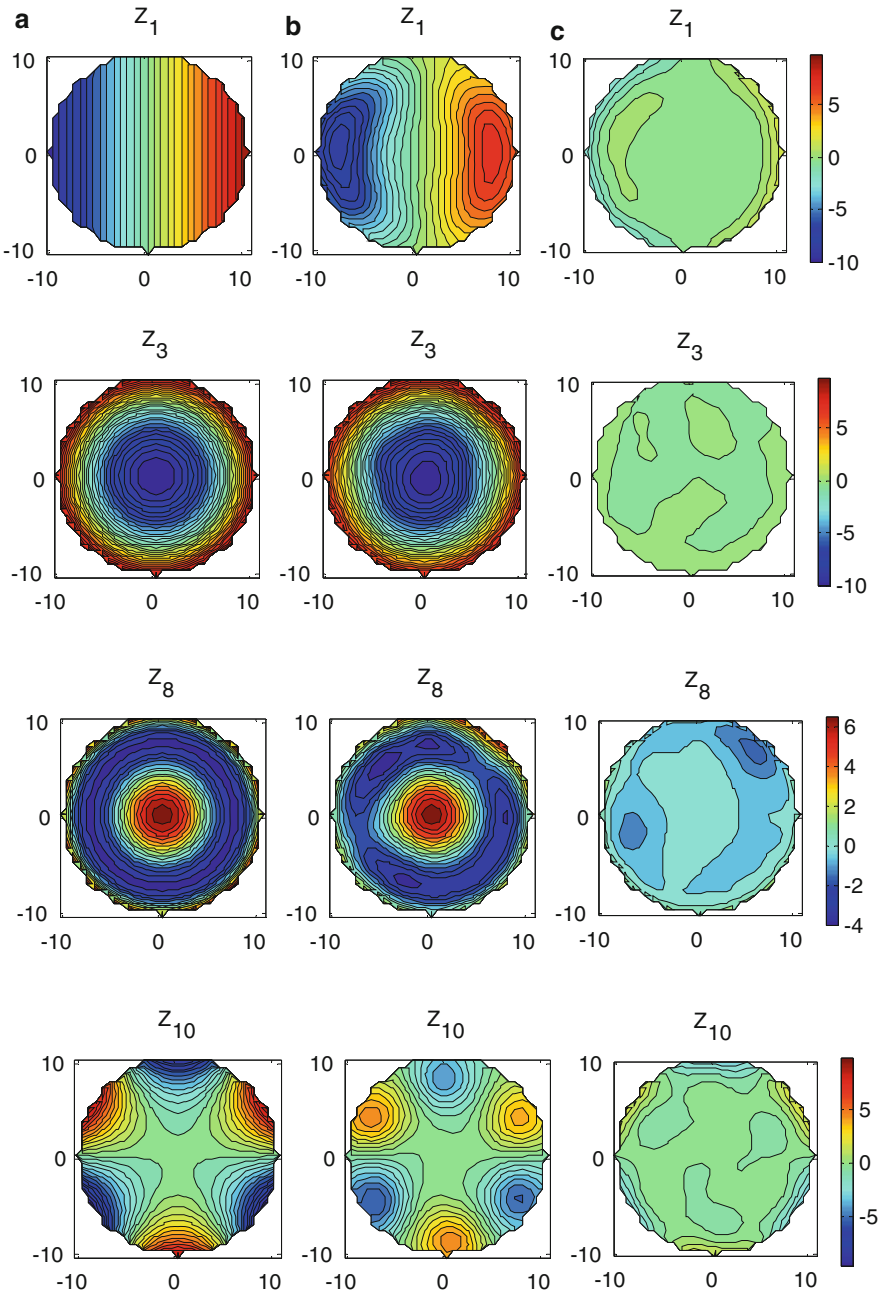


Fig. 7.14 Zernike mode shapes generated using the MFDM with decentralized PI controller: (**a**) desired shapes, (**b**) acquired shapes, (**c**) error surface

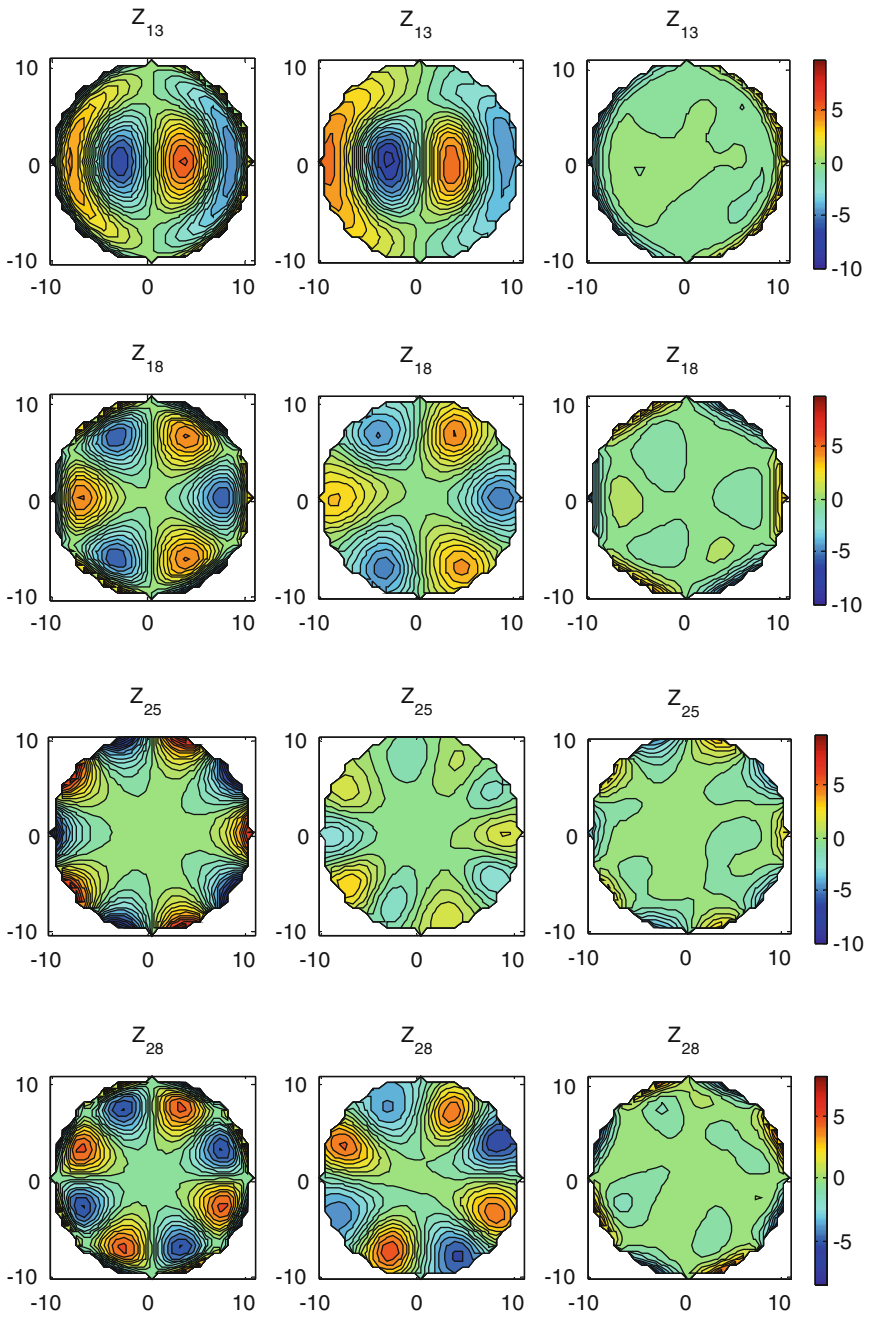


Fig. 7.14 (continued)

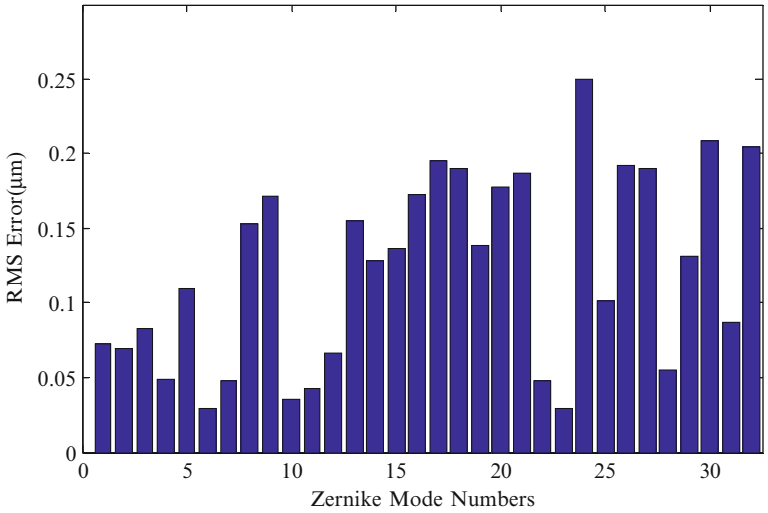


Fig. 7.15 RMS error for the Zernike mode shapes generated using the MFDM

7.4.2 Decentralized Robust PID Controller

The decentralized robust PID controller is implemented by setting up a closed-loop system as was shown in Fig. 7.2, where the plant **G** is the actual system comprising the MFDM and the sensor. The controller given in (7.44) is used with the PID parameters as given in (7.41). The error signal **e** is filtered using the discretized version of the low-pass filter $w(s)$ in (7.39). The resulting PID controller was discretized using Tustin's approximation algorithm.

7.4.2.1 Tracking of a Generalized Reference Wavefront Shape

The performance of the PID controller is tested by tracking a surface shape formed by the following set of constant deflections:

$$\mathbf{r} = \mathbf{r}_o = [2.0, 4.0, 5.0, 1.0, 5.5, 0.5, 3.5, 8.5, 6.0, 3.0, 2.5, 2.0, 3.0, 4.0, -2.0, \\ -4.0, -5.0, 1.0, 6.0]^{\mathrm{T}}(\mu\mathrm{m})$$

which forms a generalized aspheric surface shape. Figure 7.16a shows the time history of the mirror surface deflections measured at the 19 points coinciding with the location of the coils, where only selected channels (# 2, 4, 8, and 17) are shown. The RMS error in tracking the reference shape is also given in Fig. 7.16b. The experimental results show a residual steady-state error of 0.18 μm. As a comparison with the performance of the control system presented in Sect. 7.4.1, it should be noted that with respect to the reference signal given by r_o, the RMS error values

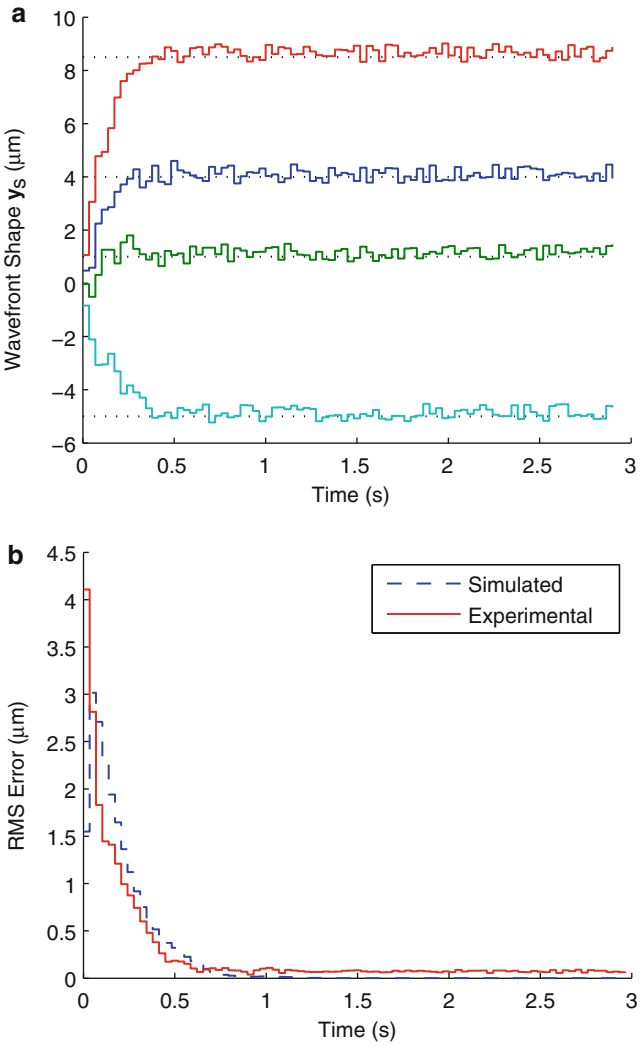


Fig. 7.16 Experimental tracking of a static reference shape using the decentralized robust PID controller: (**a**) wavefront shape, (**b**) RMS error

were 0.21 μm using the decentralized PI controller. Hence, it can be seen that with an RMS error value of 0.18 μm, the proposed controller provides a superior performance to the controllers presented in Sect. 7.4.1. Figure 7.17 shows the contour plot of the error surface, that is, the difference of the reference surface shape and the measured wavefront shape. For clarification, only the reference surface obtained by fitting the first 15 Zernike polynomials on to the discrete reference displacements r_o is shown here. It can be seen that the decentralized robust PID controller is superior to the decentralized PI controller.

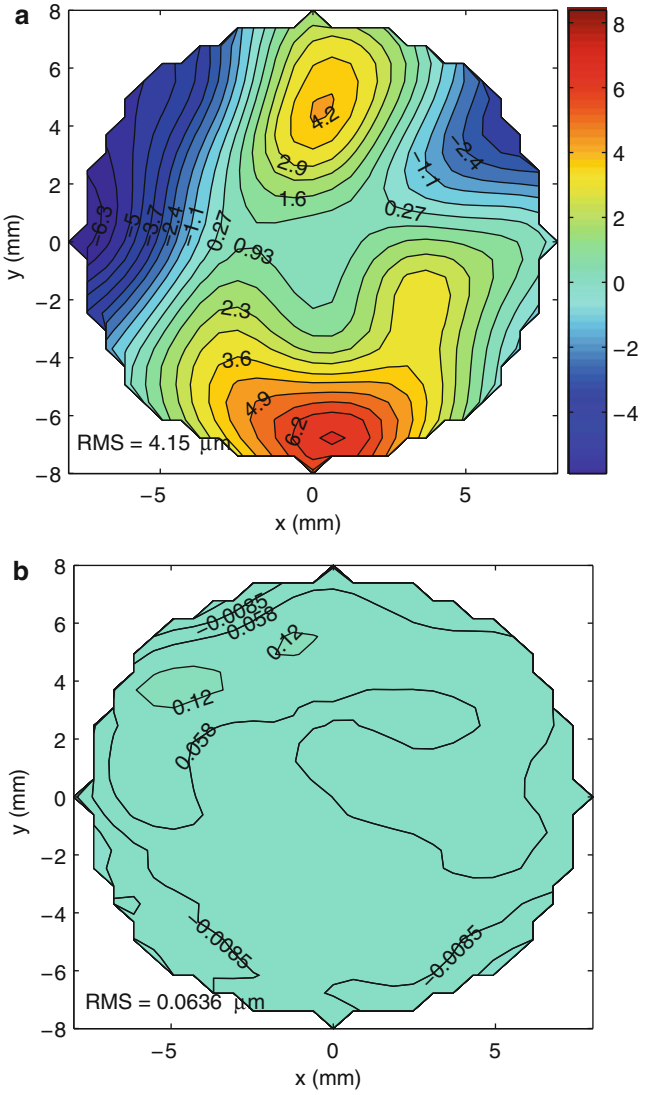


Fig. 7.17 Contour plot of the error surface: (**a**) before correction, (**b**) after correction

7.4.2.2 Tracking of Zernike Mode Shapes

The Zernike modes present a complete set of basis functions that can be used to represent any desired surface shape using a linear combination of these modes. The number of Zernike modes that the mirror surface can produce is a good estimate of the spatial resolution of the wavefront correction offered by the mirror. Therefore, the closed-loop system comprising the proposed controller and the

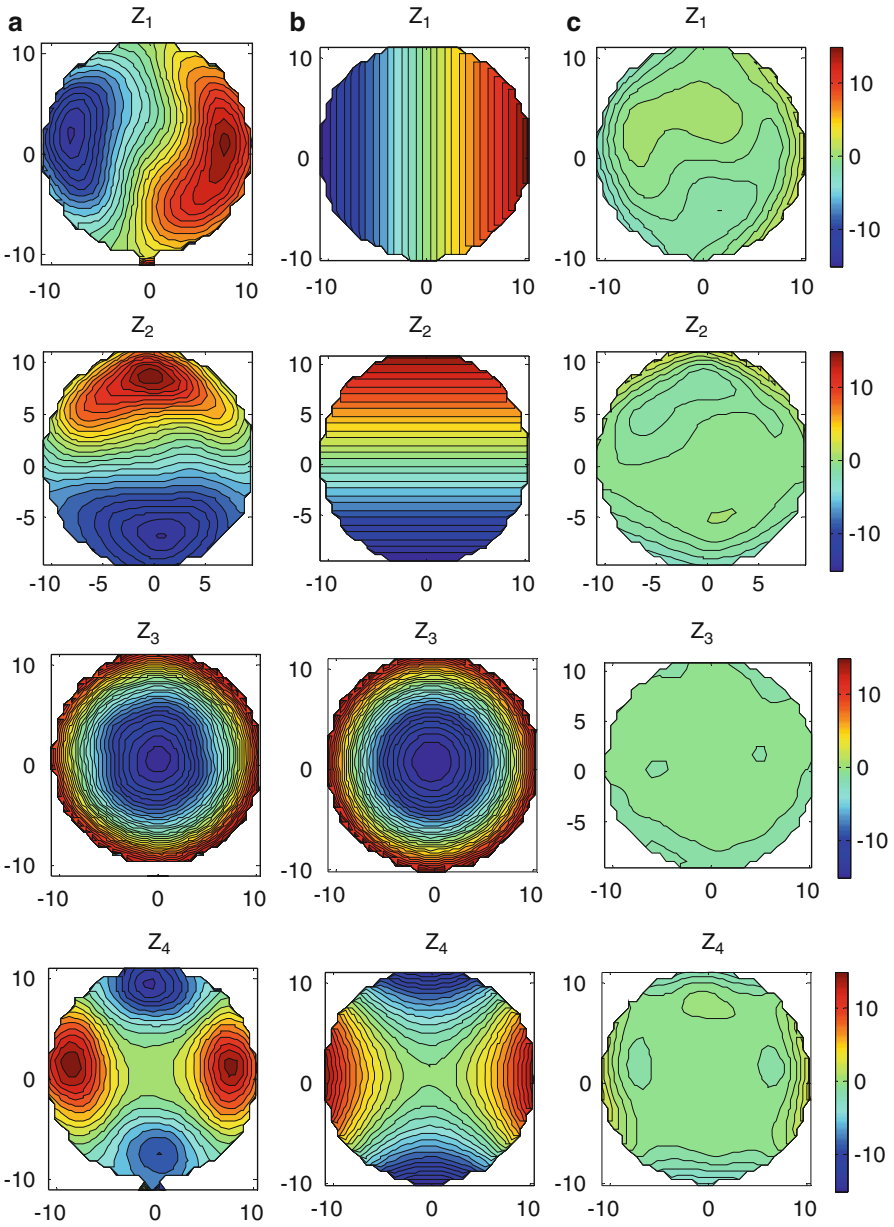


Fig. 7.18 Zernike mode shapes generated using the MFDM with decentralized robust PID controller: (**a**) acquired shapes, (**b**) desired shapes, (**c**) error surface

prototype MFDM was also tested for its ability to generate a series of Zernike mode shapes. Figure 7.18 shows the results where the surface plots of the first 15 Zernike modes are shown as desired shapes and are compared to the shapes obtained using

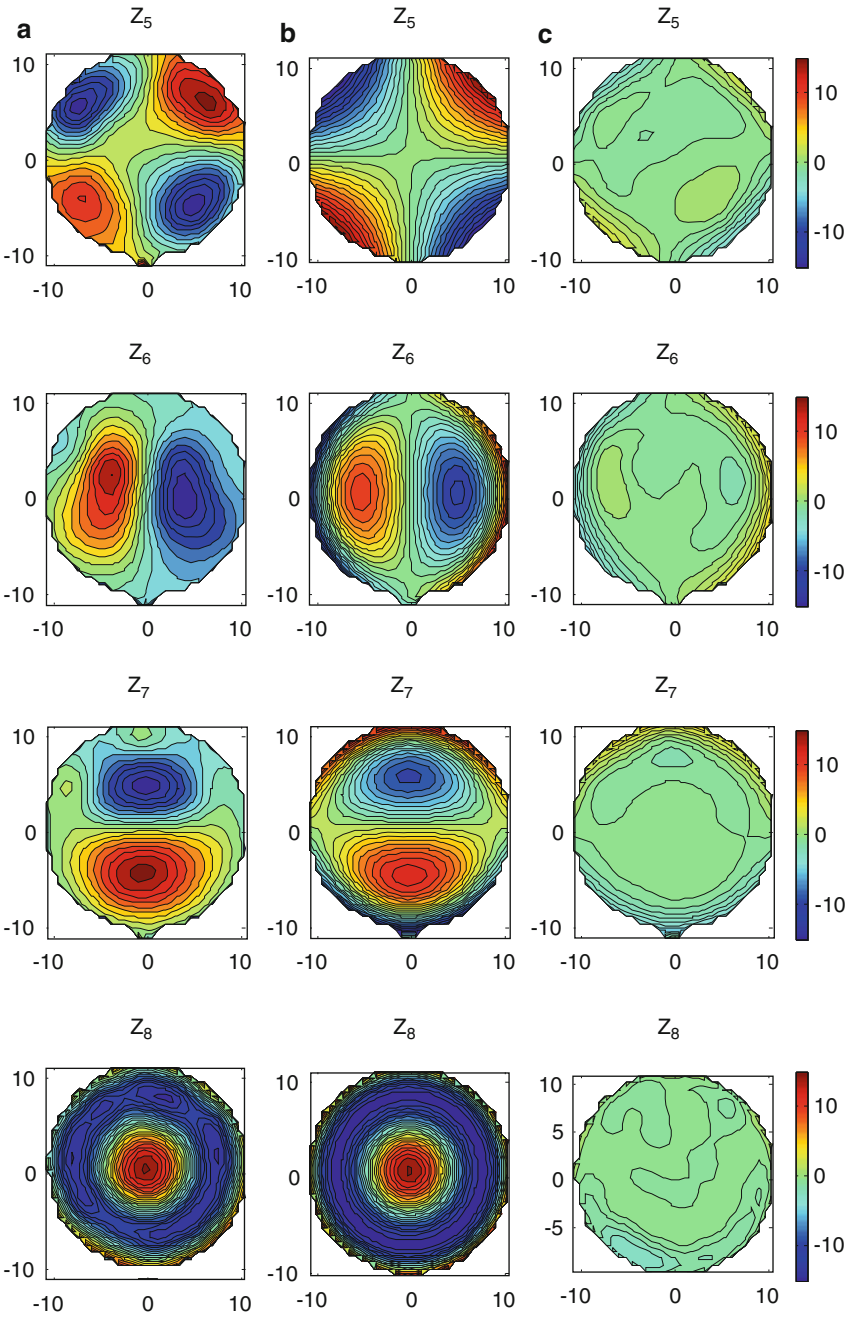


Fig. 7.18 (continued)

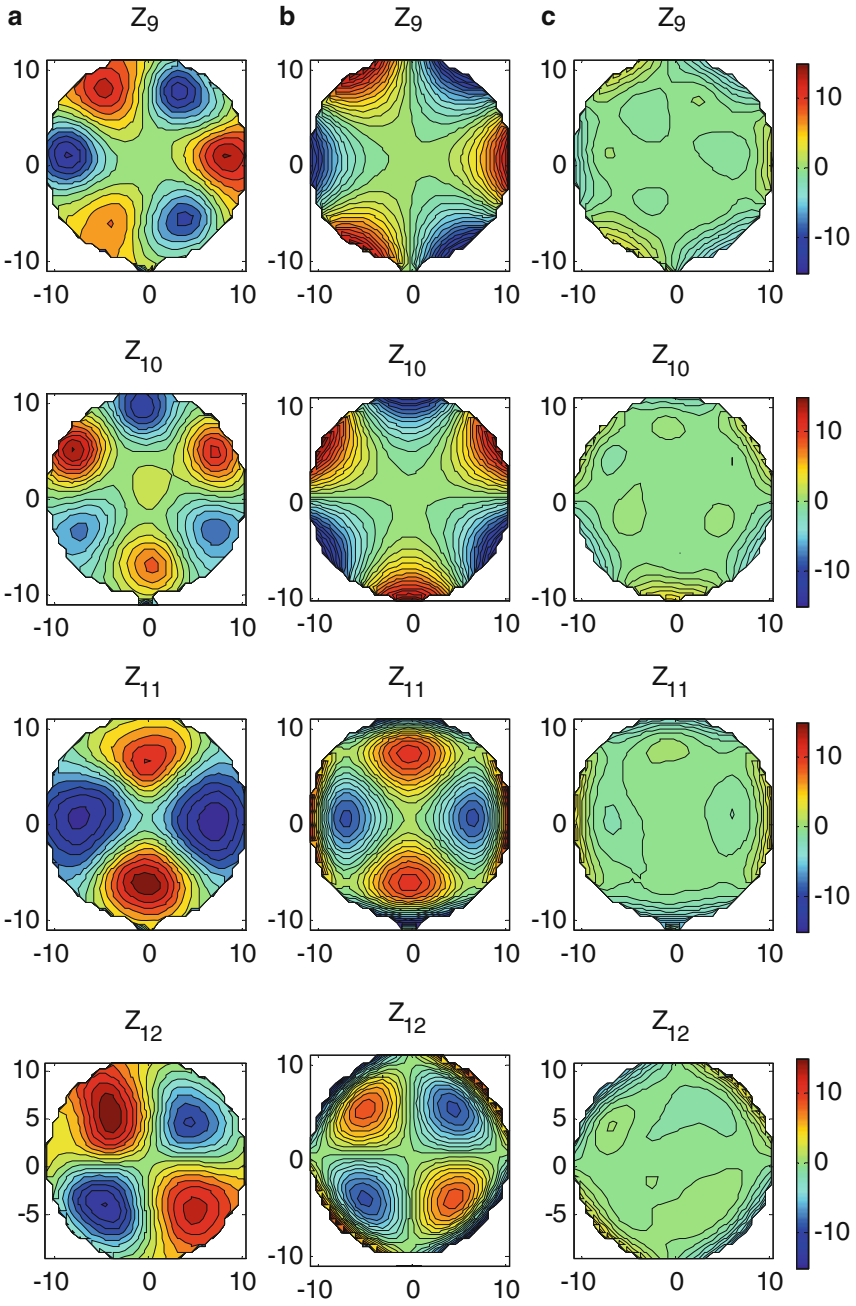


Fig. 7.18 (continued)

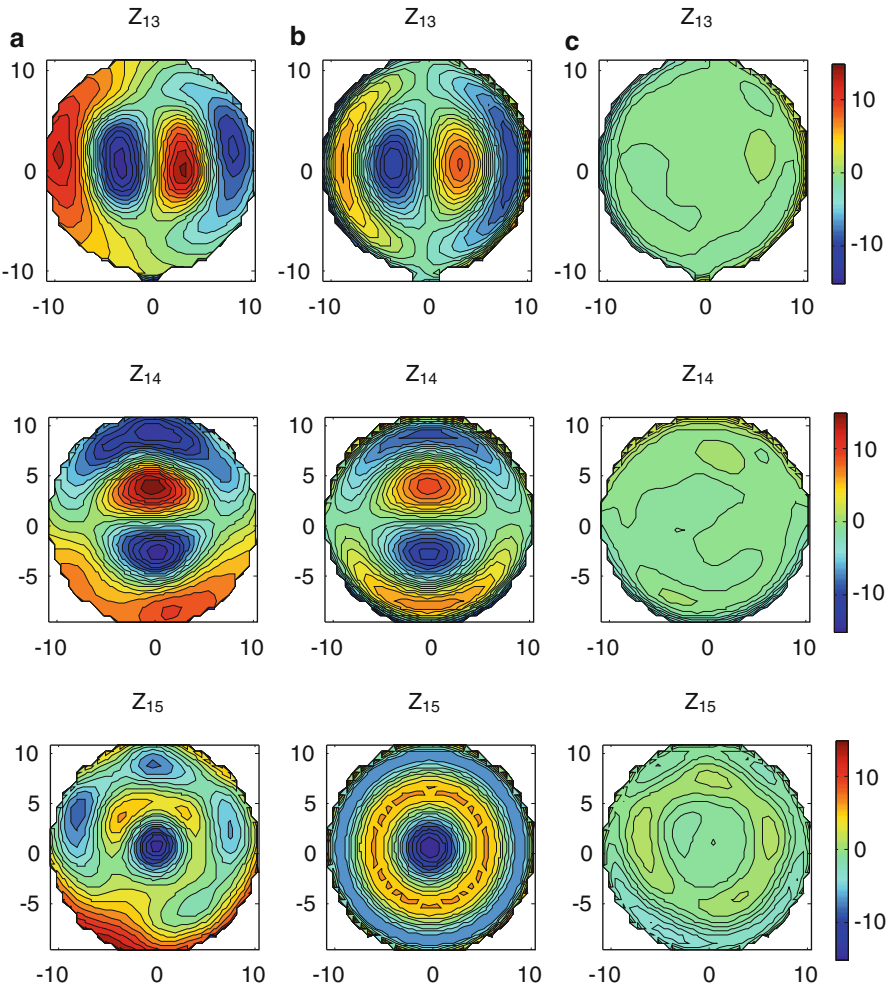


Fig. 7.18 (continued)

the closed-loop system. The difference is shown as an error surface for each mode shape. The RMS of the error surfaces is plotted in Fig. 7.19. The RMS error values are computed only at the 19 control points since the number of actuators used in the prototype is small and computing the RMS error over the whole surface will not in this case be reflective of the capabilities of the control algorithm. It can be seen in Fig. 7.19 that the closed-loop system performs well in achieving the desired shapes. It is therefore expected that, using an MFDM with a larger number of actuators, the RMS error values computed over the whole surface will be comparable to the values computed at the actuator locations.

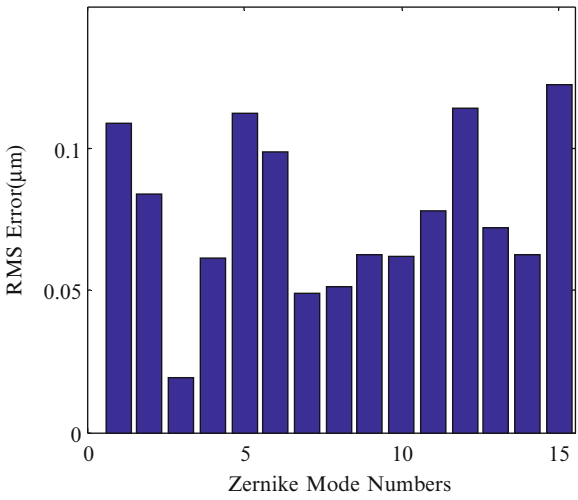


Fig. 7.19 RMS error for the Zernike mode shapes experimentally generated using the MFDM with the decentralized robust PID controller

7.4.2.3 Tracking of Dynamic Shapes

The ability of the controller to track dynamic surface shapes was tested using the reference wavefront shape signals specified in (7.13) where $f_0 = 3$ Hz. Figure 7.20a shows the time history of the experimentally obtained wavefront shape measurements $\mathbf{y}_s$, where only selected channels (# 2, 4, 8, and 17) are shown. The RMS of the wavefront shape error is shown in Fig. 7.20b. As can be observed from the figure, the controller does manage to track the specified reference wavefront shape; however, a relatively significant error still remains uncorrected.

7.5 Summary

In this chapter, two decentralized PID control algorithms are proposed to control the surface shape of a magnetic fluid deformable mirror in a closed-loop adaptive optics system. A decentralized proportional-plus-integral (PI) controller is first designed based on the static decoupling of the plant model using the plant DC gain. This represents a classical approach in the design of decentralized controllers for adaptive optics systems and assumes that the system is fully decoupled. To overcome stability robustness issues and to minimize the effect of the model uncertainties especially in the high frequency range, a decentralized robust proportional-integral-derivative (PID) controller with DC decoupling is then presented. The controller is designed based on a decoupled nominal dynamic model of the plant. To account for the modeling errors in the decoupled plant model, a robust PID controller design problem is transformed into a mixed H_2/H_∞ multi-objective optimization problem

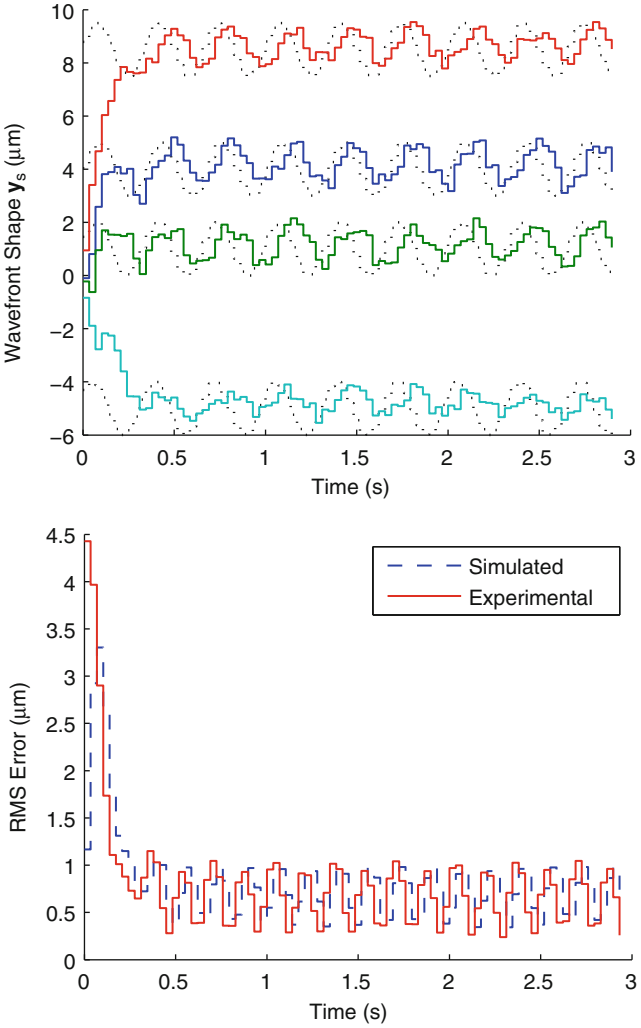


Fig. 7.20 Experimental tracking of a dynamic reference shape using the decentralized robust PID controller: (**a**) wavefront shape, (**b**) RMS error

and solved iteratively based on some properly formulated linear matrix inequalities (LMIs). The two control algorithms mentioned above provide good performance in removing static or very slowly changing aberrations. However, the simulation and experimental verifications show that the decentralized PID control algorithms have limited capabilities in dealing with fast-changing aberrations. To overcome this limitation, centralized optimal control algorithms are developed in the next chapter.

References

Albertos P, Prez PA, Sala A (2004) Multivariable control systems: an engineering approach. Springer, London

Boyd S, Ghaoui LE, Feron E, Balakrishnan V (1994) Linear matrix inequalities in system and control theory. SIAM, Philadelphia

Brousseau D, Borra EF, Thibault S (2007) Wavefront correction with a 37–actuator ferrofluid deformable mirror. Opt Express 15:18190–18199

Cao YY, Lam J, Sun YX (1998) Static output feedback stabilization: an LMI approach. Automatica 34(12):1641–1645

Chen T, Francis B (1996) Optimal sampled data control systems. Springer, London

Laird P, Caron N, Rioux M, Borra EF, Ritcey AM (2006) Ferrofluid adaptive mirrors. Appl Opt 45(15):3495–3500

Phillips CL (2000) Feedback control system. Prentice-Hall, Englewood Cliffs

Zhou K, Doyle J, Golver K (1995) Robust and optimal control. Prentice Hall, Upper Saddle River

Chapter 8
Centralized Optimal Controller Design

Contents

8.1 Multivariable PID Controller

The decentralized control strategy for AO systems presented in the previous chapter is simple and effective in attenuating the effects of static or slowly time-varying wavefront aberrations. However, a significant drawback of this approach is that the decentralized controller design is based on an approximate decoupled model of the plant dynamics, resulting in a reduction in the level of stability and performance robustness of the closed-loop AO system. The dynamic model of the magnetic fluid deformable mirror cannot be fully decoupled using the traditional approach based on the influence function matrix, especially when operating in a relatively high-frequency range. Therefore, it is necessary to develop the centralized control strategy for the AO system that accounts for the full dynamics of the MFDM. A new multivariable PID controller design approach is presented in this section.

Z. Wu et al., *Modeling and Control of Magnetic Fluid Deformable Mirrors for Adaptive Optics Systems*, DOI 10.1007/978-3-642-32229-7_8,

A proportional–integral–derivative controller (PID controller) is a generic type of controllers widely used in industrial control systems. Besides their simplicity, PID controllers can specifically achieve asymptotic tracking of static reference inputs due to the presence of integrators in their structure. The popularity of PID controllers in the industrial world resulted in many approaches to tune the PID controller gains (O'Dwyer 2006; Åström and Hägglund 2005). Most of the existing approaches are developed for the cases where the plant is or can be decoupled into a set of single-input single-output low-order systems. However, in industrial applications, there exist many plants that can only be modeled as coupled multiple-input multiple-output (MIMO) high-order systems. Therefore, it is highly desirable to develop effective methods to design multivariable PID controllers for such MIMO high-order systems.

The design of MIMO PID controllers in the frequency domain using generalized Nyquist stability theorem is addressed in Morari (1985), Nwokah and Perez (1983), and MacFarlane and Belletrutti (1973). However these methods are not convenient to work with due to their complexity. Recently, rigorous multivariable PID controller design methods have been proposed in Bianchi et al. (2008), Lin et al. (2004), Zheng et al. (2003), and Mattei (2001) and share the following common features: first, the main idea behind these approaches is to transform the MIMO PID controller design problem into an equivalent static output feedback (SOF) controller design problem. Second, the controller synthesis procedures are developed based on solving properly formulated bilinear matrix inequalities (BMIs) that can be solved using existing numerical algorithms (Apkarian et al. 2003; Cao et al. 1998; Ghaoui et al. 1997). The PID controller gains can then be further tuned by considering H_∞ or H_2 optimal controller design techniques to optimize different control objectives. However, the search for the optimal controller is heavily dependent on the initial values used in the controller synthesis numerical algorithms. Consequently, the resulting controller performance may be unacceptable. Third, the aforementioned results are mainly proposed for continuous time systems. In this section, the approach to the design of multivariable PID controllers for discrete-time systems is presented. Three main features characterize the approach proposed in this section. First, the multivariable PID controller parameters are designed directly by considering the actual closed-loop system involving the plant and the controller, without transforming the problem into an equivalent static output feedback controller design problem. Second, the proposed controller synthesis procedures are based on solving properly formulated linear matrix inequalities (LMIs). The LMI-based formulation of the controller synthesis procedure results from introducing some assumptions on the state-space representation of the plant, from properly formulating the controller state-space representation, and from introducing slack variables with special structure in the controller synthesis results. Third, the controller design procedure proposed in this section is developed specifically for discrete-time systems. It should be noted that the use of slack variables having a special structure allows the controller synthesis to be formulated based on LMIs instead of BMIs but at the same time introduces some level of conservatism in the controller design, which results in a suboptimal solution to the original nonconvex PID controller design problem.

Multivariable PID controller design approaches for continuous time systems are discussed in Bianchi et al. (2008), Lin et al. (2004), Zheng et al. (2003), Mattei (2001). Given that the control algorithm for an AO system is typically implemented in a computer system, it is more convenient to design the controller directly in discrete time. In the rest of this section, a new multivariable PID controller design approach for discrete time systems is presented. The proposed design approach is formulated as the problem of designing PID controllers that yield closed-loop systems with H_∞ and H_2 performance specifications. Corresponding synthesis algorithms based on properly formulated linear matrix inequalities are developed. Finally, the performance of the proposed controllers is successfully evaluated on an adaptive optics experimental setup, where it is desired to control the surface shape of the novel modified magnetic fluid deformable mirror. The experimental results illustrate the effectiveness of the proposed controller in providing the desired tracking performance and illustrate the potential offered by this controller in closed-loop MFDM-based retinal imaging adaptive optics systems.

8.1.1 Closed-Loop System Presentation with a PID Controller

PID controllers are popular and have been widely used in industrial control systems, mainly due to their simplicity and the ability to asymptotically drive the steady-state error to zero. However, most of the existing PID design approaches are developed for SISO systems. New approaches have been recently developed for continuous time MIMO systems by transforming the MIMO PID controller design problem into an equivalent static output feedback (SOF) controller design problem; however, these approaches are proposed based on solving bilinear matrix inequalities (BMIs) that should be solved iteratively with some derived linear matrix inequalities (LMIs). The search for the optimal controller is heavily dependent on the initial values fed into the synthesis procedure, and most often the obtained solution is unacceptable. In the following, the multivariable PID controller is formulated as a fixed structure controller and is included as part of an augmented plant. Then by introducing proper slack variables, the controller synthesis procedure can be easily derived. Consider the following generalized discrete-time linear time-invariant system:

$$\Sigma : \begin{cases} \mathbf{x}_p(k+1) = \mathbf{A}\mathbf{x}_p(k) + \mathbf{B}_1\mathbf{w}(k) + \mathbf{B}_2\mathbf{u}(k) \\ \mathbf{e}(k) = \mathbf{C}_1\mathbf{x}_p(k) + \mathbf{D}_{11}\mathbf{w}(k) \\ \mathbf{y}(k) = \mathbf{C}_2\mathbf{x}_p(k) + \mathbf{D}_{21}\mathbf{w}(k) \end{cases} \tag{8.1}$$

where $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{B}_1 \in \mathbb{R}^{n\times h}$, $\mathbf{B}_2 \in \mathbb{R}^{n\times l}$, $\mathbf{C}_1 \in \mathbb{R}^{r\times n}$, $\mathbf{C}_2 \in \mathbb{R}^{m\times n}$, $\mathbf{D}_{11} \in \mathbb{R}^{r\times h}$, $\mathbf{D}_{21} \in \mathbb{R}^{m\times h}$, $\mathbf{x}_p(k) \in \mathbb{R}^n$ is the state vector, $\mathbf{u}(k) \in \mathbb{R}^l$ is the control input, $\mathbf{w}(k) \in \mathbb{R}^h$ is an exogenous input signal representing reference or disturbance signals, $\mathbf{y}(k) \in \mathbb{R}^m$ is a measurement signal to be fedback to the controller, and

$\mathbf{e}(k) \in \mathbb{R}^r$ is the performance variable to be regulated. The following assumption is invoked in the rest of the developments presented in this section.

Assumption 8.1 The matrix $\mathbf{B}_2$ in (8.1) is of full column rank. □

Remark 8.1 There is no loss of generality in considering Assumption 8.1 for the plant model (8.1). In fact, if the plant model (8.1) is such that the $\mathbf{B}_2$ matrix is not of full column rank, a new model satisfying Assumption 8.1 can be developed and used for controller design as follows. Consider the case where $\mathbf{B}_2 \in \mathbb{R}^{n\times l}$ is of rank s and $\mathbf{B}_2$ does not have full column rank. A singular value decomposition of $\mathbf{B}_2$ can be written as $\mathbf{B}_2 = \mathbf{U}_1\mathbf{V}\mathbf{U}_2$ where $\mathbf{U}_1 \in \mathbb{R}^{n\times n}$ and $\mathbf{U}_2 \in \mathbb{R}^{l\times l}$ are unitary matrices and where $\mathbf{V} \in \mathbb{R}^{n\times l}$ is given by $\mathbf{V} = \begin{bmatrix} \mathbf{V}_s & \mathbf{0}_{s\times(l-s)} \\ \mathbf{0}_{(n-s)\times s} & \mathbf{0}_{(n-s)\times(l-s)} \end{bmatrix}$, where $\mathbf{V}_s \in \mathbb{R}^{s\times s}$ is a diagonal matrix with entries of the singular values of $\mathbf{B}_2$. Let $\mathbf{v} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix} = \mathbf{U}_2\mathbf{u}$ where $\mathbf{v}_1 \in \mathbb{R}^s$. Consider the full column rank matrix $\bar{\mathbf{B}}_2 = \mathbf{U}_1 \begin{bmatrix} \mathbf{V}_s \\ \mathbf{0}_{(n-s)\times s} \end{bmatrix} \in \mathbb{R}^{n\times s}$. It follows that $\mathbf{B}_2\mathbf{u}(k) = \bar{\mathbf{B}}_2\mathbf{v}_1(k)$. Hence, the plant model can be rewritten as in (8.1) but where the term $\mathbf{B}_2\mathbf{u}(k)$ is replaced with $\bar{\mathbf{B}}_2\mathbf{v}_1(k)$, and can be used in the design of the controller. In the actual implementation of the control system, the controller yields the vector $\mathbf{v}_1(k)$. Since $\mathbf{v}_2(k)$ does not contribute to the vector $\mathbf{B}_2\mathbf{u}(k)$, it can be chosen arbitrarily and set to the zero vector. Hence, the actual control vector $\mathbf{u}(k)$ can then be computed as $\mathbf{u}(k) = \mathbf{U}_2^{-1}\begin{bmatrix} \mathbf{v}_1(k) \\ \mathbf{0}_{(l-s)\times 1} \end{bmatrix}$. That there is no loss of generality in considering Assumption 8.1 for the plant model (8.1) is also due to the fact that the matrix $\mathbf{D}_{12} \in \mathbb{R}^{r\times l}$ that would normally appear in the expression $\mathbf{e}(k) = \mathbf{C}_1\mathbf{x}_p(k) + \mathbf{D}_{11}\mathbf{w}(k) + \mathbf{D}_{12}\mathbf{u}(k)$ in (8.1) is such that $\mathbf{D}_{12} = \mathbf{0}_{r\times l}$. □

In this section, it is desired to regulate the performance variable $\mathbf{e}(k)$ against the exogenous input $\mathbf{w}(k)$ using a multivariable discrete-time PID controller $\mathbf{C}$ given by

$$\begin{aligned} \mathbf{C}(z) &= \mathbf{K}_p + \mathbf{K}_i\frac{z}{z-1} + \mathbf{K}_d\frac{z-1}{z} \\ &= \frac{1}{z^2-z}\left(\left(\mathbf{K}_p + \mathbf{K}_i + \mathbf{K}_d\right)z^2 - \left(\mathbf{K}_p + 2\mathbf{K}_d\right)z + \mathbf{K}_d\right) \end{aligned} \tag{8.2}$$

where $\mathbf{K}_p$, $\mathbf{K}_i$, and $\mathbf{K}_d$ are all in $\mathbb{R}^{l\times m}$ and represent the gain matrices of the PID controller to be designed. A corresponding state-space realization of $\mathbf{C}$ can be written as follows:

$$\mathbf{C}: \begin{cases} \mathbf{x}_c(k+1) = \mathbf{A}_c\mathbf{x}_c(k) + \mathbf{B}_c\mathbf{y}(k) \\ \mathbf{u}(k) = \mathbf{C}_c\mathbf{x}_c(k) + \mathbf{D}_c\mathbf{y}(k) \end{cases} \tag{8.3}$$

where $\mathbf{x}_c(k) \in \mathbb{R}^{2l}$, $\mathbf{A}_c = \begin{bmatrix} \mathbf{I}_{l\times l} & \mathbf{I}_{l\times l} \\ \mathbf{0}_{l\times l} & \mathbf{0}_{l\times l} \end{bmatrix}_{2l\times 2l}$, $\mathbf{B}_c = \begin{bmatrix} \mathbf{K}_i - \mathbf{K}_d \\ \mathbf{K}_d \end{bmatrix}_{2l\times m}$, $\mathbf{C}_c = \begin{bmatrix} \mathbf{I}_{l\times l} & \mathbf{0}_{l\times l} \end{bmatrix}_{l\times 2l}$, and $\mathbf{D}_c = \begin{bmatrix} \mathbf{K}_p + \mathbf{K}_i + \mathbf{K}_d \end{bmatrix}_{l\times m}$. Let $\mathbf{x}(k) = \begin{bmatrix} \mathbf{x}_p(k) \\ \mathbf{x}_c(k) \end{bmatrix} \in \mathbb{R}^N$, where $N = n + 2l$. Then combining (8.1) and (8.3) yields the following closed-loop system representation:

$$\Sigma_{cl} : \begin{cases} \mathbf{x}(k+1) = \mathbf{A}_{\mathrm{cl}}\mathbf{x}(k) + \mathbf{B}_{\mathrm{cl}}\mathbf{w}(k) \\ \mathbf{e}(k) = \mathbf{C}_{\mathrm{cl}}\mathbf{x}(k) + \mathbf{D}_{\mathrm{cl}}\mathbf{w}(k) \end{cases} \tag{8.4}$$

where $\mathbf{A}_{cl} = \begin{bmatrix} \mathbf{A} + \mathbf{B}_2\mathbf{D}_c\mathbf{C}_2 & \mathbf{B}_2\mathbf{C}_c \\ \mathbf{B}_c\mathbf{C}_2 & \mathbf{A}_c \end{bmatrix}_{N\times N}$, $\mathbf{B}_{\mathrm{cl}} = \begin{bmatrix} \mathbf{B}_2\mathbf{D}_c\mathbf{D}_{21} + \mathbf{B}_1 \\ \mathbf{B}_c\mathbf{D}_{21} \end{bmatrix}_{N\times h}$, $\mathbf{C}_{\mathrm{cl}} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{0}_{r\times(2l)} \end{bmatrix}_{r\times N}$, and $\mathbf{D}_{\mathrm{cl}} = \mathbf{D}_{11} \in \mathbb{R}^{r\times h}$. Let $\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A} & \mathbf{B}_2\mathbf{C}_c \\ \mathbf{0}_{2l\times n} & \mathbf{A}_c \end{bmatrix}_{N\times N}$, $\tilde{\mathbf{B}} = \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{0}_{2l\times h} \end{bmatrix}_{N\times h}$, and $\mathbf{K} = \begin{bmatrix} \mathbf{D}_c \\ \mathbf{B}_c \end{bmatrix}_{3l\times m}$. Since $\mathbf{B}_2$ is assumed to be of full column rank, there exists a nonsingular transformation matrix $\mathbf{T}_b \in \mathbb{R}^{n\times n}$ such that $\mathbf{T}_b\mathbf{B}_2 = \begin{bmatrix} \mathbf{0}_{(n-l)\times l} \\ \mathbf{I}_{l\times l} \end{bmatrix}$. Let $\mathbf{T} = \begin{bmatrix} \mathbf{T}_b & \mathbf{0}_{n\times 2l} \\ \mathbf{0}_{2l\times n} & \mathbf{I}_{2l\times 2l} \end{bmatrix} \in \mathbb{R}^{N\times N}$.

Remark 8.2 It should be noted, based on the state-space representation (8.3) of the controller, that only the matrices $\mathbf{B}_c$ and $\mathbf{D}_c$ represent the unknown controller parameters. Hence, the approach presented in this section is valid for the design of any controller with a state-space representation of the form (8.3) where the matrices $\mathbf{A}_c$ and $\mathbf{C}_c$ are fixed and where the matrices $\mathbf{B}_c$ and $\mathbf{D}_c$ represent the unknown controller parameters to be determined in the controller synthesis procedure, as outlined in the following sections. For example, consider a controller given by the following input–output representation:

$$\left(\mathbf{I}z^n + \mathbf{A}_1 z^{n-1} + \mathbf{A}_2 z^{n-2} + \cdots + \mathbf{A}_n\right)\mathbf{U}(z) = \left(\mathbf{B}_0 z^n + \mathbf{B}_1 z^{n-1} + \cdots + \mathbf{B}_n\right)\mathbf{Y}(z)$$

where $\mathbf{U}(z)$ and $\mathbf{Y}(z)$ represent the Z transform of $\mathbf{u}(k)$ and $\mathbf{y}(k)$, respectively; $\mathbf{A}_i \in \mathbb{R}^{l\times l}$, $i = 1, \ldots, n$; $\mathbf{B}_j \in \mathbb{R}^{l\times m}$, $j = 0, \ldots, n$. A state-space representation of the controller can then be given as in (8.3) where

$$\mathbf{A}_c = \begin{bmatrix} -\mathbf{A}_1 & \mathbf{I}_{l\times l} & \mathbf{0}_{l\times l} & \cdots & \mathbf{0}_{l\times l} \\ -\mathbf{A}_2 & \mathbf{0}_{l\times l} & \mathbf{I}_{l\times l} & \cdots & \mathbf{0}_{l\times l} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\mathbf{A}_{n-1} & \mathbf{0}_{l\times l} & \cdots & \ddots & \mathbf{I}_{l\times l} \\ -\mathbf{A}_n & \mathbf{0}_{l\times l} & \cdots & \cdots & \mathbf{0}_{l\times l} \end{bmatrix}, \quad \mathbf{B}_c = \begin{bmatrix} \mathbf{B}_1 - \mathbf{A}_1\mathbf{B}_0 \\ \mathbf{B}_2 - \mathbf{A}_2\mathbf{B}_0 \\ \vdots \\ \vdots \\ \mathbf{B}_n - \mathbf{A}_n\mathbf{B}_0 \end{bmatrix}$$

$$\mathbf{C}_c = \begin{bmatrix} \mathbf{I}_{l\times l} & \mathbf{0}_{l\times((n-1)\times l)} \end{bmatrix}, \qquad \mathbf{D}_c = \mathbf{B}_0$$

If the matrices $\mathbf{A}_i \in \mathbb{R}^{l\times l}$, $i = 1, \ldots, n$, are fixed, then the controller state-space realization is such that $\mathbf{A}_c$ and $\mathbf{C}_c$ are fixed and the matrices $\mathbf{B}_c$ and $\mathbf{D}_c$ are to be determined through the controller synthesis procedure. □

In the following, based on the closed-loop system representation (8.4), the design of PID controllers that stabilize the closed-loop system and that optimize the resulting H_∞ or H_2 closed-loop system performance is investigated.

8.1.2 Stabilization with PID Controllers

To study the stabilization problem using PID controllers, consider the closed-loop system (8.4) with $\mathbf{w}(k) = 0, \forall k \geq 0$. It follows that

$$\mathbf{x}(k+1) = \mathbf{A}_{cl}\mathbf{x}(k) \tag{8.5}$$

It is well known that the system (8.5) is exponentially stable if and only if there exists a matrix $\mathbf{P} > 0$ such that

$$\mathbf{A}_{\mathrm{cl}}^{\mathrm{T}}\mathbf{P}\mathbf{A}_{\mathrm{cl}} - \mathbf{P} < 0 \tag{8.6}$$

Then, we have the following result.

Theorem 8.1 *Consider the system (8.1) where Assumption 8.1 is satisfied. If there exist a positive-definite matrix* $\mathbf{P} \in \mathbb{R}^{N\times N}$, *matrices* $\mathbf{J} = \begin{bmatrix} \mathbf{J}_{11} & \mathbf{0}_{(N-q)\times 3l} \\ \mathbf{J}_{21} & \mathbf{J}_{22} \end{bmatrix} \in \mathbb{R}^{N\times N}$, $\mathbf{J}_{22} \in \mathbb{R}^{q\times 3l}$, $1 \leq q \leq 3l$, $\mathbf{L} = \begin{bmatrix} \mathbf{L}_1 & \mathbf{0}_{N\times 3l} \end{bmatrix} \in \mathbb{R}^{N\times N}$, $\mathbf{L}_1 \in \mathbb{R}^{N\times(N-3l)}$ *and* $\mathbf{X} \in \mathbb{R}^{q\times m}$ *such that*

$$\begin{bmatrix} \mathbf{P} - \mathbf{J}\mathbf{T} - \mathbf{T}^{\mathrm{T}}\mathbf{J}^{\mathrm{T}} & \begin{bmatrix} \mathbf{0}_{(N-q)\times n} & \mathbf{0}_{(N-q)\times 2l} \\ \mathbf{X}\mathbf{C}_2 & \mathbf{0}_{q\times 2l} \end{bmatrix} + \mathbf{J}\mathbf{T}\tilde{\mathbf{A}} - \mathbf{T}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}} \\ \begin{bmatrix} \mathbf{0}_{n\times(N-q)} & (\mathbf{X}\mathbf{C}_2)^{\mathrm{T}} \\ \mathbf{0}_{2l\times(N-q)} & \mathbf{0}_{2l\times q} \end{bmatrix} + \tilde{\mathbf{A}}^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}\mathbf{J}^{\mathrm{T}} - \mathbf{L}\mathbf{T} & -\mathbf{P} + \mathbf{L}\mathbf{T}\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}} \end{bmatrix} < 0 \tag{8.7}$$

and if there exists $\mathbf{K}$ *such that*

$$\mathbf{J}_{22}\mathbf{K} = \mathbf{X} \tag{8.8}$$

then the system (8.5) is exponentially stable. □

Proof Inequality (8.7) is equivalent to

$$\begin{bmatrix} \mathbf{P} & \begin{bmatrix} \mathbf{0}_{(N-q)\times n} & \mathbf{0}_{(N-q)\times 2l} \\ \mathbf{X}\mathbf{C}_2 & \mathbf{0}_{q\times 2l} \end{bmatrix} \\ \begin{bmatrix} \mathbf{0}_{n\times(N-q)} & (\mathbf{X}\mathbf{C}_2)^{\mathrm{T}} \\ \mathbf{0}_{2l\times(N-q)} & \mathbf{0}_{2l\times q} \end{bmatrix} & -\mathbf{P} \end{bmatrix} + \begin{bmatrix} -\mathbf{J}\mathbf{T} - \mathbf{T}^{\mathrm{T}}\mathbf{J}^{\mathrm{T}} & \mathbf{J}\mathbf{T}\tilde{\mathbf{A}} - \mathbf{T}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}} \\ \tilde{\mathbf{A}}^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}\mathbf{J}^{\mathrm{T}} - \mathbf{L}\mathbf{T} & \mathbf{L}\mathbf{T}\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}} \end{bmatrix} < 0$$

It follows that

$$\begin{bmatrix} \mathbf{P} & \begin{bmatrix} \mathbf{0}_{(N-q)\times n} & \mathbf{0}_{(N-q)\times 2l} \\ \mathbf{X}\mathbf{C}_2 & \mathbf{0}_{q\times 2l} \end{bmatrix} \\ \begin{bmatrix} \mathbf{0}_{n\times(N-q)} & (\mathbf{X}\mathbf{C}_2)^{\mathrm{T}} \\ \mathbf{0}_{2l\times(N-q)} & \mathbf{0}_{2l\times q} \end{bmatrix} & -\mathbf{P} \end{bmatrix} + \begin{bmatrix} \mathbf{J}\mathbf{T} \\ \mathbf{L}\mathbf{T} \end{bmatrix} \begin{bmatrix} -\mathbf{I}_{N\times N} & \tilde{\mathbf{A}} \end{bmatrix} + \begin{bmatrix} -\mathbf{I}_{N\times N} & \tilde{\mathbf{A}} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \mathbf{J}\mathbf{T} \\ \mathbf{L}\mathbf{T} \end{bmatrix}^{\mathrm{T}} < 0 \tag{8.9}$$

Using the expression for $\mathbf{X}$ from (8.8), it follows that

$$\begin{aligned} \begin{bmatrix} \mathbf{0}_{(N-q)\times n} & \mathbf{0}_{(N-q)\times 2l} \\ \mathbf{X}\mathbf{C}_2 & \mathbf{0}_{q\times 2l} \end{bmatrix} &= \begin{bmatrix} \mathbf{0}_{(N-q)\times n} & \mathbf{0}_{(N-q)\times 2l} \\ \mathbf{J}_{22}\mathbf{K}\mathbf{C}_2 & \mathbf{0}_{q\times 2l} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{J}_{11} & \mathbf{0}_{(N-q)\times 3l} \\ \mathbf{J}_{21} & \mathbf{J}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{0}_{(N-3l)\times n} & \mathbf{0}_{(N-3l)\times 2l} \\ \mathbf{K}\mathbf{C}_2 & \mathbf{0}_{3l\times 2l} \end{bmatrix} \\ &= \mathbf{J} \begin{bmatrix} \mathbf{0}_{(N-3l)\times n} & \mathbf{0}_{(N-3l)\times 2l} \\ \begin{bmatrix} \mathbf{D}_c \\ \mathbf{B}_c \end{bmatrix} \mathbf{C}_2 & \mathbf{0}_{3l\times 2l} \end{bmatrix} \\ &= \mathbf{J} \begin{bmatrix} \begin{bmatrix} \mathbf{0}_{(n-l)\times l} \\ \mathbf{I}_{l\times l} \end{bmatrix} \mathbf{D}_c\mathbf{C}_2 & \mathbf{0}_{n\times 2l} \\ \mathbf{B}_c\mathbf{C}_2 & \mathbf{0}_{2l\times 2l} \end{bmatrix} \\ &= \mathbf{J} \begin{bmatrix} \mathbf{T}_b\mathbf{B}_2\mathbf{D}_c\mathbf{C}_2 & \mathbf{0}_{n\times 2l} \\ \mathbf{B}_c\mathbf{C}_2 & \mathbf{0}_{2l\times 2l} \end{bmatrix} \\ &= \mathbf{J} \begin{bmatrix} \mathbf{T}_b & \mathbf{0}_{n\times 2l} \\ \mathbf{0}_{2l\times n} & \mathbf{I}_{2l\times 2l} \end{bmatrix} \begin{bmatrix} \mathbf{B}_2\mathbf{D}_c\mathbf{C}_2 & \mathbf{0}_{n\times 2l} \\ \mathbf{B}_c\mathbf{C}_2 & \mathbf{0}_{2l\times 2l} \end{bmatrix} \\ &= \mathbf{J}\mathbf{T} \begin{bmatrix} \mathbf{B}_2\mathbf{D}_c\mathbf{C}_2 & \mathbf{0}_{n\times 2l} \\ \mathbf{B}_c\mathbf{C}_2 & \mathbf{0}_{2l\times 2l} \end{bmatrix} \end{aligned}$$

Therefore, we have that $\begin{bmatrix} \mathbf{0}_{(N-q)\times n} & \mathbf{0}_{(N-q)\times 2l} \\ \mathbf{X}\mathbf{C}_2 & \mathbf{0}_{q\times 2l} \end{bmatrix} + \mathbf{J}\mathbf{T}\tilde{\mathbf{A}} = \mathbf{J}\mathbf{T}\mathbf{A}_{\mathrm{cl}}$. It can also be shown that $\mathbf{L}\mathbf{T}\mathbf{A}_{\mathrm{cl}} = \mathbf{L}\mathbf{T}\tilde{\mathbf{A}}$. Hence, (8.9) is rewritten as

$$\begin{bmatrix} \mathbf{P} & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & -\mathbf{P} \end{bmatrix} + \begin{bmatrix} \mathbf{J}\mathbf{T} \\ \mathbf{L}\mathbf{T} \end{bmatrix} \begin{bmatrix} -\mathbf{I}_{N\times N} & \mathbf{A}_{\mathrm{cl}} \end{bmatrix} + \begin{bmatrix} -\mathbf{I}_{N\times N} & \mathbf{A}_{\mathrm{cl}} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \mathbf{J}\mathbf{T} \\ \mathbf{L}\mathbf{T} \end{bmatrix}^{\mathrm{T}} < 0 \quad (8.10)$$

Multiplying by $\begin{bmatrix} \mathbf{x}(k+1) \\ \mathbf{x}(k) \end{bmatrix}^{\mathrm{T}}$ and $\begin{bmatrix} \mathbf{x}(k+1) \\ \mathbf{x}(k) \end{bmatrix}$ from the left- and right-hand sides of (8.10), respectively, we have

$$\mathbf{x}^{\mathrm{T}}(k+1)\mathbf{P}\mathbf{x}(k+1) - \mathbf{x}^{\mathrm{T}}(k)\mathbf{P}\mathbf{x}(k) < 0$$

which implies (8.6). Hence, the closed-loop system (8.5) is exponentially stable. □

Remark 8.3 Based on Assumption 8.1, we always have $n \geq l$, and therefore, $N = n + 2l \geq 3l$. For the special case of $n = l$, the matrices $\mathbf{J}$ and $\mathbf{L}$ in Theorem 8.1 are simplified to $\mathbf{J} = \mathbf{J}_{22}$ and $\mathbf{L} = \mathbf{0}_{3l\times 3l}$. The same holds true for similar matrices in the rest of the theorems presented in this section. □

Remark 8.4 All the matrices that appear in the LMI (8.7) are independent of the controller gains. Once $\mathbf{K}$ is determined based on (8.8), the corresponding parameters in the PID controller can be obtained directly from $\mathbf{K}$. The above comments hold true for the LMIs involved in all of the other theorems presented in this section. □

Remark 8.5 Once an admissible solution to (8.7) is found, then equality (8.8) is first solved without requiring the matrix $\mathbf{J}_{22}$ to be nonsingular. If equality (8.8) does not admit a solution, then the following approach can be used to deal with this situation. Select $\mathbf{J}_{22} \in \mathbb{R}^{3l\times 3l}$ and solve (8.7) while taking into account an added constraint of the form $\mathbf{J}\mathbf{T} = \mathbf{T}^{\mathrm{T}}\mathbf{J}^{\mathrm{T}}$. The resulting $\mathbf{J}_{22}$ is guaranteed to be nonsingular, and the solution will be $\mathbf{K} = \mathbf{J}_{22}^{-1}\mathbf{X}$. The proof of the invertibility of $\mathbf{J}_{22}$ follows from $\mathbf{P} - \mathbf{J}\mathbf{T} - \mathbf{T}^{\mathrm{T}}\mathbf{J}^{\mathrm{T}} < 0$ which is derived from the (1,1) block in the left-hand side of (8.7). Since $\mathbf{P} > 0$, then $0 < \mathbf{J}\mathbf{T} + \mathbf{T}^{\mathrm{T}}\mathbf{J}^{\mathrm{T}} = 2\mathbf{J}\mathbf{T}$, which implies that $0 < \mathbf{J}\mathbf{T}$ and therefore $\mathbf{J}\mathbf{T}$ is invertible. Since $\mathbf{T}$ is nonsingular, then so is $\mathbf{J}$. Due to the block diagonal structure of $\mathbf{J}$, it follows that the matrix $\mathbf{J}_{22}$ is invertible. The same idea can be applied to similar situations arising from the rest of the theorems presented in this section. □

In typical closed-loop systems involving PID controllers, satisfactory time response of the closed-loop system is very important. However, the above proposed design method provides little control over the closed-loop system pole location and transient behavior. In the following, the PID controller design method is modified so that the closed-loop system poles are forced into a suitable subregion of the unit circle having desired properties. Let λ be an eigenvalue of the system matrix $\mathbf{A}_{\mathrm{cl}}$, λ_r and λ_i, the real and imaginary parts of λ, respectively. Consider a disk region

symmetrical with respect to the real axis as $\Im_D : \left\{(\lambda_r, \lambda_i) | \, (\lambda_r + a)^2 + {\lambda_i}^2 < b^2\right\}$. All eigenvalues of the system matrix $\mathbf{A}_{\rm cl}$ lie in the region $\Im_D$ if and only if there exists a $\mathbf{P} > 0$ such that (Masubuchi et al. 1998)

$$\begin{bmatrix} b\mathbf{P} & \mathbf{A}_{\rm cl}^{\rm T}\mathbf{P} + a\mathbf{P} \\ \mathbf{P}\mathbf{A}_{\rm cl} + a\mathbf{P} & b\mathbf{P} \end{bmatrix} > 0 \tag{8.11}$$

Then we have the following result to add a pole-clustering constraint in the design procedure.

Theorem 8.2 *Consider the system (8.1) where Assumption 8.1 is satisfied. If there exist a positive-definite matrix* $\mathbf{P} \in \mathbb{R}^{N\times N}$, *a matrix* $\mathbf{J} = \begin{bmatrix} \mathbf{J}_{11} & \mathbf{0}_{(N-3l)\times 3l} \\ \mathbf{J}_{21} & \mathbf{J}_{22} \end{bmatrix} \in \mathbb{R}^{N\times N}$, *where* $\mathbf{J}_{22} \in \mathbb{R}^{3l\times 3l}$, *and* $\mathbf{X} \in \mathbb{R}^{3l\times m}$ *such that*

$$\begin{bmatrix} b\mathbf{P} & \begin{bmatrix} \mathbf{0}_{n\times(N-3l)} & (\mathbf{X}\mathbf{C}_2)^{\rm T} \\ \mathbf{0}_{2l\times(N-3l)} & \mathbf{0}_{2l\times 3l} \end{bmatrix} + \tilde{\mathbf{A}}^{\rm T}\mathbf{T}^{\rm T}\mathbf{J}^{\rm T} + a\mathbf{T}^{\rm T}\mathbf{J}^{\rm T} \\ \begin{bmatrix} \mathbf{0}_{(N-3l)\times n} & \mathbf{0}_{(N-3l)\times 2l} \\ \mathbf{X}\mathbf{C}_2 & \mathbf{0}_{3l\times 2l} \end{bmatrix} + \mathbf{J}\mathbf{T}\tilde{\mathbf{A}} + a\mathbf{J}\mathbf{T} & b\left(\mathbf{J}\mathbf{T} + \mathbf{T}^{\rm T}\mathbf{J}^{\rm T} - \mathbf{P}\right) \end{bmatrix} > 0 \tag{8.12}$$

and if $\mathbf{J}$ *is nonsingular, then with*

$$\mathbf{K} = \mathbf{J}_{22}^{-1}\mathbf{X} \tag{8.13}$$

the eigenvalues of the system matrix $\mathbf{A}_{cl}$ *lie in the region* $\Im_D$. □

Proof First, as in the proof of Theorem 8.1, using (8.13) and the fact that $\begin{bmatrix} \mathbf{0}_{(N-3l)\times n} & \mathbf{0}_{(N-3l)\times 2l} \\ \mathbf{X}\mathbf{C}_2 & \mathbf{0}_{3l\times 2l} \end{bmatrix} + \mathbf{J}\mathbf{T}\tilde{\mathbf{A}} = \mathbf{J}\mathbf{T}\mathbf{A}_{\rm cl}$, (8.12) implies that

$$\begin{bmatrix} b\mathbf{P} & \mathbf{A}_{\rm cl}^{\rm T}\mathbf{T}^{\rm T}\mathbf{J}^{\rm T} + a\mathbf{T}^{\rm T}\mathbf{J}^{\rm T} \\ \mathbf{J}\mathbf{T}\mathbf{A}_{\rm cl} + a\mathbf{J}\mathbf{T} & b\left(\mathbf{J}\mathbf{T} + \mathbf{T}^{\rm T}\mathbf{J}^{\rm T} - \mathbf{P}\right) \end{bmatrix} > 0 \tag{8.14}$$

Since $\mathbf{P} > 0$, then we have $(\mathbf{P} - \mathbf{J}\mathbf{T})\,\mathbf{P}^{-1}(\mathbf{P} - \mathbf{J}\mathbf{T})^{\rm T} \geq 0$, or equivalently $\mathbf{J}\mathbf{T}\mathbf{P}^{-1}\mathbf{T}^{\rm T}\mathbf{J}^{\rm T} \geq \mathbf{J}\mathbf{T} + \mathbf{T}^{\rm T}\mathbf{J}^{\rm T} - \mathbf{P}$. Therefore, (8.14) implies that

$$\begin{bmatrix} b\mathbf{P} & \mathbf{A}_{\rm cl}^{\rm T}\mathbf{T}^{\rm T}\mathbf{J}^{\rm T} + a\mathbf{T}^{\rm T}\mathbf{J}^{\rm T} \\ \mathbf{J}\mathbf{T}\mathbf{A}_{\rm cl} + a\mathbf{J}\mathbf{T} & b\mathbf{J}\mathbf{T}\mathbf{P}^{-1}\mathbf{T}^{\rm T}\mathbf{J}^{\rm T} \end{bmatrix} > 0 \tag{8.15}$$

Multiplying (8.15) with $diag\left[\mathbf{I}, \mathbf{P}(\mathbf{J}\mathbf{T})^{-1}\right]$ on the left side and $diag\left[\mathbf{I}, \left(\mathbf{T}^{\rm T}\mathbf{J}^{\rm T}\right)^{-1}\mathbf{P}\right]$ on the right-hand side, we immediately have (8.11). □

In the following, the PID design problem with H_∞ and H_2 performance specifications in the closed-loop system is investigated, where the parameters in the PID controller are tuned using H_∞ and H_2 constraints.

8.1.3 PID Controller Design with an H_∞ Performance Constraint

This section presents the design of a PID controller that yields a closed-loop system with a desired H_∞ performance level. Let $\mathbf{T}_{ew}$ denote the closed-loop system transfer function relating $\mathbf{w}$ to $\mathbf{e}$ in (8.4). Consider a real number $\gamma > 0$, and assume the initial condition in the system (8.4) is $\mathbf{x}(0) = 0$. If $\|\mathbf{e}(k)\|_2 \leq \gamma \|\mathbf{w}(k)\|_2$, then we say that the worst case attenuation of the exogenous signal $\mathbf{w}(k)$ is given by γ and the H_∞ norm of the closed-loop system (8.4) satisfies

$$\|\mathbf{T}_{ew}\|_\infty < \gamma \tag{8.16}$$

The following Lemma will be used to design a PID controller that yields a desired closed-loop system with an H_∞ performance level γ.

Lemma 8.1 *(Cuzzola and Morari 2002) Consider the closed-loop system (8.4) with zero initial condition* $\mathbf{x}(0) = 0$. *If there exists a matrix* $\mathbf{P} > 0$ *such that*

$$\mathbf{x}^{\mathrm{T}}(k+1)\mathbf{P}\mathbf{x}(k+1) - \mathbf{x}^{\mathrm{T}}(k)\mathbf{P}\mathbf{x}(k) < \gamma^2 \mathbf{w}^{\mathrm{T}}(k)\mathbf{w}(k) - \mathbf{e}^{\mathrm{T}}(k)\mathbf{e}(k) \tag{8.17}$$

then the H_∞ *norm of the closed-loop system (8.4) satisfies (8.16).* □

Based on the above result, the following LMI conditions can be used to design the multivariable PID controller.

Theorem 8.3 *Consider the system (8.1) where Assumption 8.1 is satisfied. If there exist a positive-definite matrix* $\mathbf{P} \in \mathbb{R}^{N\times N}$, $\mathbf{F} = \begin{bmatrix} \mathbf{F}_{11} & \mathbf{0}_{(N-q)\times 3l} \\ \mathbf{F}_{21} & \mathbf{F}_{22} \end{bmatrix} \in \mathbb{R}^{N\times N}$, $\mathbf{F}_{22} \in \mathbb{R}^{q\times 3l}$, $1 \leq q \leq 3l$, $\mathbf{G}_1 = \begin{bmatrix} \mathbf{G}_{11} & \mathbf{0} \end{bmatrix} \in R^{N\times N}$, $\mathbf{G}_{11} \in \mathbb{R}^{N\times(N-3l)}$, $\mathbf{G}_2 = \begin{bmatrix} \mathbf{G}_{21} & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{h\times N}$, $\mathbf{G}_{21} \in \mathbb{R}^{h\times(N-3l)}$, $\mathbf{G}_3 = \begin{bmatrix} \mathbf{G}_{31} & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{r\times N}$, $\mathbf{G}_{31} \in \mathbb{R}^{r\times(N-3l)}$, $\mathbf{H}_1 \in \mathbb{R}^{N\times r}$, $\mathbf{H}_2 \in \mathbb{R}^{N\times r}$, $\mathbf{H}_3 \in \mathbb{R}^{h\times r}$, $\mathbf{H}_4 \in \mathbb{R}^{r\times r}$, *and* $\mathbf{Y} \in \mathbb{R}^{q\times m}$ *such that*

$$\begin{bmatrix} \mathbf{P} - \mathbf{F}\mathbf{T} - \mathbf{T}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}} & * & * & * \\ \mathbf{N}_1^{\mathrm{T}} + \tilde{\mathbf{A}}^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}} - \mathbf{G}_1\mathbf{T} + \mathbf{C}_{\mathrm{cl}}^{\mathrm{T}}\mathbf{H}_1^{\mathrm{T}} & -\mathbf{P} + \mathbf{G}_1\mathbf{T}\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}\mathbf{G}_1^{\mathrm{T}} + \mathbf{H}_2\mathbf{C}_{\mathrm{cl}} + \mathbf{C}_{\mathrm{cl}}^{\mathrm{T}}\mathbf{H}_2^{\mathrm{T}} & * & * \\ \mathbf{N}_2^{\mathrm{T}} + \tilde{\mathbf{B}}^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}} - \mathbf{G}_2\mathbf{T} + \mathbf{D}_{\mathrm{cl}}^{\mathrm{T}}\mathbf{H}_1^{\mathrm{T}} & \mathbf{G}_2\mathbf{T}\tilde{\mathbf{A}} + \mathbf{H}_3\mathbf{C}_{\mathrm{cl}} + \tilde{\mathbf{B}}^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}\mathbf{G}_1^{\mathrm{T}} + \mathbf{D}_{\mathrm{cl}}^{\mathrm{T}}\mathbf{H}_2^{\mathrm{T}} & -\gamma^2\mathbf{I} + \mathbf{G}_2\mathbf{T}\tilde{\mathbf{B}} + \mathbf{H}_3\mathbf{D}_{\mathrm{cl}} + \tilde{\mathbf{B}}^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}\mathbf{G}_2^{\mathrm{T}} + \mathbf{D}_{\mathrm{cl}}^{\mathrm{T}}\mathbf{H}_3^{\mathrm{T}} & * \\ -\mathbf{G}_3\mathbf{T} - \mathbf{H}_1^{\mathrm{T}} & \mathbf{G}_3\mathbf{T}\tilde{\mathbf{A}} + \mathbf{H}_4\mathbf{C}_{cl} - \mathbf{H}_2^{\mathrm{T}} & \mathbf{G}_3\mathbf{T}\tilde{\mathbf{B}} + \mathbf{H}_4\mathbf{D}_{\mathrm{cl}} - \mathbf{H}_3^{T} & \mathbf{I} - \mathbf{H}_4 - \mathbf{H}_4^{\mathrm{T}} \end{bmatrix} < 0 \tag{8.18}$$

where $\mathbf{N}_1 = \begin{bmatrix} \mathbf{0}_{(N-q)\times n} & \mathbf{0}_{(N-q)\times 2l} \\ \mathbf{YC}_2 & \mathbf{0}_{q\times 2l} \end{bmatrix}$ *and* $\mathbf{N}_2 = \begin{bmatrix} \mathbf{0}_{(N-q)\times h} \\ \mathbf{YD}_{21} \end{bmatrix}$, *and if there exists* $\mathbf{K}$ *such that*

$$\mathbf{F}_{22}\mathbf{K} = \mathbf{Y} \tag{8.19}$$

then the H_∞ *norm of the closed-loop system (8.4) satisfies (8.16).* □

Proof First, by simple calculation, (8.18) is equivalent to

$$\begin{bmatrix} \mathbf{P} & \begin{bmatrix} \mathbf{0}_{(N-q)\times n} & \mathbf{0}_{(N-q)\times 2l} \\ \mathbf{YC}_2 & \mathbf{0}_{q\times 2l} \end{bmatrix} & \begin{bmatrix} \mathbf{0}_{(N-q)\times h} \\ \mathbf{YD}_{21} \end{bmatrix} & \mathbf{0}_{N\times r} \\ \begin{bmatrix} \mathbf{0}_{(N-q)\times n} & \mathbf{0}_{(N-q)\times 2l} \\ \mathbf{YC}_2 & \mathbf{0}_{q\times 2l} \end{bmatrix}^{\mathrm{T}} & -\mathbf{P} & \mathbf{0}_{N\times h} & \mathbf{0}_{N\times r} \\ \begin{bmatrix} \mathbf{0}_{(N-q)\times h} \\ \mathbf{YD}_{21} \end{bmatrix}^{\mathrm{T}} & \mathbf{0}_{h\times N} & -\gamma^2\mathbf{I}_{h\times h} & \mathbf{0}_{h\times r} \\ \mathbf{0}_{r\times N} & \mathbf{0}_{r\times N} & \mathbf{0}_{r\times h} & \mathbf{I}_{r\times r} \end{bmatrix}$$

$$+ \begin{bmatrix} \mathbf{FT} & \mathbf{H}_1 \\ \mathbf{G}_1\mathbf{T} & \mathbf{H}_2 \\ \mathbf{G}_2\mathbf{T} & \mathbf{H}_3 \\ \mathbf{G}_3\mathbf{T} & \mathbf{H}_4 \end{bmatrix} \begin{bmatrix} -\mathbf{I}_{N\times N} & \tilde{\mathbf{A}} & \tilde{\mathbf{B}} & \mathbf{0}_{N\times r} \\ \mathbf{0}_{r\times N} & \mathbf{C}_{cl} & \mathbf{D}_{cl} & -\mathbf{I}_{r\times r} \end{bmatrix} + \begin{bmatrix} -\mathbf{I}_{N\times N} & \tilde{\mathbf{A}} & \tilde{\mathbf{B}} & \mathbf{0}_{N\times r} \\ \mathbf{0}_{r\times N} & \mathbf{C}_{\mathrm{cl}} & \mathbf{D}_{\mathrm{cl}} & -\mathbf{I}_{r\times r} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \mathbf{FT} & \mathbf{H}_1 \\ \mathbf{G}_1\mathbf{T} & \mathbf{H}_2 \\ \mathbf{G}_2\mathbf{T} & \mathbf{H}_3 \\ \mathbf{G}_3\mathbf{T} & \mathbf{H}_4 \end{bmatrix}^{\mathrm{T}} < 0, \tag{8.20}$$

As in the proof of Theorem 8.1 and using $\mathbf{F}_{22}\mathbf{K} = \mathbf{Y}$ from (8.19), we have $\begin{bmatrix} \mathbf{0}_{(N-q)\times n} & \mathbf{0}_{(N-q)\times 2l} \\ \mathbf{YC}_2 & \mathbf{0}_{q\times 2l} \end{bmatrix} = \mathbf{FT}\begin{bmatrix} \mathbf{B}_2\mathbf{D}_c\mathbf{C}_2 & \mathbf{0}_{n\times 2l} \\ \mathbf{B}_c\mathbf{C}_2 & \mathbf{0}_{2l\times 2l} \end{bmatrix}$ and $\begin{bmatrix} \mathbf{0}_{(N-q)\times h} \\ \mathbf{YD}_{21} \end{bmatrix} = \mathbf{FT}\begin{bmatrix} \mathbf{B}_2\mathbf{D}_c\mathbf{D}_{21} \\ \mathbf{B}_c\mathbf{D}_{21} \end{bmatrix}$. We also have that $\begin{bmatrix} \mathbf{0}_{(N-q)\times n} & \mathbf{0}_{(N-q)\times 2l} \\ \mathbf{YC}_2 & \mathbf{0}_{q\times 2l} \end{bmatrix} + \mathbf{FT}\tilde{\mathbf{A}} = \mathbf{FTA}_{\mathrm{cl}}$, $\begin{bmatrix} \mathbf{0}_{(N-q)\times h} \\ \mathbf{YD}_{21} \end{bmatrix} + \mathbf{FT}\tilde{\mathbf{B}} = \mathbf{FTB}_{\mathrm{cl}}$, $\mathbf{G}_i\mathbf{T}\tilde{\mathbf{A}} = \mathbf{G}_i\mathbf{TA}_{\mathrm{cl}}$, and $\mathbf{G}_i\mathbf{T}\tilde{\mathbf{B}} = \mathbf{G}_i\mathbf{TB}_{\mathrm{cl}}$, $i = 1, \ldots, 3$. It follows from (8.20) that

$$\begin{bmatrix} \mathbf{P} & \mathbf{0}_{N\times N} & \mathbf{0}_{N\times h} & \mathbf{0}_{N\times r} \\ \mathbf{0}_{N\times N} & -\mathbf{P} & \mathbf{0}_{N\times h} & \mathbf{0}_{N\times r} \\ \mathbf{0}_{h\times N} & \mathbf{0}_{h\times N} & -\gamma^2\mathbf{I}_{h\times h} & \mathbf{0}_{h\times r} \\ \mathbf{0}_{r\times N} & \mathbf{0}_{r\times N} & \mathbf{0}_{r\times h} & \mathbf{I}_{r\times r} \end{bmatrix}$$

$$+ \begin{bmatrix} \mathbf{FT} & \mathbf{H}_1 \\ \mathbf{G}_1\mathbf{T} & \mathbf{H}_2 \\ \mathbf{G}_2\mathbf{T} & \mathbf{H}_3 \\ \mathbf{G}_3\mathbf{T} & \mathbf{H}_4 \end{bmatrix} \begin{bmatrix} -\mathbf{I}_{N\times N} & \mathbf{A}_{\mathrm{cl}} & \mathbf{B}_{\mathrm{cl}} & \mathbf{0}_{N\times r} \\ \mathbf{0}_{r\times N} & \mathbf{C}_{\mathrm{cl}} & \mathbf{D}_{\mathrm{cl}} & -\mathbf{I}_{r\times r} \end{bmatrix} + \begin{bmatrix} -\mathbf{I}_{N\times N} & \mathbf{A}_{\mathrm{cl}} & \mathbf{B}_{\mathrm{cl}} & \mathbf{0}_{N\times r} \\ \mathbf{0}_{r\times N} & \mathbf{C}_{\mathrm{cl}} & \mathbf{D}_{\mathrm{cl}} & -\mathbf{I}_{r\times r} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \mathbf{FT} & \mathbf{H}_1 \\ \mathbf{G}_1\mathbf{T} & \mathbf{H}_2 \\ \mathbf{G}_2\mathbf{T} & \mathbf{H}_3 \\ \mathbf{G}_3\mathbf{T} & \mathbf{H}_4 \end{bmatrix}^{\mathrm{T}} < 0 \tag{8.21}$$

Multiplying (8.21) from the left- and right-hand sides by $\begin{bmatrix} \mathbf{x}(k+1) \\ \mathbf{x}(k) \\ \mathbf{w}(k) \\ \mathbf{e}(k) \end{bmatrix}^{\mathrm{T}}$ and $\begin{bmatrix} \mathbf{x}(k+1) \\ \mathbf{x}(k) \\ \mathbf{w}(k) \\ \mathbf{e}(k) \end{bmatrix}$, respectively, we obtain (8.17). Hence, (8.16) is satisfied. □

Remark 8.6 A multi-objective controller design problem involving both the H_∞ and the pole-clustering constraints can be solved using (8.12) and (8.18) simultaneously by setting $\mathbf{J}_{22} = \mathbf{F}_{22}$ and $\mathbf{X} = \mathbf{Y}$. The same idea can be applied to the case where the H_2 or generalized H_2 controller design constraints to be presented in the following sections are considered with the pole-clustering constraint. □

8.1.4 PID Controller Design with an H_2 Performance Constraint

Consider the closed-loop system (8.4). The H_2 norm of $\mathbf{T}_{ew}$ is given by $\|\mathbf{T}_{ew}\|_2^2 = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{Tr}\left(\mathbf{T}_{ew}(e^{j\omega})^* \mathbf{T}_{ew}(e^{j\omega})\right) d\omega$. Then we have $\|\mathbf{T}_{ew}\|_2 < \gamma$ for $\gamma > 0$ if and only if $\mathrm{Tr}\left(\mathbf{D}_{\mathrm{cl}}\mathbf{D}_{\mathrm{cl}}^{\mathrm{T}} + \mathbf{C}_{\mathrm{cl}}\mathbf{Z}\mathbf{C}_{\mathrm{cl}}^{\mathrm{T}}\right) < \gamma^2$, where $\mathbf{Z} > 0$ satisfies $\mathbf{A}_{\mathrm{cl}}\mathbf{Z}\mathbf{A}_{\mathrm{cl}}^{\mathrm{T}} - \mathbf{Z} + \mathbf{B}_{\mathrm{cl}}\mathbf{B}_{\mathrm{cl}}^{\mathrm{T}} < 0$ (Zhou et al. 1995). The following result, obtained using the Schur complement formula, can then be used to calculate the H_2 norm.

Lemma 8.2 *Consider the system (8.4). Then $\|\mathbf{T}_{ew}\|_2 < \gamma$ if and only if there exist positive definite matrices $\mathbf{Z} \in \mathbb{R}^{N\times N}$ and $\mathbf{S} \in \mathbb{R}^{r\times r}$ such that*

$$\begin{bmatrix} -\mathbf{Z} & \mathbf{Z}\mathbf{A}_{\mathrm{cl}}^{\mathrm{T}} & \mathbf{0}_{N\times h} \\ \mathbf{A}_{\mathrm{cl}}\mathbf{Z} & -\mathbf{Z} & \mathbf{B}_{cl} \\ \mathbf{0}_{h\times N} & \mathbf{B}_{\mathrm{cl}}^{\mathrm{T}} & -\mathbf{I}_{h\times h} \end{bmatrix} < 0 \tag{8.22}$$

$$\begin{bmatrix} \mathbf{Z} & \mathbf{0}_{N\times h} & \mathbf{Z}\mathbf{C}_{\mathrm{cl}}^{\mathrm{T}} \\ \mathbf{0}_{h\times N} & \mathbf{I}_{h\times h} & \mathbf{D}_{\mathrm{cl}}^{\mathrm{T}} \\ \mathbf{C}_{\mathrm{cl}}\mathbf{Z} & \mathbf{D}_{\mathrm{cl}} & \mathbf{S} \end{bmatrix} > 0 \tag{8.23}$$

$$Tr\,(\mathbf{S}) < \gamma^2 \tag{8.24}$$

□

The design of a PID controller that satisfies a closed-loop H_2 norm constraint can be performed using the following result.

Theorem 8.4 *Consider the system (8.1) where Assumption 8.1 is satisfied. If there exist a positive-definite matrix* $\mathbf{P} \in \mathbb{R}^{N\times N}$, $\mathbf{F} = \begin{bmatrix} \mathbf{F}_{11} & \mathbf{0}_{(N-q)\times 3l} \\ \mathbf{F}_{21} & \mathbf{F}_{22} \end{bmatrix} \in \mathbb{R}^{N\times N}$, $\mathbf{F}_{22} \in \mathbb{R}^{q\times 3l}$, $1 \le q \le 3l$, $\mathbf{G}_1 = \begin{bmatrix} \mathbf{G}_{11} & \mathbf{0} \end{bmatrix} \in R^{N\times N}$, $\mathbf{G}_{11} \in \mathbb{R}^{N\times(N-3l)}$, $\mathbf{G}_2 = \begin{bmatrix} \mathbf{G}_{21} & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{h\times N}$, $\mathbf{G}_{21} \in \mathbb{R}^{h\times(N-3l)}$, $\gamma > 0$, *and* $\mathbf{Y} \in \mathbb{R}^{q\times m}$ *such that*

$$\begin{bmatrix} \mathbf{P} - \mathbf{F}\mathbf{T} - \mathbf{T}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}} & * & * \\ \mathbf{N}_1^{\mathrm{T}} - \mathbf{G}_1\mathbf{T} + \tilde{\mathbf{A}}^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}} & -\mathbf{P} + \mathbf{G}_1\mathbf{T}\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}\mathbf{G}_1^{\mathrm{T}} & * \\ \mathbf{N}_2^{\mathrm{T}} - \mathbf{G}_2\mathbf{T} + \tilde{\mathbf{B}}^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}} & \mathbf{G}_2\mathbf{T}\tilde{\mathbf{A}} + \tilde{\mathbf{B}}^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}\mathbf{G}_1^{\mathrm{T}} & -\mathbf{I} + \mathbf{G}_2\mathbf{T}\tilde{\mathbf{B}} + \tilde{\mathbf{B}}^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}\mathbf{G}_2^{\mathrm{T}} \end{bmatrix} < 0 \tag{8.25}$$

$$\begin{bmatrix} \mathbf{P} & \mathbf{0}_{N\times h} & \mathbf{C}_{\mathrm{cl}}^{\mathrm{T}} \\ \mathbf{0}_{h\times N} & \mathbf{I}_{h\times h} & \mathbf{D}_{\mathrm{cl}}^{\mathrm{T}} \\ \mathbf{C}_{\mathrm{cl}} & \mathbf{D}_{\mathrm{cl}} & \mathbf{S} \end{bmatrix} > 0 \tag{8.26}$$

$$Tr\,(\mathbf{S}) < \gamma^2 \tag{8.27}$$

where $\mathbf{N}_1 = \begin{bmatrix} \mathbf{0}_{(N-q)\times n} & \mathbf{0}_{(N-q)\times 2l} \\ \mathbf{Y}\mathbf{C}_2 & \mathbf{0}_{q\times 2l} \end{bmatrix}$ *and* $\mathbf{N}_2 = \begin{bmatrix} \mathbf{0}_{(N-q)\times h} \\ \mathbf{Y}\mathbf{D}_{21} \end{bmatrix}$, *and if there exists* $\mathbf{K}$ *such that*

$$\mathbf{F}_{22}\mathbf{K} = \mathbf{Y} \tag{8.28}$$

then the H_2 *norm of the closed-loop system (8.4) satisfies* $\|\mathbf{T}_{ew}\|_2 < \gamma$. □

Proof Inequality (8.25) is equivalent to

$$\begin{bmatrix} \mathbf{P} & \begin{bmatrix} \mathbf{0}_{(N-q)\times n} & \mathbf{0}_{(N-q)\times 2l} \\ \mathbf{Y}\mathbf{C}_2 & \mathbf{0}_{q\times 2l} \end{bmatrix} & \begin{bmatrix} \mathbf{0}_{(N-q)\times h} \\ \mathbf{Y}\mathbf{D}_{21} \end{bmatrix} \\ \begin{bmatrix} \mathbf{0}_{(N-q)\times n} & \mathbf{0}_{(N-q)\times 2l} \\ \mathbf{Y}\mathbf{C}_2 & \mathbf{0}_{q\times 2l} \end{bmatrix}^{\mathrm{T}} & -\mathbf{P} & \mathbf{0}_{N\times h} \\ \begin{bmatrix} \mathbf{0}_{(N-q)\times h} \\ \mathbf{Y}\mathbf{D}_{21} \end{bmatrix}^{\mathrm{T}} & \mathbf{0}_{h\times N} & -\mathbf{I}_{h\times h} \end{bmatrix} + \begin{bmatrix} \mathbf{F} \\ \mathbf{G}_1 \\ \mathbf{G}_2 \end{bmatrix} \mathbf{T} \begin{bmatrix} -\mathbf{I}_{N\times N} & \tilde{\mathbf{A}} & \tilde{\mathbf{B}} \end{bmatrix} + \begin{bmatrix} -\mathbf{I}_{N\times N} & \tilde{\mathbf{A}} & \tilde{\mathbf{B}} \end{bmatrix}^{\mathrm{T}} \mathbf{T}^{\mathrm{T}} \begin{bmatrix} \mathbf{F} \\ \mathbf{G}_1 \\ \mathbf{G}_2 \end{bmatrix}^{\mathrm{T}} < 0 \tag{8.29}$$

Using $\mathbf{F}_{22}\mathbf{K} = \mathbf{Y}$ from (8.28), we have $\begin{bmatrix} \mathbf{0}_{(N-q)\times n} & \mathbf{0}_{(N-q)\times 2l} \\ \mathbf{Y}\mathbf{C}_2 & \mathbf{0}_{q\times 2l} \end{bmatrix} = \mathbf{F}\mathbf{T}\begin{bmatrix} \mathbf{B}_2\mathbf{D}_c\mathbf{C}_2 & \mathbf{0}_{n\times 2l} \\ \mathbf{B}_c\mathbf{C}_2 & \mathbf{0}_{2l\times 2l} \end{bmatrix}$ and $\begin{bmatrix} \mathbf{0}_{(N-q)\times h} \\ \mathbf{Y}\mathbf{D}_{21} \end{bmatrix} = \mathbf{F}\mathbf{T}\begin{bmatrix} \mathbf{B}_2\mathbf{D}_c\mathbf{D}_{21} \\ \mathbf{B}_c\mathbf{D}_{21} \end{bmatrix}$. We also have that

$\begin{bmatrix} \mathbf{0}_{(N-q)\times n} & \mathbf{0}_{(N-q)\times 2l} \\ \mathbf{YC}_2 & \mathbf{0}_{q\times 2l} \end{bmatrix} + \mathbf{FT}\tilde{\mathbf{A}} = \mathbf{FTA}_{\mathrm{cl}}$, $\begin{bmatrix} \mathbf{0}_{(N-q)\times h} \\ \mathbf{YD}_{21} \end{bmatrix} + \mathbf{FT}\tilde{\mathbf{B}} = \mathbf{FTB}_{\mathrm{cl}}$, $\mathbf{G}_i\mathbf{T}\tilde{\mathbf{A}} = \mathbf{G}_i\mathbf{TA}_{\mathrm{cl}}$, and $\mathbf{G}_i\mathbf{T}\tilde{\mathbf{B}} = \mathbf{G}_i\mathbf{TB}_{\mathrm{cl}}$, $i = 1, 2$. It follows from (8.29) that

$$\begin{bmatrix} \mathbf{P} & \mathbf{0}_{N\times N} & \mathbf{0}_{N\times h} \\ \mathbf{0}_{N\times N} & -\mathbf{P} & \mathbf{0}_{N\times h} \\ \mathbf{0}_{h\times N} & \mathbf{0}_{h\times N} & -\mathbf{I}_{h\times h} \end{bmatrix} + \begin{bmatrix} \mathbf{F} \\ \mathbf{G}_1 \\ \mathbf{G}_2 \end{bmatrix} \mathbf{T} \begin{bmatrix} -\mathbf{I}_{N\times N} & \mathbf{A}_{\mathrm{cl}} & \mathbf{B}_{\mathrm{cl}} \end{bmatrix} + \begin{bmatrix} -\mathbf{I}_{N\times N} & \mathbf{A}_{\mathrm{cl}} & \mathbf{B}_{\mathrm{cl}} \end{bmatrix}^{\mathrm{T}} \mathbf{T}^{\mathrm{T}} \begin{bmatrix} \mathbf{F} \\ \mathbf{G}_1 \\ \mathbf{G}_2 \end{bmatrix}^{\mathrm{T}} < 0 \tag{8.30}$$

Multiplying (8.30) from the left- and right-hand sides by $\begin{bmatrix} \mathbf{x}(k+1) \\ \mathbf{x}(k) \\ \mathbf{w}(k) \end{bmatrix}^{\mathrm{T}}$ and $\begin{bmatrix} \mathbf{x}(k+1) \\ \mathbf{x}(k) \\ \mathbf{w}(k) \end{bmatrix}$, respectively, we have

$$\mathbf{x}(k+1)^{\mathrm{T}}\mathbf{P}\mathbf{x}(k+1) < \mathbf{x}(k)^{\mathrm{T}}\mathbf{P}\mathbf{x}(k) + \mathbf{w}(k)^{\mathrm{T}}\mathbf{w}(k) \tag{8.31}$$

which is equivalent to

$$\begin{bmatrix} \mathbf{x}(k) \\ \mathbf{w}(k) \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \mathbf{A}_{\mathrm{cl}}^{\mathrm{T}}\mathbf{P}\mathbf{A}_{\mathrm{cl}} - \mathbf{P} & \mathbf{A}_{\mathrm{cl}}^{\mathrm{T}}\mathbf{P}\mathbf{B}_{\mathrm{cl}} \\ \mathbf{B}_{\mathrm{cl}}^{\mathrm{T}}\mathbf{P}\mathbf{A}_{\mathrm{cl}} & \mathbf{B}_{\mathrm{cl}}^{\mathrm{T}}\mathbf{P}\mathbf{B}_{\mathrm{cl}} - \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x}(k) \\ \mathbf{w}(k) \end{bmatrix} < 0$$

Then we have $\begin{bmatrix} \mathbf{A}_{\mathrm{cl}}^{\mathrm{T}}\mathbf{P}\mathbf{A}_{\mathrm{cl}} - \mathbf{P} & \mathbf{A}_{\mathrm{cl}}^{\mathrm{T}}\mathbf{P}\mathbf{B}_{\mathrm{cl}} \\ \mathbf{B}_{\mathrm{cl}}^{\mathrm{T}}\mathbf{P}\mathbf{A}_{\mathrm{cl}} & \mathbf{B}_{\mathrm{cl}}^{\mathrm{T}}\mathbf{P}\mathbf{B}_{\mathrm{cl}} - \mathbf{I} \end{bmatrix} < 0$, which is equivalent to

$$\begin{bmatrix} -\mathbf{P} & \mathbf{0}_{N\times h} & \mathbf{A}_{\mathrm{cl}}^{\mathrm{T}}\mathbf{P} \\ \mathbf{0}_{h\times N} & -\mathbf{I}_{h\times h} & \mathbf{B}_{\mathrm{cl}}^{\mathrm{T}}\mathbf{P} \\ \mathbf{P}\mathbf{A}_{\mathrm{cl}} & \mathbf{P}\mathbf{B}_{\mathrm{cl}} & -\mathbf{P} \end{bmatrix} < 0 \tag{8.32}$$

Consider a congruence transformation on (8.32) by multiplying by $\begin{bmatrix} \mathbf{Z} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \\ \mathbf{0} & \mathbf{Z} & \mathbf{0} \end{bmatrix}$ on the right-hand side and $\begin{bmatrix} \mathbf{Z} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Z} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \end{bmatrix}$ on left-hand side, where $\mathbf{Z} = \mathbf{P}^{-1}$, then we immediately have (8.22). Similarly, performing a congruence transformation on

(8.26) using $\begin{bmatrix} \mathbf{Z} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix}$, we get (8.23). Therefore, based on Lemma 8.2, we obtain $\|\mathbf{T}_{ew}\|_2 < \gamma$. □

8.1.5 Extension to Generalized H_2 Control

The objective of this section is to design a PID controller for the system (8.1) such that a generalized H_2 performance level is achieved. Consider the closed-loop system (8.4) with zero initial condition $\mathbf{x}(0) = 0$, and assume $\mathbf{D}_{11} = \mathbf{0}_{r\times h}$. Given a $\gamma > 0$, the generalized H_2 performance of the closed-loop system (8.4) is defined as the l_2 to l_∞ induced norm of the closed-loop system and is given by Rotea (1993):

$$\|\mathbf{T}_{ew}\|_{2,\infty} := \sup_{0<\|\mathbf{w}\|_2<\infty} \frac{\|\mathbf{e}\|_\infty}{\|\mathbf{w}\|_2} < \gamma \tag{8.33}$$

Then we have the following result to design PID controllers that yield closed-loop systems satisfying a generalized H_2 performance constraint.

Theorem 8.5 *Consider the system (8.1) where Assumption 8.1 is satisfied. If there exist a positive-definite matrix* $\mathbf{P} \in \mathbb{R}^{N\times N}$, $\mathbf{F} = \begin{bmatrix} \mathbf{F}_{11} & \mathbf{0}_{(N-q)\times 3l} \\ \mathbf{F}_{21} & \mathbf{F}_{22} \end{bmatrix} \in \mathbb{R}^{N\times N}$, $\mathbf{F}_{22} \in \mathbb{R}^{q\times 3l}$, $1 \le q \le 3l$, $\mathbf{G}_1 = \begin{bmatrix} \mathbf{G}_{11} & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{N\times N}$, $\mathbf{G}_{11} \in \mathbb{R}^{N\times(N-3l)}$, $\mathbf{G}_2 = \begin{bmatrix} \mathbf{G}_{21} & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{h\times N}$, $\mathbf{G}_{21} \in \mathbb{R}^{h\times(N-3l)}$, $\gamma > 0$, *and* $\mathbf{Y} \in \mathbb{R}^{q\times m}$ *such that*

$$\begin{bmatrix} \mathbf{P} - \mathbf{F}\mathbf{T} - \mathbf{T}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}} & * & * \\ \mathbf{N}_1^{\mathrm{T}} - \mathbf{G}_1\mathbf{T} + \tilde{\mathbf{A}}^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}} & -\mathbf{P} + \mathbf{G}_1\mathbf{T}\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}\mathbf{G}_1^{\mathrm{T}} & * \\ \mathbf{N}_2^{\mathrm{T}} - \mathbf{G}_2\mathbf{T} + \tilde{\mathbf{B}}^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}\mathbf{F}^{T} & \mathbf{G}_2\mathbf{T}\tilde{\mathbf{A}} + \tilde{\mathbf{B}}^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}\mathbf{G}_1^{\mathrm{T}} & -\mathbf{I} + \mathbf{G}_2\mathbf{T}\tilde{\mathbf{B}} + \tilde{\mathbf{B}}^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}\mathbf{G}_2^{\mathrm{T}} \end{bmatrix} < 0 \tag{8.34}$$

$$\begin{bmatrix} -\mathbf{P} & \mathbf{C}_{\mathrm{cl}}^{\mathrm{T}} \\ \mathbf{C}_{\mathrm{cl}} & -\gamma^2\mathbf{I} \end{bmatrix} < 0 \tag{8.35}$$

where $\mathbf{N}_1 = \begin{bmatrix} \mathbf{0}_{(N-q)\times n} & \mathbf{0}_{(N-q)\times 2l} \\ \mathbf{Y}\mathbf{C}_2 & \mathbf{0}_{q\times 2l} \end{bmatrix}$ *and* $\mathbf{N}_2 = \begin{bmatrix} \mathbf{0}_{(N-q)\times h} \\ \mathbf{Y}\mathbf{D}_{21} \end{bmatrix}$, *and if there exists* $\mathbf{K}$ *such that*

$$\mathbf{F}_{22}\mathbf{K} = \mathbf{Y} \tag{8.36}$$

then the generalized H_2 *norm of the closed-loop system (8.4) satisfies (8.33).* □

Proof Using the same procedure as that presented in the proof of Theorem 8.4, (8.34) implies that

$$\mathbf{x}(k+1)^{\mathrm{T}}\mathbf{P}\mathbf{x}(k+1) < \mathbf{x}(k)^{\mathrm{T}}\mathbf{P}\mathbf{x}(k) + \mathbf{w}(k)^{\mathrm{T}}\mathbf{w}(k) \tag{8.37}$$

Sum up (8.37) from $k=0$ to $\mathrm{k}=\mathrm{T}-1>0$, and with the initial condition $\mathbf{x}(0)=0$, we have

$$\mathbf{x}(T)^{\mathrm{T}}\mathbf{P}\mathbf{x}(T) < \sum_{k=0}^{T-1}\mathbf{w}(k)^{\mathrm{T}}\mathbf{w}(k) \tag{8.38}$$

If (8.35) is satisfied, then we have

$$-\gamma^2\mathbf{P} + \mathbf{C}_{\mathrm{cl}}^{\mathrm{T}}\mathbf{C}_{\mathrm{cl}} < 0$$

which implies

$$\mathbf{e}^{\mathrm{T}}(k)\mathbf{e}(k) < \gamma^2\mathbf{x}^{\mathrm{T}}(k)\mathbf{P}\mathbf{x}(k) \tag{8.39}$$

Then based on (8.38), taking the supremum in (8.39) over $k > 0$, we have

$$\|\mathbf{e}\|_\infty^2 < \gamma^2\|\mathbf{w}\|_2^2$$

Hence, (8.33) is satisfied. □

8.1.6 *The MFDM Surface Shape Control with the Multivariable PID Controller*

The control algorithms presented above are used in this section in a practical application aimed at controlling the surface shape of a magnetic fluid deformable mirror (MFDM) in an adaptive optics (AO) system. Experimental results showing the performance of the resulting closed-loop system with the designed PID controllers are presented.

The control system design is based on an analytical model of the MFDM already presented in Chap. 4. The model represents the MFDM as an MIMO system with the currents applied to the electromagnetic coils as the system inputs and the sampled mirror surface deflections as the system outputs. To test the performance of the proposed control algorithms, a modified closed-loop system as shown in Fig. 8.1, is considered. In this system, a laser beam with a planar wavefront is directed toward the deformable mirror with surface shape ζ which causes the wavefront shape in the reflected beam to be 2ζ. A hypothetical aberrated wavefront shape $\mathbf{r}$ is stored inside the control computer and is sampled at the same locations at which the

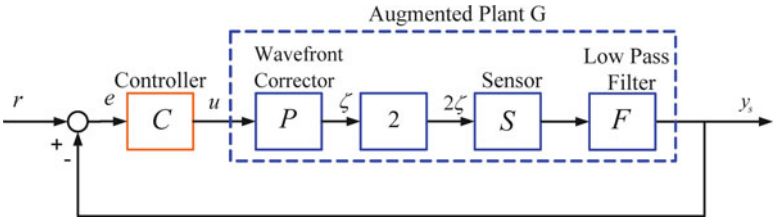


Fig. 8.1 Block diagram of the closed-loop AO system

measured wavefront shape vector $\mathbf{y}_s$ is sampled. The signal $\mathbf{r}$ serves as a reference wavefront shape in the closed-loop operation of the system. It should be noted that the MFDM surface is required to track half the aberrated wavefront shape signals $\mathbf{r}$. The wavefront shape error $\mathbf{e}$ fed to the controller $\mathbf{C}$ can be written as

$$\mathbf{e} = \mathbf{r} - \mathbf{y}_s \tag{8.40}$$

With respect to the feedback system shown in Fig. 8.1, it is desired to design a PID controller $\mathbf{C}$ to provide input currents $\mathbf{u}$ such that the wavefront shape signals $\mathbf{y}_s$ track the reference wavefront shape signals $\mathbf{r}$ which represent the aberrated wavefront shape. The operation of the system shown in Fig. 8.1 is along the same lines as that of a typical adaptive optics system. The system allows for the generation of the aberrated wavefront shapes analytically inside the control computer, hence simplifying the testing of the control algorithms.

The development of the control algorithms is based on an analytical model of the augmented system $\mathbf{G}$, as shown in Fig. 8.1. The analytical model is developed by considering a transfer function $\mathbf{P}(z)$ of the MFDM, representing the sensor dynamics as a unit time step delay $1/z$ and incorporating a low-pass filter dynamics $\mathbf{F}(z)$ to filter out high-frequency measurement noise in the signal $\mathbf{y}_s$. The augmented plant $\mathbf{G}$, as shown in Fig. 8.1, can then be represented using the transfer function $\mathbf{G}(z) = 2\frac{1}{z}\mathbf{P}(z)\mathbf{F}(z)$ and is given in state-space form as:

$$\mathbf{G}: \begin{cases} \mathbf{x}(k+1) = \mathbf{A}_g\mathbf{x}(k) + \mathbf{B}_g\mathbf{u}(k) \\ \mathbf{y}_s(k) = \mathbf{C}_g\mathbf{x}(k) \end{cases} \tag{8.41}$$

where $\mathbf{A}_g$, $\mathbf{B}_g$, and $\mathbf{C}_g$ are system matrices with appropriate dimensions. In this experimental evaluation, the system model has 19 inputs and 19 outputs, and the dimension of the state vector is truncated as 19.

The control objective is for the signal $\mathbf{y}_s$ to track a reference signal $\mathbf{r}$. It follows that the model (8.1) to be used in the controller design can be obtained by setting $\mathbf{A} = \mathbf{A}_g$, $\mathbf{B}_2 = \mathbf{B}_g$, $\mathbf{C}_1 = \mathbf{C}_2 = -\mathbf{C}_g$, and $\mathbf{D}_{11} = \mathbf{D}_{21} = \mathbf{I}$. In the following, the proposed multivariable PID controller design methods that yield closed-loop systems with H_∞ and H_2 performance are both evaluated. Consider $a = 0$ and $b = 0.91$ first; the controller parameters of the H_∞ PID controller are determined

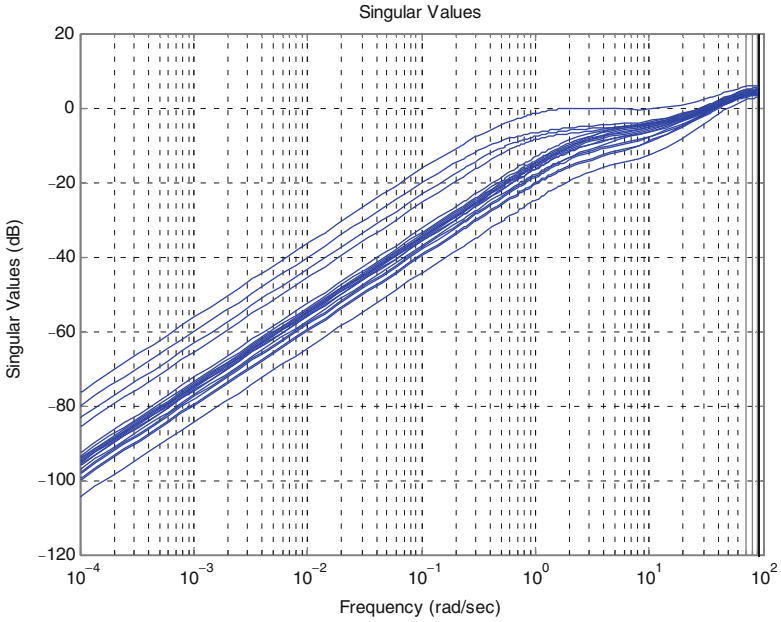


Fig. 8.2 The singular values of the closed-loop system with the H_∞ PID controller

using (8.18) by minimizing the H_∞ performance variable γ. The singular values of the resulting closed-loop system with the designed H_∞ PID controller are shown in Fig. 8.2. In retinal imaging applications, the reference signal $\mathbf{r}$ usually has a low-frequency content typically ranging from 0 to 3 Hz (Hofer et al. 2001). As seen in Fig. 8.2, the closed-loop system with the designed H_∞ PID controller has a good performance in the specified low-frequency range. Similarly, the parameters of the H_2 PID controller are determined using (8.25), (8.26), and (8.27) by minimizing the H_2 performance variable γ. In the following, both PID controllers are evaluated in the experimental setup.

The experimental setup consists mainly of an MFDM wavefront corrector, a Shack–Hartmann-type wavefront sensor used to measure the shape of the wavefront reflected off the MFDM, a control computer running the proposed control algorithms, and a laser diode with 661 nm wavelength used as the reference light. The wavefront corrector is the prototype MFDM consisting of a 60-mm-diameter, 1-mm-thick layer of EFH1 magnetic fluid, and an array of 19 electromagnetic coils used to control the shape of the fluid surface. In the experimental evaluation, the reference signal $\mathbf{r}(k)$ is given by

$$\mathbf{r}(k) = \mathbf{r}_0 + \mathbf{r}_1(k) \tag{8.42}$$

Where $\mathbf{r}_0$ is a static signal chosen to represent a typical static reference wavefront shape signal and is given as

$$\mathbf{r}_0 = [2.0, 4.0, 5.0, 1.0, 5.5, 0.5, 3.5, 8.5, 6.0, 3.0, 2.5, 2.0, 3.0, 4.0, -2.0, -4.0, -5.0, 1.0, 6.0]^{\mathrm{T}} (\mu\mathrm{m}) \tag{8.43}$$

The signal $\mathbf{r}_1(k)$ represents the time-varying component of the wavefront aberrations and can be represented as a low-frequency sinusoidal signal or a low-pass filtered random signal to be consistent with the low-frequency content of the aberrations in the eye (Hofer et al. 2001). The performance of the designed H_∞ PID controller in tracking static reference signals is shown in Fig. 8.3. Figure 8.3a shows the time history of the wavefront shape signals $\mathbf{y}_s$ plotted alongside the reference signals given as dotted lines. The root mean square (RMS) of the error computed over the 19 control points is plotted in Fig. 8.3b. Figure 8.4 shows the results for the case where the H_∞ PID controller is used to track a dynamic reference signal comprising the static offset given in (8.43) plus sinusoidal signals with an amplitude of 1 μm and a frequency of 2 Hz. For clarity only selected channels (#2, 3, 15 and 16) are shown. As can be seen in Fig. 8.3a, the closed-loop system outputs (i.e., wavefront shape signals $\mathbf{y}_s$) track the static reference signals fairly well and are rapidly driven to their steady states. The RMS of the wavefront shape error for the static reference signal, as shown in Fig. 8.3b, has a steady-state value of 0.09 ± 0.05 μm. The RMS of the wavefront shape error for the sinusoidal reference signal is shown in Fig. 8.4b. Although there is a significant ripple in the steady-state RMS error, the average values remain around 0.15 μm. Figure 8.5 shows the results for the case of tracking randomly varying reference signals superimposed on the static offset signals given in (8.43). Although the controller does provide for the successful tracking of the reference signals by the MFDM surface, the average RMS error for this case is 0.22 μm. The controller performance with the designed H_2 PID controller is shown in Figs. 8.6 and 8.7. As can be seen in Fig. 8.7b, the RMS error in tracking the randomly varying reference signals decreases to 0.16 μm, showing an improvement over the H_∞ PID controller. The experimental results given above show that the designed controller can effectively compensate for typical high-order wavefront aberrations in the human eye.

8.2 Mixed-Sensitivity H_∞ Controller

This section presents an optimal H_∞ controller designed to cancel dynamic wavefront aberrations in an MFDM-based adaptive optics system. The control algorithm is developed using the mixed-sensitivity H_∞ design approach that enables the tracking of the desired MFDM surface shape and also limits the magnitude of the control currents applied to the MFDM actuators. Experimental results show the performance of the closed-loop system comprising the developed controller and the 19-channel prototype MFDM.

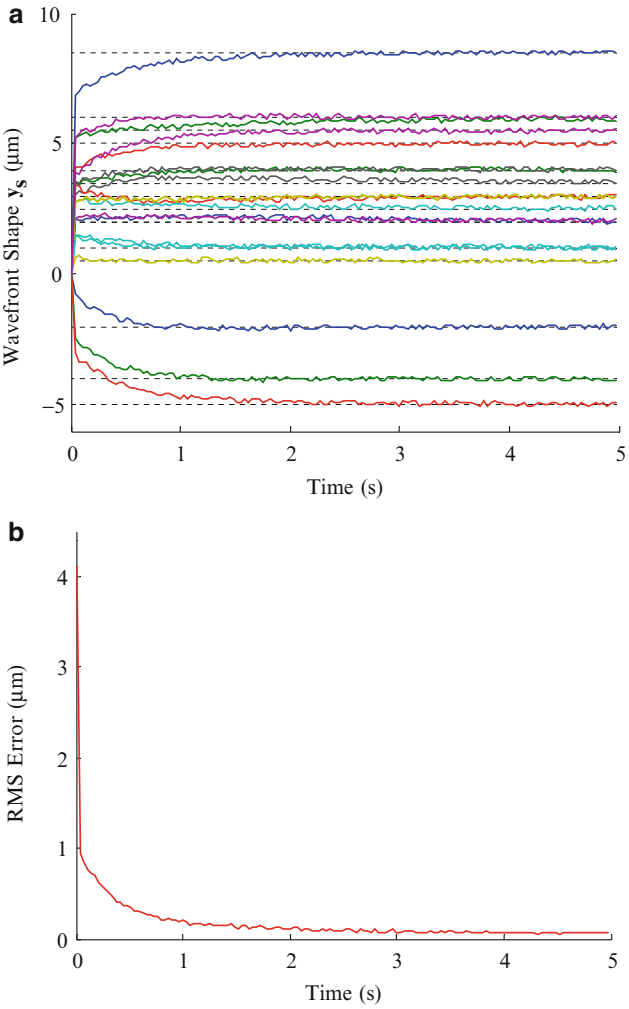


Fig. 8.3 Experimental tracking of the static wavefront shape with the H_∞ PID controller: (**a**) wavefront shape, (**b**) RMS error

The mixed-sensitivity H_∞ controller provides the desired performance in tracking the reference wavefront surface shapes using an MFDM and limits the current inputs to the MFDM. The choice of the H_∞ controller design methodology is motivated by the following two main considerations related to the possible use of this controller in ophthalmic MFDM-based AO imaging systems. First, the temporal frequency content of the aberrations in the human eye is typically limited to within the range of 0–3 Hz. Moreover, the eye's aberrations can be modeled as a combination of static aberration signals, harmonic aberration signals associated with physiological processes such the heartbeat and respiration and that can be

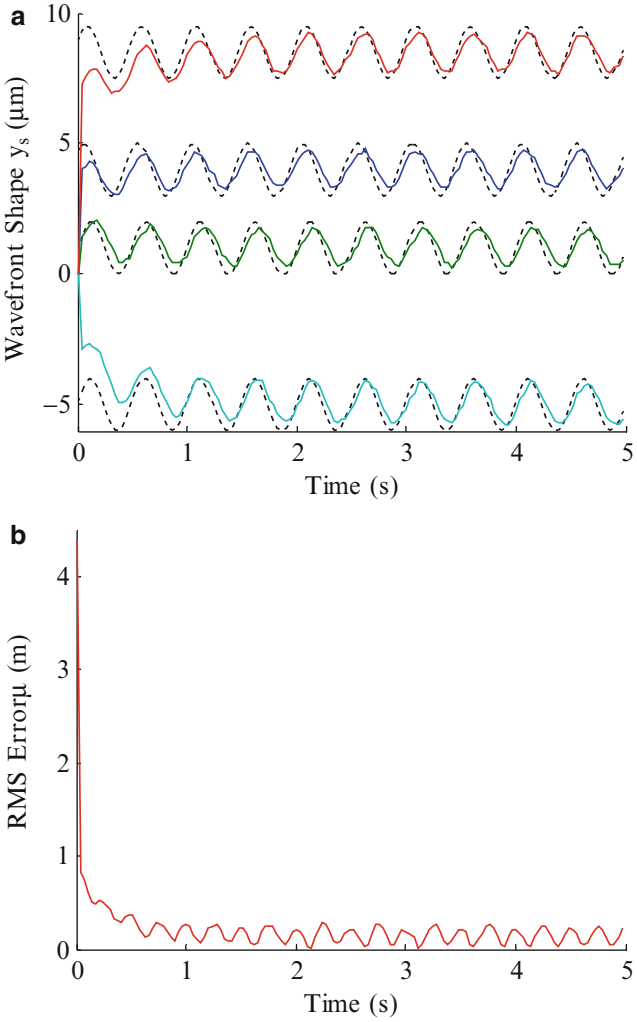


Fig. 8.4 Experimental tracking of the sinusoidal dynamic wavefront shape with the H_∞ PID controller: (**a**) wavefront shape, (**b**) RMS error

represented using sinusoids, and random components. These signals can best be dealt with using an H_∞ type controller to account for the mixed nature of the aberrations (harmonic and random components) as well as for the limited frequency content of the aberrations. The second motivation for considering an H_∞ type controller is that, besides the conventional performance criterion of minimizing the residual wavefront error, the mixed-sensitivity H_∞ controller design is used to minimize the magnitude of the high-frequency components in the controller output, that is, in the current input to the MFDM. Since the electromagnetic coils in the MFDMs can sustain only a limited amount of current, this feature helps

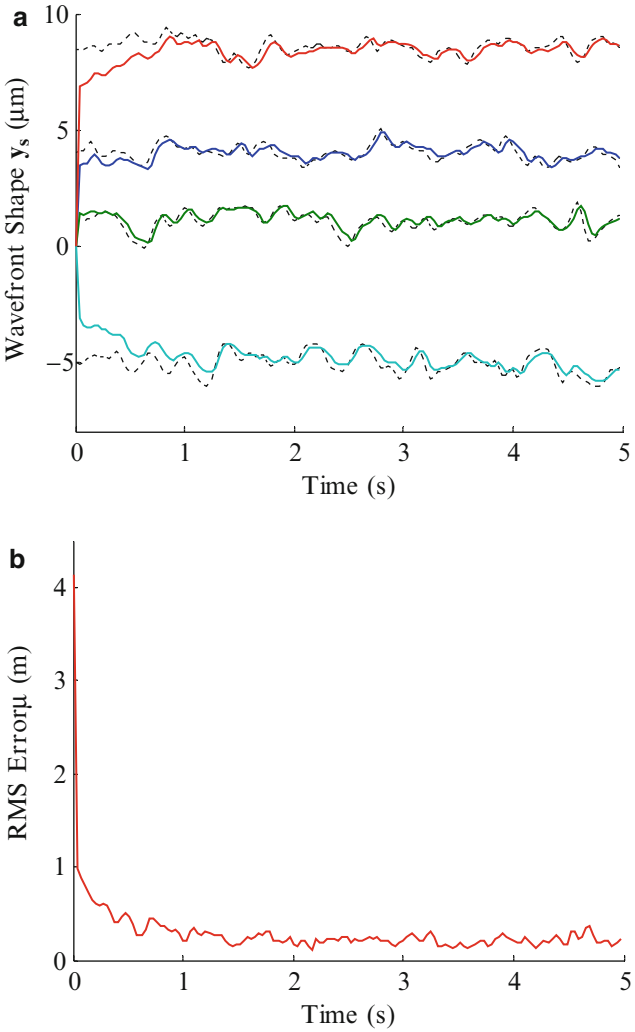


Fig. 8.5 Experimental tracking of the random dynamic wavefront shape with the H_∞ PID controller: (**a**) wavefront shape, (**b**) RMS error

with alleviating the practical problem of actuator saturation. It also contributes toward the robustness of the closed-loop system by minimizing the controller output in the frequency range beyond the desired closed-loop system bandwidth and therefore helps avoid exciting unmodeled high-frequency dynamics in the wavefront corrector. In fact, an analytical model of the MFDM presented in Chap. 4 shows that the system model is infinite dimensional. For practical reasons, the controller is usually designed based on a truncated model where high-frequency dynamics are ignored. Consequently, the mixed-sensitivity H_∞ design method provides the

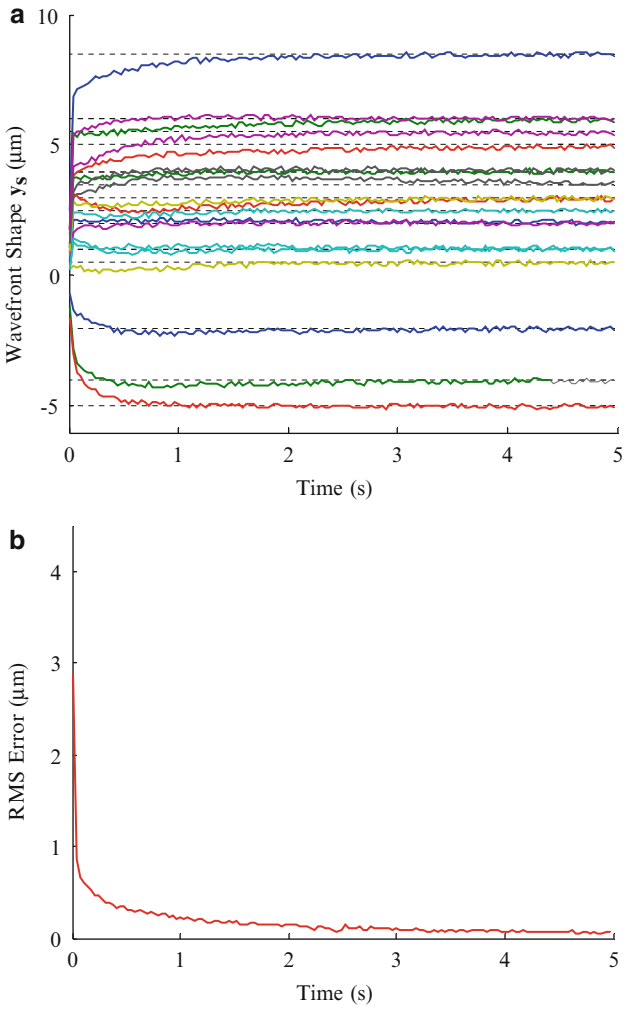


Fig. 8.6 Experimental tracking of the static wavefront shape with the H_2 PID controller: (**a**) wavefront shape, (**b**) RMS error

means to deal with this model uncertainty in the high-frequency range. It is also worth noting that this method can be extended to multi-objective H_∞ control, as shown in this section.

8.2.1 Mixed-Sensitivity H_∞ Control Problem

A control system should remain stable and meet certain performance objectives in the presence of possible uncertainties. The H_∞ optimization approach has been

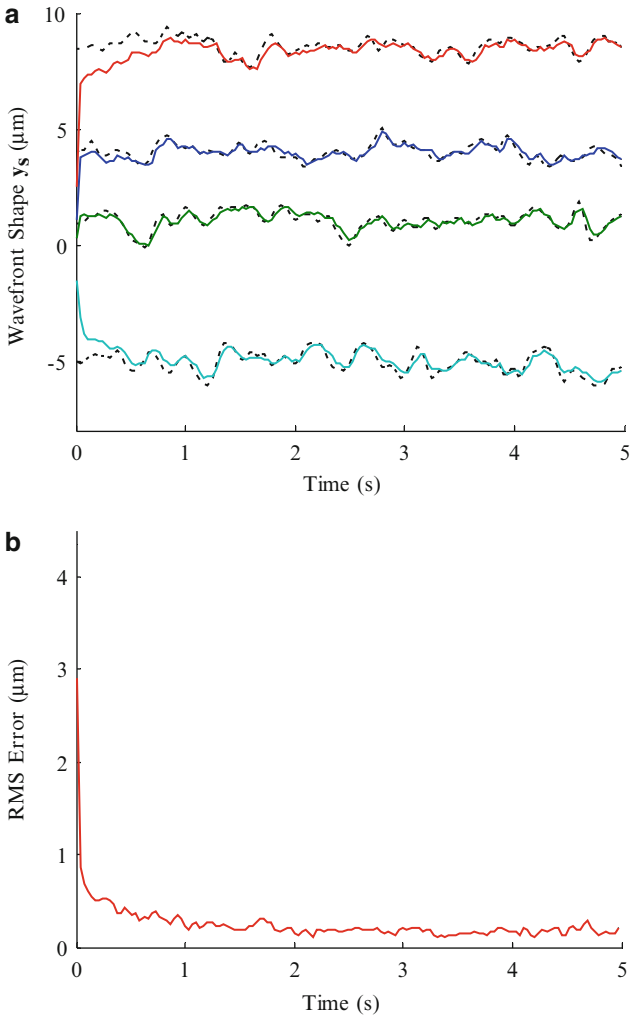


Fig. 8.7 Experimental tracking of the random dynamic wavefront shape with the H_2 PID controller: (**a**) wavefront shape, (**b**) RMS error

shown to be an effective and efficient robust design method for linear, time-invariant control systems. Consider the typical closed-loop system configuration shown in Fig. 8.8, where **G** is the plant, **K** the controller to be designed, **r** the reference input, **y** the system output, **u** the control signal, **e** the performance error signal, **d** the disturbance signal, and **n** the measurement noise, respectively. The input–output relationships are then given by

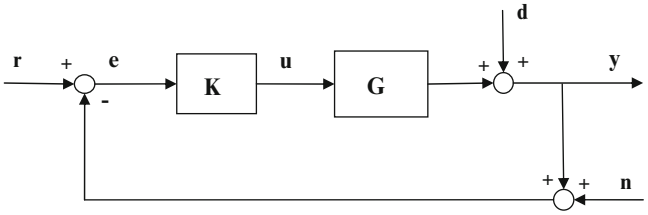


Fig. 8.8 A closed-loop feedback configuration involving the plant **G** and controller **K**

$$\mathbf{y} = (\mathbf{I} + \mathbf{GK})^{-1}\mathbf{GKr} + (\mathbf{I} + \mathbf{GK})^{-1}\mathbf{d} - (\mathbf{I} + \mathbf{GK})^{-1}\mathbf{GKn}$$

$$\mathbf{u} = \mathbf{K}(\mathbf{I} + \mathbf{GK})^{-1}\mathbf{r} - \mathbf{K}(\mathbf{I} + \mathbf{GK})^{-1}\mathbf{d} - \mathbf{K}(\mathbf{I} + \mathbf{GK})^{-1}\mathbf{n}$$

$$\mathbf{e} = (\mathbf{I} + \mathbf{GK})^{-1}\mathbf{r} - (\mathbf{I} + \mathbf{GK})^{-1}\mathbf{d} - (\mathbf{I} + \mathbf{GK})^{-1}\mathbf{n}$$

The usual performance specifications, such as tracking, disturbance attenuation and noise rejection, can be optimized by minimizing the H_∞ norm of corresponding transfer function matrices. For example, minimizing $\left\|(\mathbf{I} + \mathbf{GK})^{-1}\right\|_\infty$ can achieve good tracking and good disturbance attenuation; minimizing $\left\|\mathbf{K}(\mathbf{I} + \mathbf{GK})^{-1}\right\|_\infty$ can achieve less control energy; minimizing $\left\|(\mathbf{I} + \mathbf{GK})^{-1}\mathbf{GK}\right\|_\infty$ can achieve good noise rejection. Let $\mathbf{S} := (\mathbf{I} + \mathbf{GK})^{-1}$ denote the sensitivity function and $\mathbf{T} := (\mathbf{I} + \mathbf{GK})^{-1}\mathbf{GK}$ the complementary sensitivity function. It is easy to obtain a relationship $\mathbf{T} = \mathbf{I} - \mathbf{S}$. In general, weighting functions would be used in the above minimization to meet the design specifications. For instance, instead of minimizing the sensitivity function **S**, the following optimization problem will be formulated:

$$\min_{\mathbf{K}} \|\mathbf{W}_e\mathbf{S}\mathbf{W}_d\|_\infty$$

where $\mathbf{W}_e$ is chosen to tailor the tracking requirement and is usually a high-gain low-pass filter type and $\mathbf{W}_d$ can be regarded as a generator that characterizes all relevant disturbances in the case considered. The weighting functions are usually stable and of minimum phase.

Mixed-sensitivity H_∞ control problem is solved by shaping the corresponding transfer functions **S**, **KS**, and **T** together, where usually the sensitivity function **S** is shaped along with **KS** or the complementary sensitivity function **T**. Consider the regulation problem where it is desired to reject a disturbance **d** entering at the plant output or track a reference **r** and assume that the measurement noise is relatively insignificant. For this problem, it makes sense to shape the closed-loop transfer functions **S** along with **KS** (see Fig. 8.9). Recall that **S** is the transfer function between **d** and the output and **KS** the transfer function between **d** and the control signals; therefore, in order to achieve good disturbance rejection, it is important to include **KS** as a mechanism for limiting the size and bandwidth of the controller

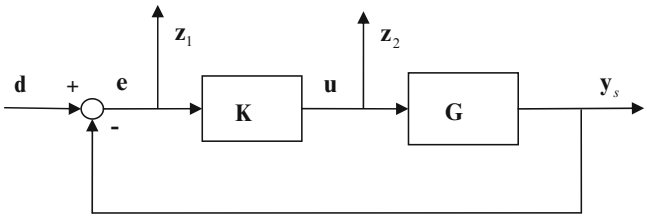
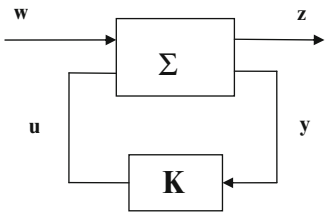


Fig. 8.9 **S/KS** mixed sensitivity optimization configuration

Fig. 8.10 The standard H_∞ configuration



and the control energy used. The size of **KS** is also important for robust stability with respect to plant model uncertainties modeled as additive plant perturbations. Therefore, it is useful in practice to minimize

$$\min_{\mathbf{K}} \left\| \begin{array}{c} \mathbf{W}_e \mathbf{S} \mathbf{W}_d \\ \mathbf{W}_u \mathbf{K} \mathbf{S} \mathbf{W}_d \end{array} \right\|_\infty \tag{8.44}$$

where $\mathbf{W}_u$ is a high-pass filter with a crossover frequency approximately equal to that of the desired closed-loop system bandwidth. This cost function can also be interpreted as the design objectives of nominal performances including good tracking and disturbance attenuation, and robust stabilization with regard to additive perturbations.

In order to adopt a unified solution procedure, the above cost function (8.44) can be formulated in the standard H_∞ configuration as in Fig. 8.10, where all the external inputs are denoted by **w**, **z** denotes the output signals to be minimized that include both performance and robustness, **y** is the vector of measurements available to the controller **K,** and **u** is the vector of control signals. Σ is called the generalized plant or interconnected system. The objective is to find a stabilizing controller **K** to minimize the H_∞ norm of the transfer function from **w** to **z**. Partitioning the interconnected system Σ as

$$\Sigma = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix}$$

it follows then that

$$\begin{aligned} \mathbf{z} &= \left[\mathbf{P}_{11} + \mathbf{P}_{12} \mathbf{K} (\mathbf{I} - \mathbf{P}_{22} \mathbf{K})^{-1} \mathbf{P}_{21} \right] \mathbf{w} \\ &=: F_l(\Sigma, \mathbf{K})\, \mathbf{w} \end{aligned}$$

where $F_l(\Sigma, \mathbf{K})$ is defined as the lower linear fractional transformation of Σ and $\mathbf{K}$. The H_∞ optimization problem (8.44) now is equivalent to

$$\min_{\mathbf{K}} \|F_l(\Sigma, \mathbf{K})\|_\infty \tag{8.45}$$

We can consider the disturbance $\mathbf{d}$ as a single exogenous input $\mathbf{w}$ and define an error signal $\mathbf{z} = \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{W}_u \mathbf{u} \\ \mathbf{W}_e \mathbf{e} \end{bmatrix}$, $\mathbf{e} = \mathbf{d} - \mathbf{y}_s$, $\mathbf{y} = \mathbf{e}$, and $\mathbf{u} = \mathbf{u}$. Then the interconnected system Σ can be presented as

$$\Sigma = \left[\begin{array}{c|c} \mathbf{0} & \mathbf{W}_u \\ \hline \mathbf{W}_e \mathbf{W}_d & -\mathbf{W}_e \mathbf{G} \\ \hline \mathbf{W}_e \mathbf{W}_d & -\mathbf{W}_e \mathbf{G} \end{array}\right]$$

The system of Fig. 8.10 can be described by

$$\begin{aligned} \begin{bmatrix} \mathbf{z} \\ \mathbf{y} \end{bmatrix} &= \sum \begin{bmatrix} \mathbf{w} \\ \mathbf{u} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \mathbf{u} \end{bmatrix} \\ &= \left[\begin{array}{c|c} \mathbf{0} & \mathbf{W}_u \\ \hline \mathbf{W}_e \mathbf{W}_d & -\mathbf{W}_e \mathbf{G} \\ \hline \mathbf{W}_e \mathbf{W}_d & -\mathbf{W}_e \mathbf{G} \end{array}\right] \begin{bmatrix} \mathbf{w} \\ \mathbf{u} \end{bmatrix} \end{aligned}$$

and

$$\mathbf{u} = \mathbf{K}\mathbf{y}$$

Let the state space description of the generalized system Σ in Fig. 8.10 be given by

$$\Sigma : \begin{cases} \mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}_1 \mathbf{w}(k) + \mathbf{B}_2 \mathbf{u}(k) \\ \mathbf{z}(k) = \mathbf{C}_1 \mathbf{x}(k) + \mathbf{D}_{11} \mathbf{w}(k) + \mathbf{D}_{12} \mathbf{u}(k) \\ \mathbf{y}(k) = \mathbf{C}_2 \mathbf{x}(k) + \mathbf{D}_{21} \mathbf{w}(k) \end{cases} \tag{8.46}$$

which is simply presented as

$$\begin{aligned} \Sigma : &= \left[\begin{array}{c|cc} \mathbf{A} & \mathbf{B}_1 & \mathbf{B}_2 \\ \hline \mathbf{C}_1 & \mathbf{D}_{11} & \mathbf{D}_{12} \\ \mathbf{C}_2 & \mathbf{D}_{21} & \mathbf{0} \end{array}\right] \\ &:= \left[\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array}\right] \end{aligned}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the state vector, $\mathbf{w} \in \mathbb{R}^{m_1}$ the exogenous input vector, $\mathbf{u} \in \mathbb{R}^{m_2}$ the control input vector, $\mathbf{z} = \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix} \in \mathbb{R}^{p_1}$ the error vector, and $\mathbf{y} \in \mathbb{R}^{p_2}$ the measurement vector, with $p_1 \geq m_2$, and $p_2 \leq m_1$. According to performance variable $\mathbf{z}_1$ and $\mathbf{z}_2$, $\mathbf{C}_1$, $\mathbf{D}_{11}$ and $\mathbf{D}_{12}$ can be further divided as $\mathbf{C}_1 = \begin{bmatrix} \mathbf{C}_{11} \\ \mathbf{C}_{12} \end{bmatrix}$, $\mathbf{D}_{11} = \begin{bmatrix} \mathbf{D}_{111} \\ \mathbf{D}_{112} \end{bmatrix}$, and $\mathbf{D}_{12} = \begin{bmatrix} \mathbf{D}_{121} \\ \mathbf{D}_{122} \end{bmatrix}$.

8.2.2 Controller Design Based on Algebraic Riccati Equations

The most general and widely used algorithms for H_∞ controller design in a state-space setting have been developed by Glover and Doyle (Zhou et al. 1995) and have been derived based on two algebraic Riccati equations (ARE). In the following, the algorithm for H_∞ controller design for discrete-time systems is presented (Gu et al. 2005; Zhou et al. 1995). The control problem is to find an internally stabilizing controller $\mathbf{K}$ such that for a prespecified positive value γ (representing a bound on the performance level of the closed-loop system), the following performance is satisfied:

$$\|F_l(\Sigma, \mathbf{K})\|_\infty < \gamma \tag{8.47}$$

Consider the following discrete-time Riccati equation

$$\mathbf{X} = \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{0} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \mathbf{I}_{p1} & \mathbf{0} \\ \mathbf{0} & -\gamma^2 \mathbf{I}_{m1} \end{bmatrix} \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{0} \end{bmatrix} + \mathbf{A}^{\mathrm{T}}\mathbf{X}\mathbf{A} - \mathbf{L}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{L} \tag{8.48}$$

where

$$\mathbf{R} = \begin{bmatrix} \mathbf{D}_{11} & \mathbf{D}_{12} \\ \mathbf{I}_{m1} & \mathbf{0} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \mathbf{I}_{p1} & \mathbf{0} \\ \mathbf{0} & -\gamma^2 \mathbf{I}_{m1} \end{bmatrix} \begin{bmatrix} \mathbf{D}_{11} & \mathbf{D}_{12} \\ \mathbf{I}_{m1} & \mathbf{0} \end{bmatrix} + \mathbf{B}^{\mathrm{T}}\mathbf{X}\mathbf{B} =: \begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_2^{\mathrm{T}} \\ \mathbf{R}_2 & \mathbf{R}_3 \end{bmatrix}$$

$$\mathbf{L} = \begin{bmatrix} \mathbf{D}_{11} & \mathbf{D}_{12} \\ \mathbf{I}_{m1} & \mathbf{0} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \mathbf{I}_{p1} & \mathbf{0} \\ \mathbf{0} & -\gamma^2 \mathbf{I}_{m1} \end{bmatrix} \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{0} \end{bmatrix} + \mathbf{B}^{\mathrm{T}}\mathbf{X}\mathbf{A} =: \begin{bmatrix} \mathbf{L}_1 \\ \mathbf{L}_2 \end{bmatrix}$$

Assume that there exist a matrix $\mathbf{V}_{12} \in \mathbb{R}^{m_2 \times m_2}$ such that

$$\mathbf{V}_{12}^{\mathrm{T}}\mathbf{V}_{12} = \mathbf{R}_3$$

and a matrix $\mathbf{V}_{21} \in \mathbb{R}^{m_1 \times m_1}$ such that

$$\mathbf{N} = \mathbf{R}_1 - \mathbf{R}_2^{\mathrm{T}}\mathbf{R}_3^{-1}\mathbf{R}_2 < 0$$

$$\mathbf{V}_{21}^{\mathrm{T}}\mathbf{V}_{21} = -\gamma^2 \mathbf{N}$$

Define the matrices

$$\begin{bmatrix} \mathbf{A}_Y & \mathbf{B}_Y \\ \mathbf{C}_Y & \mathbf{D}_Y \end{bmatrix} =: \left[\begin{array}{c|cc} \mathbf{A}_Y & \mathbf{B}_{Y1} & \mathbf{B}_{Y2} \\ \hline \mathbf{C}_{Y1} & \mathbf{D}_{Y11} & \mathbf{D}_{Y12} \\ \mathbf{C}_{Y2} & \mathbf{D}_{Y21} & \mathbf{D}_{Y22} \end{array}\right]$$

$$= \left[\begin{array}{c|cc} \mathbf{A}-\mathbf{B}_1\mathbf{N}^{-1}\mathbf{F} & \mathbf{B}_1\mathbf{V}_{21}^{-1} & \mathbf{0} \\ \hline \mathbf{V}_{12}\mathbf{R}_3^{-1}\left(\mathbf{L}_2-\mathbf{R}_2\mathbf{N}^{-1}\mathbf{F}\right) & \mathbf{V}_{12}\mathbf{R}_3^{-1}\mathbf{R}_2\mathbf{V}_{21}^{-1} & \mathbf{I} \\ \mathbf{C}_2-\mathbf{D}_{21}\mathbf{N}^{-1}\mathbf{F} & \mathbf{D}_{21}\mathbf{V}_{21}^{-1} & \mathbf{0} \end{array}\right]$$

where

$$\mathbf{F} = \mathbf{L}_1 - \mathbf{R}_2^{\mathrm{T}}\mathbf{R}_3^{-1}\mathbf{L}_2$$

Let $\mathbf{Y}$ be the solution to the following discrete-time Riccati equation:

$$\mathbf{Y} = \mathbf{B}_Y\mathbf{J}\mathbf{B}_Y^{\mathrm{T}} + \mathbf{A}_Y\mathbf{Y}\mathbf{A}_Y^{\mathrm{T}} - \mathbf{M}\mathbf{S}\mathbf{M}^{\mathrm{T}} \tag{8.49}$$

where

$$\mathbf{J} = \begin{bmatrix} \mathbf{I}_{m1} & \mathbf{0} \\ \mathbf{0} & -\gamma^2\mathbf{I}_{m2} \end{bmatrix}$$

$$\mathbf{S} = \mathbf{D}_Y\mathbf{J}\mathbf{D}_Y^{\mathrm{T}} + \mathbf{C}_Y\mathbf{Y}\mathbf{C}_Y^{\mathrm{T}} =: \begin{bmatrix} \mathbf{S}_1 & \mathbf{S}_2 \\ \mathbf{S}_2^{\mathrm{T}} & \mathbf{S}_3 \end{bmatrix}$$

$$\mathbf{M} = \mathbf{B}_Y\mathbf{J}\mathbf{D}_Y^{\mathrm{T}} + \mathbf{A}_Y\mathbf{Y}\mathbf{C}_Y^{\mathrm{T}} =: \begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2 \end{bmatrix}$$

Consider the following assumptions for the generalized system (8.46).

A1: $(\mathbf{A}, \mathbf{B}_2)$ is stabilizable and $(\mathbf{C}_2, \mathbf{A})$ is detectable:

A2: $\begin{bmatrix} \mathbf{A}-e^{j\omega}\mathbf{I} & \mathbf{B}_2 \\ \mathbf{C}_1 & \mathbf{D}_{12} \end{bmatrix}$ has full column rank for all $\omega \in [0, 2\pi)$.

A3: $\begin{bmatrix} \mathbf{A}-\mathrm{e}^{j\omega}\mathbf{I} & \mathbf{B}_1 \\ \mathbf{C}_2 & \mathbf{D}_{21} \end{bmatrix}$ has full row rank for all $\omega \in [0, 2\pi)$.

Based on the above assumptions, the algorithm for the design of an H_∞ controller is summarized as follows:

General H_∞ algorithm: For the general control configuration of Fig. 8.10 described by equation (8.46), with Assumptions A1 to A3, there exists a stabilizing controller $\mathbf{K}$ such that $\|F_l(\Sigma, \mathbf{K})\|_\infty < \gamma$ if and only if

(i) There exists a solution to the Riccati equation (8.48) satisfying

$$\mathbf{X} \geq 0 \quad \text{and} \quad \mathrm{N} < 0$$

such that $\mathbf{A} - \mathbf{B}\mathbf{R}^{-1}\mathbf{L}$ is asymptotically stable.

(ii) There exists a solution to the Riccati equation (8.49) satisfying

$$\mathbf{Y} \geq 0 \quad \text{and} \quad \mathbf{S}_1 - \mathbf{S}_2\mathbf{S}_3^{-1}\mathbf{S}_2^{\mathrm{T}} < 0$$

such that $\mathbf{A}_Y - \mathbf{M}\mathbf{S}^{-1}\mathbf{C}_Y$ is asymptotically stable.

The optimal controller is then given by

$$\mathbf{K}_c := \left[\begin{array}{c|c} \mathbf{A}_c & \mathbf{B}_c \\ \hline \mathbf{C}_c & \mathbf{D}_c \end{array}\right], \tag{8.50}$$

where

$$\begin{aligned}
\mathbf{A}_c &= \mathbf{A}_Y - \mathbf{B}_2\mathbf{V}_{12}^{-1}\left(\mathbf{C}_{Y1} - \mathbf{S}_2\mathbf{S}_3^{-1}\mathbf{C}_{Y2}\right) - \mathbf{M}_2\mathbf{S}_3^{-1}\mathbf{C}_{Y2} \\
\mathbf{B}_c &= -\mathbf{B}_2\mathbf{V}_{12}^{-1}\mathbf{S}_2\mathbf{S}_3^{-1} + \mathbf{M}_2\mathbf{S}_3^{-1} \\
\mathbf{C}_c &= -\mathbf{V}_{12}^{-1}\left(\mathbf{C}_{Y1} - \mathbf{S}_2\mathbf{S}_3^{-1}\mathbf{C}_{Y2}\right) \\
\mathbf{D}_c &= -\mathbf{V}_{12}^{-1}\mathbf{S}_2\mathbf{S}_3^{-1}
\end{aligned}$$

$\mathbf{K}_c$ is the so-called central H_∞ controller that is widely used in practice and has the same state-space dimension of the plant Σ.

The above algorithm can be used only when the bound γ on the closed-loop system performance is known. Therefore, an optimal value of γ cannot be obtained directly. One way to achieve the minimum value of γ to within a specified tolerance is by performing bisection search on γ until its value is sufficiently accurate. However, when γ is approaching its minimum value, the problem would become more and more ill conditioned numerically. Hence, the solution thus obtained might be very unreliable. Furthermore, singular H_∞ problems (those violating (A1)–(A3)) require regularization by small perturbation which makes the algorithm even more complicated. In recent years, linear matrix inequalities (LMIs) have emerged as powerful tool to approach control problems that appear hard if impossible to solve in an analytic fashion. Currently, several commercial and noncommercial software packages are available, which allow a simple coding of general LMI problems to solve typical control problems in a numerically efficient manner. In comparison with the Riccati method for H_∞ problems, the LMI approach is applicable to any plant and does not involve γ-iterations to find the optimal solution. Rather, the H_∞ performance is directly optimized by solving properly formulated LMIs. The LMI-based approach usually involves more unknown variables to be determined and consequently results in a heavier computational burden, especially for systems with large state-space dimension. In the following, the controller design method based on the LMI approach is briefly presented.

8.2.3 Controller Design Based on the LMI Approach

Consider a dynamical output feedback controller $\mathbf{K}$ as

$$\Sigma_K : \begin{cases} \varsigma(k+1) = \mathbf{A}_c\varsigma(k) + \mathbf{B}_c\mathbf{y}(k) \\ \mathbf{u}(k) = \mathbf{C}_c\varsigma + \mathbf{D}_c\mathbf{y}(k) \end{cases} \tag{8.51}$$

With the plant Σ and controller $\mathbf{K}$ defined as above, the closed-loop system is given by the state-space equation

$$\Sigma_{\text{cl}} : \begin{cases} \mathbf{x}_{\text{cl}}(k+1) = \mathbf{A}_{\text{cl}}\mathbf{x}_{\text{cl}}(k) + \mathbf{B}_{\text{cl}}\mathbf{w}(k) \\ \mathbf{z}(k) = \mathbf{C}_{\text{cl}}\mathbf{x}_{\text{cl}}(k) + \mathbf{D}_{\text{cl}}\mathbf{w}(k) \end{cases} \tag{8.52}$$

where

$$\left[\begin{array}{c|c} \mathbf{A}_{\text{cl}} & \mathbf{B}_{\text{cl}} \\ \hline \mathbf{C}_{\text{cl}} & \mathbf{D}_{\text{cl}} \end{array}\right] = \left[\begin{array}{cc|c} \mathbf{A} + \mathbf{B}_2\mathbf{D}_c\mathbf{C}_2 & \mathbf{B}_2\mathbf{C}_c & \mathbf{B}_1 + \mathbf{B}_2\mathbf{D}_c\mathbf{D}_{21} \\ \mathbf{B}_c\mathbf{C}_2 & \mathbf{A}_c & \mathbf{B}_c\mathbf{D}_{21} \\ \hline \mathbf{C}_1 + \mathbf{D}_{12}\mathbf{D}_c\mathbf{C}_2 & \mathbf{D}_{12}\mathbf{C}_c & \mathbf{D}_{11} + \mathbf{D}_{12}\mathbf{D}_c\mathbf{D}_{21} \end{array}\right].$$

Based on the bounded real lemma, it is well known that $\mathbf{A}_{\text{cl}}$ is stable and the H_∞ norm of Σ_{cl} is smaller than γ if and only if there exists a symmetric $\mathbf{P}$ such that

$$\begin{bmatrix} -\mathbf{P} & \mathbf{P}\mathbf{A}_{\text{cl}} & \mathbf{P}\mathbf{B}_{\text{cl}} & \mathbf{0} \\ \mathbf{A}_{\text{cl}}^{\text{T}}\mathbf{P} & -\mathbf{P} & \mathbf{0} & \mathbf{C}_{\text{cl}}^{\text{T}} \\ \mathbf{B}_{\text{cl}}^{\text{T}}\mathbf{P} & \mathbf{0} & -\gamma\mathbf{I} & \mathbf{D}_{\text{cl}}^{\text{T}} \\ \mathbf{0} & \mathbf{C}_{\text{cl}} & \mathbf{D}_{\text{cl}} & -\gamma\mathbf{I} \end{bmatrix} < 0 \tag{8.53}$$

$$\mathbf{P} > 0 \tag{8.54}$$

However, matrix inequality (8.53) is nonlinear in the unknown variables $\mathbf{A}_c$, $\mathbf{B}_c$, $\mathbf{C}_c$, and $\mathbf{P}$ and cannot be solved directly using the available LMI tools. In the following, with the proper change of variables and congruence transformation, (8.53) can be transformed into an LMI optimization problem (Scherer et al. 1997). Let n be the number of states of $\mathbf{A}$ and m be the number of states of $\mathbf{A}_c$. Partition $\mathbf{P}$ and $\mathbf{P}^{-1}$ as

$$\mathbf{P} = \begin{bmatrix} \mathbf{Y} & \mathbf{N} \\ \mathbf{N}^T & \bullet \end{bmatrix}, \quad \mathbf{P}^{-1} = \begin{bmatrix} \mathbf{X} & \mathbf{M} \\ \mathbf{M}^T & \bullet \end{bmatrix},$$

where $\mathbf{X}$ and $\mathbf{Y}$ are $n \times n$ symmetric matrices. Define $\Pi_1 := \begin{bmatrix} \mathbf{X} & \mathbf{I} \\ \mathbf{M}^{\text{T}} & \mathbf{0} \end{bmatrix}$, $\Pi_2 := \begin{bmatrix} \mathbf{I} & \mathbf{Y} \\ \mathbf{0} & \mathbf{N}^{\text{T}} \end{bmatrix}$, and

$$\begin{cases} \hat{\mathbf{A}} := \mathbf{N}\mathbf{A}_c\mathbf{M}^{\mathrm{T}} + \mathbf{N}\mathbf{B}_c\mathbf{C}_2\mathbf{X} + \mathbf{Y}\mathbf{B}_2\mathbf{C}_c\mathbf{M}^{\mathrm{T}} + \mathbf{Y}\left(\mathbf{A} + \mathbf{B}_2\mathbf{D}_c\mathbf{C}_2\right)\mathbf{X} \\ \hat{\mathbf{B}} := \mathbf{N}\mathbf{B}_c + \mathbf{Y}\mathbf{B}_2\mathbf{D}_c \\ \hat{\mathbf{C}} := \mathbf{C}_c\mathbf{M}^{\mathrm{T}} + \mathbf{D}_c\mathbf{C}_2\mathbf{X} \\ \hat{\mathbf{D}} := \mathbf{D}_c \end{cases} \tag{8.55}$$

Then the following identities can be obtained:

$$\Pi_1^{\mathrm{T}}\mathbf{P}\mathbf{A}_{\mathrm{cl}}\Pi_1 = \Pi_2^{\mathrm{T}}\mathbf{A}_{\mathrm{cl}}\Pi_1 = \begin{bmatrix} \mathbf{A}\mathbf{X} + \mathbf{B}_2\hat{\mathbf{C}} & \mathbf{A} + \mathbf{B}_2\hat{\mathbf{D}}\mathbf{C}_2 \\ \hat{\mathbf{A}} & \mathbf{Y}\mathbf{A} + \hat{\mathbf{B}}\mathbf{C}_2 \end{bmatrix}$$

$$\Pi_1^{\mathrm{T}}\mathbf{P}\mathbf{B}_{\mathrm{cl}} = \Pi_2^{\mathrm{T}}\mathbf{B}_{\mathrm{cl}} = \begin{bmatrix} \mathbf{B}_1 + \mathbf{B}_2\hat{\mathbf{D}}\mathbf{D}_{21} \\ \mathbf{Y}\mathbf{B}_1 + \hat{\mathbf{B}}\mathbf{D}_{21} \end{bmatrix}$$

$$\mathbf{C}_{\mathrm{cl}}\Pi_1 = \begin{bmatrix} \mathbf{C}_1\mathbf{X} + \mathbf{D}_{12}\hat{\mathbf{C}} & \mathbf{C}_1 + \mathbf{D}_{12}\hat{\mathbf{D}}\mathbf{C}_2 \end{bmatrix}$$

$$\Pi_1^{\mathrm{T}}\mathbf{P}\Pi_1 = \Pi_1^{\mathrm{T}}\Pi_2 = \begin{bmatrix} \mathbf{X} & \mathbf{I} \\ \mathbf{I} & \mathbf{Y} \end{bmatrix}$$

Perform congruence transformation on (8.53) and (8.54) with $diag\left(\Pi_1, \Pi_1, \mathbf{I}, \mathbf{I}\right)$ and Π_1, respectively; (8.53) and (8.54) are then transformed to

$$\begin{bmatrix} -\mathbf{X} & \mathbf{I} & \mathbf{A}\mathbf{X} + \mathbf{B}_2\hat{\mathbf{C}} & \mathbf{A} + \mathbf{B}_2\hat{\mathbf{D}}\mathbf{C}_2 & \mathbf{B}_1 + \mathbf{B}_2\hat{\mathbf{D}}\mathbf{D}_{21} & \mathbf{0} \\ * & -\mathbf{Y} & \hat{\mathbf{A}} & \mathbf{Y}\mathbf{A} + \hat{\mathbf{B}}\mathbf{C}_2 & \mathbf{Y}\mathbf{B}_1 + \hat{\mathbf{B}}\mathbf{D}_{21} & \mathbf{0} \\ * & * & -\mathbf{X} & \mathbf{I} & \mathbf{0} & \left(\mathbf{C}_1\mathbf{X} + \mathbf{D}_{12}\hat{\mathbf{C}}\right)^{\mathrm{T}} \\ * & * & * & -\mathbf{Y} & \mathbf{0} & \left(\mathbf{C}_1 + \mathbf{D}_{12}\hat{\mathbf{D}}\mathbf{C}_2\right)^{\mathrm{T}} \\ * & * & * & * & -\gamma\mathbf{I} & \mathbf{D}_{\mathrm{cl}}^{\mathrm{T}} \\ * & * & * & * & * & -\gamma\mathbf{I} \end{bmatrix} < 0 \tag{8.56}$$

$$\begin{bmatrix} \mathbf{X} & \mathbf{I} \\ \mathbf{I} & \mathbf{Y} \end{bmatrix} > 0 \tag{8.57}$$

(8.56) and (8.57) are linear in the unknown variables $\mathbf{X}$, $\mathbf{Y}$, $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$, $\hat{\mathbf{C}}$, $\hat{\mathbf{D}}$, and γ, which can be solved efficiently using the available LMI tools. After solving LMIs of (8.56) and (8.57), nonsingular matrices $\mathbf{M}$ and $\mathbf{N}$ to satisfying $\mathbf{M}\mathbf{N}^{\mathrm{T}} = \mathbf{I} - \mathbf{X}\mathbf{Y}$ can be found first; then the controller $\mathbf{K}$ is constructed as follows:

$$\begin{cases} \mathbf{D}_c := \hat{\mathbf{D}} \\ \mathbf{C}_c := \left(\hat{\mathbf{C}} - \mathbf{D}_c\mathbf{C}_2\mathbf{X}\right)\mathbf{M}^{-\mathrm{T}} \\ \mathbf{B}_c := \mathbf{N}^{-1}\left(\hat{\mathbf{B}} - \mathbf{Y}\mathbf{B}_2\mathbf{D}_c\right) \\ \mathbf{A}_c := \mathbf{N}^{-1}\left(\hat{\mathbf{A}} - \mathbf{N}\mathbf{B}_c\mathbf{C}_2\mathbf{X} - \mathbf{Y}\mathbf{B}_2\mathbf{C}_c\mathbf{M}^{\mathrm{T}} - \mathbf{Y}\left(\mathbf{A} + \mathbf{B}_2\mathbf{D}_c\mathbf{C}_2\right)\mathbf{X}\right)\mathbf{M}^{-\mathrm{T}} \end{cases} \tag{8.58}$$

Then based on (8.56), (8.57), and (8.58), we can obtain the controller $\mathbf{K}$ such that (8.45) is solved.

8.2.4 Some Modifications to the Controller Synthesis

8.2.4.1 Regional Pole Constraint

By properly selecting the weight functions for the mixed-sensitivity H_∞ control, we can obtain the desired H_∞ performance. However, the closed-loop system may possess an unacceptable dynamic response. It is well known that pure H_∞ synthesis can enforce closed-loop stability; however, it does not allow for direct placement of the closed-loop poles in more specific regions, in which the pole location is usually related to the time response and transient behavior of the feedback system. Therefore, it is often desirable to impose additional damping and clustering constraints on the closed-loop dynamics. In this case, a regional pole constraint can be imposed on the closed-loop system, and pole assignment can also be expressed as LMI constraints on the Lyapunov matrix $\mathbf{P}$. LMI constraints can be used for any region $\Re$ of the complex plane that can be defined as

$$\Re = \left\{ z \in \mathbb{C} : \mathbf{L} + z\mathbf{M} + \bar{z}\mathbf{M}^{\mathrm{T}} < 0 \right\} \tag{8.59}$$

Where $\mathrm{L} = \mathrm{L}^T$ and $\mathbf{M}$ are fixed real matrices. $\Re$ is a convex region in the complex plane and can be easily formulated for special cases including vertical strips, disks, horizontal strips, conic sectors, ellipsoids, and arbitrary intersections thereof (more details can be found in Chilali and Gahinet (1996)). Then the matrix $\mathbf{A}_{cl}$ in (8.52) has all its eigenvalues in the LMI region $\Re$ if and only if some LMIs involving $\mathbf{A}_{cl}$ are solvable, as presented in the following theorem:

Theorem 8.6 *(Chilali and Gahinet 1996) The matrix $\mathbf{A}_{cl}$ has all its eigenvalues in the LMI region $\left\{ z \in \mathbb{C} : \left[l_{ij} + m_{ij} z + m_{ji} \bar{z} \right]_{ij} < 0 \right\}$ if and only if there exists a matrix $\mathbf{P} > 0$ such that*

$$\left[l_{ij}\mathbf{P} + m_{ij}\mathbf{A}_{\mathrm{cl}}^{\mathrm{T}}\mathbf{P} + m_{ji}\mathbf{P}\mathbf{A}_{\mathrm{cl}} \right]_{ij} < 0 \tag{8.60}$$

□

Using the congruence transformation with Π_1, (8.60) can be transformed to the LMI constraint

$$\left[l_{ij} \begin{bmatrix} \mathbf{X} & \mathbf{I} \\ \mathbf{I} & \mathbf{Y} \end{bmatrix} + m_{ij} \begin{bmatrix} \mathbf{A}\mathbf{X} + \mathbf{B}_2\hat{\mathbf{C}} & \mathbf{A} + \mathbf{B}_2\hat{\mathbf{D}}\mathbf{C}_2 \\ \hat{\mathbf{A}} & \mathbf{Y}\mathbf{A} + \hat{\mathbf{B}}\mathbf{C}_2 \end{bmatrix}^{\mathrm{T}} + m_{ji} \begin{bmatrix} \mathbf{A}\mathbf{X} + \mathbf{B}_2\hat{\mathbf{C}} & \mathbf{A} + \mathbf{B}_2\hat{\mathbf{D}}\mathbf{C}_2 \\ \hat{\mathbf{A}} & \mathbf{Y}\mathbf{A} + \hat{\mathbf{B}}\mathbf{C}_2 \end{bmatrix} \right]_{ij} < 0$$

This inequality then can be included with (8.56) and (8.57) to enforce the eigenvalues of $\mathbf{A}_{cl}$ be contained in the desired LMI region (8.59).

8.2.4.2 Synthesis with Integral Control

It should be noted that the H_∞ design frameworks do not in general produce integral control. However, for adaptive optics system, it is useful to include the integral function in the controller in order to track a constant reference (or to cancel a static aberration). There are several ways to achieve the integral design. One approach is to introduce an integral in the performance weight $\mathbf{W}_e$. Then the transfer function between $\mathbf{w}$ and $\mathbf{z}_1$ is given by

$$\mathbf{z}_1 = \mathbf{W}_e(\mathbf{I} + \mathbf{PK})^{-1}\mathbf{W}_d\mathbf{w}$$

Now if the resulting controller $\mathbf{K}$ stabilizes the plant $\mathbf{P}$ and makes the H_∞ norm between $\mathbf{w}$ and $\mathbf{z}_1$ finite, then $\mathbf{K}$ must have a pole at zero which is the zero of the sensitivity function. The problem with this approach is that the weight functions used in the mixed-sensitivity H_∞ control design must all be stable. If they are not, Assumption A1 in Sect. 8.2.2 is not satisfied because the zero pole of $\mathbf{W}_e$ becomes an uncontrollable pole of the feedback system, and the general H_∞ algorithm is not applicable. Therefore, we need to approximate the integral control with a small perturbation. For example, we can approximate $1/s$ as $1/s + \varepsilon, 0 < \varepsilon \ll 1$, based on which the weight function $\mathbf{W}_e$ can then be formulated both in continuous or discrete-time domain.

Another way to produce the integral control is to reformulate the problem by properly adding the integral into the feedback loop directly (Zhou et al. 1995). This can be achieved by expressing the controller as $\mathbf{K} = \mathbf{M}_i\mathbf{K}_\infty$ where $\mathbf{M}_i$ is a preselected diagonal function with integral terms along the main diagonal and $\mathbf{K}_\infty$ is the H_∞ controller designed for a new plant obtained by cascading the function $\mathbf{M}_i$ with the old plant.

8.2.4.3 Synthesis with H_2 Performance

In many real-world applications, standard H_∞ synthesis cannot adequately capture all design specifications. For instance, noise attenuation or regulation against random disturbances is more naturally expressed in H_2 or LQG terms. This makes mixed H_∞/H_2 multi-objective synthesis highly desirable in practice, and LMI theory offers powerful tools to attack such problems. The control problem is sketched in Fig. 8.11. The output channel $\mathbf{z}_\infty$ is associated with the H_∞ performance, while the channel $\mathbf{z}_2$ is associated with the H_2 performance. For system (8.46), if the disturbance $\mathbf{w}$ is the random signal, for example, the random aberration caused by the turbulent air in the astronomic adaptive optics system, then it is natural to consider the H_2 performance for the performance error $\mathbf{e}$, which is argumented

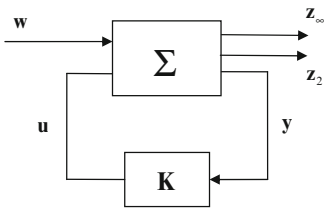


Fig. 8.11 Multiple objective H_∞/H_2 synthesis

as $\mathbf{z}_2$ in (8.46). In order to cast the control design problem into a mixed H_∞/H_2 multi-objective synthesis problem, redefine $\mathbf{z}_\infty = \mathbf{z}_1$ and $\mathbf{z}_2 = \mathbf{z}_2$, and then the closed-loop system (8.52) can be reformulated as

$$\Sigma_{\text{cl}}: \begin{cases} \mathbf{x}_{\text{cl}}(k+1) = \mathbf{A}_{\text{cl}}\mathbf{x}_{\text{cl}}(k) + \mathbf{B}_{\text{cl}}\mathbf{w}(k) \\ \begin{bmatrix} \mathbf{z}_\infty(k) \\ \mathbf{z}_2(k) \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{\text{cl1}} \\ \mathbf{C}_{\text{cl2}} \end{bmatrix} \mathbf{x}_{\text{cl}}(k) + \begin{bmatrix} \mathbf{D}_{\text{cl1}} \\ \mathbf{D}_{\text{cl2}} \end{bmatrix} \mathbf{w}(k) \end{cases}$$

where $\begin{bmatrix} \mathbf{C}_{\text{cl1}} \\ \mathbf{C}_{\text{cl2}} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{11} + \mathbf{D}_{121}\mathbf{D}_c\mathbf{C}_2 & \mathbf{D}_{121}\mathbf{C}_c \\ \mathbf{C}_{12} + \mathbf{D}_{122}\mathbf{D}_c\mathbf{C}_2 & \mathbf{D}_{122}\mathbf{C}_c \end{bmatrix}$ and $\begin{bmatrix} \mathbf{D}_{\text{cl1}} \\ \mathbf{D}_{\text{cl2}} \end{bmatrix} = \begin{bmatrix} \mathbf{D}_{111} + \mathbf{D}_{121}\mathbf{D}_c\mathbf{D}_{21} \\ \mathbf{D}_{112} + \mathbf{D}_{122}\mathbf{D}_c\mathbf{D}_{21} \end{bmatrix}$.
Then the mixed H_∞/H_2 multi-objective suboptimal problem can be solved based on the following linear matrix inequalities (Gahinet et al. 1995).

Mixed H_∞/H_2 multi-objective problem:

OP1: Minimizing $\alpha\gamma^2 + (1-\alpha)\,Trace\,(\mathbf{Q})$ over $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{A}_c$, $\mathbf{B}_c$, $\mathbf{C}_c$, $\mathbf{D}_c$, and γ, satisfying

$$\begin{bmatrix} -\mathbf{P} & \mathbf{P}\mathbf{A}_{\text{cl}} & \mathbf{P}\mathbf{B}_{\text{cl}} & \mathbf{0} \\ \mathbf{A}_{\text{cl}}^{\text{T}}\mathbf{P} & -\mathbf{P} & \mathbf{0} & \mathbf{C}_{\text{cl1}}^{\text{T}} \\ \mathbf{B}_{\text{cl}}^{\text{T}}\mathbf{P} & \mathbf{0} & -\gamma\mathbf{I} & \mathbf{D}_{\text{cl1}}^{\text{T}} \\ \mathbf{0} & \mathbf{C}_{\text{cl1}} & \mathbf{D}_{\text{cl1}} & -\gamma\mathbf{I} \end{bmatrix} < 0 \tag{8.61}$$

$$\begin{bmatrix} -\mathbf{P} & \mathbf{P}\mathbf{A}_{\text{cl}}^{\text{T}} & \mathbf{0} \\ \mathbf{A}_{\text{cl}}\mathbf{P} & -\mathbf{P} & \mathbf{B}_{\text{cl}} \\ \mathbf{0} & \mathbf{B}_{\text{cl}}^{\text{T}} & -\mathbf{I} \end{bmatrix} < 0 \tag{8.62}$$

$$\begin{bmatrix} \mathbf{P} & \mathbf{0} & \mathbf{P}\mathbf{C}_{\text{cl2}}^{\text{T}} \\ \mathbf{0} & \mathbf{I} & \mathbf{D}_{\text{cl2}}^{\text{T}} \\ \mathbf{C}_{\text{cl2}}\mathbf{P} & \mathbf{D}_{\text{cl2}} & \mathbf{Q} \end{bmatrix} > 0 \tag{8.63}$$

$$\mathbf{P} > 0 \tag{8.64}$$

$$Trace\,(\mathbf{Q}) < \upsilon_0 \tag{8.65}$$

$$\gamma < \gamma_0 \tag{8.66}$$

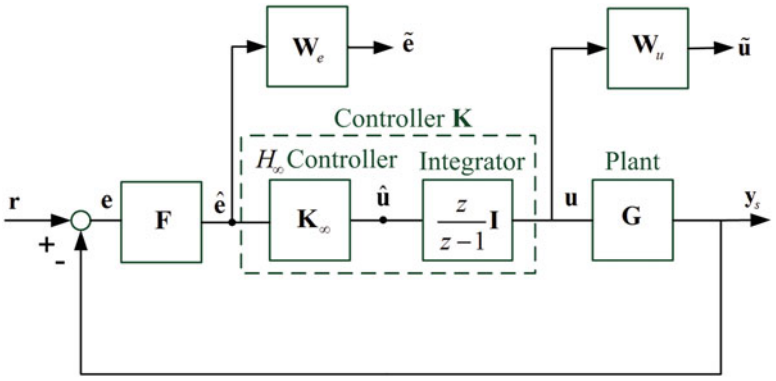


Fig. 8.12 Closed-loop system structure for mixed sensitivity H_∞ controller design

where υ_0 and γ_0 are preset optimal performance values and $\alpha \in (0, 1)$.

Since the above matrix inequalities are nonlinear in the unknown parameters $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{A}_c$, $\mathbf{B}_c$, $\mathbf{C}_c$, $\mathbf{D}_c$, and γ, the similar procedure as in Sect. 8.2.3 can be used to transform the above matrix inequalities into LMIs. Using the congruence transformation with $diag\,(\Pi_1, \Pi_1, \mathbf{I}, \mathbf{I})$, $diag\,(\Pi_1, \Pi_1, \mathbf{I})$, $(\Pi_1, \mathbf{I}, \mathbf{I})$, and Π_1 on (8.61), (8.62), (8.63), and (8.64), respectively, the OP1 problem is then transformed to an optimization problem in LMIs and can be solved effectively.

8.2.5 *The MFDM Surface Shape Control with the Mixed-Sensitivity H_∞ Controller*

In order to design an effective controller that minimizes the shape error $\mathbf{e}$ in the closed-loop adaptive optics system, a loop-shaping approach is used where weighting functions are employed to guide the optimization process which yields the desired controller parameters. The closed-loop system used in the controller design process is shown in Fig. 8.12, where $\mathbf{G}$ is the discretized augmented plant model representing the wavefront corrector and the sensor. The diagonal matrices $\mathbf{W}_e$ and $\mathbf{W}_u$ are the weighting functions which influence the selection of the desired controller parameters. The weighting function $\mathbf{W}_e = w_e\mathbf{I}$, where w_e is a scalar low-pass filter, is used to shape the closed-loop sensitivity function $\mathbf{S}_c$, that is, the transfer function from the input signal $\mathbf{r}$ to the performance variable $\mathbf{e}$ in the closed-loop system, and expressed as

$$\mathbf{S}_c = (\mathbf{I} + \mathbf{FGK})^{-1} \tag{8.67}$$

The low pass filter w_e is selected to have a relatively large gain in the frequency range covering the common wavefront aberrations in the human eye, that is, 0–3 Hz.

The weighting function $\mathbf{W}_u = w_u\mathbf{I}$ is such that w_u is a scalar high-pass filter with a corner frequency approximately equal to the desired closed-loop bandwidth of 3 Hz. The weighting function also contributes toward the robustness of the closed-loop system by minimizing the controller output in the frequency range beyond the desired closed-loop system bandwidth. The diagonal matrix $\mathbf{F} = f\mathbf{I}$, where f is a scalar low-pass filter, is introduced before the controller $\mathbf{K}$ and is used to attenuate the high-frequency measurement noise in the measurement signal being fed to the controller. The integrator shown in Fig. 8.12 is introduced to ensure that the error in tracking static reference signals is driven to zero. The incorporation of the integral action into the controller can also be realized by including integrators into the weighting function $\mathbf{W}_e$. This approach requires approximating the integrators with stable functions since the weight function must be stable. To avoid these approximations and to realize practically acceptable fast transient responses in the closed-loop system, the integrators are explicitly included in the controller. With the integrator block included, the overall controller becomes

$$\mathbf{K} = \frac{z}{z-1}\mathbf{K}_\infty$$

where $\mathbf{K}_\infty$ is the H_∞ controller to be discussed in the following. The successful tracking of the reference $\mathbf{r}$ by the closed-loop system can be achieved if the maximum singular value of the sensitivity function $\mathbf{S}_c$ is made small over a desired frequency range. In order to limit the input currents to the plant and increase the robustness of the closed-loop system, the transfer function from the reference signal $\mathbf{r}$ to the weighted controller output $\tilde{\mathbf{u}}$ is included in the transfer function to be shaped, hence the mixed-sensitivity design approach. With the structure of the closed-loop system selected as described above, the mixed-sensitivity H_∞ controller design problem is formulated as follows:

$$\min_{\mathbf{K}} \left\| \begin{bmatrix} \mathbf{W}_e\mathbf{S}_c\mathbf{F} \\ \mathbf{W}_u\mathbf{K}\mathbf{S}_c\mathbf{F} \end{bmatrix} \right\|_\infty \tag{8.68}$$

where $\mathbf{W}_e\mathbf{S}_c\mathbf{F}$ and $\mathbf{W}_u\mathbf{K}\mathbf{S}_c\mathbf{F}$ are the transfer functions from $\mathbf{r}$ to $\tilde{\mathbf{e}}$ and from $\mathbf{r}$ to $\tilde{\mathbf{u}}$, respectively. The controller is obtained by minimizing the H_∞ norm of the transfer function from the reference signals (i.e., the reference wavefront shape $\mathbf{r}$) to the weighted wavefront shape error $\tilde{\mathbf{e}}$ and the weighted controller output $\tilde{\mathbf{u}}$.

Let the state-space representations of $\mathbf{W}_e$, $\mathbf{W}_u$, and $\mathbf{F}$ be $\mathbf{W}_e : \begin{bmatrix} \mathbf{A}_{we} & \mathbf{B}_{we} \\ \mathbf{C}_{we} & \mathbf{D}_{we} \end{bmatrix}$, $\mathbf{W}_u : \begin{bmatrix} \mathbf{A}_{wu} & \mathbf{B}_{wu} \\ \mathbf{C}_{wu} & \mathbf{D}_{wu} \end{bmatrix}$, $\mathbf{F} : \begin{bmatrix} \mathbf{A}_f & \mathbf{B}_f \\ \mathbf{C}_f & \mathbf{D}_f \end{bmatrix}$ and that of the integral term $\frac{z}{z-1}\mathbf{I}$ be $\begin{bmatrix} \mathbf{A}_i & \mathbf{B}_i \\ \mathbf{C}_i & \mathbf{D}_i \end{bmatrix}$. Define the weighted performance variable as $\mathbf{z} = \begin{bmatrix} \tilde{\mathbf{e}}^{\mathrm{T}} & \tilde{\mathbf{u}}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$. Then the problem of minimizing the H_∞ norm of the mixed-sensitivity function $\begin{bmatrix} \mathbf{W}_e\mathbf{S}_c\mathbf{F} \\ \mathbf{W}_u\mathbf{K}\mathbf{S}_c\mathbf{F} \end{bmatrix}$ with respect to the parameters of the controller $\mathbf{K}_\infty$ can be properly formulated by

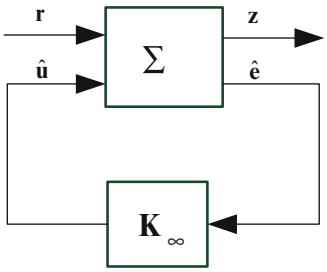


Fig. 8.13 Block diagram of the closed-loop system rearranged in a standard feedback configuration

rearranging the closed-loop system as shown in Fig. 8.13. The open-loop system $\boldsymbol{\Sigma}$ from $\left[\mathbf{r}^{\mathrm{T}}\ \hat{\mathbf{u}}^{\mathrm{T}}\right]^{\mathrm{T}}$ to $\left[\mathbf{z}^{\mathrm{T}}\ \hat{\mathbf{e}}^{\mathrm{T}}\right]^{\mathrm{T}}$, as shown in Fig. 8.13, can be written in state-space form as

$$\Sigma : \begin{cases} \tilde{\mathbf{x}}(k+1) = \mathbf{A}\tilde{\mathbf{x}}(k) + \mathbf{B}_1\mathbf{r}(k) + \mathbf{B}_2\hat{\mathbf{u}}(k) \\ \mathbf{z}(k) = \mathbf{C}_1\tilde{\mathbf{x}}(k) + \mathbf{D}_{11}\mathbf{r}(k) + \mathbf{D}_{12}\hat{\mathbf{u}}(k) \\ \hat{\mathbf{e}}(k) = \mathbf{C}_2\tilde{\mathbf{x}}(k) + \mathbf{D}_{21}\mathbf{r}(k) + \mathbf{D}_{22}\hat{\mathbf{u}}(k) \end{cases} \tag{8.69}$$

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_g & \mathbf{B}_g\mathbf{C}_i & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_i & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{B}_f\mathbf{C}_g & \mathbf{0} & \mathbf{A}_f & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_{wu}\mathbf{C}_i & \mathbf{0} & \mathbf{A}_{wu} & \mathbf{0} \\ -\mathbf{B}_{we}\mathbf{D}_f\mathbf{C}_g & \mathbf{0} & \mathbf{B}_{we}\mathbf{C}_f & \mathbf{0} & \mathbf{A}_{we} \end{bmatrix}$$

$$\mathbf{B}_1 = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{B}_f \\ \mathbf{0} \\ \mathbf{B}_{we}\mathbf{D}_f \end{bmatrix}, \quad \mathbf{B}_2 = \begin{bmatrix} \mathbf{B}_g\mathbf{D}_i \\ \mathbf{B}_i \\ \mathbf{0} \\ \mathbf{B}_{wu}\mathbf{D}_i \\ \mathbf{0} \end{bmatrix}$$

$$\mathbf{C}_1 = \begin{bmatrix} -\mathbf{D}_{we}\mathbf{D}_f\mathbf{C}_g & \mathbf{0} & \mathbf{D}_{we}\mathbf{C}_f & \mathbf{0} & \mathbf{C}_{we} \\ \mathbf{0} & \mathbf{D}_{wu}\mathbf{C}_i & \mathbf{0} & \mathbf{C}_{wu} & \mathbf{0} \end{bmatrix}$$

$$\mathbf{D}_{11} = \begin{bmatrix} \mathbf{D}_{we}\mathbf{D}_f \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{D}_{12} = \begin{bmatrix} \mathbf{0} \\ \mathbf{D}_{wu}\mathbf{D}_i \end{bmatrix}$$

$$\mathbf{C}_2 = \begin{bmatrix} -\mathbf{D}_f\mathbf{C}_g\ \mathbf{0}\ \mathbf{C}_f\ \mathbf{0}\ \mathbf{0} \end{bmatrix}, \quad \mathbf{D}_{21} = \mathbf{D}_f, \quad \mathbf{D}_{22} = \mathbf{0}$$

Since the sampling period for the closed-loop system is 0.0345 s, the weighting function $\mathbf{W}_e$ is selected such that it acts as a low-pass filter with bandwidth of 3 Hz equal to that of the expected reference signals and is given by

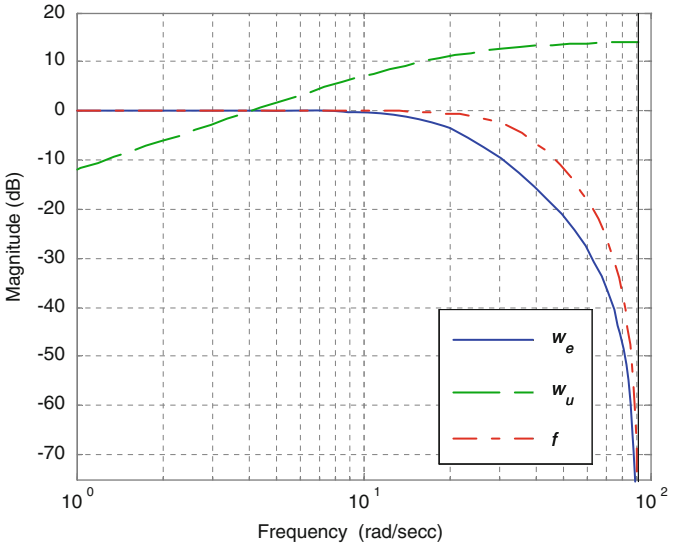


Fig. 8.14 Magnitude plots of w_e, w_u, and f

$$w_e(z) = \frac{z^2 + 2z + 1}{13.99z^2 - 15.6z + 5.605} \tag{8.70}$$

The filter f is selected as a low-pass filter with cutoff frequency at 5 Hz and is given by

$$f(z) = \frac{z^2 + 2z + 1}{6.108z^2 - 3.518z + 1.41} \tag{8.71}$$

The weight function $\mathbf{W}_u$ is selected such that

$$w_u(z) = \frac{3.717z - 3.717}{z - 0.487} \tag{8.72}$$

which represents a high-pass filter with a corner frequency of 3.2 Hz. The bode magnitude plots of the functions w_e, w_u, and f are shown in Fig. 8.14.

With the closed-loop system represented as shown in Fig. 8.13, and with the necessary weights selected as above, the H_∞ controller synthesis problem is solved based on LMIs using the Matlab Robust Control Toolbox function *dhinflmi*. For a plant model based on the prototype MFDM with 19 input electromagnetic coils arranged in a circular pattern and the mirror surface deflections measured immediately above the center of each of the 19 electromagnetic coils (i.e., $L = M = 19$), the controller design algorithm provides a controller with an H_∞ performance $\gamma_{\min} = 0.776$. The maximum singular values of the transfer functions $\mathbf{S}_c\mathbf{F}$ and $\mathbf{K}\mathbf{S}_c\mathbf{F}$ are plotted in Fig. 8.15. It can be seen that the sensitivity of

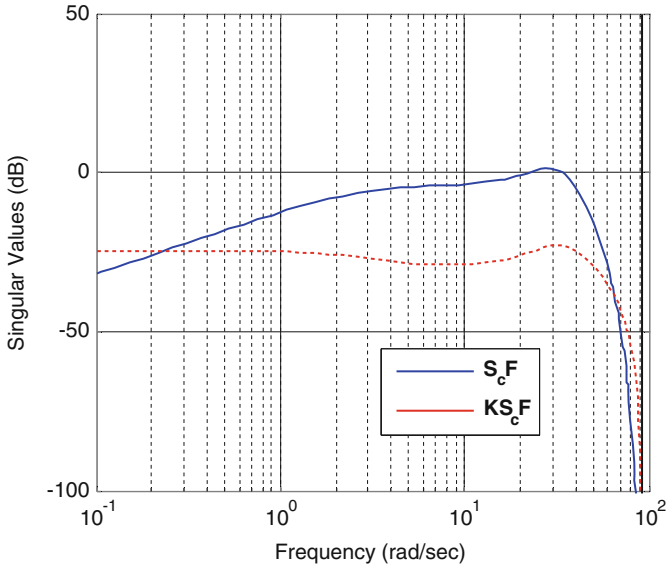


Fig. 8.15 Maximum singular values of $\mathbf{S}_c\mathbf{F}$ and $\mathbf{KS}_c\mathbf{F}$

the closed-loop system is attenuated in the desired low-frequency range. Also the controller output is attenuated in the high-frequency range, thereby increasing the robustness of the closed-loop system.

The H_∞ controller is implemented by incorporating the controller into the closed-loop system as shown in Fig. 8.12. The augmented plant $\mathbf{G}$ represents the actual system comprising the mirror and the sensor. The evaluation of the performance of the resulting closed-loop system is conducted by separately considering reference signals that are commonly present in the eye's aberrations, namely, static, harmonic (sinusoidal), and random signals. For a typical static reference wavefront shape signal $\mathbf{r}_0$ given as

$$\mathbf{r}_0 = [2.0, 4.0, 5.0, 1.0, 5.5, 0.5, 3.5, 8.5, 6.0, 3.0, 2.5, 2.0, 3.0, 4.0, -2.0, \\ -4.0, -5.0, 1.0, 6.0]^{\mathrm{T}} (\mu\mathrm{m})$$

The controller performance is shown in Fig. 8.16 where only selected channels (#2, 4, 8, and 17) are shown. As can be seen in Fig. 8.16a, the wavefront shape measurements $\mathbf{y}_s$ track the static reference signals fairly well, and the measurements are rapidly driven to their steady-states. The RMS of the wavefront shape error is shown in Fig. 8.16b and has a steady state value of 0.09 ± 0.05 μm. In order to test the performance of the closed-loop system in tracking dynamic wavefront shapes, the reference signals

$$\mathbf{r}(k) = \mathbf{r}_0 + \mathbf{r}'(k) \ (\mu\mathrm{m})$$

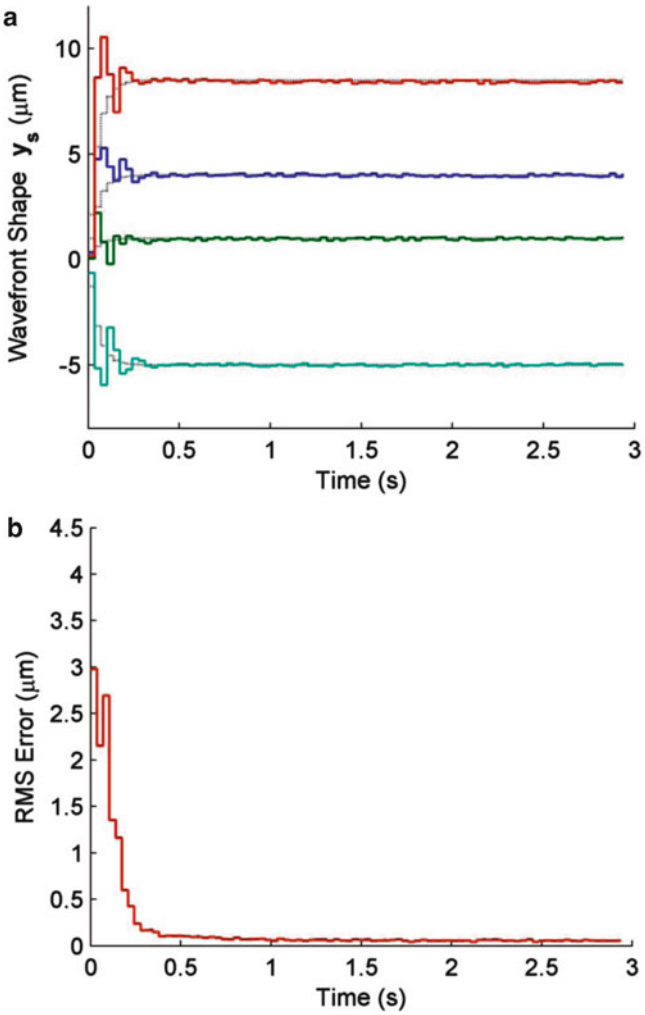


Fig. 8.16 Experimental results on the tracking of the static wavefront shape: (**a**) wavefront shape, (**b**) RMS error

are used where $\mathbf{r}_0$ is the static component and $\mathbf{r}'$ is a sinusoidally varying signal superimposed on the static component. The dynamic component $\mathbf{r}'$ is defined as follows:

$$\mathbf{r}'(k) = [\sin(f_o k + \phi_1), \sin(f_o k + \phi_2), \ldots, \sin(f_o k + \phi_{19})]^{\mathrm{T}}$$

where $f_o = 0.65$ radians are the frequency of the sinusoidal signals and ϕ_m, $m = 1, 2, ..., 19$ are the phases of the signals chosen randomly such that $0 < \phi_m < \frac{\pi}{2}$. The 2 μm peak-to-valley dynamic reference wavefront shape signals are superimposed

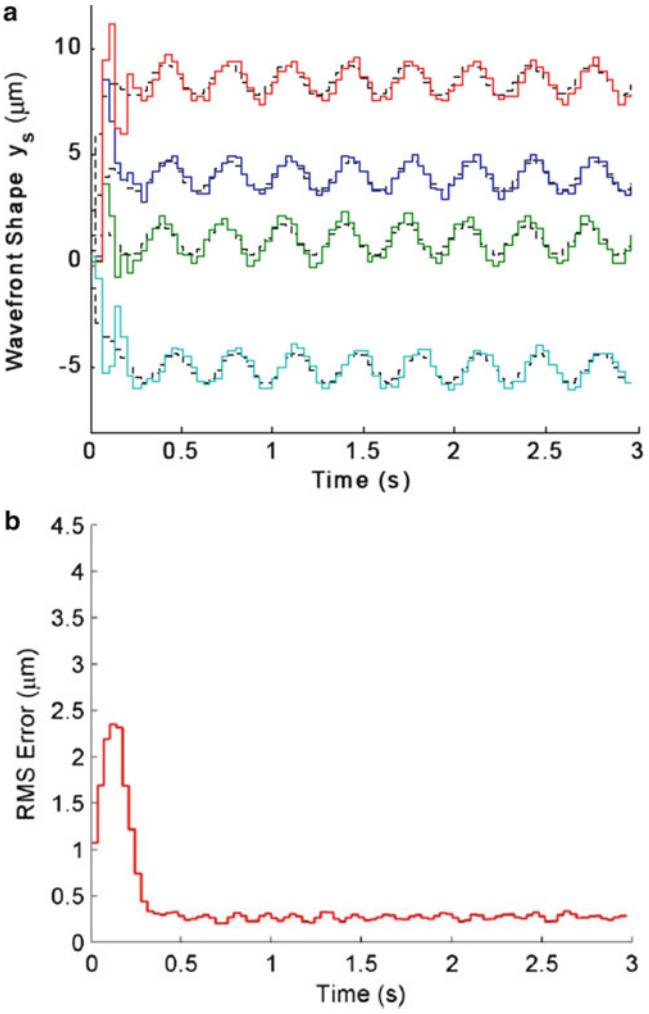


Fig. 8.17 Experimental results on the tracking of the sinusoidal wavefront shape: (**a**) wavefront shape, (**b**) RMS error

on the static offset $\mathbf{r}_0$ to represent the aberrations in the eye, which almost always have a large static component. The performance of the designed controller in ensuring the tracking of the dynamic reference signals by the wavefront shape measurements $\mathbf{y}_s$ is presented in Fig. 8.17. The wavefront shape measurements $\mathbf{y}_s$ are plotted in Fig. 8.17a where the corresponding reference wavefront shape signals $\mathbf{r}$ are shown using dotted lines. The RMS of the corresponding wavefront shape error is plotted in Fig. 8.17b. As can be observed from the results shown in Fig. 8.17, the measured wavefront shape follows the reference shape fairly well.

In order to test the performance of the closed-loop system in tracking generalized dynamic wavefront shapes, the following reference signals are used:

$$\mathbf{r}(k) = \mathbf{r}_0 + \mathbf{Z}\mathbf{c}(k)$$

Where $\mathbf{r}_0$ is the static component and $\mathbf{Z}$ is the matrix of Zernike mode shapes computed at the 19 output points coinciding with the location of electromagnetic coils in the optical pupil. The Zernike coefficients $\mathbf{c}(k)$ determine the magnitude of the resulting reference shape and are chosen as:

$$\mathbf{c}(k) = \mathbf{F}_r\mathbf{s}(k)$$

where $\mathbf{F}_r = f_r\mathbf{I}$ is a low-pass filter with $f_r = w_e$ as given in (8.70) and $\mathbf{s}(k)$ is a random signal. The low-pass filter $\mathbf{F}_r$ limits the frequency content of the coefficients $\mathbf{c}(k)$ to the range of frequencies present in the wavefront shapes presented by a typically aberrated human eye. The power of the random signals $\mathbf{s}(k)$ is chosen such that the signals corresponding to different Zernike modes are scaled according to the dynamics of the typically aberrated eye. The experimental results for tracking of the generalized dynamic reference wavefront shape are given in Fig. 8.18. The wavefront shape measurements $\mathbf{y}_s$ are plotted in Fig. 8.18a where the corresponding reference wavefront shape signals are shown using dotted lines. The RMS of the wavefront shape error is plotted in Fig. 8.18b. Although there are fluctuations in the RMS error values, the average RMS error still remains less than 0.5 μm, and the measured wavefront shape follows the reference shape fairly well. The experimental results given above show that the designed controller can effectively compensate for static as well as dynamic high-order wavefront aberrations.

8.3 Summary

In this chapter, two centralized optimal control algorithms are proposed for the surface shape control problem of MFDMs. The first control algorithm based on properly formulated LMIs is developed to design multivariable PID controllers in the discrete-time domain. Controller designs that yield closed-loop systems with H_∞ and H_2 performance specifications are investigated. Corresponding controllers are determined based on solving properly formulated linear matrix inequalities. The latter are obtained directly from different tuning constraints by introducing slack variables with special structure. The proposed design methods are verified to design PID controllers for a magnetic fluid deformable mirror which is used to track desired wavefront shapes in adaptive optics systems. The performance of the controllers is experimentally evaluated using a closed-loop adaptive optics system comprising the developed controllers and a novel 19-channel prototype magnetic fluid deformable mirror. The experimental results indicate the closed-loop system allows for the successful tracking of static and dynamic wavefronts typically present in the human eye.

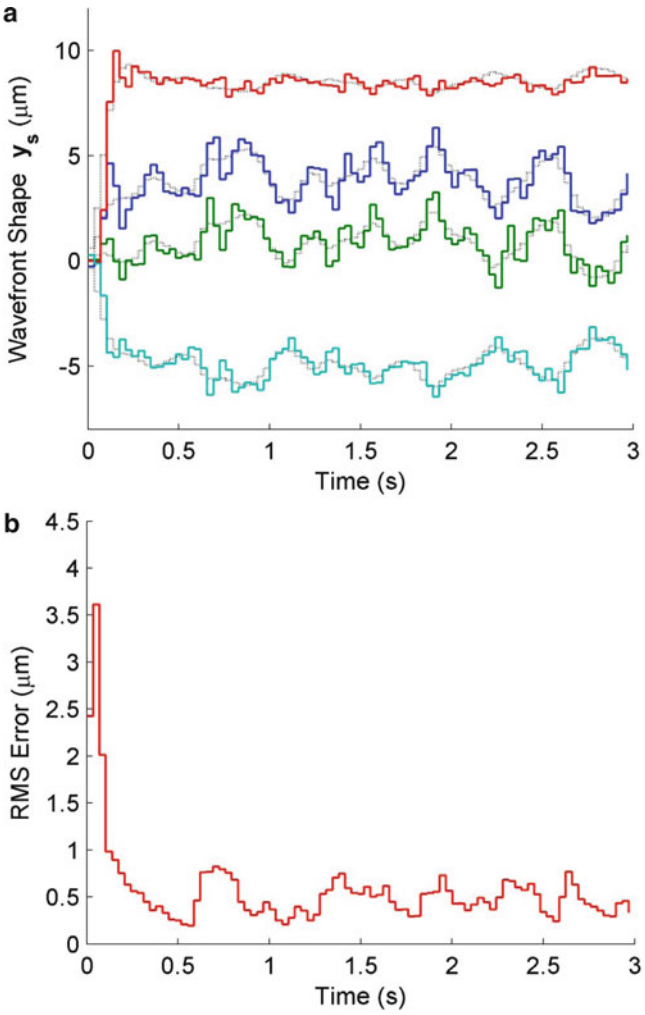


Fig. 8.18 Experimental results on the tracking of the generalized wavefront shape: (**a**) wavefront shape, (**b**) RMS error

The second control algorithm offering an H_∞ performance in tracking dynamic wavefront shapes using a magnetic fluid deformable mirror is also presented. A mixed-sensitivity H_∞ controller design approach is used to provide the desired tracking performance, limit the input currents, and enhance the robustness of the closed-loop system. The performance of the control algorithm is experimentally evaluated using a closed-loop adaptive optics system comprising the developed controller and a 19-channel prototype magnetic fluid deformable mirror. The closed-loop system was experimentally shown to provide successful tracking of dynamic wavefront shapes having a frequency content that is consistent with that of the human eye aberrations.

References

Apkarian P, Gahinet P, Becker G (2003) Fixed-order H_∞ control design via a partially augmented Lagrangian method. Int J Robust Nonlinear Control 13(12):1137–1148

Åström KJ, Hägglund T (2005) Advanced PID control. ISA-The Instrumentation, Systems, and Automation Society, Pittsburgh

Bianchi FD, Mantz RJ, Christiansen CF (2008) Multivariable PID control with set-point weighting via BMI optimisation. Automatica 44:472–478

Cao YY, Lam J, Sun YX (1998) Static output feedback stabilization: an LMI approach. Automatica 34(12):1641–1645

Chilali M, Gahinet P (1996) H_∞ design with pole placement constraints: an LMI approach. IEEE Trans Autom Control 41:358–367

Cuzzola FA, Morari M (2002) An LMI approach for H_∞ analysis and control of discrete-time piecewise affine systems. Int J Control 75:1293–1301

Gahinet P, Nemirovski A, Laub AJ, Chilali M (1995) LMI control toolbox. The MathWorks Inc., Natick

Ghaoui LE, Oustry F, Aitram M (1997) A cone complementary linearization algorithm for static output feedback and related problem. IEEE Trans Autom Control 42:1171–1176

Gu DW, Petkov PH, Konstantinov MM (2005) Robust control design with MATLAB. Springer, London

Hofer H, Artal P, Singer B, Aragon JL, Williams DR (2001) Dynamics of the eye's aberration. J Opt Soc Am A 18(3):497–505

Lin C, Wang QG, Lee TH (2004) An improvement on multivariable PID controller design via iterative LMI approach. Automatica 40:519–525

MacFarlane AGJ, Belletrutti JJ (1973) The characteristic locus design method. Automatica 9:575–588

Masubuchi I, Ohara A, Suda N (1998) LMI based controller synthesis: a unified formulation and solution. Int J Robust Nonlinear Control 8:669–686

Mattei M (2001) Robust multivariable PID control for linear parameter varying systems. Automatica 37:1997–2003

Morari M (1985) Robust stability of systems with integral control. IEEE Trans Autom Control 30(6):574–577

Nwokah ODI, Perez RA (1983) The stability of linear multivariable systems. Int J Control 37:623–629

O'Dwyer A (2006) Handbook of PI and PID controller tuning rules. Imperial College Press, London

Rotea MA (1993) The generalized H_2 control problem. Automatica 29:373–385

Scherer C, Gahinet P, Chilali M (1997) Multiobjective output feedback control via LMI optimization. IEEE Trans Autom Control 42(7):896–907

Zheng F, Wang QG, Lee TH (2003) On the design of multivariable PID controllers via LMI approach. Automatica 38:517–526

Zhou K, Doyle J, Golver K (1995) Robust and optimal control. Prentice Hall, Upper Saddle River

Chapter 9
Conclusions and Perspectives

Contents

9.1 Conclusions

Liquid deformable mirrors represent an emerging and promising technology and offer a major intrinsic advantage over solid mirrors, such as the extremely smooth surfaces, the high strokes that can be easily produced, the low cost per actuator, and the ease of scalability. Mercury-based liquid mirrors can yield high reflective paraboloidal surface by rotation and have been successfully used to carry out astronomical surveys. However, mercury mirrors have a number of major disadvantages. First, the rotating liquid can only produce parabolic shapes and will also be limited by winds generated by the rotation of the mirror itself. Second, mercury suffers from serious limitations when it is desired to deform its surface electromagnetically, and it is very difficult to obtain a stable mercury-based ferrofluid due to its chemical properties. Third, the high density of mercury necessitates very large electromagnetic forces to achieve even minimum required surface deflections, which means prohibitively large currents running through the electromagnetic coils used to generate the magnetic field. Confronted with these limitations of mercury, researchers came up with an innovative new technology that utilized the concept of ferrofluid deformation shaped with magnetic fields. Borra et al. (2006, 2008) at Laval University have pioneered and experimented with ferrofluidic mirrors coated with a metal liquid-like film (MELLF) to produce complex surface shapes. These mirrors are inexpensive and are capable of generating large deformations. However, the early generation of MFDMs had a major inconvenience in that the surface deformations of the liquid depend on the square of the applied magnetic vector field, which requires novel complicated control algorithms and precludes the possibility

Z. Wu et al., *Modeling and Control of Magnetic Fluid Deformable Mirrors for Adaptive Optics Systems*, DOI 10.1007/978-3-642-32229-7_9,

of directly using any linear control algorithms. In this backdrop, the closed-loop operation of the MFDMs is a difficult task and brings forward the control of the mirror as the most pressing requirement that needs to be addressed before they can be considered for practical applications. The research results presented in this book aim at facilitating the fulfillment of these requirements.

In this book, a dynamic model of a magnetic fluid deformable mirror and effective methods to control the surface shape of the mirror have been developed and validated experimentally. A significant development presented in this book is related to the proposed design of the mirror. It has been verified analytically as well as experimentally that the proposed design enables the use of linear control algorithms to control the mirror surface, which was otherwise not possible. The modified design also permits large surface displacements using very small electromagnetic coils. These features of the modified design are expected to open new opportunities for MFDMs in different applications. For example, the contributed work is expected to help realize the concept of an MFDM-based retinal imaging adaptive optics system in the future. The important developments presented in this book are summarized as follows:

- An accurate analytical model of the dynamics of an MFDM surface shape has been presented. The model represents the MFDM as an MIMO system with the currents applied to the array of electromagnetic coils as the input signals and the deflections of mirror surface provided by the model as the output signals. The model is derived by solving a coupled system of fluid-electromagnetic governing equations and is presented in state-space form. The model was experimentally validated using a prototype MFDM. It was demonstrated that the model can be used in the simulation of the response of the mirror as well as in the design of surface shape controllers for the mirror.
- A novel design of an MFDM, which linearizes the response of the surface shape of the mirror, has been presented. The proposed design features a uniform vertical magnetic field, which is superimposed on the magnetic field of the array of coils conventionally used to control the surface shape. The design change, which was prompted by the findings of the analytical work undertaken to develop the model of the response of the mirror, was implemented using a Helmholtz coil. The effectiveness of the proposed design in linearizing the response of the mirror surface shape was experimentally demonstrated. It was also shown that the proposed change in the design of an MFDM results in a significant amplification of the deflections of the mirror surface and allows for the bidirectional control of the mirror surface. These enhanced features facilitate the control of the mirror surface shape and also open new application avenues for these mirrors.
- The first-ever successful use of an MFDM in a closed-loop system has been reported. Two types of control algorithms are developed to control the surface shape of an MFDM, and their performance is experimentally evaluated. The classical decentralized control methods are first investigated, where two decentralized PID control algorithms are proposed to control the surface shape of a magnetic fluid deformable mirror in a closed-loop adaptive optics system.

A decentralized proportional-plus-integral (PI) controller is designed based on the static decoupling of the plant model using the plant DC gain. To overcome stability robustness issues and minimize the effect of the model uncertainties, especially in the high-frequency range, a decentralized robust proportional-integral-derivative (PID) controller is further developed. These two control algorithms provide good performance in removing static or very slowly changing aberrations. However, the simulation and experimental verifications show that the decentralized PID control algorithms have limited capabilities in dealing with fast-changing aberrations. To overcome this limitation, two centralized optimal control algorithms, including a multivariable PID controller and a mixed-sensitivity H_∞ controller, are proposed for the MFDM. The centralized optimal control algorithms allow for the tracking of generalized dynamic surface shapes. The closed-loop system was experimentally shown to provide successful tracking of dynamic wavefront shapes having a frequency content that is consistent with that of the human eye aberrations.

9.2 Perspectives

The scope of the study presented in this book was limited to the development of an analytical model for the MFDM and the novel modification in the conceptual design of an MFDM, as well as to the design of surface shape controllers for the mirror. The effectiveness of this new conceptual design approach for MFDMs has recently been further verified by Brousseau et al. (2011, 2010) and Parent and Thibault (2011). Achievement of these goals has certainly brought the MFDMs much closer to their deployment in practical AO systems. However, there are other aspects of MFDMs that will have to be addressed before they can be finally used practically in clinical or astronomical applications. The following are the important recommendations for future work to be carried out on MFDMs:

- *Size Reduction.* As stated in Sect. 2.3, one of the important requirements of ophthalmic AO systems is that the deformable mirror should be comparable to a dilated eye's pupil in size. Due to time and resource constraints, the prototype MFDM presented in this book could not be fabricated as per the size requirement of the ophthalmic AO systems. Development of a new prototype that meets the size requirements is recommended. The dynamics of the mirror, particularly the effects of linearization and surface tension, should be validated in the miniaturized device. Since a design using conventional coils is not likely to provide the required actuator density, the possibility of using microfabricated arrays of coils should be explored.
- *Study of effects of the reflective film.* The dynamic model presented in this book is based on the assumption that the nanoscale film that acts as a reflective coating on the surface of the MFDM does not have any significant effect on the dynamics of the mirror. Although this prospect has been construed in previous studies as

well (Brousseau et al. 2006), it needs to be verified experimentally. It is therefore recommended that the effects of the reflective film on the dynamics and control of the mirror be studied and experimentally validated.

- *Enhancement of bandwidth.* Although the MFDM presented in this book meets the closed-loop bandwidth requirements of ophthalmic AO systems, it does so marginally. In order for the MFDMs to find a place in practical applications, their response speed will have to be improved substantially. One approach to increase the response speed is by increasing the viscosity of the magnetic fluid along with an overdrive technique (Parent et al. 2009). This approach should be experimentally tested using the proposed MFDM design, and its impact on the response speed in the closed-loop system should be ascertained.
- *Physical properties of ferrofluid.* The ferrofluidic mirror can produce large deformations using large magnetic field; however, the increasing magnitude of the magnetic field will increase the risk of hitting the limit imposed by the Rosensweig instability. Fortunately, the limit of Rosensweig instability depends greatly on the physical parameters of the ferrofluid. It might be possible to synthesize a ferrofluid that has physical characteristics that do not show the instability. However, this prediction needs to be further studied and examined.
- *Design of effective coil actuators.* It is possible for the magnetic fluid deformable mirror to generate deformations having very large amplitudes, such that time-varying concave or convex mirrors with directly added aberration compensation can be made. As an example tested by Brousseau et al. (2006), by applying a sufficiently high current to concentric coils, a 5-cm-diameter spherical mirror can be easily made, and the additional defocus or spherical compensation can be further realized by modifying the currents in the coils. However, for large mirrors, the extremely large currents need high-voltage supply and huge power consumption. The large amount of heat generated by the coils could also deteriorate the image quality. The possible solutions are to use superconducting quantum wires or wires spun from carbon nanotubes (Borra 2009), while considerable research efforts need to be seen.
- *Tiltability of the ferrofluidic mirror.* The magnetic fluid deformable mirror could not be tilted, let alone be used facing down. Borra et al. (2008) had made a simple experiment by putting an MELLF-coated ferrofluid on a permanent magnet, and no obvious surface shape changes had been observed by turning it upside down. Changing the inclination of such a mirror would introduce a slight changing wedge into the surface, which could be eliminated by additional current-carrying wires of the mirror. The experiment confirmed that further examination of the implications of tilting this kind of mirror needs to be investigated. The tiltability would render MFDMs far more versatile and increase the appeal of these inexpensive devices in more optical system applications.

Magnetic fluid deformable mirrors are characterized by their robustness, low fabrication costs, and ease of maintenance. With additional research and development work, this type of mirrors could be practically used in the near future in AO applications such as ophthalmology, astronomy, photolithography, imaging, and projection.

References

Borra EF (2009) Liquid mirrors in engineering. Optics & Photonics News, pp 14–17, September 2009

Borra EF, Brousseau D, Vincent A (2006) Large magnetic liquid mirrors. Astron Astrophys 446(1):389–393

Borra EF, Brousseau D, Cliche M, Parent J (2008) Aberration correction with a magnetic liquid active mirror. Mon Notice R Astron Soc 391(4):1925–1930

Brousseau D, Borra EF, Ruel HJ, Parent J (2006) A magnetic liquid deformable mirror for high stroke and low order axially symmetrical aberrations. Opt Express 14:11486–11493

Brousseau D, Borra EF, Rochette M, Landry DB (2010) Linearization of the response of a 91-actuator magnetic liquid deformable mirror. Opt Express 18(8):8239–8250

Brousseau D, Drapeau J, Piché M, Borra EF (2011) Generation of Bessel beams using a magnetic liquid deformable mirror. Appl Opt 50:4005–4010

Parent J, Thibault S (2011) Locally magnifying imager. Opt Express 19(6):5676–5689

Parent J, Borra EF, Brousseau D, Ritcey AM, Dery JP, Thibault S (2009) Dynamic response of ferrofluidic deformable mirrors. Appl Opt 48(1):1–6

Appendix A

Contents

Derivation of the Analytical Model

A.1 Free Surface Kinematics

In the Cartesian coordinate system, for example, the surface can be represented using $z = \zeta(x, y)$. Accordingly, the expression $z - \zeta(x, y) = C$ represents contours having the same shape as the surface but displaced by magnitude C, and the gradient $\nabla[z - \zeta(x, y)]$ yields a vector normal to the surface. Therefore, the unit normal vector can be

$$\hat{\mathbf{n}} = -(\partial\zeta/\partial x)\hat{\mathbf{i}} - (\partial\zeta/\partial y)\hat{\mathbf{j}} + \hat{\mathbf{k}} \tag{A1}$$

The mean curvature κ of a surface can be found using

$$\kappa = \tfrac{1}{2}\nabla \cdot \hat{\mathbf{n}} \tag{A2}$$

Employing the linearized form (A1) of the unit normal vector, (A2) simplifies to

$$\kappa = -\frac{1}{2}\left(\frac{\partial^2\zeta}{\partial x^2} + \frac{\partial^2\zeta}{\partial y^2}\right) \tag{A3}$$

The deflection of a fluid surface can be related to the motion of the adjacent fluid using what is commonly known as the kinematic condition. The condition is derived as follows.

Z. Wu et al., *Modeling and Control of Magnetic Fluid Deformable Mirrors for Adaptive Optics Systems*, DOI 10.1007/978-3-642-32229-7,

When the shape of the fluid surface changes with time, the location of any point on the surface may be expressed as

$$z = \zeta\,(x, y, t) \tag{A4}$$

The corresponding Monge function can be written as

$$F\,(x, y, z, t) = z - \zeta\,(x, y, t) \tag{A5}$$

For any point on the surface $(z = \zeta)$,

$$\frac{DF}{Dt} = \frac{\partial F}{\partial t} + \mathbf{V} \cdot \nabla F = 0 \tag{A6}$$

where $\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{V} \cdot \nabla$ is an operator referred to as the material derivative and $\mathbf{V} = V_x\hat{\mathbf{i}} + V_y\hat{\mathbf{j}} + V_z\hat{\mathbf{k}}$ is the velocity of the fluid particle constituting the surface. Since ∇F defines a vector-oriented normal to the surface, using (A1), condition (A6) yields

$$-\frac{\partial \zeta}{\partial t} + V_x\left(-\frac{\partial \zeta}{\partial x}\right) + V_y\left(-\frac{\partial \zeta}{\partial y}\right) + V_z = 0 \tag{A7}$$

which can be rearranged to

$$V_z = \frac{\partial \zeta}{\partial t} + V_x\frac{\partial \zeta}{\partial x} + V_y\frac{\partial \zeta}{\partial y} \tag{A8}$$

Ignoring higher-order terms, (A8) can be written as

$$V_z = \frac{\partial \zeta}{\partial t} \tag{A9}$$

The condition (A9) states that, at any point on the free surface of a fluid, the rate of deflection of the surface is equal to the vertical component of the fluid velocity at that point.

A.2 Surface Dynamic Condition

The balance of interfacial stress at the free surface of an incompressible, inviscid magnetic fluid that is exposed to air (considered to be non-magnetizable material) yields the following dynamic condition at the surface:

$$p + p_{\mathrm{s}} + p_{\mathrm{m}} + p_{\mathrm{n}} = p^{\mathrm{a}} + p_{\mathrm{c}} \tag{A10}$$

where

$$p = \text{Thermodynamic pressure of the fluid}$$

$$p_{\mathrm{s}} = \text{Magnetostrictive pressure} = \mu_0 \int_0^H \upsilon\left(\frac{\partial M}{\partial \upsilon}\right)\mathrm{d}H$$

$$
\begin{aligned}
p_{\mathrm{m}} &= \text{Fluid} - \text{magnetic pressure} = \mu_0 \int_0^H M \, \mathrm{d}H \\
p_{\mathrm{n}} &= \text{Magnetic normal traction} = \frac{1}{2}\mu_0 (\mathbf{M} \cdot \hat{\mathbf{n}})^2 \\
p_{\mathrm{c}} &= \text{Capillary pressure} = 2\sigma\kappa \\
p^{\mathrm{a}} &= \text{Thermodynamic pressure of the air}
\end{aligned} \tag{A11}
$$

In the definition of the magnetostrictive pressure p_{s}, $\upsilon = \rho^{-1}$ is the volume density. This type of pressure arises from any volume change brought about by the applied magnetic field and, for the incompressible liquid carriers, can be neglected, that is,

$$
p_{\mathrm{s}} \approx 0 \tag{A12}
$$

The fluid-magnetic pressure p_{m} in a linearly magnetizable fluid ($\mathbf{M} = \chi \mathbf{H}$) can be evaluated as follows:

$$
\begin{aligned}
p_{\mathrm{m}} &= \mu_0 \int_0^H M \, \mathrm{d}H \\
&= \mu_0 \chi \int_0^H H \, \mathrm{d}H \\
&= \frac{1}{2}\mu_0 \chi H^2 \\
&= \frac{1}{2}\mu_0 \chi \, (\mathbf{H} \cdot \mathbf{H})
\end{aligned} \tag{A13}
$$

Similarly, the normal traction p_n can be evaluated as

$$
\begin{aligned}
p_{\mathrm{n}} &= \frac{1}{2}\mu_0 (\mathbf{M} \cdot \hat{\mathbf{n}})^2 \\
&= \frac{1}{2}\mu_0 \chi^2 (\mathbf{H} \cdot \hat{\mathbf{n}})^2
\end{aligned} \tag{A14}
$$

Using (A11), (A12), and (A13), the free surface dynamic condition (A10) can be written as

$$
p + \frac{\mu_0 \chi}{2} \left((\mathbf{H} \cdot \mathbf{H}) + \chi (\mathbf{H} \cdot \hat{\mathbf{n}})^2 \right) = p^{\mathrm{a}} + 2\sigma\kappa \tag{A15}
$$

The condition specifies a jump in the thermodynamic pressure above and below the interface of two fluids. In ordinary fluids, this jump in the pressure is determined

by surface tension of the fluid. As seen in (A15), in case of the magnetic fluids, the magnetic field also contributes toward the difference of the pressure.

A.3 Simplification of the Fluid Dynamic Equations

For convenience, the equations governing the fluid dynamic phenomena in the magnetic fluids are reproduced as follows:

$$\nabla \cdot \mathbf{V} = 0 \tag{A16}$$

$$\rho\left(\frac{\partial \mathbf{V}}{\partial t} + \mathbf{V} \cdot \nabla \mathbf{V}\right) = -\nabla\left(p + p_{\mathrm{s}} + p_{\mathrm{m}}\right) + \eta \nabla^2 \mathbf{V} + \rho \mathbf{g} + \mu_0 M \nabla H \tag{A17}$$

Consider the following:

(a) The application of magnetic field produces no appreciable change in the volume of the incompressible magnetic fluids. Therefore, the contribution of the magnetostrictive pressure toward the momentum balance can be ignored, that is,

$$\nabla p_{\mathrm{s}} \approx 0 \tag{A18}$$

(b) Using the definition of p_{m},

$$\nabla p_{\mathrm{m}} = \nabla\left(\mu_0 \int_0^H M \,\mathrm{d}H\right) \tag{A19}$$

Considering that the magnetization M depends only on magnetic field H and that the magnetic permeability μ_{o} remains constant, the right-hand side of (A19) can be written as

$$\nabla\left(\mu_0 \int_0^H M \,\mathrm{d}H\right) = \mu_0 M \nabla H \tag{A20}$$

It follows from (A19) and (A20) that

$$-\nabla p_m + \mu_0 M \nabla H = 0 \tag{A21}$$

(c) If the fluid is assumed to be nonviscous, that is, $\eta = 0$, then the viscous force term in (A17) is identically zero:

$$\eta \nabla^2 \mathbf{V} = 0 \tag{A22}$$

(d) Choosing a coordinate system with the gravitational force acting opposite to the z-axis and using the fact that the gravitational field is conservative, the gravitational acceleration vector can be written as

$$\mathbf{g} = \nabla(-gz) \tag{A23}$$

(e) Assuming that fluid is initially static and that the applied magnetic field results in only small perturbations of the fluid velocity, the term $\mathbf{V}\cdot\nabla\mathbf{V}$ in (A17), which is nonlinear in $\mathbf{V}$, can be ignored, that is,

$$\mathbf{V}\cdot\nabla\mathbf{V} \approx \mathbf{0} \tag{A24}$$

Applying the assumptions/simplifications (A18), (A21), (A22), (A23), and (A24), the momentum equation (A17) can be written as

$$\rho\left(\frac{\partial\mathbf{V}}{\partial t}\right) = -\nabla(p + \rho gz) \tag{A25}$$

A further assumption of an irrotational flow implies that

$$\nabla\times\mathbf{V} = \mathbf{0} \tag{A26}$$

Making use of a known vector identity, (A16) and (A26) allow the velocity $\mathbf{V}$ to be written as

$$\mathbf{V} = -\nabla\Phi \tag{A27}$$

such that Φ satisfies the Laplace equation

$$\nabla^2\Phi = 0 \tag{A28}$$

The substitution of (A27) in (A25) yields

$$\nabla\left(-\rho\frac{\partial\Phi}{\partial t} + p + \rho gz\right) = \mathbf{0} \tag{A29}$$

The integration of (A29) gives

$$-\rho\frac{\partial\Phi}{\partial t} + p + \rho gz = c(t) \tag{A30}$$

where $c(t)$ is the *constant* of integration and can be lumped into the potential $\Phi(x, y, z, t)$ resulting in

$$-\rho\frac{\partial\Phi}{\partial t} + p + \rho gz = 0 \tag{A31}$$

Note that the four equations governing the fluid field (the scalar equation (A16) and the three vector equations (A17)) have been reduced to two scalar equations (A28) and (A31), with two unknowns p and Φ.

A.4 Perturbation Analysis

The magnetic fluid layer in an initial equilibrium state is perturbed by the input magnetic field applied at the bottom of the layer. Assuming that the applied field results in small perturbations of the field variables, the perturbed part of the equations governing the fluid flow is extracted as follows.

Equilibrium State

The initial equilibrium state of the magnetic fluid layer is characterized by

$$\begin{aligned}
&\zeta = 0\\
&\kappa = 0\\
&\hat{\mathbf{n}} = \hat{\mathbf{k}}\\
&\mathbf{V} = \mathbf{V}_0 = -\nabla\Phi_0 = \mathbf{0}\\
&\mathbf{H} = \mathbf{H}_0 = -\nabla\Psi_0 = [0, 0, H_0]^{\mathrm{T}} = \left[0, 0, \frac{B_0}{\mu}\right]^{\mathrm{T}}\\
&\mathbf{H}^{\mathrm{a}} = \mathbf{H}_0^{\mathrm{a}} = -\nabla\Psi_0^{\mathrm{a}} = \left[0, 0, H_0^{\mathrm{a}}\right]^{\mathrm{T}} = \left[0, 0, \frac{B_0}{\mu_0}\right]^{\mathrm{T}}
\end{aligned} \tag{A32}$$

where Φ_0, $\Psi_{0,}$ and Ψ_0^{a} have been introduced as the scalar potentials corresponding to the initial velocity $\mathbf{V}_0$ and the magnetic fields $\mathbf{H}_0$ and $\mathbf{H}_0^{\mathrm{a}}$, respectively. Using (A32), the surface dynamic equation (A15) can be written as follows:

$$-\rho\frac{\partial\Phi_0}{\partial t} + \rho g\zeta - \frac{\mu_0\chi}{2}\left((\nabla\Psi_0\cdot\nabla\Psi_0) + \chi\left(\nabla\Psi_0\cdot\hat{\mathbf{k}}\right)^2\right) + p^{\mathrm{a}} + 2\sigma k = 0$$

Namely,

$$-\rho\frac{\partial\Phi_0}{\partial t} - \frac{\mu_0\chi\,(1+\chi)}{2}H_0^2 + p^{\mathrm{a}} = 0 \tag{A33}$$

It may be noted that the pressure p^{a} above the interface has been considered to be constant.

Perturbed State

The application of the input magnetic potential results in

$$\begin{aligned}
\hat{\mathbf{n}} &= -\frac{\partial \zeta}{\partial x}\hat{\mathbf{i}} - \frac{\partial \zeta}{\partial y}\hat{\mathbf{j}} + \hat{\mathbf{k}} \\
\kappa &= -\frac{1}{2}\left(\frac{\partial^2 \zeta}{\partial x^2} + \frac{\partial^2 \zeta}{\partial y^2}\right) \\
\mathbf{V} &= -\nabla\Phi = \mathbf{V}_0 + \mathbf{v} \\
\mathbf{H} &= -\nabla\Psi = \mathbf{H}_0 + \mathbf{h} \\
\mathbf{H}^{\mathrm{a}} &= -\nabla\Psi^{\mathrm{a}} = \mathbf{H}_0^{\mathrm{a}} + \mathbf{h}^{\mathrm{a}}
\end{aligned} \tag{A34}$$

where $\mathbf{v}$, $\mathbf{h}$, and $\mathbf{h}^{\mathrm{a}}$ are small perturbations and Φ, Ψ and Ψ^{a} are the scalar potentials. Since the velocity potential Φ obeys the Laplace equation (4.10), it can be written as a linear summation of the potentials corresponding to the component velocities. Therefore,

$$\Phi = \Phi_0 + \phi \tag{A35}$$

where ϕ is the velocity potential corresponding to the perturbation velocity component $\mathbf{v}$ which can be written as

$$\mathbf{v} = -\nabla\phi \tag{A36}$$

such that

$$\nabla^2\phi = 0 \tag{A37}$$

Similarly, since the magnetic potentials Ψ and Ψ^{a} too obey the Laplace equations (4.13) and (4.15), respectively, they can be written as

$$\Psi = \Psi_0 + \psi \tag{A38}$$

$$\Psi^{\mathrm{a}} = \Psi_0^{\mathrm{a}} + \psi^{\mathrm{a}} \tag{A39}$$

where ψ and ψ^{a} are the magnetic potentials corresponding to perturbations $\mathbf{h}$ and $\mathbf{h}^{\mathrm{a}}$ which can be written as

$$\begin{aligned}
\mathbf{h} &= -\nabla\psi \\
&= -\frac{\partial \psi}{\partial x}\hat{\mathbf{i}} - \frac{\partial \psi}{\partial y}\hat{\mathbf{j}} - \frac{\partial \psi}{\partial z}\hat{\mathbf{k}}
\end{aligned} \tag{A40}$$

$$\begin{aligned}\mathbf{h}^{\mathrm{a}} &= -\nabla\psi^{\mathrm{a}} \\ &= -\frac{\partial\psi^{\mathrm{a}}}{\partial x}\hat{\mathbf{i}} - \frac{\partial\psi^{\mathrm{a}}}{\partial y}\hat{\mathbf{j}} - \frac{\partial\psi^{\mathrm{a}}}{\partial z}\hat{\mathbf{k}}\end{aligned} \tag{A41}$$

such that

$$\nabla^2\psi = 0 \tag{A42}$$

$$\nabla^2\psi^{\mathrm{a}} = 0 \tag{A43}$$

Using (A34), the surface dynamic equation (A15) may be written for the perturbed state as follows:

$$-\rho\frac{\partial\Phi}{\partial t} + \rho g\zeta - \frac{\mu_0\chi}{2}\left((\nabla\Psi\cdot\nabla\Psi) + \chi(\nabla\Psi\cdot\hat{\mathbf{n}})^2\right) + p^{\mathrm{a}} - \sigma\left(\frac{\partial^2\zeta}{\partial x^2} + \frac{\partial^2\zeta}{\partial y^2}\right) = 0 \tag{A44}$$

Introducing (A35), (A38), and (A39), and keeping the first-order terms only, (A44) can be simplified to

$$\begin{aligned}&\underbrace{-\rho\frac{\partial\Phi_0}{\partial t} - \frac{\mu_0\chi\,(1+\chi)}{2}H_0^2 + p^{\mathrm{a}}}_{0} - \rho\frac{\partial\phi}{\partial t} + \rho g\zeta \\ &\quad + \mu_0\chi\,(1+\chi)\,H_0\frac{\partial\psi}{\partial z} - \sigma\left(\frac{\partial^2\zeta}{\partial x^2} + \frac{\partial^2\zeta}{\partial y^2}\right) = 0\end{aligned} \tag{A45}$$

which, using (A33), reduces to

$$-\rho\frac{\partial\phi}{\partial t} + \rho g\zeta + \mu_0\chi\,(1+\chi)\,H_0\frac{\partial\psi}{\partial z} - \sigma\left(\frac{\partial^2\zeta}{\partial x^2} + \frac{\partial^2\zeta}{\partial y^2}\right) = 0 \tag{A46}$$

Since

$$\begin{aligned}B_0 &= \mu H_0 \\ &= \mu_0\,(1+\chi)\,H_0\end{aligned} \tag{A47}$$

(A46) may be written as

$$-\rho\frac{\partial\phi}{\partial t} + \rho g\zeta + \chi B_0\frac{\partial\psi}{\partial z} - \sigma\left(\frac{\partial^2\zeta}{\partial x^2} + \frac{\partial^2\zeta}{\partial y^2}\right) = 0 \tag{A48}$$

This equation, along with the Laplace equations (A37), (A42), and (A43), describe the dynamics of the surface deflection ζ.

A.5 Solution of Laplace Equations

For convenience, the Laplace equations to be solved are reproduced as follows:

$$\nabla^2\phi = 0 \quad \text{for} \quad -h < z < \zeta \tag{A49}$$

$$\nabla^2\psi^{\mathrm{a}} = 0 \quad \text{for} \quad \zeta < z \tag{A50}$$

$$\nabla^2\psi = 0 \quad \text{for} \quad -h < z < \zeta \tag{A51}$$

We assume the following solutions for the surface deflection ζ and the Laplace equations given above:

$$\zeta\,(x, y, t) = \tilde{\zeta}(t)E\,(x, y) \tag{A52}$$

$$\phi\,(x, y, z, t) = \tilde{\phi}\,(z, t)\,E\,(x, y) \tag{A53}$$

$$\psi^{\mathrm{a}}\,(x, y, z, t) = \tilde{\psi}^{\mathrm{a}}\,(z, t)\,E\,(x, y) \tag{A54}$$

$$\psi\,(x, y, z, t) = \tilde{\psi}\,(z, t)\,E\,(x, y) \tag{A55}$$

where

$$E\,(x, y) = \mathrm{e}^{-i\left(k_x x + k_y y\right)} \tag{A56}$$

Equation (A49) is solved using the following boundary conditions:

$$-\frac{\partial\phi}{\partial z} = \frac{\partial\zeta}{\partial t} \quad \text{at } z = \zeta \tag{A57}$$

$$-\frac{\partial\phi}{\partial z} = 0 \quad \text{at } z = -h \tag{A58}$$

Further separating the variables, the assumed solution ϕ is written as

$$\phi\,(x, y, z, t) = Z(z)T(t)E\,(x, y) \tag{A59}$$

and is substituted into (A49), resulting in the following ordinary differential equation (ODE):

$$\frac{\mathrm{d}^2 Z}{\mathrm{d}z^2} - \left(k_x^2 + k_y^2\right) Z = 0 \tag{A60}$$

The differential equation (A60) has a general solution of the form

$$\mathrm{Z} = A\mathrm{e}^{-kz} + B\mathrm{e}^{kz} \tag{A61}$$

where

$$k = \sqrt{k_x^2 + k_y^2} \tag{A62}$$

and where A and B are constants of integration that can be determined using the boundary conditions (A57) and (A58). Using the general solution (A61) and lumping T in (A59) into A and B, ϕ can now be written as

$$\phi\left(x, y, z, t\right) = \left(A(t)\mathrm{e}^{-kz} + B(t)\mathrm{e}^{kz}\right) E\left(x, y\right) \tag{A63}$$

Applying the boundary condition (A57) and (A58), and solving simultaneously for $A(t)$ and $B(t)$, we obtain

$$A(t) = -\frac{1}{k}\frac{\mathrm{e}^{-kh}}{\mathrm{e}^{kh} - \mathrm{e}^{-kh}}\frac{\mathrm{d}\tilde{\zeta}}{\mathrm{d}t} \tag{A64}$$

$$B(t) = -\frac{1}{k}\frac{\mathrm{e}^{kh}}{\mathrm{e}^{kh} - \mathrm{e}^{-kh}}\frac{\mathrm{d}\tilde{\zeta}}{\mathrm{d}t} \tag{A65}$$

where (A57) has been evaluated at $z = 0$ instead of $z = \zeta$. Substituting for $A(t)$ and $B(t)$ using (A64) and (A65), it follows from (A63) that

$$\phi\left(x, y, z, t\right) = -\frac{1}{k}\frac{\cosh\left(k\left(z + h\right)\right)}{\sinh\left(kh\right)}\frac{\mathrm{d}\tilde{\zeta}(t)}{\mathrm{d}t} E\left(x, y\right) \tag{A66}$$

Equations (A50) and (A51) are solved for the magnetic potentials ψ^a and ψ using the boundary conditions

$$\hat{\mathbf{n}} \times \left(\mathbf{H} - \mathbf{H}^{\mathrm{a}}\right) = 0 \quad \text{at } z = \zeta \tag{A67}$$

$$\hat{\mathbf{n}} \cdot \left(\mathbf{B} - \mathbf{B}^{\mathrm{a}}\right) = 0 \quad \text{at } z = \zeta \tag{A68}$$

$$\lim_{z \to \infty} \psi^{\mathrm{a}} < \infty. \tag{A69}$$

$$\psi = \sum_{j=1}^{J} \psi_j(t)\delta^2\left(x - x_j\right)\left(y - y_j\right) \quad \text{at } z = -h \tag{A70}$$

Using (A34), (A40), and (A41), the boundary conditions (A67) and (A68) can be written more concisely as

$$\frac{\partial \psi^{\mathrm{a}}}{\partial x} - \frac{\partial \psi}{\partial x} - \frac{\chi}{\mu} B_0 \frac{\partial \zeta}{\partial x} = 0 \quad \text{at } z = \zeta \tag{A71}$$

$$\mu_0 \frac{\partial \psi^{\mathrm{a}}}{\partial z} = \mu \frac{\partial \psi}{\partial z} \quad \text{at } z = \zeta \tag{A72}$$

where only linear terms have been retained and, as needed, only one component of the vector equation (A67) has been used. Using the assumed solutions (A54) and (A55), a similar exercise as done for the velocity potential $\phi(x, y, z, t)$ above results in the following general solutions for $\psi^{\mathrm{a}}(x, y, z, t)$ and $\psi(x, y, z, t)$:

$$\psi^{\mathrm{a}}(x, y, z, t) = \left(A(t)\mathrm{e}^{-kz} + B(t)\mathrm{e}^{kz}\right) E(x, y) \tag{A73}$$

$$\psi(x, y, z, t) = \left(C(t)\mathrm{e}^{-kz} + D(t)\mathrm{e}^{kz}\right) E(x, y) \tag{A74}$$

where $A(t)$, $B(t)$, $C(t)$ and $D(t)$ are integration *constants* to be determined using the boundary conditions (A70), (A71), and (A72). Condition (A69) can be satisfied only if $B(t) = 0$. The boundary condition (A72) results in

$$D(t) - C(t) = -\frac{\mu_0}{\mu} A(t) \tag{A75}$$

where the condition has been evaluated at $z = 0$ instead of $z = \zeta$. Similarly, the application of (A71) yields

$$C(t) + D(t) = A(t) - \frac{\chi}{\mu} B_0 \tilde{\zeta} \tag{A76}$$

Simultaneous solution of (A75) and (A76) yields

$$C(t) = \frac{1}{2}\left[A(t)\left(1 + \frac{\mu_0}{\mu}\right) - \frac{\chi}{\mu} B_0 \tilde{\zeta}\right] \tag{A77}$$

$$D(t) = \frac{1}{2}\left[A(t)\left(1 - \frac{\mu_0}{\mu}\right) - \frac{\chi}{\mu} B_0 \tilde{\zeta}\right] \tag{A78}$$

Substituting (A77) and (A78) into (A74) and making use of known hyperbolic identities, $\psi^{\mathrm{a}}$and ψ can be written in terms of $A(t)$ as follows:

$$\psi^{\mathrm{a}}(x, y, z, t) = A(t)\mathrm{e}^{-kz}E(x, y) \tag{A79}$$

$$\psi(x, y, z, t) = \left(A(t)\left(\cosh(kz) - \frac{\mu_0}{\mu}\sinh(kz)\right) - \frac{\chi}{\mu}B_0\tilde{\zeta}(t)\cosh(kz)\right)E(x, y) \tag{A80}$$

The remaining unknown $A(t)$ can be evaluated using the last boundary condition (A70). Note that the only magnetic field variable required for the evaluation of the surface dynamic equation (4.25) is $(\partial\psi/\partial z)$ at $z = \zeta$, which is obtained using (A80) as follows:

$$\left.\frac{\partial\psi}{\partial z}\right|_{z=0} = -A(t)\frac{\mu_0}{\mu}kE(x, y) \tag{A81}$$

In what follows, $A(t)$ will be evaluated using the input magnetic potential (A70) as a boundary condition on ψ. In order to be able to apply the condition, the admissible values of k_x and k_y will be determined first. To that end, we apply the physically plausible boundary condition in the horizontal plane that the components of the velocity and magnetic fields normal to the container walls must be zero. Considering that the boundary condition must hold for all z and t, and for all mode numbers k_x and k_y, the condition translates to

$$\frac{\partial E}{\partial x} = 0 \quad \text{at} \quad x = 0,\ L_x \tag{A82}$$

and

$$\frac{\partial E}{\partial y} = 0 \quad \text{at} \quad y = 0,\ L_y \tag{A83}$$

These conditions are satisfied by the mode shapes

$$E = \cos(k_x x)\cos(k_y y) \tag{A84}$$

with the following characteristic equations:

$$\sin k_x x = 0 \quad \text{at } x = L_x \tag{A85}$$

$$\sin k_y y = 0 \quad \text{at } y = L_y \tag{A86}$$

Equations (A85) and (A86) can be satisfied by multiple discrete values of k_x and k_y, respectively, which are written in series form as follows:

$$k_x = \frac{(m-1)\,\pi}{L_x}, \quad m = 1, 2, \ldots$$
$$k_y = \frac{(n-1)\,\pi}{L_y}, \quad n = 1, 2, \ldots \tag{A87}$$

The discrete nature of k_x, k_y, and k — and of the variables dependent on these parameters — is indicated by writing them henceforth as k_m, k_n, and k_{mn}, respectively.

Using (A80), the total magnetic potential ψ at any point in the magnetic fluid can now be written as

$$\psi = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\left\{\left(A_{mn}\left(\cosh(k_{mn}z) - \frac{\mu_0}{\mu}\sinh(k_{mn}z)\right) - \frac{\chi}{\mu}B_0\tilde{\zeta}_{mn}\cosh(k_{mn}z)\right)E_{mn}\right\} \tag{A88}$$

However, the magnetic potential at the bottom of the fluid layer is specified by the boundary condition (A70). It follows from the comparison of conditions (A70) and (A88) (evaluated at $z = -h$) that

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\left\{\left(A_{mn}\left(\cosh(k_{mn}z) - \frac{\mu_o}{\mu}\sinh(k_{mn}z)\right) - \frac{\chi}{\mu}B_0\tilde{\zeta}_{mn}\cosh(k_{mn}z)\right)E_{mn}\right\}_{z=-h}$$
$$= \sum_{j=1}^{J}\psi_j\,\delta^2\left(x - x_j\right)\left(y - y_j\right) \tag{A89}$$

In order to determine the constants A_{mn}, we multiply both sides of (A89) with $\cos(k'_m x)\cos(k'_n y)$, such that

$$k'_m = \frac{(m'-1)\,\pi}{L_x}, \quad m' = 1, 2, \ldots$$
$$k'_n = \frac{(n'-1)\,\pi}{L_y}, \quad n' = 1, 2, \ldots, \tag{A90}$$

and integrate w.r.t. x and y over the domain $0 < x < L_x$, $0 < y < L_y$. The orthogonality of the mode shapes E_{mn} results in

$$\int_0^{L_x} \cos(k_m x)\cos(k'_m x)dx = \begin{cases} 0 & m \neq m' \\ L_x & m = m' = 1 \\ \frac{L_x}{2} & m = m' > 1 \end{cases}$$

$$\int_0^{L_y} \cos(k_n y)\cos(k'_n y)dy = \begin{cases} 0 & n \neq n' \\ L_y & n = n' = 1 \\ \frac{L_y}{2} & n = n' > 1 \end{cases} \tag{A91}$$

which allows to write the constants A_{mn} as

$$A_{mn} = \frac{1}{\cosh(k_{mn}h) + \frac{\mu_0}{\mu}\sinh(k_{mn}h)} \times \left(\frac{\chi}{\mu_0} B_0 \cosh(k_{mn}h)\,\tilde{\xi}_{mn} + \frac{c_m c_n}{L_x L_y} \sum_{j=1}^{J} \psi_j \cos\left(k_m x_j\right)\cos\left(k_n y_j\right) \right) \tag{A92}$$

where

$$c_m = \begin{cases} 1 \text{ for } m = 1 \\ 2 \text{ for } m > 1 \end{cases}$$

$$c_n = \begin{cases} 1 \text{ for } n = 1 \\ 2 \text{ for } n > 1 \end{cases} \tag{A93}$$

Substituting A_{mn} from (A92) into (A79) and (A80), the magnetic potentials ψ^{a} and ψ are fully determined, thus completing the solution of the Laplace equations (A49), (A50), and (A51).

Appendix B

Contents

Model Parameters

The parameters of the analytical model given in (4.102) are presented in this section. The model is given by

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$$
$$\mathbf{y} = \mathbf{C}\mathbf{x} \tag{B.1}$$

For $M = 4$, $N = 4$, the system order $N_d = 2N(2M + 1) = 72$. The system matrix $A \in \mathbb{R}^{N_d \times N_d}$ can be obtained using the following diagonal entries:

$-\omega_{mn}^2 \times 10^{-4} =$

$-0.026, -0.163, -0.513, -1.221, -0.071, -0.071, -0.296, -0.296,$
$-0.802, -0.802, -1.745, -1.745, -0.138, -0.138, -0.476, -0.476,$
$-1.171, -1.171, -2.381, -2.381, -0.233, -0.233, -0.714, -0.714, -1.629, -1.629,$
$-3.137, -3.137, -0.362, -0.362, -1.017, -1.017, -2.184, -2.184, -4.018, -4.018$

and

$-\omega_{d_{mn}}^2 \times 10^{-2} =$

$-0.138, -0.727, -1.787, -3.318, -0.350, -0.350, -1.174, -1.174, -2.470$
$-2.470, -4.236, -4.236, -0.629, -0.629, -1.691, -1.691, -3.222, -3.222$
$-5.224, -5.224, -0.971, -0.971, -2.273, -2.273 -4.042, -4.042, -6.280$
$-6.280, -1.374, -1.374, -2.921, -2.921, -4.929, -4.929, -7.404, -7.404$

The matrices $\mathbf{B} \in \mathbb{R}^{N_d \times L}$ and $\mathbf{C} \in \mathbb{R}^{M \times N_d}$, where L and M are the number of inputs and outputs, respectively, are produced using the following block matrices:

Z. Wu et al., *Modeling and Control of Magnetic Fluid Deformable Mirrors for Adaptive Optics Systems*, DOI 10.1007/978-3-642-32229-7,

$$\mathbf{B} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \times 10^3$$

$$\mathbf{C} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} \end{bmatrix}$$

The block matrices are given on the following pages:

B_{11} =

0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.011	0.010	0.010	0.010	0.010	0.009	0.009	0.007	0.007	0.008	0.007	0.007	0.008	0.006	0.007
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.185	0.098	0.116	0.098	0.098	0.059	0.078	-0.031	-0.031	-0.016	-0.031	-0.031	-0.016	-0.056	-0.045
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.652	0.034	0.139	0.034	0.034	-0.142	-0.062	-0.232	-0.232	-0.257	-0.232	-0.232	-0.257	-0.136	-0.190
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.294	-0.395	-0.213	-0.395	-0.395	-0.519	-0.498	0.156	0.156	-0.008	0.156	0.156	-0.008	0.367	0.287
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	-0.000	0.036	0.033	-0.009	-0.057	-0.043	0.012	0.034	0.059	0.068	0.059	0.023	-0.011	-0.043
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	-0.052	-0.030	0.040	0.051	0.021	-0.036	-0.067	-0.059	-0.034	0.000	0.034	0.064	0.064	0.051
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	-0.000	0.332	0.283	-0.076	-0.379	-0.329	0.018	0.051	0.152	0.101	0.088	0.060	0.007	-0.018
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	-0.440	-0.279	0.337	0.433	0.138	-0.276	-0.100	-0.088	-0.088	0.000	0.051	0.164	-0.042	0.021
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	-0.000	0.772	0.539	-0.146	-0.329	-0.469	-0.112	-0.322	-0.479	-0.644	-0.558	-0.189	0.105	0.424
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	-0.838	-0.647	0.642	0.826	0.120	-0.393	0.634	0.558	0.277	0.000	-0.322	-0.520	-0.596	-0.506
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	-0.000	0.737	0.239	-0.065	0.653	0.161	-0.060	-0.173	-0.640	-0.345	-0.299	-0.253	-0.083	-0.056
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	-0.372	-0.618	0.285	0.367	-0.238	0.135	0.340	0.299	0.370	0.000	-0.173	-0.695	0.470	0.067
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	-0.064	0.009	-0.011	-0.060	0.070	0.013	-0.136	-0.073	0.069	0.145	0.073	-0.105	-0.145	-0.026
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	-0.050	0.063	-0.022	-0.058	0.076	-0.050	-0.126	-0.119	0.000	0.126	0.088	-0.053	-0.149
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	-0.505	0.075	-0.088	-0.475	0.458	0.098	-0.362	-0.193	0.239	0.385	0.193	-0.366	-0.149	-0.048
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	-0.425	0.497	-0.173	-0.384	0.554	-0.132	-0.334	-0.413	0.000	0.334	0.307	-0.054	-0.273
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	-1.242	0.209	-0.216	-1.167	0.770	0.204	0.483	0.257	-0.120	-0.514	-0.257	0.183	0.747	0.123
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	-1.183	1.223	-0.425	-0.646	1.158	0.176	0.445	0.207	0.000	-0.445	-0.154	0.272	0.695

B_{21} =

0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	-1.443	0.308	-0.251	-1.356	0.206	0.158	0.896	0.477	-0.589	-0.954	-0.477	0.902	0.032	0.094
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	-1.749	1.421	-0.494	-0.173	0.896	0.326	0.826	1.020	0.000	-0.826	-0.757	0.012	0.532
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	-0.019	-0.048	0.027	-0.048	0.037	-0.110	-0.219	-0.000	0.219	0.000	-0.170	0.129	0.208
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.055	-0.033	0.027	-0.048	0.083	-0.064	0.190	0.000	-0.196	0.000	0.219	-0.098	-0.223	0.120
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	-0.160	-0.373	0.215	-0.321	0.270	-0.350	-0.700	-0.000	0.700	0.000	-0.665	0.226	0.514
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.430	-0.278	0.215	-0.373	0.555	-0.468	0.607	0.000	-0.767	0.000	0.700	-0.384	-0.392	0.297
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	-0.502	-1.056	0.610	-0.688	0.677	0.060	0.120	-0.000	-0.120	-0.000	-0.255	-0.383	-0.424
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	1.219	-0.870	0.610	-1.056	1.192	-1.173	-0.104	-0.000	-0.295	0.000	-0.120	-0.147	0.663	-0.245
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	-0.906	-1.630	0.941	-0.607	0.839	0.631	1.261	0.000	-1.261	-0.000	1.004	-0.302	-0.914
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	1.882	-1.570	0.941	-1.630	1.052	-1.454	-1.092	-0.000	1.159	0.000	-1.261	0.579	0.523	-0.528
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.038	-0.022	-0.036	0.029	0.014	-0.055	0.207	-0.135	-0.114	0.271	-0.135	0.040	0.269	-0.293
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	-0.000	-0.008	-0.013	0.025	-0.082	0.020	0.174	0.234	-0.198	0.000	0.234	-0.225	0.225	0.107
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.307	-0.188	-0.288	0.235	0.099	-0.407	0.740	-0.483	-0.486	0.966	-0.483	0.169	0.595	-0.847
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	-0.000	-0.069	-0.105	0.197	-0.562	0.148	0.621	0.837	-0.841	0.000	0.837	-0.957	0.499	0.308
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.967	-0.648	-0.909	0.741	0.248	-1.156	0.315	-0.206	-0.434	0.411	-0.206	0.151	-0.401	0.073
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	-0.000	-0.236	-0.331	0.622	-1.405	0.421	0.264	0.356	-0.752	0.000	0.356	-0.855	-0.337	-0.027
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	1.769	-1.339	-1.663	1.355	0.310	-1.800	-0.916	0.598	0.372	-1.196	0.598	-0.129	-0.816	1.230
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	-0.000	-0.488	-0.605	1.137	-1.756	0.655	-0.768	-1.035	0.644	0.000	-1.035	0.732	-0.685	-0.448

B_{12} =

0.000	0.000	0.000	0.000
0.006	0.006	0.007	0.008
0.000	0.000	0.000	0.000
-0.064	-0.064	-0.045	-0.016
0.000	0.000	0.000	0.000
-0.074	-0.074	-0.190	-0.257
0.000	0.000	0.000	0.000
0.388	0.388	0.287	-0.008
0.000	0.000	0.000	0.000
-0.055	-0.063	-0.058	-0.034
0.000	0.000	0.000	0.000
0.031	0.000	-0.033	-0.059
0.000	0.000	0.000	0.000
0.093	0.107	-0.024	-0.088
0.000	0.000	0.000	0.000
-0.054	-0.000	-0.014	-0.152
0.000	0.000	0.000	0.000
0.423	0.488	0.572	0.277
0.000	0.000	0.000	0.000
-0.244	-0.000	0.330	0.479
0.000	0.000	0.000	0.000
-0.655	-0.756	-0.075	0.370
0.000	0.000	0.000	0.000
0.378	0.000	-0.044	0.640
0.000	0.000	0.000	0.000
0.078	0.155	0.075	-0.069
0.000	0.000	0.000	0.000
-0.134	-0.000	0.131	0.119
0.000	0.000	0.000	0.000
0.018	0.037	0.138	-0.239
0.000	0.000	0.000	0.000
-0.032	-0.000	0.240	0.413
0.000	0.000	0.000	0.000
-0.388	-0.775	-0.353	0.120
0.000	0.000	0.000	0.000
0.671	0.000	-0.611	-0.207

B_{22} =

0.000	0.000	0.000	0.000
0.225	0.450	-0.270	0.589
0.000	0.000	0.000	0.000
-0.390	-0.000	-0.468	-1.020
0.000	0.000	0.000	0.000
0.000	-0.270	-0.000	0.196
0.000	0.000	0.000	0.000
0.270	0.000	-0.240	0.000
0.000	0.000	0.000	0.000
0.000	-0.287	-0.000	0.767
0.000	0.000	0.000	0.000
0.287	0.000	-0.593	0.000
0.000	0.000	0.000	0.000
-0.000	0.910	0.000	0.295
0.000	0.000	0.000	0.000
-0.910	-0.000	0.490	0.000
0.000	0.000	0.000	0.000
-0.000	0.029	0.000	-1.159
0.000	0.000	0.000	0.000
-0.029	-0.000	1.055	-0.000
0.000	0.000	0.000	0.000
-0.192	0.384	-0.156	-0.114
0.000	0.000	0.000	0.000
-0.333	-0.000	0.270	-0.198
0.000	0.000	0.000	0.000
-0.299	0.599	-0.451	-0.486
0.000	0.000	0.000	0.000
-0.519	-0.000	0.781	-0.841
0.000	0.000	0.000	0.000
0.429	-0.857	0.039	-0.434
0.000	0.000	0.000	0.000
0.742	0.000	-0.067	-0.752
0.000	0.000	0.000	0.000
0.272	-0.543	0.654	0.372
0.000	0.000	0.000	0.000
0.471	0.000	-1.133	0.644

C_{11} =

21.000	0.000	21.000	0.000	21.000	0.000	21.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
18.896	0.000	11.075	0.000	1.080	0.000	-6.4060	0.000	-0.000	0.000	-9.388	0.000	-0.000	0.000	-12.215
19.379	0.000	13.153	0.000	4.469	0.000	-3.4590	0.000	6.496	0.000	-5.451	0.000	9.222	0.000	-7.738
18.896	0.000	11.075	0.000	1.080	0.000	-6.4060	0.000	6.035	0.000	7.192	0.000	7.852	0.000	9.358
18.896	0.000	11.075	0.000	1.080	0.000	-6.4060	0.000	-1.630	0.000	9.246	0.000	-2.121	0.000	12.030
17.760	0.000	6.651	0.000	-4.588	0.000	-8.4170	0.000	-10.201	0.000	3.713	0.000	-10.512	0.000	3.826
18.355	0.000	8.890	0.000	-1.994	0.000	-8.0850	0.000	-7.801	0.000	-6.546	0.000	-9.134	0.000	-7.664
14.069	0.000	-3.536	0.000	-7.482	0.000	2.5360	0.000	2.118	0.000	-12.010	0.000	0.489	0.000	-2.774
14.069	0.000	-3.536	0.000	-7.482	0.000	2.5360	0.000	6.098	0.000	-10.561	0.000	1.408	0.000	-2.439
14.891	0.000	-1.794	0.000	-8.281	0.000	-0.1240	0.000	10.572	0.000	-6.104	0.000	4.207	0.000	-2.429
14.069	0.000	-3.536	0.000	-7.482	0.000	2.5360	0.000	12.195	0.000	0.000	0.000	2.816	0.000	0.000
14.069	0.000	-3.536	0.000	-7.482	0.000	2.5360	0.000	10.561	0.000	6.098	0.000	2.439	0.000	1.408
14.891	0.000	-1.794	0.000	-8.281	0.000	-0.1240	0.000	4.175	0.000	11.471	0.000	1.662	0.000	4.565
12.325	0.000	-6.307	0.000	-4.380	0.000	5.9490	0.000	-2.042	0.000	11.579	0.000	0.207	0.000	-1.172
13.212	0.000	-5.050	0.000	-6.136	0.000	4.6510	0.000	-7.742	0.000	9.226	0.000	-0.495	0.000	0.590
11.412	0.000	-7.285	0.000	-2.372	0.000	6.2910	0.000	-9.822	0.000	5.670	0.000	2.580	0.000	-1.489
11.412	0.000	-7.285	0.000	-2.372	0.000	6.2910	0.000	-11.341	0.000	0.000	0.000	2.979	0.000	-0.000
13.212	0.000	-5.050	0.000	-6.136	0.000	4.6510	0.000	-10.431	0.000	-6.022	0.000	-0.667	0.000	-0.385
14.891	0.000	-1.794	0.000	-8.281	0.000	-0.1240	0.000	-6.104	0.000	-10.572	0.000	-2.429	0.000	-4.207

C_{12} =

21.000	0.000	21.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.080	0.000	-6.406	0.000	0.000	0.000	-9.3880	0.000	-0.000	0.000	-12.215	0.000	-0.000	0.000	-9.190
4.469	0.000	-3.459	0.000	6.496	0.000	-5.4510	0.000	9.222	0.000	-7.738	0.000	8.457	0.000	-7.096
1.080	0.000	-6.406	0.000	6.035	0.000	7.1920	0.000	7.852	0.000	9.358	0.000	5.907	0.000	7.040
1.080	0.000	-6.406	0.000	-1.630	0.000	9.2460	0.000	-2.121	0.000	12.030	0.000	-1.596	0.000	9.050
-4.588	0.000	-8.417	0.000	-10.201	0.000	3.7130	0.000	-10.512	0.000	3.826	0.000	-3.610	0.000	1.314
-1.994	0.000	-8.085	0.000	-7.801	0.000	-6.5460	0.000	-9.134	0.000	-7.664	0.000	-5.138	0.000	-4.312
-7.482	0.000	2.536	0.000	2.118	0.000	-12.010	0.000	0.489	0.000	-2.774	0.000	-1.225	0.000	6.949
-7.482	0.000	2.536	0.000	6.098	0.000	-10.561	0.000	1.408	0.000	-2.439	0.000	-3.528	0.000	6.111
-8.281	0.000	-0.124	0.000	10.572	0.000	-6.104	0.000	4.207	0.000	-2.429	0.000	-5.251	0.000	3.032
-7.482	0.000	2.536	0.000	12.195	0.000	0.000	0.000	2.816	0.000	0.000	0.000	-7.057	0.000	0.000
-7.482	0.000	2.536	0.000	10.561	0.000	6.098	0.000	2.439	0.000	1.408	0.000	-6.111	0.000	-3.528
-8.281	0.000	-0.124	0.000	4.175	0.000	11.471	0.000	1.662	0.000	4.565	0.000	-2.074	0.000	-5.698
-4.380	0.000	5.949	0.000	-2.042	0.000	11.579	0.000	0.207	0.000	-1.172	0.000	1.152	0.000	-6.534
-6.136	0.000	4.651	0.000	-7.742	0.000	9.226	0.000	-0.495	0.000	0.590	0.000	4.652	0.000	-5.544
-2.372	0.000	6.291	0.000	-9.822	0.000	5.670	0.000	2.580	0.000	-1.489	0.000	4.637	0.000	-2.677
-2.372	0.000	6.291	0.000	-11.341	0.000	0.000	0.000	2.979	0.000	-0.000	0.000	5.354	0.000	-0.000
-6.136	0.000	4.651	0.000	-10.431	0.000	-6.022	0.000	-0.667	0.000	-0.385	0.000	6.268	0.000	3.619
-8.281	0.000	-0.124	0.000	-6.104	0.000	-10.572	0.000	-2.429	0.000	-4.207	0.000	3.032	0.000	5.251

C_{13} =

0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	-9.388	0.000	0.000	0.000	-12.215	0.000	0.000	0.000	-9.190	0.000	0.000	0.000	-2.418
6.496	0.000	-5.451	0.000	9.222	0.000	-7.738	0.000	8.457	0.000	-7.096	0.000	4.785	0.000	-4.015
6.035	0.000	7.192	0.000	7.852	0.000	9.358	0.000	5.907	0.000	7.040	0.000	1.554	0.000	1.853
-1.630	0.000	9.246	0.000	-2.121	0.000	12.030	0.000	-1.596	0.000	9.050	0.000	-0.420	0.000	2.382
-10.201	0.000	3.713	0.000	-10.512	0.000	3.826	0.000	-3.610	0.000	1.314	0.000	4.238	0.000	-1.543
-7.801	0.000	-6.546	0.000	-9.134	0.000	-7.664	0.000	-5.138	0.000	-4.312	0.000	1.045	0.000	0.877
2.118	0.000	-12.010	0.000	0.489	0.000	-2.774	0.000	-1.225	0.000	6.949	0.000	-0.389	0.000	2.207
6.098	0.000	-10.561	0.000	1.408	0.000	-2.439	0.000	-3.528	0.000	6.111	0.000	-1.121	0.000	1.941
10.572	0.000	-6.104	0.000	4.207	0.000	-2.429	0.000	-5.251	0.000	3.032	0.000	-4.158	0.000	2.401
12.195	0.000	0.000	0.000	2.816	0.000	0.000	0.000	-7.057	0.000	0.000	0.000	-2.241	0.000	0.000
10.561	0.000	6.098	0.000	2.439	0.000	1.408	0.000	-6.111	0.000	-3.528	0.000	-1.941	0.000	-1.121
4.175	0.000	11.471	0.000	1.662	0.000	4.565	0.000	-2.074	0.000	-5.698	0.000	-1.642	0.000	-4.512
-2.042	0.000	11.579	0.000	0.207	0.000	-1.172	0.000	1.152	0.000	-6.534	0.000	-0.538	0.000	3.050
-7.742	0.000	9.226	0.000	-0.495	0.000	0.590	0.000	4.652	0.000	-5.544	0.000	-0.363	0.000	0.433
-9.822	0.000	5.670	0.000	2.580	0.000	-1.489	0.000	4.637	0.000	-2.677	0.000	-4.252	0.000	2.455
-11.341	0.000	0.000	0.000	2.979	0.000	-0.000	0.000	5.354	0.000	0.000	0.000	-4.910	0.000	0.000
-10.431	0.000	-6.022	0.000	-0.667	0.000	-0.385	0.000	6.268	0.000	3.619	0.000	-0.490	0.000	-0.283
-6.104	0.000	-10.572	0.000	-2.429	0.000	-4.207	0.000	3.032	0.000	5.251	0.000	2.401	0.000	4.158

C_{14} =

0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	-12.215	0.000	0.000	0.000	-9.190	0.000	0.000	0.000	-2.418	0.000	-4.198	0.000	0.000
9.222	0.000	-7.738	0.000	8.457	0.000	-7.096	0.000	4.785	0.000	-4.015	0.000	0.580	0.000	-3.287
7.852	0.000	9.358	0.000	5.907	0.000	7.040	0.000	1.554	0.000	1.853	0.000	-0.729	0.000	4.134
-2.121	0.000	12.030	0.000	-1.596	0.000	9.050	0.000	-0.420	0.000	2.382	0.000	-3.944	0.000	-1.436
-10.512	0.000	3.826	0.000	-3.610	0.000	1.314	0.000	4.238	0.000	-1.543	0.000	4.579	0.000	-3.843
-9.134	0.000	-7.664	0.000	-5.138	0.000	-4.312	0.000	1.045	0.000	0.877	0.000	0.883	0.000	5.009
0.489	0.000	-2.774	0.000	-1.225	0.000	6.949	0.000	-0.389	0.000	2.207	0.000	-8.980	0.000	-3.268
1.408	0.000	-2.439	0.000	-3.528	0.000	6.111	0.000	-1.121	0.000	1.941	0.000	-4.778	0.000	-8.276
4.207	0.000	-2.429	0.000	-5.251	0.000	3.032	0.000	-4.158	0.000	2.401	0.000	4.520	0.000	-7.828
2.816	0.000	0.000	0.000	-7.057	0.000	0.000	0.000	-2.241	0.000	0.000	0.000	9.556	0.000	0.000
2.439	0.000	1.408	0.000	-6.111	0.000	-3.528	0.000	-1.941	0.000	-1.121	0.000	4.778	0.000	8.276
1.662	0.000	4.565	0.000	-2.074	0.000	-5.698	0.000	-1.642	0.000	-4.512	0.000	-6.924	0.000	5.810
0.207	0.000	-1.172	0.000	1.152	0.000	-6.534	0.000	-0.538	0.000	3.050	0.000	-9.544	0.000	-3.474
-0.495	0.000	0.590	0.000	4.652	0.000	-5.544	0.000	-0.363	0.000	0.433	0.000	-1.725	0.000	-9.783
2.580	0.000	-1.489	0.000	4.637	0.000	-2.677	0.000	-4.252	0.000	2.455	0.000	5.107	0.000	-8.846
2.979	0.000	-0.000	0.000	5.354	0.000	-0.000	0.000	-4.910	0.000	0.000	0.000	10.214	0.000	0.000
-0.667	0.000	-0.385	0.000	6.268	0.000	3.619	0.000	-0.490	0.000	-0.283	0.000	4.967	0.000	8.603
-2.429	0.000	-4.207	0.000	3.032	0.000	5.251	0.000	2.401	0.000	4.158	0.000	-4.520	0.000	7.828

C_{15} =

$$\begin{bmatrix}
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
2.636 & 0.000 & -0.000 & 0.000 & 5.371 & 0.000 & 0.000 & 0.000 & 7.750 & 0.000 & 0.000 & 0.000 \\
-1.620 & 0.000 & -0.590 & 0.000 & -3.596 & 0.000 & -1.309 & 0.000 & -5.867 & 0.000 & -2.136 & 0.000 \\
-2.477 & 0.000 & -0.902 & 0.000 & -5.047 & 0.000 & -1.837 & 0.000 & -7.283 & 0.000 & -2.651 & 0.000 \\
2.019 & 0.000 & 1.695 & 0.000 & 4.114 & 0.000 & 3.452 & 0.000 & 5.937 & 0.000 & 4.982 & 0.000 \\
0.851 & 0.000 & -4.828 & 0.000 & 1.376 & 0.000 & -7.801 & 0.000 & 1.357 & 0.000 & -7.693 & 0.000 \\
-3.496 & 0.000 & 1.272 & 0.000 & -6.417 & 0.000 & 2.335 & 0.000 & -7.885 & 0.000 & 2.870 & 0.000 \\
6.364 & 0.000 & 5.340 & 0.000 & 1.750 & 0.000 & 1.468 & 0.000 & -4.012 & 0.000 & -3.366 & 0.000 \\
-4.154 & 0.000 & 7.194 & 0.000 & -1.142 & 0.000 & 1.978 & 0.000 & 2.618 & 0.000 & -4.535 & 0.000 \\
-4.175 & 0.000 & -7.231 & 0.000 & -2.411 & 0.000 & -4.176 & 0.000 & 1.628 & 0.000 & 2.820 & 0.000 \\
8.307 & 0.000 & 0.000 & 0.000 & 2.284 & 0.000 & 0.000 & 0.000 & -5.237 & 0.000 & 0.000 & 0.000 \\
-4.154 & 0.000 & 7.194 & 0.000 & -1.142 & 0.000 & 1.978 & 0.000 & 2.618 & 0.000 & -4.535 & 0.000 \\
1.450 & 0.000 & -8.223 & 0.000 & 0.837 & 0.000 & -4.749 & 0.000 & -0.565 & 0.000 & 3.207 & 0.000 \\
5.115 & 0.000 & 4.292 & 0.000 & -2.229 & 0.000 & -1.871 & 0.000 & -3.574 & 0.000 & -2.999 & 0.000 \\
-7.281 & 0.000 & 2.650 & 0.000 & 0.405 & 0.000 & -0.148 & 0.000 & 5.387 & 0.000 & -1.961 & 0.000 \\
-2.574 & 0.000 & -4.458 & 0.000 & 2.380 & 0.000 & 4.122 & 0.000 & 1.190 & 0.000 & 2.062 & 0.000 \\
5.148 & 0.000 & -0.000 & 0.000 & -4.760 & 0.000 & 0.000 & 0.000 & -2.381 & 0.000 & 0.000 & 0.000 \\
-3.874 & 0.000 & 6.711 & 0.000 & 0.216 & 0.000 & -0.374 & 0.000 & 2.866 & 0.000 & -4.964 & 0.000 \\
-4.175 & 0.000 & -7.231 & 0.000 & -2.411 & 0.000 & -4.176 & 0.000 & 1.628 & 0.000 & 2.820 & 0.000
\end{bmatrix}$$

Appendix C

Contents

Modal Reconstruction of the Wavefront Shape

The wavefront slope data provided by the WFS is used to reconstruct the wavefront shape. Any appropriate set of 2D basis functions may be used for the reconstruction purpose. A standard reconstruction algorithm using Zernike mode shapes is described in the following.

Assume that for any wavefront profile $\zeta(x, y)$ measured by the SHWS, the wavefront slope readings are available at K lenslets subapertures. The wavefront slope readings can be assembled into a vector $\mathbf{s} \in \mathbb{R}^{2K}$ as

$$\mathbf{s} = \left[s_{1x}, s_{1y}, s_{2x}, s_{2y}, \ldots, s_{kx}, s_{ky}, \ldots, s_{Kx}, s_{Ky}\right]^{\mathrm{T}} \tag{C.1}$$

where $\left(s_{kx}, s_{ky}\right)$ is the local average slope of the wavefront, $\zeta(x, y)$, at the kth lenslet, and can be computed as follows:

$$\begin{aligned} s_{kx} &= \frac{\iint\limits_{k} \frac{\partial \zeta(x,y)}{\partial x} \mathrm{d}x\mathrm{d}y}{\iint\limits_{k} \mathrm{d}x\mathrm{d}y} = \frac{\Delta x_{sk}}{f} \\ s_{ky} &= \frac{\iint\limits_{k} \frac{\partial \zeta(x,y)}{\partial y} \mathrm{d}x\mathrm{d}y}{\iint\limits_{k} \mathrm{d}x\mathrm{d}y} = \frac{\Delta y_{sk}}{f} \end{aligned} \tag{C.2}$$

Z. Wu et al., *Modeling and Control of Magnetic Fluid Deformable Mirrors for Adaptive Optics Systems*, DOI 10.1007/978-3-642-32229-7,

Here, f is the focal length of the lenslet, and Δx_{sk} and Δy_{sk} are the displacement of the sensor focal spots in the horizontal and vertical directions, respectively, from the center of the lenslet k.

For any number J of Zernike modes specified for the reconstruction of the wavefront shape, the reconstruction matrix $\tilde{\mathbf{Z}} \in \mathbb{R}^{2K\times J}$ is given as

$$\tilde{\mathbf{Z}} = \begin{bmatrix} Z_{11x} & \cdots & Z_{j1x} & \cdots & Z_{J1x} \\ Z_{11y} & \cdots & Z_{j1y} & \cdots & Z_{J1y} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ Z_{1kx} & \cdots & Z_{jkx} & \cdots & Z_{Jkx} \\ Z_{1ky} & \cdots & Z_{jky} & \cdots & Z_{Jky} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ Z_{1Kx} & \cdots & Z_{jKx} & \cdots & Z_{JKx} \\ Z_{1Ky} & \cdots & Z_{jKy} & \cdots & Z_{JKy} \end{bmatrix}$$

where

$$Z_{jkx} = \frac{\iint\limits_k \frac{\partial Z_j(x,y)}{\partial x}\mathrm{d}x\mathrm{d}y}{\iint\limits_k \mathrm{d}x\mathrm{d}y}$$

$$Z_{jky} = \frac{\iint\limits_k \frac{\partial Z_j(x,y)}{\partial y}\mathrm{d}x\mathrm{d}y}{\iint\limits_k \mathrm{d}x\mathrm{d}y}$$

are the average slopes of the jth Zernike mode ($j = 1, 2, \ldots, J$) at the kth lenslet ($k = 1, 2, \ldots, K$). The wavefront constructed from the measured slope vector, $\mathbf{s}$, is

$$\mathbf{c} = \tilde{\mathbf{Z}}^{\dagger}\mathbf{s} \tag{C.3}$$

where $\tilde{\mathbf{Z}}^{\dagger} \in \mathbb{R}^{J\times 2K}$ is the pseudo-inverse of the reconstruction matrix $\tilde{\mathbf{Z}}$ and $\mathbf{c} = [c_1, c_2, \ldots, c_J]^{\mathrm{T}}$ is the vector of the reconstructed Zernike coefficients for the measured wavefront $\zeta(x, y)$. The set of Zernike coefficients can be used to estimate the wavefront displacement at any point in the pupil. For a given set of points (x_m, y_m), $m = 1, 2, \ldots, M$ a vector of wavefront displacements, $\mathbf{y} = [y_1, y_2, \ldots, y_m, \ldots, y_M]^T$, may be computed as

$$\mathbf{y} = \mathbf{Z}\mathbf{c} \tag{C.4}$$

where $\mathbf{Z} \in \mathbb{R}^{M \times J}$ is the matrix of Zernike mode shapes computed at the specified points (x_m, y_m):

$$\mathbf{Z} = \begin{bmatrix} Z_{11} & \cdots & Z_{1j} & \cdots & Z_{1J} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ Z_{m1} & \cdots & Z_{mj} & \cdots & Z_{mJ} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ Z_{M1} & \cdots & Z_{Mj} & \cdots & Z_{MJ} \end{bmatrix}$$

Z_{mj} being the jth Zernike mode shape computed at (x_m, y_m).

Appendix D

Contents

Alignment Procedure

In this Appendix, the design of optical setup used in the AO system proposed in Chap. 5 is presented. The steps necessary to calculate the focal lengths and positions of the various lenses are presented. A schematic of the system can be seen in Fig. D.1, where L_i represents lens i. In the rest of this Appendix, f_i is used to represent the focal length of lens i. A collimated laser light, which has a wavelength of 661 nm, leaves the laser at a diameter of h_l and travels through a set of lenses 1 and 2, used to magnify the laser light to match the pupil size of the eye, h_e. The laser light is directed into the eye using a pellicle beam splitter and reflects off the retina. The light beam leaves the eye at a diameter of h_e and travels through lens 3 and lens 4. Lenses 3 and 4 are used to magnify the laser light diameter to match the active surface area of the DM, h_m. The laser light reflects off the DM and passes through lenses 5 and 6, which magnifies the light beam to match the aperture of the wavefront sensor, h_s. The light beam then travels from lens 6 to the wavefront sensor.

With respect to the AO system described above, this section begins by discussing the computations necessary to create a general two-lens optical relay telescope used throughout the AO system and then applies these steps to assemble the entire AO system. The first-order Gaussian (or paraxial) theory is used to determine the focal lengths and placement of different lenses in the AO system. In the following, a brief discussion of geometrical optics and paraxial theory is first presented (Hetch 2002). Figure D.2 depicts a wave from point source S on a spherical interface of radius r centered at C. The ray (SA) will be refracted at the interface toward the optical axis. Assume that at some point P, it will cross the axis, as will all other rays incident at the same angle θ_i. Fermat's principle maintains that the optical path length (OPL)

Z. Wu et al., *Modeling and Control of Magnetic Fluid Deformable Mirrors for Adaptive Optics Systems*, DOI 10.1007/978-3-642-32229-7,

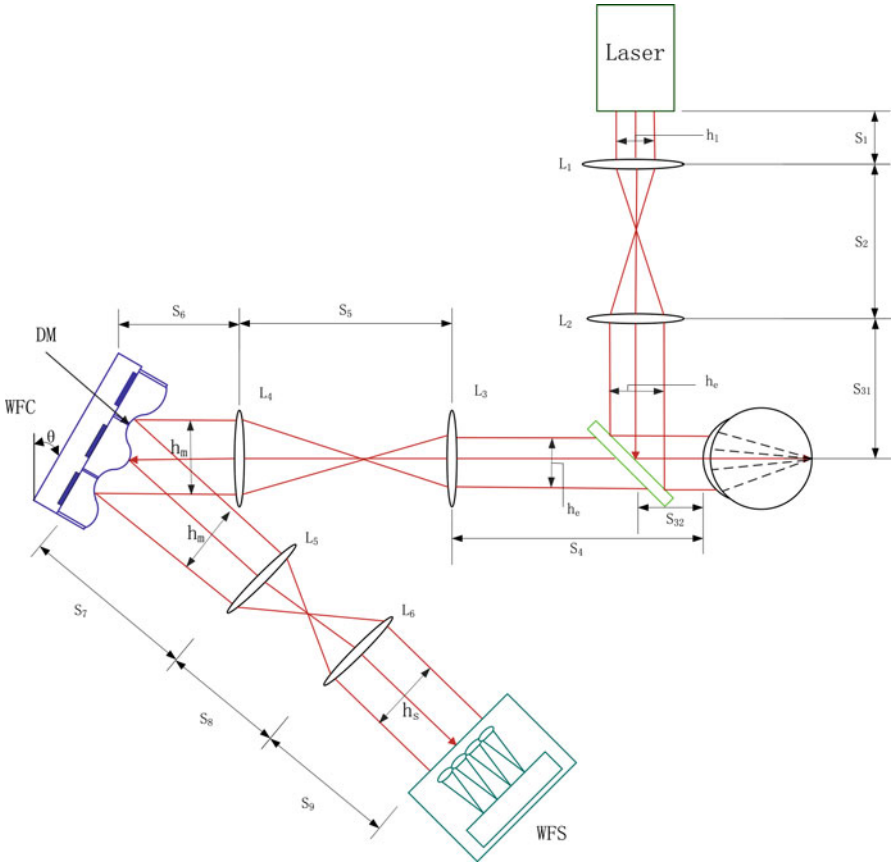


Fig. D.1 Schematic of the experimental AO system used in the laboratory

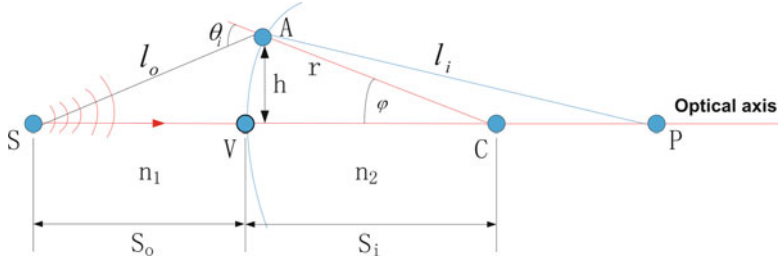


Fig. D.2 Refraction at a spherical interface

will be stationary, that is, its derivative with respect to the position variable φ (see Fig. D.2) will be zero. For the light ray in question,

$$(\text{OPL}) = n_1 l_0 + n_2 l_i \tag{D1}$$

where n_i is the index of refraction of medium i. Using the law of cosines in triangles SAC and ACP along with the fact that

$$\cos\varphi = -\cos(180 - \varphi)$$

$$l_0 = \left[r^2 + (s_0 + r)^2 - 2r(s_0 + r)\cos\varphi\right]^{1/2}$$

$$l_i = \left[r^2 + (s_i + r)^2 - 2r(s_i + r)\cos\varphi\right]^{1/2}$$

The OPL can be rewritten as

$$(\text{OPL}) = n_1\left[r^2 + (s_0 + r)^2 - 2r(s_0 + r)\cos\varphi\right]^{1/2} + n_2\left[r^2 + (s_i + r)^2 - 2r(s_i + r)\cos\varphi\right]^{1/2}$$

From Fig. D.2, it can be seen that point A moves at the end of a fixed radius r, φ is the position variable, and therefore setting $\frac{\mathrm{d}(\text{OPL})}{\mathrm{d}\varphi} = 0$, via Fermat's principle

$$\frac{n_1 r(s_0 + r)\sin\varphi}{2l_0} - \frac{n_2 r(s_i - r)\sin\varphi}{2l_i} = 0$$

from which it follows

$$\frac{n_l}{l_0} + \frac{n_2}{l_i} = \frac{1}{r}\left(\frac{n_2 s_i}{l_i} - \frac{n_1 s_0}{l_0}\right) \tag{D2}$$

This relationship must hold among the parameters for a ray going from S to P by way of refraction at the spherical interface. It is known that if A is moved to a new location by changing φ, the new light ray will not intercept the optical axis at P. The approximations used to represent l_i and l_0 are vital in the following calculations. Recall the following approximations:

$$\cos\varphi = 1 - \frac{\varphi^2}{2!} + \frac{\varphi^4}{4!} - \frac{\varphi^6}{6!} + \cdots$$

and

$$\sin\varphi = \varphi - \frac{\varphi^3}{3!} + \frac{\varphi^5}{5!} - \frac{\varphi^7}{7!} + \cdots$$

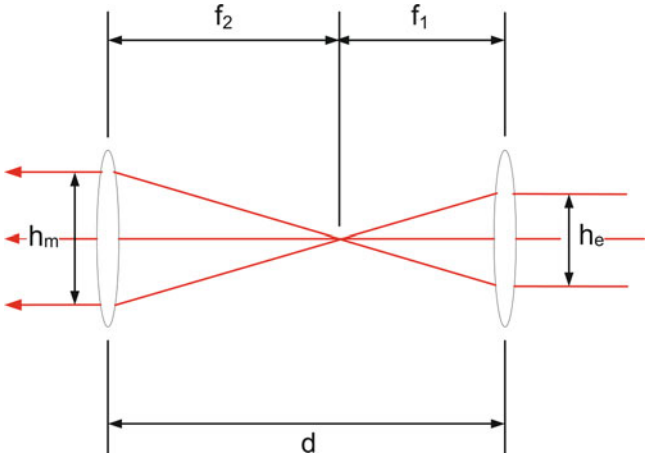


Fig. D.3 Two thin lenses separated by a distance $f_1 + f_2$

If small values for φ are assumed (i.e., A is close to V), $\cos\varphi \approx 1$. The expression for l_i and l_0 yields $l_0 \approx s_0$, $l_i \approx s_i$, and (D2) can be approximated as

$$\frac{n_1}{l_0} + \frac{n_2}{l_i} = \frac{n_2 - n_1}{r} \tag{D3}$$

This approximation is the foundation of what is called *first-order theory* (similarly third-order theory uses the approximation $\sin\varphi = \varphi - (\varphi^3/3!)$). Light rays that arrive at a shallow angle with respect to the optical axis (such that φ and h are appropriately small) are known as *paraxial rays*. The emerging wavefront segments corresponding to these paraxial rays are essentially spherical and will form a "perfect" image at its center P located at s_i. It should be noted that equation (D3) is independent of the location of A over a small area about the optical axis, called the paraxial region. Under this approximation, the analysis is known as first-order, paraxial, or Gaussian optics. If the optical system is well corrected, an incident spherical wave will emerge in a form closely resembling a spherical wave. As the system becomes free of defects, it more closely resembles first-order theory. Use of paraxial analysis provides a convenient measure of the quality of an optical system.

The discussion begins by describing the relay telescope consisting of lenses 3 and 4. In order to ensure a collimated beam is directed toward the DM with a diameter of h_m, first-order Gaussian (or paraxial) theory is used to determine the focal lengths and placement of lenses 3 and 4. The focal lengths and placements of lenses 1, 2, 5, and 6 can then be determined in a similar way. Figure D.3 shows two thin lenses separated by a distance d and are used to magnify a collimated beam of light from a diameter h_e to a diameter h_m. In order to magnify the beam of light from a diameter h_e to a diameter h_m, the following magnifications are required:

$$M_T = \frac{h_m}{h_e} \tag{D4}$$

Since the distance between the lenses must be equal to the sum of their focal lengths in order to collimate the beam traversing between the lenses, the magnification can also be related to the ratio of the focal lengths. From the geometry of the light rays shown in Fig. D.3, it is apparent that

$$\frac{h_m}{f_4} = \frac{h_e}{f_3} \tag{D5}$$

Using the ratio in equation (D5), the magnification can be rewritten as

$$M_T = \frac{h_m}{h_e} = \frac{f_4}{f_3} \tag{D6}$$

In order for the collimated incident light ray, traveling from the eye to lens 3, to remain collimated when it travels from lens 4 to the DM, two conditions must be satisfied:

(i) The pupil of the eye, the DM, and the wavefront sensors lenslet array must be in conjugate planes.
(ii) The distance d between the two lenses must equal $f_3 + f_4$.

Using equation (D6), the specifications for the components in the AO system can be calculated as follows. The focal lengths of lenses 3 and 4 can be chosen based on the magnification needed for the laser light leaving the eye and directed toward the DM, M_{T1}:

$$M_{T1} = \frac{f_3}{f_4} = \frac{h_e}{h_m} \tag{D7}$$

The distance between the lenses 3 and 4 in Fig. D.1 is chosen as

$$S_5 = f_3 + f_4 \tag{D8}$$

The distances S_4 and S_6 are chosen to be f_3 and f_4, respectively. This is done to satisfy condition (i). Similarly, the focal lengths of lenses 5 and 6 can be chosen based on the magnification needed for the laser light leaving the DM and directed toward the wavefront sensor, M_{T2}:

$$M_{T2} = \frac{f_5}{f_6} = \frac{h_m}{h_s} \tag{D9}$$

The distance between lenses 5 and 6, Fig. D.1, is chosen as

$$S_8 = f_5 + f_6 \tag{D10}$$

The distances S_7 and S_9 must be chosen equal be to f_5 and f_6, respectively, in order to satisfy condition (i). Next, the focal lengths of lenses 1 and 2, shown

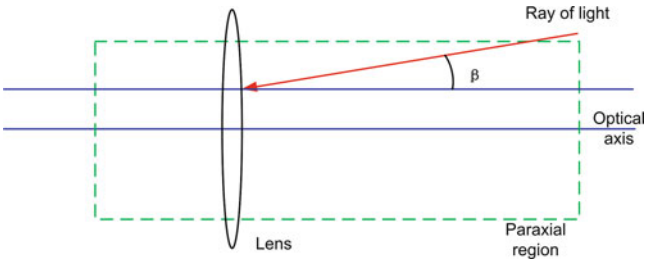


Fig. D.4 Ray of light entering the lens

in Fig. D.1, can be chosen based on the magnification needed for the laser light entering the eye. Namely,

$$M_{T3} = \frac{f_1}{f_2} = \frac{h_l}{h_e} \tag{D11}$$

The distance between lenses 1 and 2 is chosen as

$$S_2 = f_1 + f_2 \tag{D12}$$

The distances S_1 and S_3 (where $S_3 = S_{31} + S_{32}$) are chosen according to the boundary conditions

(1) $S_4 + S_5 + S_6 < OpWidth$ and $(S_7 + S_8 + S_9)\sin(90 - \theta) < OpWidth$
(2) $((S_7 + S_8 + S_9)\cos(90 - \theta)) + S_1 + S_2 + S_3 < OpLen$

where *OpWidth* and *OpLen* are the width and length of the optical table holding the AO system and θ is the angle of tilt of the DM in Fig. D.1. Next, given that the first-order Gaussian theory only holds within the paraxial region, the following conditions must be satisfied:

(iii) That the angle β in Fig. D.4 between the incident ray entering a lens and the optical axis must be small.
(iv) The incident ray should be close to the optical axis.

Given assumptions (iii) and (iv), the diameter of the lenses should be chosen at least two to three times larger than the diameter of the beam propagating through the lenses. For example,

$$\begin{aligned} \phi_1 &= 3h_l \\ \phi_2 &= 3h_e \\ \phi_3 &= 3h_e \\ \phi_4 &= 3h_m \\ \phi_5 &= 3h_m \\ \phi_6 &= 3h_s \end{aligned} \tag{D13}$$

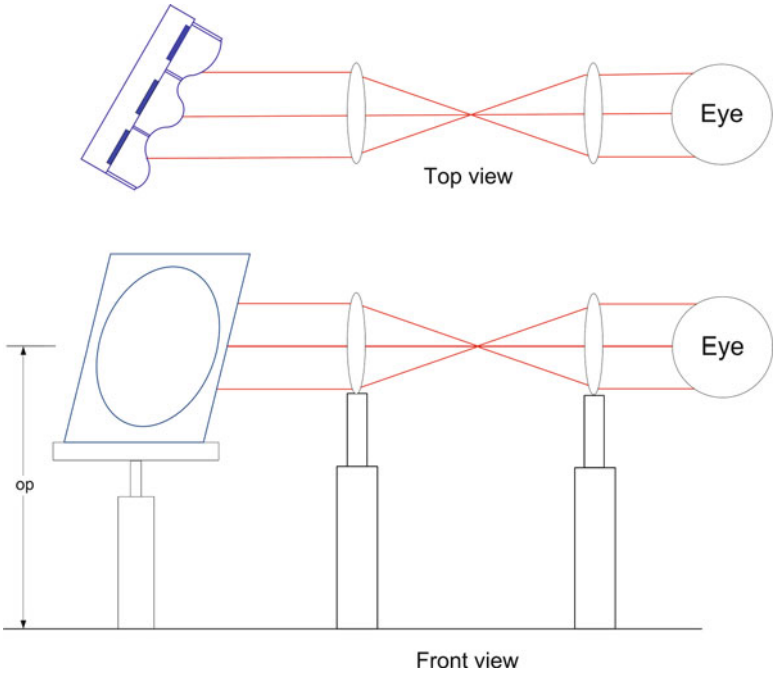


Fig. D.5 AO system experimental setup, showing the path of light from the eye to the DM

Table D.1 Lenses used in the experimental AO system

Lens	Diameter (mm)	Focal length (mm)
1	12.50	40.00
2	50.80	200.00
3	25.40	100.00
4	50.80	400.00
5	50.80	300.00
6	12.50	50.00

Table D.2 The diameter of the laser light in the AO system

Laser position	Diameter (mm)
h_l	1.1
h_e	5.5
h_m	22
h_s	3.6

where ϕ_i is the diameter of lens i. The height of the optical path, Op, can be seen in Fig. D.5, which shows the path of light from the eye to the DM. The height can be set at the time of the development of the optical setup since the post holders and mounts offer adjustable heights.

Therefore, given the above information, a method for choosing the lenses for the AO experimental setup in the laboratory can be written in a series of general steps as follows:

(1) Separate the setup into optical paths.
(2) Determine h_e,h_m,h_s, and h_l.
(3) Determine the magnification needed to achieve the diameters in step 2, using equations (D7), (D9), and (D11).
(4) Determine the diameter of each of the lenses using equation (D13).
(5) Find the focal length of an appropriate lens in a vendor's catalogue.
(6) Calculate distances S_1, S_2, S_3, S_4, S_5, S_6, S_7, $S_{8,}$ and S_9 using equations (D8), (D10), and (D12) with condition (i) and boundary conditions (1) and (2).

It should be noted that the angle θ in Fig. D.1 should be as small as possible to minimize any distortion of the image reflecting from the DM to the wavefront sensor. The lenses used in the AO system presented in Chap. 5 were chosen using steps (1)–(6). The results are summarized in Table D.1. The diameter of the laser light as it passes through the AO system is shown in Table D.2.

Index

Z. Wu et al., *Modeling and Control of Magnetic Fluid Deformable Mirrors for Adaptive Optics Systems*, DOI 10.1007/978-3-642-32229-7,