Ben M. Chen,
Tong H. Lee
and
Venkatakrishnan Venkataramanan

Hard Disk Drive Servo Systems

•ADVANCES IN INDUSTRIAL CONTROL•

Springer

Advances in Industrial Control

Springer-Verlag London Ltd.

Other titles published in this Series:

Feedback Control Theory for Dynamic Traffic Assignment
Pushkin Kachroo and Kaan Özbay

Autotuning of PID Controllers
Cheng-Ching Yu

Robust Aeroservoelastic Stability Analysis
Rick Lind and Marty Brenner

Performance Assessment of Control Loops: Theory and Applications
Biao Huang and Sirish L. Shah

Data Mining and Knowledge Discovery for Process Monitoring and Control
Xue Z. Wang

Advances in PID Control
Tan Kok Kiong, Wang Quing-Guo and Hang Chang Chieh with Tore J. Hägglund

Advanced Control with Recurrent High-order Neural Networks: Theory and Industrial Applications
George A. Rovithakis and Manolis A. Christodoulou

Structure and Synthesis of PID Controllers
Aniruddha Datta, Ming-Tzu Ho and Shankar P. Bhattacharyya

Data-driven Techniques for Fault Detection and Diagnosis in Chemical Processes
Evan L. Russell, Leo H. Chiang and Richard D. Braatz

Bounded Dynamic Stochastic Systems: Modelling and Control
Hong Wang

Non-linear Model-based Process Control
Rashid M. Ansari and Moses O. Tadé

Identification and Control of Sheet and Film Processes
Andrew P. Featherstone, Jeremy G. VanAntwerp and Richard D. Braatz

Precision Motion Control
Tan Kok Kiong, Lee Tong Heng, Dou Huifang and Huang Sunan

Nonlinear Identification and Control: A Neural Network Approach
Guoping Liu

Digital Controller Implementation and Fragility: A Modern Perspective
Robert S.H. Istepanian and James F. Whidborne

Optimisation of Industrial Processes at Supervisory Level
Doris Sáez, Aldo Cipriano and Andrzej W. Ordys

Applied Predictive Control
Huang Sunan, Tan Kok Kiong and Lee Tong Heng

Ben M. Chen, Tong H. Lee and
Venkatakrishnan Venkataramanan

Hard Disk Drive Servo Systems

With 80 Figures

Springer

Ben M. Chen, PhD
Tong H. Lee, PhD
Venkatakrishnan Venkataramanan, PhD
Department of Electrical & Computer Engineering,
National University of Singapore, 4 Engineering Drive 3, Singapore 117576

British Library Cataloguing in Publication Data
Chen, Ben., 1963-
Hard disk drive servo systems. - (Advances in industrial control)
1.Servomechanisms 2.Data disk drives - Design 3.Hard disks (Computer science)
I.Title II.Lee, Tong H. III.Venkataramanan, Venkatakrishnan
629.8'323

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress.

ISBN 978-1-4471-3714-6 ISBN 978-1-4471-3712-2 (eBook)
DOI 10.1007/978-1-4471-3712-2

http://www.springer.co.uk

Originally published by Springer-Verlag London Limited in 2002.
Softcover reprint of the hardcover 1st edition 2002

Typesetting: Electronic text files prepared by authors

69/3830-543210 Printed on acid-free paper SPIN 10836568

Advances in Industrial Control

Professor Emeritus O. P. Malik
Department of Electrical and Computer Engineering
University of Calgary
2500, University Drive, NW
Calgary
Alberta
T2N 1N4
Canada

Doctor K.-F. Man
Electronic Engineering Department
City University of Hong Kong
Tat Chee Avenue
Kowloon
Hong Kong

Professor G. Olsson
Department of Industrial Electrical Engineering and Automation
Lund Institute of Technology
Box 118
S-221 00 Lund
Sweden

Professor A. Ray
Pennsylvania State University
Department of Mechanical Engineering
0329 Reber Building
University Park
PA 16802
USA

Professor D. E. Seborg
Chemical Engineering
3335 Engineering II
University of California Santa Barbara
Santa Barbara
CA 93106
USA

Doctor I. Yamamoto
Technical Headquarters
Nagasaki Research & Development Center
Mitsubishi Heavy Industries Ltd
5-717-1, Fukahori-Machi
Nagasaki 851-0392
Japan

To our families

SERIES EDITORS' FOREWORD

The series *Advances in Industrial Control* aims to report and encourage technology transfer in control engineering. The rapid development of control technology has an impact on all areas of the control discipline. New theory, new controllers, actuators, sensors, new industrial processes, computer methods, new applications, new philosophies..., new challenges. Much of this development work resides in industrial reports, feasibility study papers and the reports of advanced collaborative projects. The series offers an opportunity for researchers to present an extended exposition of such new work in all aspects of industrial control for wider and rapid dissemination.

From time to time a particular practical control problem emerges as a challenge to the design capabilities of the control community. One example has been the activated sludge process in wastewater systems where the process is highly nonlinear and measurements are few. A second example is the hard disk drive servo system. These widely used systems are critical to the operation of modern computing devices. They are nonlinear and demand a high-precision control system for the operations of track seeking and track following. There are also alternative actuation systems available to achieve these objectives. In this *Advances in Industrial Control* monograph B.M. Chen, T.H. Lee and V. Venkataramanan at the National University of Singapore provide a thorough presentation of the technical background, the modelling and control solutions for this benchmark problem of hard disk drive servo control systems. It is a monograph which encompasses physical system descriptions, modelling, identification, linear control and nonlinear control. The issue of implementing discrete control solutions makes an important appearance. Professor Chen and his colleagues also describe the test facilities that they have used for assessing the performance of proposed control solutions.

It would be difficult to find a better example of the full range of software and hardware tools used by the modern control engineer to solve a challenging advanced technical problem. The monograph should be an inspiration to many students studying control engineering today. There is much to learn from the monograph on the validity of advanced linear (H_2 and H_∞) control and how nonlinear control can be applied. Thus the monograph should be of considerable

interest to postgraduate students, academics and practising industrial engineers alike.

M.J. Grimble and M.A. Johnson
Industrial Control Centre
Glasgow, Scotland, U.K.

PREFACE

Nowadays, it is pretty hard for us to imagine what life would be like without computers and what computers would be like without hard disks. Hard disks provide an important data-storage medium for computers and other data-processing systems. Many of us can still recall that the storage medium used on computers in the 1960s and 1970s was actually paper, which was later replaced by magnetic tapes. The key technological breakthrough that enabled the creation of the modern hard disks came in the 1950s, when a group of researchers and engineers in IBM made the very first production hard disk, IBM 305 RAMAC (Random Access Method of Accounting and Control). The first generation of hard disks used in personal computers in the early 1980s had a capacity of 10 megabytes and cost over $100 per megabyte. Modern hard disks have capacities approaching 100 gigabytes and cost less than 1 cent per megabyte.

In modern hard disk drives (HDDs), rotating disks coated with a thin magnetic layer or recording medium are written with data that are arranged in concentric circles or tracks. Data are read or written with a read/write (R/W) head, which consists of a small horseshoe-shaped electromagnet. The two main functions of the R/W head positioning servomechanism in disk drives are track seeking and track following. Track seeking moves the R/W head from the current track to a specified destination track in minimum time using a bounded control effort. Track following maintains the head as close as possible to the destination track centre while information is being read from or written to the disk. Track density is the reciprocal of the track width. It is suggested that, on a disk surface, tracks should be written as closely spaced as possible so that we can maximise the usage of the disk surface. This means an increase in the track density, which subsequently means a more stringent requirement on the allowable variations of the position of the heads from the true track centre. The prevalent trend in hard disk design is towards smaller hard disks with increasingly larger capacities. This implies that the track width has to be smaller, leading to lower error tolerance in the positioning of the head. The controller for track following has to achieve tighter regulation in the control of the servomechanism.

The scope of this book is to provide a systematic treatment on the design of modern HDD servo systems. In particular, we will focus on the applications of some newly developed results in control theory, namely the robust and perfect tracking control, and the composite nonlinear feedback control. Emphasis will be made on HDD servo systems with either a single-stage voice-coil-motor actuator or a dual-stage actuator in which an additional micro-actuator is attached to a conventional voice-coil-motor actuator to provide faster response and hence higher bandwidth in the track following stage. Most of the results presented in this book are from research carried out by the authors and their co-workers over the last few years.

The intended audience of this book includes practising engineers in hard disk and CD-ROM drive industries and researchers in areas related to servo systems and engineering. An appropriate background for this monograph would be some senior level and/or first year graduate level courses in linear systems and multivariable control. Some knowledge of control techniques for systems with actuator nonlinearities would certainly be helpful.

We have the benefit of the collaboration of several co-workers, from whom we have learnt a great deal. Many of the results presented in this monograph are the results of our collaboration. Among these co-workers are Professor Chang C. Hang of the National University of Singapore, Dr Siri Weerasooriya of Quantum Corporation, Dr Tony Huang of Seagate Technology Colorado, and Dr Guoxiao Guo of the Data Storage Institute of Singapore. We are indebted to all of them for their contributions.

We are grateful to Professor Zongli Lin of the University of Virginia, for his invaluable comments and discussions on the subject related to the composite nonlinear feedback control technique of Chapter 5. This technique, originally proposed by Zongli and his workers and later enhanced by us, has emerged as an effective tool in designing HDD servo systems. We are also indebted to Professor Iven Mareels of the University of Melbourne, who is now visiting our department here at the National University of Singapore, for many beneficial discussions on related subjects, especially the issue on HDD servo systems with a dual actuator.

The first two authors of this monograph would like to thank their current and former graduate students, especially Yi Guo, Xiaoping Hu, Lan Wang, Teck-Beng Goh, Kexiu Liu, Zhongming Li and Chen Lin, for their help and contributions. We are also indebted to Dr Kemao Peng, our Research Fellow, for various help throughout the preparation of this book, especially for his help in obtaining the experimental results of Chapters 7 and 9, and to Kavitha and the Copy-Editor of Springer for their kindly help in proof reading the whole monograph.

We would like to acknowledge the National University of Singapore for providing us with the funds for two research projects (R-263-000-038-112 and R-263-000-096-112) on the development of the dual-stage actuated servo systems of HDDs. We are also grateful to Dr Teck-Seng Low, Dr Tow-Chong Chong and Dr Guoxiao Guo of the Data Storage Institute of Singapore for their support to our projects.

Last, but certainly not the least, we owe a debt of gratitude to our families for their sacrifice, understanding and encouragement. It is very natural that we dedicate this book to all of them.

Ben M. Chen
Tong H. Lee
Venkatakrishnan Venkataramanan

Singapore, 2001

CONTENTS

CHAPTER 1
INTRODUCTION AND PREVIEW

1.1 Introduction

Hard disk drives (HDDs) provide an important data-storage medium for computers and other data-processing systems. In most HDDs, rotating disks coated with a thin magnetic layer or recording medium are written with data that are arranged in concentric circles or tracks. Data are read or written with a read/write (R/W) head, which consists of a small horseshoe-shaped electromagnet. Figure 1.1 shows a simple illustration of a typical hard disk servo system with a voice-coil-motor (VCM) actuator.

The two main functions of the R/W head positioning servomechanism in disk drives are track seeking and track following. Track seeking moves the R/W head from the present track to a specified destination track in minimum time using a bounded control effort. Track following maintains the head as close as possible to the destination track centre while information is being read from or written to the disk. Track density is the reciprocal of the track width. It is suggested that, on a disk surface, tracks should be written as closely spaced as possible so that we can maximise the usage of the disk surface. This means an increase in the track density, which subsequently means a more stringent requirement on the allowable variations of the position of the heads from the true track centre.

The prevalent trend in hard disk design is towards smaller hard disks with increasingly larger capacities. This implies that the track width has to be smaller, which leads to lower error tolerance in the positioning of the head. The controller for track following has to achieve tighter regulation in the control of the servomechanism. Basically, the servo system of an HDD can be divided into three stages, *i.e.*, the track seeking, track settling and track following stages (see Figure 1.2 for a detailed illustration). Current HDDs use a combination of classical control techniques, such as the proximate time optimal control technique in the tracking seeking stage, and lead-lag compensators, proportional-integral-derivative (PID) compensators in the track following stage, plus some notch filters to reduce the effects of high-frequency resonant modes (see *e.g.*, [1–16] and references cited therein). These classical

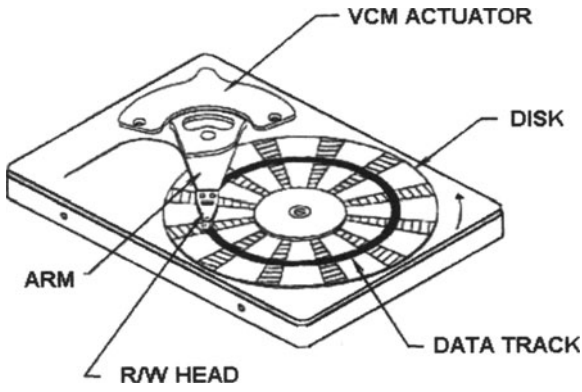


Figure 1.1. A typical HDD with a VCM actuator servo system.

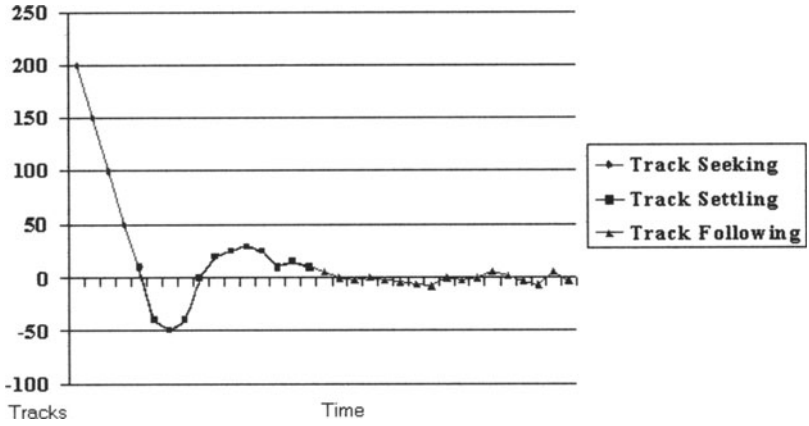


Figure 1.2. Track seeking and following of an HDD servo system.

methods can no longer meet the demand for HDDs of higher performance. Thus, many control approaches have been tried, such as the linear quadratic Gaussian (LQG) with loop transfer Recovery (LTR) approach (see *e.g.*, [17–19]), H_∞ control approach (see *e.g.*, [20–25]), and adaptive control (see *e.g.*, [26–29]) and so on. Although much work has been conducted to date, more studies need to be done to achieve better performance in HDDs.

The scope of this book is to provide a systematic treatment on the design of modern HDD servo systems. In particular, we will focus on the applications of some newly developed results in control theory, *i.e.*, robust and perfect tracking (RPT) control, which is suitable for track following, and composite nonlinear feedback (CNF) control, which is for track seeking and following.

The emphasis will be on HDD servo systems with either a single-stage VCM actuator or a dual-stage actuator in which an additional microactuator is attached to a conventional VCM actuator to provide faster response and hence higher bandwidth in the track following stage. Most of the results presented in this book are from research carried out by the authors and their co-workers over the last few years. The purpose of this book is to discuss various aspects of the subject under a single cover.

1.2 Mechanical Structure of HDDs

The mechanical structure of a typical modern hard disk drive is depicted in Figure 1.3. The authors of this book are thankful to Seagate Technology for granting permission to use this figure in our work. A brief description (see also, *e.g.*, [30]) is given below:

1. DEVICE ENCLOSER. This is the most important component, as it determines the reliability of the disk drives. It helps to keep the contamination low. With the aid of recirculation and a breather filter, it keeps out dust and other contamination that could enter between the R/W heads and the platters over which they float, and reduces the possibility of head crashes. The two major parts, the base casting and top cover, are sealed with a gasket. The base casting provides supports for the spindle, actuator, VCM yoke and electronics card.

2. DISK. Every hard disk will have one or more flat rotating disks, each with two magnetic surfaces, called platters. These are made of either an Al–Mg alloy substrate material electro-less plated with Ni–P, or a mixture of glass and ceramic. The magnetic material, to allow data storage, is applied as a thin coating on both sides of each platter together with a carbon overcoat. The surfaces of each platter are precision machined and treated to remove any imperfections, and attention is paid during the manufacturing process to ensure a very smooth surface.

3. ACTUATOR ASSEMBLY. This consists of a VCM, flex cable or printed circuit cable, actuator arms and crash-stops at both ends of travel. The data are read/written from/to the platters using the R/W heads mounted on the top and bottom surfaces of each platter. The heads are supported by the actuator arm. The actuator in HDDs, *i.e.*, the VCM actuator, is so named as it works like a loudspeaker. The electrical input to the VCM is supplied through a flex cable. The coil of the VCM actuator extends between a yoke/magnets. The write-driver/pre-amplifier is often part of the actuator assembly, which is mounted on a flex cable.

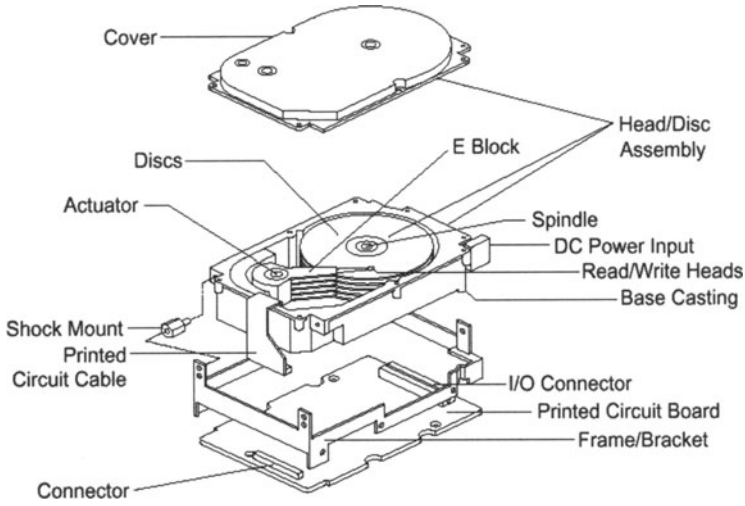


Figure 1.3. Mechanical structure of a typical HDD.

4. HEAD/SUSPENSION ASSEMBLY. The R/W heads are of ferrite, metal-in-gap, thin film or magnetoresistive (MR) types. Older types, *i.e.*, ferrite, metal-in-gap, and thin film, used the principle of electromagnetic induction, whereas the modern disk drive heads use MR heads, which use the principle of change of magnetoresistance. Both read and write operations in older disk drives were performed by a single head, but the modern HDDs use separate heads for read and write operations. These heads are positioned only micro-inches above the recording medium on an air bearing surface, which is often referred to as a slider. A gimbal attaches the slider to a stainless steel suspension to allow for pitch and roll, and the suspension is attached to the arm of the actuator by a ball swaging.

5. SPINDLE AND MOTOR ASSEMBLY. These are responsible for turning the hard disk platters with stable, reliable and consistent turning power for thousands of hours of often continuous use. All hard disks use servo-controlled DC spindle motors and are configured for direct connection, *i.e.*, there are no belts or gears that are used to connect them to the hard disk platter spindle. The critical component of the hard disk's spindle motor is the set of spindle motor bearings at each end of the spindle shaft. These bearings are used to turn the platters smoothly. The disk clamper and spacers are other important parts of this assembly.

6. ELECTRONICS CARD: This provides an interface to the host personal computer (PC). The most common interfaces used are the integrated

drive electronics (IDE), the advanced technology attachment (ATA), and the small computer systems interface (SCSI), which all use integrated electronic circuits. These integrated circuits have a power driver for the spindle motor, VCM, R/W electronics, servo demodulator, controller chip for timing control and control of interface, micro-controller/digital signal processor (DSP) for servo control and control interface, and ROM and RAM for micro-code and data transfer.

Lastly, we note that a fairly complete report on the basic mechanical and electrical structures of hard and floppy disk drives can be found in a book by Zhang [31].

1.3 Historical Development of HDDs

The first generation of hard disks used in PCs had a capacity of 10 megabytes (MB) and cost over $100 per MB. Modern hard disks have capacities approaching 100 gigabytes (GB) and cost less than 1 cent per MB. This represents an improvement of 1,000,000% in less than 20 years and now it is cumulatively improving at 70% per year. At the same time, the speed of the hard disk and its interfaces has also increased dramatically.

Some of the very earliest computers had no storage at all. Each time a program had to be run it would have to be entered manually. It was realised then that, to utilise the power of computers fully, there was a need for permanent storage.

During the initial search for permanent storage, paper played a major role in human life. The computer scientists were also psychologically influenced by *paper*. This led to the use of paper as the first storage medium on computers, though magnetic storage had already gained momentum by that time. Programs and data were recorded using holes punched into paper tapes or punch cards to represent a "1", and paper blocks to represent a "0" (or *vice versa*). This type of storage was used for many years until the creation of magnetic tapes. However, these tapes also lost their place when random access to the data was needed for quick and efficient usage of data stored. Thus, an improvement needed to be sought. Disk drive development took an eventful spin, when IBM announced, in May 1955, a product that offered unprecedented random-access storage to 5 million characters each of 7-bit.

These early prototypes had the heads of the hard disk in contact with the disk surface. This was done to allow the low-sensitivity electronics to be able to better read the magnetic fields on the disk surface. However, owing to the fact that manufacturing techniques were not nearly as sophisticated as

they are now, it was not possible to produce a disk surface that was smooth enough for the head to slide smoothly over it at high speed while in contact with the surface. As a result, the heads and the magnetic coating on the surface of the disk would wear out over time. Thus the problem of reliability was not addressed.

IBM engineers working under R. Johnson at IBM in San Jose, California, between 1952 and 1954, realised that, with the proper design, the R/W heads could be suspended above the disk surface and read the bits as they passed underneath. This critical discovery, that contact with the surface of the disk was no longer necessary, was implemented as IBM 305 RAMAC (Random Access Method of Accounting and Control), introduced on September 13, 1956. This early version stored 5 million characters on 50 disks, each 24″ in diameter. The capacity was approximately 5 MB. Its bit density was about 2,000 bits per square inch and the data transfer rate was an impressive 8,800 bytes per second. Over the succeeding years, the technology improved incrementally; bit density, capacity and performance all increased.

Next, we summarise the interesting history of the hard disk. In what follows, we present lists of some historical "firsts" and new trends in the development of HDDs. These lists are generated from the following sources on the net: www.pcguide.com, www.storage.ibm.com, www.storagereview.com and www.mkdata.dk [32, 33].

1.3.1 Chronological List of Developments in HDDs

There have been a number of important "firsts" in the world of hard disks over their first 40 years or so. The following is a list, in chronological order, of some of the products developed during the past half-century that introduced key or important technologies in HDDs.

- FIRST HARD DISK (1956): IBM 305 RAMAC was introduced. It had a capacity of about 5 MB, stored on fifty 24″ disks. Its bit density was a mere 2,000 bits per square inch and its data throughput was about 8,800 MB per second.

- FIRST AIR BEARING HEADS (1962): IBM's model 1301 lowered the flying height of the R/W heads to 250 micro-inches. It had a 28 MB capacity on half as many heads as the original RAMAC, and increased both bit density and throughput by about 1000%.

- FIRST REMOVABLE DISK DRIVE (1965): IBM's model 2310 was the first disk drive with a removable disk pack. While many PC users think of removable hard disks as being a modern invention, in fact they were very popular in the 1960s and 1970s.

- FIRST FERRITE HEADS (1966): IBM's model 2314 was the first hard disk to use ferrite core heads, the first type later used on PC hard disks.

- FIRST MODERN HARD DISK DESIGN (1973): IBM's model 3340, nicknamed the *Winchester*, was introduced. With a capacity of 60 MB, it introduced several key technologies that led to it being considered by many as the ancestor of the modern disk drives.

- FIRST THIN FILM HEADS (1979): IBM's model 3370 was the first with thin film heads, which would for many years be the standard in the HDD industry.

- FIRST 8″ FORM FACTOR DISK DRIVE (1979): IBM's model 3310 was the first disk drive with 8″ platters, greatly reduced in size from the 14″ that had been the standard for over a decade.

- FIRST 5.25″ FORM FACTOR DISK DRIVE (1980): Seagate's ST-506 was the first drive in the 5.25″ form factor, used in the earliest PCs.

- FIRST 3.5″ FORM FACTOR DISK DRIVE (1983): Rodime introduced RO352, the first disk drive to use the 3.5″ form factor, which became one of the most important industry standards.

- FIRST EXPANSION CARD DISK DRIVE (1985): Quantum introduced the Hardcard, a 10.5 MB hard disk mounted on an industry standard architecture (ISA) expansion card for PCs that were originally built without a hard disk. This product put Quantum "on the map" so to speak.

- FIRST VOICE-COIL-ACTUATOR 3.5″ DRIVE (1986): Conner Peripherals introduced CP340, the first disk drive to use a voice-coil actuator.

- FIRST "LOW-PROFILE" 3.5″ DISK DRIVE (1988): Conner Peripherals introduced CP3022, which was the first 3.5″ drive to use the reduced 1″ height, now called *low profile* and the standard for modern 3.5″ drives.

- FIRST 2.5″ FORM FACTOR DISK DRIVE (1988): PrairieTek introduced a drive using 2.5″ platters. This size later became a standard for portable computing.

- FIRST DRIVE WITH MR HEADS AND PARTIAL RESPONSE AND MAXIMUM LIKELIHOOD (PRML) DATA DECODING (1990): IBM's model 681 (Redwing), an 857 MB drive, was the first to use MR heads and PRML data decoding.

- FIRST THIN FILM DISKS (1991): IBM's *Pacifica* mainframe drive was the first to replace oxide media with thin film media on the platter surface.

- FIRST 1.8″ FORM FACTOR DISK DRIVE (1991): Integral Peripherals' 1820 was the first hard disk with 1.8″ platters, later used for PC-card disk drives.
- FIRST 1.3″ FORM FACTOR DISK DRIVE (1992): Hewlett Packard's C3013A is the first 1.3″ drive.
- FIRST 1″ HIGH 1 GB DISK DRIVE (1993): IBM unveiled the world's first 1″ high 1 GB disk drive, storing 354 million bits per square inch.
- FIRST 7,200 RPM ULTRA ATA-INTERFACE DISK DRIVE (1997): Industry's first of this kind for desktop computers from Seagate Technology.
- FIRST 10,000 RPM DISK DRIVE (1998): Seagate Technology introduced the first 10,000 rpm drives, *i.e.*, the 9.1 GB (ST19101) and 4.55 GB (ST34501) Cheetah family.
- FIRST ULTRA ATA/100 DISK DIVES (2000): Seagate announced the first Ultra ATA/100 interface on its Barracuda ATA II disk drive, the industry's fastest desktop PC disk drive.
- LARGEST HDD (2000): At the time of the preparation of this monograph, Seagate's Barracuda 180 is the largest single drive in the world. It has a capacity of 180 GB.

1.3.2 Trends in the Development of HDD Systems

In spite of a slow change in the basic design of hard disks over the years, accelerated improvements in terms of their capacity, storage, reliability and other characteristics have been made. In what follows, the various trends are highlighted.

- BIT DENSITY: The bit density of hard disk platters continues to increase at an amazing rate, even exceeding some of the optimistic predictions of a few years ago. Densities in the laboratory are now exceeding 35 Gbits per square inch, and modern disks are now packing as much as 20 GB of data onto a single 3.5″ platter.
- CAPACITY: Hard disk capacity continues to increase at an accelerating rate. From 10 MB in 1981, the normal capacity is now well over 20-30 GB. Consumer drives would most likely have a capacity of 100 GB within a couple of years.
- SPINDLE SPEED: The move to faster and faster spindle speeds continues. Since increasing the spindle speed improves both random access and sequential performance, this is likely to continue. 7,200 rpm spindles are now

standard on mainstream IDE/ATA drives. A 15,000 rpm SCSI drive was announced by Seagate in early 2000.

- FORM FACTOR: The trend in form factors is downward: to smaller and smaller drives. 5.25″ drives have now all but disappeared from the mainstream market, with 3.5″ drives dominating the desktop and server segment. In the mobile world, 2.5″ drives are the standard, with smaller sizes becoming more prevalent; IBM in 1999 announced its Microdrive, a tiny 170 MB or 340 MB device, only 1″ in diameter and less than 0.25″ thick. Over the next few years, desktop and server drives are likely to make a transition to the 2.5″ form factor as well. The primary reasons for this "shrinking trend" include the enhanced rigidity of smaller platters, reduction of mass to enable faster spin speeds, and improved reliability due to enhanced ease of manufacturing.

- PERFORMANCE: Both positioning and transfer performance factors are improving. The speed with which data can be pulled from the disk is increasing more rapidly than the improvement of positioning performance, suggesting that, over the next few years, addressing seek time and latency will be the areas of greatest value to hard disk engineers.

- REDUNDANT ARRAYS OF INEXPENSIVE DISKS (RAID): In the province of only high-end servers, the use of multiple disk arrays to improve performance and reliability is becoming increasingly common, and is now seen even in consumer desktop machines.

- RELIABILITY: The reliability of hard disks is improving slowly as manufacturers refine their processes and add new reliability-enhancing features, but this characteristic is not changing nearly as rapidly as the others above. It is simply very hard to improve the reliability of a product when it is changing rapidly.

- INTERFACES: Despite the introduction to the PC world of new interfaces, such as the IEEE-1394 and universal serial bus (USB), the mainstream interfaces are the same as they were through the 1990s: IDE/ATA and SCSI. The interfaces themselves continue to create new and improved standards with higher maximum transfer rates, to match the increase in performance of the hard disks themselves.

1.4 Implementation Setup

To make our work more complete, we have implemented almost all of our designs on actual HDDs with some highly advanced and accurate equipment.

In what follows, we briefly summarise the key software and hardware tools used to obtain the simulation and implementation results.

1. MATLAB AND SIMULINK. All off-line computation and simulation of the results in this book are done using the well-known products from Math Work, MATLAB 5.3 with its simulation package SIMULINK 3.0.

2. LINEAR SYSTEMS AND CONTROL TOOLBOX. The linear systems and control toolbox [34], developed under a MATLAB environment by the first author of the book and his co-worker, Zongli Lin of the University of Virginia, collects a few tens of m-functions. These m-functions realise algorithms for computing linear system structures (such as the finite and infinite zero structures, invertibility structures, and many other properties) and algorithms for computing H_2 and H_∞ optimal controllers, as well as controllers that solve the H_∞ almost disturbance decoupling problem, and robust and perfect tracking problem. The toolbox has been used intensively to carry out the design of the HDD servo systems throughout the book.

3. THE dSPACE DSP SYSTEM. A dSPACE (DSP) system is used in the actual implementation throughout the book. The system has the following main components:

   - *The dSPACE Add-on Card.* The main component of the dSPACE DSP system is its add-on card, DS1102, which is built upon a Texas Instruments TMS320C31 floating-point DSP. The DSP has been supplemented by a set of analog-to-digital (A/D) and digital-to-analog (D/A) converters, a DSP micro-controller-based digital I/O subsystem and incremental sensor interfaces. Some major features of this add-on card are:

     a) a TMS320C31 floating-point DSP;

     b) two 16-bit 250 kHz and two 12-bit 800 kHz sampling A/D converters with input span of ± 10 V;

     c) a quad 12-bit D/A converter with programmable output voltages;

     d) a 16-bit fixed point digital I/O, a bit-selectable-parallel I/O port, four timers, six PWM circuits, and a serial interface.

   - *Real-Time Interface (RTI) and Real-Time Workshop (RTW).* The RTI acts as a link between SIMULINK and the dSPACE hardware. It has built-in hardware control functions and blocks for DS1102 add-on card based on SIMULINK. This, together with the RTW, automatically generates real-time codes from SIMULINK off-line models and implements these codes on the dSPACE real-time hardware.

- *The dSPACE Control Desk.* This is a software platform that combines all the above tools of dSPACE for controlling, monitoring, and automating the implementation process on the actual HDDs.

4. POLYTEC LASER DOPPLER VIBROMETER (LDV). The Polytec LDV is an optical instrument for accurately measuring velocity and displacement of vibrating surfaces completely without contact. The LDV system consists of two main components: (1) an optical sensor head or fibre optic unit (both are laser interferometers), which measures the dynamic Doppler shift from the vibrating object; and (2) a controller (processor), which provides power to the optics and demodulates the Doppler information using various types of Doppler signal decoder electronics, thereby producing an analog vibration signal (velocity and/or displacement) that can be viewed/measured by the customer using commercially available fast Fourier transform (FFT) analysers and oscilloscopes. This instrument is used to measure the displacement and velocity of the R/W heads of HDDs.

5. DYNAMIC SIGNAL ANALYSER (DSA). The HP dynamic signal analyser, HP35670A, is a dynamic monitoring and measuring instrument that can be used for characterising the performance and stability of a control system. Performance parameters, such as rise time, overshoot, and settling time, are generally specified in the time domain. Stability criteria, gain/phase margins, are generally specified in the frequency domain. The HP35670A DSA is capable of measuring in both the time and frequency domains. The instrument can also be used for system identification.

6. VIBRATION-FREE TABLE. Since the success of the actual implementation depends largely on the accurate measurement of very small displacements of less than 1μm, there is a need to isolate the HDD implementation setup from the external vibrations. A Vibraplane Model 9100/9200 series vibration-free workstation was used. These are designed and constructed to provide very effective isolation of vibrations at frequencies above 5 Hz and low amplification at low frequencies of 2–3 Hz. Hence, the use of this vibration-free table shows significant improvements in resolution and repeatability of the measurement.

The overall hardware setup in our laboratory is depicted in Figure 1.4.

1.5 Preview of Each Chapter

A preview of each chapter is given next. Chapter 2 recalls some commonly used system identification and modelling techniques, such as the prediction

Figure 1.4. Implementation setup for HDD servo systems.

error identification and least squares estimation methods, applicable in the frequency domain, and the impulse response analysis and step response analysis in the time domain, which will be employed to identify the models of VCM actuators and microactuators later in the book. Chapter 3 recalls some linear system tools such as bilinear transformation, Jordan canonical forms and several structural decompositions of linear systems, which have the distinct feature of displaying the finite and infinite zero structures as well as the invertibility structures of a given system. They play a dominant role in the development of several control methods in Chapter 4, which will be utilised in the design of HDD servo systems. More precisely, Chapter 4 deals with linear control techniques, which include the well-known classical PID control, H_2 optimal control, H_∞ control and almost disturbance decoupling, robust and perfect tracking (RPT) control, and loop transfer recovery (LTR) technique. These methods are suitable for track following control and have been used extensively in designing HDD servo systems in the literature. Chapter 5 focuses on nonlinear control techniques such as the proximate time optimal servomechanism (PTOS), mode switching control (MSC) and composite nonlinear feedback (CNF) control. PTOS is generally used to design a control law in the track seeking stage of HDD servo systems, whereas the MSC and CNF design techniques can be used to find a controller that is applicable for both track seeking and track following.

Chapters 6 and 7 focus on the design of HDD servo systems with a single-stage VCM actuator. In particular, Chapter 6 deals with the modelling of the VCM actuator and design of track following controllers using both the

conventional PID control method and the recently developed RPT control method. The integrated single-stage actuator servo systems combining the track seeking and track following controllers will be given in Chapter 7. Three different approaches will be presented in this chapter and their results will be carefully compared.

Likewise, Chapters 8 and 9 deal with the modelling and design of HDD servo systems with a dual-stage actuator. In particular, Chapter 8 considers a robust controller design for a piezoelectric bimorph nonlinear actuator using an H_∞ almost disturbance decoupling approach. A complete HDD servo system with a dual-stage actuator will then be presented in Chapter 9. Conventional HDDs with a single-stage VCM actuator usually have resonances in the positioning arm and low-frequency bearing effects. It is believed that the performances of such HDDs have been pushed almost to their limits. Dual-stage servo systems with high bandwidth and high accuracy control are a possible solution to overcome the problems associated with conventional HDDs.

Finally, some issues on disturbance rejection, such as repeatable and non-repeatable runout rejection, and resonance compensation will be discussed in Chapter 10. Concluding remarks and further discussions on some key issues related to HDD servo systems will also be addressed. This will conclude the whole monograph.

1.6 Nomenclature

Throughout this monograph, we adopt the following abbreviations and notation, which are fairly standard.

$\mathbb{R}$	the set of real numbers
$\mathbb{C}$	the entire complex plane
$\mathbb{C}^{\odot}$	the set of complex numbers inside the unit circle
$\mathbb{C}^{\otimes}$	the set of complex numbers outside the unit circle
$\mathbb{C}^{\circ}$	the unit circle in the complex plane
$\mathbb{C}^{-}$	the open left-half complex plane
$\mathbb{C}^{+}$	the open right-half complex plane
$\mathbb{C}^{0}$	the imaginary axis in the complex plane
I	an identity matrix
I_k	an identity matrix of dimension $k \times k$
X'	the transpose of X
X^{H}	the complex conjugate transpose of X

$\mathrm{Im}\,(X)$	the range space of X
$\mathrm{Ker}\,(X)$	the null space of X
$X^{\dagger}$	the Moore–Penrose (pseudo) inverse of X
$\lambda(X)$	the set of eigenvalues of X
$\lambda_{\max}(X)$	the maximum eigenvalue of X
$\sigma_{\max}(X)$	the maximum singular value of X
$\|X\|$	the usual 2-norm of a matrix X
$\|\|G\|\|_2$	the H_2-norm of a stable system $G(s)$ or $G(z)$
$\|\|g\|\|_2$	the l_2-norm of a signal $g(t)$ or $g(k)$
L_2	the set of all functions whose l_2 norms are finite
$\|\|g\|\|_p$	the l_p-norm of a signal $g(t)$ or $g(k)$
L_p	the set of all functions whose l_p-norms are finite
$\|\|G\|\|_\infty$	the H_∞-norm of a stable system $G(s)$ or $G(z)$
$\dim\,(\mathcal{X})$	the dimension of a subspace $\mathcal{X}$
$\mathcal{X}^{\perp}$	the orthogonal complement of a subspace $\mathcal{X}$ of $\mathbb{R}^n$
ARE	algebraic Riccati equation
CNF	composite nonlinear feedback
DSP	digital signal processor
LDV	laser Doppler vibrometer
LQG	linear quadratic Gaussian
LQR	linear quadratic regulator
LTR	loop transfer recovery
MSC	mode switching control
N/RRO	non-/repeatable runouts
PES	position error signal
PID	proportional-integral-derivative
PTOS	proximate time optimal servomechanism
RPT	robust and perfect tracking
TMR	track mis-registration
TOC	time optimal control
TPI	track per inch (kTPI = kilo TPI)
ZOH	zero-order hold

Also, $C^{-1}\{\mathcal{X}\} := \{x \,|\, Cx \in \mathcal{X}\}$, where $\mathcal{X}$ is a subspace and C is a matrix. Finally, we append a $\diamondsuit$ at the end of a proof or a result statement.

CHAPTER 2
SYSTEM IDENTIFICATION TECHNIQUES

2.1 Introduction

The purpose of this chapter is to revisit some basic theories and solutions of system identification, which will be used later in the coming chapters to model various HDD systems. In general, the goal of system identification is to determine a mathematical model for a system or a process. Mathematical models may be developed either by use of "laws of nature", commonly known as *modelling* or based on experimentation, which is known as *system identification* [35]. In order to achieve a certain desirable performance for a given plant, it is necessary to derive a model for the plant that is adequate for controller design. The conventional design techniques in linear control systems require either parametric or nonparametric models. For example, design methods via root locus or robust control technique require a transfer function or a state space description of the plant to be controlled. The plant model is either described by the coefficients of certain polynomials or by the elements of state space matrices. In either case, we call these polynomial coefficients or matrix elements the *parameters* of the model. The category of such models is a *parametric description* of the plant model. On the other hand, design based on Nyquist, Bode and Nichols methods requires curves of amplitude and phase of transfer function from input to the output as functions of real frequency ω. If we have experimental data from a typical frequency response test, then we will be able to obtain certain functional curves for the plant. These curves are called nonparametric models of the plant, as there is no finite set of numbers that describes it exactly (see *e.g.*, [1]).

Thus, for a given plant, the problem of system identification is to determine a system model from the relationship (either in the time or the frequency domain) between its input and output. The problem can be represented graphically as shown in the Figure 2.1, in which $u(t)$ is the known input signal, $n(t)$ is the observation noise, and $y(t)$ is the measured output. A large variety of methods have been developed for solving such a problem (see *e.g.*, [36] and references cited therein). These methods include classical identification techniques (such as the impulse response analysis, step response analysis,

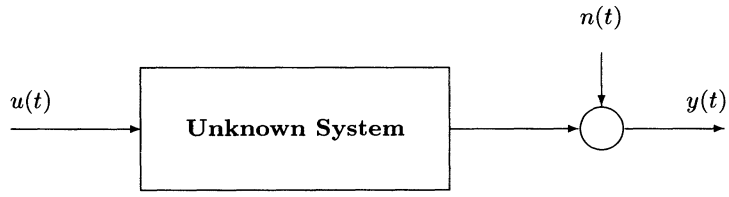


Figure 2.1. The unknown system to be identified.

frequency response identification) and equation error approaches and model adjustment techniques (such as the least squares estimation, maximum likelihood, and stochastic approximation, to name a few). The detailed derivations of these techniques can be found in a number of advanced texts devoted to system identification, *e.g.*, [35, 37–39].

2.2 Time Domain Methods

In this section, we will restrict our attention to identifying both parametric and nonparametric models through some commonly used time domain techniques. Interested readers are referred to [35, 36, 40] for detailed materials on the identification through impulse and step response characteristics.

2.2.1 Impulse Response Analysis

Parametric Models. Parametric models are described by parameters of differential equations or transfer functions. From these analytic representations, plots or values of interest of frequency response can in general be generated without much difficulty, whereas the reverse process of deriving parameters from nonparametric model descriptions is much more difficult.

A fairly general parametric model of a single-input and single-output (SISO) system can be described by the following differential equation (see *e.g.*, [40]),

$$\begin{aligned} a_n y^{(n)}(t) + \cdots + a_1 \dot{y}(t) + a_0 y(t) \\ = b_0 u(t - \tau_d) + b_1 \dot{u}(t - \tau_d) + \cdots + b_m u^{(m)}(t - \tau_d). \end{aligned} \tag{2.1}$$

Solving the differential equation for the input signal

$$u(t) = \delta(t), \tag{2.2}$$

with $\delta(t)$ being the unit impulse function, gives the impulse transfer function $h(t)$,

$$y(t) = h(t), \tag{2.3}$$

as the corresponding output function. Note that it is hard to generate an impulse input in the continuous-time domain, and hence this method is impractical.

Nonparametric Models. Consider again a SISO system as in Figure 2.1 with a scalar input signal $u(t)$ and a scalar output signal $y(t)$. Assume that the system is linear, time invariant and causal.

It is well known that a linear, time-invariant causal system can be described by its impulse response $g(\tau)$ as follows:

$$y(t) = \int_0^\infty g(\tau)u(t-\tau)\,\mathrm{d}\tau + n(t). \tag{2.4}$$

Knowing $\{g(\tau)\}_{\tau=0}^{\infty}$ and knowing $u(s)$ for $s \leq t$, we can consequently compute the corresponding output $y(s)$, $s \leq t$ for any input. The impulse response is thus a complete characterisation of the system.

The discrete equivalent of the output $y(t)$ can be written at the sampling instants $t_k = kT$, $k = 0, 1, 2, \cdots$, as

$$y(t_k) = \int_0^\infty g(\tau)u(t_k-\tau)\,\mathrm{d}\tau + n(t_k), \tag{2.5}$$

where T is the sampling period. Since, the input $u(t_k)$ is kept constant between the sampling instants:

$$u(t) = u(t_k) = u_k, \qquad kT \leq t < (k+1)T, \tag{2.6}$$

we can derive that

$$y(t_k) = \sum_{i=1}^{\infty} g(i)u(t_k - i) + n(t_k), \quad t_k = kT,\ k = 0, 1, 2, \cdots \tag{2.7}$$

Now, let $G(z)$ be the transfer function of the system from input to output with z being the usual forward shift operator, *i.e.*,

$$G(z) = \sum_{k=1}^{\infty} g(k) z^{-k}. \tag{2.8}$$

Then, Equation (2.7) can be written as

$$y(t_k) = G(z)u(t_k) + n(t_k). \tag{2.9}$$

If the system Equation (2.9) is subjected to a pulse input

$$u(t_k) = \begin{cases} \alpha, & t_k = 0, \\ 0, & t_k \neq 0, \end{cases} \tag{2.10}$$

then the output will be

$$y(t_k) = \alpha g(t_k) + n(t_k). \tag{2.11}$$

If the noise level is low, then the estimates of the coefficients of the impulse response $\{g(t_k)\}$ from an experiment will be

$$\hat{g}(t_k) = \frac{y(t_k)}{\alpha} \tag{2.12}$$

and the errors $n(t_k)/\alpha$. This simple analysis is called impulse response analysis. Unfortunately, many physical processes do not allow the error $n(t_k)/\alpha$ to be insignificant compared with the impulse response coefficients. Moreover, such an input could induce nonlinear effects that would disturb the linearised behaviour of the model. As such, identification methods depend on impulse inputs are rarely used in practical situations.

2.2.2 Step Response Analysis

Parametric Models. Parametric models usually are described by their frequency response function,

$$G(\omega) = \frac{b_0 + b_1 j\omega + ...b_m(j\omega)^m}{a_0 + a_1 j\omega + ...a_n(j\omega)^n} \cdot \exp(-j\omega\tau), \tag{2.13}$$

where a_i, b_i, m, n and τ are parameters to be identified.

Many researchers proposed methods for determining parameter values from time functions of the process output provided that the process input is a

well-determined signal. Of primary importance are step functions as input signals and step responses as output signals.

Almost all methods for the evaluation of step response functions use a small number of characteristic values of the response function. A first-order lag model with time delay and a frequency response

$$\hat{G}(\omega) = \frac{K}{1 + j\omega\tau_1} \cdot \exp(-j\omega\tau) \tag{2.14}$$

is widely used in step response analysis. In fact, there are quite a number of systems, especially in process control that can be approximated by a first-order model with an appropriate delay. The parameters K, τ_1 and τ can be derived from the step response shown in Figure 2.2 with u_0 being the amplitude of the input step signal (see *e.g.*, [40]). In short, if the system model is of the first order, one may need only obtain two pieces of information: (i) the steady-state response to the step input, and (ii) the time constant. The latter can be obtained either from the tangent with maximum slope of the step response or from the 10 to 90% rise time.

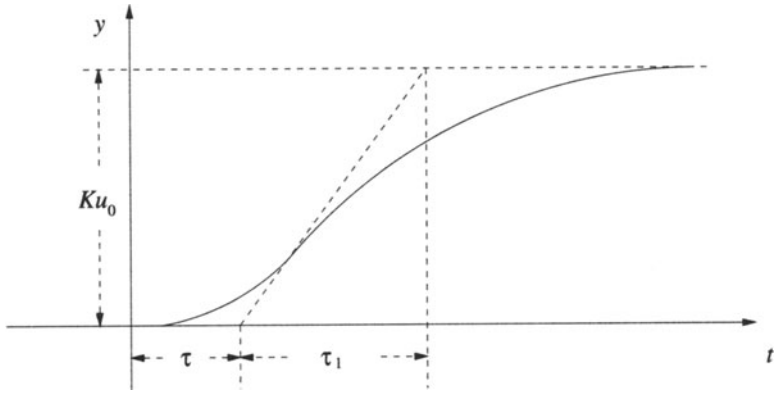


Figure 2.2. A typical step response.

For a second-order system model (with two poles and no zero), there are two possible situations: (1) when the two poles are real and (2) when the poles are a complex conjugate pair. Formulae for finding these from measurements of the (a) steady-state response, (b) maximum overshoot, (c) time required to reach the first-peak, and (d) time required to reach 50% of the steady-state value (for overdamped systems) can be easily derived. For the general case of higher-order practical systems, it is perhaps best to use a gradient method to find the parameters of the model of a given order such that the integral of the square of the error is minimised (see e.g., [35]).

Nonparametric Models. Since the impulse response of a system is the derivative of the step response, the identification problem in this case may be regarded as the determination of the transfer function from the impulse response. Alternatively, a step function

$$u(t_k) = \begin{cases} \alpha, & t_k \geq 0 \\ 0, & t_k < 0 \end{cases} \tag{2.15}$$

when applied to Equation (2.9) gives the output

$$y(t_k) = \alpha \sum_{i=1}^{k} g(i) + n(t_k), \quad t_k = kT, \; k = 0, 1, \cdots \tag{2.16}$$

Then, the estimates can be obtained as

$$\hat{g}(t_k) = \frac{y(t_k) - y(t_{k-1})}{\alpha}, \tag{2.17}$$

which has an error $[n(t_k) - n(t_{k-1})]/\alpha$. Hence, we would suffer again from large errors in most practical applications. But, if the goal is to determine some basic control-related characteristics, then the step responses from Equation (2.16) can very well furnish that information to a sufficient degree of accuracy. In fact, some well-known rules for tuning simple regulators, such as the Ziegler–Nichols rule, are based on model information reached in step responses. Based on plots of the step responses, some key characteristics of the system can be graphically constructed, which in turn can be used to determine system parameters.

2.3 Frequency Domain Methods

We recall in this section two identification methods in the frequency domain, namely, the predication error identification approach and the least squares estimation method, Both are particularly important to our studies in modelling the micro and VCM actuators in HDD servo systems in the coming chapters. The theories behind these techniques can be found in various references (see *e.g.*, [35, 41]).

2.3.1 Prediction Error Identification Approach

The prediction error approach is one of a black-box identification method. It includes the following three steps.

1. PARAMETER IDENTIFICATION. Suppose a system can be described as

$$y(k) = G(z^{-1})u(k) + H(z^{-1})n(k), \tag{2.18}$$

where $u(k)$ and $y(k)$ are its process input and output; $n(k)$ is noise input and supposed to be white; and

$$\left.\begin{aligned} G(z^{-1}) &= B(z^{-1})/A(z^{-1}) \\ H(z^{-1}) &= D(z^{-1})/A(z^{-1}) \\ A(z^{-1}) &= 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_{n_a} z^{-n_a} \\ B(z^{-1}) &= b_1 z^{-1} + b_2 z^{-2} + \cdots + b_{n_b} z^{-n_b} \\ D(z^{-1}) &= d_1 z^{-1} + d_2 z^{-2} + \cdots + d_{n_b} z^{-n_d} \end{aligned}\right\} \tag{2.19}$$

The predictor is:

$$\hat{y}(k|\theta) = [1 - H^{-1}(z^{-1}, \theta)]y(k) + H^{-1}(z^{-1}, \theta)G(z^{-1}, \theta)u(k), \tag{2.20}$$

where

$$\theta = \begin{pmatrix} a_1 \\ \vdots \\ a_{n_a} \\ b_1 \\ \vdots \\ b_{n_b} \\ d_1 \\ \vdots \\ d_{n_d} \end{pmatrix} \tag{2.21}$$

is the parameter vector of the system. Then, the prediction error given by a model is

$$e(k, \theta) = y(k) - \hat{y}(k|\theta). \tag{2.22}$$

Next, we define a loss function as

$$V_N(\theta, Z_N) = \frac{1}{N} \sum_{k=1}^{N} \ell(e(k, \theta)), \tag{2.23}$$

where $\ell(\cdot)$ is a scalar-valued positive function, and

$$Z_N = \left[y(1), \cdots, y(N), u(1), \cdots, u(N)\right] \tag{2.24}$$

is a set of input and output data from the experiment test. The desired system parameters can then be obtained by the minimisation of this loss function, *i.e.*,

$$\hat{\theta}_N = \arg\min V_N(\theta, Z_N). \tag{2.25}$$

2. DETERMINATION OF MODEL ORDER. The loss function $V_N(\theta, Z_N)$ can also be used to determine the order of a system. If the order of a model is lower than that of the system, then the value of the loss function will decrease significantly with the increase of the order of the model. However, when the order of the model is higher than that of the system, the increase of model order will not provide any more innovation for parameter identification, thus the value of $V_N(\theta, Z_N)$ will not decrease much. Therefore, the order of the system to be identified can be determined based on the decrease rate of $V_N(\theta, Z_N)$. Figure 2.3 shows a typical plot of the loss function versus identified model orders. It is clear from the plot that the order of the corresponding system to be identified is four.

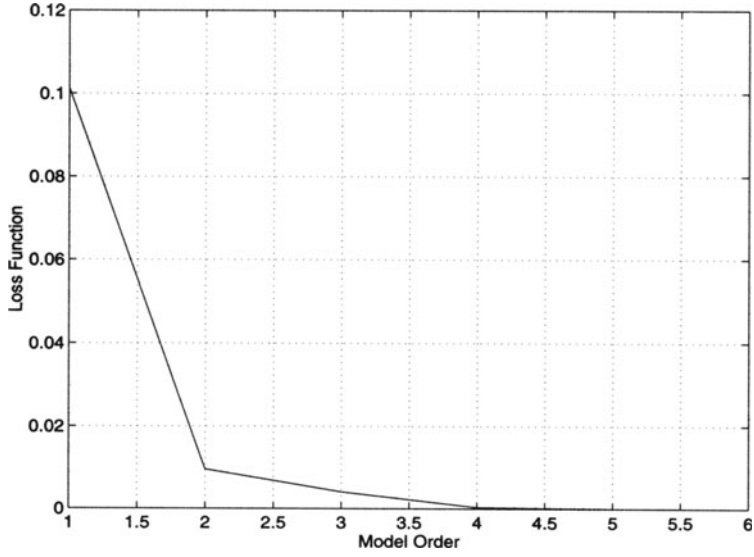


Figure 2.3. Values of loss function versus identified model orders.

3. MODEL VALIDATION. The third step of the prediction error identification method is to verify the correctness of the model obtained. It is clear that the residuals of the model can be obtained as

$$\begin{aligned} e(k,\theta) &= y(k) - \hat{y}(k|\theta) \\ &= \hat{H}^{-1}(z^{-1},\theta)\left[y(k) - \hat{G}(z^{-1},\theta)u(k)\right]. \end{aligned} \tag{2.26}$$

Obviously, if the model is correct, *i.e.*,

$$\hat{G}(z^{-1},\theta) = G_0(z^{-1}) \quad \text{and} \quad \hat{H}^{-1}(z^{-1},\theta) = H_0(z^{-1}), \tag{2.27}$$

the residual will tend to a white noise sequence $n(k)$. However, the non-whiteness of the residuals does not necessarily mean that the model is incorrect. In that case, the cross-correlation of the input u and residuals e can be used to verify the model. If u and e are independent, this means that all information in the residuals is explained by the process model $\hat{G}$, and we can conclude that the estimate is correct. Otherwise the result is incorrect.

The cross-correlation of u and e is

$$R_{eu}(\tau) = \boldsymbol{E}\{e(t+\tau)u(t)\} \tag{2.28}$$

where $\boldsymbol{E}\{\cdot\}$ is the expected value. If the residuals and input are independent, we have

$$\sqrt{N}R_{eu} \to \mathcal{N}(0,P), \quad \text{as } N \to \infty, \tag{2.29}$$

where $P = \sum_{k=-\infty}^{\infty} R_e(k)R_u(k)$, and $\mathcal{N}(0,P)$ denotes the normal random distribution with zero mean and a variance P. Let N_α be the α-level of the $\mathcal{N}(0,P)$ such that

$$\boldsymbol{P}\left\{|R_{eu}| \le \sqrt{\frac{P}{N}}N_\alpha\right\} = \alpha, \tag{2.30}$$

where $\boldsymbol{P}\{\cdot\}$ is the probability. Define the following null hypothesis:

$$H_0 : |R_{eu}| \le \sqrt{\frac{P}{N}}N_\alpha. \tag{2.31}$$

If H_0 is accepted, then we can say that the model is acceptable with a probability of $1-\alpha$. Figure 2.4 shows a typical plot of the values of cross-correlation function between the input and the error residual. It can be seen that, for such a model, all the data are within the 95% confidence region. Hence, we can say that the corresponding identified model is acceptable with a probability of 95%.

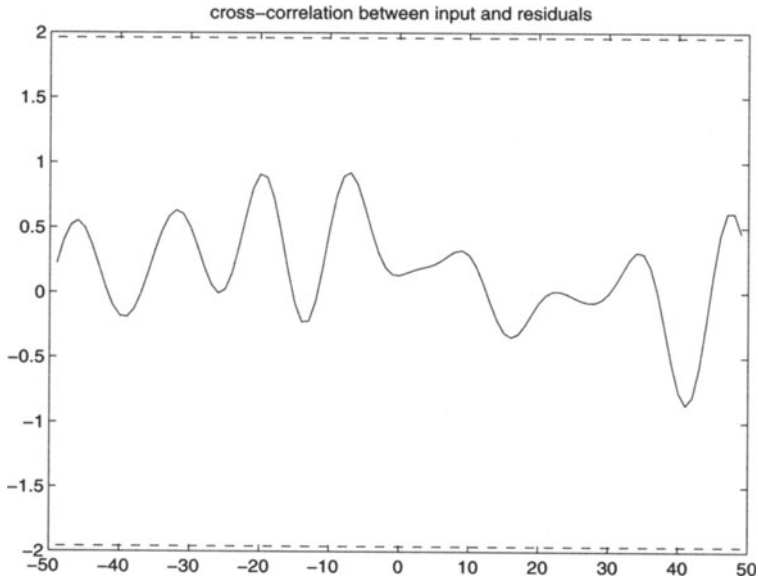


Figure 2.4. Model validation test.

2.3.2 Least Squares Estimation Method

We will utilise the frequency response identification method (see *e.g.*, [37]) to model our actuator. Such a method is applicable to minimum phase processes. We expect from the properties of the physical system that the VCM actuator should be of minimum phase. The detailed procedure proceeds as follows: we first assume that the transfer function of a minimum phase plant is given by

$$G(s) = \frac{N(s)}{D(s)} = \frac{b_0 + b_1 s + b_2 s^2 + \cdots + b_m s^m}{1 + a_1 s + a_2 s^2 + \cdots + a_n s^n}, \tag{2.32}$$

for some appropriate coefficients a_k, $k = 1, 2, \cdots, n$, and b_k, $k = 0, 1, \cdots, m$, with $n \geq m$. These parameters are to be identified. Then, its corresponding frequency response is given by

$$G(j\omega) = \frac{\alpha(\omega) + j\omega\beta(\omega)}{\sigma(\omega) + j\omega\tau(\omega)} = \frac{N(j\omega)}{D(j\omega)}, \tag{2.33}$$

where

$$\left.\begin{aligned}
\alpha(\omega_i) &= b_0 - b_2\omega_i^2 + b_4\omega_i^4 - \cdots \\
\beta(\omega_i) &= b_1 - b_3\omega_i^2 + b_5\omega_i^4 - \cdots \\
\sigma(\omega_i) &= 1 - a_2\omega_i^2 + a_4\omega_i^4 - \cdots \\
\tau(\omega_i) &= a_1 - a_3\omega_i^2 + a_5\omega_i^4 - \cdots
\end{aligned}\right\} \tag{2.34}$$

Let $R(\omega)$ and $I(\omega)$ be the real and imaginary parts of the measured frequency response of the actuator system. The frequency response error between the model and the actual measurement data is given by

$$\mathcal{E}(j\omega) = [R(\omega) + jI(\omega)] - \frac{N(j\omega)}{D(j\omega)}. \tag{2.35}$$

Thus, the parameters of the system can be obtained by minimising the following index:

$$J = \sum_{i=1}^{L} |\mathcal{E}(j\omega_i)|^2, \tag{2.36}$$

where L is the total number of points of the measured data. Unfortunately, this is a nonlinear optimisation problem, and it is difficult to solve. We then follow the results of [37] to modify the error norm as

$$J = \sum_{i=1}^{L} |D(j\omega_i)\mathcal{E}(j\omega_i)|^2. \tag{2.37}$$

The original problem now becomes a linear optimisation problem. Using Equations (2.33) and (2.35), we can rewrite Equation (2.37) as follows

$$J = \sum_{i=1}^{L} \Big\{ [X(\omega_i)]^2 + [Y(\omega_i)]^2 \Big\}, \tag{2.38}$$

where

$$X(\omega_i) = \sigma(\omega_i)R(\omega_i) - \omega_i\tau(\omega_i)I(\omega_i) - \alpha(\omega_i), \tag{2.39}$$

and

$$Y(\omega_i) = \omega_i\tau(\omega_i)R(\omega_i) + \sigma(\omega_i)I(\omega_i) - \omega_i\beta(\omega_i). \tag{2.40}$$

Therefore, J can be minimised by finding $\hat{b}_0$, $\hat{b}_1$, $\cdots$, $\hat{b}_m$ and $\hat{a}_1$, $\hat{a}_2$, $\cdots$, $\hat{a}_n$ such that

$$\left.\begin{aligned}
&\frac{\partial J}{\partial b_0}\Big|_{b_0=\hat{b}_0} = \sum_{i=1}^{L}\Big\{2X(\omega_i)(-1)\Big\}\Big|_{b_0=\hat{b}_0} = 0\\
&\frac{\partial J}{\partial b_1}\Big|_{b_1=\hat{b}_1} = \sum_{i=1}^{L}\Big\{2Y(\omega_i)(-\omega_i)\Big\}\Big|_{b_1=\hat{b}_1} = 0\\
&\qquad\vdots\\
&\frac{\partial J}{\partial a_1}\Big|_{a_1=\hat{a}_1} = \sum_{i=1}^{L} 2\omega_i\Big[Y(\omega_i)R(\omega_i) - X(\omega_i)I(\omega_i)\Big]\Big|_{a_1=\hat{a}_1} = 0\\
&\frac{\partial J}{\partial a_2}\Big|_{a_2=\hat{a}_2} = \sum_{i=1}^{L} -2\omega_i^2\Big[X(\omega_i)R(\omega_i) + Y(\omega_i)I(\omega_i)\Big]\Big|_{a_2=\hat{a}_2} = 0\\
&\qquad\vdots
\end{aligned}\right\} \tag{2.41}$$

Rearranging the above equations, we obtain

$$\begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{a} \end{pmatrix} = \begin{pmatrix} \boldsymbol{B}_1 \\ \boldsymbol{B}_2 \end{pmatrix}, \tag{2.42}$$

where

$$\boldsymbol{A}_{11} = \begin{bmatrix} V_0 & 0 & -V_2 & 0 & V_4 & \cdots \\ 0 & V_2 & 0 & -V_4 & 0 & \cdots \\ V_2 & 0 & -V_4 & 0 & V_6 & \cdots \\ 0 & V_4 & 0 & -V_6 & 0 & \cdots \\ V_4 & 0 & -V_6 & 0 & V_8 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \\ \hat{b}_4 \\ \vdots \end{bmatrix}, \tag{2.43}$$

$$\boldsymbol{A}_{12} = \begin{bmatrix} T_1 & S_2 & -T_3 & -S_4 & T_5 & \cdots \\ -S_2 & T_3 & S_4 & -T_5 & -S_6 & \cdots \\ T_3 & S_4 & -T_5 & -S_6 & T_7 & \cdots \\ -S_4 & T_5 & S_6 & -T_7 & -S_8 & \cdots \\ T_5 & S_6 & -T_7 & -S_8 & T_9 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \quad \boldsymbol{a} = \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \vdots \end{bmatrix}, \tag{2.44}$$

$$\boldsymbol{A}_{21} = \begin{bmatrix} T_1 & -S_2 & -T_3 & S_4 & T_5 & \cdots \\ S_2 & T_3 & -S_4 & -T_5 & S_6 & \cdots \\ T_3 & -S_4 & -T_5 & S_6 & T_7 & \cdots \\ S_4 & T_5 & -S_6 & -T_7 & S_8 & \cdots \\ T_5 & -S_6 & -T_7 & S_8 & T_9 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \quad \boldsymbol{B}_1 = \begin{bmatrix} S_0 \\ T_1 \\ S_2 \\ T_3 \\ S_4 \\ \vdots \end{bmatrix}, \tag{2.45}$$

$$\boldsymbol{A}_{22} = \begin{bmatrix} U_2 & 0 & -U_4 & 0 & U_6 & \cdots \\ 0 & U_4 & 0 & -U_6 & 0 & \cdots \\ U_4 & 0 & -U_6 & 0 & U_8 & \cdots \\ 0 & U_6 & 0 & -U_8 & 0 & \cdots \\ U_6 & 0 & -U_8 & 0 & U_{10} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \quad \boldsymbol{B}_2 = \begin{bmatrix} 0 \\ U_2 \\ 0 \\ U_4 \\ 0 \\ \vdots \end{bmatrix}, \tag{2.46}$$

and where

$$V_k = \sum_{i=0}^{L} \omega_i^k, \qquad S_k = \sum_{i=0}^{L} \omega_i^k R(\omega_i), \tag{2.47}$$

$$T_k = \sum_{i=0}^{L} \omega_i^k I(\omega_i), \qquad U_k = \sum_{i=0}^{L} \omega_i^k [R^2(\omega_i) + I^2(\omega_i)]. \tag{2.48}$$

The desired parameters of the corresponding transfer function model can be obtained by solving the above equations.

Finally, we note that the methods recalled above are merely for the identification of the HDD VCM and micro-actuators in the coming chapters. If systems to be identified are highly uncertain with disturbances, it would be more appropriate to use the methods reported in a recent monograph by Chen and Gu [42] to yield more accurate results.

CHAPTER 3
LINEAR SYSTEM TOOLS

3.1 Introduction

It is our belief that a good unambiguous understanding of linear system structures, *i.e.*, the finite and infinite zero structures as well as the invertibility structures of linear systems, is essential for a meaningful control system design. As a matter of fact, the performance and limitation of an overall control system is primarily dependent on the structural properties of the given open-loop system. In our opinion, a control system engineer should thoroughly study the properties of a given plant before carrying out any meaningful design. Many of the difficulties one might face in the design stage may be avoided if the designer has fully understood the system properties or limitations. For example, it is well understood in the literature that a nonminimum phase zero would generally yield a bad overall performance no matter what design methodology is used. A good control engineer should try to avoid these kinds of problem at the initial stage by adding or adjusting sensors or actuators in the system. Sometimes, a simple rearrangement of existing sensors and/or actuators could totally change the system properties. We refer interested readers to the recent work by Liu *et al.* [43] for details.

As such, we recall in this chapter several system decomposition techniques that can be utilised to display all the above-mentioned structural properties. More specifically, we will recall: (1) the Jordan and real Jordan canonical forms for a square constant matrix; (2) the controllability structural decomposition (CSD) and block diagonal control canonical form for a constant matrix pair; and (3) the special coordinate basis of a linear time invariant system characterised by either a matrix triple or a matrix quadruple. These canonical forms and the special coordinate basis will form a transformer for linear systems. Once a linear system is touched by this transformer, all its structural properties become clear and transparent. Lastly, we will recall at the end of the chapter some key results of bilinear transformation and inverse bilinear transformation. Mappings of the structural properties of general linear systems under bilinear and inverse bilinear transformations will also be identified. These results serve as a bridge between the continuous-

and discrete-time systems. In fact, many results for discrete-time systems, such as discrete-time H_∞ optimisation and H_∞ almost disturbance decoupling problems, can be derived from their continuous-time counterparts by the bilinear transformation.

3.2 Jordan Canonical Forms

We recall in this section the Jordan canonical form and the real Jordan canonical form of a square constant matrix. We first have the following theorem.

3.2.1 Jordan Canonical Form

Theorem 3.2.1. Consider a constant matrix $A \in \mathbb{R}^{n\times n}$. There exists a nonsingular transformation $T \in \mathbb{C}^{n\times n}$ and an integer k such that

$$T^{-1}AT = \text{blkdiag}\Big\{J_1, J_2, \cdots, J_k\Big\}, \tag{3.1}$$

where J_i, $i = 1, 2, \cdots, k$, are some $n_i \times n_i$ Jordan blocks, *i.e.*,

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_i & 1 \\ & & & \lambda_i \end{bmatrix}. \tag{3.2}$$

Obviously, $\lambda_i \in \lambda(A)$, $i = 1, 2, \cdots, k$, and $\sum_{i=1}^{k} n_i = n$. ◇

The result of Theorem 3.2.1 is very well known. The realisation of this Jordan canonical form in MATLAB can be found in Lin and Chen [34] (see *e.g.*, the m-function `jordan.m`). The following theorem is to find a real Jordan canonical form.

3.2.2 Real Jordan Canonical Form

Theorem 3.2.2. Consider a constant matrix $A \in \mathbb{R}^{n\times n}$. There exists a nonsingular transformation $P \in \mathbb{R}^{n\times n}$ and an integer k such that

$$P^{-1}AP = \text{blkdiag}\Big\{J_1, J_2, \cdots, J_k\Big\}, \tag{3.3}$$

where each block J_i, $i = 1, 2, \cdots, k$, has the following form: if $\lambda_i \in \lambda(A)$ is real,

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_i & 1 \\ & & & \lambda_i \end{bmatrix}, \tag{3.4}$$

or if $\lambda_i = \mu_i + j\omega_i \in \lambda(A)$ and $\bar{\lambda}_i = \mu_i - j\omega_i \in \lambda(A)$ with $\omega_i \neq 0$,

$$J_i = \begin{bmatrix} \Lambda_i & I_2 & & \\ & \ddots & \ddots & \\ & & \Lambda_i & I_2 \\ & & & \Lambda_i \end{bmatrix}, \quad \Lambda_i = \begin{bmatrix} \mu_i & \omega_i \\ -\omega_i & \mu_i \end{bmatrix}. \tag{3.5}$$

The above structure of $P^{-1}AP$ is called *the real Jordan canonical form.* $\diamondsuit$

Again, the above result can be found in many standard textbooks. An m-function, `r_jordan.m`, realising the above real Jordan canonical form has been reported in Lin and Chen [34].

3.3 Structural Decompositions of Matrix Pairs

In this section, two structural decompositions of a constant matrix pair, namely the so-called controllability structural decomposition (CSD) and the block diagonal control canonical form (see *e.g.*, [44]), will be presented. We will first recall the CSD for a linear system characterised by a matrix pair (A, B), which was called a Brunovsky canonical form by many researchers in the literature (see *e.g.*, [45]). However, it is noted that such a decomposition was actually first discovered by Luenberger [46] in 1967, which was 3 years earlier than the publication of Brunovsky's results [47] in 1970. We have the following theorems regarding the CSD and the block diagonal control canonical form for a given constant matrix pair.

3.3.1 Controllability Structural Decomposition

Theorem 3.3.1. Consider a constant matrix pair (A, B) with $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$ with B being of full rank. There exist nonsingular state and input

transformations T_s and T_i such that $(\tilde{A}, \tilde{B}) := (T_s^{-1}AT_s, T_s^{-1}BT_i)$ has the following form:

$$\left(\begin{bmatrix} A_o & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & I_{k_1-1} & \cdots & 0 & 0 \\ \star & \star & \star & \cdots & \star & \star \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & I_{k_m-1} \\ \star & \star & \star & \cdots & \star & \star \end{bmatrix}, \begin{bmatrix} 0 & \cdots & 0 \\ 0 & \cdots & 0 \\ 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \\ 0 & \cdots & 1 \end{bmatrix} \right), \tag{3.6}$$

where $k_i > 0$, $i = 1, \cdots, m$, A_o is of dimension $n_o := n - \sum_{i=1}^{m} k_i$ and its eigenvalues are the uncontrollable modes of (A, B). Moreover, the set of integers, $\mathcal{C} := \{ n_o, k_1, \cdots, k_m \}$, is hereafter called the *controllability index* of (A, B). ◇

3.3.2 Block Diagonal Control Canonical Form

Theorem 3.3.2. Consider a constant matrix pair (A, B) with $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times m}$ and with (A, B) being completely controllable. Then there exist an integer $k \leq m$, a set of κ integers $k_1, k_2, \cdots, k_\kappa$, and nonsingular transformations T_s and T_i such that

$$T_s^{-1}AT_s = \begin{bmatrix} A_1 & 0 & 0 & \cdots & 0 \\ 0 & A_2 & 0 & \cdots & 0 \\ 0 & 0 & A_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & A_\kappa \end{bmatrix}, \tag{3.7}$$

$$T_s^{-1}BT_i = \begin{bmatrix} B_1 & \star & \star & \cdots & \star & \star \\ 0 & B_2 & \star & \cdots & \star & \star \\ 0 & 0 & B_3 & \cdots & \star & \star \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & B_\kappa & \star \end{bmatrix}, \tag{3.8}$$

where the $\star$ symbols represent some matrices of less interest, and matrices A_i and B_i, $i = 1, 2, \cdots, \kappa$, have the following control canonical form:

$$A_i = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_{k_i}^i & -a_{k_i-1}^i & -a_{k_i-2}^i & \cdots & -a_1^i \end{bmatrix}, \quad B_i = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \tag{3.9}$$

for some scalars a_1^i, a_2^i, $\cdots$, $a_{k_i}^i$. Obviously, $\sum_{i=1}^{\kappa} k_i = n$. We call the above structure of A and B a *block diagonal control canonical form.* $\diamondsuit$

The software realisations of these canonical forms in MATLAB can be found in Lin and Chen [34].

3.4 Structural Decomposition of Linear Systems

Consider a general proper linear time-invariant system Σ_*, which could be of either continuous-time or discrete-time, characterised by a matrix quadruple (A_*, B_*, C_*, D_*) or in the state-space form

$$\Sigma_* : \begin{cases} \delta(x) = A_* x + B_* u, \\ y = C_* x + D_* u, \end{cases} \tag{3.10}$$

where $\delta(x) = \dot{x}(t)$ if Σ_* is a continuous-time system, or $\delta(x) = x(k+1)$ if Σ_* is a discrete-time system. Similarly, $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$ and $y \in \mathbb{R}^p$ are the state, input and output of Σ_*. They represent respectively $x(t)$, $u(t)$ and $y(t)$ if the given system is of continuous-time, or represent respectively $x(k)$, $u(k)$ and $y(k)$ if Σ_* is of discrete-time. Without loss of any generality, we assume throughout this section that both $[B_*' \;\; D_*']$ and $[C_* \;\; D_*]$ are of full rank. The transfer function of Σ_* is then given by

$$H_*(\varsigma) = C_*(\varsigma I - A_*)^{-1} B_* + D_*, \tag{3.11}$$

where $\varsigma = s$, the Laplace transform operator, if Σ_* is of continuous-time, or $\varsigma = z$, the z-transform operator, if Σ_* is of discrete-time. It is simple to verify that there exist nonsingular transformations U and V such that

$$UD_*V = \begin{bmatrix} I_{m_0} & 0 \\ 0 & 0 \end{bmatrix}, \tag{3.12}$$

where m_0 is the rank of matrix D_*. In fact, U can be chosen as an orthogonal matrix. Hence, hereafter, without loss of generality, it is assumed that the matrix D_* has the form given on the right hand-side of Equation (3.12). One can now rewrite system Σ_* of Equation (3.10) as

$$\begin{cases} \delta(x) = A_* x + [B_{*,0} \;\; B_{*,1}] \begin{pmatrix} u_0 \\ u_1 \end{pmatrix}, \\ \begin{pmatrix} y_0 \\ y_1 \end{pmatrix} = \begin{bmatrix} C_{*,0} \\ C_{*,1} \end{bmatrix} x + \begin{bmatrix} I_{m_0} & 0 \\ 0 & 0 \end{bmatrix} \begin{pmatrix} u_0 \\ u_1 \end{pmatrix}, \end{cases} \tag{3.13}$$

where the matrices $B_{*,0}$, $B_{*,1}$, $C_{*,0}$ and $C_{*,1}$ have appropriate dimensions.

Theorem 3.4.1 on the special coordinate basis (SCB) of linear systems (see also [48, 49]) is mainly due to the results of Sannuti and Saberi. The proofs of all its properties can be found in a recent monograph by Chen [44].

Theorem 3.4.1. Given the linear system Σ_* of Equation (3.10), there exist

1. coordinate-free nonnegative integers n_a^-, n_a^0, n_a^+, n_b, n_c, n_d, $m_d \leq m - m_0$ and q_i, $i = 1, \cdots, m_d$, and
2. nonsingular state, output and input transformations Γ_s, Γ_o and Γ_i that take the given Σ_* into a special coordinate basis that displays explicitly both the finite and infinite zero structures of Σ_*.

The special coordinate basis is described by the following set of equations:

$$x = \Gamma_s \tilde{x}, \quad y = \Gamma_o \tilde{y}, \quad u = \Gamma_i \tilde{u}, \tag{3.14}$$

$$\tilde{x} = \begin{pmatrix} x_a \\ x_b \\ x_c \\ x_d \end{pmatrix}, \quad x_a = \begin{pmatrix} x_a^- \\ x_a^0 \\ x_a^+ \end{pmatrix}, \quad x_d = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{m_d} \end{pmatrix}, \tag{3.15}$$

$$\tilde{y} = \begin{pmatrix} y_0 \\ y_d \\ y_b \end{pmatrix}, \; y_d = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{m_d} \end{pmatrix}, \; \tilde{u} = \begin{pmatrix} u_0 \\ u_d \\ u_c \end{pmatrix}, \; u_d = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_{m_d} \end{pmatrix}, \tag{3.16}$$

$$\delta(x_a^-) = A_{aa}^- x_a^- + B_{0a}^- y_0 + L_{ad}^- y_d + L_{ab}^- y_b, \tag{3.17}$$

$$\delta(x_a^0) = A_{aa}^0 x_a^0 + B_{0a}^0 y_0 + L_{ad}^0 y_d + L_{ab}^0 y_b, \tag{3.18}$$

$$\delta(x_a^+) = A_{aa}^+ x_a^+ + B_{0a}^+ y_0 + L_{ad}^+ y_d + L_{ab}^+ y_b, \tag{3.19}$$

$$\delta(x_b) = A_{bb} x_b + B_{0b} y_0 + L_{bd} y_d, \quad y_b = C_b x_b, \tag{3.20}$$

$$\begin{aligned} \delta(x_c) = {} & A_{cc} x_c + B_{0c} y_0 + L_{cb} y_b + L_{cd} y_d \\ & + B_c \left[E_{ca}^- x_a^- + E_{ca}^0 x_a^0 + E_{ca}^+ x_a^+ \right] + B_c u_c, \end{aligned} \tag{3.21}$$

$$y_0 = C_{0c} x_c + C_{0a}^- x_a^- + C_{0a}^+ x_a^0 + C_{0a}^+ x_a^+ + C_{0d} x_d + C_{0b} x_b + u_0 \tag{3.22}$$

and for each $i = 1, \cdots, m_d$,

$$\delta(x_i) = A_{q_i} x_i + L_{i0} y_0 + L_{id} y_d$$

$$+ B_{q_i}\left[u_i + E_{ia}x_a + E_{ib}x_b + E_{ic}x_c + \sum_{j=1}^{m_d} E_{ij}x_j\right], \quad (3.23)$$

$$y_i = C_{q_i}x_i, \quad y_d = C_d x_d. \quad (3.24)$$

Here the states x_a^-, x_a^0, x_a^+, x_b, x_c and x_d are respectively of dimensions n_a^-, n_a^0, n_a^+, n_b, n_c and $n_d = \sum_{i=1}^{m_d} q_i$, and x_i is of dimension q_i for each $i = 1, \cdots, m_d$. The control vectors u_0, u_d and u_c are respectively of dimensions m_0, m_d and $m_c = m - m_0 - m_d$, and the output vectors y_0, y_d and y_b are respectively of dimensions $p_0 = m_0$, $p_d = m_d$ and $p_b = p - p_0 - p_d$. The matrices A_{q_i}, B_{q_i} and C_{q_i} have the following form:

$$A_{q_i} = \begin{bmatrix} 0 & I_{q_i-1} \\ 0 & 0 \end{bmatrix}, \quad B_{q_i} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad C_{q_i} = [1,\ 0,\ \cdots,\ 0]. \quad (3.25)$$

Assuming that x_i, $i = 1, 2, \cdots, m_d$, are arranged such that $q_i \leq q_{i+1}$, the matrix L_{id} has the particular form

$$L_{id} = [L_{i1} \quad L_{i2} \quad \cdots \quad L_{ii-1} \quad 0 \quad \cdots \quad 0]. \quad (3.26)$$

The last row of each L_{id} is identically zero. Moreover:

1. If Σ_* is a continuous-time system, then

$$\lambda(A_{aa}^-) \subset \mathbb{C}^-, \quad \lambda(A_{aa}^0) \subset \mathbb{C}^0, \quad \lambda(A_{aa}^+) \subset \mathbb{C}^+. \quad (3.27)$$

2. If Σ_* is a discrete-time system, then

$$\lambda(A_{aa}^-) \subset \mathbb{C}^\odot, \quad \lambda(A_{aa}^0) \subset \mathbb{C}^\circ, \quad \lambda(A_{aa}^+) \subset \mathbb{C}^\otimes. \quad (3.28)$$

Also, the pair (A_{cc}, B_c) is controllable and the pair (A_{bb}, C_b) is observable. $\Diamond$

Note that a procedure of constructing the original version of the above structural decomposition of a strictly proper linear system was given by Sannuti and Saberi [48]. The required modifications for non-strictly proper systems were given by the same authors in [49]. Here, in Theorem 3.4.1 by another change of basis, the variable x_a is further decomposed into x_a^-, x_a^0 and x_a^+. For continuous-time systems, one can use the real Schur algorithm to obtain such a decomposition. For discrete-time systems, the algorithm of Chen [50] can be used. The realisation of this unified decomposition can be found in Lin and Chen [34]. Finally, the proofs of all properties of this decomposition listed below have been reported by Chen [44, 51].

We can rewrite the special coordinate basis of the quadruple (A_*, B_*, C_*, D_*) given by Theorem 3.4.1 in a more compact form:

$$\begin{aligned}\tilde{A}_* &= \Gamma_{\rm s}^{-1}(A_* - B_{*,0}C_{*,0})\Gamma_{\rm s} \\ &= \begin{bmatrix} A_{aa}^- & 0 & 0 & L_{ab}^- C_b & 0 & L_{ad}^- C_d \\ 0 & A_{aa}^0 & 0 & L_{ab}^0 C_b & 0 & L_{ad}^0 C_d \\ 0 & 0 & A_{aa}^+ & L_{ab}^+ C_b & 0 & L_{ad}^+ C_d \\ 0 & 0 & 0 & A_{bb} & 0 & L_{bd} C_d \\ B_c E_{ca}^- & B_c E_{ca}^0 & B_c E_{ca}^+ & L_{cb} C_b & A_{cc} & L_{cd} C_d \\ B_d E_{da}^- & B_d E_{da}^0 & B_d E_{da}^+ & B_d E_{db} & B_d E_{dc} & A_{dd} \end{bmatrix},\end{aligned} \tag{3.29}$$

$$\tilde{B}_* = \Gamma_{\rm s}^{-1}\begin{bmatrix} B_{*,0} & B_{*,1} \end{bmatrix}\Gamma_{\rm i} = \begin{bmatrix} B_{0a}^- & 0 & 0 \\ B_{0a}^0 & 0 & 0 \\ B_{0a}^+ & 0 & 0 \\ B_{0b} & 0 & 0 \\ B_{0c} & 0 & B_c \\ B_{0d} & B_d & 0 \end{bmatrix}, \tag{3.30}$$

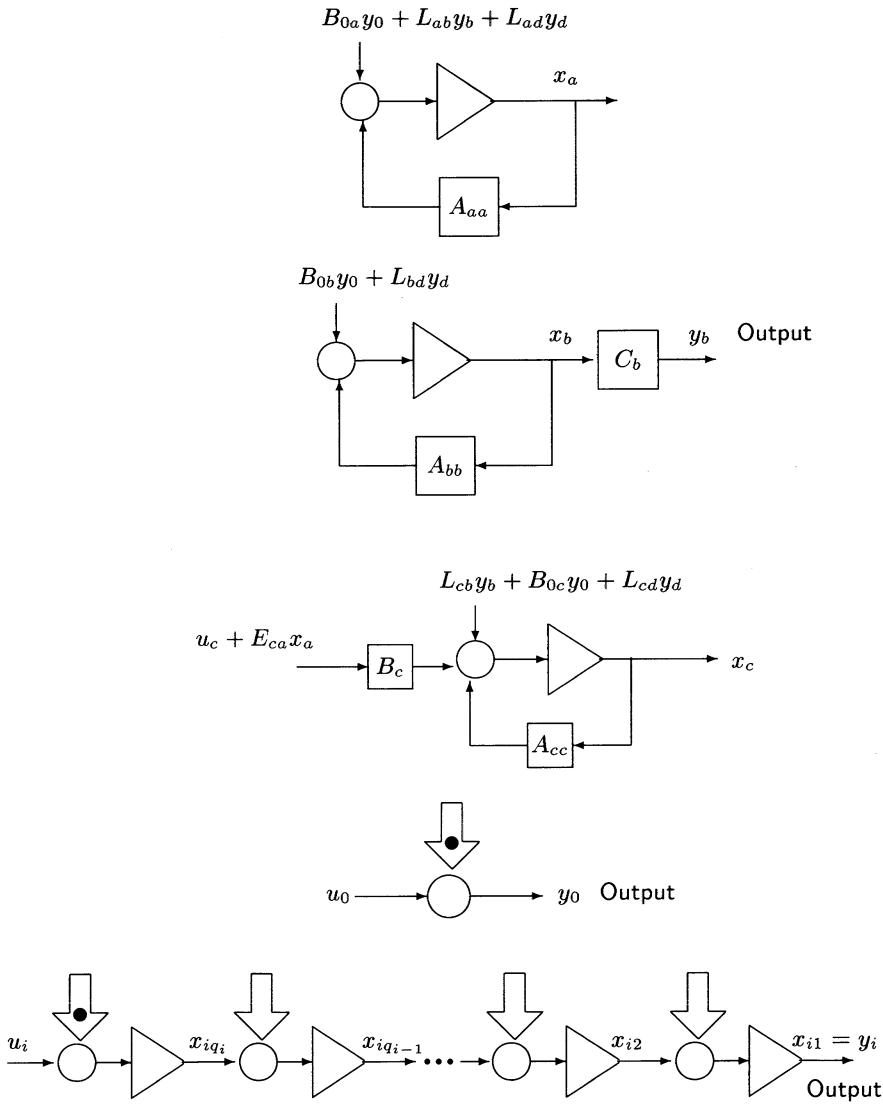
$$\tilde{C}_* = \Gamma_{\rm o}^{-1}\begin{bmatrix} C_{*,0} \\ C_{*,1} \end{bmatrix}\Gamma_{\rm s} = \begin{bmatrix} C_{0a}^- & C_{0a}^0 & C_{0a}^+ & C_{0b} & C_{0c} & C_{0d} \\ 0 & 0 & 0 & 0 & 0 & C_d \\ 0 & 0 & 0 & C_b & 0 & 0 \end{bmatrix}, \tag{3.31}$$

$$\tilde{D}_* = \Gamma_{\rm o}^{-1} D_* \Gamma_{\rm i} = \begin{bmatrix} I_{m_0} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{3.32}$$

A block diagram of the special coordinate basis of Theorem 3.4.1 is given in Figure 3.1. In this figure, a signal given by a double-edged arrow is some linear combination of outputs y_i, $i = 0$ to m_d, whereas a signal given by the double-edged arrow with a solid dot is some linear combination of all the states. Also, the block $\triangleright$ is either an integrator if Σ_* is of continuous-time or a backward shifting operator if Σ_* is of discrete-time.

We note the following intuitive points regarding the special coordinate basis.

1. The variable u_i controls the output y_i through a stack of q_i integrators (or backward shifting operators), whereas x_i is the state associated with those integrators (or backward shifting operators) between u_i and y_i. Moreover, (A_{q_i}, B_{q_i}) and (A_{q_i}, C_{q_i}) respectively form controllable and observable pairs. This implies that all the states x_i are both controllable and observable.



Note that a signal given by a double-edged arrow with a solid dot is some linear combination of all the states, whereas a signal given by a simple double-edged arrow is some linear combination of only output y_d. Also, matrices B_{0a}, L_{ab}, L_{ad} and E_{ca} are to be defined in Property 3.4.1.

Figure 3.1. A block diagram representation of the special coordinate basis.

2. The output y_b and the state x_b are not directly influenced by any inputs; however, they could be indirectly controlled through the output y_d. Moreover, (A_{bb}, C_b) forms an observable pair. This implies that the state x_b is observable.

3. The state x_c is directly controlled by the input u_c, but it does not directly affect any output. Moreover, (A_{cc}, B_c) forms a controllable pair. This implies that the state x_c is controllable.

4. The state x_a is neither directly controlled by any input nor does it directly affect any output.

In what follows, we state some important properties of the above special coordinate basis that are pertinent to our present work. As mentioned earlier, the proofs of these properties can be found in Chen [44].

Property 3.4.1. The given system Σ_* is observable (detectable) if and only if the pair $(A_{\text{obs}}, C_{\text{obs}})$ is observable (detectable), where

$$A_{\text{obs}} := \begin{bmatrix} A_{aa} & 0 \\ B_c E_{ca} & A_{cc} \end{bmatrix}, \quad C_{\text{obs}} := \begin{bmatrix} C_{0a} & C_{0c} \\ E_{da} & E_{dc} \end{bmatrix}, \tag{3.33}$$

and where

$$A_{aa} := \begin{bmatrix} A_{aa}^- & 0 & 0 \\ 0 & A_{aa}^0 & 0 \\ 0 & 0 & A_{aa}^+ \end{bmatrix}, \quad C_{0a} := [\, C_{0a}^- \quad C_{0a}^0 \quad C_{0a}^+ \,], \tag{3.34}$$

$$E_{da} := [\, E_{da}^- \quad E_{da}^0 \quad E_{da}^+ \,], \quad E_{ca} := [\, E_{ca}^- \quad E_{ca}^0 \quad E_{ca}^+ \,]. \tag{3.35}$$

Also, define

$$A_{\text{con}} := \begin{bmatrix} A_{aa} & L_{ab} C_b \\ 0 & A_{bb} \end{bmatrix}, \quad B_{\text{con}} := \begin{bmatrix} B_{0a} & L_{ad} \\ B_{0b} & L_{bd} \end{bmatrix}, \tag{3.36}$$

$$B_{0a} := \begin{bmatrix} B_{0a}^- \\ B_{0a}^0 \\ B_{0a}^+ \end{bmatrix}, \quad L_{ab} := \begin{bmatrix} L_{ab}^- \\ L_{ab}^0 \\ L_{ab}^+ \end{bmatrix}, \quad L_{ad} := \begin{bmatrix} L_{ad}^- \\ L_{ad}^0 \\ L_{ad}^+ \end{bmatrix}. \tag{3.37}$$

Similarly, Σ_* is controllable (stabilisable) if and only if the pair $(A_{\text{con}}, B_{\text{con}})$ is controllable (stabilisable). ◊

The invariant zeros of a system Σ_* characterised by (A_*, B_*, C_*, D_*) can be defined via the Smith canonical form of the (Rosenbrock) system matrix [52] of Σ_*:

$$P_{\Sigma_*}(\varsigma) := \begin{bmatrix} \varsigma I - A_* & -B_* \\ C_* & D_* \end{bmatrix}. \tag{3.38}$$

We have the following definition for the invariant zeros (see also [53]).

Definition 3.4.1. **(Invariant Zeros).** A complex scalar $\alpha \in \mathbb{C}$ is said to be an invariant zero of Σ_* if

$$\text{rank } \{P_{\Sigma_*}(\alpha)\} < n + \text{normrank } \{H_*(\varsigma)\}, \tag{3.39}$$

where normrank $\{H_*(\varsigma)\}$ denotes the normal rank of $H_*(\varsigma)$, which is defined as its rank over the field of rational functions of ς with real coefficients. $\Diamond$

The special coordinate basis of Theorem 3.4.1 shows explicitly the invariant zeros and the normal rank of Σ_*. To be more specific, we have the following properties.

Property 3.4.2.

1. The normal rank of $H_*(\varsigma)$ is equal to $m_0 + m_d$.
2. Invariant zeros of Σ_* are the eigenvalues of A_{aa}, which are the unions of the eigenvalues of A_{aa}^-, A_{aa}^0 and A_{aa}^+. Moreover, the given system Σ_* is of minimum phase if and only if A_{aa} has only stable eigenvalues, marginal minimum phase if and only if A_{aa} has no unstable eigenvalue but has at least one marginally stable eigenvalue, and nonminimum phase if and only if A_{aa} has at least one unstable eigenvalue. $\Diamond$

In order to display various multiplicities of invariant zeros, let X_{a} be a nonsingular transformation matrix such that A_{aa} can be transformed into a Jordan canonical form (see Theorem 3.2.1), *i.e.*,

$$X_{\mathrm{a}}^{-1} A_{aa} X_{\mathrm{a}} = J = \text{blkdiag} \left\{ J_1, J_2, \cdots, J_k \right\}, \tag{3.40}$$

where J_i, $i = 1, 2, \cdots, k$, are some $n_i \times n_i$ Jordan blocks:

$$J_i = \text{diag} \left\{ \alpha_i, \alpha_i, \cdots, \alpha_i \right\} + \begin{bmatrix} 0 & I_{n_i - 1} \\ 0 & 0 \end{bmatrix}. \tag{3.41}$$

For any given $\alpha \in \lambda(A_{aa})$, let there be τ_α Jordan blocks of A_{aa} associated with α. Let $n_{\alpha,1}$, $n_{\alpha,2}$, $\cdots$, n_{α,τ_α} be the dimensions of the corresponding Jordan blocks. Then we say α is an invariant zero of Σ_* with multiplicity structure $S^\star_\alpha(\Sigma_*)$,

$$S^\star_\alpha(\Sigma_*) = \left\{ n_{\alpha,1}, n_{\alpha,2}, \cdots, n_{\alpha,\tau_\alpha} \right\}. \tag{3.42}$$

The geometric multiplicity of α is then simply given by τ_α, and the algebraic multiplicity of α is given by $\sum_{i=1}^{\tau_\alpha} n_{\alpha,i}$. Here we should note that the invariant zeros, together with their structures of Σ_*, are related to the structural invariant indices list $\mathcal{I}_1(\Sigma_*)$ of Morse [54].

The special coordinate basis can also reveal the infinite zero structure of Σ_*. We note that the infinite zero structure of Σ_* can be either defined in association with root-locus theory or as Smith–McMillan zeros of the transfer function at infinity. For the sake of simplicity, we only consider the infinite zeros from the point of view of Smith–McMillan theory here. To define the zero structure of $H_*(\varsigma)$ at infinity, one can use the familiar Smith–McMillan description of the zero structure at finite frequencies of a general not necessarily square but strictly proper transfer function matrix $H_*(\varsigma)$. Namely, a rational matrix $H_*(\varsigma)$ possesses an infinite zero of order k when $H_*(1/z)$ has a finite zero of precisely that order at $z = 0$ (see [52, 55–57]). The number of zeros at infinity, together with their orders, indeed defines an infinite zero structure. Owens [58] related the orders of the infinite zeros of the root-loci of a square system with a nonsingular transfer function matrix to the $\mathcal{C}^*$ structural invariant indices list $\mathcal{I}_4$ of Morse [54]. This connection reveals that, even for general not necessarily strictly proper systems, the *structure at infinity is in fact the topology of inherent integrations between the input and the output variables.* The special coordinate basis of Theorem 3.4.1 explicitly shows this topology of inherent integrations. The following property pinpoints this.

Property 3.4.3. Σ_* has $m_0 =$ rank (D_*) infinite zeros of order 0. The infinite zero structure (of order greater than 0) of Σ_* is given by

$$S^\star_\infty(\Sigma_*) = \left\{ q_1, q_2, \cdots, q_{m_d} \right\}. \tag{3.43}$$

That is, each q_i corresponds to an infinite zero of Σ_* of order q_i. Note that for an SISO system Σ_*, we have $S^\star_\infty(\Sigma_*) = \{q_1\}$, where q_1 is the *relative degree* of Σ_*. $\Diamond$

The special coordinate basis can also exhibit the invertibility structure of a given system Σ_*. The formal definitions of right invertibility and left invertibility of a linear system can be found in [59]. Basically, for the usual case

when $[B_*' \quad D_*']$ and $[C_* \quad D_*]$ are of maximal rank, the system Σ_*, or equivalently $H_*(\varsigma)$, is said to be left invertible if there exists a rational matrix function, say $L_*(\varsigma)$, such that

$$L_*(\varsigma)H_*(\varsigma) = I_m. \tag{3.44}$$

Σ_* or $H_*(\varsigma)$ is said to be right invertible if there exists a rational matrix function, say $R_*(\varsigma)$, such that

$$H_*(\varsigma)R_*(\varsigma) = I_p. \tag{3.45}$$

Σ_* is invertible if it is both left and right invertible, and Σ_* is degenerate if it is neither left nor right invertible.

Property 3.4.4. The given system Σ_* is right invertible if and only if x_b (and hence y_b) are nonexistent, left invertible if and only if x_c (and hence u_c) are nonexistent, and invertible if and only if both x_b and x_c are nonexistent. Moreover, Σ_* is degenerate if and only if both x_b and x_c are present. $\Diamond$

The special coordinate basis can also be modified to obtain the structural invariant indices lists $\mathcal{I}_2$ and $\mathcal{I}_3$ of Morse [54] of the given system Σ_*. In order to display $\mathcal{I}_2(\Sigma_*)$, we let X_{c} and X_{i} be nonsingular matrices such that the controllable pair (A_{cc}, B_c) is transformed into the CSD (see Theorem 3.3.1), *i.e.*,

$$X_{\mathrm{c}}^{-1}A_{cc}X_{\mathrm{c}} = \begin{bmatrix} 0 & I_{\ell_1-1} & \cdots & 0 & 0 \\ \star & \star & \cdots & \star & \star \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & I_{\ell_{m_c}-1} \\ \star & \star & \cdots & \star & \star \end{bmatrix}, \tag{3.46}$$

$$X_{\mathrm{c}}^{-1}B_cX_{\mathrm{i}} = \begin{bmatrix} 0 & \cdots & 0 \\ 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \\ 0 & \cdots & 1 \end{bmatrix}, \tag{3.47}$$

where the $\star$ symbols denote constant scalars or row vectors. Then we have

$$\mathcal{I}_2(\Sigma_*) = \left\{\ell_1, \cdots, \ell_{m_c}\right\}, \tag{3.48}$$

which is also called the controllability index of (A_{cc}, B_c). Similarly, we have

$$\mathcal{I}_3(\Sigma_*) = \left\{\mu_1, \cdots, \mu_{p_b}\right\}, \tag{3.49}$$

where $\left\{\mu_1, \cdots, \mu_{p_b}\right\}$ is the controllability index of the controllable pair (A'_{bb}, C'_b).

By now it is clear that the special coordinate basis decomposes the state-space into several distinct parts. In fact, the state-space $\mathcal{X}$ is decomposed as

$$\mathcal{X} = \mathcal{X}_a^- \oplus \mathcal{X}_a^0 \oplus \mathcal{X}_a^+ \oplus \mathcal{X}_b \oplus \mathcal{X}_c \oplus \mathcal{X}_d. \tag{3.50}$$

Here $\mathcal{X}_a^-$ is related to the stable invariant zeros, *i.e.*, the eigenvalues of A_{aa}^- are the stable invariant zeros of Σ_*. Similarly, $\mathcal{X}_a^0$ and $\mathcal{X}_a^+$ are respectively related to the invariant zeros of Σ_* located in the marginally stable and unstable regions. On the other hand, $\mathcal{X}_b$ is related to the right invertibility, *i.e.*, the system is right invertible if and only if $\mathcal{X}_b = \{0\}$, whereas $\mathcal{X}_c$ is related to left invertibility, *i.e.*, the system is left invertible if and only if $\mathcal{X}_c = \{0\}$. Finally, $\mathcal{X}_d$ is related to zeros of Σ_* at infinity.

There are interconnections between the special coordinate basis and various invariant geometric subspaces. To show these interconnections, we introduce the following geometric subspaces.

Definition 3.4.2. **(Geometric Subspaces $\mathcal{V}^{\mathrm{x}}$ and $\mathcal{S}^{\mathrm{x}}$).** The weakly unobservable subspaces of Σ_*, $\mathcal{V}^{\mathrm{x}}$, and the strongly controllable subspaces of Σ_*, $\mathcal{S}^{\mathrm{x}}$, are defined as follows:

1. $\mathcal{V}^{\mathrm{x}}(\Sigma_*)$ is the maximal subspace of $\mathbb{R}^n$ that is $(A_*+B_*F_*)$-invariant and contained in $\mathrm{Ker}\,(C_*+D_*F_*)$ such that the eigenvalues of $(A_*+B_*F_*)|\mathcal{V}^{\mathrm{x}}$ are contained in $\mathbb{C}^{\mathrm{x}} \subseteq \mathbb{C}$ for some constant matrix F_*.

2. $\mathcal{S}^{\mathrm{x}}(\Sigma_*)$ is the minimal $(A_* + K_*C_*)$-invariant subspace of $\mathbb{R}^n$ containing $\mathrm{Im}\,(B_* + K_*D_*)$ such that the eigenvalues of the map that is induced by $(A_* + K_*C_*)$ on the factor space $\mathbb{R}^n/\mathcal{S}^{\mathrm{x}}$ are contained in $\mathbb{C}^{\mathrm{x}} \subseteq \mathbb{C}$ for some constant matrix K_*.

Moreover, we let $\mathcal{V}^-=\mathcal{V}^{\mathrm{x}}$ and $\mathcal{S}^-=\mathcal{S}^{\mathrm{x}}$, if $\mathbb{C}^{\mathrm{x}}=\mathbb{C}^- \cup \mathbb{C}^0$; $\mathcal{V}^+=\mathcal{V}^{\mathrm{x}}$ and $\mathcal{S}^+=\mathcal{S}^{\mathrm{x}}$, if $\mathbb{C}^{\mathrm{x}}=\mathbb{C}^+$; $\mathcal{V}^\odot=\mathcal{V}^{\mathrm{x}}$ and $\mathcal{S}^\odot=\mathcal{S}^{\mathrm{x}}$, if $\mathbb{C}^{\mathrm{x}}=\mathbb{C}^\odot \cup \mathbb{C}^\circ$; $\mathcal{V}^\otimes=\mathcal{V}^{\mathrm{x}}$ and $\mathcal{S}^\otimes=\mathcal{S}^{\mathrm{x}}$, if $\mathbb{C}^{\mathrm{x}}=\mathbb{C}^\otimes$; and finally $\mathcal{V}^*=\mathcal{V}^{\mathrm{x}}$ and $\mathcal{S}^*=\mathcal{S}^{\mathrm{x}}$, if $\mathbb{C}^{\mathrm{x}}=\mathbb{C}$. ◇

Property 3.4.5.

1. $\mathcal{X}_a^- \oplus \mathcal{X}_a^0 \oplus \mathcal{X}_c$ spans $\begin{cases} \mathcal{V}^-(\Sigma_*), & \text{if } \Sigma_* \text{ is of continuous-time,} \\ \mathcal{V}^\odot(\Sigma_*), & \text{if } \Sigma_* \text{ is of discrete-time.} \end{cases}$

2. $\mathcal{X}_a^+ \oplus \mathcal{X}_c$ spans $\begin{cases} \mathcal{V}^+(\Sigma_*), & \text{if } \Sigma_* \text{ is of continuous-time,} \\ \mathcal{V}^{\otimes}(\Sigma_*), & \text{if } \Sigma_* \text{ is of discrete-time.} \end{cases}$

3. $\mathcal{X}_a^- \oplus \mathcal{X}_a^0 \oplus \mathcal{X}_a^+ \oplus \mathcal{X}_c$ spans $\mathcal{V}^*(\Sigma_*)$.

4. $\mathcal{X}_a^+ \oplus \mathcal{X}_c \oplus \mathcal{X}_d$ spans $\begin{cases} \mathcal{S}^-(\Sigma_*), & \text{if } \Sigma_* \text{ is of continuous-time,} \\ \mathcal{S}^{\odot}(\Sigma_*), & \text{if } \Sigma_* \text{ is of discrete-time.} \end{cases}$

5. $\mathcal{X}_a^- \oplus \mathcal{X}_a^0 \oplus \mathcal{X}_c \oplus \mathcal{X}_d$ spans $\begin{cases} \mathcal{S}^+(\Sigma_*), & \text{if } \Sigma_* \text{ is of continuous-time,} \\ \mathcal{S}^{\otimes}(\Sigma_*), & \text{if } \Sigma_* \text{ is of discrete-time.} \end{cases}$

6. $\mathcal{X}_c \oplus \mathcal{X}_d$ spans $\mathcal{S}^*(\Sigma_*)$. ◊

Finally, for future development on deriving solvability conditions for H_∞ almost disturbance decoupling problems, we introduce two more subspaces of Σ_*. The original definitions of these subspaces were given by Scherer [60].

Definition 3.4.3. **(Geometric Subspaces $\mathcal{V}_\lambda$ and $\mathcal{S}_\lambda$).** For any $\lambda \in \mathbb{C}$, we define

$$\mathcal{V}_\lambda(\Sigma_*) := \left\{ \zeta \in \mathbb{C}^n \;\middle|\; \exists\, \omega \in \mathbb{C}^m : 0 = \begin{bmatrix} A_* - \lambda I & B_* \\ C_* & D_* \end{bmatrix} \begin{pmatrix} \zeta \\ \omega \end{pmatrix} \right\} \tag{3.51}$$

and

$$\mathcal{S}_\lambda(\Sigma_*) := \left\{ \zeta \in \mathbb{C}^n \;\middle|\; \exists\, \omega \in \mathbb{C}^{n+m} : \begin{pmatrix} \zeta \\ 0 \end{pmatrix} = \begin{bmatrix} A_* - \lambda I & B_* \\ C_* & D_* \end{bmatrix} \omega \right\} \tag{3.52}$$

$\mathcal{V}_\lambda(\Sigma_*)$ and $\mathcal{S}_\lambda(\Sigma_*)$ are associated with the so-called state zero directions of Σ_* if λ is an invariant zero of Σ_*. ◊

These subspaces $\mathcal{S}_\lambda(\Sigma_*)$ and $\mathcal{V}_\lambda(\Sigma_*)$ can also be easily obtained using the special coordinate basis. We have the following new property of the special coordinate basis.

Property 3.4.6.

$$\mathcal{S}_\lambda(\Sigma_*) = \operatorname{Im} \left\{ \Gamma_s \begin{bmatrix} \lambda I - A_{aa} & 0 & 0 & 0 \\ 0 & Y_{b\lambda} & 0 & 0 \\ 0 & 0 & I_{n_c} & 0 \\ 0 & 0 & 0 & I_{n_d} \end{bmatrix} \right\}, \tag{3.53}$$

where

$$\text{Im}\,\{Y_{b\lambda}\} = \text{Ker}\,\left[C_b(A_{bb} + K_bC_b - \lambda I)^{-1}\right], \tag{3.54}$$

and where K_b is any appropriately dimensional matrix subject to the constraint that $A_{bb} + K_bC_b$ has no eigenvalue at λ. We note that such a K_b always exists, as (A_{bb}, C_b) is completely observable.

$$\mathcal{V}_\lambda(\Sigma_*) = \text{Im}\left\{\Gamma_s\begin{bmatrix} X_{a\lambda} & 0 \\ 0 & 0 \\ 0 & X_{c\lambda} \\ 0 & 0 \end{bmatrix}\right\}, \tag{3.55}$$

where $X_{a\lambda}$ is a matrix whose columns form a basis for the subspace,

$$\Big\{\zeta_a \in \mathbb{C}^{n_a} \;\Big|\; (\lambda I - A_{aa})\zeta_a = 0\Big\}, \tag{3.56}$$

and

$$X_{c\lambda} := \Big(A_{cc} + B_cF_c - \lambda I\Big)^{-1}B_c, \tag{3.57}$$

with F_c being any appropriately dimensional matrix subject to the constraint that $A_{cc} + B_cF_c$ has no eigenvalue at λ. Again, we note that the existence of such an F_c is guaranteed by the controllability of (A_{cc}, B_c). ◇

Clearly, if $\lambda \notin \lambda(A_{aa})$, then we have

$$\mathcal{V}_\lambda(\Sigma_*) \subseteq \mathcal{V}^{\text{x}}(\Sigma_*), \quad \mathcal{S}_\lambda(\Sigma_*) \supseteq \mathcal{S}^{\text{x}}(\Sigma_*). \tag{3.58}$$

It is interesting to note that the subspaces $\mathcal{V}^{\text{x}}(\Sigma_*)$ and $\mathcal{S}^{\text{x}}(\Sigma_*)$ are dual in the sense that $\mathcal{V}^{\text{x}}(\Sigma_*^\star) = \mathcal{S}^{\text{x}}(\Sigma_*)^\perp$, where $\Sigma_*^\star$ is characterised by the quadruple (A_*', C_*', B_*', D_*'). Also, $\mathcal{S}_\lambda(\Sigma_*) = \mathcal{V}_{\bar{\lambda}}(\Sigma_*^\star)^\perp$. Finally, we conclude this section by summarising in Figure 3.2 some major properties of the tools of linear systems, which combines the mechanisms of the special coordinate basis, the Jordan canonical form and the CSD. Such tools have been used in the literature to solve many system and control problems (see Chen [44] for details).

3.5 Bilinear Transformations

We recall in this section the work of Chen and Weller [61] on bilinear and inverse bilinear transformations of linear time-invariant systems (see also Chen

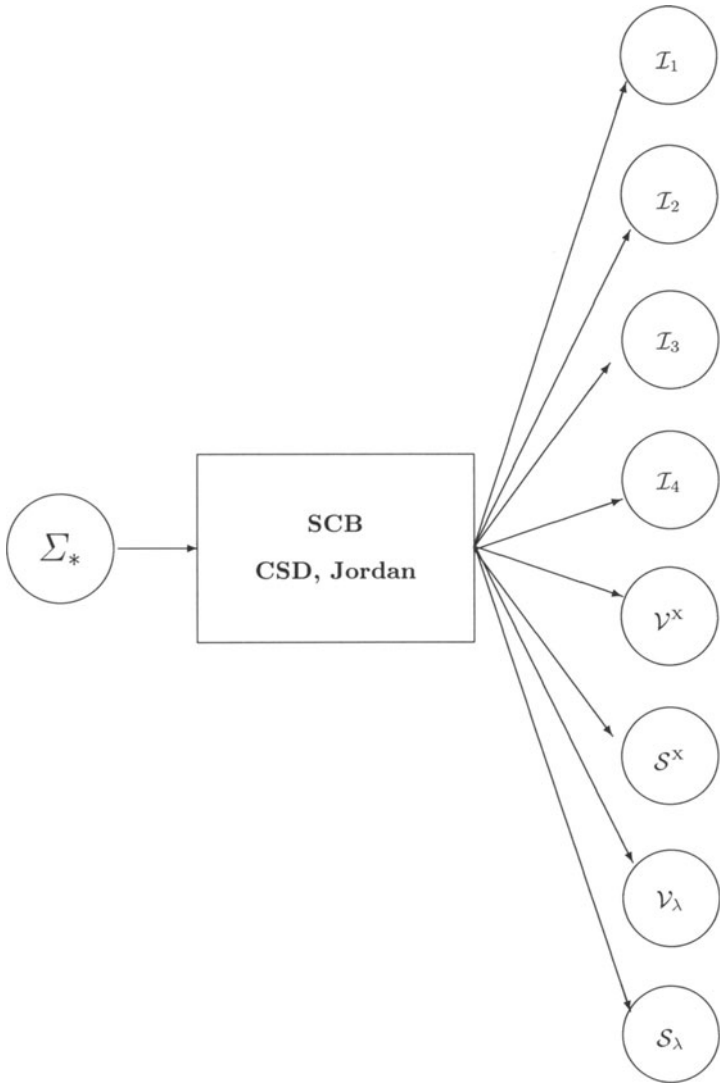


Figure 3.2. Tools and structural properties of linear time-invariant systems.

[44]). Their result presents a comprehensive picture of the mapping of structural properties associated with general linear multivariable systems under bilinear and inverse bilinear transformations. They have completely investigated the problem of how the finite and infinite zero structures, as well as invertibility structures of a general continuous-time (discrete-time) linear time-invariant multivariable system, are mapped to those of its discrete-time (continuous-time) counterpart under the bilinear (inverse bilinear) transfor-

mation. In what follows, we present a comprehensive study on the structural mappings of the well-known bilinear (inverse bilinear) transformations

$$s = a\left(\frac{z-1}{z+1}\right) \quad \text{and} \quad z = \frac{a+s}{a-s} \tag{3.59}$$

respectively.

3.5.1 Continuous to Discrete

In this subsection, we will consider a continuous-time linear time-invariant system Σ_c characterised by

$$\Sigma_c : \begin{cases} \dot{x} = A\,x + B\,u, \\ y = C\,x + D\,u, \end{cases} \tag{3.60}$$

where $x \in \mathbb{R}^n$, $y \in \mathbb{R}^p$, $u \in \mathbb{R}^m$ and A, B, C and D are matrices of appropriate dimensions. Without loss of any generality, we assume that both matrices $[C \;\; D]$ and $[B' \;\; D']$ are of full rank. Σ_c has a transfer function

$$G_c(s) = C(sI - A)^{-1}B + D. \tag{3.61}$$

Let us apply a bilinear transformation to the above continuous-time system, by replacing s in Equation (3.61) with

$$s = \frac{2}{T}\left(\frac{z-1}{z+1}\right) = a\left(\frac{z-1}{z+1}\right), \tag{3.62}$$

where $T = 2/a$ is the sampling period. As presented in Equation (3.62), the bilinear transformation is often called Tustin's approximation [62], whereas the choice

$$a = \frac{\omega_1}{\tan(\omega_1 T/2)} \tag{3.63}$$

yields the pre-warped Tustin approximation, in which the frequency responses of the continuous-time system and its discrete-time counterpart are matched at frequency ω_1. In this way, we obtain a discrete-time system

$$G_d(z) = C\left(a\frac{z-1}{z+1}I - A\right)^{-1}B + D. \tag{3.64}$$

The following lemma provides a direct state-space realisation of $G_d(z)$.

Lemma 3.5.1. A state-space realisation of $G_{\mathrm{d}}(z)$, the discrete-time counterpart of the continuous-time system Σ_{c} of Equation (3.60) under the bilinear transformation of Equation (3.62), is given by

$$\Sigma_{\mathrm{d}} : \begin{cases} x(k+1) = \tilde{A}\, x(k) + \tilde{B}\, u(k), \\ y(k) \quad = \tilde{C}\, x(k) + \tilde{D}\, u(k), \end{cases} \tag{3.65}$$

where

$$\left.\begin{aligned} \tilde{A} &= (aI + A)(aI - A)^{-1}, \\ \tilde{B} &= \sqrt{2a}\,(aI - A)^{-1}B, \\ \tilde{C} &= \sqrt{2a}\,C(aI - A)^{-1}, \\ \tilde{D} &= D + C(aI - A)^{-1}B. \end{aligned}\right\} \tag{3.66}$$

Here we clearly assume that matrix A has no eigenvalue at a. ◊

The following theorem establishes the interconnection of the structural properties of Σ_{c} and Σ_{d}.

Theorem 3.5.1. Consider the continuous-time system Σ_{c} of Equation (3.60) characterised by the quadruple (A, B, C, D) with matrix A having no eigenvalue at a, and its discrete-time counterpart under the bilinear transformation of Equation (3.62), *i.e.*, Σ_{d} of Equation (3.65) characterised by the quadruple $(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D})$ of Equation (3.66). We have the following properties.

1. Controllability (stabilisability) and observability (detectability) of Σ_{d}:
   a) the pair $(\tilde{A}, \tilde{B})$ is controllable (stabilisable) if and only if the pair (A, B) is controllable (stabilisable);
   b) the pair $(\tilde{A}, \tilde{C})$ is observable (detectable) if and only if the pair (A, C) is observable (detectable).
2. Effects of nonsingular state, output and input transformations, together with state feedback and output injection laws:
   a) for any given nonsingular state, output and input transformations T_{s}, T_{o} and T_{i}, the quadruple

$$(T_{\mathrm{s}}^{-1}\tilde{A}T_{\mathrm{s}}, T_{\mathrm{s}}^{-1}\tilde{B}T_{\mathrm{i}}, T_{\mathrm{o}}^{-1}\tilde{C}T_{\mathrm{s}}, T_{\mathrm{o}}^{-1}\tilde{D}T_{\mathrm{i}}), \tag{3.67}$$

   is the discrete-time counterpart under the bilinear transformation of Equation (3.62), of the continuous-time system

$$(T_{\mathrm{s}}^{-1}AT_{\mathrm{s}}, T_{\mathrm{s}}^{-1}BT_{\mathrm{i}}, T_{\mathrm{o}}^{-1}CT_{\mathrm{s}}, T_{\mathrm{o}}^{-1}DT_{\mathrm{i}}); \tag{3.68}$$

b) for any $F \in \mathbb{R}^{m\times n}$ with $A+BF$ having no eigenvalue at a, define a nonsingular matrix

$$\begin{aligned}\tilde{T}_{\mathrm{i}} &:= I + F(aI - A - BF)^{-1}B \\ &= [I - F(aI - A)^{-1}B]^{-1} \in \mathbb{R}^{m\times m},\end{aligned} \tag{3.69}$$

and a constant matrix

$$\tilde{F} := \sqrt{2a}\, F(aI - A - BF)^{-1} \in \mathbb{R}^{m\times n}. \tag{3.70}$$

Then a continuous-time system Σ_{cF} characterised by

$$(A + BF, B, C + DF, D), \tag{3.71}$$

is mapped to a discrete-time system Σ_{dF}, characterised by

$$(\tilde{A} + \tilde{B}\tilde{F}, \tilde{B}\tilde{T}_{\mathrm{i}}, \tilde{C} + \tilde{D}\tilde{F}, \tilde{D}\tilde{T}_{\mathrm{i}}), \tag{3.72}$$

under the bilinear transformation of Equation (3.62). Here we note that Σ_{cF} is the closed-loop system comprising Σ_{c} and a state feedback law with gain matrix F, and Σ_{dF} is the closed-loop system comprising Σ_{d} and a state feedback law with gain matrix $\tilde{F}$, together with a nonsingular input transformation $\tilde{T}_{\mathrm{i}}$;

c) for any $K \in \mathbb{R}^{n\times p}$ with $A+KC$ having no eigenvalue at a, define a nonsingular matrix

$$\tilde{T}_{\mathrm{o}} := [I + C(aI - A - KC)^{-1}K]^{-1} \in \mathbb{R}^{p\times p}, \tag{3.73}$$

and a constant matrix

$$\tilde{K} := \sqrt{2a}\,(aI - A - KC)^{-1}K. \tag{3.74}$$

Then a continuous-time system Σ_{cK} characterised by

$$(A + KC, B + KD, C, D), \tag{3.75}$$

is mapped to a discrete-time system Σ_{dK}, characterised by

$$(\tilde{A} + \tilde{K}\tilde{C}, \tilde{B} + \tilde{K}\tilde{D}, \tilde{T}_{\mathrm{o}}^{-1}\tilde{C}, \tilde{T}_{\mathrm{o}}^{-1}\tilde{D}), \tag{3.76}$$

under the bilinear transformation of Equation (3.62). We note that Σ_{cK} is the closed-loop system comprising Σ_{c} and an output injection law with gain matrix K, and Σ_{dK} is the closed-loop system comprising Σ_{d} and an output injection law with gain matrix $\tilde{K}$, together with a nonsingular output transformation $\tilde{T}_{\mathrm{o}}$.

3. Invertibility and structural invariant indices lists $\mathcal{I}_2$ and $\mathcal{I}_3$ of Σ_{d}:

a) $\mathcal{I}_2(\Sigma_d) = \mathcal{I}_2(\Sigma_c)$, and $\mathcal{I}_3(\Sigma_d) = \mathcal{I}_3(\Sigma_c)$.

b) Σ_d is left (right) invertible if and only if Σ_c is left (right) invertible.

c) Σ_d is invertible (degenerate) if and only if Σ_c is invertible (degenerate).

4. The invariant zeros of Σ_d and their associated structures consist of the following two parts.

   a) Let the infinite zero structure (of order greater than 0) of Σ_c be given by $S^\star_\infty(\Sigma_c) = \{q_1, q_2, \cdots, q_{m_d}\}$. Then $z = -1$ is an invariant zero of Σ_d with the multiplicity structure $S^\star_{-1}(\Sigma_d) = \{q_1, q_2, \cdots, q_{m_d}\}$.

   b) Let $s = \alpha \neq a$ be an invariant zero of Σ_c with the multiplicity structure $S^\star_\alpha(\Sigma_c) = \{n_{\alpha,1}, n_{\alpha,2}, \cdots, n_{\alpha,\tau_\alpha}\}$. Then $z = \beta = (a+\alpha)/(a-\alpha)$ is an invariant zero of its discrete-time counterpart Σ_d with the multiplicity structure $S^\star_\beta(\Sigma_d) = \{n_{\alpha,1}, n_{\alpha,2}, \cdots, n_{\alpha,\tau_\alpha}\}$.

5. The infinite zero structure of Σ_d consists of the following two parts:

   a) Let $m_0 =$ rank (D), and let m_d be the total number of infinite zeros of Σ_c of order greater than 0. Also, let τ_a be the geometric multiplicity of the invariant zero of Σ_c at $s = a$. Then we have rank $(\tilde{D}) = m_0 + m_d - \tau_a$.

   b) Let $s = a$ be an invariant zero of the given continuous-time system Σ_c with a multiplicity structure $S^\star_a(\Sigma_c) = \{n_{a,1}, n_{a,2}, \cdots, n_{a,\tau_a}\}$. Then the discrete-time counterpart Σ_d has an infinite zero (of order greater than 0) structure $S^\star_\infty(\Sigma_d) = \{n_{a,1}, n_{a,2}, \cdots, n_{a,\tau_a}\}$.

6. The mappings of geometric subspaces:

   a) $\mathcal{V}^+(\Sigma_c) = \mathcal{S}^\odot(\Sigma_d)$.

   b) $\mathcal{S}^+(\Sigma_c) = \mathcal{V}^\odot(\Sigma_d)$. ◇

We have the following two interesting observations. The first is with regard to the minimum phase and nonminimum phase properties of Σ_d, and the second concerns the asymptotic behaviour of Σ_d as the sampling period T tends to zero (or, equivalently, as $a \to \infty$).

Observation 3.5.1. Consider a general continuous-time system Σ_c and its discrete-time counterpart Σ_d under the bilinear transformation of Equation (3.62). Then it follows from 4(a) and 4(b) of Theorem 3.5.1 that

1. Σ_d has all its invariant zeros inside the unit circle if and only if Σ_c has all its invariant zeros in the open left-half plane and has no infinite zero of order greater than 0;

2. Σ_d has invariant zeros on the unit circle if and only if Σ_c has invariant zeros on the imaginary axis, and/or Σ_c has at least one infinite zero of order greater than 0;

3. Σ_d has invariant zeros outside the unit circle if and only if Σ_c has invariant zeros in the open right-half plane. $\Diamond$

Observation 3.5.2. Consider a general continuous-time system Σ_c and its discrete-time counterpart Σ_d under the bilinear transformation of Equation (3.62). Then as a consequence of Theorem 3.5.1, Σ_d has the following asymptotic properties as the sampling period T tends to zero (but not equal to zero):

1. Σ_d has no infinite zero of order greater than 0, *i.e.*, no delays from the input to the output;

2. Σ_d has one invariant zero at $z = -1$ with an appropriate multiplicity structure if Σ_c has any infinite zero of order greater than 0; and

3. The remaining invariant zeros of Σ_d, if any, tend to the point $z = 1$. More interestingly, the invariant zeros of Σ_d corresponding to the stable invariant zeros of Σ_c are always stable, and approach the point $z = 1$ from inside the unit circle. Conversely, the invariant zeros of Σ_d corresponding to the unstable invariant zeros of Σ_c are always unstable, and approach the point $z = 1$ from outside the unit circle. Finally, those associated with the imaginary axis invariant zeros of Σ_c are always mapped onto the unit circle and move towards to the point $z = 1$. $\Diamond$

3.5.2 Discrete to Continuous

We present in this subsection a similar result as in the previous section, but for the inverse bilinear transformation mapping a discrete-time system to a continuous-time system. We begin with a discrete-time linear time-invariant system $\tilde{\Sigma}_d$ characterised by

$$\tilde{\Sigma}_d : \begin{cases} x(k+1) = \tilde{A}\, x(k) + \tilde{B}\, u(k), \\ y(k) \quad = \tilde{C}\, x(k) + \tilde{D}\, u(k), \end{cases} \tag{3.77}$$

where $x \in \mathbb{R}^n$, $y \in \mathbb{R}^p$, $u \in \mathbb{R}^m$ and $\tilde{A}$, $\tilde{B}$, $\tilde{C}$ and $\tilde{D}$ are matrices of appropriate dimensions. Without loss of any generality, we assume that both matrices $[\tilde{C} \;\; \tilde{D}]$ and $[\tilde{B}' \;\; \tilde{D}']$ are of full rank. Σ_{d} has a transfer function

$$H_{\mathrm{d}}(z) = \tilde{C}(zI - \tilde{A})^{-1}\tilde{B} + \tilde{D}. \tag{3.78}$$

The inverse bilinear transformation corresponding to (3.62) replaces z in Equation (3.78) with

$$z = \frac{a+s}{a-s}, \tag{3.79}$$

to obtain the following continuous-time system:

$$H_{\mathrm{c}}(s) = \tilde{C}\left(\frac{a+s}{a-s}I - \tilde{A}\right)^{-1}\tilde{B} + \tilde{D}. \tag{3.80}$$

The following lemma is analogous to Lemma 3.5.1, and provides a state-space realisation of $H_{\mathrm{c}}(s)$.

Lemma 3.5.2. A state-space realisation of $H_{\mathrm{c}}(s)$, the continuous-time counterpart of the discrete-time system $\tilde{\Sigma}_{\mathrm{d}}$ of Equation (3.77) under the inverse bilinear transformation of Equation (3.79), is given by

$$\tilde{\Sigma}_{\mathrm{c}} \;:\; \begin{cases} \dot{x} = A\,x + B\,u, \\ y = C\,x + D\,u, \end{cases} \tag{3.81}$$

where

$$\left.\begin{aligned} A &= a(\tilde{A}+I)^{-1}(\tilde{A}-I), \\ B &= \sqrt{2a}\,(\tilde{A}+I)^{-1}\tilde{B}, \\ C &= \sqrt{2a}\,\tilde{C}(\tilde{A}+I)^{-1}, \\ D &= \tilde{D} - \tilde{C}(\tilde{A}+I)^{-1}\tilde{B}. \end{aligned}\right\} \tag{3.82}$$

Here we clearly assume that the matrix $\tilde{A}$ has no eigenvalue at -1. ◊

The following theorem is analogous to Theorem 3.5.1.

Theorem 3.5.2. Consider the discrete-time system $\tilde{\Sigma}_{\mathrm{d}}$ of Equation (3.77) characterised by the quadruple $(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D})$ with matrix $\tilde{A}$ having no eigenvalue at -1, and its continuous-time counterpart under the inverse bilinear

transformation of Equation (3.79), *i.e.*, $\tilde{\Sigma}_{\mathrm{c}}$ of Equation (3.81) characterised by the quadruple (A, B, C, D) of Equation (3.82). We have the following properties.

1. Controllability (stabilisability) and observability (detectability) of $\tilde{\Sigma}_{\mathrm{c}}$:

   a) the pair (A, B) is controllable (stabilisable) if and only if the pair $(\tilde{A}, \tilde{B})$ is controllable (stabilisable).

   b) the pair (A, C) is observable (detectable) if and only if the pair $(\tilde{A}, \tilde{C})$ is observable (detectable).

2. Effects of nonsingular state, output and input transformations, together with state feedback and output injection laws:

   a) for any given nonsingular state, output and input transformations T_{s}, T_{o} and T_{i}, the quadruple

   $$(T_{\mathrm{s}}^{-1} A T_{\mathrm{s}}, T_{\mathrm{s}}^{-1} B T_{\mathrm{i}}, T_{\mathrm{o}}^{-1} C T_{\mathrm{s}}, T_{\mathrm{o}}^{-1} D T_{\mathrm{i}}), \tag{3.83}$$

   is the continuous-time counterpart of the inverse bilinear transformation, *i.e.*, (3.79), of the discrete-time system

   $$(T_{\mathrm{s}}^{-1} \tilde{A} T_{\mathrm{s}}, T_{\mathrm{s}}^{-1} \tilde{B} T_{\mathrm{i}}, T_{\mathrm{o}}^{-1} \tilde{C} T_{\mathrm{s}}, T_{\mathrm{o}}^{-1} \tilde{D} T_{\mathrm{i}}); \tag{3.84}$$

   b) for any $\tilde{F} \in \mathbb{R}^{m \times n}$ with $\tilde{A} + \tilde{B}\tilde{F}$ having no eigenvalue at -1, define a nonsingular matrix

   $$T_{\mathrm{i}} := I - \tilde{F}(I + \tilde{A} + \tilde{B}\tilde{F})^{-1} \tilde{B} \in \mathbb{R}^{m \times m}, \tag{3.85}$$

   and a constant matrix

   $$F := \sqrt{2a}\, \tilde{F}(I + \tilde{A} + \tilde{B}\tilde{F})^{-1} \in \mathbb{R}^{m \times n}. \tag{3.86}$$

   Then a discrete-time system $\tilde{\Sigma}_{\mathrm{dF}}$ characterised by

   $$(\tilde{A} + \tilde{B}\tilde{F}, \tilde{B}, \tilde{C} + \tilde{D}\tilde{F}, \tilde{D}), \tag{3.87}$$

   is mapped to a continuous-time counterpart $\tilde{\Sigma}_{\mathrm{cF}}$ characterised by

   $$(A + BF, BT_{\mathrm{i}}, C + DF, DT_{\mathrm{i}}), \tag{3.88}$$

   under the inverse bilinear transformation of Equation (3.79). Note that $\tilde{\Sigma}_{\mathrm{dF}}$ is the closed-loop system comprising $\tilde{\Sigma}_{\mathrm{d}}$ and a state feedback law with gain matrix $\tilde{F}$, and $\tilde{\Sigma}_{\mathrm{dF}}$ is the closed-loop system comprising $\tilde{\Sigma}_{\mathrm{d}}$ and a state feedback law with gain matrix F, together with a nonsingular input transformation T_{i};

c) for any $\tilde{K} \in \mathbb{R}^{n\times p}$ with $\tilde{A}+\tilde{K}\tilde{C}$ having no eigenvalue at -1, define a nonsingular matrix

$$T_{\mathrm{o}} := [I - \tilde{C}(I+\tilde{A}+\tilde{K}\tilde{C})^{-1}\tilde{K}]^{-1} \in \mathbb{R}^{p\times p}, \tag{3.89}$$

and a constant matrix

$$K := \sqrt{2a}\,(I+\tilde{A}+\tilde{K}\tilde{C})^{-1}\tilde{K}. \tag{3.90}$$

Then a discrete-time system $\tilde{\Sigma}_{\mathrm{dK}}$ characterised by

$$(\tilde{A}+\tilde{K}\tilde{C}, \tilde{B}+\tilde{K}\tilde{D}, \tilde{C}, \tilde{D}), \tag{3.91}$$

is mapped to a continuous-time $\tilde{\Sigma}_{\mathrm{cK}}$, characterised by

$$(A+KC, B+KD, T_{\mathrm{o}}^{-1}C, T_{\mathrm{o}}^{-1}D), \tag{3.92}$$

under the inverse bilinear transformation of Equation (3.79). We note that $\tilde{\Sigma}_{\mathrm{dK}}$ is the closed-loop system comprising $\tilde{\Sigma}_{\mathrm{d}}$ and an output injection law with gain matrix $\tilde{K}$, and $\tilde{\Sigma}_{\mathrm{cK}}$ is the closed-loop system comprising $\tilde{\Sigma}_{\mathrm{c}}$ and an output injection law with gain matrix K, together with a nonsingular output transformation T_{o}.

3. Invertibility and structural invariant indices lists $\mathcal{I}_2$ and $\mathcal{I}_3$ of $\tilde{\Sigma}_{\mathrm{c}}$:

   a) $\mathcal{I}_2(\tilde{\Sigma}_{\mathrm{c}}) = \mathcal{I}_2(\tilde{\Sigma}_{\mathrm{d}})$ and $\mathcal{I}_3(\tilde{\Sigma}_{\mathrm{c}}) = \mathcal{I}_3(\tilde{\Sigma}_{\mathrm{d}})$;

   b) $\tilde{\Sigma}_{\mathrm{c}}$ is left (right) invertible if and only if $\tilde{\Sigma}_{\mathrm{d}}$ is left (right) invertible;

   c) $\tilde{\Sigma}_{\mathrm{c}}$ is invertible (degenerate) if and only if $\tilde{\Sigma}_{\mathrm{d}}$ is invertible (degenerate).

4. Invariant zeros of $\tilde{\Sigma}_{\mathrm{c}}$ and their structures consist of the following two parts.

   a) Let the infinite zero structure (of order greater than 0) of $\tilde{\Sigma}_{\mathrm{d}}$ be given by $S^{\star}_{\infty}(\tilde{\Sigma}_{\mathrm{d}}) = \{q_1, q_2, \cdots, q_{m_d}\}$. Then $s = a$ is an invariant zero of $\tilde{\Sigma}_{\mathrm{c}}$ with the multiplicity structure $S^{\star}_{a}(\tilde{\Sigma}_{\mathrm{c}}) = \{q_1, q_2, \cdots, q_{m_d}\}$.

   b) Let $z = \alpha \neq -1$ be an invariant zero of $\tilde{\Sigma}_{\mathrm{d}}$ with the multiplicity structure $S^{\star}_{\alpha}(\tilde{\Sigma}_{\mathrm{d}}) = \{n_{\alpha,1}, n_{\alpha,2}, \cdots, n_{\alpha,\tau_\alpha}\}$. Then $s = \beta = a\frac{\alpha-1}{\alpha+1}$ is an invariant zero of its continuous-time counterpart $\tilde{\Sigma}_{\mathrm{c}}$ with the multiplicity structure $S^{\star}_{\beta}(\tilde{\Sigma}_{\mathrm{c}}) = \{n_{\alpha,1}, n_{\alpha,2}, \cdots, n_{\alpha,\tau_\alpha}\}$.

5. The infinite zero structure of $\tilde{\Sigma}_{\mathrm{c}}$ consists of the following two parts.

   a) Let $m_0 = \text{rank}\,(\tilde{D})$, and let m_d be the total number of infinite zeros of $\tilde{\Sigma}_{\mathrm{d}}$ of order greater than 0. Also, let τ_{-1} be the geometric

multiplicity of the invariant zero of $\tilde{\Sigma}_{\mathrm{d}}$ at $z = -1$. Then we have rank $(D) = m_0 + m_d - \tau_{-1}$.

b) Let $z = -1$ be an invariant zero of the given discrete-time system $\tilde{\Sigma}_{\mathrm{d}}$ with the multiplicity structure $S^{\star}_{-1}(\tilde{\Sigma}_{\mathrm{d}}) = \{n_{-1,1}, n_{-1,2}, \cdots, n_{-1,\tau_{-1}}\}$. Then $\tilde{\Sigma}_{\mathrm{c}}$ has an infinite zero (of order greater than 0) structure $S^{\star}_{\infty}(\tilde{\Sigma}_{\mathrm{c}}) = \{n_{-1,1}, n_{-1,2}, \cdots, n_{-1,\tau_{-1}}\}$.

6. The mappings of geometric subspaces:

a) $\mathcal{V}^{\odot}(\tilde{\Sigma}_{\mathrm{d}}) = \mathcal{S}^{+}(\tilde{\Sigma}_{\mathrm{c}})$.

b) $\mathcal{S}^{\odot}(\tilde{\Sigma}_{\mathrm{d}}) = \mathcal{V}^{+}(\tilde{\Sigma}_{\mathrm{c}})$. ◊

Lastly, we summarise in a graphical form in Figure 3.3 the detailed mappings associated with the bilinear and inverse bilinear transformations.

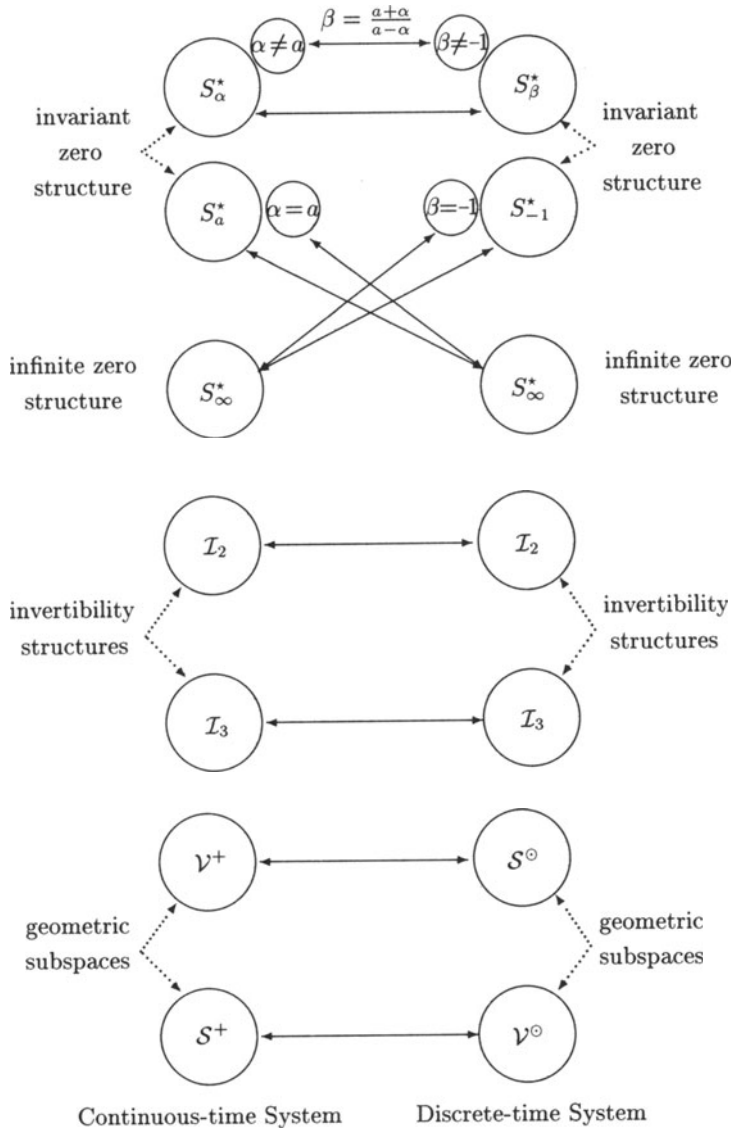


Figure 3.3. Structural mappings of bilinear transformations.

CHAPTER 4
LINEAR CONTROL TECHNIQUES

4.1 Introduction

We now present some common linear control system design techniques, such as the well-known PID control, H_2 and H_∞ optimal control, linear quadratic regulator (LQR) with loop transfer recovery design (LTR), together with some newly developed design techniques, such as the robust and perfect tracking (RPT) method. We will first introduce the precise problem definitions of these techniques and then provide detailed solutions explicitly constructed by closely examining the structural properties of the given systems. Most of these results will be intensively used later in the design of HDD servo systems, though some are presented here for the purpose of easy reference for general readers.

We have noticed that it is some kind of tradition or fashion in the HDD servo system research community in which researchers and practising engineers prefer to carry out a control system design in the discrete-time setting. In this case, the designer would have to discretise the plant to be controlled (mostly using the ZOH technique) first and then use some discrete-time control system design technique to obtain a discrete-time control law. However, in our personal opinion, it is easier to design a controller directly in the continuous-time setting and then use some continuous-to-discrete transformations, such as the bilinear transformation as given in Chapter 3, to discretise it when it is to be implemented in the real system. The advantage of such an approach follows from the following fact: the bilinear transformation does not introduce unstable invariant zeros to its discrete-time counterpart. On the other hand, it is well-known in the literature that the ZOH approach almost always produce some additional nonminimum phase invariant zeros for higher-order systems with faster sampling rates. These nonminimum phase zeros will cause some additional limitations on the overall performance of the system to be controlled. Nevertheless, we will present both continuous-time and discrete-time versions of these control techniques for completeness. It is up to the reader to choose the appropriate approach in designing their own servo system.

Lastly, we would like to note that the design methods presented in this chapter are well studied in the literature. As such, all results are quoted without detailed proofs and derivations. Interested readers are referred to the related references for details.

4.2 PID Control

PID control is the most popular technique used in industries because it is relatively easy and simple to design and implement. Most importantly, it works in most practical situations, although its performance is somewhat limited owing to its restricted structure. Nevertheless, in what follows, we recall this well-known classical control system design methodology for ease of reference.

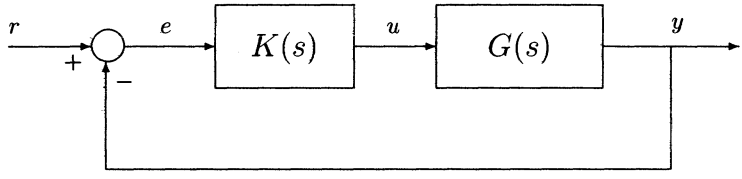


Figure 4.1. The typical PID control configuration.

To be more specific, we consider the control system as depicted in Figure 4.1, in which $G(s)$ is the plant to be controlled and $K(s)$ is the PID controller characterised by the following transfer function

$$K(s) = K_{\mathrm{p}}\left(1 + \frac{1}{T_{\mathrm{i}}s} + T_{\mathrm{d}}s\right). \tag{4.1}$$

The control system design is then to determine the parameters K_{p}, T_{i} and T_{d} such that the resulting closed-loop system yields a certain desired performance, *i.e.*, it meets certain prescribed design specifications.

Ziegler–Nichols tuning is one of the most common techniques used in practical situations to design an appropriate PID controller for the class of systems that can be exactly modelled as, or approximated by, the following first-order system:

$$G(s) = \frac{Y(s)}{U(s)} = \frac{K}{\tau s + 1}\mathrm{e}^{-t_{\mathrm{d}}s}. \tag{4.2}$$

One of the methods proposed by Ziegler and Nichols ([63, 64]) is first to replace the controller $K(s)$ in Figure 4.1 by a simple proportional gain. We then increase this proportional gain to a value, say K_u, for which we observe continuous oscillations in its step response, *i.e.*, the system becomes marginally stable. Assume that the corresponding oscillating frequency is ω_u. The PID controller parameters are then given as follows:

$$K_p = \frac{3K_u}{5}, \quad T_i = \frac{\pi}{\omega_u}, \quad T_d = \frac{\pi}{4\omega_u}. \tag{4.3}$$

Experience has shown that such controller settings provide a good closed-loop response for many systems. Unfortunately, it will be seen shortly in the coming chapters that the typical model of a VCM actuator is actually a double integrator and thus Ziegler–Nichols tuning cannot be directly applied to design a servo system for the VCM actuator.

Another common way to design a PID controller is the pole assignment method, in which the parameters K_p, T_i and T_d are chosen such that the dominant roots of the closed-loop characteristic equation, *i.e.*,

$$1 + K(s)G(s) = 0, \tag{4.4}$$

are assigned to meet certain desired specifications (such as overshoot, rise time, settling time, *etc.*), while its remaining roots are placed far away to the left on the complex plane (roughly three to four times faster compared with the dominant roots). The detailed procedure of this method can be found in most of classical control engineering texts (see *e.g.*, [65]). For the PID control of discrete-time systems, interested readers are referred to [1] for more information.

4.3 H_2 Optimal Control

Most of the feedback design tools provided by the classical Nyquist–Bode frequency domain theory are restricted to single-feedback-loop designs. Modern multivariable control theory based on state-space concepts has the capability to deal with multi feedback-loop designs, and as such has emerged as an alternative to the classical Nyquist–Bode theory. Although it does have shortcomings of its own, a great asset of modern control theory utilising the state-space description of systems is that the design methods derived from it are easily amenable to computer implementation. Owing to this, rapid progress has been made during the last two or three decades in developing a number of multivariable analysis and design tools using the state-space

description of systems. One of the foremost and most powerful design tools developed in this connection is based on what is called linear quadratic Gaussian (LQG) control theory. Here, given a linear model of the plant in a state-space description, and assuming that the disturbance and measurement noise are Gaussian stochastic processes with known power spectral densities, the designer translates the design specifications into a quadratic performance criterion consisting of some state variables and control signal inputs. The object of design then is to minimise the performance criterion by using appropriate state or measurement feedback controllers while guaranteeing the closed-loop stability. A ubiquitous architecture for a measurement feedback controller has been observer-based, wherein a state feedback control law is implemented by utilising an estimate of the state. Thus, the design of a measurement feedback controller here is worked out in two stages. In the first stage, an optimal internally stabilising static state feedback controller is designed, and in the second stage a state estimator is designed. The estimator, otherwise called an observer or filter, is traditionally designed to yield the least mean square error estimate of the state of the plant, utilising only the measured output, which is often assumed to be corrupted by an additive white Gaussian noise. The LQG control problem as described above is posed in a stochastic setting. The same can be posed in a deterministic setting, known as an H_2 optimal control problem, in which the H_2 norm of a certain transfer function from an exogenous disturbance to a pertinent controlled output of a given plant is minimised by appropriate use of an internally stabilising controller.

Much research effort has been expended in the area of H_2 optimal control or optimal control in general during the last few decades (see *e.g.*, Anderson and Moore [66], Fleming and Rishel [67], Kwakernaak and Sivan [68], and Saberi, Sannuti and Chen [69], and references cited therein). In what follows, we focus mainly on the formulation and solution to both continuous- and discrete-time H_2 optimal control problems. Interested readers are referred to [69] for more detailed treatments of such problems.

4.3.1 Continuous-time Systems

We consider a generalised system Σ with a state-space description,

$$\Sigma : \begin{cases} \dot{x} = A\ x + B\ u + E\ w, \\ y = C_1\ x + D_{11}\ u + D_1\ w, \\ h = C_2\ x + D_2\ u + D_{22}\ w, \end{cases} \tag{4.5}$$

where $x \in \mathbb{R}^n$ is the state, $u \in \mathbb{R}^m$ is the control input, $w \in \mathbb{R}^q$ is the external disturbance input, $y \in \mathbb{R}^p$ is the measurement output, and $h \in \mathbb{R}^\ell$ is the controlled output of Σ. For the sake of simplicity in future development,

throughout this chapter, we let $\Sigma_{\rm P}$ be the subsystem characterised by the matrix quadruple (A, B, C_2, D_2) and $\Sigma_{\rm Q}$ be the subsystem characterised by (A, E, C_1, D_1). Throughout this section, we assume that (A, B) is stabilisable and (A, C_1) is detectable.

Generally, we can assume that matrix D_{11} in Equation (4.5) is zero. This can be justified as follows: If $D_{11} \neq 0$, we define a new measurement output

$$y_{\rm new} = y - D_{11}u = C_1 x + D_1 w, \tag{4.6}$$

which does not have a direct feedthrough term from u. Suppose we carry on our control system design using this new measurement output to obtain a proper control law, say, $u = K(s)y_{\rm new}$. Then, it is straightforward to verify that this control law is equivalent to the following one

$$u = [I + K(s)D_{11}]^{-1}K(s)y, \tag{4.7}$$

provided that $[I + K(s)D_{11}]^{-1}$ is well posed, *i.e.*, the inverse exists for almost all $s \in \mathbb{C}$. Thus, for simplicity, we will assume that $D_{11} = 0$.

The standard H_2 optimal control problem is to find an internally stabilising proper measurement feedback control law,

$$\Sigma_{\rm cmp} \; : \; \begin{cases} \dot{v} = A_{\rm cmp}\, v + B_{\rm cmp}\, y, \\ u = C_{\rm cmp}\, v + D_{\rm cmp}\, y, \end{cases} \tag{4.8}$$

such that the H_2-norm of the overall closed-loop transfer matrix function from w to h is minimised (see also Figure 4.2). To be more specific, we will say that the control law $\Sigma_{\rm cmp}$ of Equation (4.8) is internally stabilising when applied to the system Σ of Equation (4.5), if the following matrix is asymptotically stable:

$$A_{\rm cl} := \begin{bmatrix} A + BD_{\rm cmp}C_1 & BC_{\rm cmp} \\ B_{\rm cmp}C_1 & A_{\rm cmp} \end{bmatrix}, \tag{4.9}$$

i.e., all its eigenvalues lie in the open left-half complex plane. It is straightforward to verify that the closed-loop transfer matrix from the disturbance w to the controlled output h is given by

$$T_{hw}(s) = C_e(sI - A_e)^{-1}B_e + D_e, \tag{4.10}$$

where

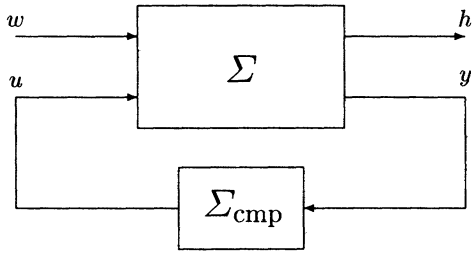


Figure 4.2. The typical control configuration in state-space setting.

$$A_e := \begin{bmatrix} A + BD_{\text{cmp}}C_1 & BC_{\text{cmp}} \\ B_{\text{cmp}}C_1 & A_{\text{cmp}} \end{bmatrix}, \tag{4.11}$$

$$B_e := \begin{bmatrix} E + BD_{\text{cmp}}D_1 \\ B_{\text{cmp}}D_1 \end{bmatrix}, \tag{4.12}$$

$$C_e := [\, C_2 + D_2D_{\text{cmp}}C_1 \quad D_2C_{\text{cmp}} \,], \tag{4.13}$$

$$D_e := D_2D_{\text{cmp}}D_1 + D_{22}. \tag{4.14}$$

It is simple to note that if Σ_{cmp} is a static state feedback law, *i.e.*, $u = Fx$, then the closed-loop transfer matrix from w to h is given by

$$T_{hw}(s) = (C_2 + D_2F)(sI - A - BF)^{-1}E + D_{22}. \tag{4.15}$$

The H_2-norm of a stable continuous-time transfer matrix, *e.g.*, $T_{hw}(s)$, is defined as follows:

$$\|T_{hw}\|_2 := \left(\frac{1}{2\pi}\text{trace}\left[\int_{-\infty}^{\infty} T_{hw}(j\omega)T_{hw}(j\omega)^{\text{H}}\,\text{d}\omega\right]\right)^{1/2}. \tag{4.16}$$

By Parseval's theorem, $\|T_{hw}\|_2$ can equivalently be defined as

$$\|T_{hw}\|_2 = \left(\text{trace}\left[\int_0^{\infty} g(t)g(t)'\,\text{d}t\right]\right)^{1/2}, \tag{4.17}$$

where $g(t)$ is the unit impulse response of $T_{hw}(s)$. Thus, $\|T_{hw}\|_2 = \|g\|_2$.

The H_2 optimal control is to design a proper controller Σ_{cmp} such that, when it is applied to the plant Σ, the resulting closed-loop is asymptotically stable and the H_2-norm of $T_{hw}(s)$ is minimised. For future use, we define

$$\gamma_2^* := \inf\left\{\, \|T_{hw}(\Sigma \times \Sigma_{\text{cmp}})\|_2 \mid \Sigma_{\text{cmp}} \text{ internally stabilises } \Sigma \,\right\}. \tag{4.18}$$

Furthermore, a control law Σ_{cmp} is said to be an H_2 optimal controller for Σ of Equation (4.5) if its resulting closed-loop transfer function from w to h has an H_2-norm equal to γ_2^*, *i.e.*, $||T_{hw}||_2 = \gamma_2^*$.

It is clear to see from the definition of the H_2-norm that, in order to have a finite $||T_{hw}||_2$, the following must be satisfied:

$$D_e = D_2 D_{\text{cmp}} D_1 + D_{22} = 0, \tag{4.19}$$

which is equivalent to the existence of a static measurement pre-feedback law $u = Sy + n_{\text{new}}$ to the system in Equation (4.5) such that

$$D_2 S D_1 + D_{22} = 0. \tag{4.20}$$

We note that the minimisation of $||T_{hw}||_2$ is meaningful only when it is finite. As such, it is without loss of any generality to assume that the feed-forward matrix $D_{22} = 0$ hereafter in this section. In fact, in this case, $||T_{hw}||_2$ can be easily obtained. Solving either one of the following Lyapunov equations:

$$A_e' P_e + P_e A_e + C_e' C_e = 0, \quad A_e Q_e + Q_e A_e' + B_e B_e' = 0, \tag{4.21}$$

for P_e or Q_e, then the H_2-norm of $T_{hw}(s)$ can be computed by

$$||T_{hw}||_2 = \text{trace}\Big[B_e' P_e B_e\Big] = \text{trace}\Big[C_e Q_e C_e'\Big]. \tag{4.22}$$

In what follows, we present solutions to the problem without detailed proofs. We start first with the simplest case, when the given system Σ satisfies the following assumptions of the so-called *regular case*:

1. Σ_{P} has no invariant zeros on the imaginary axis and D_2 is of maximal column rank.
2. Σ_{Q} has no invariant zeros on the imaginary axis and D_1 is of maximal row rank.

The problem is called the *singular case* if Σ does not satisfy these conditions.

The solution to the regular case of the H_2 optimal control problem is very simple. The optimal controller is given by (see, *e.g.*, [70]),

$$\Sigma_{\text{cmp}} : \begin{cases} \dot{v} = (A + BF + KC_1)\, v + K\, y, \\ u = \qquad -F \qquad\quad v \end{cases} \tag{4.23}$$

where

$$F = -(D_2'D_2)^{-1}(D_2'C_2 + B'P), \tag{4.24}$$

$$K = -(QC_1' + ED_1')(D_1D_1')^{-1}, \tag{4.25}$$

and where $P = P' \geq 0$ and $Q = Q' \geq 0$ are respectively the stabilising solutions of the following Riccati equations:

$$A'P+PA+C_2'C_2-(PB+C_2'D_2)(D_2'D_2)^{-1}(D_2'C_2+B'P) = 0 \tag{4.26}$$

$$QA'+AQ+EE'-(QC_1'+ED_1')(D_1D_1')^{-1}(D_1E'+C_1Q) = 0. \tag{4.27}$$

Moreover, the optimal value γ_2^* can be computed as follows:

$$\gamma_2^* = \left\{\text{trace}\left[E'PE\right] + \text{trace}\left[\left(A'P + PA + C_2'C_2\right)Q\right]\right\}^{1/2}. \tag{4.28}$$

We note that if all the states of Σ are available for feedback, then the optimal controller is reduced to a static law $u = Fx$ with F being given as in Equation (4.24).

Next, we present two methods that solve the singular H_2 optimal control problem. As a matter of fact, in the singular case, it is in general infeasible to obtain an optimal controller, although it is possible under certain restricted conditions (see *e.g.*, [69, 71]). The solutions to the singular case are generally suboptimal, and usually parameterised by a certain tuning parameter, say ε. A controller parameterised by ε is said to be suboptimal if there exists an $\varepsilon^* > 0$ such that for all $\varepsilon \leq \varepsilon^*$ the closed-loop system comprising the given plant and the controller is asymptotically stable, and the resulting closed-loop transfer function from w to h, which is obviously a function ε, has an H_2-norm arbitrarily close to γ_2^* as ε approaches to 0.

The following is a so-called *perturbation approach* (see *e.g.*, [72]) that would yield a suboptimal controller for the general singular case. We should note that such an approach is numerically unstable. The problem becomes very serious when the given system is ill-conditioned or has multiple time scales. In principle, the desired solution can be obtained by introducing some small perturbations to the matrices E, D_1, C_2 and D_2, *i.e.*,

$$\tilde{E} := [E \quad \varepsilon I \quad 0], \quad \tilde{D}_1 := [D_1 \quad 0 \quad \varepsilon I], \tag{4.29}$$

and

$$\tilde{C}_2 := \begin{bmatrix} C_2 \\ \varepsilon I \\ 0 \end{bmatrix}, \quad \tilde{D}_2 := \begin{bmatrix} D_2 \\ 0 \\ \varepsilon I \end{bmatrix}. \tag{4.30}$$

A full order H_2 suboptimal output feedback controller is given by

$$\tilde{\Sigma}_{\text{cmp}}(\varepsilon) \; : \; \begin{cases} \dot{v} = A + B\tilde{F} + \tilde{K}C_1 \; v + \tilde{K} \; y, \\ u = \qquad -\tilde{F} \qquad\quad v \end{cases} \tag{4.31}$$

where

$$\tilde{F} = -(\tilde{D}_2'\tilde{D}_2)^{-1}(\tilde{D}_2'\tilde{C}_2 + B'\tilde{P}), \tag{4.32}$$

$$\tilde{K} = -(\tilde{Q}C_1' + \tilde{E}\tilde{D}_1')(\tilde{D}_1\tilde{D}_1')^{-1}, \tag{4.33}$$

and where $\tilde{P} = \tilde{P}' > 0$ and $\tilde{Q} = \tilde{Q}' > 0$ are respectively the solutions of the following Riccati equations:

$$A'\tilde{P}+\tilde{P}A+\tilde{C}_2'\tilde{C}_2-(\tilde{P}B+\tilde{C}_2'\tilde{D}_2)(\tilde{D}_2'\tilde{D}_2)^{-1}(\tilde{D}_2'\tilde{C}_2+B'\tilde{P}) = 0 \tag{4.34}$$

$$\tilde{Q}A'+A\tilde{Q}+\tilde{E}\tilde{E}'-(\tilde{Q}C_1'+\tilde{E}\tilde{D}_1')(\tilde{D}_1\tilde{D}_1')^{-1}(\tilde{D}_1\tilde{E}'+C_1\tilde{Q}) = 0. \tag{4.35}$$

Alternatively, one could solve the singular case by using numerically stable algorithms (see *e.g.*, [69]) that are based on a careful examination of the structural properties of the given system. We separate the problem into three distinct situations: (1) the state feedback case, (2) the full order measurement feedback case, and (3) the reduced order measurement feedback case. The software realisation of these algorithms in MATLAB can be found in [34]. For simplicity, we assume throughout the rest of this subsection that both subsystems Σ_{P} and Σ_{Q} *have no invariant zeros on the imaginary axis.* We believe that such a condition is always satisfied for most HDD servo systems. However, most servo systems can be represented as certain chains of integrators and thus could not be formulated as a regular problem without adding dummy terms. Nevertheless, interested readers are referred to the monograph in [69] for the complete treatment of H_2 optimal control using the approach given below.

i. State Feedback Case. For the case when $y = x$ in the given system Σ of Equation (4.5), *i.e.*, all the state variables of Σ are available for feedback, we have the following step-by-step algorithm that constructs an H_2 suboptimal static feedback control law $u = F(\varepsilon)x$ for Σ.

STEP 4.3.C.S.1: transform the system Σ_{P} into the special coordinate basis as given by Theorem 3.4.1. To all submatrices and transformations in the special coordinate basis of Σ_{P}, we append the subscript P to signify their relation to the system Σ_{P}. We also choose the output transformation Γ_{OP} to have the following form:

$$\Gamma_{0P} = \begin{bmatrix} I_{m_{0P}} & 0 \\ 0 & \Gamma_{orP} \end{bmatrix}, \tag{4.36}$$

where $m_{0P} = \text{rank}\,(D_2)$. Next, define

$$A_{11P} = \begin{bmatrix} A^+_{aaP} & L^+_{abP}C_{bP} \\ 0 & A_{bbP} \end{bmatrix}, \quad B_{11P} = \begin{bmatrix} B^+_{0aP} \\ B_{0bP} \end{bmatrix}, \tag{4.37}$$

$$A_{13P} = \begin{bmatrix} L^+_{adP} \\ L_{bdP} \end{bmatrix}, \tag{4.38}$$

$$C_{21P} = \Gamma_{orP} \begin{bmatrix} 0 & 0 \\ 0 & C_{bP} \end{bmatrix}, \quad C_{23P} = \Gamma_{orP} \begin{bmatrix} C_{dP}C'_{dP} \\ 0 \end{bmatrix}, \tag{4.39}$$

$$A_{xP} = A_{11P} - A_{13P}(C'_{23P}C_{23P})^{-1}C'_{23P}C_{21P}, \tag{4.40}$$

$$B_{xP}B'_{xP} = B_{11P}B'_{11P} + A_{13P}(C'_{23P}C_{23P})^{-1}A'_{13P}, \tag{4.41}$$

$$C'_{xP}C_{xP} = C'_{21P}C_{21P} - C'_{21P}C_{21P}(C'_{23P}C_{23P})^{-1}C'_{23P}C_{21P}. \tag{4.42}$$

STEP 4.3.C.S.2: solve the following algebraic matrix Riccati equation:

$$P_xA_{xP} + A'_{xP}P_x - P_xB_{xP}B'_{xP}P_x + C'_{xP}C_{xP} = 0 \tag{4.43}$$

for $P_x > 0$ and define

$$F_{11} := \begin{bmatrix} F^+_{a0} & F_{b0} \\ F^+_{a1} & F_{b1} \end{bmatrix} = \begin{bmatrix} B'_{11P}P_x \\ (C'_{23P}C_{23P})^{-1}(A'_{13P}P_x + C'_{23P}C_{21P}) \end{bmatrix}. \tag{4.44}$$

Then, partition $[\,F^+_{a1} \quad F_{b1}\,]$ as

$$[\,F^+_{a1} \quad F_{b1}\,] = \begin{bmatrix} F^+_{a11} & F_{b11} \\ F^+_{a12} & F_{b12} \\ \vdots & \vdots \\ F^+_{a1m_{dP}} & F_{b1m_{dP}} \end{bmatrix}, \tag{4.45}$$

where F^+_{a1i} and F_{b1i} are of dimensions $1 \times n^+_{aP}$ and $1 \times n_{bP}$, respectively.

STEP 4.3.C.S.3: let Δ_{cP} be any arbitrary $m_{cP} \times n_{cP}$ matrix subject to the constraint that

$$A^c_{ccP} = A_{ccP} - B_{cP}\Delta_{cP}, \tag{4.46}$$

is a stable matrix. Note that the existence of such a Δ_{cP} is guaranteed by the property that (A_{ccP}, B_{cP}) is controllable.

STEP 4.3.C.S.4: this step makes use of subsystems, $i = 1$ to m_{dP}, represented by Equation (3.23). Let $\Lambda_i = \{ \lambda_{i1}, \lambda_{i2}, \cdots, \lambda_{iq_i} \}$, $i = 1$ to m_{dP}, be the sets of q_i elements all in $\mathbb{C}^-$, which are closed under complex conjugation, where q_i and m_{dP} are as defined in Theorem 3.4.1 but associated with the special coordinate basis of Σ_P. Let $\Lambda_{dP} := \Lambda_1 \cup \Lambda_2 \cup \cdots \cup \Lambda_{m_{dP}}$. For $i = 1$ to m_{dP}, we define

$$p_i(s) := \prod_{j=1}^{q_i} (s - \lambda_{ij}) = s^{q_i} + F_{i1} s^{q_i - 1} + \cdots + F_{iq_i - 1} s + F_{iq_i} \quad (4.47)$$

and

$$\tilde{F}_i(\varepsilon) := \frac{1}{\varepsilon^{q_i}} \left[F_{iq_i}, \ \varepsilon F_{iq_i - 1}, \ \cdots, \ \varepsilon^{q_i - 1} F_{i1} \right]. \quad (4.48)$$

STEP 4.3.C.S.5: in this step, various gains calculated in Steps 4.3.C.S.2 to 4.3.C.S.4 are put together to form a composite state feedback gain for the given system Σ_P. Let

$$\tilde{F}_{a1}^{+}(\varepsilon) := \begin{bmatrix} F_{a11}^{+} F_{1q_1} / \varepsilon^{q_1} \\ F_{a12}^{+} F_{2q_2} / \varepsilon^{q_2} \\ \vdots \\ F_{a1m_{dP}}^{+} F_{m_{dP} q_{m_{dP}}} / \varepsilon^{q_{m_{dP}}} \end{bmatrix}, \quad (4.49)$$

and

$$\tilde{F}_{b1}(\varepsilon) := \begin{bmatrix} F_{b11} F_{1q_1} / \varepsilon^{q_1} \\ F_{b12} F_{2q_2} / \varepsilon^{q_2} \\ \vdots \\ F_{b1m_{dP}} F_{m_{dP} q_{m_{dP}}} / \varepsilon^{q_{m_{dP}}} \end{bmatrix}. \quad (4.50)$$

Then, the H_2 suboptimal state feedback gain is given by

$$F(\varepsilon) = -\Gamma_{iP} \left(\tilde{F}(\varepsilon) + \tilde{F}_0 \right) \Gamma_{sP}^{-1}, \quad (4.51)$$

where

$$\tilde{F}(\varepsilon) = \begin{bmatrix} 0 & F_{a0}^{+} & F_{b0} & 0 & 0 \\ 0 & \tilde{F}_{a1}^{+}(\varepsilon) & \tilde{F}_{b1}(\varepsilon) & 0 & \tilde{F}_d(\varepsilon) \\ 0 & 0 & 0 & \Delta_{cP} & 0 \end{bmatrix}, \quad (4.52)$$

$$\tilde{F}_0 = \begin{bmatrix} C^-_{0a\mathrm{P}} & C^+_{0a\mathrm{P}} & C_{0b\mathrm{P}} & C_{0c\mathrm{P}} & C_{0d\mathrm{P}} \\ E^-_{da\mathrm{P}} & E^+_{da\mathrm{P}} & E_{db\mathrm{P}} & E_{dc\mathrm{P}} & E_{d\mathrm{P}} \\ E^-_{ca\mathrm{P}} & E^+_{ca\mathrm{P}} & E_{cb\mathrm{P}} & 0 & 0 \end{bmatrix}, \tag{4.53}$$

and where

$$E_{d\mathrm{P}} = \begin{bmatrix} E_{11} & \cdots & E_{1m_{d\mathrm{P}}} \\ \vdots & \ddots & \vdots \\ E_{m_{d\mathrm{P}}1} & \cdots & E_{m_{d\mathrm{P}}m_{d\mathrm{P}}} \end{bmatrix}, \tag{4.54}$$

$$\tilde{F}_d(\varepsilon) = \mathrm{diag}\Big[\tilde{F}_1(\varepsilon),\ \tilde{F}_2(\varepsilon),\ \cdots,\ \tilde{F}_{m_{d\mathrm{P}}}(\varepsilon)\Big]. \tag{4.55}$$

This completes the algorithm. ◊

Theorem 4.3.1. Consider the given system in Equation (4.5) with $D_{22} = 0$ and $y = x$, *i.e.*, all states are measurable. Assume that Σ_{P} has no invariant zeros on the imaginary axis. Then, the closed-loop system comprising that of Equation (4.5) and $u = F(\varepsilon)x$ with $F(\varepsilon)$ being given as in Equation (4.51) has the following properties:

1. it is internally stable for sufficiently small ε;
2. the closed-loop transfer matrix from the disturbance w to the controlled output h, $T_{zh}(s,\varepsilon)$, possesses $\|T_{zh}(s,\varepsilon)\|_2 \to \gamma_2^*$ as $\varepsilon \to 0$.

Clearly, $u = F(\varepsilon)x$ is an H_2 suboptimal controller for the system in Equation (4.5). ◊

ii. Full Order Output Feedback Case. The following is a step-by-step algorithm for constructing a parameterised full order output feedback controller that solves the general H_2 optimisation problem.

STEP 4.3.C.F.1: (construction of the gain matrix $F_{\mathrm{P}}(\varepsilon)$). Define an auxiliary system

$$\begin{cases} \dot{x} = A\ x + B\ u + E\ w \\ y = \quad x \\ h = C_2\ x + D_2\ u \end{cases} \tag{4.56}$$

and then perform Steps 4.3.C.S.1 to 4.3.C.S.5 of the previous algorithm on the above system to obtain a parameterised gain matrix $F(\varepsilon)$. We let $F_{\mathrm{P}}(\varepsilon) = F(\varepsilon)$.

STEP 4.3.C.F.2: (construction of the gain matrix $K_{\mathrm{Q}}(\varepsilon)$). Define another auxiliary system

$$\begin{cases} \dot{x} = A'\,x + C_1'\,u + C_2'\,w \\ y = \quad x \\ h = E'\,x + D_1'\,u \end{cases} \tag{4.57}$$

and then perform Steps 4.3.C.S.1 to 4.3.C.S.6 on the above system to get the parameterised gain matrix $F(\varepsilon)$. We let $K_{\mathrm{Q}}(\varepsilon) = F(\varepsilon)'$.

STEP 4.3.C.F.3: (construction of the full order controller Σ_{FC}). Finally, the parameterised full order output feedback controller is given by

$$\Sigma_{\mathrm{FC}} \; : \; \begin{cases} \dot{v} = A_{\mathrm{FC}}(\varepsilon)\,v + B_{\mathrm{FC}}(\varepsilon)\,y, \\ u = C_{\mathrm{FC}}(\varepsilon)\,v + D_{\mathrm{FC}}(\varepsilon)\,y, \end{cases} \tag{4.58}$$

where

$$\left.\begin{aligned} A_{\mathrm{FC}}(\varepsilon) &:= A + BF_{\mathrm{P}}(\varepsilon) + K_{\mathrm{Q}}(\varepsilon)C_1, \\ B_{\mathrm{FC}}(\varepsilon) &:= -K_{\mathrm{Q}}(\varepsilon), \\ C_{\mathrm{FC}}(\varepsilon) &:= F_{\mathrm{P}}(\varepsilon), \\ D_{\mathrm{FC}}(\varepsilon) &:= 0. \end{aligned}\right\} \tag{4.59}$$

This concludes the algorithm for constructing the full order measurement feedback controller. ◊

Theorem 4.3.2. Consider the given system in Equation (4.5) with $D_{22} = 0$. Assume that Σ_{P} and Σ_{Q} have no invariant zeros on the imaginary axis. Then the closed-loop system comprising Equation (4.5) and the full order output feedback controller of Equation (4.58) has the following properties:

1. it is internally stable for sufficiently small ε;
2. the closed-loop transfer matrix from the disturbance w to the controlled output h, $T_{zh}(s,\varepsilon)$, possesses $\|T_{zh}(s,\varepsilon)\|_2 \to \gamma_2^*$ as $\varepsilon \to 0$.

By definition, Equation (4.58) is an H_2 suboptimal controller for the system in Equation (4.5). ◊

iii. Reduced Order Output Feedback Case. For the case when some measurement output channels are clean, *i.e.*, they are not mixed with disturbances, then we could design an output feedback control law that has a dynamical order less than that of the given plant and yet has an identical

performance compared with that of full order control law. Such a control law is called the reduced order output feedback controller. We note that the construction of a reduced order controller was first reported by Chen *et al.* [73] for general linear systems, in which the direct feedthrough matrix from input is nonzero. It was shown in [73] that the reduced order output feedback controller has the following advantages over the full order counterpart:

1. the dynamical order of the reduced order controller is generally smaller than that of the full order counterpart;
2. the gain required for the same degree of performance for the reduced order controller is smaller compared with that of the full order counterpart.

We now proceed to design a reduced order controller, which solves the general H_2 suboptimal problem. First, without loss of generality and for simplicity of presentation, we assume that the matrices C_1 and D_1 are already in the form

$$C_1 = \begin{bmatrix} 0 & C_{1,02} \\ I_k & 0 \end{bmatrix} \quad \text{and} \quad D_1 = \begin{bmatrix} D_{1,0} \\ 0 \end{bmatrix}, \tag{4.60}$$

where $k = \ell - \text{rank}(D_1)$ and $D_{1,0}$ is of full rank. Then the given system in Equation (4.5) can be written as

$$\begin{cases} \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} u + \begin{bmatrix} E_1 \\ E_2 \end{bmatrix} w, \\ \begin{pmatrix} y_0 \\ y_1 \end{pmatrix} = \begin{bmatrix} 0 & C_{1,02} \\ I_k & 0 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{bmatrix} D_{1,0} \\ 0 \end{bmatrix} w, \\ h = [C_{2,1} \quad C_{2,2}] \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + D_2 u + D_{22} w, \end{cases} \tag{4.61}$$

where the original state x is partitioned into two parts, x_1 and x_2; and y is partitioned into y_0 and y_1 with $y_1 \equiv x_1$. Thus, one needs to estimate only the state x_2 in the reduced order controller design. Next, define an auxiliary subsystem Σ_{QR} characterised by a matrix quadruple $(A_{\text{R}}, E_{\text{R}}, C_{\text{R}}, D_{\text{R}})$, where

$$(A_{\text{R}}, E_{\text{R}}, C_{\text{R}}, D_{\text{R}}) = \left(A_{22}, E_2, \begin{bmatrix} C_{1,02} \\ A_{12} \end{bmatrix}, \begin{bmatrix} D_{1,0} \\ E_1 \end{bmatrix} \right). \tag{4.62}$$

The following is a step-by-step algorithm that constructs the reduced order output feedback controller for the general H_2 optimisation.

STEP 4.3.C.R.1: (construction of the gain matrix $F_{\mathrm{P}}(\varepsilon)$). Define an auxiliary system

$$\begin{cases} \dot{x} = A\ x + B\ u + E\ w \\ y = \quad x \\ h = C_2\ x + D_2\ u \end{cases} \tag{4.63}$$

and then perform Steps 4.3.C.R.1 to 4.3.C.R.5 on the above system to get the parameterised gain matrix $F(\varepsilon)$. We let $F_{\mathrm{P}}(\varepsilon) = F(\varepsilon)$.

STEP 4.3.C.R.2: (construction of the gain matrix $K_{\mathrm{R}}(\varepsilon)$). Define another auxiliary system

$$\begin{cases} \dot{x} = A'_{\mathrm{R}}\ x + C'_{\mathrm{R}}\ u + C'_{2,2}\ w \\ y = \quad x \\ h = E'_{\mathrm{R}}\ x + D'_{\mathrm{R}}\ u \end{cases} \tag{4.64}$$

and then perform Steps 4.3.C.R.1 to 4.3.C.R.6 on the above system to get the parameterised gain matrix $F(\varepsilon)$. We let $K_{\mathrm{R}}(\varepsilon) = F(\varepsilon)'$.

STEP 4.3.C.R.3: (construction of the reduced order controller Σ_{RC}). Let us partition $F_{\mathrm{P}}(\varepsilon)$ and $K_{\mathrm{R}}(\varepsilon)$ as

$$F_{\mathrm{P}}(\varepsilon) = [\,F_{\mathrm{P}1}(\varepsilon) \quad F_{\mathrm{P}2}(\varepsilon)\,], \quad K_{\mathrm{R}}(\varepsilon) = [\,K_{\mathrm{R}0}(\varepsilon) \quad K_{\mathrm{R}1}(\varepsilon)\,] \tag{4.65}$$

in conformity with the partitions of $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and $y = \begin{pmatrix} y_0 \\ y_1 \end{pmatrix}$ of Equation (4.61), respectively. Then define

$$G_{\mathrm{R}}(\varepsilon) = [\,-K_{\mathrm{R}0}(\varepsilon),\ A_{21} + K_{\mathrm{R}1}(\varepsilon)A_{11} - (A_{\mathrm{R}} + K_{\mathrm{R}}(\varepsilon)C_{\mathrm{R}})K_{\mathrm{R}1}(\varepsilon)\,].$$

Finally, the reduced order output feedback controller is given by

$$\Sigma_{\mathrm{RC}} : \begin{cases} \dot{v} = A_{\mathrm{RC}}(\varepsilon)\ v + B_{\mathrm{RC}}(\varepsilon)\ y, \\ u = C_{\mathrm{RC}}(\varepsilon)\ v + D_{\mathrm{RC}}(\varepsilon)\ y, \end{cases} \tag{4.66}$$

where

$$\left.\begin{aligned} A_{\mathrm{RC}}(\varepsilon) &:= A_{\mathrm{R}} + B_2F_{\mathrm{P}2}(\varepsilon) + K_{\mathrm{R}}(\varepsilon)C_{\mathrm{R}} + K_{\mathrm{R}1}(\varepsilon)B_1F_{\mathrm{P}2}(\varepsilon) \\ B_{\mathrm{RC}}(\varepsilon) &:= G_{\mathrm{R}}(\varepsilon) + [B_2 + K_{\mathrm{R}1}(\varepsilon)B_1] \\ &\qquad \times [\,0,\ F_{\mathrm{P}1}(\varepsilon) - F_{\mathrm{P}2}(\varepsilon)K_{\mathrm{R}1}(\varepsilon)\,] \\ C_{\mathrm{RC}}(\varepsilon) &:= F_{\mathrm{P}2}(\varepsilon) \\ D_{\mathrm{RC}}(\varepsilon) &:= [\,0,\ F_{\mathrm{P}1}(\varepsilon) - F_{\mathrm{P}2}(\varepsilon)K_{\mathrm{R}1}(\varepsilon)\,]. \end{aligned}\right\} \tag{4.67}$$

This concludes the algorithm for constructing the reduced order output feedback controller. ◇

Theorem 4.3.3. Consider the given system in Equation (4.5) with $D_{22} = 0$. Assume that Σ_{P} and Σ_{Q} have no invariant zeros on the imaginary axis. Then, the closed-loop system comprising Equation (4.5) and the reduced order output feedback controller in Equation (4.66) has the following properties:

1. it is internally stable for sufficiently small ε;
2. the closed-loop transfer matrix from the disturbance w to the controlled output h, $T_{zh}(s,\varepsilon)$, possesses $\|T_{zh}(s,\varepsilon)\|_2 \to \gamma_2^*$ as $\varepsilon \to 0$.

By definition, Equation (4.66) is an H_2 suboptimal controller for the system in Equation (4.5). ◇

4.3.2 Discrete-time Systems

We now consider a generalised discrete-time system Σ characterised by the following state-space equations

$$\Sigma : \begin{cases} x(k+1) = A\ x(k) + B\ u(k) + E\ w(k), \\ y(k) = C_1\ x(k) + D_{11}\ u(k) + D_1\ w(k), \\ h(k) = C_2\ x(k) + D_2\ u(k) + D_{22}\ w(k), \end{cases} \tag{4.68}$$

where $x \in \mathbb{R}^n$ is the state, $u \in \mathbb{R}^m$ is the control input, $w \in \mathbb{R}^q$ is the external disturbance input, $y \in \mathbb{R}^p$ is the measurement output, and $h \in \mathbb{R}^\ell$ is the controlled output of Σ. As usual, we let Σ_{P} be the subsystem characterised by the matrix quadruple (A, B, C_2, D_2) and Σ_{Q} be the subsystem characterised by (A, E, C_1, D_1). Without loss of any generality, we assume that $D_{11} = 0$, (A, B) is stabilisable and (A, C_1) is detectable.

The standard H_2 optimal control problem for a discrete-time system is to find an internally stabilising proper measurement feedback control law,

$$\Sigma_{\mathrm{cmp}} : \begin{cases} v(k+1) = A_{\mathrm{cmp}}\ v(k) + B_{\mathrm{cmp}}\ y(k), \\ u(k) = C_{\mathrm{cmp}}\ v(k) + D_{\mathrm{cmp}}\ y(k), \end{cases} \tag{4.69}$$

such that the H_2-norm of the overall closed-loop transfer matrix function from w to h is minimised. To be more specific, we will say that the control law Σ_{cmp} of Equation (4.69) is internally stabilising when applied to the system Σ of Equation (4.68), if the following matrix is asymptotically stable:

$$A_{\mathrm{cl}} := \begin{bmatrix} A + BD_{\mathrm{cmp}}C_1 & BC_{\mathrm{cmp}} \\ B_{\mathrm{cmp}}C_1 & A_{\mathrm{cmp}} \end{bmatrix}, \tag{4.70}$$

i.e., all its eigenvalues lie inside the open unit disc. The closed-loop transfer matrix from the disturbance w to the controlled output h is given by

$$T_{hw}(z) = C_e(zI - A_e)^{-1}B_e + D_e, \tag{4.71}$$

where

$$A_e := \begin{bmatrix} A + BD_{\mathrm{cmp}}C_1 & BC_{\mathrm{cmp}} \\ B_{\mathrm{cmp}}C_1 & A_{\mathrm{cmp}} \end{bmatrix}, \tag{4.72}$$

$$B_e := \begin{bmatrix} E + BD_{\mathrm{cmp}}D_1 \\ B_{\mathrm{cmp}}D_1 \end{bmatrix}, \tag{4.73}$$

$$C_e := [\, C_2 + D_2D_{\mathrm{cmp}}C_1 \quad D_2C_{\mathrm{cmp}} \,], \tag{4.74}$$

$$D_e := D_2D_{\mathrm{cmp}}D_1 + D_{22}. \tag{4.75}$$

The H_2-norm of a stable discrete-time transfer matrix, *e.g.*, $T_{hw}(z)$, is defined as follows:

$$\|T_{hw}\|_2 := \left(\frac{1}{2\pi} \mathrm{trace} \left[\int_{-\pi}^{\pi} T_{hw}(\mathrm{e}^{j\omega}) T_{hw}(\mathrm{e}^{j\omega})^{\mathrm{H}} \, \mathrm{d}\omega \right] \right)^{1/2}. \tag{4.76}$$

By Parseval's theorem, $\|T_{hw}\|_2$ can equivalently be defined as

$$\|T_{hw}\|_2 = \left(\mathrm{trace} \left[\sum_{k=0}^{\infty} g(k)g(k)' \, \mathrm{d}t \right] \right)^{1/2}, \tag{4.77}$$

where $g(k)$ is the impulse response of $T_{hw}(k)$. Thus, $\|T_{hw}\|_2 = \|g\|_2$.

The H_2 optimal control for the discrete-time system of Equation (4.68) is to design a proper controller Σ_{cmp} such that, when it is applied to the plant Σ, the resulting closed-loop is asymptotically stable and the H_2-norm of $T_{hw}(z)$ is minimised. For future use, we define

$$\gamma_2^* := \inf \left\{ \, \|T_{hw}(\Sigma \times \Sigma_{\mathrm{cmp}})\|_2 \mid \Sigma_{\mathrm{cmp}} \text{ internally stabilises } \Sigma \, \right\}. \tag{4.78}$$

Again, a control law Σ_{cmp} is said to be an H_2 optimal controller for Σ of Equation (4.68) if its resulting closed-loop transfer function from w to h has an H_2-norm equal to γ_2^*, *i.e.*, $\|T_{hw}\|_2 = \gamma_2^*$.

For the case when $D_e = 0$, $\|T_{hw}\|_2$ can be computed by

$$\|T_{hw}\|_2 = \text{trace}\left[B_e' P_e B_e\right] = \text{trace}\left[C_e Q_e C_e'\right], \tag{4.79}$$

where P_e and Q_e are respectively the solutions of the following Lyapunov equations:

$$A_e' P_e A - P_e + C_e' C_e = 0, \quad A_e Q_e A_e' - Q_e + B_e B_e' = 0. \tag{4.80}$$

We now present solutions to the discrete-time H_2 optimal control problem without detailed proofs. As usual, we assume that $D_{22} = 0$ for convenience. We start first with the simplest case when the given system Σ satisfies the following assumptions of the so-called *regular case*:

1. Σ_{P} has no invariant zeros on the imaginary axis and is left invertible;
2. Σ_{Q} has no invariant zeros on the imaginary axis and is right invertible.

The problem is called the *singular case* if Σ does not satisfy these conditions.

Again, the solution to the regular case of the discrete-time H_2 optimal control problem is very simple as well. The optimal controller is given by Σ_{cmp}:

$$\begin{cases} v(k+1) = (A+BF+KC_1-BNC_1)\, v(k) + (BN-K)\, y(k) \\ u(k) = (F-NC_1)\, v(k) + N\, y(k) \end{cases} \tag{4.81}$$

where

$$F = -(B'PB + D_2'D_2)^{-1}(B'PA + D_2'C_2), \tag{4.82}$$

$$K = -(AQC_1' + ED_1')(D_1D_1' + C_1QC_1')^{-1}, \tag{4.83}$$

and

$$N = -(B'PB + D_2'D_2)^{-1}\left[(B'PA + D_2'D_2)QC_1 + B'PED_1'\right] \times (D_1D_1' + C_1QC_1')^{-1}, \tag{4.84}$$

and where $P = P' \geq 0$ and $Q = Q' \geq 0$ are respectively the stabilising solutions of the following Riccati equations:

$$P = A'PA + C_2'C_2 - (C_2'D_2 + A'PB)(D_2'D_2 + B'PB)^{-1} \times (D_2'C_2 + B'PA), \tag{4.85}$$

and

$$Q = AQA' + EE' - (ED_1' + AQC_1')(D_1D_1' + C_1QC_1')^{-1} \times (D_1E' + C_1QA'). \quad (4.86)$$

We note that the above discrete-time Riccati equations can be solved using the noniterative method given in [44]. If all the states of Σ are available for feedback, then the optimal controller is reduced to a static law $u(k) = Fx(k)$ with F being given as in Equation (4.82).

Next, we present solutions to the singular H_2 optimal control problem. Similarly, the solutions to the singular case are generally suboptimal, and usually parameterised by a certain tuning parameter, say ε. Again, a discrete-time controller parameterised by ε is said to be suboptimal for the system in Equation (4.68) if there exists an $\varepsilon^* > 0$ such that for all $\varepsilon \leq \varepsilon^*$ the closed-loop system comprising the given plant and the controller is asymptotically stable, and the resulting closed-loop transfer function from w to h, which is obviously a function ε, has an H_2-norm arbitrarily close to γ_2^* as ε approaches to 0.

The following *perturbation approach* would yield a suboptimal controller for the general discrete-time singular case. Given any $\varepsilon > 0$, define

$$\tilde{E} := [E \quad \varepsilon I \quad 0], \quad \tilde{D}_1 := [D_1 \quad 0 \quad \varepsilon I], \quad (4.87)$$

and

$$\tilde{C}_2 := \begin{bmatrix} C_2 \\ \varepsilon I \\ 0 \end{bmatrix}, \quad \tilde{D}_2 := \begin{bmatrix} D_2 \\ 0 \\ \varepsilon I \end{bmatrix}. \quad (4.88)$$

A full order H_2 suboptimal output feedback controller is given by

$$\begin{cases} v(k+1) = (A+B\tilde{F}+\tilde{K}C_1-B\tilde{N}C_1)\, v(k) + (B\tilde{N}-\tilde{K})\, y(k) \\ u(k) \quad = \qquad (\tilde{F}-\tilde{N}C_1) \qquad v(k) + \quad \tilde{N} \quad y(k) \end{cases} \quad (4.89)$$

where

$$\tilde{F} = -(B'\tilde{P}B + \tilde{D}_2'\tilde{D}_2)^{-1}(B'\tilde{P}A + \tilde{D}_2'\tilde{C}_2), \quad (4.90)$$

$$\tilde{K} = -(A\tilde{Q}C_1' + \tilde{E}\tilde{D}_1')(\tilde{D}_1\tilde{D}_1' + C_1\tilde{Q}C_1')^{-1}, \quad (4.91)$$

and

$$\tilde{N} = -(B'\tilde{P}B + \tilde{D}_2'\tilde{D}_2)^{-1}\left[(B'\tilde{P}A + \tilde{D}_2'\tilde{D}_2)\tilde{Q}C_1 + B'\tilde{P}\tilde{E}\tilde{D}_1'\right] \times (\tilde{D}_1\tilde{D}_1' + C_1\tilde{Q}C_1')^{-1}, \quad (4.92)$$

and where $\tilde{P} = \tilde{P}' > 0$ and $\tilde{Q} = \tilde{Q}' > 0$ are respectively the stabilising solutions of the following Riccati equations:

$$\tilde{P} = A'\tilde{P}A + \tilde{C}_2'\tilde{C}_2 - (\tilde{C}_2'\tilde{D}_2 + A'\tilde{P}B)(\tilde{D}_2'\tilde{D}_2 + B'\tilde{P}B)^{-1} \times (\tilde{D}_2'\tilde{C}_2 + B'\tilde{P}A) \quad (4.93)$$

and

$$\tilde{Q} = A\tilde{Q}A' + \tilde{E}\tilde{E}' - (\tilde{E}\tilde{D}_1' + A\tilde{Q}C_1')(\tilde{D}_1\tilde{D}_1' + C_1\tilde{Q}C_1')^{-1} \times (\tilde{D}_1\tilde{E}' + C_1\tilde{Q}A'). \quad (4.94)$$

The following are alternative methods based on the structural decompositions of the given systems. Similarly, we separate the problem into three distinct situations: (1) the state feedback case, (2) the full order measurement feedback case, and (3) the reduced order measurement feedback case.

Similarly, for convenience, we assume throughout the rest of this subsection that both subsystems Σ_P and Σ_Q *have no invariant zeros on the unit circle.* The complete treatment of H_2 optimal control using the approach given below can be found in [69]. Interestingly, it turns out that for this case, although it is singular, we can always obtain a set of H_2 optimal controllers that need not be parameterised by any tuning scalar.

i. State Feedback Case. For the case when $y = x$ in the given system Σ of Equation (4.68), we have the following step-by-step algorithm that constructs an H_2 suboptimal static feedback control law $u = F(\varepsilon)x$ for Σ.

STEP 4.3.D.S.1: (decomposition of Σ_P). Transform the subsystem Σ_P, *i.e.*, the matrix quadruple (A, B, C_2, D_2), into the special coordinate basis as given by Theorem 3.4.1. Denote the state, output and input transformation matrices as Γ_{SP}, Γ_{OP} and Γ_{IP}, respectively.

STEP 4.3.D.S.2: (gain matrix for the subsystem associated with $\mathcal{X}_c$). Let F_c be any constant matrix subject to the constraint that

$$A_{cc}^c = A_{cc} - B_cF_c \quad (4.95)$$

is a stable matrix. Note that the existence of such an F_c is guaranteed by the property of the special coordinate basis, *i.e.*, (A_{cc}, B_c) is controllable.

STEP 4.3.D.S.3: (gain matrix for the subsystem associated with $\mathcal{X}_a^+$, $\mathcal{X}_b$ and $\mathcal{X}_d$). Let

$$A_x = \begin{bmatrix} A_{aa}^+ & L_{ad}^+ C_d & L_{ab}^+ C_b \\ B_d E_{da}^+ & A_{dd} & B_d E_{db} \\ 0 & L_{bd} C_d & A_{bb} \end{bmatrix}, \quad E_x = \begin{bmatrix} E_a^+ \\ E_d \\ E_b \end{bmatrix}, \tag{4.96}$$

$$B_x = \begin{bmatrix} B_{0a}^+ & 0 \\ B_{0d} & B_d \\ B_{0b} & 0 \end{bmatrix}, \tag{4.97}$$

and

$$C_x = \Gamma_{\text{OP}} \begin{bmatrix} 0 & 0 & 0 \\ 0 & C_d & 0 \\ 0 & 0 & C_b \end{bmatrix}, \quad D_x = \Gamma_{\text{OP}} \begin{bmatrix} I & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}. \tag{4.98}$$

Solve the following discrete-time Riccati equation

$$P_x = A_x' P_x A_x + C_x' C_x - (C_x' D_x + A_x' P_x B_x)(D_x' D_x + B_x' P_x B_x)^{-1} \times (D_x' C_x + B_x' P_x A_x) \tag{4.99}$$

for $P_x > 0$. Then partition

$$\begin{aligned} F_x &= -(B_x' P_x B_x + D_x' D_x)^{-1}(B_x' P_x A_x + D_x' C_x) \\ &= -\begin{bmatrix} F_{a0}^+ & F_{d0} & F_{b0} \\ F_{ad}^+ & F_{dd} & F_{bd} \end{bmatrix}. \end{aligned} \tag{4.100}$$

STEP 4.3.D.S.4: (composition of gain matrix F). In this step, various gains calculated in the previous steps are put together to form a composite state feedback gain matrix F. It is given by

$$F = -\Gamma_{\text{IP}} \begin{bmatrix} C_{0a}^- & C_{0a}^+ + F_{a0}^+ & C_{0b} + F_{b0} & C_{0c} & C_{0d} + F_{d0} \\ E_{da}^- & F_{ad}^+ & F_{bd} & E_{dc} & F_{dd} \\ E_{ca}^- & E_{ca}^+ & 0 & F_c & 0 \end{bmatrix} \Gamma_{\text{SP}}^{-1}. \tag{4.101}$$

This completes the algorithm. ◊

Theorem 4.3.4. Consider the given system in Equation (4.68) with $D_{22} = 0$ and $y = x$, *i.e.*, all states are measurable. Assume that Σ_P has no invariant zeros on the unit circle. Then, the closed-loop system comprising Equation (4.68) and $u(k) = Fx(k)$ with F being given as in Equation (4.101) has the following properties:

1. it is internally stable;
2. the closed-loop transfer matrix from the disturbance w to the controlled output h, $T_{zh}(z)$, possesses $\|T_{zh}\|_2 = \gamma_2^*$.

Thus, $u(k) = Fx(k)$ is an H_2 optimal controller for the system in Equation (4.68). ◇

ii. Full Order Output Feedback Case. The following is a step-by-step algorithm for constructing an H_2 optimal full order output feedback controller.

STEP 4.3.D.F.1: (computation of N). Utilise the properties of the special coordinate basis to compute two constant matrices X and Y such that $\mathcal{V}^{\odot}(\Sigma_P) = \mathrm{Ker}\,(X)$ and $\mathcal{S}^{\odot}(\Sigma_Q) = \mathrm{Im}\,(Y)$. Then, compute

$$N = -(B'X'XB + D_2'D_2)^{\dagger} \begin{bmatrix} B'X' & D_2' \end{bmatrix} \begin{bmatrix} XAY & XE \\ C_2Y & D_{22} \end{bmatrix} \times \begin{bmatrix} Y'C_1' \\ D_1' \end{bmatrix} (C_1YY'C_1' + D_1D_1')^{\dagger}. \quad (4.102)$$

STEP 4.3.D.F.2: (construction of the gain matrix F_P). Define an auxiliary system

$$\begin{cases} x(k+1) = \tilde{A}\, x(k) + B\, u(k) + \tilde{E}\, w(k), \\ y(k) = x(k) \\ h(k) = \tilde{C}_2\, x(k) + D_2\, u(k) + 0\, w(k), \end{cases} \quad (4.103)$$

where

$$\tilde{A} = A + BNC_1,\ \tilde{E} = E + BND_1,\ \tilde{C}_2 = C_2 + D_2NC_1, \quad (4.104)$$

and then perform Steps 4.3.D.S.1 to 4.3.D.S.4 of the previous algorithm on the above system in Equation (4.103) to obtain a gain matrix F. We let $F_P = F$.

STEP 4.3.D.F.3: (construction of the gain matrix K_Q). Define another auxiliary system

$$\begin{cases} x(k+1) = \tilde{A}' x(k) + C_1' u(k) + \tilde{C}_2' w(k), \\ y(k) = x(k) \\ h(k) = \tilde{E}' x(k) + D_1' u(k) + 0 \ w(k), \end{cases} \tag{4.105}$$

and then perform Steps 4.3.D.S.1 to 4.3.D.S.4 of the previous algorithm on the above system to get a gain matrix F. Similarly, we let $K_Q = F'$.

STEP 4.3.D.F.4: (construction of the full order controller Σ_{FC}). Finally, the parameterised full order output feedback controller is given by

$$\Sigma_{FC} : \begin{cases} v(k+1) = A_{FC} v(k) + B_{FC} y(k), \\ u(k) = C_{FC} v(k) + D_{FC} y(k), \end{cases} \tag{4.106}$$

where

$$\left.\begin{aligned} A_{FC} &:= A + BNC_1 + BF_P + K_Q C_1, \\ B_{FC} &:= -K_Q, \\ C_{FC} &:= F_P, \\ D_{FC} &:= N. \end{aligned}\right\} \tag{4.107}$$

This concludes the algorithm for constructing the full order measurement feedback controller. ♢

Theorem 4.3.5. Consider the given system in Equation (4.68) with $D_{22} = 0$. Assume that Σ_P and Σ_Q have no invariant zeros on the unit circle. Then the closed-loop system comprising Equation (4.68) and the full order output feedback controller of Equation (4.106) has the following properties:

1. it is internally stable;
2. the closed-loop transfer matrix from the disturbance w to the controlled output h, $T_{zh}(z)$, possesses $\|T_{zh}\|_2 = \gamma_2^*$.

Hence, Equation (4.106) is an H_2 optimal controller for the system of Equation (4.68). ♢

iii. Reduced Order Output Feedback Case. We now follow the procedure as in the continuous-time case to design a reduced order output feedback

controller. For simplicity of presentation, we assume that the matrices C_1 and D_1 are already in the form

$$C_1 = \begin{bmatrix} 0 & C_{1,02} \\ I_k & 0 \end{bmatrix} \quad \text{and} \quad D_1 = \begin{bmatrix} D_{1,0} \\ 0 \end{bmatrix}, \tag{4.108}$$

where $k = \ell - \text{rank}(D_1)$ and $D_{1,0}$ is of full rank. Next, we follow Steps 4.3.D.F.1 and 4.3.D.F.2 of the previous subsection to compute the constant matrix N, and form the following system:

$$\begin{cases} x(k+1) = \tilde{A}\ x(k) + B\ u(k) + \tilde{E}\ w(k), \\ y(k) \ = C_1\ x(k) \qquad\quad + D_1\ w(k), \\ h(k) \ = \tilde{C}_2\ x(k) + D_2\ u(k) + 0\ w(k), \end{cases} \tag{4.109}$$

where $\tilde{A}$, $\tilde{E}$ and $\tilde{C}_2$ are defined as in Equation (4.104). Then, partition Equation (4.109) as follows:

$$\begin{cases} \begin{pmatrix} x_1(k+1) \\ x_2(k+1) \end{pmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{pmatrix} x_1(k) \\ x_2(k) \end{pmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} u(k) + \begin{bmatrix} E_1 \\ E_2 \end{bmatrix} w(k) \\ \begin{pmatrix} y_0(k) \\ y_1(k) \end{pmatrix} = \begin{bmatrix} 0 & C_{1,02} \\ I_k & 0 \end{bmatrix} \begin{pmatrix} x_1(k) \\ x_2(k) \end{pmatrix} + \begin{bmatrix} D_{1,0} \\ 0 \end{bmatrix} w(k) \\ h(k) = [\, C_{2,1} \quad C_{2,2}\,] \begin{pmatrix} x_1(k) \\ x_2(k) \end{pmatrix} + D_2\ u(k) + 0\ w(k) \end{cases}$$

where the state x of Equation (4.109) is partitioned into two parts, x_1 and x_2; and y is partitioned to y_0 and y_1 with $y_1 \equiv x_1$. Thus, one needs to estimate only the state x_2 in the reduced order controller design. Next, define an auxiliary subsystem Σ_{QR} characterised by a matrix quadruple $(A_{\text{R}}, E_{\text{R}}, C_{\text{R}}, D_{\text{R}})$, where

$$(A_{\text{R}}, E_{\text{R}}, C_{\text{R}}, D_{\text{R}}) = \left(A_{22}, E_2, \begin{bmatrix} C_{1,02} \\ A_{12} \end{bmatrix}, \begin{bmatrix} D_{1,0} \\ E_1 \end{bmatrix} \right). \tag{4.110}$$

The following is a step-by-step algorithm that constructs the reduced order output feedback controller for the general H_2 optimisation.

STEP 4.3.D.R.1: (construction of the gain matrix F_{P}). Define an auxiliary system

$$\begin{cases} x(k+1) = \tilde{A}\ x(k) + B\ u(k) + \tilde{E}\ w(k), \\ y(k) \ = \quad x(k) \\ h(k) \ = \tilde{C}_2\ x(k) + D_2\ u(k) + 0\ w(k), \end{cases} \tag{4.111}$$

and then perform Steps 4.3.D.R.1 to 4.3.D.R.4 of the previous algorithm on the above system to obtain a parameterised gain matrix F. Furthermore, we let $F_{\mathrm{P}} = F$.

STEP 4.3.D.R.2: (construction of the gain matrix K_{R}). Define another auxiliary system

$$\left\{\begin{array}{rcl} x(k+1) &=& A'_{\mathrm{R}}\, x(k) + C'_{\mathrm{R}}\, u(k) + C'_{2,2}\, w(k), \\ y(k) &=& x(k) \\ h(k) &=& E'_{\mathrm{R}}\, x(k) + D'_{\mathrm{R}}\, u(k) + 0 \quad w(k), \end{array}\right. \tag{4.112}$$

and then perform Steps 4.3.D.R.1 to 4.3.D.R.4 of the previous algorithm on the above system to obtain a parameterised gain matrix F. Similarly, we let $K_{\mathrm{R}} = F'$.

STEP 4.3.D.R.3: (construction of the reduced order controller Σ_{RC}). Let us partition F_{P} and K_{R} as

$$F_{\mathrm{P}} = [\, F_{\mathrm{P1}} \quad F_{\mathrm{P2}} \,] \quad \text{and} \quad K_{\mathrm{R}} = [\, K_{\mathrm{R0}} \quad K_{\mathrm{R1}} \,] \tag{4.113}$$

in conformity with $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and $y = \begin{pmatrix} y_0 \\ y_1 \end{pmatrix}$ respectively. Then define

$$G_{\mathrm{R}} = [\, -K_{\mathrm{R0}},\ A_{21} + K_{\mathrm{R1}} A_{11} - (A_{\mathrm{R}} + K_{\mathrm{R}} C_{\mathrm{R}}) K_{\mathrm{R1}} \,]. \tag{4.114}$$

Finally, the parameterised reduced order output feedback controller is given by

$$\Sigma_{\mathrm{RC}} : \left\{\begin{array}{rcl} v(k+1) &=& A_{\mathrm{RC}}\, v(k) + B_{\mathrm{RC}}\, y(k), \\ u(k) &=& C_{\mathrm{RC}}\, v(k) + D_{\mathrm{RC}}\, y(k), \end{array}\right. \tag{4.115}$$

where

$$\left.\begin{array}{l} A_{\mathrm{RC}} := A_{\mathrm{R}} + B_2 F_{\mathrm{P2}} + K_{\mathrm{R}} C_{\mathrm{R}} + K_{\mathrm{R1}} B_1 F_{\mathrm{P2}}, \\ B_{\mathrm{RC}} := G_{\mathrm{R}} + [B_2 + K_{\mathrm{R1}} B_1]\,[0,\ F_{\mathrm{P1}} - F_{\mathrm{P2}} K_{\mathrm{R1}}\,], \\ C_{\mathrm{RC}} := F_{\mathrm{P2}}, \\ D_{\mathrm{RC}} := [\, 0,\ F_{\mathrm{P1}} - F_{\mathrm{P2}} K_{\mathrm{R1}}\,] + N. \end{array}\right\} \tag{4.116}$$

This concludes the algorithm for constructing the reduced order output feedback controller. ◊

Theorem 4.3.6. Consider the given system in Equation (4.68) with $D_{22} = 0$. Assume that Σ_P and Σ_Q have no invariant zeros on the unit circle. Then, the closed-loop system comprising Equation (4.68) and the reduced order output feedback controller in Equation (4.115) has the following properties:

1. it is internally stable;
2. the closed-loop transfer matrix from the disturbance w to the controlled output h, $T_{zh}(z)$, possesses $\|T_{zh}(z)\|_2 = \gamma_2^*$.

Thus, Equation (4.115) is an H_2 optimal controller for the system in Equation (4.68). ◇

Lastly, we note that the result presented in this section, although it is not totally complete, should be sufficient enough to obtain appropriate solutions for HDD servo systems and many engineering problems. We next move to issues of H_∞ control and its related problems.

4.4 H_∞ Control and Disturbance Decoupling

The ultimate goal of a control system designer is to build a system that will work in a real environment. Since the real environment may change and the operating conditions may vary from time to time, the control system must be able to withstand these variations. Even if the environment does not change, other factors of life are the model uncertainties, as well as noises. Any mathematical representation of a system often involves simplifying assumptions. Nonlinearities are either unknown, and hence unmodelled, or are modelled and later ignored in order to simplify analysis. High-frequency dynamics are often ignored at the design stage as well. In consequence, control systems designed based on simplified models may not work on real plants in real environments. The particular property that a control system must possess for it to operate properly in realistic situations is commonly called *robustness*. Mathematically, this means that the controller must perform satisfactorily not just for one plant, but for a family of plants. If a controller can be designed such that the whole system to be controlled remains stable when its parameters vary within certain expected limits, the system is said to possess robust stability. In addition, if it can satisfy performance specifications such as steady state tracking, disturbance rejection and speed of response requirements, it is said to possess robust performance. The problem of designing controllers that satisfy both robust stability and performance requirements is called robust control. H_∞ control theory is one of the cornerstones of modern control theory and was developed in an attempt to solve

such a problem. Many robust control problems (such as the robust stability problem of unstructurally perturbed systems, the mixed-sensitivity problem, robust stabilisation with additive and multiplicative perturbations, to name a few) can be cast into a standard H_∞ control problem (see *e.g.*, [44]).

Since the original formulation of the H_∞ control problem by Zames [74], a great deal of work has been done on finding the solution to this problem. Practically all the research results of the early years involved a mixture of time-domain and frequency-domain techniques, including the following: *Interpolation approach* (see *e.g.*, Limebeer and Anderson [75]); *Frequency domain approach* (see *e.g.*, Doyle [76], Francis [77] and Glover [78]); *Polynomial approach* (see *e.g.*, Kwakernaak [79]); and *J-spectral factorisation approach* (see *e.g.*, Kimura [80]). Recently, considerable attention has been focused on purely *time-domain methods* based on algebraic Riccati equations and/or singular perturbation approach (see *e.g.* [44, 70, 81] and references cited therein).

We also recall in this section the solutions to the problem of H_∞ almost disturbance decoupling with measurement feedback and internal stability. Although, in principle, it is a special case of the general H_∞ control problem, the problem of almost disturbance decoupling has a vast history behind it, occupying a central part of classical as well as modern control theory. Several important problems, such as robust control, decentralised control, non-interactive control, model reference or tracking control, H_2 and H_∞ optimal control problems can all be recast into an almost disturbance decoupling problem. Roughly speaking, the basic almost disturbance decoupling problem is to find an output feedback control law such that in the closed-loop system the disturbances are quenched, say in an L_p sense, up to any pre-specified degree of accuracy while maintaining internal stability. Such a problem was originally formulated by Willems [82, 83] and termed almost disturbance decoupling problem with measurement feedback and internal stability (ADDPMS).

The formulation of H_∞ control is very similar to that of H_2 optimal control. In order to avoid unnecessary repetitions, we will make use of some terms defined in the previous section, *e.g.*, the state-space equations of the given system and its subsystems $\Sigma_{\rm P}$ and $\Sigma_{\rm Q}$, the format of the control law and its corresponding closed-loop transfer matrix, as well as the definitions of the regular and singular problems.

4.4.1 Continuous-time Systems

We consider a continuous-time linear time-invariant system as given in Equation (4.5). For simplicity, we assume that (A, B) is stabilisable, (A, C_1) is detectable, $D_{11} = 0$ and $D_{22} = 0$. The standard H_∞ control problem for

continuous-time systems is to find an internally stabilising proper measurement feedback control law of the format in Equation (4.8) such that when it is applied to Equation (4.5) the resulting closed-loop system is internally stable and the H_∞-norm of the overall closed-loop transfer matrix function from w to h, *i.e.*, $T_{hw}(s)$, is minimised. The H_∞-norm of a stable continuous-time transfer matrix, *e.g.*, $T_{hw}(s)$, is defined as follows:

$$\|T_{hw}\|_\infty := \sup_{\omega\in[0,\infty)} \sigma_{\max}\left[T_{hw}(j\omega)\right] = \sup_{\|w\|_2=1} \frac{\|h\|_2}{\|w\|_2}, \tag{4.117}$$

where w and h are respectively the input and output of $T_{hw}(s)$, and $\|\cdot\|_2$ is the l_2-norm of the corresponding signal. It is clear that the H_∞-norm of $T_{hw}(s)$ corresponds to the worst case gain from its input to its output. For future use, we define

$$\gamma^*_\infty := \inf\Big\{\|T_{hw}(\Sigma\times\Sigma_{\rm cmp})\|_\infty \mid \Sigma_{\rm cmp} \text{ internally stabilises } \Sigma\Big\}. \tag{4.118}$$

We note that the determination of this γ^*_∞ is rather tedious. For a fairly large class of systems, γ^*_∞ can be exactly computed using some numerically stable algorithms. In general, an iterative scheme is required to determine γ^*_∞. We refer interested readers to the work of Chen [44] for a detailed treatment of this particular issue. For simplicity, we assume throughout this section that γ^*_∞ has been determined and hence it is known.

For the case when $\gamma^*_\infty = 0$, the corresponding H_∞ control problem is commonly known in the literature as the problem of H_∞ almost disturbance decoupling with internal stability. It can be shown that such a problem is solvable for Σ of Equation (4.5) if and only if the following conditions hold (see *e.g.*, [44, 60]):

1. (A, B) is stabilisable;
2. (A, C_1) is detectable;
3. $D_{22} + D_2 S D_1 = 0$, where $S = -(D_2'D_2)^\dagger D_2' D_{22} D_1'(D_1 D_1')^\dagger$;
4. $\mathrm{Im}\,(E + BSD_1) \subset \mathcal{S}^+(\Sigma_{\rm P}) \cap \{\cap_{\lambda\in\mathbb{C}^0} \mathcal{S}_\lambda(\Sigma_{\rm P})\}$;
5. $\mathrm{Ker}\,(C_2 + D_2 S C_1) \supset \mathcal{V}^+(\Sigma_{\rm Q}) \cup \{\cup_{\lambda\in\mathbb{C}^0} \mathcal{V}_\lambda(\Sigma_{\rm Q})\}$; and
6. $\mathcal{V}^+(\Sigma_{\rm Q}) \subset \mathcal{S}^+(\Sigma_{\rm P})$.

We note that if $\Sigma_{\rm P}$ is right invertible and of minimum phase, and $\Sigma_{\rm Q}$ is left invertible and of minimum phase, then conditions 4–6 are automatically satisfied.

It transpires that, for H_∞ control, it is almost impossible to find a control law with a finite gain to achieve the optimal performance, *i.e.*, γ_∞^*. As such, we will focus on designing H_∞ suboptimal controllers instead. To be more specific, given a scalar $\gamma > \gamma_\infty^*$, we will focus on finding a control law that yields $\|T_{hw}\|_\infty < \gamma$, where $T_{hw}(s)$ is the corresponding closed-loop transfer matrix. Hereafter, we will call a control law that possesses such a property an H_∞ γ-suboptimal controller.

Next, we proceed to construct a solution to the regular problem (for its definition see the previous section). Given a scalar $\gamma > \gamma_\infty^*$, we solve for positive semi-definite stabilising solutions $P \geq 0$ and $Q \geq 0$ respectively to the following Riccati equations:

$$\begin{aligned} A'P + PA + C_2'C_2 + \gamma^{-2}PEE'P \\ -(PB + C_2'D_2)(D_2'D_2)^{-1}(B'P + D_2'C_2) = 0 \end{aligned} \tag{4.119}$$

and

$$\begin{aligned} AQ + QA' + EE' + \gamma^{-2}QC_2'C_2Q \\ -(QC_1' + ED_1')(D_1D_1')^{-1}(C_1Q + D_1E') = 0. \end{aligned} \tag{4.120}$$

The H_∞ γ-suboptimal control law is given by (see also [70]),

$$\Sigma_{\text{cmp}} : \begin{cases} \dot{v} = A_{\text{cmp}}\, v + B_{\text{cmp}}\, y, \\ u = C_{\text{cmp}}\, v + \quad 0 \quad y, \end{cases} \tag{4.121}$$

where

$$A_{\text{cmp}} = A + \gamma^{-2}EE'P + BF + \left(I - \gamma^{-2}QP\right)^{-1}K\left(C_1 + \gamma^{-2}D_1E'P\right) \tag{4.122}$$

$$B_{\text{cmp}} = -\left(I - \gamma^{-2}QP\right)^{-1}K \tag{4.123}$$

$$C_{\text{cmp}} = F \tag{4.124}$$

and where

$$F = -(D_2'D_2)^{-1}(D_2'C_2 + B'P), \tag{4.125}$$

$$K = -(QC_1' + ED_1')(D_1D_1')^{-1}. \tag{4.126}$$

Note that, for the state feedback case, the H_∞ γ-suboptimal control law is given by $u = Fx$ with F being given as in Equation (4.125).

For the singular case, the following perturbation method can be utilised. For $\gamma > \gamma_\infty^*$ and a positive scalar $\varepsilon > 0$, define $\tilde{E}$, $\tilde{D}_1$, $\tilde{C}_2$ and $\tilde{D}_2$ as in Equations

(4.29) and (4.30), and solve the following Riccati equations:

$$A'\tilde{P} + \tilde{P}A + \tilde{C}_2'\tilde{C}_2 + \gamma^{-2}\tilde{P}\tilde{E}\tilde{E}'\tilde{P} - (\tilde{P}B + \tilde{C}_2'\tilde{D}_2)(\tilde{D}_2'\tilde{D}_2)^{-1}(B'\tilde{P} + \tilde{D}_2'\tilde{C}_2) = 0 \tag{4.127}$$

and

$$AQ + QA' + \tilde{E}\tilde{E}' + \gamma^{-2}Q\tilde{C}_2'\tilde{C}_2Q - (QC_1' + \tilde{E}\tilde{D}_1')(\tilde{D}_1\tilde{D}_1')^{-1}(C_1Q + \tilde{D}_1\tilde{E}') = 0 \tag{4.128}$$

for $\tilde{P} > 0$ and $\tilde{Q} > 0$. Then, it can be shown that there exists an $\varepsilon^* > 0$ such that for all $\varepsilon \in (0, \varepsilon^*]$ the following control law is an H_∞ γ-suboptimal for the given system:

$$\tilde{\Sigma}_{\text{cmp}} : \begin{cases} \dot{v} = \tilde{A}_{\text{cmp}}\, v + \tilde{B}_{\text{cmp}}\, y, \\ u = \tilde{C}_{\text{cmp}}\, v + \quad 0 \quad y, \end{cases} \tag{4.129}$$

where

$$\tilde{A}_{\text{cmp}} = A + \gamma^{-2}EE'\tilde{P} + B\tilde{F} + \left(I - \gamma^{-2}\tilde{Q}\tilde{P}\right)^{-1}\tilde{K}\left(C_1 + \gamma^{-2}D_1E'\tilde{P}\right) \tag{4.130}$$

$$\tilde{B}_{\text{cmp}} = -\left(I - \gamma^{-2}\tilde{Q}\tilde{P}\right)^{-1}\tilde{K} \tag{4.131}$$

$$\tilde{C}_{\text{cmp}} = \tilde{F} \tag{4.132}$$

and where

$$\tilde{F} = -(\tilde{D}_2'\tilde{D}_2)^{-1}(\tilde{D}_2'\tilde{C}_2 + B'\tilde{P}), \tag{4.133}$$

$$\tilde{K} = -(\tilde{Q}C_1' + \tilde{E}\tilde{D}_1')(\tilde{D}_1\tilde{D}_1')^{-1}. \tag{4.134}$$

Note that for the state feedback case, the H_∞ γ-suboptimal control law is given by $u = \tilde{F}x$ with $\tilde{F}$ being given as in Equation (4.133).

Alternatively, the singular H_∞ control problem can also be solved using the singular perturbation approach as in the previous section for H_2 optimal control. In fact, we only need to modify the algorithms slightly for H_2 control to yield required H_∞ γ-suboptimal controllers. We will treat separately the state feedback case and the measurement feedback case. For simplicity, we assume that both subsystems Σ_{P} and Σ_{Q} do not have invariant zeros on the imaginary axis.

i. State Feedback Case. Given a scalar $\gamma > \gamma_\infty^*$, the following algorithm will yield an H_∞ γ-suboptimal state feedback gain matrix for Σ of Equation (4.5) with $y = x$.

STEP 4.4.C.S.1: transform the system Σ_P into the special coordinate basis as given by Theorem 3.4.1. To all submatrices and transformations in the special coordinate basis of Σ_P, we append the subscript P to signify their relation to the system Σ_P. We also choose the output transformation Γ_{0P} to have the following form:

$$\Gamma_{0P} = \begin{bmatrix} I_{m_{0P}} & 0 \\ 0 & \Gamma_{orP} \end{bmatrix}, \tag{4.135}$$

where $m_{0P} = \text{rank}\,(D_2)$. Next, compute

$$\Gamma_{sP}^{-1} E = \begin{bmatrix} E_{aP}^- \\ E_{aP}^+ \\ E_{bP} \\ E_{cP} \\ E_{dP} \end{bmatrix} \tag{4.136}$$

and define A_{11P}, B_{11P}, A_{13P}, C_{21P}, A_{xP}, $B_{xP}B_{xP}$, and $C'_{xP}C_{xP}$ as in Equations (4.37)–(4.42). Finally, define

$$E_{xP} = \begin{bmatrix} E_{aP}^+ \\ E_{bP} \end{bmatrix}. \tag{4.137}$$

STEP 4.4.C.S.2: solve the following algebraic matrix Riccati equation,

$$\begin{aligned} P_x A_{xP} + A'_{xP} P_x + P_x E_{xP} E'_{xP} P_x/\gamma^2 \\ - P_x B_{xP} B'_{xP} P_x + C'_{xP} C_{xP} = 0 \end{aligned} \tag{4.138}$$

for $P_x > 0$ and define

$$F_{11} := \begin{bmatrix} F_{a0}^+ & F_{b0} \\ F_{a1}^+ & F_{b1} \end{bmatrix} = \begin{bmatrix} B'_{11P} P_x \\ (C'_{23P} C_{23P})^{-1} (A'_{13P} P_x + C'_{23P} C_{21P}) \end{bmatrix}. \tag{4.139}$$

Then, partition $[\, F_{a1}^+ \quad F_{b1} \,]$ as

$$[\,F_{a1}^{+} \quad F_{b1}\,] = \begin{bmatrix} F_{a11}^{+} & F_{b11} \\ F_{a12}^{+} & F_{b12} \\ \vdots & \vdots \\ F_{a1m_{d\mathrm{P}}}^{+} & F_{b1m_{d\mathrm{P}}} \end{bmatrix}, \tag{4.140}$$

where F_{a1i}^{+} and F_{b1i} are of dimensions $1 \times n_{a\mathrm{P}}^{+}$ and $1 \times n_{b\mathrm{P}}$ respectively.

STEP 4.4.C.S.3: follow exactly Steps 4.3.C.S.3 to 4.3.C.S.5 of the previous section to construct a state feedback gain matrix $F(\varepsilon)$.

This completes the algorithm. ◊

We have the following result.

Theorem 4.4.1. Consider the given system in Equation (4.5) with $D_{22} = 0$ and $y = x$, *i.e.*, all states are measurable. Assume that Σ_{P} has no invariant zeros on the imaginary axis. Then, the closed-loop system comprising Equation (4.5) and $u = F(\varepsilon)x$ with $F(\varepsilon)$ being given as in the above algorithm has the following properties:

1. it is internally stable for sufficiently small ε;
2. the closed-loop transfer matrix from the disturbance w to the controlled output h, $T_{zh}(s,\varepsilon)$, possesses $\|T_{zh}(s,\varepsilon)\|_\infty < \gamma$, for sufficiently small ε.

Clearly, $u = F(\varepsilon)x$ is an H_∞ γ-suboptimal controller for the system in Equation (4.5). ◊

ii. Measurement Feedback Case. Similarly, one can design either a full order or a reduced order output feedback control law that solves the H_∞ γ-suboptimal problem. Unfortunately, the reduced order controller design for H_∞ control is quite different from its counterpart in H_2 control and is quite complicated. We will only focus below on the design of a full order H_∞ γ-suboptimal controller. Interested readers are referred to [44] for a more complete treatment. The following is a step-by-step algorithm to construct a full order H_∞ γ-suboptimal control law for any $\gamma > \gamma_\infty^*$.

STEP 4.4.C.F.1: define an auxiliary full state feedback system

$$\begin{cases} \dot{x} = A\ x + B\ u + E\ w \\ y = \quad x \\ h = C_2\ x + D_2\ u \end{cases} \tag{4.141}$$

and perform Steps 4.4.C.S.1 to 4.4.C.S.3 of the previous algorithm to obtain a gain matrix $F(\varepsilon)$. Also, define

$$P := (\Gamma_{\text{SP}}^{-1})' \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & P_x & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \Gamma_{\text{SP}}^{-1} \geq 0. \tag{4.142}$$

Note that $P_x > 0$ is the solution of Equation (4.138).

STEP 4.4.C.F.2: define another auxiliary full state feedback system as follows:

$$\begin{cases} \dot{x} = A'\, x + C_1'\, u + C_2'\, w \\ y = \quad x \\ h = E'\, x + D_1'\, u \end{cases} \tag{4.143}$$

and perform Steps 4.4.C.S.1 to 4.4.C.S.3 of the previous algorithm, but for this auxiliary system, to obtain a gain matrix $F(\varepsilon)$. Let us define $K(\varepsilon) := F(\varepsilon)'$. Similarly, define a positive semi-definite matrix Q as in (4.142), but for the current auxiliary system.

STEP 4.4.C.F.3: the full order H_∞ γ-suboptimal controller is given by

$$\Sigma_{\text{cmp}} : \begin{cases} \dot{v} = A_{\text{cmp}}\, v + B_{\text{cmp}}\, y, \\ u = C_{\text{cmp}}\, v + \quad 0 \quad y, \end{cases} \tag{4.144}$$

where

$$\begin{aligned} A_{\text{cmp}} &= A + \gamma^{-2} EE'P + BF(\varepsilon) \\ &\quad + \big(I - \gamma^{-2} QP\big)^{-1} \Big[K(\varepsilon)\big(C_1 + \gamma^{-2} D_1 E'P\big) \\ &\quad + \gamma^{-2} Q\big(A'P + PA + C_2'C_2 + \gamma^{-2} PEE'P\big) \\ &\quad + \gamma^{-2} Q\big(PB + C_2'D_2\big)F(\varepsilon)\Big], \end{aligned} \tag{4.145}$$

$$B_{\text{cmp}} = -\big(I - \gamma^{-2} QP\big)^{-1} K(\varepsilon), \tag{4.146}$$

$$C_{\text{cmp}} = F(\varepsilon). \tag{4.147}$$

This completes the algorithm. ◊

We have the following theorem.

Theorem 4.4.2. Consider the given system in Equation (4.5) with $D_{22} = 0$. Assume that Σ_P and Σ_Q have no invariant zeros on the imaginary axis. Then, the closed-loop system comprising Equation (4.5) and the measurement feedback controller in Equation (4.144) has the following properties:

1. it is internally stable for sufficiently small ε;
2. the closed-loop transfer matrix from the disturbance w to the controlled output h, $T_{zh}(s,\varepsilon)$, possesses $\|T_{zh}(s,\varepsilon)\|_\infty < \gamma$, for sufficiently small ε.

Hence, Equation (4.144) is an H_∞ γ-suboptimal controller for the system in Equation (4.5). ◊

Remark 4.4.1. For the case when $\gamma_\infty^* = 0$, *i.e.*, the problem of H_∞ almost disturbance decoupling with internal stability for Σ of Equation (4.5) (with both subsystems Σ_P and Σ_Q having no invariant zeros on the imaginary axis) is solvable, then either the above algorithm or the algorithm in the previous section for H_2 optimal control would yield desired solutions. We note that, in general, suboptimal solutions are nonunique. In fact, one could utilise the algorithm for constructing the reduced order output feedback H_2 suboptimal controller in the previous section to construct a reduced order solution for the H_∞ almost disturbance decoupling problem. ◊

4.4.2 Discrete-time Systems

We now consider a discrete-time linear time-invariant system as given in Equation (4.68). Again, for simplicity, we assume that (A, B) is stabilisable, (A, C_1) is detectable, $D_{11} = 0$ and $D_{22} = 0$. The standard H_∞ control problem for discrete-time systems is to find an internally stabilising proper measurement feedback control law of the format in Equation (4.69) such that, when it is applied to the system in Equation (4.68), the resulting closed-loop system is internally stable and the H_∞-norm of the overall closed-loop transfer matrix function from w to h, *i.e.*, $T_{hw}(z)$, is minimised. The H_∞-norm of a stable discrete-time transfer matrix, *e.g.*, $T_{hw}(z)$, is defined as follows:

$$\|T_{hw}\|_\infty := \sup_{\omega\in[0,2\pi]} \sigma_{\max}\left[T_{hw}(e^{j\omega})\right] = \sup_{\|w\|_2=1} \frac{\|h\|_2}{\|w\|_2}, \tag{4.148}$$

where w and h are respectively the input and output of $T_{hw}(s)$, and $\|\cdot\|_2$ is the l_2-norm of the corresponding signal. Next, we define

$$\gamma_\infty^* := \inf\Big\{\|T_{hw}(\Sigma \times \Sigma_{\text{cmp}})\|_\infty \mid \Sigma_{\text{cmp}} \text{ internally stabilises } \Sigma\Big\}. \quad (4.149)$$

Again, we refer interested readers to the work of Chen [44] for the computation of γ_∞^*. For simplicity, we assume throughout this section that γ_∞^* has been determined and hence it is known.

It can be shown that the problem of H_∞ almost disturbance decoupling with internal stability is solvable for Σ of Equation (4.5), *i.e.*, $\gamma_\infty^* = 0$, if and only if the following conditions hold (see *e.g.*, Chen [44]):

1. (A, B) is stabilisable;
2. (A, C_1) is detectable;
3. $D_{22} + D_2 S D_1 = 0$, where $S = -(D_2'D_2)^\dagger D_2' D_{22} D_1'(D_1 D_1')^\dagger$;
4. $\text{Im}\,(E + BSD_1) \subset \Big\{\mathcal{V}^\odot(\Sigma_{\text{P}}) + B\text{Ker}\,(D_2)\Big\} \cap \Big\{\bigcap_{|\lambda|=1} \mathcal{S}_\lambda(\Sigma_{\text{P}})\Big\}$;
5. $\text{Ker}\,(C_2 + D_2 S C_1) \supset \Big\{\mathcal{S}^\odot(\Sigma_{\text{Q}}) \cap C_1^{-1}\{\text{Im}\,(D_1)\}\Big\} \cup \Big\{\bigcup_{|\lambda|=1} \mathcal{V}_\lambda(\Sigma_{\text{Q}})\Big\}$;
6. $\mathcal{S}^\odot(\Sigma_{\text{Q}}) \subset \mathcal{V}^\odot(\Sigma_{\text{P}})$.

We note that if Σ_{P} is right invertible and of minimum phase with no infinite zeros, and Σ_{Q} is left invertible and of minimum phase with no infinite zeros, then conditions 4–6 are automatically satisfied.

The problems of discrete-time H_∞ control and H_∞ almost disturbance decoupling can be solved explicitly in the discrete-time domain. Complete solutions to these problems have been reported by Chen [44]. However, by utilising the bilinear transformations of Chapter 3, we can convert these discrete-time problems into equivalent continuous-time problems, and thus all algorithms presented in the previous subsection can be readily applied to yield desired solutions. The procedure is pretty simple.

1. We first apply the result of Lemma 3.5.2 with $a = 1$ to the given discrete-time system in Equation (4.68) to obtain an equivalent continuous-time counterpart, *i.e.*,

$$\left.\begin{aligned} \dot{x} &= A_{\text{d}}\, x + B_{\text{d}}\, u + E_{\text{d}}\, w \\ y &= C_{1\text{d}}\, x + D_{11\text{d}}\, u + D_{1\text{d}}\, w \\ h &= C_{2\text{d}}\, x + D_{2\text{d}}\, u + D_{22\text{d}}\, w \end{aligned}\right\} \quad (4.150)$$

   where

$$\left.\begin{aligned}
A_{\mathrm{d}} &= (A+I)^{-1}(A-I),\\
B_{\mathrm{d}} &= \sqrt{2}(A+I)^{-1}B,\\
E_{\mathrm{d}} &= \sqrt{2}(A+I)^{-1}E,\\
C_{1\mathrm{d}} &= \sqrt{2}C_1(A+I)^{-1},\\
D_{11\mathrm{d}} &= D_{11} - C_1(A+I)^{-1}B,\\
D_{1\mathrm{d}} &= D_1 - C_1(A+I)^{-1}E,\\
C_{2\mathrm{d}} &= \sqrt{2}C_2(A+I)^{-1},\\
D_{2\mathrm{d}} &= D_2 - C_2(A+I)^{-1}B,\\
D_{22\mathrm{d}} &= D_{22} - C_2(A+I)^{-1}E.
\end{aligned}\right\} \tag{4.151}$$

Note that if A has eigenvalues at -1, then one should apply some pre-feedback laws to remove them.

2. Next, utilise any method of the previous subsection to Equation (4.150) to find an appropriate H_∞ γ-suboptimal controller, say,

$$\left.\begin{aligned}
\dot{v} &= A_{\mathrm{c}}\, v + B_{\mathrm{c}}\, y,\\
u &= C_{\mathrm{c}}\, v + D_{\mathrm{c}}\, y.
\end{aligned}\right\} \tag{4.152}$$

Note that some pre-feedback might be necessary to wash out $D_{22\mathrm{d}}$.

3. Lastly, apply the result of Lemma 3.5.1 to convert Equation (4.152) to a discrete-time equivalent system. It is known in the literature (see *e.g.*, [78]) that the discrete-time controller obtained is H_∞ γ-suboptimal to the original discrete-time system in Equation (4.68). $\diamondsuit$

The above procedure is also applicable to finding discrete-time H_∞ almost disturbance decoupling controllers. As mentioned earlier, all results presented in the previous section on H_2 optimal control are applicable to solving the H_∞ almost disturbance decoupling problem when both Σ_{P} and Σ_{Q} have no invariant zeros on the unit circle.

4.5 Robust and Perfect Tracking Control

We present in this section a robust and perfect tracking (RPT) problem, which was proposed and solved by Liu *et al.* [84] for continuous-time systems and Chen *et al.* [85] for discrete-time systems (see also Chen [44]). The RPT problem is to design a controller such that the resulting closed-loop system is asymptotically stable and the controlled output almost perfectly tracks a

given reference signal in the presence of any initial conditions and external disturbances. By almost perfect tracking we mean the ability of a controller to track a given reference signal with an arbitrarily fast settling time in the face of external disturbances and initial conditions. Of course, in real life, a certain trade-off has to be made in order to design a physically implementable control law. The results of this section will be utilised heavily in later chapters to solve track following problems in hard disk drive servo systems.

4.5.1 Continuous-time Systems

Consider the following continuous-time system:

$$\Sigma : \begin{cases} \dot{x} = A\ x + B\ u + E\ w, & x(0) = x_0, \\ y = C_1\ x \qquad\quad + D_1\ w, \\ h = C_2\ x + D_2\ u + D_{22}\ w, \end{cases} \tag{4.153}$$

where $x \in \mathbb{R}^n$ is the state, $u \in \mathbb{R}^m$ is the control input, $w \in \mathbb{R}^q$ is the external disturbance, $y \in \mathbb{R}^p$ is the measurement output, and $h \in \mathbb{R}^\ell$ is the output to be controlled. We also assume that the pair (A, B) is stabilisable and (A, C_1) is detectable. For future reference, we define Σ_P and Σ_Q to be the subsystems characterised by the matrix quadruples (A, B, C_2, D_2) and (A, E, C_1, D_1) respectively. Given the external disturbance $w \in L_p$, $p \in [1, \infty)$, and any reference signal vector $r \in \mathbb{R}^\ell$ with $r, \dot{r}, \cdots, r^{(\kappa-1)}$, $\kappa \geq 1$, being available, and $r^{(\kappa)}$ being either a vector of delta functions or in L_p, the RPT problem for the system in Equation (4.153) is to find a parameterised dynamic measurement control law of the following form

$$\begin{cases} \dot{v} = A_{\text{cmp}}(\varepsilon)v + B_{\text{cmp}}(\varepsilon)y + G_0(\varepsilon)r + \cdots + G_{\kappa-1}(\varepsilon)r^{(\kappa-1)} \\ u = C_{\text{cmp}}(\varepsilon)v + D_{\text{cmp}}(\varepsilon)y + H_0(\varepsilon)r + \cdots + H_{\kappa-1}(\varepsilon)r^{(\kappa-1)} \end{cases} \tag{4.154}$$

such that when Equation (4.154) is applied Equation (4.153), we have

1. there exists an $\varepsilon^* > 0$ such that the resulting closed-loop system with $r = 0$ and $w = 0$ is asymptotically stable for all $\varepsilon \in (0, \varepsilon^*]$; and
2. let $h(t, \varepsilon)$ be the closed-loop controlled output response and let $e(t, \varepsilon)$ be the resulting tracking error, *i.e.*, $e(t, \varepsilon) := h(t, \varepsilon) - r(t)$. Then, for any initial condition of the state, $x_0 \in \mathbb{R}^n$,

$$\|e\|_p = \left(\int_0^\infty |e(t)|^p \, \mathrm{d}t \right)^{1/p} \to 0 \quad \text{as } \varepsilon \to 0. \tag{4.155}$$

We introduce in the above formulation some additional information besides the reference signal r, *i.e.*, $\dot{r}, \ddot{r}, \cdots, r^{(\kappa-1)}$, as additional controller inputs. Note that, in general, these additional signals can easily be generated without any extra costs. For example, if $r(t) = t^2 \cdot 1(t)$, where $1(t)$ is a unit step function, then one can easily obtain its first-order derivative

$$\dot{r}(t) = 2t \cdot 1(t) + t^2 \cdot \delta(t) = 2t \cdot 1(t), \tag{4.156}$$

where $\delta(t)$ is a unit impulse function, and the second-order derivative

$$\ddot{r}(t) = 2 \cdot 1(t). \tag{4.157}$$

These $\dot{r}(t)$ and $\ddot{r}(t)$ can be used to improve the overall tracking performance, whereas $r^{(3)}(t) = 2 \cdot \delta(t)$ does not exist in the real world and hence cannot be used. We also note that our formulation covers all possible reference signals that have the form $r(t) = t^k$, $0 \leq k < \infty$. Thus, our method could be applied to track approximately those reference signals that have a Taylor series expansion at $t = 0$. This can be done by truncating the higher-order terms of the Taylor series of the given signal. Also, it is simple to see that, when $r(t) \equiv 0$, the proposed problem reduces to the well-known perfect regulation problem with measurement feedback.

It is shown that the RPT problem for Equation (4.153) is solvable if and only if the following conditions hold:

1. (A, B) is stabilisable and (A, C_1) is detectable;
2. $D_{22} + D_2 S D_1 = 0$, where $S = -(D_2' D_2)^{\dagger} D_2' D_{22} D_1' (D_1 D_1')^{\dagger}$;
3. Σ_{P}, *i.e.*, (A, B, C_2, D_2), is right invertible and of minimum phase;
4. $\mathrm{Ker}\,(C_2 + D_2 S C_1) \supset C_1^{-1}\{\mathrm{Im}\,(D_1)\}$.

We assume throughout the rest of this subsection that the above conditions are satisfied, and move on to construct solutions to the RPT problem. As usual, we will focus on the following three cases: (1) the state feedback case; (2) the full order measurement feedback case; and (3) the reduced order measurement feedback case.

i. State Feedback Case. When all states of the plant are measured for feedback, the problem can be solved by a static control law. We construct in this subsection a parameterised state feedback control law,

$$u = F(\varepsilon)x + H_0(\varepsilon)r + \cdots + H_{\kappa-1}(\varepsilon)r^{(\kappa-1)} \tag{4.158}$$

that solves the RPT problem for the system in Equation (4.153). It is simple to note that we can rewrite the given reference in the following form:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} r \\ \vdots \\ r^{(\kappa-2)} \\ r^{(\kappa-1)} \end{pmatrix} = \begin{bmatrix} 0 & I_\ell & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_\ell \\ 0 & 0 & \cdots & 0 \end{bmatrix} \begin{pmatrix} r \\ \vdots \\ r^{(\kappa-2)} \\ r^{(\kappa-1)} \end{pmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ I_\ell \end{bmatrix} r^{(\kappa)}. \tag{4.159}$$

Combining Equation (4.159) with the given system, we obtain the following augmented system:

$$\Sigma_{\text{AUG}} : \begin{cases} \dot{\boldsymbol{x}} = \boldsymbol{A}\ \boldsymbol{x} + \boldsymbol{B}\ u + \boldsymbol{E}\ \boldsymbol{w} \\ \boldsymbol{y} = \quad \boldsymbol{x} \\ e = \boldsymbol{C}_2\ \boldsymbol{x} + \boldsymbol{D}_2\ u \end{cases} \tag{4.160}$$

where

$$\boldsymbol{w} := \begin{pmatrix} w \\ r^{(\kappa)} \end{pmatrix}, \quad \boldsymbol{x} := \begin{pmatrix} r \\ \vdots \\ r^{(\kappa-2)} \\ r^{(\kappa-1)} \\ x \end{pmatrix}, \tag{4.161}$$

$$\boldsymbol{A} = \begin{bmatrix} 0 & I_\ell & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_\ell & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & A \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ B \end{bmatrix}, \quad \boldsymbol{E} = \begin{bmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & I_\ell \\ E & 0 \end{bmatrix}, \tag{4.162}$$

and

$$\boldsymbol{C}_2 = [\,-I_\ell \quad 0 \quad 0 \quad \cdots \quad 0 \quad C_2\,], \quad \boldsymbol{D}_2 = D_2. \tag{4.163}$$

It is then straightforward to show that the subsystem from u to e in the augmented system of Equation (4.160), *i.e.*, the quadruple $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}_2, \boldsymbol{D}_2)$, is right invertible and has the same infinite zero structure as that of Σ_{P}. Furthermore, its invariant zeros contain those of Σ_{P} and $\ell \times \kappa$ extra ones at $s = 0$. We are now ready to present a step-by-step algorithm to construct the required control law of the form in Equation (4.158).

STEP 4.5.C.S.1: this step transforms the subsystem from u to e of the augmented system in Equation (4.160) into the special coordinate basis of

Theorem 3.4.1, *i.e.*, finds nonsingular state, input and output transformations Γ_s, Γ_i and Γ_o to put the subsystem into the structural form of Theorem 3.4.1 and in a small variation of the compact form of Equations (3.29) to (3.32). It can be shown that the compact form of Equations (3.29) to (3.32) for the subsystem from u to e of Equation (4.160) can be written as

$$\tilde{A} = \begin{bmatrix} A^0_{aa} & 0 & 0 & 0 \\ 0 & A^-_{aa} & 0 & L^-_{ad}C_d \\ B_cE^0_{ca} & B_cE^-_{ca} & A_{cc} & L_{cd}C_d \\ B_dE^0_{da} & B_dE^-_{da} & B_dE_{dc} & A_{dd} \end{bmatrix}, \tag{4.164}$$

$$A^0_{aa} = \begin{bmatrix} 0 & I_\ell & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_\ell \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \quad \tilde{B} = \begin{bmatrix} 0 & 0 & 0 \\ B^-_{0a} & 0 & 0 \\ B_{0c} & 0 & B_c \\ B_{0d} & B_d & 0 \end{bmatrix}, \tag{4.165}$$

and

$$\tilde{C} = \begin{bmatrix} C^0_{0a} & C^-_{0a} & C_{0c} & C_{0d} \\ 0 & 0 & 0 & C_d \end{bmatrix}, \quad \tilde{D} = \begin{bmatrix} I_{m_0} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{4.166}$$

STEP 4.5.C.S.2: choose an appropriate dimensional matrix F_c such that

$$A^c_{cc} = A_{cc} - B_cF_c \tag{4.167}$$

is asymptotically stable. The existence of such an F_c is guaranteed by the property that (A_{cc}, B_c) is completely controllable.

STEP 4.5.C.S.3: for each x_i of x_d, which is associated with the infinite zero structure of Σ_P or the subsystem from u to e of Equation (4.160), we choose an F_i such that

$$p_i(s) = \prod_{j=1}^{q_i}(s - \lambda_{ij}) = s^{q_i} + F_{i1}s^{q_i-1} + \cdots + F_{iq_{i-1}}s + F_{iq_i} \tag{4.168}$$

with all λ_{ij} being in $\mathbb{C}^-$. Let

$$F_i = [\, F_{iq_i} \quad F_{iq_{i-1}} \quad \cdots \quad F_{i1} \,], \quad i = 1, \cdots, m_d. \tag{4.169}$$

STEP 4.5.C.S.4: next, we construct

$$\boldsymbol{F}(\varepsilon) = -\Gamma_{\mathrm{i}} \begin{bmatrix} C_{0a}^{0} & C_{0a}^{-} & C_{0c} & C_{0d} \\ E_{da}^{0} & E_{da}^{-} & E_{dc} & E_{d} + F_{d}(\varepsilon) \\ E_{ca}^{0} & E_{ca}^{-} & F_{c} & 0 \end{bmatrix} \Gamma_{\mathrm{s}}^{-1}, \tag{4.170}$$

where

$$E_{d} = \begin{bmatrix} E_{11} & \cdots & E_{1m_d} \\ \vdots & \ddots & \vdots \\ E_{m_d 1} & \cdots & E_{m_d m_d} \end{bmatrix}, \tag{4.171}$$

$$F_{d}(\varepsilon) = \mathrm{blkdiag}\left\{ \frac{F_1}{\varepsilon^{q_1}} S_1(\varepsilon), \frac{F_2}{\varepsilon^{q_2}} S_2(\varepsilon), \cdots, \frac{F_{m_d}}{\varepsilon^{q_{m_d}}} S_{m_d}(\varepsilon) \right\}, \tag{4.172}$$

and where

$$S_i(\varepsilon) = \mathrm{diag}\left\{1, \varepsilon, \varepsilon^2, \cdots, \varepsilon^{q_i - 1}\right\}. \tag{4.173}$$

STEP 4.5.C.S.5: finally, we partition

$$\boldsymbol{F}(\varepsilon) = [\, H_0(\varepsilon) \quad \cdots \quad H_{\kappa-1}(\varepsilon) \quad F(\varepsilon) \,], \tag{4.174}$$

where $H_i(\varepsilon) \in \mathbb{R}^{m \times \ell}$ and $F(\varepsilon) \in \mathbb{R}^{m \times n}$. This ends the constructive algorithm. ◇

We have the following result.

Theorem 4.5.1. Consider the given system in Equation (4.153) with its external disturbance $w \in L_p$, $p \in [1, \infty)$, its initial condition $x(0) = x_0$. Assume that all its states are measured for feedback, *i.e.*, $C_1 = I$ and $D_1 = 0$, and the RPT problem for the system in Equation (4.153) is solvable. Then, for any reference signal $r(t)$, which has all its ith-order derivatives, $i = 0, 1, \cdots, \kappa - 1$, $\kappa \geq 1$, being available and $r^{(\kappa)}(t)$ being either a vector of delta functions or in L_p, the RPT problem is solved by the control law of Equation (4.158) with $F(\varepsilon)$ and $H_i(\varepsilon)$, $i = 0, 1, \cdots, \kappa - 1$, as given in Equation (4.174). ◇

The following remark gives an alternative approach for solving the proposed RPT problem via full state feedback. We leave the proof of this method to readers as an exercise.

Remark 4.5.1. Note that the required gain matrices for the state feedback RPT problem might be computed by solving the following Riccati equation:

$$\boldsymbol{P}\tilde{\boldsymbol{A}}+\tilde{\boldsymbol{A}}'\boldsymbol{P}+\tilde{\boldsymbol{C}}_2'\tilde{\boldsymbol{C}}_2-\left(\boldsymbol{P}\boldsymbol{B}+\tilde{\boldsymbol{C}}_2'\tilde{\boldsymbol{D}}_2\right)\left(\tilde{\boldsymbol{D}}_2'\tilde{\boldsymbol{D}}_2\right)^{-1}\left(\boldsymbol{P}\boldsymbol{B}+\tilde{\boldsymbol{C}}_2'\tilde{\boldsymbol{D}}_2\right)'=0$$

for a positive definite solution $\boldsymbol{P} > 0$, where

$$\tilde{\boldsymbol{C}}_2 = \begin{bmatrix} \boldsymbol{C}_2 \\ \varepsilon I_{\kappa\ell+n} \\ 0 \end{bmatrix}, \quad \tilde{\boldsymbol{D}}_2 = \begin{bmatrix} \boldsymbol{D}_2 \\ 0 \\ \varepsilon I_m \end{bmatrix}, \tag{4.175}$$

$$\tilde{\boldsymbol{A}} = \begin{bmatrix} \tilde{A}_0 & 0 \\ 0 & A \end{bmatrix}, \quad \tilde{A}_0 = -\varepsilon I_{\kappa\ell} + \begin{bmatrix} 0 & I_\ell & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_\ell \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \tag{4.176}$$

and where $\boldsymbol{B}$, $\boldsymbol{C}_2$ and $\boldsymbol{D}_2$ are as defined in Equations (4.162) and (4.163). The required gain matrix is then given by

$$\begin{aligned} \tilde{\boldsymbol{F}}(\varepsilon) &= -\left(\tilde{\boldsymbol{D}}_2'\tilde{\boldsymbol{D}}_2\right)^{-1}\left(\boldsymbol{P}\boldsymbol{B}+\tilde{\boldsymbol{C}}_2'\tilde{\boldsymbol{D}}_2\right)' \\ &= [\,H_0(\varepsilon) \quad \cdots \quad H_{\kappa-1}(\varepsilon) \quad F(\varepsilon)\,], \end{aligned} \tag{4.177}$$

where $H_i(\varepsilon) \in \mathbb{R}^{m\times\ell}$ and $F(\varepsilon) \in \mathbb{R}^{m\times n}$. ◇

Finally, we note that solutions to the Riccati equation in Remark 4.5.1 might have severe numerical problems as ε becomes smaller and smaller.

We will consider two types of measurement feedback control laws: one is of full order controllers whose dynamical order is equal to the order of the given system; the other is of reduced order controllers with a dynamical order that is less than the order of the given system. Without loss of generality, we assume throughout this subsection that $D_{22} = 0$. If it is nonzero, it can always be washed out by the following pre-output feedback: $u = Sy$.

ii. Full Order Output Feedback Case. The following is a step-by-step algorithm for constructing a parameterised full order measurement feedback controller, which solves the RPT problem.

STEP 4.5.C.F.1: for the given reference $r(t)$ and the given system in Equation (4.153), we first assume that all the state variables of Equation (4.153) are measurable and follow the procedures of the state feedback case to define an auxiliary system,

$$\begin{cases} \dot{\boldsymbol{x}} = \boldsymbol{A}\ \boldsymbol{x} + \boldsymbol{B}\ u + \boldsymbol{E}\ \boldsymbol{w} \\ \boldsymbol{y} = \quad \boldsymbol{x} \\ e = \boldsymbol{C}_2\ \boldsymbol{x} + \boldsymbol{D}_2\ u\ . \end{cases} \tag{4.178}$$

Then, we follow Steps 4.5.C.S.1 to 4.5.C.S.5 of the algorithm of the state feedback case to construct a state feedback gain matrix:

$$\boldsymbol{F}(\varepsilon) = [\, H_0(\varepsilon) \quad \cdots \quad H_{\kappa-1}(\varepsilon) \quad F(\varepsilon) \,]\,. \tag{4.179}$$

STEP 4.5.C.F.2: let Σ_{Qa} be characterised by a matrix quadruple

$$(A_{Qa}, E_{Qa}, C_{Qa}, D_{Qa}) := (A, [\, E \quad I_n \,], C_1, [\, D_1 \quad 0 \,])\,. \tag{4.180}$$

This step is to transform this Σ_{Qa} into the special coordinate basis of Theorem 3.4.1. Because of the special structure of the matrix E_{Qa}, it is simple to show that Σ_{Qa} is always right invertible and is free of invariant zeros. Utilise the results of Theorem 3.4.1 to find nonsingular state, input and output transformation Γ_{sQ}, Γ_{iQ} and Γ_{oQ} such that

$$\Gamma_{sQ}^{-1} A \Gamma_{sQ} = \begin{bmatrix} A_{ccQ} & L_{cdQ} \\ E_{dcQ} & A_{ddQ} \end{bmatrix} + \begin{bmatrix} B_{0cQ} \\ B_{0dQ} \end{bmatrix} [\, C_{0cQ} \quad 0 \,], \tag{4.181}$$

$$\Gamma_{sQ}^{-1} E_{Qa} \Gamma_{iQ} = \begin{bmatrix} B_{0cQ} & 0 & I_{n-k} & 0 \\ B_{0dQ} & I_k & 0 & 0 \end{bmatrix}, \tag{4.182}$$

$$\Gamma_{oQ}^{-1} C_1 \Gamma_{sQ} = \begin{bmatrix} C_{0cQ} & 0 \\ 0 & I_k \end{bmatrix}, \tag{4.183}$$

$$\Gamma_{oQ}^{-1} [\, D_1 \quad 0 \,] \Gamma_{iQ} = \begin{bmatrix} I_{p-k} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{4.184}$$

where $k = p - \text{rank}(D_1)$. It can be verified that the pair (A, C_1) is detectable if and only if the pair

$$\left(A_{ccQ}, \begin{bmatrix} C_{0cQ} \\ E_{dcQ} \end{bmatrix} \right) \tag{4.185}$$

is detectable.

STEP 4.5.C.F.3: let K_{cQ} be an appropriate dimensional constant matrix such that the eigenvalues of the matrix

$$A_{ccQ}^{c} = A_{ccQ} - K_{cQ} \begin{bmatrix} C_{0cQ} \\ E_{dcQ} \end{bmatrix} = A_{ccQ} - [\, K_{c0Q} \quad K_{cdQ} \,] \begin{bmatrix} C_{0cQ} \\ E_{dcQ} \end{bmatrix} \tag{4.186}$$

are all in $\mathbb{C}^-$. Next, we define a parameterised observer gain matrix,

$$K(\varepsilon) = -\Gamma_{sQ} \begin{bmatrix} B_{0cQ} + K_{c0Q} & L_{cdQ} + K_{cdQ}/\varepsilon \\ B_{0dQ} & A_{ddQ} + I_k/\varepsilon \end{bmatrix} \Gamma_{oQ}^{-1}. \tag{4.187}$$

STEP 4.5.C.F.4: finally, we obtain the following full order measurement feedback control law:

$$\begin{cases} \dot{v} = A_{cmp}(\varepsilon)v - K(\varepsilon)y + BH_0(\varepsilon)r + \cdots + BH_{\kappa-1}(\varepsilon)r^{(\kappa-1)} \\ u = F(\varepsilon)\ v + H_0(\varepsilon)\, r + \cdots + H_{\kappa-1}(\varepsilon)\, r^{(\kappa-1)} \end{cases} \tag{4.188}$$

where $A_{cmp}(\varepsilon) = A + BF(\varepsilon) + K(\varepsilon)C_1$. This completes the construction of the full order measurement feedback controller. ◊

We have the following theorem.

Theorem 4.5.2. Consider the given system in Equation (4.153) with its external disturbance $w \in L_p$, $p \in [1, \infty)$, its initial condition $x(0) = x_0$. Assume that the RPT problem is solvable for the system in Equation (4.153). Then, for any reference signal $r(t)$, which has all its ith-order derivatives, $i = 0, 1, \cdots, \kappa - 1$, $\kappa \geq 1$, being available and $r^{(\kappa)}(t)$ being either a vector of delta functions or in L_p, the proposed RPT problem is solved by the parameterised full order measurement feedback control laws as given in Equation (4.188). ◊

The following remark yields an alternative way to compute the gain matrix $K(\varepsilon)$ in Step 4.5.C.F.3.

Remark 4.5.2. The gain matrix $K(\varepsilon)$ in Step 4.5.C.F.3 can also be computed by solving the following Riccati equation:

$$\begin{aligned} AQ + QA' + (EE' + I) \\ - (QC_1' + ED_1')(D_1D_1' + \varepsilon I)^{-1}(C_1Q + D_1E') = 0, \end{aligned} \tag{4.189}$$

for a positive definite solution $Q > 0$. The required gain matrix $K(\varepsilon)$ is then given by

$$K(\varepsilon) = -(QC_1' + ED_1')(D_1D_1' + \varepsilon I)^{-1}. \tag{4.190}$$

Again, this approach might have some numerical problems. ◊

iii. Reduced Order Output Feedback Case. We now present solutions to the RPT problem via reduced order measurement feedback control laws. For

simplicity of presentation, we assume that matrices C_1 and D_1 have already been transformed into the following forms:

$$C_1 = \begin{bmatrix} 0 & C_{1,02} \\ I_k & 0 \end{bmatrix} \quad \text{and} \quad D_1 = \begin{bmatrix} D_{1,0} \\ 0 \end{bmatrix}, \tag{4.191}$$

where $D_{1,0}$ is of full row rank. Before we present a step-by-step algorithm to construct a parameterised reduced order measurement feedback controller, we first partition the following system

$$\begin{cases} \dot{x} = A\ x + B\ u + [E \quad I_n]\ \tilde{w} \\ y = C_1\ x \qquad\quad + [D_1 \quad 0]\ \tilde{w} \end{cases} \tag{4.192}$$

in conformity with the structures of C_1 and D_1 in Equation (4.191), *i.e.*,

$$\begin{cases} \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} u + \begin{bmatrix} E_1 & I_k & 0 \\ E_2 & 0 & I_{n-k} \end{bmatrix} \tilde{w} \\ \begin{pmatrix} y_0 \\ y_1 \end{pmatrix} = \begin{bmatrix} 0 & C_{1,02} \\ I_k & 0 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \qquad\qquad + \begin{bmatrix} D_{1,0} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tilde{w} \end{cases} \tag{4.193}$$

where

$$\tilde{w} = \begin{pmatrix} w \\ x_0 \cdot \delta(t) \end{pmatrix}. \tag{4.194}$$

Obviously, $y_1 = x_1$ is directly available and hence need not be estimated. Next, we define Σ_{QR} to be characterised by

$$(A_{\text{R}}, E_{\text{R}}, C_{\text{R}}, D_{\text{R}}) = \left(A_{22}, [E_2 \ \ 0 \ \ I_{n-k}], \begin{bmatrix} C_{1,02} \\ A_{12} \end{bmatrix}, \begin{bmatrix} D_{1,0} & 0 & 0 \\ E_1 & I_k & 0 \end{bmatrix} \right).$$

It is again straightforward to verify that Σ_{QR} is right invertible with no finite and infinite zeros. Moreover, $(A_{\text{R}}, C_{\text{R}})$ is detectable if and only if (A, C_1) is detectable. We are ready to present the following algorithm.

STEP 4.5.C.R.1: for the given reference $r(t)$ and the given system in Equation (4.153), we again assume that all the state variables of Equation (4.153) are measurable and follow the procedures of the state feedback case to define an auxiliary system,

$$\begin{cases} \dot{\boldsymbol{x}} = \boldsymbol{A}\ \boldsymbol{x} + \boldsymbol{B}\ u + \boldsymbol{E}\ \boldsymbol{w} \\ \boldsymbol{y} = \quad \boldsymbol{x} \\ e = \boldsymbol{C}_2\ \boldsymbol{x} + \boldsymbol{D}_2\ u\ . \end{cases} \tag{4.195}$$

Then, we follow Steps 4.5.C.S.1 to 4.5.C.S.5 of the algorithm of the state feedback case to construct a state feedback gain matrix

$$\boldsymbol{F}(\varepsilon) = [\, H_0(\varepsilon) \quad \cdots \quad H_{\kappa-1}(\varepsilon) \quad F(\varepsilon) \,]. \tag{4.196}$$

Let us partition $F(\varepsilon)$ in conformity with x_1 and x_2 of Equation (4.193) as follows,

$$F(\varepsilon) = [\, F_1(\varepsilon) \quad F_2(\varepsilon) \,]. \tag{4.197}$$

STEP 4.5.C.R.2: let K_{R} be an appropriate dimensional constant matrix such that the eigenvalues of

$$A_{\mathrm{R}} + K_{\mathrm{R}} C_{\mathrm{R}} = A_{22} + [\, K_{\mathrm{R}0} \quad K_{\mathrm{R}1} \,] \begin{bmatrix} C_{1,02} \\ A_{12} \end{bmatrix} \tag{4.198}$$

are all in $\mathbb{C}^-$. This can be done because $(A_{\mathrm{R}}, C_{\mathrm{R}})$ is detectable.

STEP 4.5.C.R.3: let

$$G_{\mathrm{R}} = [\, -K_{\mathrm{R}0}, \quad A_{21} + K_{\mathrm{R}1} A_{11} - (A_{\mathrm{R}} + K_{\mathrm{R}} C_{\mathrm{R}}) K_{\mathrm{R}1} \,] \tag{4.199}$$

and

$$\left.\begin{aligned} A_{\mathrm{cmp}}(\varepsilon) &= A_{\mathrm{R}} + B_2 F_2(\varepsilon) + K_{\mathrm{R}} C_{\mathrm{R}} + K_{\mathrm{R}1} B_1 F_2(\varepsilon) \\ B_{\mathrm{cmp}}(\varepsilon) &= G_{\mathrm{R}} + (B_2 + K_{\mathrm{R}1} B_1)\,[\, 0, \quad F_1(\varepsilon) - F_2(\varepsilon) K_{\mathrm{R}1} \,] \\ C_{\mathrm{cmp}}(\varepsilon) &= F_2(\varepsilon) \\ D_{\mathrm{cmp}}(\varepsilon) &= [\, 0, \quad F_1(\varepsilon) - F_2(\varepsilon) K_{\mathrm{R}1} \,]. \end{aligned}\right\} \tag{4.200}$$

STEP 4.5.C.R.4: finally, we obtain the following reduced order measurement feedback control law:

$$\left\{\begin{aligned} \dot{v} &= A_{\mathrm{cmp}}(\varepsilon) v + B_{\mathrm{cmp}}(\varepsilon) y + G_0(\varepsilon) r + \cdots + G_{\kappa-1}(\varepsilon) r^{(\kappa-1)} \\ u &= C_{\mathrm{cmp}}(\varepsilon) v + D_{\mathrm{cmp}}(\varepsilon) y + H_0(\varepsilon) r + \cdots + H_{\kappa-1}(\varepsilon) r^{(\kappa-1)} \end{aligned}\right. \tag{4.201}$$

where for $i = 0, 1, \cdots, \kappa - 1$, $G_i(\varepsilon) = (B_2 + K_{\mathrm{R}1} B_1) H_i(\varepsilon)$. This completes the constructing procedure. ◇

Theorem 4.5.3. Consider the given system in Equation (4.153) with its external disturbance $w \in L_p$, $p \in [1, \infty)$, its initial condition $x(0) = x_0$. Assume that the RPT problem is solvable for the system in Equation (4.153).

Then, for any reference signal $r(t)$, which has all its ith-order derivatives, $i = 0, 1, \cdots, \kappa - 1$, $\kappa \geq 1$, being available and $r^{(\kappa)}(t)$ being either a vector of delta functions or in L_p, the proposed RPT problem is solved by the parameterised reduced order measurement feedback control laws of Equation (4.201). ◇

4.5.2 Discrete-time Systems

We present in this subsection the RPT problem for the following discrete-time system:

$$\Sigma : \begin{cases} x(k+1) = A\,x(k) + B\,u(k) + E\,w(k), \quad x(0) = x_0, \\ y(k) \;\; = C_1\,x(k) \qquad\qquad + D_1\,w(k), \\ h(k) \;\; = C_2\,x(k) + D_2\,u(k) + D_{22}\,w(k), \end{cases} \tag{4.202}$$

where $x \in \mathbb{R}^n$ is the state, $u \in \mathbb{R}^m$ is the control input, $w \in \mathbb{R}^q$ is the external disturbance, $y \in \mathbb{R}^p$ is the measurement output, and $h \in \mathbb{R}^\ell$ is the output to be controlled. We also assume that the pair (A, B) is stabilisable and (A, C_1) is detectable. For future reference, we define Σ_P and Σ_Q to be the subsystems characterised by the matrix quadruples (A, B, C_2, D_2) and (A, E, C_1, D_1) respectively. Given the external disturbance $w \in L_p$, $p \in [1, \infty]$, and any reference signal vector $r \in \mathbb{R}^\ell$, the RPT problem for the discrete-time system in Equation (4.202) is to find a parameterised dynamic measurement feedback control law of the following form:

$$\begin{cases} v(k+1) = A_{\text{cmp}}(\varepsilon)v(k) + B_{\text{cmp}}(\varepsilon)y(k) + G(\varepsilon)r(k), \\ u(k) \;\; = C_{\text{cmp}}(\varepsilon)v(k) + D_{\text{cmp}}(\varepsilon)y(k) + H(\varepsilon)r(k), \end{cases} \tag{4.203}$$

such that, when Equation (4.203) is applied to Equation (4.202),

1. there exists an $\varepsilon^* > 0$ such that the resulting closed-loop system with $r = 0$ and $w = 0$ is asymptotically stable for all $\varepsilon \in (0, \varepsilon^*]$; and
2. let $h(k, \varepsilon)$ be the closed-loop controlled output response and let $e(k, \varepsilon)$ be the resulting tracking error, *i.e.*, $e(k, \varepsilon) := h(k, \varepsilon) - r(k)$. Then, for any initial condition of the state, $x_0 \in \mathbb{R}^n$, $\|e\|_p \to 0$ as $\varepsilon \to 0$.

It has been shown by Chen [44] that the above RPT problem is solvable for Equation (4.202) if and only if the following conditions hold:

1. (A, B) is stabilisable and (A, C_1) is detectable;

2. $D_{22} + D_2 S D_1 = 0$, where $S = -(D_2' D_2)^{\dagger} D_2' D_{22} D_1' (D_1 D_1')^{\dagger}$;
3. $\Sigma_{\rm P}$ is right invertible and of minimum phase with no infinite zeros;
4. $\mathrm{Ker}\,(C_2 + D_2 S C_1) \supset C_1^{-1}\{\mathrm{Im}\,(D_1)\}$.

It turns out that the control laws, which solve the RPT for the given plant in Equation (4.202) under the solvability conditions, need not be parameterised by any tuning parameter. Thus, Equation (4.203) can be replaced by

$$\begin{cases} v(k+1) = A_{\rm cmp} v(k) + B_{\rm cmp} y(k) + G r(k) \\ u(k) \quad = C_{\rm cmp} v(k) + D_{\rm cmp} y(k) + H r(k) \end{cases} \tag{4.204}$$

and, furthermore, the resulting tracking error $e(k)$ can be made identically zero for all $k \geq 0$.

Assume that all the solvability conditions are satisfied. We present in the following solutions to the discrete-time RPT problem.

i. State Feedback Case. When all states of the plant are measured for feedback, the problem can be solved by a static control law. We construct in this subsection a state feedback control law,

$$u = Fx + Hr, \tag{4.205}$$

that solves the RPT problem for the system in Equation (4.202). We have the following algorithm.

STEP 4.5.D.S.1: this step transforms the subsystem from u to h of the given system in Equation (4.202) into the special coordinate basis of Theorem 3.4.1, *i.e.*, finds nonsingular state, input and output transformations $\Gamma_{\rm s}$, $\Gamma_{\rm i}$ and $\Gamma_{\rm o}$ to put it into the structural form of Theorem 3.4.1 as well as in the compact form of Equations (3.29) to (3.32), *i.e.*,

$$\tilde{A} = \Gamma_{\rm s}^{-1}(A - B_0 C_{2,0})\Gamma_{\rm s} = \begin{bmatrix} A_{aa}^{-} & 0 \\ B_c E_{ca}^{-} & A_{cc} \end{bmatrix}, \tag{4.206}$$

$$\tilde{B} = \Gamma_{\rm s}^{-1} B \Gamma_{\rm i} = \Gamma_{\rm s}^{-1} \begin{bmatrix} B_0 & B_1 \end{bmatrix} \Gamma_{\rm i} = \begin{bmatrix} B_{0a}^{-} & 0 \\ B_{0c} & B_c \end{bmatrix}, \tag{4.207}$$

$$\tilde{D}_2 = \Gamma_{\rm o}^{-1} D_2 \Gamma_{\rm i} = \begin{bmatrix} I_{m_0} & 0 \end{bmatrix}, \tag{4.208}$$

$$\tilde{C}_2 = \Gamma_{\rm o}^{-1} C_2 \Gamma_{\rm s} = \Gamma_{\rm o}^{-1} C_{2,0} \Gamma_{\rm s} = \begin{bmatrix} C_{2,0a}^{-} & C_{2,0c} \end{bmatrix}. \tag{4.209}$$

STEP 4.5.D.S.2: choose an appropriate dimensional matrix F_c such that

$$A_{cc}^{c} = A_{cc} - B_c F_c \tag{4.210}$$

is asymptotically stable. The existence of such an F_c is guaranteed by the property that (A_{cc}, B_c) is completely controllable.

STEP 4.5.D.S.3: finally, we let

$$F = -\Gamma_{\mathrm{i}}^{-} \begin{bmatrix} C_{2,0a}^{-} & C_{2,0c} \\ E_{ca}^{-} & F_c \end{bmatrix} \Gamma_{\mathrm{s}}^{-1} \quad \text{and} \quad H = \Gamma_{\mathrm{i}}^{-} \begin{bmatrix} I \\ 0 \end{bmatrix} \Gamma_{\mathrm{o}}^{-1}. \tag{4.211}$$

This ends the constructive algorithm. ◇

We have the following result.

Theorem 4.5.4. Consider the given discrete-time system in Equation (4.202) with any external disturbance $w(k)$ and any initial condition $x(0)$. Assume that all its states are measured for feedback, *i.e.*, $C_1 = I$ and $D_1 = 0$, and the solvability conditions for the RPT problem hold. Then, for any reference signal $r(k)$, the proposed RPT problem is solved by the control law of Equation (4.205) with F and H as given in Equation (4.211). ◇

ii. Measurement Feedback Case. Without loss of generality, we assume throughout this subsection that matrix $D_{22} = 0$. If it is nonzero, it can always be washed out by the following pre-output feedback $u(k) = Sy(k)$. It turns out that, for discrete-time systems, the full-order observer-based control law is not capable of achieving the RPT performance, because there is a delay of one step in the observer itself. Thus, we will focus on the construction of a reduced order measurement feedback control law to solve the RPT problem. For simplicity of presentation, we assume that matrices C_1 and D_1 have already been transformed into the following forms,

$$C_1 = \begin{bmatrix} 0 & C_{1,02} \\ I_\kappa & 0 \end{bmatrix} \quad \text{and} \quad D_1 = \begin{bmatrix} D_{1,0} \\ 0 \end{bmatrix}, \tag{4.212}$$

where $D_{1,0}$ is of full row rank. Before we present a step-by-step algorithm to construct a reduced order measurement feedback controller, we first partition the following system

$$\begin{cases} x(k+1) = A\ x(k) + B\ u(k) + [E \quad I_n]\ \tilde{w}(k) \\ y(k) \quad = C_1\ x(k) \qquad\qquad + [D_1 \quad 0]\ \tilde{w}(k) \end{cases} \tag{4.213}$$

in conformity with the structures of C_1 and D_1 in Equation (4.212), *i.e.*,

$$\begin{cases} \begin{pmatrix} \delta(x_1) \\ \delta(x_2) \end{pmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} u + \begin{bmatrix} E_1 & I_\kappa & 0 \\ E_2 & 0 & I_{n-\kappa} \end{bmatrix} \tilde{w} \\ \begin{pmatrix} y_0 \\ y_1 \end{pmatrix} = \begin{bmatrix} 0 & C_{1,02} \\ I_\kappa & 0 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{bmatrix} D_{1,0} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tilde{w} \end{cases}$$

where $\delta(x_1) = x_1(k+1)$ and $\delta(x_2) = x_2(k+1)$. Obviously, $y_1 = x_1$ is directly available and hence need not be estimated. Next, let Σ_{QR} be characterised by

$$(A_{\text{R}}, E_{\text{R}}, C_{\text{R}}, D_{\text{R}}) = \left(A_{22}, [E_2 \quad 0 \quad I_{n-\kappa}], \begin{bmatrix} C_{1,02} \\ A_{12} \end{bmatrix}, \begin{bmatrix} D_{1,0} & 0 & 0 \\ E_1 & I_\kappa & 0 \end{bmatrix} \right).$$

It is straightforward to verify that Σ_{QR} is right invertible with no finite and infinite zeros. Moreover, $(A_{\text{R}}, C_{\text{R}})$ is detectable if and only if (A, C_1) is detectable. We are ready to present the following algorithm.

STEP 4.5.D.R.1: for the given system in Equation (4.202), we again assume that all the state variables of Equation (4.202) are measurable and then follow Steps 4.5.D.S.1 to 4.5.D.S.3 of the algorithm of the previous subsection to construct gain matrices F and H. We also partition F in conformity with x_1 and x_2 as follows:

$$F = [F_1 \quad F_2]. \tag{4.214}$$

STEP 4.5.D.R.2: let K_{R} be an appropriate dimensional constant matrix such that the eigenvalues of

$$A_{\text{R}} + K_{\text{R}} C_{\text{R}} = A_{22} + [K_{\text{R0}} \quad K_{\text{R1}}] \begin{bmatrix} C_{1,02} \\ A_{12} \end{bmatrix} \tag{4.215}$$

are all in $\mathbb{C}^{\odot}$. This can be done because $(A_{\text{R}}, C_{\text{R}})$ is detectable.

STEP 4.5.D.R.3: let

$$G_{\text{R}} = [-K_{\text{R0}}, \quad A_{21} + K_{\text{R1}} A_{11} - (A_{\text{R}} + K_{\text{R}} C_{\text{R}}) K_{\text{R1}}], \tag{4.216}$$

$$\left.\begin{aligned} A_{\text{cmp}} &= A_{\text{R}} + B_2 F_2 + K_{\text{R}} C_{\text{R}} + K_{\text{R1}} B_1 F_2, \\ B_{\text{cmp}} &= G_{\text{R}} + (B_2 + K_{\text{R1}} B_1)[0, \quad F_1 - F_2 K_{\text{R1}}], \\ C_{\text{cmp}} &= F_2, \\ D_{\text{cmp}} &= [0, \quad F_1 - F_2 K_{\text{R1}}], \end{aligned}\right\} \tag{4.217}$$

$$G = (B_2 + K_{\text{R1}} B_1) H. \tag{4.218}$$

STEP 4.5.D.R.4: finally, we obtain the following reduced order measurement feedback control law:

$$\begin{cases} v(k+1) = A_{\text{cmp}}\, v(k) + B_{\text{cmp}}\, y(k) + G\, r(k), \\ u(k) \quad = C_{\text{cmp}}\, v(k) + D_{\text{cmp}}\, y(k) + H\, r(k). \end{cases} \tag{4.219}$$

This completes the algorithm. ◊

Theorem 4.5.5. Consider the given system in Equation (4.202) with any external disturbance $w(k)$ and any initial condition $x(0)$. Assume that the solvability conditions for the RPT problem hold. Then, for any reference signal $r(k)$, the proposed RPT problem is solved by the reduced order measurement feedback control laws of Equation (4.219). ◊

As has been seen in the previous section, the solvability conditions for the RPT problem, especially the restriction on the infinite zeros of the given system, are too strong to be satisfied in most practical situations. We introduce in this section a modified problem, the so-called almost perfect tracking problem, which can be solved for a much larger class of discrete-time systems with any infinite zero structure. This modified formulation will yield internally stabilising control laws that are capable of tracking reference signal $r(k)$ with some delays. If we know the reference signal a few steps ahead, the modified tracking control scheme will then track the reference precisely after certain steps.

For simplicity, we consider in the following the discrete-time system of Equation (4.202) without external disturbances, *i.e.*,

$$\Sigma : \begin{cases} x(k+1) = A\, x(k) + B\, u(k), \qquad x(0) = x_0, \\ y(k) \quad = C_1\, x(k), \\ h(k) \quad = C_2\, x(k) + D_2\, u(k). \end{cases} \tag{4.220}$$

Let us first consider the reference $r(k) \in \mathbb{R}^\ell$ to be tracked is a known vector sequence, which implies that $r(k+d)$, $0 \le d \le \kappa_d$, is known for some integer $\kappa_d \ge 0$. This is a quite reasonable assumption in most practical situations when one wants to track references such as step functions, ramp functions and sinusoidal functions. We will later deal with the case when $r(k+d)$, $d > 0$, is unknown. We are ready to define formally the almost perfect tracking problem. Given the discrete-time system in Equation (4.220) with $x(0) = x_0$ and the reference $r(k)$ with $r(k+d)$, $0 \le d \le \kappa_d$, being known for a nonnegative integer κ_d, the (κ_d, κ_0) almost perfect tracking problem, where κ_0 is another nonnegative integer, is to find a dynamic measurement feedback control law of the following form:

$$\begin{cases} v(k+1) = A_{\text{cmp}}v(k) + B_{\text{cmp}}y(k) + G_0 r(k) + \cdots + G_{\kappa_d} r(k+\kappa_d) \\ u(k) \quad = C_{\text{cmp}}v(k) + D_{\text{cmp}}y(k) + H_0 r(k) + \cdots + H_{\kappa_d} r(k+\kappa_d) \end{cases} \tag{4.221}$$

such that, when Equation (4.221) is applied to Equation (4.220),

1. the resulting closed-loop system is internally stable; and
2. for any initial condition $x_0 \in \mathbb{R}^n$, the resulting tracking error satisfies:

$$J(x_0, u, \kappa_d, \kappa_0) := \sum_{k=\kappa_0}^{\infty} |e(k)| = 0, \tag{4.222}$$

i.e., $e(k) = 0$, or $h(k) = r(k)$, for all $k \geq \kappa_0$.

We have the following theorem.

Theorem 4.5.6. Consider the discrete-time plant in Equation (4.220) with $x(0) = x_0$, and with (i) (A, B) being stabilisable and (A, C_1) being observable; and (ii) Σ_{P} being right invertible and of minimum phase. Let the infinite zero structure (see Chapter 3 for its definition) of Σ_{P} be given as $S_\infty^\star(\Sigma_{\text{P}}) = \{q_1, \cdots, q_{m_d}\}$, with $q_1 \leq \cdots \leq q_{m_d}$, and let the controllability index of the pair (A', C_1') be $\mathcal{C} = \{k_1, \cdots, k_{\text{p}}\}$, with $k_1 \leq \cdots \leq k_{\text{p}}$. Then, the (κ_d, κ_0) almost perfect tracking problem is solvable for any reference with $\kappa_0 = q_{m_d} + k_{\text{p}} - 1$ and $\kappa_d = q_{m_d}$. ◇

Proof. We prove this theorem by explicitly constructing the required control law. Let us first construct the special coordinate basis of Σ_{P}. It follows from Theorem 3.4.1 that there exist nonsingular state, output and input transformations Γ_{s}, Γ_{o} and Γ_{i}, which will take Σ_{P} into the standard format of the special coordinate basis, *i.e.*,

$$x = \Gamma_{\text{s}}\tilde{x}, \quad h = \Gamma_{\text{o}}\tilde{h}, \quad u = \Gamma_{\text{i}}\tilde{u}, \quad r = \Gamma_{\text{o}}\tilde{r}, \tag{4.223}$$

$$\tilde{x} = \begin{pmatrix} x_a^- \\ x_c \\ x_d \end{pmatrix}, \quad \tilde{h} = \begin{pmatrix} h_0 \\ h_d \end{pmatrix}, \quad \tilde{u} = \begin{pmatrix} u_0 \\ u_d \\ u_c \end{pmatrix}, \quad \tilde{r} = \begin{pmatrix} r_0 \\ r_d \end{pmatrix}, \tag{4.224}$$

$$x_d = \begin{pmatrix} x_1 \\ \vdots \\ x_{m_d} \end{pmatrix}, \quad x_i = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{iq_i} \end{pmatrix}, \quad h_d = \begin{pmatrix} h_1 \\ \vdots \\ h_{m_d} \end{pmatrix}, \tag{4.225}$$

$$r_d = \begin{pmatrix} r_1 \\ \vdots \\ r_{m_d} \end{pmatrix}, \quad u_d = \begin{pmatrix} u_1 \\ \vdots \\ u_{m_d} \end{pmatrix}, \tag{4.226}$$

$$\delta(x_a^-) = A_{aa}^- x_a^- + B_{0a}^- h_0 + L_{ad}^- h_d, \tag{4.227}$$

$$\delta(x_c) = A_{cc}x_c + B_{0c}h_0 + L_{cd}h_d + B_c E_{ca}^- x_a^- + B_c u_c, \tag{4.228}$$

$$h_0 = C_{2,0a}^- x_a^- + C_{2,0c}x_c + C_{2,0d}x_d + u_0, \quad u_0 \in \mathbb{R}^{m_0}, \tag{4.229}$$

and for each $i = 1, \cdots, m_d$, $x_i \in \mathbb{R}^{q_i}$ and

$$\delta(x_i) = A_{q_i}x_i + L_{i0}h_0 + L_{id}h_d + B_{q_i}\left[u_i + E_{ia}^- x_a^- + E_{ic}x_c + \sum_{j=1}^{m_d} E_{ij}x_j\right] \tag{4.230}$$

$$h_i = C_{q_i}x_i = x_{i1}, \quad h_d = C_d x_d, \tag{4.231}$$

where $\delta(\star) = \star(k+1)$; the triple $(A_{q_i}, B_{q_i}, C_{q_i})$ has the special structure as given in Equation (3.25). It follows from Theorem 3.4.1 that L_{id} has the special format, $L_{id} = [\, L_{i1} \quad L_{i2} \quad \cdots \quad L_{ii-1} \quad 0 \quad \cdots \quad 0\,]$, with its last row always being identically zero. Next, we partition L_{i0} and L_{id}, $i = 1, \cdots, m_d$, as follows:

$$L_{i0} = \begin{bmatrix} L_{i0,1} \\ \vdots \\ L_{i0,q_i} \end{bmatrix}, \quad L_{id} = \begin{bmatrix} L_{id,1} \\ \vdots \\ L_{id,q_i} \end{bmatrix}, \tag{4.232}$$

and define a new controlled output,

$$\tilde{h}_{\mathrm{n}}(k) = \begin{bmatrix} h_0(k) \\ h_1(k+q_1) - \sum_{j=1}^{q_1} \begin{bmatrix} L_{10,j} & L_{1d,j} \end{bmatrix} \tilde{h}(k+q_1-j) \\ \vdots \\ h_{m_d}(k+q_{m_d}) - \sum_{j=1}^{q_{m_d}} \begin{bmatrix} L_{m_d 0,j} & L_{m_d d,j} \end{bmatrix} \tilde{h}(k+q_{m_d}-j) \end{bmatrix}. \tag{4.233}$$

Then, it is straightforward to verify that $\tilde{y}_{\mathrm{n}}$ can be expressed as

$$\tilde{h}_{\mathrm{n}}(k) = \check{C}_2 \tilde{x}(k) + \check{D}_2 \tilde{u}(k), \tag{4.234}$$

with

$$\check{C}_2 = \begin{bmatrix} C_{2,0a}^- & C_{2,0c} & C_{2,0d} \\ E_{da}^- & E_{dc} & E_{dd} \end{bmatrix} \quad \text{and} \quad \check{D}_2 = \begin{bmatrix} I_{m_0} & 0 & 0 \\ 0 & I_{m_d} & 0 \end{bmatrix}, \tag{4.235}$$

where

$$E_{da}^{-}=\begin{bmatrix} E_{1a}^{-} \\ \vdots \\ E_{m_d a}^{-} \end{bmatrix},\ E_{dc}=\begin{bmatrix} E_{1c} \\ \vdots \\ E_{m_d c} \end{bmatrix},\ E_{dd}=\begin{bmatrix} E_{11} & \cdots & E_{1m_d} \\ \vdots & \ddots & \vdots \\ E_{m_d 1} & \cdots & E_{m_d m_d} \end{bmatrix}. \quad (4.236)$$

Let $\tilde{A}=\Gamma_{\mathrm{s}}^{-1}A\Gamma_{\mathrm{s}}$, $\tilde{B}=\Gamma_{\mathrm{s}}^{-1}B\Gamma_{\mathrm{i}}$, and let $\tilde{\Sigma}_{\mathrm{P}}$ characterised by $(\tilde{A},\tilde{B},\check{C}_2,\check{D}_2)$. It is simple to show that the auxiliary system $\tilde{\Sigma}_{\mathrm{P}}$ is right invertible and of minimum phase with no infinite zeros.

We first assume that $C_1=I$ and follow Steps 4.5.D.S.1 to 4.5.D.S.3 to obtain a state feedback control law

$$\tilde{u}(k)=\tilde{F}\tilde{x}(k)+\tilde{H}\tilde{r}_{\mathrm{n}}(k), \quad (4.237)$$

where

$$\tilde{r}_{\mathrm{n}}(k)=\begin{bmatrix} r_0(k) \\ r_1(k+q_1)-\sum_{j=1}^{q_1}\begin{bmatrix} L_{10,j} & L_{1d,j} \end{bmatrix}\tilde{r}(k+q_1-j) \\ \vdots \\ r_{m_d}(k+q_{m_d})-\sum_{j=1}^{q_{m_d}}\begin{bmatrix} L_{m_d 0,j} & L_{m_d d,j} \end{bmatrix}\tilde{r}(k+q_{m_d}-j) \end{bmatrix}, \quad (4.238)$$

which has the following properties: (i) $\tilde{A}+\tilde{B}\tilde{F}$ is asymptotically stable, and (ii) the resulting $\tilde{h}_{\mathrm{n}}(k)=\tilde{r}_{\mathrm{n}}(k)$. This implies that the actual controlled output h is capable of precisely tracking the given reference $r(k)$ after q_{m_d} steps. Rewriting Equation (4.237), we obtain

$$\begin{aligned} u(k)&=\Gamma_{\mathrm{i}}\tilde{u}(k)=\Gamma_{\mathrm{i}}\left[\tilde{F}\tilde{x}(k)+\tilde{H}\tilde{r}_{\mathrm{n}}(k)\right] \\ &=\Gamma_{\mathrm{i}}\left[\tilde{F}\tilde{x}(k)+\tilde{H}L_0\tilde{r}(k)+\cdots+\tilde{H}L_{m_d}\tilde{r}(k+q_{m_d})\right] \end{aligned} \quad (4.239)$$

for some appropriate matrices $L_0, L_1, \cdots, L_{m_d}$. Let $F=\Gamma_{\mathrm{i}}\tilde{F}\Gamma_{\mathrm{s}}^{-1}$, and

$$H_j=\Gamma_{\mathrm{i}}\tilde{H}L_j\Gamma_{\mathrm{o}}^{-1}, \quad (4.240)$$

for $j=0,1,\cdots,m_d$. We have

$$u(k)=Fx(k)+H_0r(k)+\cdots+H_{m_d}r(k+q_{m_d}). \quad (4.241)$$

Next, we proceed to construct a reduced order measurement feedback controller. We follow Steps 4.5.D.R.1 to 4.5.D.R.3 to obtain matrices B_1, B_2, F_1, F_2 and gain matrices A_{cmp}, B_{cmp}, C_{cmp} and D_{cmp} as given in Equation (4.217). Noting that the pair (A, C_1) is observable and (A', C_1') has a controllability index $\{k_1, \cdots, k_p\}$, it is simple to show that (A_R, C_R) is also observable and the controllability index of (A_R', C_R') is given by $\{k_1-1, \cdots, k_p-1\}$. It follows from Theorem 3.3.1 that there exists a gain matrix K_R such that $A_R + K_R C_R$ has all its eigenvalues at the origin and

$$(A_R + K_R C_R)^{k_p-2} \equiv 0. \tag{4.242}$$

We thus choose such a K_R in constructing gain matrices A_{cmp}, B_{cmp}, C_{cmp} and D_{cmp}. The reduced order measurement feedback law is then given by

$$\left.\begin{aligned} v(k+1) &= A_{\text{cmp}} v(k) + B_{\text{cmp}} y(k) + \sum_{j=0}^{m_d} G_j r(k+j) \\ u(k) &= C_{\text{cmp}} v(k) + D_{\text{cmp}} y(k) + \sum_{j=0}^{m_d} H_j r(k+j) \end{aligned}\right\} \tag{4.243}$$

where $G_j = (B_2 + K_R B_1) H_j$, $i = 0, 1, \cdots, m_d$. It can be further shown that such a controller solves the (κ_d, κ_0) almost perfect tracking problem for Equation (4.220) with $\kappa_d = q_{m_d}$ and $\kappa_0 = q_{m_d} + k_p - 1$. The result follows. ◇

Remark 4.5.3. Consider the given plant in Equation (4.220) which has all properties as stated in Theorem 4.5.6. Then, the (κ_d, κ_0) almost perfect tracking problem is solvable by a full order measurement feedback controller of the form in Equation (4.221) with $\kappa_d = q_{m_d}$ and $\kappa_0 = q_{m_d} + k_p$. The required control law is given by

$$\left.\begin{aligned} v(k+1) &= A_{\text{cmp}} v(k) + \sum_{j=0}^{m_d} B H_j r(k+j) - K y(k) \\ u(k) &= F v(k) + \sum_{j=0}^{m_d} H_j r(k+j) \end{aligned}\right\} \tag{4.244}$$

where $A_{\text{cmp}} = A + BF + KC_1$ with K being chosen such that all the eigenvalues of $A + KC_1$ are at the origin and $(A + KC_1)^{k_p-1} = 0$. ◇

Remark 4.5.4. For simplicity, we consider Σ_P being a single output system, *i.e.*, $\ell = 1$, with a relative degree q_1. Clearly, if the reference $r(k+d)$ is unknown for all $d > 0$, then the full order measurement feedback controller

(4.244) with $r(k+\star)$ being replaced by $r(k)$ will be capable of tracking the reference with a delay of q_1 steps after q_1+k_p initial steps. Similarly, under the same situation, the reduced order measurement feedback controller in Equation (4.243) with $r(k+\star)$ being replaced by $r(k)$ will track the reference with a delay of q_1 steps after q_1+k_p-1 initial steps. $\diamondsuit$

4.6 Loop Transfer Recovery Technique

Another popular design methodology for multivariable systems, which is based on 'loop shaping' concept, is linear quadratic Gaussian (LQG) with loop transfer recovery (LTR). It involves two separate designs of a state feedback controller and an observer or an estimator. The exact design procedure depends on the point where the unstructured uncertainties are modelled and where the loop is broken to evaluate the open-loop transfer matrices. Commonly, either the input point or the output point of the plant is taken as such a point. We should focus on the case when the loop is broken at the input point of the plant. The required results for the output point can be easily obtained by appropriate dualisation. Thus, in the two-step procedure of LQG/LTR, the first step of design involves loop shaping by a state feedback design to obtain an appropriate loop transfer function, called the target loop transfer function. Such a loop shaping is an engineering art and often involves the use of linear quadratic regulator (LQR) design, in which the cost matrices are used as free design parameters to generate the target loop transfer function, and thus the desired sensitivity and complementary sensitivity functions. However, when such a feedback design is implemented via an observer-based controller (or Kalman filter) that uses only the measurement feedback, the loop transfer function obtained, in general, is not the same as the target loop transfer function, unless proper care is taken in designing the observers. This is when the second step of LQG/LTR design philosophy comes into picture. In this step, the required observer design is attempted so as to recover the loop transfer function of the full state feedback controller. This second step is known as LTR.

The topic of LTR was heavily studied in 1980s. Major contributions came from [86–96]. We present in the following the methods of LTR design at both the input point and output point of the given plant.

4.6.1 LTR at Input Point

It turns out that it is very simple to formulate the LTR design technique for both continuous- and discrete-time systems into a single framework. Thus,

we will do it in one shot. Let us consider a linear time-invariant multivariable system characterised by

$$\Sigma : \begin{cases} \delta(x) = A\,x + B\,u \\ y \quad = C\,x + D\,u \end{cases} \tag{4.245}$$

where $\delta(x) = \dot{x}(t)$, if Σ is a continuous-time system, or $\delta(x) = x(k+1)$, if Σ is a discrete-time system. Similarly, $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$ and $y \in \mathbb{R}^p$ are the state, input and output of Σ. They represent respectively $x(t)$, $u(t)$ and $y(t)$ if the given system is of continuous-time, or represent respectively $x(k)$, $u(k)$ and $y(k)$ if Σ is of discrete-time. Without loss of any generality, we assume throughout this section that both $[\,B' \;\; D'\,]$ and $[\,C \;\; D\,]$ are of full rank. The transfer function of Σ is then given by

$$P(\varsigma) = C(\varsigma I - A)^{-1}B + D, \tag{4.246}$$

where $\varsigma = s$, the Laplace transform operator, if Σ is of continuous-time, or $\varsigma = z$, the z-transform operator, if Σ is of discrete-time.

As mentioned earlier, there are two steps involved in LQG/LTR design. In the first step, we assume that all state variables of Equation (4.245) are available and design a full state feedback control law

$$u = -Fx \tag{4.247}$$

such that

1. the closed-loop system is asymptotically stable, and
2. the open-loop transfer function when the loop is broken at the input point of the given system, *i.e.*,

$$L_{\mathrm{t}}(\varsigma) = F(\varsigma I - A)^{-1}B, \tag{4.248}$$

meets some frequency-dependent specifications.

Arriving at an appropriate value for F is concerned with the issue of loop shaping, which often includes the use of LQR design in which the cost matrices are used as free design parameters to generate $L_{\mathrm{t}}(\varsigma)$ that satisfies the given specifications.

To be more specific, if Σ is a continuous-time system, the target loop transfer function $L_{\mathrm{t}}(s)$ can be generated by minimising the following cost function:

$$J_{\mathrm{C}} = \int_0^{\infty} (x'Qx + u'Ru)\ \mathrm{d}t, \tag{4.249}$$

where $Q \geq 0$ and $R > 0$ are free design parameters provided that $(A, Q^{1/2})$ has no unobservable modes on the imaginary axis. The solution to the above problem is given by

$$F = R^{-1}B'P, \tag{4.250}$$

where $P \geq 0$ is the stabilising solution of the following algebraic Riccati equation (ARE):

$$PA + A'P - PBR^{-1}B'P + Q = 0. \tag{4.251}$$

It is known in the literature that a target loop transfer function $L_{\mathrm{t}}(s)$ with F be given as in Equation (4.250) has a phase margin greater than 60° and an infinite gain margin.

Similarly, if Σ is a discrete-time system, we can generate a target loop transfer function $L_{\mathrm{t}}(z)$ by minimising

$$J_{\mathrm{D}} = \sum_{k=0}^{\infty} \Big(x'(k)Qx(k) + u'(k)Ru(k) \Big) \tag{4.252}$$

where $Q \geq 0$ and $R > 0$ are free design parameters provided that $(A, Q^{1/2})$ has no unobservable modes on the unit circle.

$$F = (R + B'PB)^{-1}B'PA, \tag{4.253}$$

where $P \geq 0$ is the stabilising solution of the following ARE:

$$P = A'PA - A'PB(R + B'PB)^{-1}B'PA + Q. \tag{4.254}$$

Unfortunately, there are no guaranteed phase and gain margins for the target loop transfer function $L_{\mathrm{t}}(z)$ resulting from the discrete-time linear quadratic regulator.

Generally, it is unreasonable to assume that all the state variables of a given system can be measured. Thus, we will have to implement the control law obtained in the first step by a measurement feedback controller. The technique of LTR is to design an appropriate measurement feedback control (see Figure 4.3) such that the resulting system is asymptotically stable and the achieved

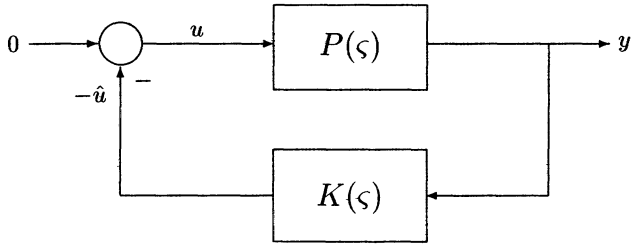


Figure 4.3. Plant-controller closed-loop configuration.

open-loop transfer function $L_a(\varsigma)$ from u to $-\hat{u}$ is either exactly or approximately matched with the target loop transfer function $L_t(\varsigma)$ obtained in the first step. In this way, all the nice properties associated with the target loop transfer function can be recovered by the measurement feedback controller. This is the so-called LTR design.

It is simple to observe that the achieved open-loop transfer function in the configuration of Figure 4.3 is given by

$$L_a(\varsigma) = K(\varsigma)P(\varsigma). \tag{4.255}$$

Let us define *recovery error* as

$$E(\varsigma) := L_t(\varsigma) - L_a(\varsigma) = F(\varsigma I - A)^{-1}B - K(\varsigma)P(\varsigma). \tag{4.256}$$

The LTR technique is to design an appropriate stabilising $K(\varsigma)$ such that the recovery error $E(\varsigma)$ is either identically zero or small in a certain sense. As usual, two commonly used structures for $K(\varsigma)$ are: (1) the full order observer-based controller, and (2) the reduced order observer-based controller.

i. Full Order Observer-Based Controller. The dynamic equations of a full order observer-based controller are well known and are given by

$$\begin{cases} \delta(\hat{x}) = A\hat{x} + Bu + K_f(y - C\hat{x} - Du), \\ u = -F\hat{x}, \end{cases} \tag{4.257}$$

where K_f is the full order observer gain matrix and is the only free design parameter. It should be chosen so that $A - K_f C$ is asymptotically stable. The transfer function of the full order observer-based control is given by

$$K(\varsigma) = F(\varsigma I - A + BF + K_f C - K_f DF)^{-1}K_f. \tag{4.258}$$

It has been shown [87, 94] that the recovery error resulting from the full order observer-based controller can be expressed as

$$E(\varsigma) = M_{\mathrm{f}}(\varsigma)[I + M_{\mathrm{f}}(\varsigma)]^{-1}[I + F(\varsigma I - A)^{-1}B], \tag{4.259}$$

where

$$M_{\mathrm{f}}(\varsigma) = F(\varsigma I - A + K_{\mathrm{f}}C)^{-1}(B - K_{\mathrm{f}}D). \tag{4.260}$$

Obviously, in order to render $E(\varsigma)$ to be zero or small, one will have to design an observer gain K_{f} such that $M_{\mathrm{f}}(\varsigma)$, or equivalently $M'_{\mathrm{f}}(\varsigma)$, is zero or small (in certain sense). Defining an auxiliary system,

$$\begin{cases} \dot{\tilde{x}} = A'\,\tilde{x} + C'\,\tilde{u} + F'\,\tilde{w} \\ \tilde{y} = \quad \tilde{x} \\ \tilde{h} = B'\,\tilde{x} + D'\,\tilde{u} \end{cases} \tag{4.261}$$

with a state feedback control law,

$$\tilde{u} = -K'_{\mathrm{f}}\,\tilde{x}. \tag{4.262}$$

It is straightforward to verify that the closed-loop transfer matrix from $\tilde{w}$ to $\tilde{h}$ of the above system is equivalent to $M'_{\mathrm{f}}(\varsigma)$. As such, any of the methods presented in Sections 4.3 and 4.4 for H_2 and H_∞ optimal control can be utilised to find K_{f} to minimise either the H_2-norm or H_∞-norm of $M'_{\mathrm{f}}(\varsigma)$. In particular,

1. if the given plant Σ is a continuous-time system and if Σ is left invertible and of minimum phase,
2. if the given plant Σ is a discrete-time system and if Σ is left invertible and of minimum phase with no infinite zeros,

then either the H_2-norm or H_∞-norm of $M_{\mathrm{f}}(\varsigma)$ can be made arbitrarily small, and hence LTR can be achieved. If these conditions are not satisfied, the target loop transfer function $L_{\mathrm{t}}(\varsigma)$, in general, cannot be fully recovered!

For the case when the target loop transfer function can be approximately recovered, the following full order Chen–Saberi–Sannuti (CSS) architecture-based control law (see [88, 94]),

$$\begin{cases} \delta(v) = Av + K_{\mathrm{f}}(y - Cv), \\ u \quad = -Fv, \end{cases} \tag{4.263}$$

which has a resulting recovery error,

$$E(\varsigma) = M_{\mathrm{f}}(\varsigma) = F(\varsigma I - A + K_{\mathrm{f}}C)^{-1}(B - K_{\mathrm{f}}D), \tag{4.264}$$

can be utilised to recover the target loop transfer function as well. In fact, when the same gain matrix K_{f} is used, the full order CSS architecture-based controller would yield a much better recovery compared to that of the full order observer-based controller.

ii. Reduced Order Observer-Based Controller. For simplicity, we assume that C and D have already been transformed into the form

$$C = \begin{bmatrix} 0 & C_{02} \\ I & 0 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} D_0 \\ 0 \end{bmatrix}, \tag{4.265}$$

where D_0 is of full row rank. Then, the dynamic equations of Σ can be partitioned as follows:

$$\begin{cases} \begin{pmatrix} \delta(x_1) \\ \delta(x_2) \end{pmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} u, \\ \begin{pmatrix} y_0 \\ y_1 \end{pmatrix} = \begin{bmatrix} 0 & C_{02} \\ I & 0 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{bmatrix} D_0 \\ 0 \end{bmatrix} u, \end{cases} \tag{4.266}$$

where x_1 is readily accessible. Let

$$A_{\mathrm{r}} = A_{22}, \quad B_{\mathrm{r}} = B_2, \quad C_{\mathrm{r}} = \begin{bmatrix} C_{02} \\ A_{12} \end{bmatrix}, \quad D_{\mathrm{r}} = \begin{bmatrix} D_0 \\ B_1 \end{bmatrix}, \tag{4.267}$$

and the reduced order observer gain matrix K_{r} be such that $A_{\mathrm{r}} - K_{\mathrm{r}}C_{\mathrm{r}}$ is asymptotically stable. Next, we partition

$$F = [\, F_1 \quad F_2 \,], \quad K_{\mathrm{r}} = [\, K_{\mathrm{r}0} \quad K_{\mathrm{r}1} \,], \tag{4.268}$$

in conformity with the partitions of $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and $y = \begin{pmatrix} y_0 \\ y_1 \end{pmatrix}$ respectively. Then, define

$$G_{\mathrm{r}} = [\, K_{\mathrm{r}0}, \quad A_{21} - K_{\mathrm{r}1}A_{11} + (A_{\mathrm{r}} - K_{\mathrm{r}}C_{\mathrm{r}})K_{\mathrm{r}1} \,]. \tag{4.269}$$

The reduced order observer-based controller is given by

$$\begin{cases} \delta(v) = (A_{\mathrm{r}} - K_{\mathrm{r}}C_{\mathrm{r}})v + (B_{\mathrm{r}} - K_{\mathrm{r}}D_{\mathrm{r}})u + G_{\mathrm{r}}y, \\ u = -F_2 v - [\, 0, \quad F_1 + F_2 K_{\mathrm{r}1} \,]\, y. \end{cases} \tag{4.270}$$

It is again reported in [87, 94] that the recovery error resulting from the reduced order observer-based controller can be expressed as

$$E(\varsigma) = M_{\rm r}(\varsigma)[I + M_{\rm r}(\varsigma)]^{-1}[I + F(\varsigma I - A)^{-1}B], \tag{4.271}$$

where

$$M_{\rm r}(\varsigma) = F_2(\varsigma I - A_{\rm r} + K_{\rm r}C_{\rm r})^{-1}(B_{\rm r} - K_{\rm r}D_{\rm r}). \tag{4.272}$$

Thus, making $E(\varsigma)$ to be zero or small is equivalent to design a reduced order observer gain $K_{\rm r}$ such that $M_{\rm r}(\varsigma)$, or equivalently $M_{\rm r}'(\varsigma)$, is zero or small. Following the same idea as in the full order case, we define an auxiliary system

$$\begin{cases} \dot{\tilde{x}} = A_{\rm r}'\,\tilde{x} + C_{\rm r}'\,\tilde{u} + F_2'\,\tilde{w} \\ \tilde{y} = \quad \tilde{x} \\ \tilde{h} = B_{\rm r}'\,\tilde{x} + D_{\rm r}'\,\tilde{u} \end{cases} \tag{4.273}$$

with a state feedback control law,

$$\tilde{u} = -K_{\rm r}'\,\tilde{x}. \tag{4.274}$$

Obviously, the closed-loop transfer matrix from $\tilde{w}$ to $\tilde{h}$ of the above system is equivalent to $M_{\rm r}'(\varsigma)$. Hence, the methods of Sections 4.3 and 4.4 for H_2 and H_∞ optimal control again can be used to find $K_{\rm r}$ to minimise either the H_2-norm or H_∞-norm of $M_{\rm r}'(\varsigma)$. In particular, for the case when Σ satisfies Condition 1 (for continuous-time systems) or Condition 2 (for discrete-time systems) stated in the full order case, the target loop can be either exactly or approximately recovered. In fact, in this case, the following reduced order CSS architecture-based controller

$$\begin{cases} \delta(v) = (A_{\rm r} - K_{\rm r}C_{\rm r})\,v + \quad G_{\rm r} \quad y, \\ u \quad = \quad -F_2 \quad v - [\,0, \quad F_1 + F_2K_{\rm r1}\,]\,y, \end{cases} \tag{4.275}$$

which has a resulting recovery error,

$$E(\varsigma) = M_{\rm r}(\varsigma) = F_2(\varsigma I - A_{\rm r} + K_{\rm r}C_{\rm r})^{-1}(B_{\rm r} - K_{\rm r}D_{\rm r}), \tag{4.276}$$

can also be used to recover the given target loop transfer function. Again, when the same $K_{\rm r}$ is used, the reduced order CSS architecture-based controller would yield a better recovery compared to that of the reduced order observer-based controller (see [88, 94]).

4.6.2 LTR at Output Point

For the case when uncertainties of the given plant are modelled at the output point, the following dualisation procedure can be used to find appropriate solutions. The basic idea is to convert the LTR design at the output point of the given plant into an equivalent LTR problem at the input point of an auxiliary system so that all the methods studied in the previous subsection can be readily applied.

1. Consider a plant Σ characterised by the quadruple (A, B, C, D). Let us design a Kalman filter or an observer first with a Kalman filter or observer gain matrix K_f such that $A - K_f C$ is asymptotically stable and the resulting target loop

$$L_t(\varsigma) = C(\varsigma I - A)^{-1} K_f \tag{4.277}$$

meets all design requirements specified at the output point. We are now seeking to design a measurement feedback controller $K(\varsigma)$ such that all the properties of $L_t(\varsigma)$ can be recovered.

2. Define a dual system Σ_{du} characterised by $(A_{du}, B_{du}, C_{du}, D_{du})$ where

$$A_{du} := A', \quad B_{du} := C', \quad C_{du} := B', \quad D_{du} := D'. \tag{4.278}$$

Let $F_{du} := K_f'$ and let $L_{du}(\varsigma)$ be defined as

$$L_{du}(\varsigma) := L_t'(\varsigma) = F_{du}(\varsigma I - A_{du})^{-1} B_{du}. \tag{4.279}$$

Let $L_{du}(\varsigma)$ be considered as a target loop transfer function for Σ_{du} when the loop is broken at the input point of Σ_{du}. Let a measurement feedback controller $K_{du}(\varsigma)$ be used for Σ_{du}. Here, the controller $K_{du}(\varsigma)$ could be based either on a full or a reduced order observer or CSS architecture depending upon what $K(\varsigma)$ is based on. Following the results given earlier for LTR at the input point to design an appropriate controller $K_{du}(\varsigma)$, then the required controller for LTR at the output point of the original plant Σ is given by

$$K(\varsigma) = K_{du}'(\varsigma). \tag{4.280}$$

This concludes the LTR design for the case when the loop is broken at the output point of the plant.

CHAPTER 5
NONLINEAR CONTROL TECHNIQUES

5.1 Introduction

Every physical system in our real life has nonlinearities and very little can be done to overcome them. Many practical systems are sufficiently nonlinear so that important features of their performance may be completely overlooked if they are analysed and designed through linear techniques. In HDD servo systems, major nonlinearities are frictions, high-frequency mechanical resonances and actuator saturation nonlinearities. Among all these, the actuator saturation could be the most significant nonlinearity in designing an HDD servo system. When the actuator saturates, the performance of the control system designed will seriously deteriorate. Interested readers are referred to a recent monograph by Hu and Lin [97] for a fairly complete coverage of many newly developed results on control systems with actuator nonlinearities.

The actuator saturation in the HDD has seriously limited the performance of its overall servo system, especially in the track seeking stage, in which the HDD R/W head is required to move in a wide range of tracks. It will be obvious in the coming chapters that it is impossible to design a linear controller that would achieve a desired performance in the track seeking stage. Instead, we will have no choice but to utilise some sophisticated nonlinear control techniques in the design. The most popular nonlinear control technique used in the design of HDD servo systems is the so-called proximate time-optimal servomechanism (PTOS) proposed by Workman [29], which achieves near time-optimal performance for a large class of motion control systems characterised a double integrator. The PTOS was actually modified from the well-known time-optimal control. However, it is made to yield a minimum variance with smooth switching from the track seeking to track following modes via a mode switching controller.

In this chapter, we will also introduce two newly proposed nonlinear control techniques, namely a new mode switching control (MSC) and a composite nonlinear feedback (CNF) control. The MSC we present in this chapter is actually a combination of the PTOS and the RPT control of Chapter 4. Inspired by the recent work of Lin *et al.* [98], which was introduced to improve

the tracking performance under state feedback laws for a class of second-order continuous-time linear systems subject to actuator saturation, we have developed in this chapter a nonlinear control technique, namely the composite nonlinear feedback control, to a more general class of continuous-time and discrete-time systems with measurement feedback. The advantages of this new CNF control technique are as follows: (1) it generally yields a better performance compared with those of the PTOS and MSC; and (2) it does not need a mode switching controller, which greatly simplifies the actual process in implementation.

5.2 Time Optimal Control

We recall the technique of the time optimal control (TOC) in this section. Given a dynamic system characterised by

$$\dot{x} = h(x, u, t), \tag{5.1}$$

where x is the state variable and u is the control input, the objective of optimal control is to determine a control input u that will cause a controlled process to satisfy the physical constraints and at the same time optimise a certain performance criterion,

$$J = \int_{t_0}^{t_f} f(x, u, t)\, \mathrm{d}t, \tag{5.2}$$

where t_0 and t_f are respectively initial time and final time of operation, and f is a scalar function. The TOC is a special class of optimisation problems and is defined as the transfer of the system from an arbitrary initial state $x(t_0)$ to a specified target set point in minimum time. For simplicity, we let $t_0 = 0$. Hence, the performance criterion for the time-optimal problem becomes minimising the following cost function with $f(x, u, t) = 1$, *i.e.*,

$$J = \int_0^{t_f} \mathrm{d}t = t_f. \tag{5.3}$$

Let us now derive the TOC law using Pontryagin's principle and the calculus of variation (see *e.g.*, [99]) for a simple dynamic system obeying Newton's law, *i.e.*, for a double integrator system represented by

$$\ddot{y}(t) = au(t), \tag{5.4}$$

where y is the position output, a is the acceleration constant and u is the input to the system. It will be seen later that the dynamics of the actuator of an HDD can be approximated as a double integrator model. To start with, we rewrite Equation (5.4) as the following state space model:

$$\dot{x}(t) = Ax(t) + Bu(t) \tag{5.5}$$

with

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ a \end{bmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} y \\ v \end{pmatrix} = \begin{pmatrix} y \\ \dot{y} \end{pmatrix}. \tag{5.6}$$

Note that $v = \dot{y}$ be the velocity of the system. Let the control input be constrained as follows:

$$| u(t) | \leq u_{\max}. \tag{5.7}$$

Then, the Hamiltonian (see *e.g.*, [99]) for such a problem is given by

$$H(x(t), u(t), \lambda(t)) = 1 + \lambda_1(t)x_2(t) + \lambda_2(t)au(t) \tag{5.8}$$

where $\lambda = (\lambda_1 \quad \lambda_2)'$ is a vector of the time-varying Lagrange multipliers. Pontryagin's principle states that the Hamiltonian is minimised by the optimal control, or

$$H(x^*, u^*, \lambda^*) \leq H(x^*, u, \lambda^*), \tag{5.9}$$

where superscript $*$ indicates optimality. Thus, from Equations (5.8) and (5.9), the optimal control is

$$u^*(t) = \left\{ \begin{array}{ll} -u_{\max}, & \text{for } \lambda_2^*(t) > 0 \\ \\ +u_{\max}, & \text{for } \lambda_2^*(t) < 0 \end{array} \right\} := -\operatorname{sgn}(\lambda_2^*(t))u_{\max}. \tag{5.10}$$

The calculus of variation (see [99]) yields the following necessary condition for a time-optimal solution:

$$\dot{\lambda}^*(t) = -A'\lambda^*(t), \tag{5.11}$$

which is known as a costate equation in optimal control terminology. The solution to the costate equation is of the form

$$\left.\begin{aligned} \lambda_1^*(t) &= c_1 \\ \lambda_2^*(t) &= -c_1 t + c_2 \end{aligned}\right\} \tag{5.12}$$

where c_1 and c_2 are constants of integration. Equation (5.12) indicates that λ_2^* and, therefore u^* can change sign at most once. Since there can be at most one switching, the optimal control for a specified initial state must be one of the following forms:

$$\left.\begin{aligned} 1^\circ &: \quad u^*(t) = +u_{\max}, \qquad \forall t \in [t_0,\ t^*]; \\ 2^\circ &: \quad u^*(t) = \begin{cases} +u_{\max}, & \forall t \in [t_0,\ t_1), \\ -u_{\max}, & \forall t \in [t_1,\ t^*]; \end{cases} \\ 3^\circ &: \quad u^*(t) = -u_{\max}, \qquad \forall t \in [t_0,\ t^*]; \\ 4^\circ &: \quad u^*(t) = \begin{cases} -u_{\max}, & \forall t \in [t_0,\ t_1), \\ +u_{\max}, & \forall t \in [t_1,\ t^*]. \end{cases} \end{aligned}\right\} \tag{5.13}$$

Thus, the segment of optimal trajectories can be found by integrating Equation (5.5) with $u = \pm u_{\max}$ to obtain

$$x_2(t) = \pm a\, u_{\max}\, t + c_3, \tag{5.14}$$

$$x_1(t) = \pm \frac{1}{2} a\, u_{\max}\, t^2 + c_3\, t + c_4, \tag{5.15}$$

where c_3 and c_4 are constants of integration. It is to be noted that if the initial state lies on the optimal trajectories defined by Equations (5.14) and (5.15) in the state plane, then the control will be either 1° or 3° in Equation (5.13) depending upon the direction of motion. In HDD servo systems, it will be shown later that the problem is of relative head positioning control, and hence the initial and final states must be

$$x(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad x(t^*) = \begin{pmatrix} r \\ 0 \end{pmatrix}, \tag{5.16}$$

where r is the reference set point. Because of these kinds of initial state in HDD servo systems, the optimal control must be chosen from either 2° or 4° in Equation (5.13). Note that if the control input $+u_{\max}$ produces the acceleration α, then the input $-u_{\max}$ will produce a deceleration of the same magnitude.

Hence, the minimum time performance can be achieved either with maximum acceleration for half of the travel followed by maximum deceleration

for an equal amount of time, or by first accelerating and then decelerating the system with maximum effort to follow some predefined optimal velocity trajectory to reach the final destination in minimum time. The former case results in an open-loop form of TOC which will use predetermined time-based acceleration and deceleration inputs, whereas the latter yields a closed-loop form of TOC. We note that if the area under acceleration, which is a function of time, is the same as the area under deceleration, there will be no net change in velocity after the input is removed. The final output velocity and the position will be in a steady state.

In general, the time-optimal performance can be achieved by switching the control between two extreme levels of the input, and we have shown that in the double integrator system the number of switchings is at most equal to one, *i.e.*, one less than the order of dynamics. Thus, if we extend the result to an nth-order system, it will need $n-1$ switchings between maximum and minimum inputs to achieve a time-optimal performance. Since the control must be switched between two extreme values, the TOC is also known as *bang-bang* control.

In what follows, we will discuss the bang-bang control in two versions, *i.e.*, in the open-loop and in the closed-loop forms for the double integrator model characterised by Equation (5.5) with the control constraint represented by Equation (5.7).

5.2.1 Open-loop Bang-bang Control

The open-loop method of bang-bang control uses maximum acceleration and maximum deceleration for a predetermined time period. Thus, the time required for the system to reach the target position in minimum time is predetermined from the above principles and the control input is switched between two extreme levels for this time period. We can precalculate the minimum time t^* for a specified reference set point r. Let the control be

$$u^*(t) = \begin{cases} +u_{\max}, & \text{for } t \in (0, \frac{t^*}{2}], \\ -u_{\max}, & \text{for } t \in (\frac{t^*}{2}, t^*]. \end{cases} \tag{5.17}$$

We now solve Equations (5.14) and (5.15) for the accelerating phase with zero initial condition. For the accelerating phase, *i.e.*, with $u = +u_{\max}$, we have

$$x_2(t) = a\, u_{\max}\, t, \quad x_1(t) = \frac{1}{2}\, a\, u_{\max}\, t^2. \tag{5.18}$$

At the end of the accelerating phase, *i.e.*, at $t = t^*/2$,

$$x_{2a} = a\,u_{\max}\frac{t^*}{2}, \quad x_{1a} = \frac{1}{8}a\,u_{\max}(t^*)^2. \tag{5.19}$$

Similarly, at the end of decelerating phase, we can show that

$$x_{2d} = 0, \quad x_{1d} = \frac{1}{8}a\,u_{\max}(t^*)^2. \tag{5.20}$$

Obviously, the total displacement at the end of bang-bang control must reach the target, *i.e.*, the reference set point r. Thus,

$$r = x_{1a} + x_{1d} = \frac{1}{4}a\,u_{\max}(t^*)^2, \tag{5.21}$$

which gives

$$t^* = \sqrt{\frac{4\,r}{a\,u_{\max}}}, \tag{5.22}$$

the minimum time required to reach the target set point.

5.2.2 Closed-loop Bang-bang Control

In this method, the velocity of the plant is controlled to follow a predefined trajectory and more specifically the decelerating trajectory. These trajectories can be generated from the phase-plane analysis. This analysis is explained below for the system given by Equation (5.5) and can be extended to higher-order systems (see *e.g.*, [99]). We will show later that this deceleration trajectory will bring the system to the desired set point in finite time. We now move to find the deceleration trajectory.

First, eliminating t from Equations (5.14) and (5.15), we have

$$x_1(t) = \frac{1}{2\,a\,u_{\max}}x_2^2(t) + c_5, \qquad \text{for } u = +u_{\max}, \tag{5.23}$$

$$x_1(t) = -\frac{1}{2\,a\,u_{\max}}x_2^2(t) + c_6, \qquad \text{for } u = -u_{\max}, \tag{5.24}$$

where c_5 and c_6 are appropriate constants. Note that each of the above equation defines the family of parabolas. Let us define $e(t) := r - x_1(t)$ to be the position error with r being the desired final position. Then, if we consider the

trajectories between $e(t)$ and $x_2(t)$, our desired final state in $e(t)$ and $x_2(t)$ plane must be

$$\begin{pmatrix} r - x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{5.25}$$

In this case, the constants in the above trajectories are equal to zero. Moreover, both the trajectories given by Equations (5.23) and (5.24) are the decelerating trajectories depending upon the direction of the travel. The mechanism of the TOC can be illustrated in a graphical form as given in Figure 5.1. Clearly, any initial state lying below the curve is to be driven by the positive accelerating force to bring the state to the deceleration trajectory. On the other hand, any initial state lying above the curve is to be accelerated by the negative force to the deceleration trajectory.

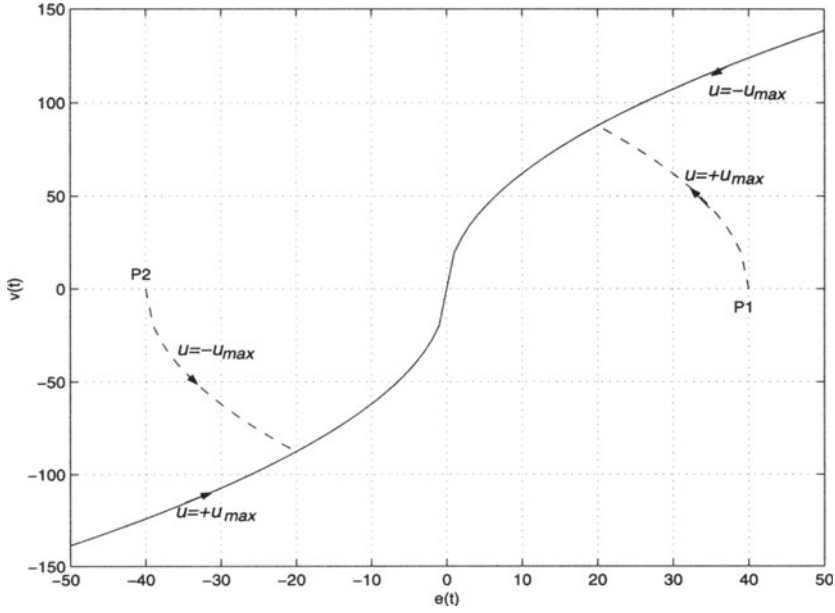


Figure 5.1. Deceleration trajectories for TOC.

Let

$$f_t(e) := \operatorname{sgn}(e)\sqrt{2\,a\,u_{\max} \mid e \mid}. \tag{5.26}$$

The control law is then given by

$$u = u_{\max} \cdot \operatorname{sgn}\Big(f_t(e) - v\Big). \tag{5.27}$$

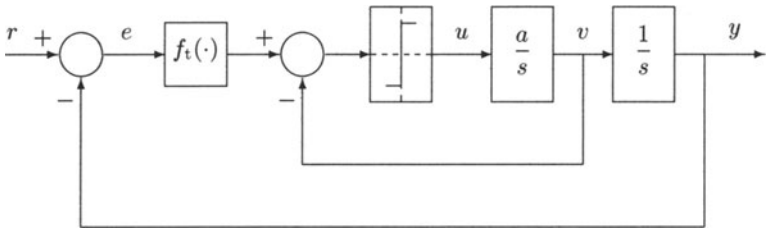


Figure 5.2. Typical scheme of TOC.

A block diagram depicting the closed-loop method of bang-bang control is shown in Figure 5.2. Unfortunately, the control law given by Equation (5.27) for the system shown in Figure 5.2, although time-optimal, is not practical. It applies maximum or minimum input to the plant to be controlled even for a small error. Moreover, this algorithm is not suited for disk drive applications for the following reasons:

1. even the smallest system process or measurement noise will cause control "chatter". This will excite the high frequency modes.

2. any error in the plant model, will cause limit cycling to occur.

As such, the TOC given above has to be modified to suit HDD applications. In the following section, we will recall a modified version of the TOC proposed by Workman [29], *i.e.*, the PTOS. Such a control scheme is widely used nowadays in designing HDD servo systems.

5.3 Proximate Time-Optimal Servomechanism

The infinite gain of the signum function in the TOC causes control chatter, as seen in the previous section. Workman [29], in 1987, proposed a modification of this technique, *i.e.*, the so-called PTOS, to overcome such a drawback. The PTOS essentially uses maximum acceleration where it is practical to do so. When the error is small, it switches to a linear control law. To do so, it replaces the signum function in TOC law by a saturation function. In the following sections, we will revisit the PTOS method in continuous-time and in discrete-time domains.

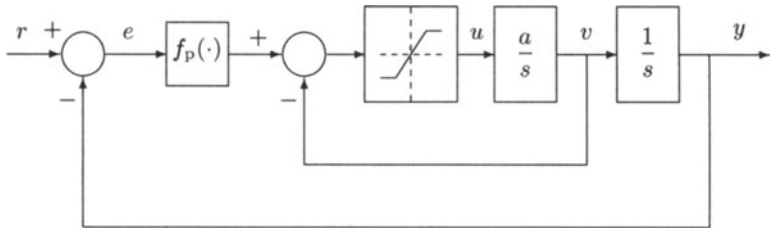


Figure 5.3. Continuous-time PTOS.

5.3.1 Continuous-time Systems

The configuration of the PTOS is shown in Figure 5.3. The function $f_p(\cdot)$ is a finite slope approximation to the switching function $f_t(\cdot)$ given by Equation (5.26). The PTOS control law for the system in Equation (5.5) is given by

$$u = u_{\max} \cdot \text{sat}\left(\frac{k_2[f_p(e) - v]}{u_{\max}}\right), \tag{5.28}$$

where $\text{sat}(x)$ is defined as

$$\text{sat}(x) = \begin{cases} +1, & \text{if} \quad x > +y_\ell, \\ x, & \text{if} \quad -y_\ell \leq x \leq y_\ell, \\ -1, & \text{if} \quad x < -y_\ell, \end{cases} \tag{5.29}$$

and the function $f_p(e)$ is given by

$$f_p(e) = \begin{cases} \dfrac{k_1}{k_2} e & \text{for } |e| \leq y_\ell, \\ \text{sgn}(e)\left[\sqrt{2\,a\,u_{\max}\,\alpha\,|\,e\,|} - \dfrac{u_{\max}}{k_2}\right] & \text{for } |e| > y_\ell. \end{cases} \tag{5.30}$$

Here we note that k_1 and k_2 are respectively the feedback gains for position and velocity, α is a constant between 0 and 1 and is referred to as the acceleration discount factor, and y_ℓ is the size of the linear region. Since the linear portion of the curve $f_p(\cdot)$ must connect the two disjoint halves of the nonlinear portion, we have constraints on the feedback gains and the linear region to guarantee the continuity of the function $f_p(\cdot)$. It was proved by Workman [29] that

$$y_\ell = \frac{u_{\max}}{k_1}, \quad k_2 = \sqrt{\frac{2k_1}{a\alpha}}. \tag{5.31}$$

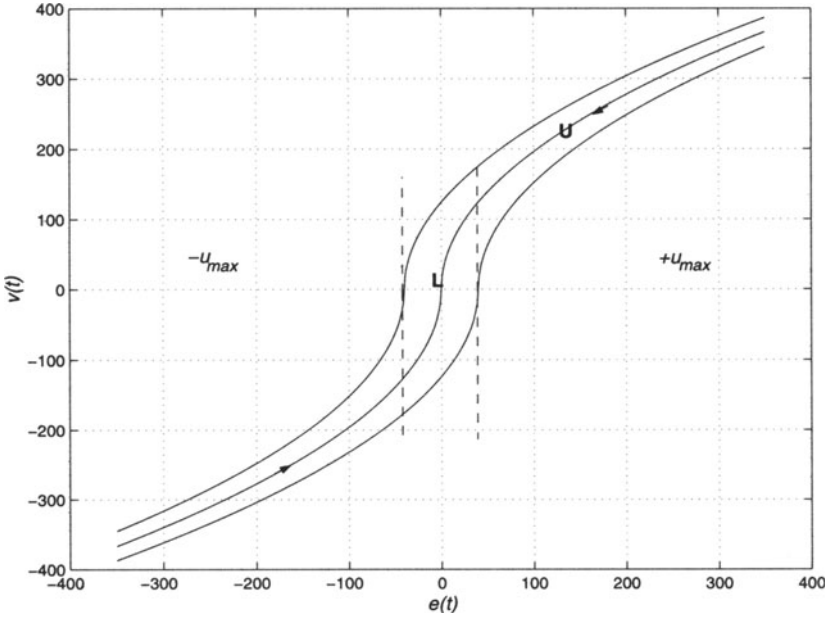


Figure 5.4. Control zones of a PTOS.

The control zones in the PTOS are shown in Figure 5.4. The two curves bounding the switching curve (central curve) now redefine the control boundaries and it is termed a linear boundary. Let this region be **U**. The region below the lower curve is the region where the control $u = +u_{\max}$, whereas the region above the upper curve is the region where the control $u = -u_{\max}$. It has been proved [29] that once the state trajectory enters the band **U** in Figure 5.4 it remains within **U** and the control signal is below the saturation. The region marked **L** is the region where the linear control is applied.

The presence of the acceleration discount factor α allows us to accommodate uncertainties in the plant accelerating factor at the cost of increase in response time. By approximating the positioning time as the time that it takes the position error to be within the linear region, one can show that the percentage increase P in time taken by the PTOS over the time taken by the TOC is given by (see [29]):

$$P = \frac{1}{2}\left(\frac{1}{\sqrt{\alpha}} - 1\right) \times 100\%. \tag{5.32}$$

Clearly, larger values of α make the response closer to that of the TOC. As a result of changing the nonlinearity from $\mathrm{sgn}(\cdot)$ to $\mathrm{sat}(\cdot)$, the control chatter is eliminated.

5.3.2 Discrete-time Systems

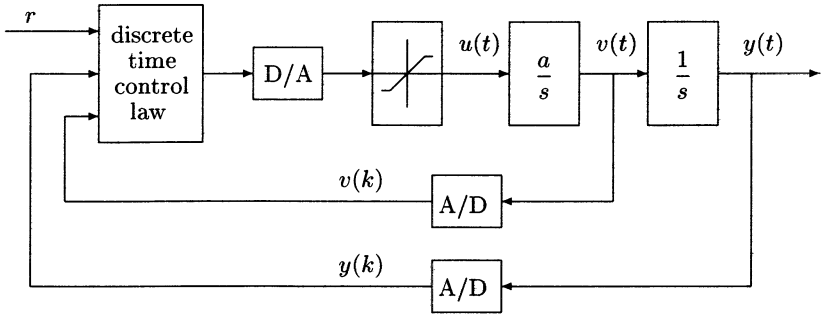


Figure 5.5. Discrete-time PTOS.

The discrete-time PTOS can be derived from its continuous-time counterpart, but with some conditions on sample time to ensure stability. In his seminal work, Workman [29] extended the continuous-time PTOS to discrete-time control of a continuous-time double integrator plant driven by a zero-order hold as shown in Figure 5.5. As in the continuous-time case, the states are defined as position and velocity. With insignificant calculation delay, the state-space description of the plant given by Equation (5.5) in the discrete-time domain is

$$x(k+1) = \begin{bmatrix} 1 & T_s \\ 0 & 1 \end{bmatrix} x(k) + \begin{bmatrix} aT_s^2/2 \\ aT_s \end{bmatrix} u(k), \tag{5.33}$$

where T_s is the sampling period. The control structure is a discrete-time mapping of the continuous-time PTOS law, but with a constraint on the sampling period to guarantee that the control will not saturate during the deceleration phase to the target position and also to guarantee its stability. Thus, the mapped control law is

$$u(k) = u_{max} \cdot \text{sat}\left(\frac{k_2[f_p(e(k)) - v(k)]}{u_{max}}\right), \tag{5.34}$$

with the following constraint on sampling frequency ω_s,

$$\frac{\omega_s}{\omega_n} > 6.3, \tag{5.35}$$

where ω_n is the desired bandwidth of the closed-loop system.

5.4 Mode Switching Control

In this section, we follow the development of [100] to introduce the design of a new mode switching control (MSC) design for a system characterised by a double integrator or in the following state-space equation:

$$\dot{x} = Ax + Bu = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ a \end{bmatrix} u, \quad x = \begin{pmatrix} y \\ v \end{pmatrix}, \tag{5.36}$$

where as usual x is the state, which consists of the displacement y and the velocity v; u is the control input constrained by

$$|u(t)| \leq u_{\max}. \tag{5.37}$$

As will be seen shortly in the coming chapters, the VCM actuators of HDDs can generally be approximated by such a model with appropriate parameters a and $u_{\max}$. In HDD servo systems, in order to achieve both high-speed track seeking and highly accurate head positioning, multimode control designs are widely used. The two commonly used multimode control designs are MSC and sliding mode control. Both control techniques in fact belong to the category of variable structure control. That is, the control is switched between two or more different controllers to achieve the two conflicting requirements. In this section, we will propose an MSC control in which the seeking mode is controlled by a PTOS and the track following mode is controlled by a RPT controller.

The MSC (see *e.g.*, [15]) is a type of variable structure control system [101], but the switching is in only one direction. Figure 5.6 shows a basic schematic diagram of MSC. There are track seeking and track following modes. Each servo mode can be designed independently, and so the main issue in MSC is the design of the switching mechanism.

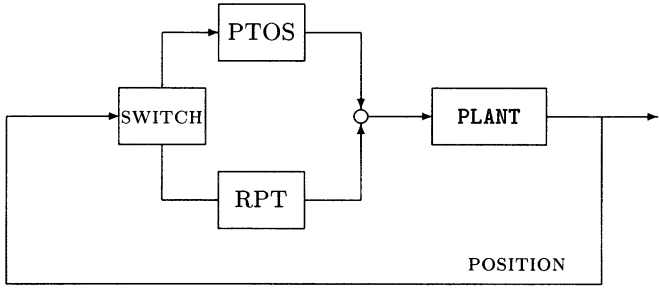


Figure 5.6. Basic schematic diagram of MSC.

This design problem has not yet been completely resolved, and many heuristic approaches have been tried so far (see *e.g.*, [102]). Several methods were proposed for mode switching from one controller to another. In [15], a method called initial value compensation is proposed. Note that, when the switch is transfered from track seeking mode to the track following mode, the final states of the track seeking controller become the initial states for the track following controller, and hence affect the settling performance of track following mode. In order to reduce the impact of these initial values during mode switching, some compensation must be worked out. Here, the initial values are referred to the values of the states during mode switching. However, the RPT controllers developed by Chen and co-workers [44, 84] (see also Chapter 4) have enough robustness against plant variations and are actually independent of initial values. Hence, the use of these controllers in track following mode eliminates the need for initial value compensation during mode switching. Moreover, the RPT controllers in a track following servo have been proved to be robust against resonance mode changes from disk to disk and work well against runout disturbances.

The MSC law that combines the PTOS and RPT controllers takes the following simple form:

$$u = \begin{cases} u_{\mathrm{P}}, & t < t_1, \\ u_{\mathrm{R}}, & t \geq t_1, \end{cases} \tag{5.38}$$

where u_{P} is a control signal generated by the PTOS control and is given as in Equation (5.28), and u_{R} is a signal generated by the reduced order RPT control as given in Equation (4.201). Furthermore, t_1, the time that MSC switches from one mode to the other, will be presented in the next subsection together with the stability analysis of the closed-loop comprising the given plant and the MSC control law.

5.4.1 Stability Analysis and Mode Switching Conditions

We show in this subsection the stability of the MSC and give a set of conditions for mode switching. First of all, we rewrite the given system Equation (5.36) as follows:

$$\begin{pmatrix} \dot{e} \\ \dot{v} \end{pmatrix} = \begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix} \begin{pmatrix} e \\ v \end{pmatrix} + \begin{bmatrix} 0 \\ a \end{bmatrix} u := A_{\mathrm{P}} x_{\mathrm{P}} + B_{\mathrm{P}} u, \quad a > 0, \tag{5.39}$$

where $e = r - y$ is the tracking error with r being the target reference. In the HDD servo systems that we deal with in the coming chapters, y will be

regarded as the displacement of an HDD R/W head, and v is its velocity. Recall the PTOS control law:

$$u_{\mathrm{P}} = u_{\max} \cdot \mathrm{sat}\left[\frac{k_2}{u_{\max}}\Big(f(e) - v\Big)\right], \tag{5.40}$$

where the $u_{\max}$ used throughout this section has a saturation level equal to unity; the function $f(\cdot)$ and the feedback gain k_2 are as defined in the previous section. It has been shown [29] that the PTOS control law will yield an asymptotically stable closed-loop system provided that the following conditions are satisfied:

1. $ak_2 > 0$;
2. $f(0) = 0$;
3. $f(e)e > 0$, for any nonzero e;
4. $\lim\limits_{e\to\infty} \int_0^e f(\delta)\,\mathrm{d}\delta = \infty$;
5. $\dot{f}(e)$ exists everywhere; and
6. for any e, $u_{\max}\left[-a + \dfrac{1}{k_2}\right]\dot{f}(e) < -\dot{f}(e)f(e) < u_{\max}\left[a - \dfrac{1}{k_2}\dot{f}(e)\right]$.

Generally, as the velocity is not measurable, the PTOS control law will have to be modified as follows if it is to be implemented in a real system:

$$\begin{cases} u_{\mathrm{P}} &= u_{\max}\ \mathrm{sat}\left[\dfrac{k_2}{u_{\max}}\Big(f(e) - \hat{v}\Big)\right], \\ \dot{z} &= -\kappa z + \kappa^2 e + a\,u_{\mathrm{P}}, \\ \hat{v} &= z - \kappa e, \end{cases} \tag{5.41}$$

where κ is the estimator feedback gain, and z is the estimator state. Next, we let $\tilde{z} = z - v - \kappa e$. Then, the dynamics of the closed-loop system with the above control law can be written as

$$\begin{cases} \dot{e} &= -v, \\ \dot{v} &= a\,u_{\max}\mathrm{sat}\left[\dfrac{k_2}{u_{\max}}\Big(f(e) - \tilde{z} - v\Big)\right], \\ \dot{\tilde{z}} &= -\kappa\tilde{z}. \end{cases} \tag{5.42}$$

It can be shown that the closed-loop system comprising the given plant and the modified PTOS control law, in which the velocity is replaced by the above

estimation, is asymptotically stable, if conditions 1 to 5 above are satisfied and condition 6 is replaced by

$$a > \frac{1}{k_2}\dot{f}(e) > 0, \tag{5.43}$$

and

$$|\dot{f}(e)f(e)| < u_{\max}\left[a - \frac{1}{k_2}\dot{f}(e)\right] - \left[\kappa + \dot{f}(e)\right] \cdot |\tilde{z}(0)|. \tag{5.44}$$

We can show that, under these new conditions, the closed-loop is stable for the case when the control input is saturated, *i.e.*, $|k_2[f(e) - (z - \kappa e)]| > u_{\max}$. For the case when $|k_2[f(e) - (z - \kappa e)]| \leq u_{\max}$, the closed-loop system in Equation (5.42) can be written as

$$\begin{cases} \dot{e} &= -v, \\ \dot{v} &= ak_2[f(e) - \tilde{z} - v], \\ \dot{\tilde{z}} &= -\kappa\tilde{z}. \end{cases} \tag{5.45}$$

Following the result of [103], we propose the following Lyapunov function for the system in Equation (5.45):

$$V_{\mathrm{P}} = \frac{v^2}{2k_2a} + \int_0^e f(\delta)\, d\delta + \frac{p_z}{2}\tilde{z}^2, \tag{5.46}$$

where $p_z > 0$ is a scalar constant. The derivative of the above Lyapunov function is given by

$$\dot{V}_{\mathrm{P}} = -\left(v + \frac{\tilde{z}}{2}\right)^2 - \tilde{z}^2\left(p_z\kappa - \frac{1}{4}\right). \tag{5.47}$$

The last term is negative for all $p_z > \frac{1}{4\kappa}$. Thus, under this choice of p_z, $\dot{V}_{\mathrm{P}} \leq 0$. It follows from LaSalle's Theorem [104] that the closed-loop system comprising the PTOS control law with the estimated velocity and the given plant is asymptotically stable.

It is pretty obvious that the closed-loop system comprising the given plant in Equation (5.39) and the reduced order RPT control law of Equation (4.201) is asymptotically stable when the control input is not saturated. For completeness, and for the analysis of the overall closed-loop system with the MSC scheme, we proceed to investigate the closed-loop system comprising the plant and the RPT controller, which can be written as

$$\begin{cases} \dot{z} &= -\kappa z + \kappa^2 e + a\, u_{\max} \cdot \text{sat}\left(\dfrac{u_{\text{R}}}{u_{\max}}\right), \\ u_{\text{R}} &= (f_1 - \kappa f_2)e + f_2 z, \end{cases} \tag{5.48}$$

where κ is again the reduced order observer gain, which is selected to be exactly the same as that used for the velocity estimation in the PTOS, and $F = [\, f_1 \quad f_2 \,]$ is the feedback gain obtained using the RPT technique of Chapter 4. Again, let $\tilde{z} = z - v - \kappa e$ and rewrite the RPT control law as

$$u_{\text{R}} = [\, f_1 \quad f_2 \,] \begin{pmatrix} e \\ v \end{pmatrix} + f_2 \tilde{z} = F x_{\text{P}} + f_2 \tilde{z}. \tag{5.49}$$

Let $W_x \in \mathbb{R}^{2\times 2}$ be a positive definite matrix and solve the following Lyapunov equation:

$$(A_{\text{P}} + B_{\text{P}} F)' P_x + P_x (A_{\text{P}} + B_{\text{P}} F) = -W_x, \tag{5.50}$$

for $P_x > 0$. Such a P_x always exists as $A_{\text{P}} + B_{\text{P}} F$ is stable. Next, let

$$w_z > \max\left\{\frac{1}{2},\, f_2 B_{\text{P}}' P_x W_x^{-1} P_x B_{\text{P}} f_2\right\} > 0, \tag{5.51}$$

and

$$p_z = \frac{w_z}{2\kappa} > \frac{1}{4\kappa} > 0. \tag{5.52}$$

Then, we define a set

$$\boldsymbol{X} := \left\{ \begin{pmatrix} x \\ z \end{pmatrix} : \begin{pmatrix} x \\ z \end{pmatrix}' \begin{bmatrix} P_x & 0 \\ 0 & p_z \end{bmatrix} \begin{pmatrix} x \\ z \end{pmatrix} \le c \right\}, \tag{5.53}$$

where $c > 0$ is the largest positive value such that

$$\begin{pmatrix} x \\ z \end{pmatrix} \in \boldsymbol{X} \Rightarrow \left| [\, F \quad f_2 \,] \begin{pmatrix} x \\ z \end{pmatrix} \right| \le u_{\max}. \tag{5.54}$$

For all

$$\begin{pmatrix} x_{\text{P}} \\ z \end{pmatrix} \in \boldsymbol{X} \tag{5.55}$$

the resulting closed-loop system can then be written as

$$\begin{pmatrix} \dot{x}_{\mathrm{P}} \\ \dot{\tilde{z}} \end{pmatrix} = \begin{bmatrix} A_{\mathrm{P}} + B_{\mathrm{P}}F & B_{\mathrm{P}}f_2 \\ 0 & -\kappa \end{bmatrix} \begin{pmatrix} x_{\mathrm{P}} \\ \tilde{z} \end{pmatrix}. \tag{5.56}$$

Define a Lyapunov function,

$$V_{\mathrm{R}} = \begin{pmatrix} x_{\mathrm{P}} \\ \tilde{z} \end{pmatrix}' \begin{bmatrix} P_x & 0 \\ 0 & p_z \end{bmatrix} \begin{pmatrix} x_{\mathrm{P}} \\ \tilde{z} \end{pmatrix} \tag{5.57}$$

and evaluate its derivative along the trajectories of the closed-loop system in Equation (5.56), *i.e.*,

$$\begin{aligned} \dot{V}_{\mathrm{R}} &= \begin{pmatrix} x_{\mathrm{P}} \\ \tilde{z} \end{pmatrix}' \begin{bmatrix} -W_x & P_x B_{\mathrm{P}} f_2 \\ f_2 B_{\mathrm{P}}' P_x & -w_z \end{bmatrix} \begin{pmatrix} x_{\mathrm{P}} \\ \tilde{z} \end{pmatrix} \\ &= \begin{pmatrix} \hat{x} \\ \tilde{z} \end{pmatrix}' \begin{bmatrix} -W_x & 0 \\ 0 & -\tilde{w}_z \end{bmatrix} \begin{pmatrix} \hat{x} \\ \tilde{z} \end{pmatrix} \leq 0, \end{aligned} \tag{5.58}$$

where

$$\hat{x} = x_{\mathrm{P}} - W_x^{-1} P_x B_p f_2 \tilde{z}, \quad \tilde{w}_z = w_z - f_2' B_{\mathrm{v}}' P_x W_x^{-1} P_x B_p f_2 > 0. \tag{5.59}$$

This shows that all trajectories of Equation (5.56) starting from $\boldsymbol{X}$ will remain there and converge asymptotically to zero. Hence, the closed-loop system comprising the plant and the reduced order RPT control law is asymptotically stable provided that the control input is not saturated.

Next, we re-express Equation (5.46) using the Taylor expansion as follows:

$$V_{\mathrm{P}} = \frac{v^2}{2k_2 a} + \frac{\dot{f}(\tau)}{2} e^2 + \frac{p_z}{2} \tilde{z}^2 = \begin{pmatrix} e \\ v \\ \tilde{z} \end{pmatrix}' \begin{bmatrix} \dfrac{\dot{f}(\tau)}{2} & 0 & 0 \\ 0 & \dfrac{1}{2ak_2} & 0 \\ 0 & 0 & \dfrac{p_z}{2} \end{bmatrix} \begin{pmatrix} e \\ v \\ \tilde{z} \end{pmatrix},$$

where τ is an appropriate scalar between 0 and e. Let

$$\sigma = \min \left\{ \frac{\dot{f}(\tau)}{2}, \frac{1}{2ak_2}, \frac{p_z}{2} \right\} \Big/ \max\Big\{ \lambda_{\max}(P_x), p_z \Big\}. \tag{5.60}$$

The MSC scheme can be obtained as follows:

$$u(t) = \begin{cases} u_{\mathrm{P}}, & t < t_1, \\ u_{\mathrm{R}}, & t \geq t_1, \end{cases} \tag{5.61}$$

where t_1 is such that

$$\begin{pmatrix} x_{\mathrm{P}}(t_1) \\ z(t_1) \end{pmatrix} \in \boldsymbol{X} \quad \text{and} \quad |e(t_1)| < y_\ell, \tag{5.62}$$

and where y_ℓ is the size of the linear region of the PTOS control law. The Lyapunov function for the overall closed-loop system can be chosen as

$$V = V_{\mathrm{P}}[1 - 1(t - t_1)] + \sigma V_{\mathrm{R}}\, 1(t - t_1), \tag{5.63}$$

where $1(t)$ is the unit step function. It is simple to verify that

$$\dot{V} = \dot{V}_{\mathrm{P}}[1 - 1(t - t_1)] + \sigma \dot{V}_{\mathrm{R}}\, 1(t - t_1) + (\sigma V_{\mathrm{R}} - V_{\mathrm{P}})\delta(t - t_1). \tag{5.64}$$

It has already been proved that the derivatives of the functions V_{P} and V_{R} are negative definite. The last term is always negative in view of the definition of σ in Equation (5.60). Hence, $\dot{V} \leq 0$ and the resulting closed-loop system comprising the given plant and the mode switching control law is stable.

5.5 Composite Nonlinear Feedback Control

The PTOS and MSC schemes discussed in the previous sections have two controllers, *i.e.*, a nonlinear controller and a linear controller, operating in two different time stages. Inspired by the recent work of Lin *et al.* [98], we develop in this section a new servo system design method, *i.e.*, the so-called composite nonlinear feedback (CNF), which again consists of two parts, a linear part and a nonlinear part. However, unlike the MSC and PTOS control laws, both the linear and nonlinear controllers in CNF are in operation all the time. The linear part in CNF is designed to yield a fast response, whereas the nonlinear part is designed to increase the damping ratio of the closed-loop system to reduce the overshoot as the system controlled output approaches the desired target position. As will be seen soon in the coming chapters, this new control method will improve the performance of the overall servo system a great deal. In what follows, we will present the CNF technique for both the continuous- and discrete-time systems. We will, as Chapter 4, consider the following three different situations: (1) the state feedback case, (2) the full order measurement feedback case, and (3) the reduced order measurement feedback case. Since this is a new technique (although some parts are under review, see *e.g.*, [105]), we will present rigorous and complete proofs for all results derived.

5.5.1 Continuous-time Systems

Consider a linear continuous-time system Σ with an amplitude-constrained actuator characterised by

$$\begin{cases} \dot{x} = A\ x + B\ \mathrm{sat}(u), \quad x(0) = x_0 \\ y = C_1\ x \\ h = C_2\ x \end{cases} \tag{5.65}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}$, $y \in \mathbb{R}^p$ and $h \in \mathbb{R}$ are respectively the state, control input, measurement output and controlled output of Σ. A, B, C_1 and C_2 are appropriate dimensional constant matrices, and sat: $\mathbb{R} \to \mathbb{R}$ represents the actuator saturation defined as

$$\mathrm{sat}(u) = \mathrm{sgn}(u) \min\{\ u_{\max},\ |\ u\ |\ \}, \tag{5.66}$$

with $u_{\max}$ being the saturation level of the input. The following assumptions on the system matrices are required:

1. (A, B) is stabilisable,
2. (A, C_1) is detectable, and
3. (A, B, C_2) is invertible and has no invariant zeros at $s = 0$.

The objective here is to design a CNF control law that will cause the output to track a high amplitude step input rapidly without experiencing large overshoot and without the adverse actuator saturation effects. This will be done through the design of a linear feedback law with a small closed-loop damping ratio and a nonlinear feedback law through an appropriate Lyapunov function to cause the closed-loop system to be highly damped as system output approaches the command input to reduce the overshoot. Similar to the design of linear controllers discussed in Chapter 4, we separate the CNF controller design into three distinct situations (1) the state feedback case, (2) the full order measurement feedback case, and (3) the reduced order measurement feedback case.

i. State Feedback Case. We proceed to develop a composite nonlinear feedback control technique for the case when all the states of the plant Σ are measurable, *i.e.*, $y = x$. It will be done in three steps. In the first step, a linear feedback control law will be designed; in the second step, the design of nonlinear feedback control will be carried out. Lastly, in the third step, the linear and nonlinear feedback laws will be combined to give a composite

nonlinear feedback control law. Again, we note that the procedure given for this case follows closely to that reported in Lin *et al.* [98], although our result is applicable to a much larger class of systems.

STEP 5.5.C.S.1: design a linear feedback law,

$$u_{\text{L}} = Fx + Gr, \tag{5.67}$$

where F is chosen such that (1) $A + BF$ is an asymptotically stable matrix, and (2) the closed-loop system $C_2(sI - A - BF)^{-1}B$ has certain desired properties, *e.g.*, having a small damping ratio. We note that such an F can be designed using methods such as the H_2 and H_∞ optimisation approaches, as well as the RPT technique given in Chapter 4. Furthermore, G is a scalar and is given by

$$G = -\left[C_2(A + BF)^{-1}B\right]^{-1}, \tag{5.68}$$

and r is a step command input. Here we note that G is well defined because $A + BF$ is stable, and the triple (A, B, C_2) is invertible and has no invariant zeros at $s = 0$.

The following lemma determines the magnitude of r that can be tracked by such a control law without exceeding the control limits.

Lemma 5.5.1. Given a positive definite matrix $W \in \mathbb{R}^{n\times n}$, let $P > 0$ be the solution of the following Lyapunov equation:

$$(A + BF)'P + P(A + BF) = -W. \tag{5.69}$$

Such a P exists since $A + BF$ is asymptotically stable. For any $\delta \in (0, 1)$, let $c_\delta > 0$ be the largest positive scalar satisfying the following condition:

$$|\, Fx \,| \leq u_{\max}(1 - \delta), \quad \forall x \in \boldsymbol{X}_\delta := \left\{x : x'Px \leq c_\delta\right\}. \tag{5.70}$$

Also, let

$$H := \left[1 - F(A + BF)^{-1}B\right] G, \tag{5.71}$$

and

$$x_{\text{e}} := G_{\text{e}}\, r := -(A + BF)^{-1}BG\, r. \tag{5.72}$$

Then, the control law of Equation (5.67) is capable of driving the controlled output h to track asymptotically a step command input r, provided that the initial state x_0 and r satisfy:

$$\tilde{x}_0 := (x_0 - x_e) \in \boldsymbol{X}_\delta \quad \text{and} \quad |H\,r| \leq \delta u_{\max}. \tag{5.73}$$

Proof. Let us first define a new state variable

$$\tilde{x} = x - x_e. \tag{5.74}$$

It is simple to verify that the linear feedback control law of Equation (5.67) can be rewritten as

$$u_L(t) = F\tilde{x}(t) + \left[1 - F(A + BF)^{-1}B\right] Gr = F\tilde{x}(t) + H\,r, \tag{5.75}$$

and hence for all $\tilde{x} \in \boldsymbol{X}_\delta$ and, provided that $|H\,r| \leq \delta u_{\max}$, the closed-loop system is linear and is given by

$$\dot{\tilde{x}} = (A + BF)\tilde{x} + Ax_e + BH\,r. \tag{5.76}$$

Noting that

$$\begin{aligned} Ax_e + BHr &= \Big\{B[1-F(A+BF)^{-1}B]G - A(A+BF)^{-1}BG\Big\}r \\ &= \Big\{[I-BF(A+BF)^{-1}]BG - A(A+BF)^{-1}BG\Big\}r \\ &= \Big\{I-BF(A+BF)^{-1} - A(A+BF)^{-1}\Big\}BGr \\ &= 0, \end{aligned} \tag{5.77}$$

the closed-loop system in Equation (5.76) can then be simplified as

$$\dot{\tilde{x}} = (A + BF)\tilde{x}. \tag{5.78}$$

Next, we define a Lyapunov function:

$$V(\tilde{x}) = \tilde{x}'P\tilde{x}. \tag{5.79}$$

Along the trajectories of the closed-loop system in Equation (5.78), $V(\tilde{x})$ satisfies

$$\dot{V}(\tilde{x}) = -\tilde{x}'W\tilde{x}, \tag{5.80}$$

which implies that $V(\tilde{x})$ is a monotonically decreasing function with respect to t along the trajectories of Equation (5.78). Thus, all trajectories of Equation (5.78) starting from $\boldsymbol{X}_\delta$ will remain there and converge asymptotically to the origin. For an initial state x_0 and the step command input r that satisfy Equation (5.73), we have

$$\lim_{t\to\infty} x(t) = x_e, \tag{5.81}$$

and hence

$$\lim_{t\to\infty} h(t) = \lim_{t\to\infty} C_2 x(t) = C_2 x_e = -C_2(A+BF)^{-1}BGr = r. \tag{5.82}$$

This completes the proof of Lemma 5.5.1. ◇

Remark 5.5.1. We would like to note that for the case when $x_0 = 0$, any step command of amplitude r can be asymptotically tracked if

$$| r | \leq \left[c_\delta (G_e' P G_e)^{-1}\right]^{1/2} \quad \text{and} \quad | Hr | \leq \delta \cdot u_{\max}. \tag{5.83}$$

Clearly, the trackable amplitudes of reference inputs by the linear feedback control law can be increased by increasing δ and/or decreasing $G_e' P G_e$ through the choice of W. However, the change in gain F will of course affect the damping ratio of the closed-loop system and hence its rising time. ◇

STEP 5.5.C.S.2: the nonlinear feedback control law $u_N(t)$ is given by

$$u_N = \rho(r,y) B' P (x - x_e) \tag{5.84}$$

where $\rho(r,y)$ is any nonpositive function locally Lipschitz in y, which is used to change the system closed-loop damping ratio as the output approaches the step command input. The choice of such a $\rho(r,y)$ will be discussed at the end of this section.

STEP 5.5.C.S.3: the linear and nonlinear feedback laws derived in the previous steps are now combined to form a CNF controller:

$$u = u_L + u_N = Fx + Gr + \rho(r,y) B' P (x - x_e). \tag{5.85}$$

We now move to prove that the closed-loop system comprising the given plant in Equation (5.65) and the CNF control law of Equation (5.85) is asymptotically stable.

Theorem 5.5.1. Consider the given system in Equation (5.65). Then, for any nonpositive function $\rho(r, y)$, locally Lipschitz in y, the composite nonlinear feedback law in Equation (5.85) will drive the system controlled output $h(t)$ to track asymptotically the step command input of amplitude r from an initial state x_0, provided that x_0 and r satisfy Equation (5.73). ◊

Proof. Again, let $\tilde{x} = x - x_e$. Then, the closed-loop system comprising the given plant in Equation (5.65) and the CNF control law of Equation (5.85) can be expressed as

$$\dot{\tilde{x}} = (A + BF)\tilde{x} + Bw, \tag{5.86}$$

where

$$w = \text{sat}\,(F\tilde{x} + H\,r + u_N) - F\tilde{x} - H\,r. \tag{5.87}$$

Clearly, for the given x_0 satisfying Equation (5.73), we have $\tilde{x}_0 = (x_0 - x_e) \in \boldsymbol{X}_\delta$. Using the Lyapunov function

$$V = \tilde{x}'P\tilde{x} \tag{5.88}$$

we can evaluate the derivative of V along the trajectories of the closed-loop system in Equation (5.86), *i.e.*,

$$\begin{aligned} \dot{V} &= \dot{\tilde{x}}'P\tilde{x} + \tilde{x}'P\dot{\tilde{x}} \\ &= \tilde{x}'(A + BF)'P\tilde{x} + \tilde{x}'P(A + BF)\tilde{x} + 2\tilde{x}'PBw \\ &= -\tilde{x}'W\tilde{x} + 2\tilde{x}'PBw. \end{aligned} \tag{5.89}$$

Note that for all

$$\tilde{x} \in \boldsymbol{X}_\delta = \{\tilde{x} : \tilde{x}'P\tilde{x} \leq c_\delta\} \;\Rightarrow\; |\, F\tilde{x} \,| \leq u_{\max}(1 - \delta). \tag{5.90}$$

We next calculate $\dot{V}$ for three different values of saturation function.

Case 1. If $|F\tilde{x} + Hr + u_N| \leq u_{\max}$, then $w = u_N = \rho B'P\tilde{x}$ and thus

$$\dot{V} = -\tilde{x}'W\tilde{x} + 2\rho\tilde{x}'PBB'P\tilde{x} \leq -\tilde{x}'W\tilde{x}. \tag{5.91}$$

Case 2. If $F\tilde{x} + Hr + u_N > u_{\max}$, and by construction $|\, F\tilde{x} + Hr \,| \leq u_{\max}$, we have

$$0 < w = u_{\max} - F\tilde{x} - Hr < u_N = \rho B'P\tilde{x}, \tag{5.92}$$

which implies that $\tilde{x}'PB < 0$ and hence $\dot{V} = -\tilde{x}'W\tilde{x} + 2\tilde{x}'PBw \leq -\tilde{x}'W\tilde{x}$.

Case 3. Finally, if $F\tilde{x} + Hr + u_{\mathrm{N}} < -u_{\max}$, we have

$$\rho B'P\tilde{x} = u_{\mathrm{N}} < w = -u_{\max} - F\tilde{x} - Hr < 0, \tag{5.93}$$

implying $\tilde{x}'PB > 0$ and hence $\dot{V} \leq -\tilde{x}'W\tilde{x}$.

In conclusion, we have shown that

$$\dot{V} \leq -\tilde{x}'W\tilde{x}, \quad \tilde{x} \in \boldsymbol{X}_{\delta}, \tag{5.94}$$

which implies that $\boldsymbol{X}_{\delta}$ is an invariant set of the closed-loop system in Equation (5.86) and all trajectories of Equation (5.86) starting from inside $\boldsymbol{X}_{\delta}$ will converge to the origin. This, in turn, indicates that, for all initial states x_0 and the step command input of amplitude r that satisfy Equation (5.73),

$$\lim_{t\to\infty} x(t) = x_{\mathrm{e}}. \tag{5.95}$$

Therefore,

$$\lim_{t\to\infty} h(t) = \lim_{t\to\infty} C_2 x(t) = C_2 x_{\mathrm{e}} = r. \tag{5.96}$$

This completes the proof of Theorem 5.5.1. ◊

Remark 5.5.2. Theorem 5.5.1 shows that the additional nonlinear feedback control law u_{N}, as given by Equation (5.84), does not affect the ability of the closed-loop system to track the command input. Any command input that can be asymptotically tracked by the linear feedback law of Equation (5.67) can also be asymptotically tracked by the CNF control law in Equation (5.85). However, this additional term u_{N} in the CNF control law can be used to improve the performance of the overall closed-loop system. This is the key property of the CNF control technique. ◊

ii. Full Order Measurement Feedback Case. The assumption that all the states of Σ are measurable is, in general, not practical. For example, in HDD servo systems, the velocity of the actuator is not usually directly measurable. Thus, one will have to implement the PTOS controller or the control law obtained in the previous case via a certain velocity estimation. In what follows, we proceed to develop a CNF design using only measurement information.

STEP 5.5.C.F.1: we first construct a linear full order measurement feedback control law:

$$\Sigma_{\text{F}}: \begin{cases} \dot{x}_{\text{v}} = (A + KC_1)x_{\text{v}} - Ky + B\,\text{sat}(u_{\text{L}}) \\ u_{\text{L}} = F(x_{\text{v}} - x_{\text{e}}) + Hr \end{cases} \tag{5.97}$$

where r is the reference input and $x_{\text{v}} \in \mathbb{R}^n$ is the state of the controller. As usual, F and K are gain matrices and are designed such that $A + BF$ and $A + KC_1$ are asymptotically stable and the resulting closed-loop system has the desired properties. As defined in Equations (5.71) and (5.72),

$$H = [1 - F(A + BF)^{-1}B]G, \quad G = -[C_2(A + BF)^{-1}B]^{-1}, \tag{5.98}$$

and $x_{\text{e}} = -(A + BF)^{-1}BGr$.

We have the following result.

Lemma 5.5.2. Given a positive definite matrix $W_{\text{P}} \in \mathbb{R}^{n\times n}$, let $P > 0$ be the solution to the Lyapunov equation

$$(A + BF)'P + P(A + BF) = -W_{\text{P}}. \tag{5.99}$$

Given another positive define matrix $W_{\text{Q}} \in \mathbb{R}^{n\times n}$ with

$$W_{\text{Q}} > F'B'PW_{\text{P}}^{-1}PBF, \tag{5.100}$$

let $Q > 0$ be the solution to the Lyapunov equation

$$(A + KC_1)'Q + Q(A + KC_1) = -W_{\text{Q}}. \tag{5.101}$$

Note that such P and Q exist as $A + BF$ and $A + KC_1$ are asymptotically stable. For any $\delta \in (0, 1)$, let c_δ be the largest positive scalar such that for all

$$\begin{pmatrix} x \\ x_{\text{v}} \end{pmatrix} \in \boldsymbol{X}_{\text{F}\delta} := \left\{ \begin{pmatrix} x \\ x_{\text{v}} \end{pmatrix} : \begin{pmatrix} x \\ x_{\text{v}} \end{pmatrix}' \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix} \begin{pmatrix} x \\ x_{\text{v}} \end{pmatrix} \le c_\delta \right\}, \tag{5.102}$$

we have

$$\left| [F \quad F] \begin{pmatrix} x \\ x_{\text{v}} \end{pmatrix} \right| \le u_{\max}(1 - \delta). \tag{5.103}$$

The linear control law in Equation (5.97) will drive the system controlled output $h(t)$ to track asymptotically a step command input of amplitude r from an initial state x_0, provided that x_0, $x_{v0} = x_v(0)$ and r satisfy:

$$|Hr| \leq \delta \cdot u_{max} \quad \text{and} \quad \begin{pmatrix} x_0 - x_e \\ x_{v0} - x_0 \end{pmatrix} \in \boldsymbol{X}_{F\delta}. \tag{5.104}$$

Proof. Let us transform the system coordinate by defining

$$\tilde{x} = x - x_e \quad \text{and} \quad \tilde{x}_v = x_v - x. \tag{5.105}$$

Then, the linear control law of Equation (5.97) can be written as

$$\dot{\tilde{x}}_v = (A + KC_1)\tilde{x}_v, \quad u_L = [F \quad F] \begin{pmatrix} \tilde{x} \\ \tilde{x}_v \end{pmatrix} + Hr \tag{5.106}$$

Hence, for all states

$$\begin{pmatrix} \tilde{x} \\ \tilde{x}_v \end{pmatrix} \in \boldsymbol{X}_{F\delta} \Rightarrow \left| [F \quad F] \begin{pmatrix} \tilde{x} \\ \tilde{x}_v \end{pmatrix} \right| \leq u_{max}(1 - \delta) \tag{5.107}$$

and for any r satisfying

$$|Hr| \leq \delta \cdot u_{max}, \tag{5.108}$$

we have

$$|u_L| = \left| [F \quad F] \begin{pmatrix} \tilde{x} \\ \tilde{x}_v \end{pmatrix} + Hr \right| \leq \left| [F \quad F] \begin{pmatrix} \tilde{x} \\ \tilde{x}_v \end{pmatrix} \right| + |Hr| = u_{max}. \tag{5.109}$$

Thus, for all $\tilde{x}$ and $\tilde{x}_v$ satisfying the condition as given in Equation (5.107), the closed-loop system comprising the given plant and the linear control law of Equation (5.97) can be rewritten as

$$\begin{pmatrix} \dot{\tilde{x}} \\ \dot{\tilde{x}}_v \end{pmatrix} = \begin{bmatrix} A + BF & BF \\ 0 & A + KC_1 \end{bmatrix} \begin{pmatrix} \tilde{x} \\ \tilde{x}_v \end{pmatrix}. \tag{5.110}$$

Next, we define a Lyapunov function for the closed-loop system in Equation (5.110):

$$V = \begin{pmatrix} \tilde{x} \\ \tilde{x}_v \end{pmatrix}' \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix} \begin{pmatrix} \tilde{x} \\ \tilde{x}_v \end{pmatrix}. \tag{5.111}$$

Along the trajectories of the closed-loop system in Equation (5.110) the derivative of the Lyapunov function is given by

$$\begin{aligned}\dot{V} &= \begin{pmatrix} \tilde{x} \\ \tilde{x}_{\mathrm{v}} \end{pmatrix}' \begin{bmatrix} -W_{\mathrm{P}} & PBF \\ F'B'P & -W_{\mathrm{Q}} \end{bmatrix} \begin{pmatrix} \tilde{x} \\ \tilde{x}_{\mathrm{v}} \end{pmatrix} \\ &= \begin{pmatrix} \hat{x} \\ \tilde{x}_{\mathrm{v}} \end{pmatrix}' \begin{bmatrix} -W_{\mathrm{P}} & 0 \\ 0 & -\tilde{W}_{\mathrm{Q}} \end{bmatrix} \begin{pmatrix} \hat{x} \\ \tilde{x}_{\mathrm{v}} \end{pmatrix}, \end{aligned} \tag{5.112}$$

where

$$\hat{x} = \tilde{x} - W_{\mathrm{P}}^{-1} PBF\tilde{x}_{\mathrm{v}}, \quad \tilde{W}_{\mathrm{Q}} = W_{\mathrm{Q}} - F'B'PW_{\mathrm{P}}^{-1}PBF. \tag{5.113}$$

With the choice of W_{Q} satisfying Equation (5.100), it is obvious that $\dot{V} \leq 0$. This shows that $\boldsymbol{X}_{\mathrm{F}\delta}$ is an invariant set of the closed-loop system in Equation (5.110) and all trajectories starting from the set converge asymptotically to the origin. Thus, for the initial states of x_0 and $x_{\mathrm{v}0}$ and step command inputs that satisfy Equation (5.104),

$$\lim_{t\to\infty} \tilde{x}_{\mathrm{v}}(t) = 0 \quad \text{and} \quad \lim_{t\to\infty} x(t) = x_{\mathrm{e}}, \tag{5.114}$$

which imply

$$\lim_{t\to\infty} h(t) = \lim_{t\to\infty} C_2 x(t) = C_2 x_{\mathrm{e}} = r. \tag{5.115}$$

This completes the proof of Lemma 5.5.2. ◊

STEP 5.5.C.F.2: as in the state feedback case, the linear control law of Equation (5.97) obtained in the above step is to be combined with a nonlinear control law to form the following CNF controller:

$$\begin{cases} \dot{x}_{\mathrm{v}} &= (A + KC_1)x_{\mathrm{v}} - Ky + B\,\mathrm{sat}(u), \\ u &= F(x_{\mathrm{v}} - x_{\mathrm{e}}) + Hr + \rho(r,y)B'P(x_{\mathrm{v}} - x_{\mathrm{e}}), \end{cases} \tag{5.116}$$

where $\rho(r,y)$ is a nonpositive scalar function, locally Lipschitz in y, and is to be chosen to improve the performance of the closed-loop system.

It turns out that, for the measurement feedback case, the choice of $\rho(r,y)$, the nonpositive scalar function, is not totally free. It is subject to certain constraints. We have the following theorem.

Theorem 5.5.2. Consider the given system in Equation (5.65). Then, there exists a scalar $\rho^* > 0$ such that for any nonpositive function $\rho(r, y)$, locally Lipschitz in y and $|\rho(r, y)| \leq \rho^*$, the CNF control law of Equation (5.116) will drive the system controlled output $h(t)$ to track asymptotically the step command input of amplitude r from an initial state x_0, provided that x_0, x_{v0} and r satisfy Equation (5.104). ◇

Proof: Again, let $\tilde{x} = x - x_e$ and $\tilde{x}_v = x_v - x$. For simplicity, we drop r and h in $\rho(r, y)$ throughout the rest of the proof of this theorem. Then, the closed-loop system with the CNF control law of Equation (5.116) can be expressed as

$$\begin{pmatrix} \dot{\tilde{x}} \\ \dot{\tilde{x}}_v \end{pmatrix} = \begin{bmatrix} A + BF & BF \\ 0 & A + KC_1 \end{bmatrix} \begin{pmatrix} \tilde{x} \\ \tilde{x}_v \end{pmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} w, \tag{5.117}$$

where

$$w = \text{sat}\left[[F \quad F] \begin{pmatrix} \tilde{x} \\ \tilde{x}_v \end{pmatrix} + Hr + \rho [B'P \quad B'P] \begin{pmatrix} \tilde{x} \\ \tilde{x}_v \end{pmatrix} \right] - [F \quad F] \begin{pmatrix} \tilde{x} \\ \tilde{x}_v \end{pmatrix} - Hr. \tag{5.118}$$

Clearly, for the given x_0 and x_{v0} satisfying Equation (5.104), we have

$$\begin{pmatrix} \tilde{x}(0) \\ \tilde{x}_v(0) \end{pmatrix} \in \boldsymbol{X}_{F\delta}. \tag{5.119}$$

Using the following Lyapunov function:

$$V = \begin{pmatrix} \tilde{x} \\ \tilde{x}_v \end{pmatrix}' \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix} \begin{pmatrix} \tilde{x} \\ \tilde{x}_v \end{pmatrix}, \tag{5.120}$$

we evaluate the derivative of V along the trajectories of the closed-loop system in Equation (5.117), *i.e.*,

$$\dot{V} = \begin{pmatrix} \tilde{x} \\ \tilde{x}_v \end{pmatrix}' \begin{bmatrix} -W_P & PBF \\ F'B'P & -W_Q \end{bmatrix} \begin{pmatrix} \tilde{x} \\ \tilde{x}_v \end{pmatrix} + 2\tilde{x}'PBw. \tag{5.121}$$

Note that for all

$$\begin{pmatrix} \tilde{x} \\ \tilde{x}_v \end{pmatrix} \in \boldsymbol{X}_{F\delta} \quad \Rightarrow \quad \left| [F \quad F] \begin{pmatrix} \tilde{x} \\ \tilde{x}_v \end{pmatrix} \right| \leq u_{\max}(1 - \delta). \tag{5.122}$$

Again, as done in the full state feedback case, let us find the above derivative of V for three different cases.

Case 1. If

$$\left|[F \quad F]\begin{pmatrix}\tilde{x}\\ \tilde{x}_{\mathrm{V}}\end{pmatrix} + Hr + \rho[B'P \quad B'P]\begin{pmatrix}\tilde{x}\\ \tilde{x}_{\mathrm{V}}\end{pmatrix}\right| \leq u_{\max}, \tag{5.123}$$

then

$$w = \rho[B'P \quad B'P]\begin{pmatrix}\tilde{x}\\ \tilde{x}_{\mathrm{V}}\end{pmatrix}, \tag{5.124}$$

which implies

$$\begin{aligned}\dot{V} &= \begin{pmatrix}\tilde{x}\\ \tilde{x}_{\mathrm{V}}\end{pmatrix}'\begin{bmatrix}-W_{\mathrm{P}} & PB(F+\rho B'P)\\ (F+\rho B'P)'B'P & -W_{\mathrm{Q}}\end{bmatrix}\begin{pmatrix}\tilde{x}\\ \tilde{x}_{\mathrm{V}}\end{pmatrix} + 2\rho\tilde{x}'PBB'P\tilde{x}\\ &\leq \begin{pmatrix}\hat{x}\\ \tilde{x}_{\mathrm{V}}\end{pmatrix}'\begin{bmatrix}-W_{\mathrm{P}} & 0\\ 0 & -\tilde{W}_{\mathrm{Q}}\end{bmatrix}\begin{pmatrix}\hat{x}\\ \tilde{x}_{\mathrm{V}}\end{pmatrix},\end{aligned} \tag{5.125}$$

where

$$\hat{x} = \tilde{x} - W_{\mathrm{P}}^{-1}PB(F+\rho B'P)\tilde{x}_{\mathrm{V}}, \tag{5.126}$$

$$\tilde{W}_{\mathrm{Q}} = W_{\mathrm{Q}} - (F+\rho B'P)'B'PW_{\mathrm{P}}^{-1}PB(F+\rho B'P). \tag{5.127}$$

Noting Equation (5.100), *i.e.*, $W_{\mathrm{Q}} > F'B'PW_{\mathrm{P}}^{-1}PBF$, and $\rho(r,y)$ is locally Lipschitz, it is clear that there exists a $\rho_1^* > 0$ such that for any scalar function satisfying $|\rho(r,y)| \leq \rho_1^*$ we have $\tilde{W}_{\mathrm{Q}} > 0$ and hence $\dot{V} \leq 0$.

Case 2. If

$$[F \quad F]\begin{pmatrix}\tilde{x}\\ \tilde{x}_{\mathrm{V}}\end{pmatrix} + Hr + \rho[B'P \quad B'P]\begin{pmatrix}\tilde{x}\\ \tilde{x}_{\mathrm{V}}\end{pmatrix} > u_{\max}, \tag{5.128}$$

then for the trajectories inside $\boldsymbol{X}_{\mathrm{F}\delta}$,

$$\left|[F \quad F]\begin{pmatrix}\tilde{x}\\ \tilde{x}_{\mathrm{V}}\end{pmatrix} + Hr\right| \leq u_{\max}, \tag{5.129}$$

which implies that

$$0 < w < \rho[B'P \quad B'P]\begin{pmatrix}\tilde{x}\\ \tilde{x}_{\mathrm{V}}\end{pmatrix}. \tag{5.130}$$

Next, let us express

$$w = q\rho \begin{bmatrix} B'P & B'P \end{bmatrix} \begin{pmatrix} \tilde{x} \\ \tilde{x}_{\mathrm{v}} \end{pmatrix}, \tag{5.131}$$

for an appropriate positive piecewise continuous function $q(t)$, bounded by 1 for all t. In this case, the derivative of V becomes

$$\begin{aligned} \dot{V} &= \begin{pmatrix} \tilde{x} \\ \tilde{x}_{\mathrm{v}} \end{pmatrix}' \begin{bmatrix} -W_{\mathrm{P}} & PB(F + q\rho B'P) \\ (F + q\rho B'P)'B'P & -W_{\mathrm{Q}} \end{bmatrix} \begin{pmatrix} \tilde{x} \\ \tilde{x}_{\mathrm{v}} \end{pmatrix} + 2q\rho \tilde{x}' PBB'P\tilde{x} \\ &\leq \begin{pmatrix} \hat{x}_{+} \\ \tilde{x}_{\mathrm{v}} \end{pmatrix}' \begin{bmatrix} -W_{\mathrm{P}} & 0 \\ 0 & -\tilde{W}_{\mathrm{Q}_{+}} \end{bmatrix} \begin{pmatrix} \hat{x}_{+} \\ \tilde{x}_{\mathrm{v}} \end{pmatrix}, \end{aligned} \tag{5.132}$$

where

$$\hat{x}_{+} = \tilde{x} - W_{\mathrm{P}}^{-1} PB(F + q\rho B'P)\tilde{x}_{\mathrm{v}}, \tag{5.133}$$

$$\tilde{W}_{\mathrm{Q}_{+}} = W_{\mathrm{Q}} - (F + q\rho B'P)'B'PW_{\mathrm{P}}^{-1} PB(F + q\rho B'P). \tag{5.134}$$

Again, noting Equation (5.100), it can be shown that there exists a ρ_2^* such that for any $\rho(r, y)$ satisfying $|\rho(r, y)| \leq \rho_2^*$ we have $\tilde{W}_{\mathrm{Q}_{+}} > 0$ and hence $\dot{V} \leq 0$.

Case 3. Similarly, for the case when

$$\begin{bmatrix} F & F \end{bmatrix} \begin{pmatrix} \tilde{x} \\ \tilde{x}_{\mathrm{v}} \end{pmatrix} + Hr + \rho \begin{bmatrix} B'P & B'P \end{bmatrix} \begin{pmatrix} \tilde{x} \\ \tilde{x}_{\mathrm{v}} \end{pmatrix} < -u_{\max}, \tag{5.135}$$

we can show that there exists a $\rho_3^* > 0$ such that for any $\rho(r, y)$ satisfying $|\rho(r, y)| \leq \rho_3^*$, we have $\dot{V} \leq 0$ for all the trajectories in $\boldsymbol{X}_{\mathrm{F}\delta}$.

Finally, let $\rho^* = \min\{\rho_1^*, \rho_2^*, \rho_3^*\}$. Then, we have for any nonpositive scalar function $\rho(r, y)$ satisfying $|\rho(r, y)| \leq \rho^*$,

$$\dot{V} \leq 0, \quad \forall \begin{pmatrix} \tilde{x} \\ \tilde{x}_{\mathrm{v}} \end{pmatrix} \in \boldsymbol{X}_{\mathrm{F}\delta}. \tag{5.136}$$

Thus, $\boldsymbol{X}_{\mathrm{F}\delta}$ is an invariant set of the closed-loop system in Equation (5.117), and all trajectories starting from $\boldsymbol{X}_{\mathrm{F}\delta}$ will remain inside and asymptotically converge to the origin. This, in turn, indicates that, for the initial state of the given system x_0, the initial state of the controller $x_{\mathrm{v}0}$, and step command input r that satisfy Equation (5.104),

$$\lim_{t\to\infty} \tilde{x}_{\mathrm{v}}(t) = 0 \quad \text{and} \quad \lim_{t\to\infty} x(t) = x_{\mathrm{e}}, \tag{5.137}$$

and hence

$$\lim_{t\to\infty} h(t) = \lim_{t\to\infty} C_2 x(t) = C_2 x_{\mathrm{e}} = r. \tag{5.138}$$

This completes the proof of Theorem 5.5.2. ◊

iii. Reduced Order Measurement Feedback Case. For the given system in Equation (5.65), it is clear that there are p states of the system measurable if C_1 is of maximal rank. Thus, in general, it is not necessary to estimate these measurable states in measurement feedback laws. As such, we will design a dynamic controller that has a dynamical order less than that of the given plant. We now proceed to construct such a control law under the CNF control framework.

For simplicity of presentation, we assume that C_1 is already in the form

$$C_1 = [\, I_p \quad 0 \,]. \tag{5.139}$$

Then, the system in Equation (5.65) can be rewritten as

$$\begin{cases} \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \mathrm{sat}(u), \; x_0 = \begin{pmatrix} x_{10} \\ x_{20} \end{pmatrix} \\ y = [\; I_p \quad 0 \;] \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ h = C_2 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \end{cases} \tag{5.140}$$

where the original state x is partitioned into two parts, x_1 and x_2 with $y \equiv x_1$. Thus, we will only need to estimate x_2 in the reduced order measurement feedback design. Next, we let F be chosen such that (1) $A+BF$ is asymptotically stable, and (2) $C_2(sI - A - BF)^{-1}B$ has desired properties, and let K_{R} be chosen such that $A_{22} + K_{\mathrm{R}}A_{12}$ is asymptotically stable. Here we note that it was shown Chen [87] that (A_{22}, A_{12}) is detectable if and only if (A, C_1) is detectable. Thus, there exists a stabilising K_{R}. Again, such F and K_{R} can be designed using any of the linear control techniques presented in Chapter 4. We then partition F in conformity with x_1 and x_2:

$$F = [\, F_1 \quad F_2 \,]. \tag{5.141}$$

As defined in Equations (5.71) and (5.72),

$$H = [1 - F(A + BF)^{-1}B]G, \quad G = -[C_2(A + BF)^{-1}B]^{-1}, \tag{5.142}$$

and $x_e = -(A+BF)^{-1}BGr$.

The reduced order CNF controller is given by

$$\dot{x}_v = (A_{22}+K_R A_{12})x_v + \left[A_{21}+K_R A_{11}-(A_{22}+K_R A_{12})K_R\right]y + (B_2+K_R B_1)\,\mathrm{sat}(u) \tag{5.143}$$

and

$$u = F\left[\begin{pmatrix} y \\ x_v - K_R y\end{pmatrix} - x_e\right] + Hr + \rho(r,y)B'P\left[\begin{pmatrix} y \\ x_v - K_R y\end{pmatrix} - x_e\right] \tag{5.144}$$

where $\rho(r,y)$ is a nonpositive scalar function locally Lipschitz in y subject to certain constraints to be discussed later.

Next, given a positive definite matrix $W_P \in \mathbb{R}^{n\times n}$, let $P > 0$ be the solution to the Lyapunov equation

$$(A+BF)'P + P(A+BF) = -W_P. \tag{5.145}$$

Given another positive define matrix $W_R \in \mathbb{R}^{(n-p)\times(n-p)}$ with

$$W_R > F_2' B' P W_P^{-1} P B F_2, \tag{5.146}$$

let $Q_R > 0$ be the solution to the Lyapunov equation

$$(A_{22}+K_R A_{12})'Q_R + Q_R(A_{22}+K_R A_{12}) = -W_R. \tag{5.147}$$

Note that such P and Q_R exist as $A+BF$ and $A_{22}+K_R A_{12}$ are asymptotically stable. For any $\delta \in (0,1)$, let c_δ be the largest positive scalar such that for all

$$\begin{pmatrix} x \\ x_v\end{pmatrix} \in \boldsymbol{X}_{R\delta} := \left\{\begin{pmatrix} x \\ x_v\end{pmatrix} : \begin{pmatrix} x \\ x_v\end{pmatrix}' \begin{bmatrix} P & 0 \\ 0 & Q_R\end{bmatrix} \begin{pmatrix} x \\ x_v\end{pmatrix} \le c_\delta\right\}, \tag{5.148}$$

we have

$$\left|\begin{bmatrix} F & F_2\end{bmatrix}\begin{pmatrix} x \\ x_v\end{pmatrix}\right| \le u_{\max}(1-\delta). \tag{5.149}$$

We have the following theorem.

Theorem 5.5.3. Consider the given system in Equation (5.65). Then, there exists a scalar $\rho^* > 0$ such that for any nonpositive function $\rho(r, y)$, locally Lipschitz in y and $|\rho(r, y)| \leq \rho^*$, the reduced order CNF law given by Equations (5.143) and (5.144) will drive the system controlled output $h(t)$ to track asymptotically the step command input of amplitude r from an initial state x_0, provided that x_0, x_{v0} and r satisfy

$$\begin{pmatrix} x_0 - x_e \\ x_{v0} - x_{20} - K_R x_{10} \end{pmatrix} \in \boldsymbol{X}_{R\delta} \quad \text{and} \quad |Hr| \leq \delta \cdot u_{max}. \tag{5.150}$$

Proof: Let $\tilde{x} = x - x_e$ and $\tilde{x}_v = x_v - x_2 - K_R x_1$. Then, the closed-loop system comprising the given plant in Equation (5.65) and the reduced order CNF control law of (5.143) and (5.144) can be expressed as

$$\begin{pmatrix} \dot{\tilde{x}} \\ \dot{\tilde{x}}_v \end{pmatrix} = \begin{bmatrix} A + BF & BF_2 \\ 0 & A_{22} + K_R A_{12} \end{bmatrix} \begin{pmatrix} \tilde{x} \\ \tilde{x}_v \end{pmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} w, \tag{5.151}$$

where

$$w = \text{sat}\left\{ [F \quad F_2] \begin{pmatrix} \tilde{x} \\ \tilde{x}_v \end{pmatrix} + Hr + \rho(r, y) B' P \left[\tilde{x} + \begin{pmatrix} 0 \\ \tilde{x}_v \end{pmatrix} \right] \right\} - [F \quad F_2] \begin{pmatrix} \tilde{x} \\ \tilde{x}_v \end{pmatrix} - Hr. \tag{5.152}$$

The rest of the proof follows along similar lines to the reasoning given in the full order measurement feedback case. ◇

5.5.2 Discrete-time Systems

Let us now consider a linear discrete-time system Σ with an amplitude-constrained actuator characterised by

$$\begin{cases} x(k+1) = A\ x(k) + B\,\text{sat}[u(k)], \quad x(0) = x_0 \\ y(k) \quad = C_1\ x(k) \\ h(k) \quad = C_2\ x(k) \end{cases} \tag{5.153}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}$, $y \in \mathbb{R}^p$ and $h \in \mathbb{R}$ are respectively the state, control input, measurement output and controlled output of Σ. A, B, C_1 and C_2 are appropriate dimensional constant matrices, and sat: $\mathbb{R} \to \mathbb{R}$ represents the actuator saturation defined as

$$\text{sat}(u) = \text{sgn}(u) \min\{ u_{\max}, \mid u \mid \}, \tag{5.154}$$

with $u_{\max}$ being the saturation level of the input. The following assumptions on the system matrices are required:

1. (A, B) is stabilisable,
2. (A, C_1) is detectable, and
3. (A, B, C_2) is invertible and has no invariant zeros at $z = 1$.

We now extend the results of the continuous-time composite nonlinear control method to the discrete-time system in Equation (5.153). Thus, the objective here is to design a discrete-time CNF law that will cause the output to track a high-amplitude step input rapidly without experiencing large overshoot and without the adverse actuator saturation effects. This can be done through the design of a discrete-time linear feedback law with a small closed-loop damping ratio and a nonlinear feedback law through an appropriate Lyapunov function to cause the closed-loop system to be highly damped as system output approaches the command input to reduce the overshoot. The result of this discrete-time version is analogous to that of its continuous-time counterpart. Here, we again separate the design of discrete-time CNF control into three distinct situations, *i.e.*, (1) the state feedback case, (2) the full order measurement case, and (3) the reduced order measurement feedback case.

i. State Feedback Case. We consider the case when $y = x$, *i.e.*, all the state variables of Σ of Equation (5.153) are available for feedback.

STEP 5.5.D.S.1: design a linear feedback law,

$$u_{\text{L}}(k) = Fx(k) + Gr, \tag{5.155}$$

where r is the input command, and F is chosen such that $A + BF$ has all its eigenvalues in $\mathbb{C}^{\odot}$ and the closed-loop system $C_2(zI - A - BF)^{-1}B$ meets certain design specifications. We note again that such an F can be designed using any of the techniques reported in Chapter 4. Furthermore,

$$G = \left[C_2(I - A - BF)^{-1}B\right]^{-1}. \tag{5.156}$$

We note that G is well defined because $A + BF$ has all its eigenvalues in $\mathbb{C}^{\odot}$, and (A, B, C_2) is invertible and has no invariant zeros at $z = 1$.

The following lemma determines the magnitude of r that can be tracked by such a control law without exceeding the control limits.

Lemma 5.5.3. Given a positive definite matrix $W \in \mathbb{R}^{n \times n}$, let $P > 0$ be the solution of the following Lyapunov equation:

$$P = (A + BF)'P(A + BF) + W. \tag{5.157}$$

Such a P exists as $A + BF$ is asymptotically stable. For any $\delta \in (0, 1)$, let $c_\delta > 0$ be the largest positive scalar such that

$$| Fx(k) | \leq u_{\max}(1 - \delta), \quad \forall\, x(k) \in \boldsymbol{X}_\delta := \left\{ x : x'Px \leq c_\delta \right\}. \tag{5.158}$$

Also, let

$$H := \left[1 + F(I - A - BF)^{-1}B\right] G, \tag{5.159}$$

and

$$x_{\mathrm{e}} := G_{\mathrm{e}}\, r := (I - A - BF)^{-1}BG\, r. \tag{5.160}$$

Then, the control law in Equation (5.155) is capable of driving the system controlled output $h(k)$ to track asymptotically a step command input of amplitude r, provided that the initial state x_0 and r satisfy:

$$(x_0 - x_{\mathrm{e}}) \in \boldsymbol{X}_\delta \quad \text{and} \quad |Hr| \leq \delta \cdot u_{\max}. \tag{5.161}$$

Proof. Let $\tilde{x} = x - x_{\mathrm{e}}$. Then, the linear feedback control law u_{L} can be rewritten as

$$u_{\mathrm{L}}(k) = F\tilde{x}(k) + \left[1 + F(I - A - BF)^{-1}B\right] Gr = F\tilde{x}(k) + Hr. \tag{5.162}$$

Hence, for all

$$\tilde{x}(k) \in \boldsymbol{X}_\delta \Rightarrow | F\tilde{x}(k) | \leq u_{\max}(1 - \delta) \tag{5.163}$$

and for any r satisfying

$$|Hr| \leq \delta \cdot u_{\max}, \tag{5.164}$$

the linear control law can be written as

$$| u_{\mathrm{L}}(k) | = |F\tilde{x}(k) + Hr| \leq | F\tilde{x}(k) | + |Hr| \leq u_{\max}, \tag{5.165}$$

which indicates that the control signal $u_{\text{L}}(k)$ will never exceed the saturation. Next, let us move to verify the asymptotic stability of the closed-loop system comprising the given plant in Equation (5.153) and the linear feedback law in Equation (5.155), which can be expressed as follows:

$$\tilde{x}(k+1) = (A + BF)\tilde{x}(k). \tag{5.166}$$

Let us define a Lyapunov function for the closed-loop system in Equation (5.166) as

$$V(k) = \tilde{x}'(k)P\tilde{x}(k). \tag{5.167}$$

Along the trajectories of the closed-loop system in Equation (5.166) the increment of the Lyapunov function in Equation (5.167) is given by

$$\begin{aligned} \nabla V(k+1) &= \tilde{x}'(k+1)P\tilde{x}(k+1) - \tilde{x}'(k)P\tilde{x}(k) \\ &= \tilde{x}'(k)(A+BF)'P(A+BF)\tilde{x}(k) - \tilde{x}'(k)P\tilde{x}(k) \\ &= -\tilde{x}'(k)W\tilde{x}(k) \\ &\leq 0. \end{aligned} \tag{5.168}$$

This shows that $\boldsymbol{X}_{\delta}$ is an invariant set of the the closed-loop system in Equation (5.166) and all trajectories of Equation (5.166) starting from $\boldsymbol{X}_{\delta}$ will converge to the origin. Thus, for any initial state x_0 and the step command input r that satisfy Equation (5.161), we have

$$\lim_{k\to\infty} x(k) = x_{\text{e}}, \tag{5.169}$$

and hence

$$\lim_{k\to\infty} h(k) = \lim_{k\to\infty} C_2 x(k) = C_2 x_{\text{e}} = r. \tag{5.170}$$

This completes the proof of Lemma 5.5.3. ◊

Remark 5.5.3. We would like to note that, for the case when $x_0 = 0$, any step command of amplitude r can be tracked asymptotically provided that

$$| r | \leq \left[c_{\delta}(G_{\text{e}}'PG_{\text{e}})^{-1}\right]^{1/2} \quad \text{and} \quad | Hr | \leq \delta \cdot u_{\max}, \tag{5.171}$$

This input command amplitude can be increased by increasing δ and/or decreasing $G_{\text{e}}'PG_{\text{e}}$ through the choice of W. However, the change in F will of course affect the damping ratio of the closed-loop system and hence its rising time. ◊

STEP 5.5.D.S.2: the nonlinear feedback control law $u_{\mathrm{N}}(k)$ is given by

$$u_{\mathrm{N}}(k) = \rho(r,y)B'P(A+BF)\Big[x(k) - x_{\mathrm{e}}\Big], \tag{5.172}$$

where $\rho(r,y)$ is a nonpositive scalar function, locally Lipschitz in y, and is to be used to change the system closed-loop damping ratio as the output approaches the step command input. The choice of $\rho(r,y)$ will be discussed later in detail.

STEP 5.5.D.S.3: the linear and nonlinear components derived above are now combined to form a discrete-time CNF law:

$$u(k) = u_{\mathrm{L}}(k) + u_{\mathrm{N}}(k). \tag{5.173}$$

We have the following result.

Theorem 5.5.4. Consider the discrete-time system in Equation (5.153). Then, for any nonpositive $\rho(r,y)$, locally Lipschitz in y and $|\rho(r,y)| \leq \rho^* := 2(B'PB)^{-1}$, the CNF law in Equation (5.173) is capable of driving the system controlled output $h(k)$ to track the step command input of amplitude r from an initial state x_0, provided that x_0 and r satisfy Equation (5.161). $\diamondsuit$

Proof. Let $\tilde{x} = x - x_{\mathrm{e}}$. Then, the closed-loop system can be written as

$$\tilde{x}(k+1) = (A+BF)\tilde{x}(k) + Bw(k), \tag{5.174}$$

where

$$w(k) = \mathrm{sat}\Big[F\tilde{x}(k) + Hr + u_{\mathrm{N}}(k)\Big] - F\tilde{x}(k) - Hr. \tag{5.175}$$

Equation (5.161) implies that $\tilde{x}_0 \in \boldsymbol{X}_\delta$. Define a Lyapunov function

$$V(k) = \tilde{x}'(k)P\tilde{x}(k). \tag{5.176}$$

Noting that

$$\tilde{x}(k) \in \boldsymbol{X}_\delta \quad \Rightarrow \quad |\, F\tilde{x}(k)\,| \leq u_{\max}(1-\delta), \tag{5.177}$$

we can evaluate the increment of $V(k)$ along the trajectories of the closed-loop system in Equation (5.174) as follows:

$$\begin{aligned}\nabla V(k+1) &= \tilde{x}'(k+1)P\tilde{x}(k+1) - \tilde{x}'(k)P\tilde{x}(k) \\ &= \tilde{x}'(k)(A+BF)'P(A+BF)\tilde{x}(k) - \tilde{x}'(k)P\tilde{x}(k) \\ &\quad +2\tilde{x}'(k)(A+BF)'PBw(k) + w'(k)B'PBw(k) \\ &= -\tilde{x}'(k)W\tilde{x}(k) + 2\tilde{x}'(k)(A+BF)'PBw(k) \\ &\quad + w'(k)B'PBw(k) \end{aligned} \tag{5.178}$$

Next, we proceed to find the increment of $V(k)$ for three different cases, as done in continuous-time systems.

If $|F\tilde{x}(k) + Hr + u_{\mathrm{N}}(k)| \leq u_{\max}$, then

$$w(k) = u_{\mathrm{N}}(k) = \rho B'P(A+BF)\tilde{x}(k). \tag{5.179}$$

Thus,

$$\begin{aligned}\nabla V(k+1) &= -\tilde{x}'(k)W\tilde{x}(k) + 2\tilde{x}'(k)(A+BF)'PBw(k) \\ &\quad + w'(k)B'PBw(k) \\ &= -\tilde{x}'(k)W\tilde{x}(k) \\ &\quad + \rho\tilde{x}'(k)(A+BF)'PB(2+\rho B'PB)B'P(A+BF)\tilde{x}(k). \end{aligned} \tag{5.180}$$

For any nonpositive $\rho(r,y)$ with $|\rho(r,y)| \leq \rho^*$, it is clear that the increment $\nabla V(k+1) \leq -\tilde{x}'(k)W\tilde{x}(k) \leq 0$.

If $F\tilde{x}(k) + Hr + u_{\mathrm{N}}(k) > u_{\max}$, then $| F\tilde{x}(k) + Hr |\leq u_{\max}$ implies that $0 < w(k) < u_{\mathrm{N}}(k)$ and $\rho(r,y) < 0$. Hence,

$$\begin{aligned}\nabla V(k+1) &= -\tilde{x}'(k)W\tilde{x}(k) + 2\tilde{x}'(k)(A+BF)'PBw(k) \\ &\quad + w'(k)B'PBw(k) \\ &= -\tilde{x}'(k)W\tilde{x}(k) + w'(k)[2\rho^{-1}u_{\mathrm{N}}(k) + B'PBw(k)] \\ &< -\tilde{x}'(k)W\tilde{x}(k) + w'(k)[2\rho^{-1}u_{\mathrm{N}}(k) + B'PBu_{\mathrm{N}}(k)] \\ &= -\tilde{x}'(k)W\tilde{x}(k) + w'(k)(2\rho^{-1} + B'PB)u_{\mathrm{N}}(k). \end{aligned} \tag{5.181}$$

Thus, for all $-\rho^* \leq \rho(r,y) < 0$, we have $2\rho^{-1} + B'PB \leq 0$, and hence

$$\nabla V(k+1) \leq -\tilde{x}'(k)W\tilde{x}(k) \leq 0. \tag{5.182}$$

Similarly, for the case when $F\tilde{x}(k) + Hr + u_{\mathrm{N}}(k) < -u_{\max}$, it can be shown that $\nabla V(k+1) \leq -\tilde{x}'(k)W\tilde{x}(k) \leq 0$.

Thus, $\boldsymbol{X}_\delta$ is an invariant set of the closed-loop system in Equation (5.174) and all trajectories of Equation (5.174) starting from $\boldsymbol{X}_\delta$ will remain there

and converge to the origin. This, in turn, indicates that, for all initial states x_0 and the step command input of amplitude r that satisfy Equation (5.161),

$$\lim_{k\to\infty} x(k) = x_e, \tag{5.183}$$

and

$$\lim_{k\to\infty} y(k) = \lim_{k\to\infty} C_2 x(k) = C_2 x_e = r. \tag{5.184}$$

This completes the proof of Theorem 5.5.4. ◊

Remark 5.5.4. Theorem 5.5.4 shows that the addition of the nonlinear feedback control law u_N as given in Equation (5.172) does not affect the ability to track the class of command inputs. Any command input that can be tracked by the linear feedback law in Equation (5.155) can also be tracked by the CNF control law (5.173). The composite feedback law in Equation (5.173) does not reduce the level of the trackable command input for any choice of the function $\rho(r, y)$. This freedom can be used to improve the performance of the overall system. The choice of $\rho(r, y)$ will be discussed in the coming subsection. ◊

ii. Full Order Measurement Feedback Case. We next proceed to construct a discrete-time full order CNF control law.

STEP 5.5.D.F.1: we first construct a linear full order measurement feedback control law

$$\begin{cases} x_v(k+1) = (A + KC_1)x_v(k) - Ky(k) + B\,\mathrm{sat}[u_L(k)] \\ u_L(k) \quad = F\Big[x_v(k) - x_e\Big] + H\,r \end{cases} \tag{5.185}$$

where r is the command input, $x_v \in \mathbb{R}^n$ is the state of the controller, F and K are chosen such that $A + BF$ and $A + KC_1$ have all their eigenvalues in $\mathbb{C}^\odot$, *i.e.*, both are stable matrices, and, furthermore, the resulting closed-loop system has met certain design specifications. As usual, we let

$$x_e = (I - A - BF)^{-1}B \cdot [C_2(I - A - BF)^{-1}B]^{-1}r, \tag{5.186}$$

$$H = [1 + F(I - A - BF)^{-1}B] \cdot [C_2(I - A - BF)^{-1}B]^{-1}. \tag{5.187}$$

We note that both x_e and H are well defined.

Lemma 5.5.4. Given a positive definite matrix $W_P \in \mathbb{R}^{n\times n}$, let $P > 0$ be the solution to the Lyapunov equation

$$P = (A + BF)'P(A + BF) + W_P. \tag{5.188}$$

Given another positive define matrix $W_Q \in \mathbb{R}^{n\times n}$ with

$$W_Q > F'B'P(A + BF)W_P^{-1}(A + BF)'PBF, \tag{5.189}$$

let $Q > 0$ be the solution to the Lyapunov equation

$$Q = (A + KC_1)'Q(A + KC_1) + W_Q. \tag{5.190}$$

Note that such P and Q exist as $A + BF$ and $A + KC_1$ are asymptotically stable. For any $\delta \in (0, 1)$, let c_δ be the largest positive scalar such that for all

$$\begin{pmatrix} x \\ x_v \end{pmatrix} \in \boldsymbol{X}_{F\delta} := \left\{ \begin{pmatrix} x \\ x_v \end{pmatrix} : \begin{pmatrix} x \\ x_v \end{pmatrix}' \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix} \begin{pmatrix} x \\ x_v \end{pmatrix} \le c_\delta \right\}, \tag{5.191}$$

we have

$$\left| \begin{bmatrix} F & F \end{bmatrix} \begin{pmatrix} x \\ x_v \end{pmatrix} \right| \le u_{\max}(1 - \delta). \tag{5.192}$$

The linear control law in Equation (5.185) will drive the system controlled output $h(k)$ to track asymptotically a step command input of amplitude r from an initial state x_0, provided that x_0, $x_{v0} = x_v(0)$ and r satisfy:

$$|Hr| \le \delta \cdot u_{\max} \quad \text{and} \quad \begin{pmatrix} x_0 - x_e \\ x_{v0} - x_0 \end{pmatrix} \in \boldsymbol{X}_{F\delta}. \tag{5.193}$$

Proof. This follows along similar lines to the reasoning given in the proofs of Lemmas 5.5.2 and 5.5.3. ◊

STEP 5.5.D.F.2: the discrete-time full order measurement composite nonlinear feedback control law is given by

$$x_v(k + 1) = (A + KC_1)x_v(k) - Ky(k) + B\text{sat}\Big[u(k)\Big] \tag{5.194}$$

and

$$u(k)=F\left[x_{\mathrm{v}}(k)-x_{\mathrm{e}}\right]+Hr+\rho(r,y)B'P(A+BF)\left[x_{\mathrm{v}}(k)-x_{\mathrm{e}}\right] \quad (5.195)$$

where $\rho(r,y)$ is a nonpositive scalar function, locally Lipschitz in y, and is to be chosen to improve the performance of the closed-loop system.

We have the following result.

Theorem 5.5.5. Consider the given discrete-time system in Equation (5.153). Then, there exists a scalar $0 < \rho^* \leq 2(B'PB)^{-1}$ such that for any nonpositive function $\rho(r,y)$, locally Lipschitz in y and $|\rho(r,y)| \leq \rho^*$, the discrete-time CNF law in Equations (5.194) and (5.195) will drive the system controlled output $h(k)$ to track asymptotically the step command input of amplitude r from an initial state x_0, provided that x_0, $x_{\mathrm{v}0}$ and r satisfy the conditions in Equation (5.193). ◊

Proof. The proof of this theorem follows along similar lines to the reasoning given in Theorems 5.5.2 and 5.5.4. ◊

iii. Reduced Order Measurement Feedback Case. As in its continuous-time counterpart, we now proceed to design a reduced order measurement feedback controller. For the given system in Equation (5.153), it is clear that p states of the system are measurable if C_1 is of maximal rank. As such, we could design a dynamic controller that has a dynamical order less than that of the given plant. We now proceed to construct such a control law under the CNF control framework.

For simplicity of presentation, we assume that C_1 is already in the form

$$C_1 = [I_p \quad 0]. \quad (5.196)$$

Then, the system in Equation (5.153) can be rewritten as

$$\begin{cases} \begin{pmatrix} x_1(k+1) \\ x_2(k+1) \end{pmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{pmatrix} x_1(k) \\ x_2(k) \end{pmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \mathrm{sat}[u(k)] \\ y(k) = [\ I_p \quad 0\ ] \begin{pmatrix} x_1(k) \\ x_2(k) \end{pmatrix} \\ h(k) = C_2 \begin{pmatrix} x_1(k) \\ x_2(k) \end{pmatrix} \end{cases} \quad (5.197)$$

and

$$x_0 = \begin{pmatrix} x_{10} \\ x_{20} \end{pmatrix}, \quad (5.198)$$

where the original state x is partitioned into two parts, x_1 and x_2 with $y \equiv x_1$. Thus, we will only need to estimate x_2 in the reduced order measurement feedback design. Next, we let F be chosen such that (1) $A + BF$ is asymptotically stable, and (2) $C_2(sI - A - BF)^{-1}B$ has the desired properties, and let K_R be chosen such that $A_{22} + K_R A_{12}$ is asymptotically stable. Again, it follows from Chen [87] that (A_{22}, A_{12}) is detectable if and only if (A, C_1) is detectable. Thus, there exists a stabilising K_R. Again, such F and K_R can be designed using any of the linear control techniques presented in Chapter 4. We then partition F in conformity with x_1 and x_2:

$$F = [\, F_1 \quad F_2 \,]. \tag{5.199}$$

As defined in Equations (5.186) and (5.186), we let

$$x_e = (I - A - BF)^{-1}B \cdot [C_2(I - A - BF)^{-1}B]^{-1} r, \tag{5.200}$$

and

$$H = [1 + F(I - A - BF)^{-1}B] \cdot [C_2(I - A - BF)^{-1}B]^{-1}. \tag{5.201}$$

The reduced order CNF controller is given by

$$\begin{aligned} x_v(k+1) = {} & (A_{22} + K_R A_{12}) x_v(k) + (B_2 + K_R B_1)\,\mathrm{sat}[u(k)] \\ & + \Big[A_{21} + K_R A_{11} - (A_{22} + K_R A_{12}) K_R\Big] y(k) \end{aligned} \tag{5.202}$$

and

$$\begin{aligned} u(k) = {} & F\left[\begin{pmatrix} y \\ x_v(k) - K_R y(k) \end{pmatrix} - x_e\right] + Hr \\ & + \rho(r, y) B' P (A + BF) \left[\begin{pmatrix} y(k) \\ x_v - K_R y(k) \end{pmatrix} - x_e\right], \end{aligned} \tag{5.203}$$

where $\rho(r, y)$ is a nonpositive scalar function locally Lipschitz in y subject to certain constraints to be discussed later.

Next, given a positive definite matrix $W_P \in \mathbb{R}^{n \times n}$, let $P > 0$ be the solution to the Lyapunov equation

$$P = (A + BF)' P (A + BF) + W_P. \tag{5.204}$$

Given another positive define matrix $W_R \in \mathbb{R}^{(n-p)\times(n-p)}$ with

$$W_{\mathrm{R}} > F_2' B' P(A+BF) W_{\mathrm{P}}^{-1} (A+BF)' P B F_2, \tag{5.205}$$

let $Q_{\mathrm{R}} > 0$ be the solution to the Lyapunov equation

$$Q_{\mathrm{R}} = (A_{22} + K_{\mathrm{R}} A_{12})' Q_{\mathrm{R}} (A_{22} + K_{\mathrm{R}} A_{12}) + W_{\mathrm{R}}. \tag{5.206}$$

Note that such P and Q_{R} exist as $A+BF$ and $A_{22}+K_{\mathrm{R}}A_{12}$ are asymptotically stable. For any $\delta \in (0,1)$, let c_δ be the largest positive scalar such that for all

$$\begin{pmatrix} x(k) \\ x_{\mathrm{v}}(k) \end{pmatrix} \in \boldsymbol{X}_{\mathrm{R}\delta} := \left\{ \begin{pmatrix} x \\ x_{\mathrm{v}} \end{pmatrix} : \begin{pmatrix} x \\ x_{\mathrm{v}} \end{pmatrix}' \begin{bmatrix} P & 0 \\ 0 & Q_{\mathrm{R}} \end{bmatrix} \begin{pmatrix} x \\ x_{\mathrm{v}} \end{pmatrix} \le c_\delta \right\}, \tag{5.207}$$

we have

$$\left| \begin{bmatrix} F & F_2 \end{bmatrix} \begin{pmatrix} x(k) \\ x_{\mathrm{v}}(k) \end{pmatrix} \right| \le u_{\max}(1-\delta). \tag{5.208}$$

We have the following theorem.

Theorem 5.5.6. Consider the given system in Equation (5.65). Then, there exists a scalar $0 < \rho^* \le 2(B'PB)^{-1}$ such that for any nonpositive function $\rho(r,y)$, locally Lipschitz in y and $|\rho(r,y)| \le \rho^*$, the reduced order CNF law given by Equations (5.202) and (5.203) will drive the system controlled output $h(k)$ to track asymptotically the step command input of amplitude r from an initial state x_0, provided that x_0, $x_{\mathrm{v}0}$ and r satisfy

$$\begin{pmatrix} x_0 - x_{\mathrm{e}} \\ x_{\mathrm{v}0} - x_{20} - K_{\mathrm{R}} x_{10} \end{pmatrix} \in \boldsymbol{X}_{\mathrm{R}\delta} \quad \text{and} \quad |Hr| \le \delta \cdot u_{\max}. \tag{5.209}$$

Proof. Again, the proof of this theorem is similar to those given earlier. ◇

5.5.3 Tuning of Nonlinear Feedback Gains

The freedom to choose the function $\rho(r,y)$, in both the continuous-time and discrete-time system CNF techniques, is used to tune the control laws so as to improve the performance of the closed-loop system as the controlled output h approaches the set point. Since the main purpose of adding the nonlinear part to the CNF controllers is to speed up the settling time, or equivalently to contribute a significant value to the control input when the tracking error, $r-h$, is small. The nonlinear part, in general, will be in action when the

control signal is far away from its saturation level, and thus it will cause the control input to hit its limits. Under such a circumstance, it is straightforward to verify that the closed-loop system comprising the given continuous-time plant in Equation (5.65) and the three different types of control law can be expressed as

$$\dot{\tilde{x}} = (A + BF)\tilde{x} + \rho BB'P\tilde{x}, \tag{5.210}$$

and the closed-loop system comprising the given discrete-time plant in Equation (5.153) and their CNF control laws can be expressed as

$$\tilde{x}(k+1) = (A + BF)\tilde{x}(k) + \rho BB'P(A + BF)\tilde{x}(k). \tag{5.211}$$

We note that the additional term $\rho(r, y)$ does not affect the stability of the estimators, and in the final stage the difference between the true state and the estimated state is almost zero. In fact, the estimation error will always be zero if the initial state of the given plant is precisely known. This turns out to be the case in the track seeking stage of HDD servo systems. It is now clear that eigenvalues of the closed-loop systems in Equations (5.210) and (5.211) can be changed by changing the function $\rho(r, y)$ and, of course, when $\rho(r, y) = 0$, the closed-loop systems are purely linear. In general, we choose $\rho(r, y)$ such that the damping ratio of the closed-loop system changes as $\rho(r, y)$ changes without affecting the stability of the closed-loop systems. It is easy to determine the function $\rho(r, y)$ using a root-locus plot of the closed-loop systems respectively on the s or the z plane.

Since, in general, the function ρ is more closely related to tracking error, *i.e.*, $r - h$, in most practical situations the tracking error is known and available for feedback. As such, we propose the following exponential function for ρ in terms of tracking error:

$$\rho(r, h) = -1.58198\alpha \left(\mathrm{e}^{-|1-h/r|} - 0.36788 \right), \tag{5.212}$$

where $\alpha \geq 0$ is a tuning parameter. This function $\rho(r, h)$ changes from 0 to $-\alpha$ as the tracking error approaches zero. At the initial stage, when the controlled output h is far away from the final set point, $|1 - h/r|$ closes to 1, which implies that $\rho(r, h)$ is small and the effect of the nonlinear part on the overall system is very limited. When the controlled output h approaches the set point, $|1 - h/r|$ closes to zero and $\rho(r, h) \approx -\alpha$, and the nonlinear control law will become effective. In general, one can play with the parameter α to yield a desired performance. Finally, we would like to note that in fact the limit for the nonlinear tuning function $\rho(r, y)$, *i.e.*, ρ^*, in both the full order and reduced order measurement feedback cases (either in continuous-time or discrete-time), can be arbitrarily pre-chosen and then one can select an

appropriate W_{Q} (for the full order case) or W_{R} (for the reduced order case) to guarantee the stability of the corresponding closed-loop system. We leave this as an exercise for interested readers.

5.6 Can We Beat Time Optimal Control?

So far, we have presented quite a number of control techniques that can be used to design control laws to track certain target references for systems with actuator saturations. The TOC technique is believed to be non-robust to system uncertainties and noise, and thus cannot be used in tackling real problems. Unfortunately, it has also been regarded as a method that would, at least theoretically, yield the best performance in terms of settling time.

Can we design a control system that would beat the performance of the TOC? Obviously, the answer to this question is no if it is required to have a precise point-to-point tracking, *i.e.*, to track a target reference precisely from a given initial point. However, surprisingly, the answer would be yes if we consider an asymptotic tracking situation, *i.e.*, if we consider the settling time to be the total time that the controlled system output takes to get from its initial position to reach a predetermined neighbourhood of the target reference. The reason that we are interested in this issue is that asymptotic tracking is widely used in almost all practical situations.

In what follows, we will show the above observation in an example. Let us consider a system characterised a double integrator, *i.e.*,

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \mathrm{sat}(u), \quad y = x, \quad h = [\,1 \quad 0\,]\, x, \tag{5.213}$$

where as usual x is the state, u is the input, and y and h are respectively the measurement and controlled outputs. Moreover, we assume that

$$\mathrm{sat}(u) = \mathrm{sgn}(u) \cdot \min\{\, 1,\, |u| \,\}. \tag{5.214}$$

Let the initial state $x(0) = 0$ and the target reference $r = 1$. Then, it follows from Equation (5.22) that the minimum time required for the controlled output to reach precisely the target reference under TOC control is exactly 2 s. Let us now consider an asymptotic tracking situation instead. As is commonly accepted in the literature (see *e.g.*, [65]), we define the settling time to be the total time that it takes for the control output h to enter the $\pm 1\%$ region of the target reference. The following control law, obtained from a variation form of the CNF control technique, would give a faster settling time than that of the TOC,

$$u = [-6.5 \quad -1]\,x + 6.5\,r - \left(e^{-|1-h|} - 0.36788\right)[1.4481 \quad 10.8609]\left(x - \begin{bmatrix}1\\0\end{bmatrix}\right). \quad (5.215)$$

Figures 5.7 and 5.8 respectively show the resulting controlled output responses and the control signals of the TOC and the modified CNF control. The resulting output response of the modified CNF control has an overshoot less than 1%. However, if we zoom in on the output responses (see Figure 5.9), we will see that the modified CNF control clearly has a faster settling time than that of the TOC when it enters the target region, *i.e.*, $0.99 \leq h \leq 1.01$. It can be computed that the modified CNF control has a settling time of 1.8453 s whereas the TOC has a settling time of 1.8586 s. Although the difference is not much, since we have not tried to optimise the solution of the modified CNF control, it is, however, significant enough to address one interesting issue: *there are control laws that can achieve a faster settling time than that of the TOC in asymptotic tracking situations.* It can also be shown that, no matter how small the target region is, say $1 \pm \varepsilon$ for any small $\varepsilon > 0$, we can always find a suitable control law that beats the TOC in settling time. Nonetheless, we believe that it would be interesting to carry out some further studies in this subject.

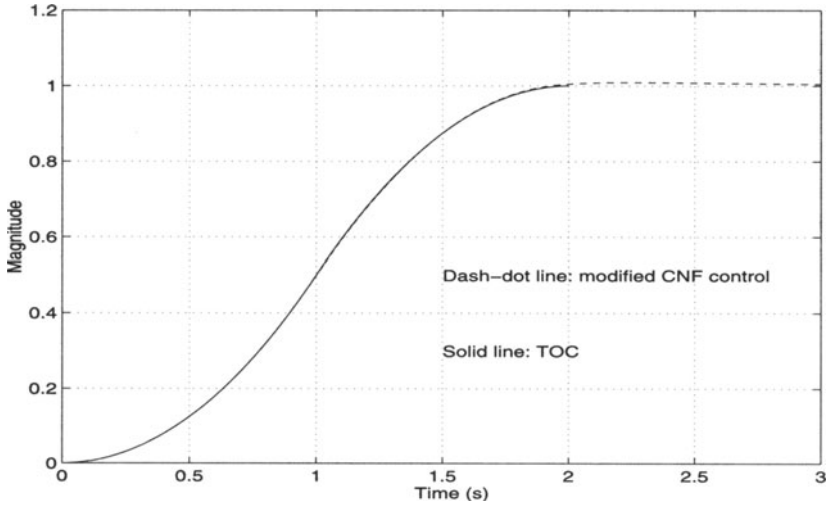


Figure 5.7. Controlled output responses of TOC and modified CNF control.

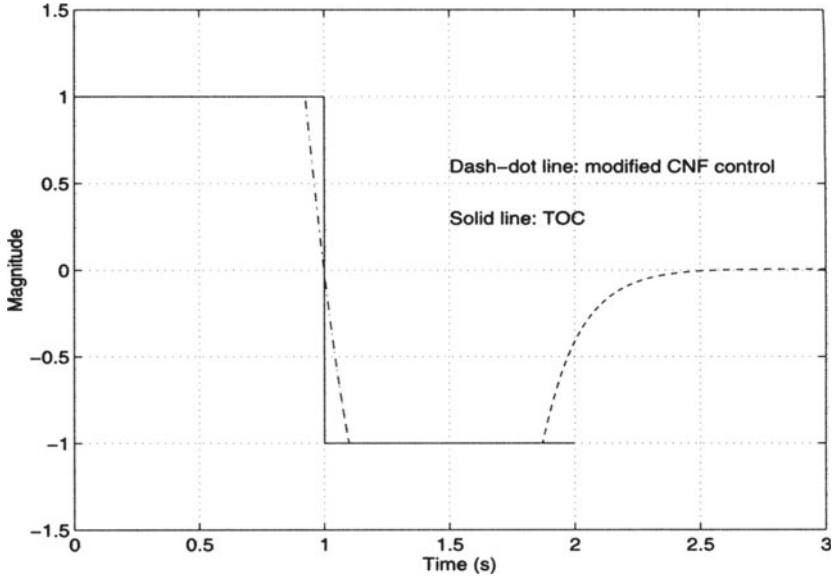


Figure 5.8. Control signals of TOC and modified CNF control.

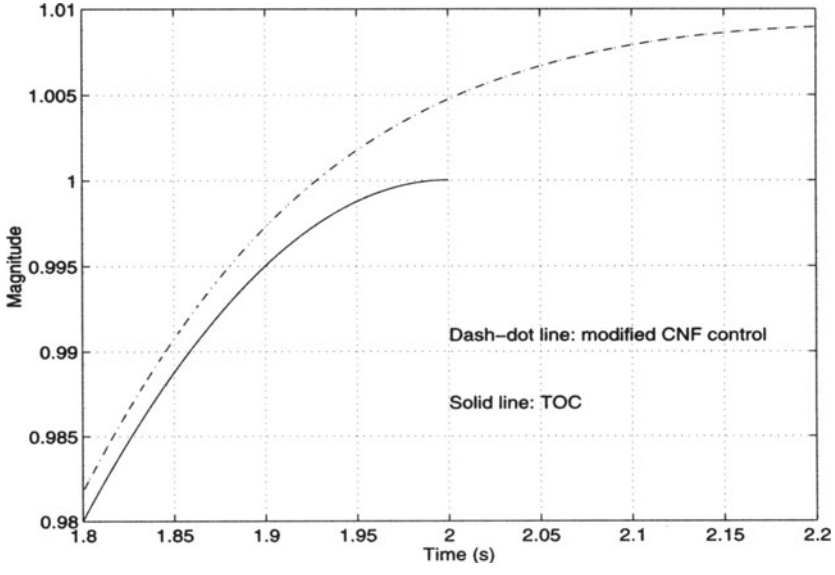


Figure 5.9. Controlled output responses around the target reference.

CHAPTER 6
TRACK FOLLOWING OF A SINGLE-STAGE ACTUATOR

6.1 Introduction

The prevalent trend in hard disk design is towards smaller hard disks with increasingly larger capacities. This implies that the track width has to be smaller, leading to lower error tolerance in the positioning of the head. The controller for track following has to achieve tighter regulation in the control of the servomechanism. Current HDDs use a combination of classical control techniques, such as lead-lag compensators, PID compensators, and notch filters. These classical methods can no longer meet the demand for HDDs of higher performance. Thus, many control approaches have been tried, such as LQG and/or LTR approach (see *e.g.*, [18, 19]), and adaptive control (see *e.g.*, [27]) and so on.

The purpose of this chapter is to use the result of the RPT control method of Chapter 4 to carry out a design of a track following controller for an HDD with a single VCM actuator. We will first obtain a model of the VCM actuator and then cast the overall track following control system design into an RPT design framework. A first-order dynamic measurement feedback controller is then designed to achieve robust and perfect tracking for any step reference. Our controller is theoretically capable of making the L_p-norm of the resulting tracking error with $1 \leq p < \infty$ arbitrarily small in faces of external disturbances and initial conditions. Some trade-offs are then made in order for the RPT controller to be implementable using the existing hardware setup and to meet physical constraints such as sampling rates and the limit of control of the system. The implementation results of the RPT controller are compared with those of a PID controller. The results show that our design is simpler and yet has faster settling times, lower overshoot and higher accuracy. As will be seen shortly, the control system with the RPT controller is highly robust with respect to the variances of resonant frequencies and runout disturbances. It also yields a better position-error-signal (PES) test compared with that of the PID control. The results presented in this chapter are rooted in earlier work reported in [3] (see also [44]). However, all the results in this

chapter are obtained using a new HDD, namely a Maxtor (Model 51536U3) HDD.

6.2 VCM Actuator Model

In this section, we present the modelling of the VCM actuator, which is well known in the research community of the HDD servo systems to have a characteristic of a double integrator cast with some high-frequency resonance, which can reduce the system stability if neglected. There are some bias forces in the HDD system that will cause steady-state errors in tracking performance. Moreover, there are also some nonlinearities in the system at low frequencies, which are primarily due to the pivot and bearing frictions. All these factors should be taken into consideration when considering the design of a controller for the VCM actuator. For the purpose of developing a model, we have to compromise between accuracy and simplicity. In this section, a relatively simplified model of the VCM actuator is identified and presented.

The dynamics of an ideal VCM actuator can be formulated as a second-order state space model as follows:

$$\begin{pmatrix} \dot{y} \\ \dot{v} \end{pmatrix} = \begin{bmatrix} 0 & k_{\mathrm{y}} \\ 0 & 0 \end{bmatrix} \begin{pmatrix} y \\ v \end{pmatrix} + \begin{pmatrix} 0 \\ k_{\mathrm{v}} \end{pmatrix} u, \tag{6.1}$$

where u is the actuator input (in volts), y and v are the position (in tracks) and the velocity of the R/W head, k_{y} is the position measurement gain and $k_{\mathrm{v}} = k_{\mathrm{t}}/m$, with k_{t} being the current–force conversion coefficient and m being the mass of the VCM actuator. Thus, the transfer function of an ideal VCM actuator model appears to be a double integrator, *i.e.*,

$$G_{\mathrm{v1}}(s) = \frac{k_{\mathrm{v}} k_{\mathrm{y}}}{s^2}. \tag{6.2}$$

However, if we also consider the high-frequency resonance modes, a more realistic model for the VCM actuator should be

$$G_{\mathrm{v}}(s) = \frac{k_{\mathrm{v}} k_{\mathrm{y}}}{s^2} \frac{\omega_{\mathrm{n}}^2}{s^2 + 2\xi\omega_{\mathrm{n}} s + \omega_{\mathrm{n}}^2}. \tag{6.3}$$

The frequency characteristics of the Maxtor (Model 51536U3) HDD have been obtained using an LDV and an HP Dynamic Signal Analyser. The actual frequency response is shown in Figure 6.1. Applying the least squares estimation identification method given in Chapter 2 (see also [13, 37]) to the

measured data from the actual system, we obtain a fourth-order model for the actuator:

$$G_{\mathrm{v}}(s) = \frac{6.4013 \times 10^7}{s^2} \frac{2.467 \times 10^8}{s^2 + 2.513 \times 10^3 s + 2.467 \times 10^8}. \tag{6.4}$$

Figure 6.1 shows that the frequency response of the identified model matches the measured data very well for the frequency range from 0 to 1.8 kHz, which far exceeds the working range of the VCM actuator.

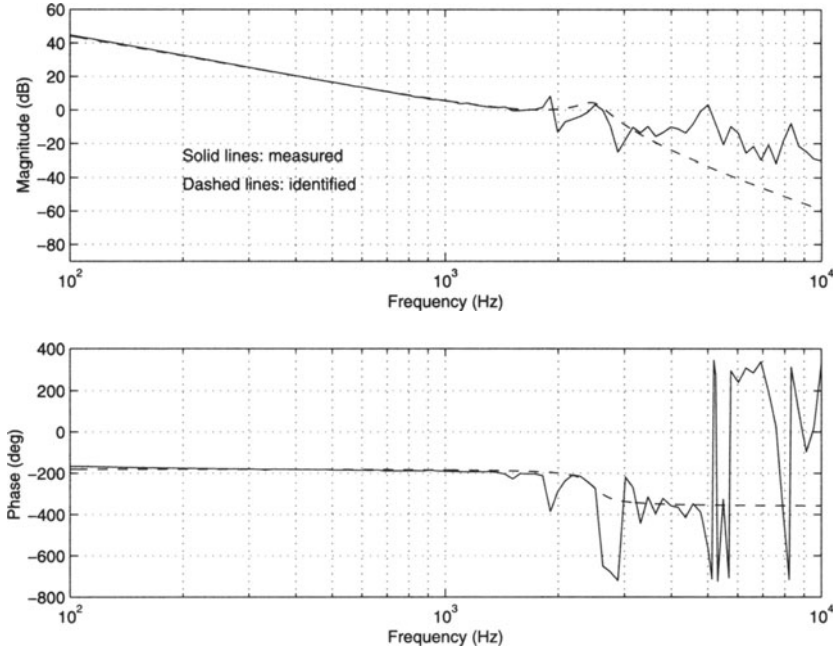


Figure 6.1. Frequency responses of the actual and identified VCM actuator models.

6.3 Track Following Controller Design

We now present the control system design for the actuator identified in the previous section. Basically, the majority of commercially available HDD servo systems todate are designed using a conventional PID approach. For drives with a single VCM actuator, designers would encounter problems if they wished to push up the track following speed. Usually, there will be some huge peak overshoot in step response. Thus, in practice, one would have to

make trade-offs between the track following speed and overshoot by selecting appropriate PID controller gains. We formulate our control system design as an RPT control problem. Such an approach will enable the designer to design a very low-order control law, and, moreover, the resulting closed-loop system will have fast track following speed and low overshoot as well as strong robustness.

We will design a control system that meets the following design specifications:

1. the control input should not exceed ± 3 V owing to physical constraints on the actual VCM actuator;
2. the overshoot and undershoot of the step response should be kept less than 5% as the R/W head can start to read or write within $\pm 5\%$ of the target;
3. the 5% settling time in the step response should be as fast as possible;
4. sampling frequency in implementing the actual controller is 10 kHz.

From the experience that we gained in designing PID controllers, we know that it is quite safe to ignore the resonance models of the VCM actuator if we are focusing on tracking performance. Thus, we will consider only a second-order model for the VCM actuator at this stage. We will then put the resonance modes back when we are to evaluate the performance of the overall design. Thus, in our design, we will first use the following simplified model of the VCM actuator:

$$\dot{x} = Ax + Bu = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ 6.4013 \times 10^7 \end{bmatrix} u, \tag{6.5}$$

and

$$y = C_1 x = [\,1 \quad 0\,]\, x. \tag{6.6}$$

Next, we define the output to be controlled as

$$h = y = C_2 x + D_2 u = [\,1 \quad 0\,]\, x. \tag{6.7}$$

Consider the reference $r(t)$ to be a step function with a magnitude α, *i.e.*, $r(t) = \alpha \cdot 1(t)$, where $1(t)$ is a unit step function. Then, we have

$$\dot{r}(t) = \alpha \cdot \delta(t), \tag{6.8}$$

where $\delta(t)$ is a unit impulse function. Following the results of Chapter 4, we obtain a corresponding auxiliary system:

$$\left\{\begin{array}{l} \dot{\boldsymbol{x}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} 0 \\ 0 \\ 6.4013 \times 10^7 \end{bmatrix} u + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \boldsymbol{w} \\ \boldsymbol{y} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} \boldsymbol{w} \\ e = [-1 \quad 1 \quad 0]\, \boldsymbol{x} + 0 \quad u \end{array}\right. \tag{6.9}$$

where

$$\boldsymbol{x} = \begin{pmatrix} r \\ x \end{pmatrix}, \quad \boldsymbol{w} = \alpha \cdot \delta(t), \quad \boldsymbol{y} = \begin{pmatrix} r \\ y \end{pmatrix}, \quad e = h - r. \tag{6.10}$$

It is simple to see that (A, B, C_2, D_2) is invertible and free of invariant zeros, and $\mathrm{Ker}\,(C_1) = \mathrm{Ker}\,(C_2)$. Hence, it follows from the result of Chapter 4 that the RPT performance is achievable. Following the results of Chapter 4, one can show that there exists a family of measurement feedback control laws, parameterised by a tuning parameter ε, such that when it is applied to the given VCM actuator:

1. the resulting closed-loop system is asymptotically stable for sufficiently small ε; and
2. for any given initial condition x_0 and any $p \in [1, \infty)$, the l_p-norm of the resulting tracking error e has the property $\|e\|_p \to 0$, as $\varepsilon \to 0$.

Following the construction algorithm for the reduced order RPT controller in Chapter 4, we obtained a parameterised first-order measurement feedback control law of the form

$$\dot{v} = A_{\mathrm{RC}}(\varepsilon) v + B_{\mathrm{RC}}(\varepsilon) \begin{pmatrix} r \\ y \end{pmatrix}, \quad u = C_{\mathrm{RC}}(\varepsilon) v + D_{\mathrm{RC}}(\varepsilon) \begin{pmatrix} r \\ y \end{pmatrix}, \tag{6.11}$$

with

$$\left.\begin{array}{l} A_{\mathrm{RC}}(\varepsilon) = -7345.8/\varepsilon \\ B_{\mathrm{RC}}(\varepsilon) = \dfrac{1}{\varepsilon^2} [\, 4.975 \times 10^6 \quad -3.4358 \times 10^7 \,] \\ C_{\mathrm{RC}}(\varepsilon) = -5.2267 \times 10^{-5}/\varepsilon \\ D_{\mathrm{RC}}(\varepsilon) = \dfrac{1}{\varepsilon^2} [\, 0.0777 \quad -0.2868 \,]. \end{array}\right\} \tag{6.12}$$

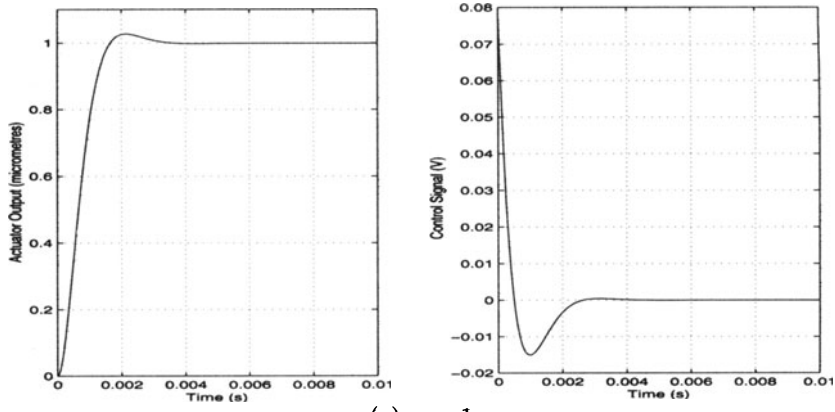
The results in Figure 6.2 were obtained using the MATLAB package. They clearly show that the RPT problem is solved as we tune the tuning parameter ε to be smaller and smaller. Unfortunately, owing to the constraints of the physical system, *i.e.*, the limits in control inputs and sampling rates, as well as resonance modes, it is impossible to implement a controller that will track the reference with zero time. We would thus have to make some compromises in the track following speed because of these limitations. After several trials, we found that the controller parameters of Equation (6.12) with $\varepsilon = 1$ would give us a satisfactory performance. We then discretise it using a bilinear transformation with a sampling frequency of 10 kHz. The discretised controller obtained using the bilinear transformation technique (see Chapter 3) is given by

$$\begin{cases} x_{\mathrm{v}}(k+1) = & 0.46275 & x_{\mathrm{v}}(k) + 363.88\, r(k) - 2512.9\, y(k) \\ u(k) & = -3.8227 \times 10^{-5}\, x_{\mathrm{v}}(k) + 0.0682\, r(k) - 0.2211\, y(k) \end{cases} \tag{6.13}$$

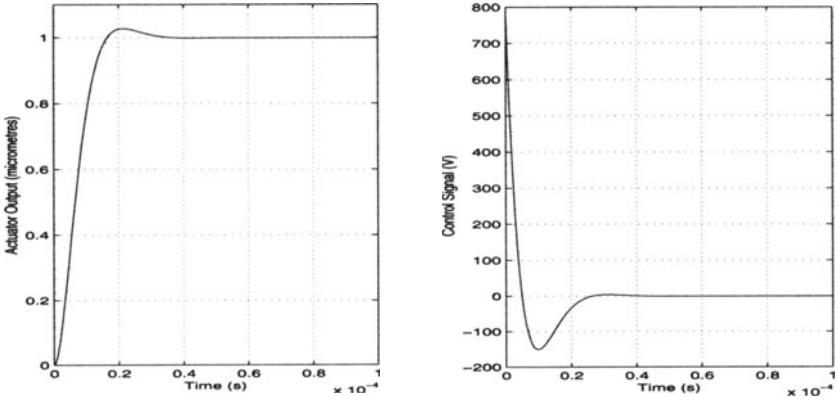
Figure 6.3 shows that the step response of the overall system, comprising the fourth-order model of the VCM actuator (we now put the resonance modes back into the VCM actuator model) and the discretised RPT controller, meets the design specifications. In actual HDD manufacturing, the resonant frequency ω_{n} of the VCM actuator, see Equation (6.4), for the same batch of drives might vary from one to another. A common practice in the disk drive industry is to add some notch filters in the servo system to attenuate these resonant peaks as much as possible. Surprisingly, our RPT controller is capable of withstanding the variation of resonance frequencies as well. Figure 6.4 shows the step responses of the closed-loop systems of our RPT controller and the VCM actuator model with two different resonant frequencies: one is $\beta = 75\%$ of the nominal value, and the other is $\beta = 150\%$ of the nominal resonant frequency. The results show that the RPT controller is very robust with respect to the change of resonant frequency in the actuator.

Although we do not consider the effects of runout disturbances in our problem formulation, it turns out that our simple first-order controller is capable of rejecting the first few modes of the runout disturbances, which are mainly due to the imperfectness of the data tracks and the spindle motor speeds, and commonly have frequencies at the multiples of about 55 Hz. We simulate these runout effects by injecting a sinusoidal signal into the measurement output, *i.e.*, the new measurement output is the sum of the actuator output and the runout disturbance. Figure 6.5 shows the simulation result of the output response of the overall control system comprising the fourth-order model of the VCM actuator model and the discretised RPT controller with a fictitious runout disturbance injection

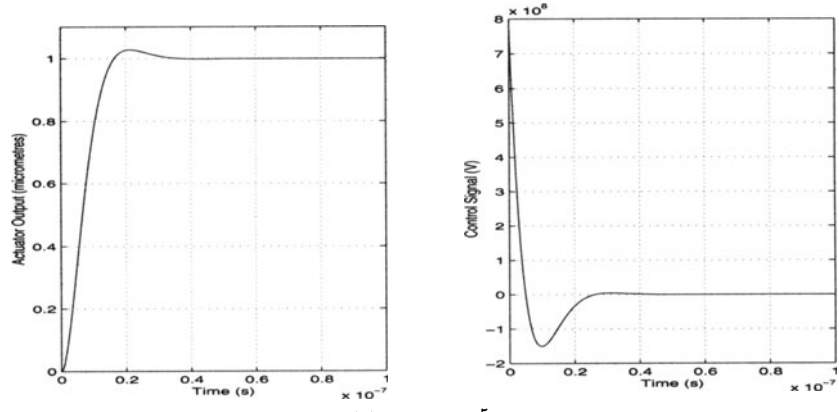
$$\tilde{w}(t) = 0.5 + 0.1\cos(110\pi t) + 0.05\sin(220\pi t), \tag{6.14}$$



(a). $\varepsilon = 1$



(b). $\varepsilon = 0.01$



(c). $\varepsilon = 10^{-5}$

Figure 6.2. Responses of the closed-loop systems with RPT controller.

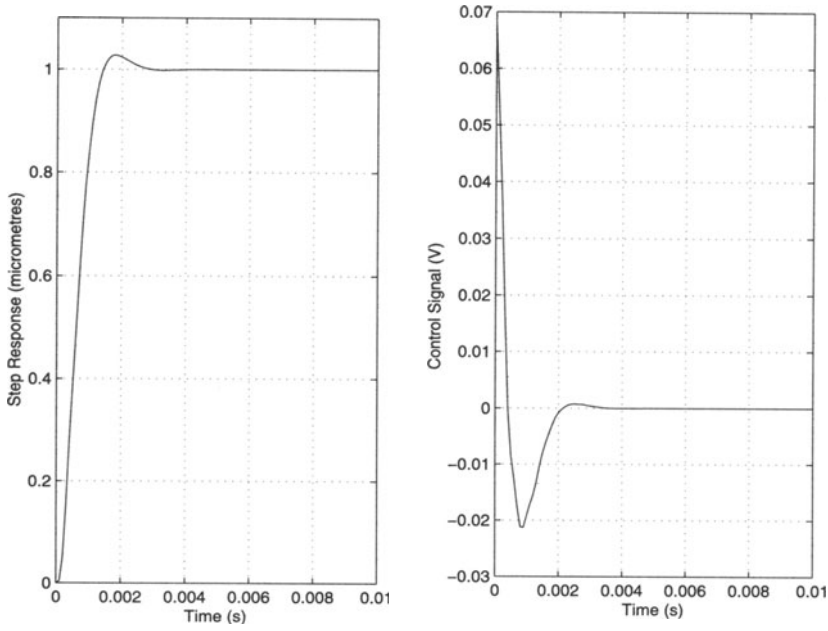


Figure 6.3. Closed-loop step response with discretised RPT controller.

and a zero reference $r(t)$. The result shows that the effects of such a disturbance on the overall response are minimal. A more comprehensive test on runout disturbances, *i.e.*, the position error signal (PES) test on the actual system will be presented in the next section.

6.4 Implementation Results

In this section we present the actual implementation results of our design and their comparison with those of a PID controller. Two major tests are presented: one is the track following of the closed-loop systems and the other is the PES test, which is considered to be a major factor in the design of HDD servo systems. Our controller was implemented on an open HDD with a sampling rate of 10 kHz. Closed-loop actuation tests were performed using an LDV to measure the R/W head position. The resolution used for LDV was 2μ m/V. This displacement output is then fed into the DSP, which would then generate the necessary control signal to the VCM actuator. A DSA was used to assist in obtaining the frequency response of the overall control system. It can inject a swept sinusoidal reference signal, then read the output displacement from the LDV and calculate the frequency Bode plot using this

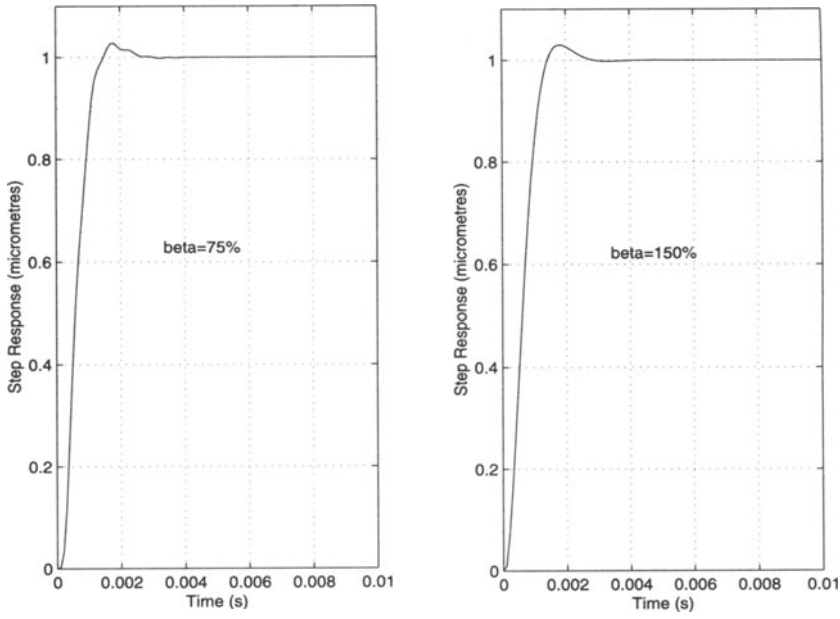


Figure 6.4. Closed-loop step responses with different resonant frequencies.

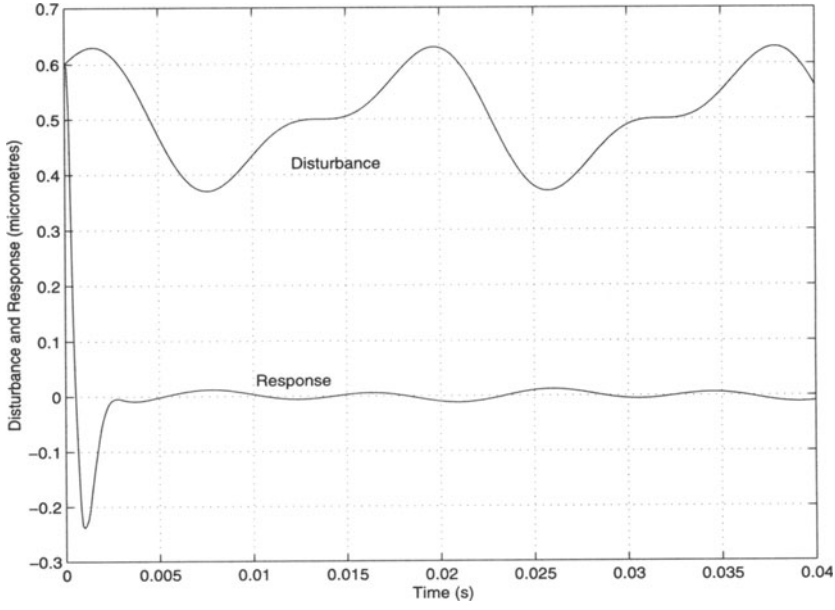


Figure 6.5. Closed-loop output response due to a runout disturbance.

information. Altogether, two sets of experiment were performed, one using the RPT controller and the other using a finely tuned PID controller.

6.4.1 Track Following Test

The solid-line curve in Figure 6.6 shows the experimental step response of the RPT controller. In this figure, the response of the RPT controller is shown together with that of a PID controller as a comparison. The 5% settling time is about 1.35 ms, which surely meets the design specifications.

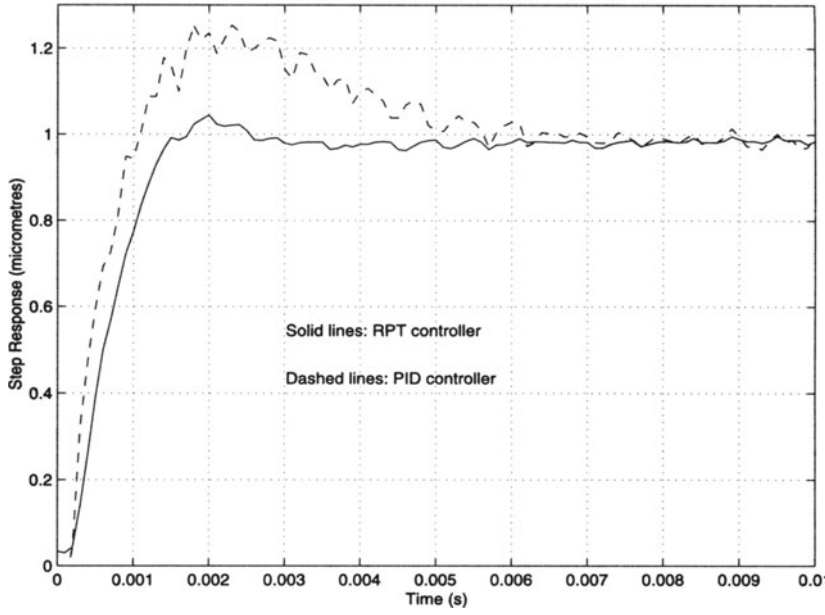
The dotted-line curve in Figure 6.6 shows the step response of the PID controller (again using a 10 kHz sampling rate). The PID controller had the usual structure and was finely tuned such that it could have a fast time response. It is given by

$$u = \frac{0.23z - 0.2156}{z - 0.091}(r - y). \tag{6.15}$$

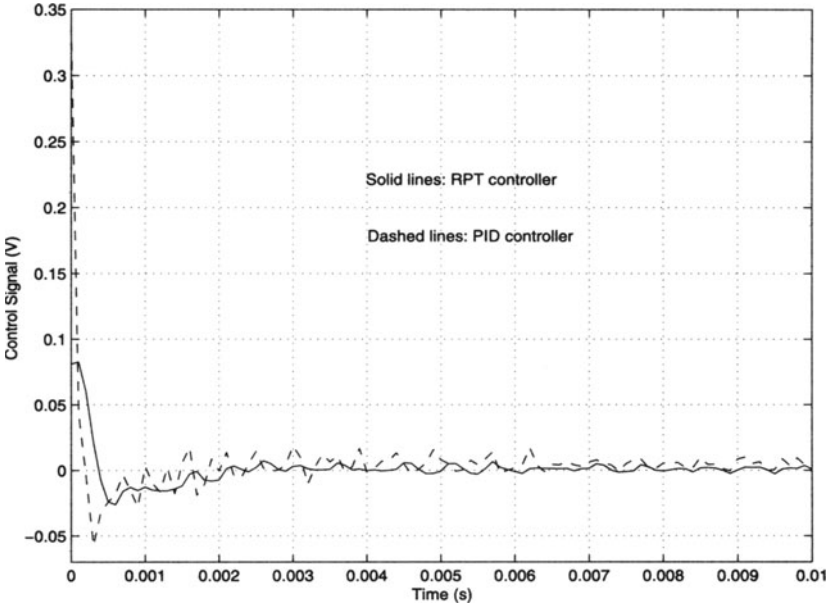
Unfortunately, the overshoot of the controller is rather high, about 25%, and this is the result of trading improved settling time at the expense of higher overshoot. To achieve a settling time of 4–5 ms, it is necessary to tune the PID controller such that the overshoot is significant. The 5% settling time for the PID control is about 4.8 ms, which is significantly larger than that of the RPT control. Nonetheless, this PID control law yields a much better performance compared with that reported in [3].

We believe that the shortcoming of the PID control is mainly due to its structure, *i.e.*, it only feeds in the error signal, $y - r$, instead of feeding in both y and r independently. We trust that the same problem might be present in other control methods if the only signal fed is $y - r$. The PID control structure might well be as simple as most researchers and engineers have claimed. However, our RPT controller is even simpler, *i.e.*, the RPT controller is of the first order and the PID controller is of the second order, but we have fully utilised all available information associated with the actual system.

Unfortunately, we could not compare our results with those of other methods mentioned in the chapter introduction. Most of the references we found in the open literature contained only simulation results in this regard. Some of implementation results we found were, however, very different in nature. For example, Hanselmann and Engelke [18] reported an implementation result of a disk drive control system design using the LQG approach with a sampling frequency of 34 kHz. The overall step response in [18] with a higher-order LQG controller and higher sampling frequency is worse than that of ours.



(a). output responses



(b). control signals

Figure 6.6. Implementation result: step responses with RPT and PID control.

6.4.2 Position Error Signal Test

The disturbances in a real HDD are usually considered as a lumped disturbance at the plant output, also known as runouts. Repeatable runouts (RROs) and nonrepeatable runouts (NRROs) are the major sources of track following errors. RROs are caused by the rotation of the spindle motor and consists of frequencies that are multiples of the spindle frequency. NRROs can be perceived as coming from three main sources: vibration shocks, mechanical disturbance and electrical noise. Static force due to flex cable bias, pivot-bearing friction and windage are all components of the vibration shock disturbance. Mechanical disturbances include spindle motor variations, disk flutter and slider vibrations. Electrical noises include quantisation errors, media noise, servo demodulator noise and power amplifier noise. NRROs are usually random and unpredictable by nature, unlike repeatable runouts. They are also of a lower magnitude (see *e.g.*, [1]). A perfect servo system for HDDs should reject both the RROs and NRROs.

In our experiment, we have simplified the system somewhat by removing many sources of disturbances, especially that of the spinning magnetic disk. Therefore, we actually have to add the runouts and other disturbances into the system manually. Based on previous experiments, we know that the runouts in real disk drives are composed mainly of RROs, which are basically sinusoidal with a frequency of about 55 Hz, equivalent to the spin rate of the spindle motor. By manually adding this "noise" to the output while keeping the reference signal at zero, we can then read off the subsequent position signal as the expected PES in the presence of runouts. In disk drive applications, the variations in the position of the R/W head from the centre of the track during track following, which can be directly read off as the PES, is very important. Track following servo systems have to ensure that the PES is kept to a minimum. Having deviations that are above the tolerance of the disk drive would result in too many read or write errors, making the disk drive unusable. A suitable measure is the standard deviation of the readings, σ_{pes}. A useful guideline is to make the $3\sigma_{\text{pes}}$ value less than 10% of the track pitch, which is about 0.1 μm for a track density of 25 kTPI.

Figures 6.7 and 6.8 show the tracking errors of the RPT controller and PID controller respectively, under the disturbance of the runouts. The $3\sigma_{\text{pes}}$ value is about 0.029 μm for the RPT controller, and about 0.063 μm for the PID controller. Again, the RPT controller does better than the PID one in the PES test.

In conclusion, the RPT controller has a much better performance in track following and in the PES tests compared with that of the PID controller. The RPT controller utilised is first order. This is one order lower in comparison with the PID controller and would allow for quicker execution of the

DSP codes during implementation. This would be an important consideration when the sampling rate of the disk drive servo is pushed higher to meet the increasing demands on the servo performance. The current results can be further improved if we used a better VCM actuator and arm assembly, with a higher resonance frequency. The control input limit has not been reached, and, theoretically, we should be able to tune the controller to achieve even faster settling time, and higher servo bandwidths.

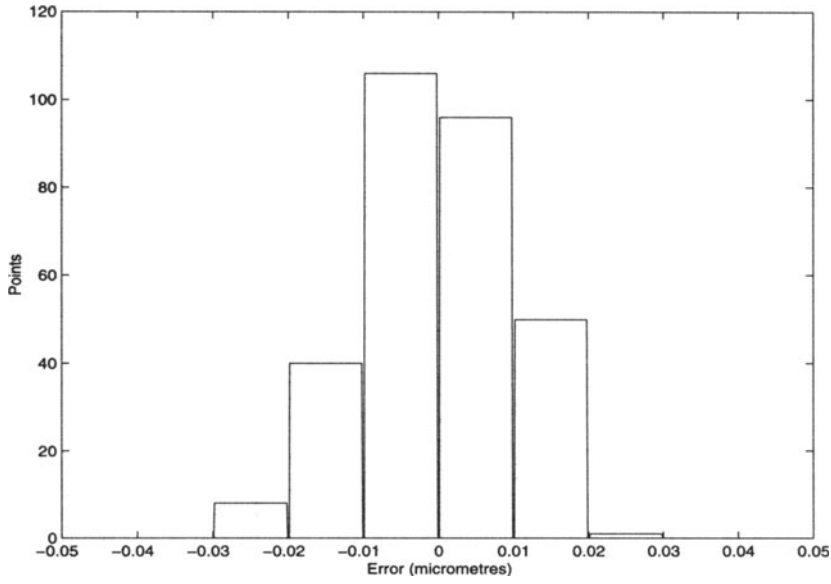


Figure 6.7. Implementation result: histogram of the PES test (RPT).

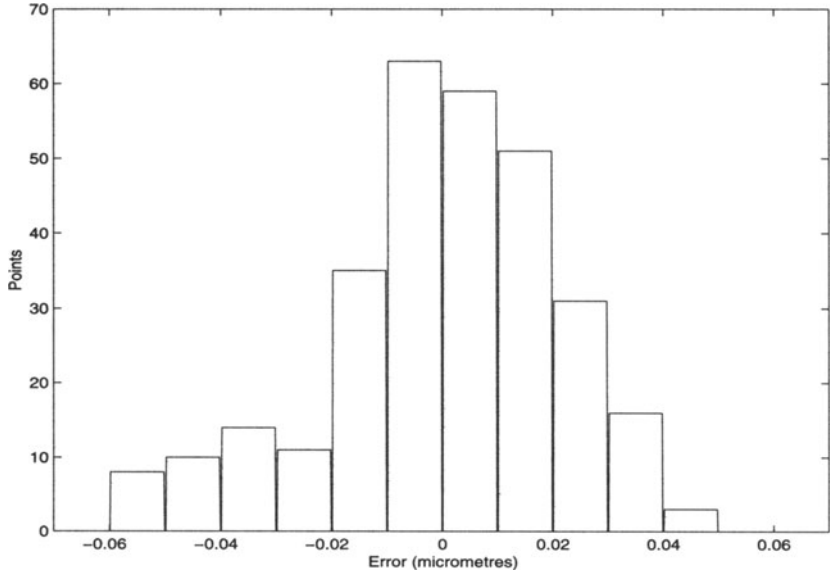


Figure 6.8. Implementation result: histogram of the PES test (PID).

CHAPTER 7
SINGLE-STAGE ACTUATED SERVO SYSTEMS

7.1 Introduction

In this chapter, we proceed to address one of the key issues of this book – the design of complete single-stage HDD servo systems that meet the two most important requirements for HDDs, *i.e.*, a high-speed seeking performance, and a highly accurate head-positioning/settling performance. We will utilise the nonlinear control techniques reported in Chapter 5 and the linear techniques reported in Chapter 4 to carry out the design of three different types of servo system for a Maxtor HDD with a single VCM actuator. More specifically, we will design the servo systems using the conventional PTOS approach, the CNF control technique, and the MSC system with PTOS and RPT controllers. As will be seen shortly, the CNF technique will yield the best performance compared with those of the other two approaches. This is due to the fact that there is no switching element involved in the CNF design.

As in the previous chapter, a Maxtor (Model 51536U3) HDD will be used to implement our design. The actual frequency response and the identified model are shown once more in Figure 7.1 for the sake of easy reference. The frequency domain model has been identified earlier in Chapter 6 and is given by

$$G_{\mathrm{v}}(s) = \frac{6.4013 \times 10^7}{s^2} \frac{2.467 \times 10^8}{s^2 + 2.513 \times 10^3 s + 2.467 \times 10^8}. \tag{7.1}$$

As mentioned earlier, it is a common practice in the design of HDD servo systems to approximate the VCM actuator model as a double integrator with an appropriate DC gain. Such an approximation simplifies the overall design procedure a great deal. Most importantly, it works very well, since, in general, we do not want to excite the resonant modes in the closed-loop system. However, in order to make our design more realistic, all our simulation results will be done using the fourth-order model. The final implementation is, of course, to be carried out on the actual system.

The following state-space model will be used throughout our design:

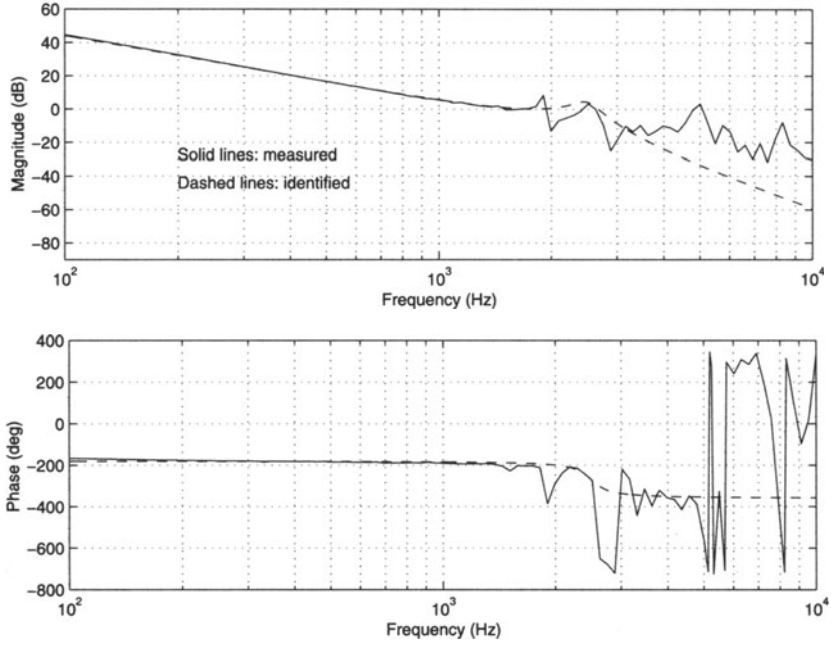


Figure 7.1. Frequency responses of the actual and identified VCM actuator models.

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ 6.4013 \times 10^7 \end{bmatrix} u, \quad x = \begin{pmatrix} y \\ v \end{pmatrix}, \tag{7.2}$$

where y and v are respectively the position of the VCM actuator head in micrometres and velocity in micrometres per second, and u is the control input in volts. In general, the velocity of the VCM actuator in the actual system is not available, and thus y will be the only measurable state. For this particular system, the controlled output is also the measurement output, *i.e.*,

$$h = y = [1 \quad 0]\, x. \tag{7.3}$$

Our objective is to design a servo controller that meets the following physical constraints and design specifications:

1. the control input should not exceed ± 3 V owing to physical constraints on the actual VCM actuator;

2. the overshoot and undershoot of the step response should be kept less than 0.05 μm, 5% of one track pitch. We note that the R/W head can start writing data on to the disk when it is within 5% of one track pitch of the target.

As mentioned earlier, three different approaches, namely the PTOS method, the MSC method and the CNF method, will be presented in the following to design appropriate servo systems for the given HDD. We will carry out control system design for each method first. All simulation and implementation results, as well as their comparison will be discussed in the last section.

7.2 Servo System with PTOS Control

We present in this section the design and implementation of an HDD servo system using the PTOS approach (see Chapter 5). The first step is to find the state feedback gains k_1 and k_2 in the PTOS control law based on the design specifications. To get specifications in terms of required closed-loop poles we need natural frequency ω_n and the damping ratio ζ. Let us choose the natural frequency to be 1067 rad/s, *i.e.*, 170 Hz, and the damping factor to be 0.898 so as to have an acceleration discount factor α of 0.62, which will yield a reasonably good performance for seek lengths up to 300 μm. It follows from Equation (5.32) that a PTOS control law with such a discount factor will only increase the total tracking time by about 13.5% from that required in the TOC. Interested readers are referred to [29, 106, 107] for detailed information on the selection of these parameters. Note that the relation between the damping ratio ζ and the acceleration discount factor in PTOS control law is given by (see [29])

$$\alpha = \frac{1}{2\zeta^2}. \tag{7.4}$$

Then, the corresponding s-plane closed-loop poles are

$$s_{1,2} = -959.19 \pm j469.98. \tag{7.5}$$

Using the m-function `acker` in MATLAB, we obtain the following feedback gains

$$k_1 = 0.0178 \quad \text{and} \quad k_2 = 2.997 \times 10^{-5}, \tag{7.6}$$

and the length of the linear region in PTOS can be found from Equation (5.31) and is given by $y_\ell = 168.32$ μm. Thus, the PTOS control law for our disk drive is as follows:

$$u = u_{\max} \cdot \text{sat}\left(\frac{k_2\,[f_p(e) - v]}{u_{\max}}\right), \tag{7.7}$$

where

$$f_{\mathrm{p}}(e) = \begin{cases} \dfrac{k_1}{k_2} e, & \text{for } |e| \leq y_\ell, \\ \mathrm{sgn}(e) \left[\sqrt{2u_{\max} a\alpha \mid e \mid} - \dfrac{u_{\max}}{k_2} \right], & \text{for } |e| > y_\ell, \end{cases} \tag{7.8}$$

with

$$a = 6.4013 \times 10^7, \quad k_1 = 0.0178, \quad k_2 = 2.997 \times 10^{-5}, \tag{7.9}$$

and

$$y_\ell = 168.32, \quad \alpha = 0.62. \tag{7.10}$$

The advantage of this control scheme is that it is quite simple to understand. The implementation of such a controller requires an estimation of the VCM actuator velocity (with the estimator pole being placed at -4000). The simulation and implementation results of the above design will be given in Section 7.5.

7.3 Servo System with MSC

In this section, we will apply the MSC method of Chapter 4 to our commercial disk drive model. The MSC scheme uses the proximate time-optimal controller in the track seeking mode, and the RPT controller in the track following mode. We note that in MSC, initially, the plant is controlled by the seeking controller and at the end of seeking mode a switch changes it to a track following controller. In [102], the mode switching was done after finding the optimal mode switching conditions such that the impact of the initial values on settling performance was minimised. But the impact of the resulting control signal on the resonance modes was not considered. It has been shown [44, 84] that the RPT controller is independent of these initial values. The optimal mode switching conditions in our scheme can just be set such that the control signal is small enough so as not to excite the resonance vibrations. The problem of unmodelled mechanical resonance can be treated more rigorously by using either minimising the jerk as defined by $\|\mathrm{d}u/\mathrm{d}t\|_2$ as reported in [108] or by using a method developed in the frequency domain in [109]. However, by utilising the features of RPT control, such as it works for a wide range of resonance frequencies (see Chapter 6), the mode switching conditions can be determined in a very simple way (see Chapter 5).

We now move to present an MSC controller for the HDD with a single VCM actuator. The control law in track seeking mode (here we label its control signal as u_{P}) is given by Equation (7.7), as this mode uses the PTOS control. The control law in the track following mode, *i.e.*, the reduced-order measurement feedback RPT control law, is given by

$$\dot{v} = A_{\text{RC}} v + B_{\text{RC}} \begin{pmatrix} r \\ y \end{pmatrix}, \quad u_{\text{R}} = C_{\text{RC}} v + D_{\text{RC}} \begin{pmatrix} r \\ y \end{pmatrix}, \tag{7.11}$$

with

$$\left.\begin{aligned} A_{\text{RC}} &= -8410.8 \\ B_{\text{RC}} &= [\,7.9952 \times 10^{6} \quad -4.1367 \times 10^{7}\,] \\ C_{\text{RC}} &= -6.8905 \times 10^{-5} \\ D_{\text{RC}} &= [\,0.1249 \quad -0.4005\,]. \end{aligned}\right\} \tag{7.12}$$

Next we find the mode switching conditions as defined in Equation (5.62). Using the RPT controller parameters, and following the results of Chapter 5, the mode switching conditions can be determined as $|e(t_1)| \leq 2\ \mu\text{m} < y_\ell = 168.32\ \mu\text{m}$ and $|v(t_1)| \leq 2703\ \mu\text{m/s}$. We select the MSC law

$$u = \begin{cases} u_{\text{P}}, & t < t_1, \\ u_{\text{R}}, & t \geq t_1, \end{cases} \tag{7.13}$$

in which t_1 is chosen such that

$$|\, y(t_1) - r \,| = 2\ \mu\text{m} \quad \text{and} \quad |\, v(t_1) \,| \leq 2703\ \mu\text{m/s}. \tag{7.14}$$

The overall closed-loop system comprising the given VCM actuated HDD and the MSC control law will be asymptotically stable. For easy comparison, the simulation and implementation of the overall system with the MSC control law will again be presented in Section 7.5.

7.4 Servo System with CNF Control

We now move to the design of a reduced order continuous-time composite nonlinear control law as given by Equations (5.143) and (5.144) for the commercial hard disk model shown in Figure 7.1. A similar result using the discrete-time CNF method has also been reported [110]. The design can be

achieved by following the procedure given in Chapter 5. Practically, the design procedure is not as complex as explained in Chapter 5. The simplified procedure is as follows.

1. Determine a state feedback gain matrix F using any appropriate method (see *e.g.*, linear control techniques in Chapter 4) to have a stable and a quick rising time with a control input that is below saturation.
2. Compute the matrix G_e and the steady-state x_e to determine the feed forward gain G_r (see CNF control in Chapter 5).
3. Select an appropriate $W_P > 0$ and solve the following Lyapunov equation:

$$(A + BF)'P + P(A + BF) = -W_P. \tag{7.15}$$

4. Select a nonlinear tuning function $\rho(r, h)$, where r and h are the step command input and the system controlled output respectively.
5. Determine an appropriate value of α such that the eigenvalues of matrix $A + B(F - \alpha B'P)$ have a large damping ratio and hence the output response will have a small overshoot.

As the control law depends on the size of the step command input, we derive, for our HDD model given by Equation (7.2), the following parameterised state feedback gain $F(\varepsilon)$:

$$F(\varepsilon) = -\frac{1}{6.4013 \times 10^7}\left[\frac{4\pi^2 f^2}{\varepsilon^2} \quad \frac{4\pi f\zeta}{\varepsilon}\right]. \tag{7.16}$$

The eigenvalues of $A + BF(\varepsilon)$, are placed at $(-\zeta \pm j\sqrt{1-\zeta^2})2\pi f/\varepsilon$. As in Chapter 5, we select the nonlinear tuning function as follows,

$$\rho(r, h) = -1.58198\alpha\left(\mathrm{e}^{-|1-h/r|} - 0.36788\right). \tag{7.17}$$

Following the design procedure given in Chapter 5 and the physical properties of the given system, we choose a low damping ratio of $\zeta = 0.3$ and $f = 350$, which corresponds roughly to the normal working frequency range of the linear part of the CNF control law with $\varepsilon = 1$.

Here we note that ε is a tuning parameter, which can be tuned with respect to the amplitude of the command reference r, *i.e.*, it can be tuned so that the given command reference can be tracked by the CNF controller. α is another tuning parameter, which is to be adjusted to yield a smooth response when the controlled output is approaching the target reference. Both tuning

parameters can be tuned with respect to the amplitude of the target reference r. Based on simulation, we come out with a set of best tuned values for ε and α. The results are given in Table 7.1. Using the least squares linear fitting technique, we obtain

$$\varepsilon = \begin{cases} 0.0594r + 1.0805, & 1 \leq r < 20\ \mu\text{m}, \\ 0.0019r + 2.2062, & 20 \leq r \leq 300\ \mu\text{m}, \end{cases} \tag{7.18}$$

and

$$\alpha = \begin{cases} 0.0515r + 1.0756, & 1 \leq r < 20\ \mu\text{m}, \\ 0.0001r + 0.8259, & 20 \leq r \leq 300\ \mu\text{m}. \end{cases} \tag{7.19}$$

Table 7.1. Tuning parameters ε and α versus track seek lengths.

r (μm)	ε	α
1	1.0000	1.00
2	1.1864	1.28
5	1.4894	1.40
10	1.7949	1.55
20	2.1875	1.55
50	2.3333	1.59
100	2.3333	1.59
150	2.5926	1.59
200	2.6923	1.59
250	2.5926	1.59
300	2.7559	1.59

Based on these functions of the tuning parameters ε and α, and following the procedure given in Chapter 5 with $K_{\text{R}} = -4000$, we obtain a reduced order CNF control law as follows:

$$\dot{x}_{\text{v}} = -4000x_{\text{v}} - 1.6 \times 10^7 y + 6.4013 \times 10^7 \,\text{sat}\,(u), \tag{7.20}$$

and

$$\begin{aligned} u = {} & \kappa_2 x_{\text{v}} + (\kappa_1 + 4000\kappa_2)y - \kappa_1 r \\ & + \rho(r,h)\Big[\kappa_3 x_{\text{v}} + (4000\kappa_3 - \kappa_1)y + \kappa_1 r\Big], \end{aligned} \tag{7.21}$$

where $\rho(r,h)$ is as given in Equation (7.17),

$$\kappa_1 = -\frac{0.0755}{\varepsilon^2}, \quad \kappa_2 = -\frac{2.0613 \times 10^{-5}}{\varepsilon}, \quad \kappa_3 = \frac{5.7257 \times 10^{-5}}{\varepsilon}. \tag{7.22}$$

Again, the simulation and implementation results of the servo system with the CNF control law will be presented in the next section for an easy comparison.

7.5 Simulation and Implementation Results

Now, we are ready to present the simulation and implementation results for all three of the servo systems discussed in the previous sections and do a full-scale comparison on the performances of these methods. In particular, we will study the following tests:

1. track seeking and following test;
2. runout disturbance test; and
3. PES test.

All simulation results presented in this section have been obtained using SIMULINK in MATLAB and all implementation results are carried out using our own experimental setup as described in Chapter 1. The sampling frequency for actual implementation is chosen as 10 kHz. Here, we note that all our controllers are discretised using the ZOH technique. As will be seen shortly, the HDD servo system with the CNF control law yields the best results in all the categories.

7.5.1 Track Seeking and Following Test

In our simulation and implementation, we use a track pitch of 1 μm for the HDD. In what follows, we present results for four selected track seek lengths (SL), *i.e.*, SL = $1, 50, 100$ and 300 μm. Unfortunately, owing to the capacity of the LDV that has been used to measure the displacement of the R/W head of the VCM actuator, the absolute errors of our implementation results given below are 0.05, 0.1, 0.2 and 0.5 μm for SL = $1, 50, 100$ and 300 μm, respectively. As such, the settling time for both implementation and simulation results is defined as the total time required for the R/W head to move from its initial position to the entrance of the region of the final target with plus and minus the respective absolute error. This is the best we can do with our current experimental setup. Nonetheless, the results we obtain here should be sufficient enough to illustrate our design ideas and philosophy.

PTOS: Figures 7.2 and 7.3 show the normalised response of the closed-loop system for the servo system with the PTOS control. The results in Figure 7.2 were obtained using SIMULINK, whereas those in Figure 7.3 are experimental results obtained from the actual system. Note that all results were obtained by replacing the velocity by a first-order estimator with its estimator pole being placed at -4000.

MSC: similarly, the simulation results using the MSC control law are shown in Figure 7.4 and the implementation results of the MSC servo system are shown in Figure 7.5.

CNF: lastly, the simulation results for the servo system with the CNF control law are shown in Figure 7.6, and the corresponding implementation results are shown in Figure 7.7.

We now summarise the overall results on settling times and percentages of improvement in Table 7.2.

Table 7.2. Simulation and implementation: settling time and improvement.

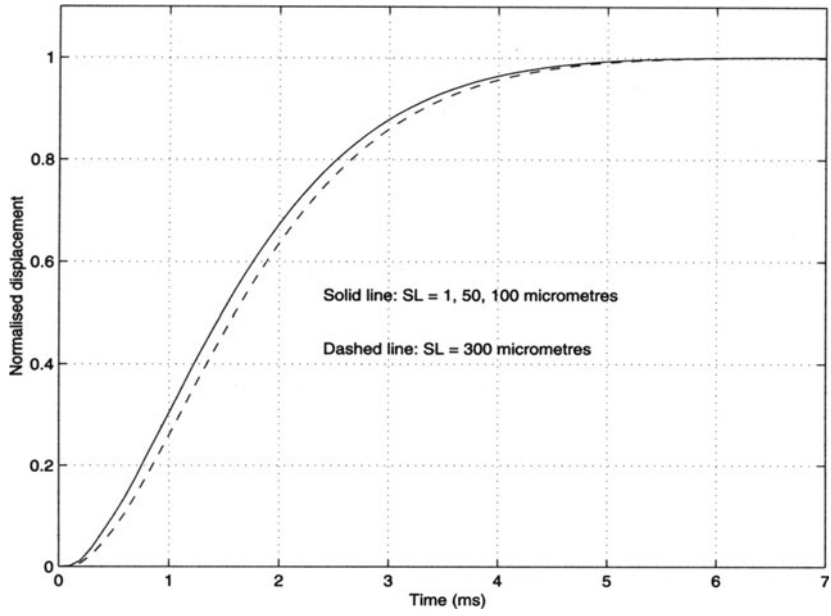
(a) settling times

Seek length (μm)	Settling time (ms) Simulation			Implementation		
	PTOS	MSC	CNF	PTOS	MSC	CNF
1	3.75	1.15	0.96	-	1.25	1.20
50	5.50	5.00	3.70	6.50	6.00	4.60
100	5.50	5.30	3.70	6.50	6.50	4.60
300	5.70	5.70	4.40	6.70	7.20	5.30

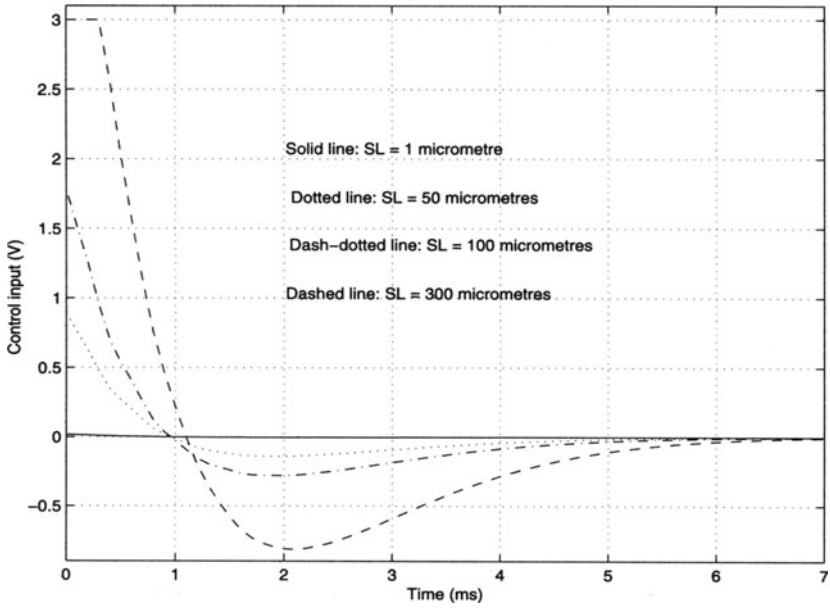
(b) percentage improvement

Seek length (μm)	Improvement vs PTOS (%) Simulation		Implementation	
	MSC	CNF	MSC	CNF
1	69	74	-	-
50	9	33	8	29
100	4	33	0	29
300	0	23	-7	21

Clearly, the simulation and implementation results show that the servo system with the CNF controller has the best performance. We believe that this is due to the fact that the CNF control law unifies the nonlinear and linear components without switching, whereas the other two servo systems involve switching elements between the nonlinear and linear parts, which degrades the overall performance.

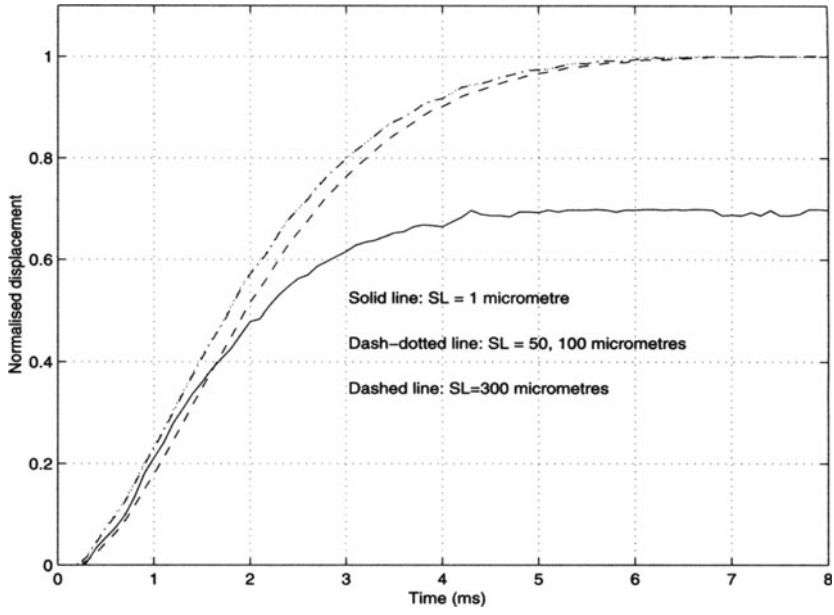


(a). normalised output responses

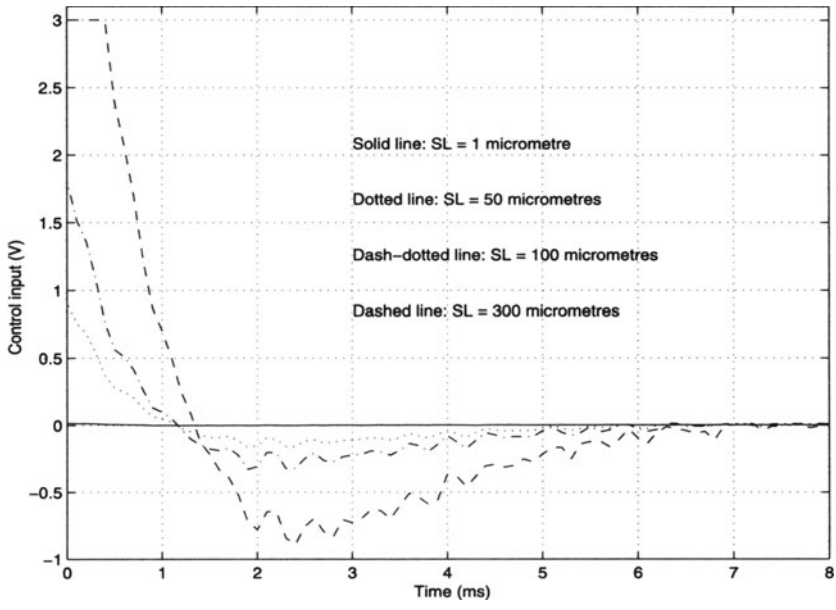


(b). control signals

Figure 7.2. Simulation result: response and control of PTOS servo system.

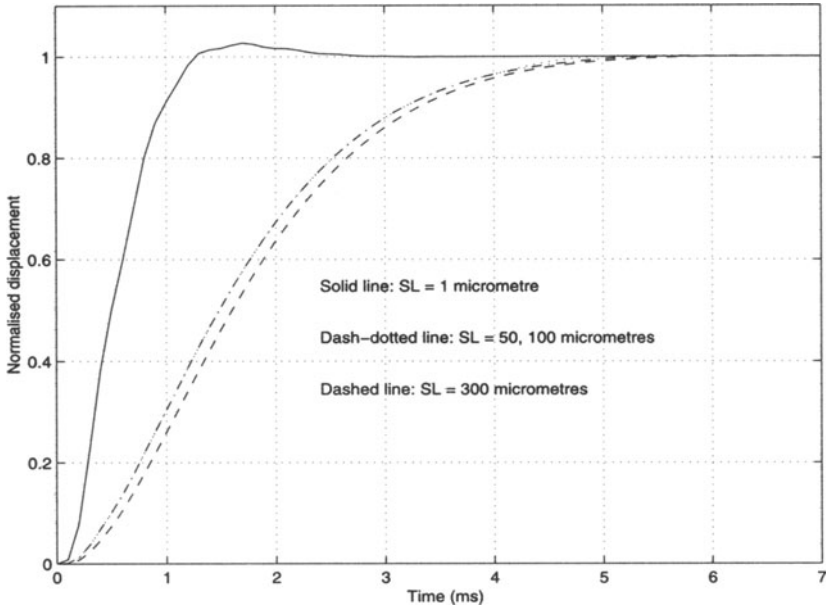


(a). normalised output responses

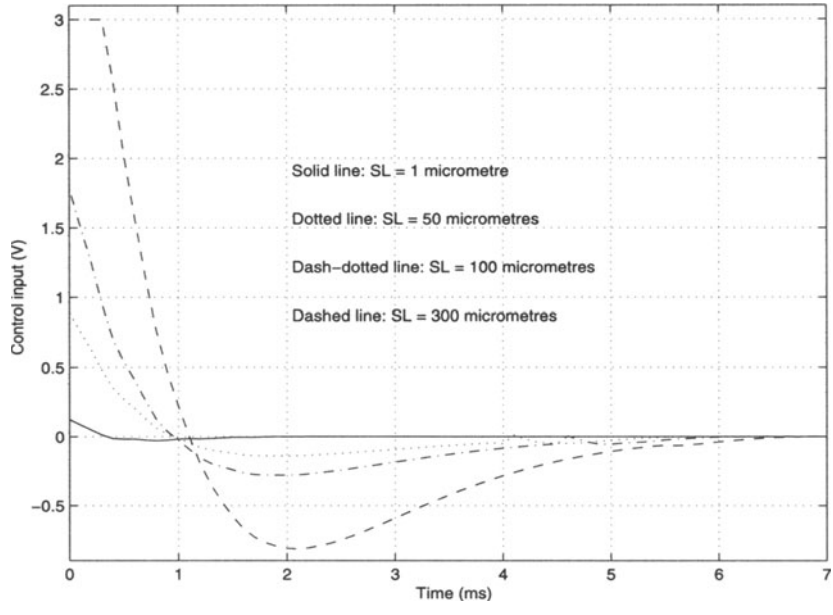


(b). control signals

Figure 7.3. Implementation result: response and control of PTOS servo system.

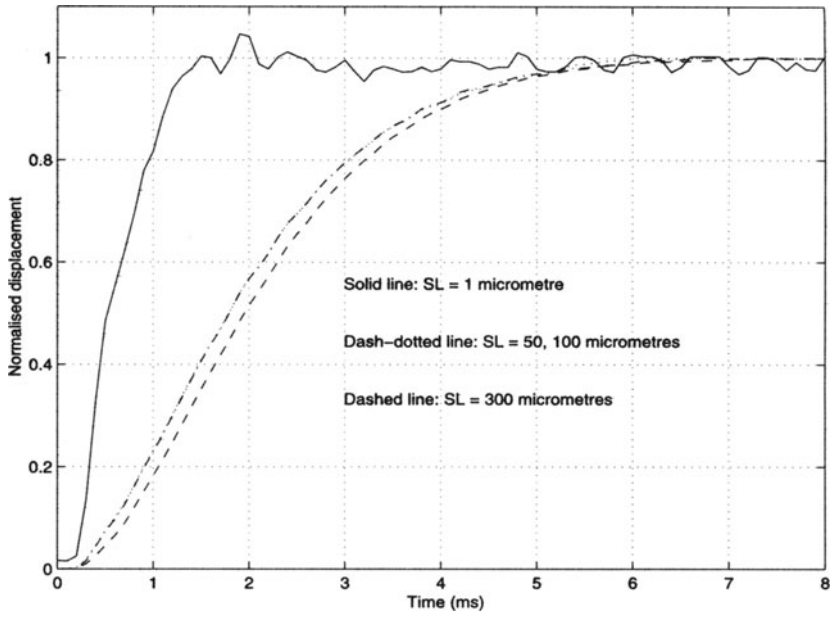


(a). normalised output responses

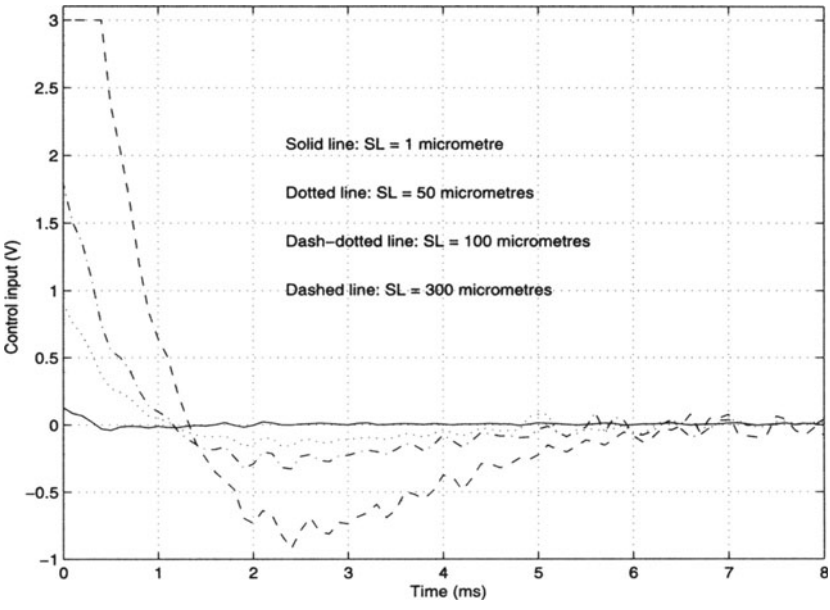


(b). control signals

Figure 7.4. Simulation results: response and control of MSC servo system.

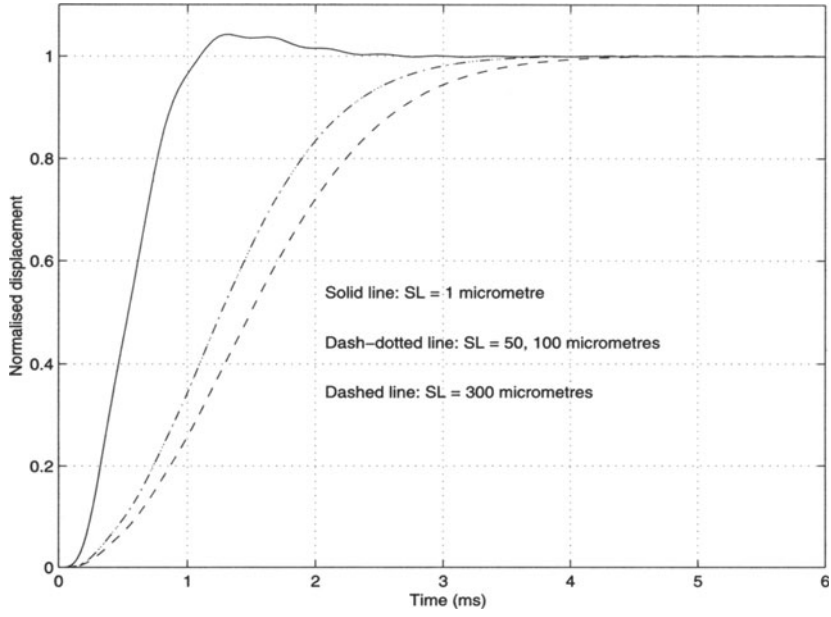


(a). normalised output responses

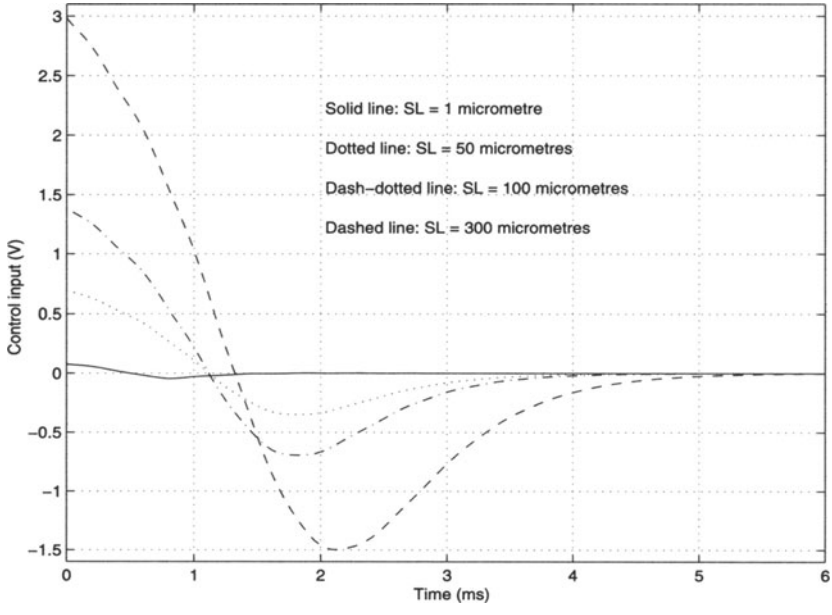


(b). control signals

Figure 7.5. Implementation results: response and control of MSC servo system.

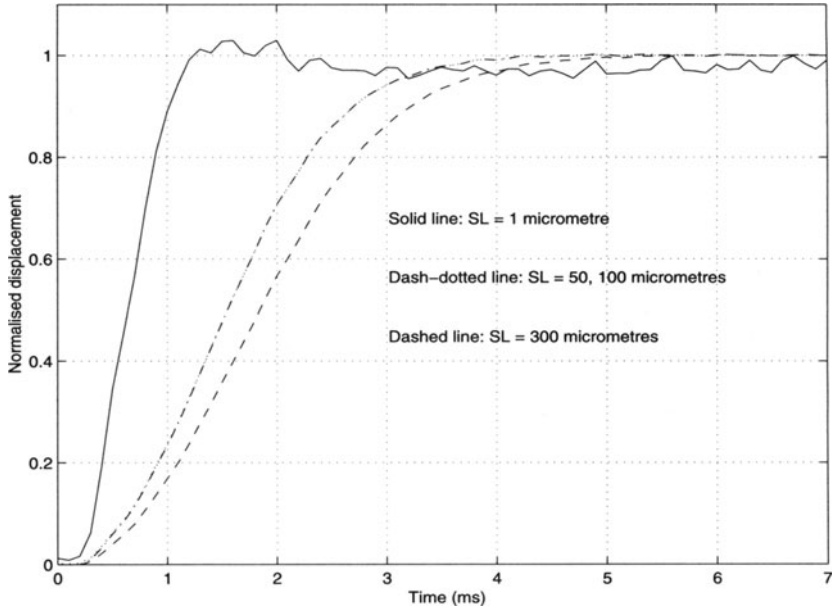


(a). normalised output responses

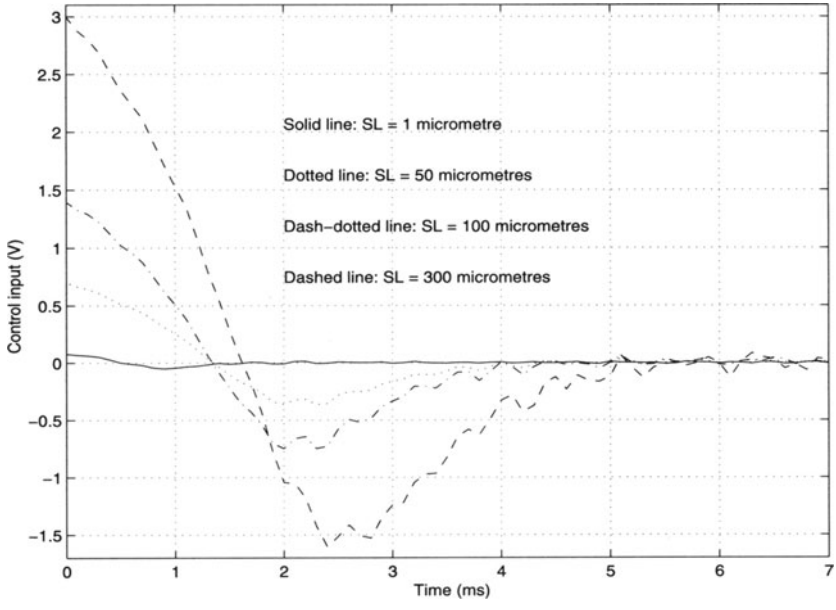


(b). control signals

Figure 7.6. Simulation result: response and control of CNF servo system.



(a). output responses



(b). control signals

Figure 7.7. Implementation result: response and control of CNF servo system.

7.5.2 Runout Disturbance Test

We now study the effects of runout disturbances on our servo systems. As pointed out in Chapter 6, runout disturbances are mainly generated by the imperfectness of the data tracks and the spindle motor speeds. In order to test the robustness of the servo systems against runout disturbances, this time we inject a fictitious runout signal, which is the same as that in Chapter 5, into the actual system (the corresponding results of Chapter 6 were based on simulation). The implementation results for the servo systems with the PTOS, MSC and CNF control laws are respectively given in Figures 7.8, 7.9 and 7.10. The standard deviations of the error signals are respectively 0.0161, 0.0084 and 0.0081. Again, the servo system with the CNF control law yields the best performance in this test.

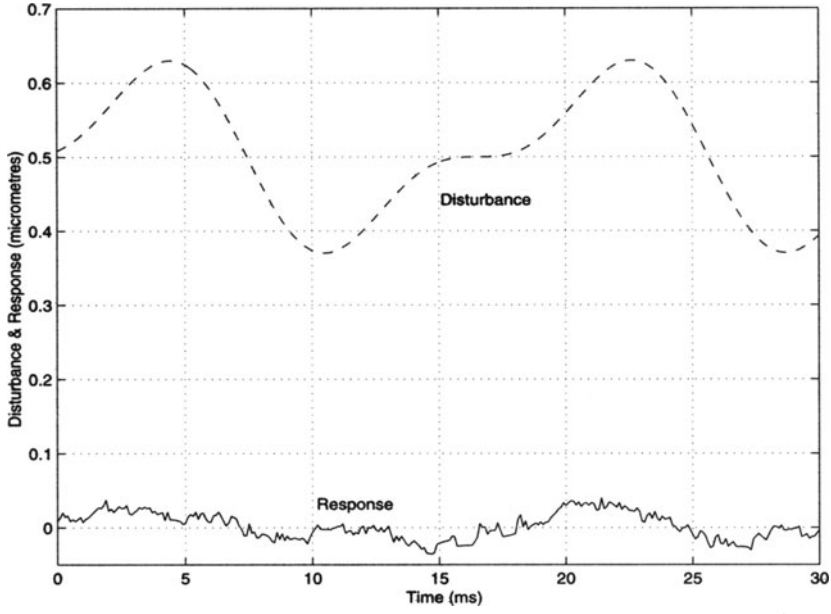


Figure 7.8. Implementation result. Response to a runout disturbance (PTOS).

7.5.3 Position Error Signal Test

As pointed out in Chapter 6, the PES test is important in designing HDD servo systems as its $3\sigma_{\text{pes}}$ value is directly related to the best achievable track density of HDDs. Figures 7.11, 7.12 and 7.13 show the histograms of the PES tests for the servo systems with the PTOS, MSC and CNF control laws

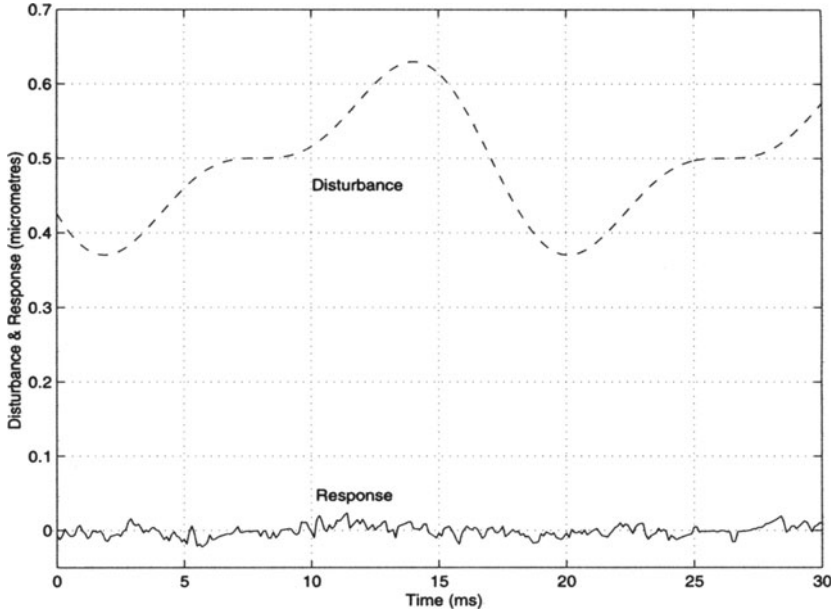


Figure 7.9. Implementation result. Response to a runout disturbance (MSC).

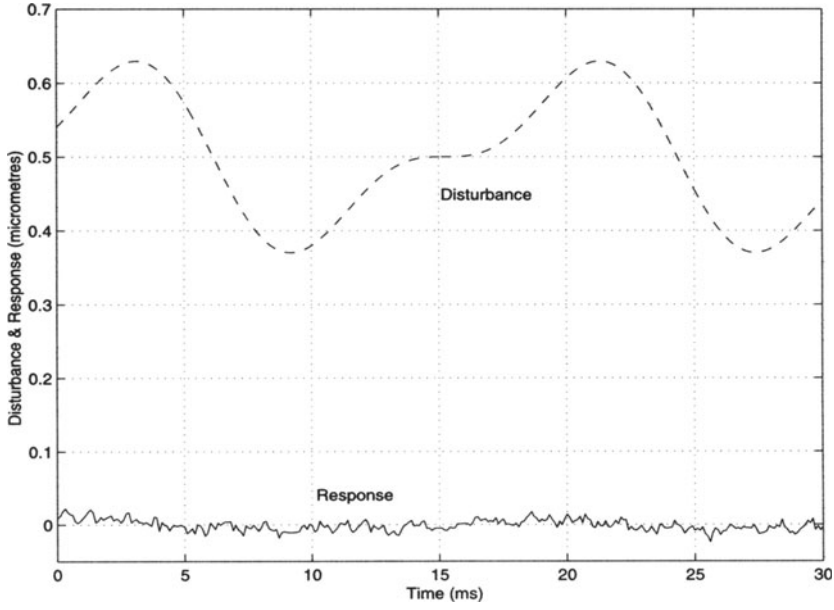


Figure 7.10. Implementation result. Response to a runout disturbance (CNF).

respectively. The $3\sigma_{\text{pes}}$ values of these tests are respectively 0.0483, 0.0252 and 0.0243. The last two servo systems in principle are capable of achieving a track density greater than 80 kTPI.

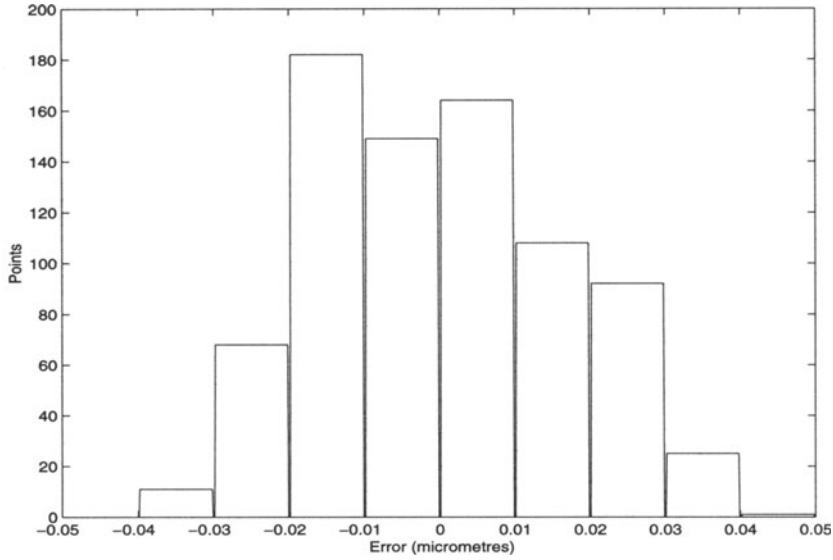


Figure 7.11. Implementation result. Histogram of PES test (PTOS).

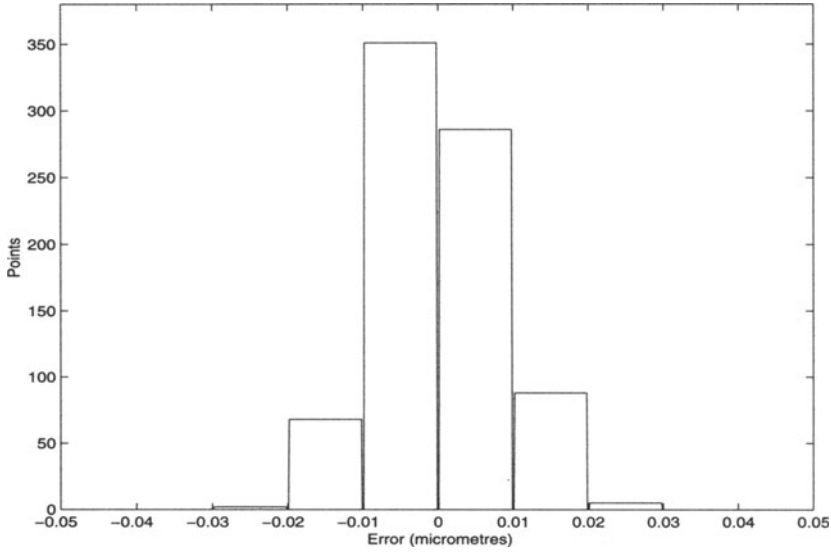


Figure 7.12. Implementation result. Histogram of PES test (MSC).

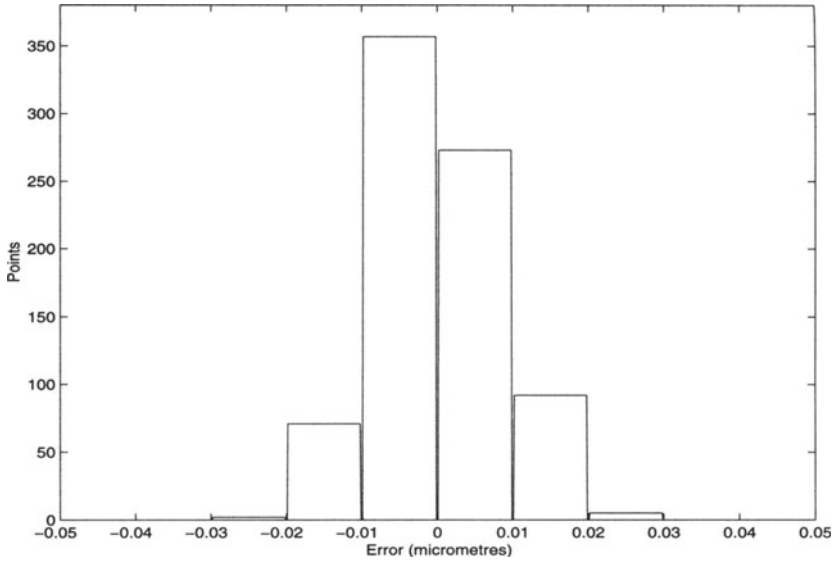


Figure 7.13. Implementation result. Histogram of PES test (CNF).

CHAPTER 8
DESIGN OF A PIEZOELECTRIC ACTUATOR SYSTEM

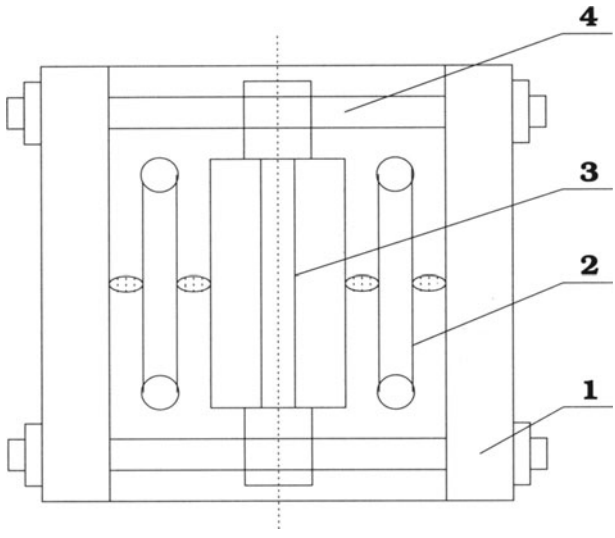
8.1 Introduction

We present in this chapter a case study on a piezoelectric bimorph actuator control system design using an H_∞ optimisation approach, which was originally reported by Chen *et al.* [20]. Piezoelectricity is a fundamental process in electromechanical energy conversion. It relates electric polarisation to mechanical stress/strain in piezoelectric materials. Under the direct piezoelectric effect, an electric charge can be observed when the materials are deformed. The converse, or the reciprocal piezoelectric effect, is when the application of an electric field can cause mechanical stress/strain in the piezo materials. There are numerous piezoelectric materials available today, including PZT (lead zirconate titanate), PLZT (lanthanum modified lead zirconate titanate), and PVDF (piezoelectric polymeric polyvinylidene fluoride) to name a few (see Low and Guo [111]).

Piezoelectric structures are widely used in applications that require electrical to mechanical energy conversion coupled with size limitations, precision, and speed of operation. Typical examples are microsensors, micropositioners, speakers, medical diagnostics, shutters and impact print hammers. In most applications, bimorph or stack piezoelectric structures are used because of the relatively high stress/strain to input electric field ratio (see Low and Guo [111]).

The present work is motivated by the possibility of applying piezoelectric microactuators in magnetic recording. A dual-stage actuated HDD was successfully demonstrated by Tsuchiura *et al.* [112] of Hitachi. In [112], a fine positioner based on a piezoelectric structure was mounted at the end of a primary VCM stage to form the dual actuator. The higher bandwidth of the fine positioner allowed the R/W heads to be positioned accurately. There have been other instances where electromagnetic [113] and electrostatic [114] microactuators have been used for fine positioning of R/W heads. Since then, many research studies have been done along this line. More detailed information about HDD servo systems with a dual-stage actuator will be found in Chapter 9.

The focus of this chapter is to concentrate on the control issues involved in dealing with the nonlinear hysteresis behaviour displayed by most piezoelectric actuators. More specifically, we consider a robust controller design for a piezoelectric bimorph actuator as depicted in Figure 8.1. A scaled-up model of this piezoelectric actuator, which is targeted for use in the secondary stage of a future dual actuator for magnetic recording, was actually built and modelled by Low and Guo [111]. It has two pairs of bimorph beams which are subjected to bipolar excitation. The dynamics of the actuator were identified in [111] as a second-order linear model coupled with a hysteresis. The linear model is given by



1-base; 2-piezoelectric bimorph beams; 3-moving plate; and 4-guides

Figure 8.1. Structure of the piezoelectric bimorph actuator.

$$m\ddot{x}_1 + b\dot{x}_1 + kx_1 = k(du - z), \tag{8.1}$$

where m, b, k and d are the tangent mass, damping, stiffness and effective piezoelectric coefficients; u is the input voltage that generates excitation forces in the actuator system. The variable x_1 is the displacement of the actuator and is also the only measurement we can have in this system. It should be noted that the working range of the displacement of this actuator is within ± 1 μm. The variable z arises from hysteretic nonlinear dynamics [111] and is governed by

$$\dot{z} = \alpha d\dot{u} - \beta|\dot{u}|z - \gamma\dot{u}|z|, \tag{8.2}$$

where α, β and γ are some constants that control the shape of the hysteresis. For the actuator system that we are considering here, the above coefficients are identified as follows:

$$\left.\begin{array}{rcl} m &=& 0.01595 \text{ kg}, \\ b &=& 1.169 \text{ N s/m}, \\ k &=& 4385 \text{ N/m}, \\ d &=& 8.209 \times 10^{-7} \text{ m/V}, \\ \alpha &=& 0.4297, \\ \beta &=& 0.03438, \\ \gamma &=& -0.002865. \end{array}\right\} \tag{8.3}$$

For a more detailed description of this piezoelectric actuator system and the identifications of the above parameters, we refer interested readers to the work of Low and Guo [111]. Our goal in this chapter is to design a robust controller, as in Figure 8.2, that meets the following design specifications.

1. The steady-state tracking errors of the displacement should be less than 1% for any input reference signals that have frequencies ranging from 0 to 30 Hz, as the actuator is to be used to track certain coloured noise types of signal in disk drive systems.

2. The 1% settling time should be as fast as possible (we are able to achieve a 1% settling time of less than 0.003 s in our design).

3. The control input signal $u(t)$ should not exceed 112.5 V because of the physical limitations of the piezoelectric materials.

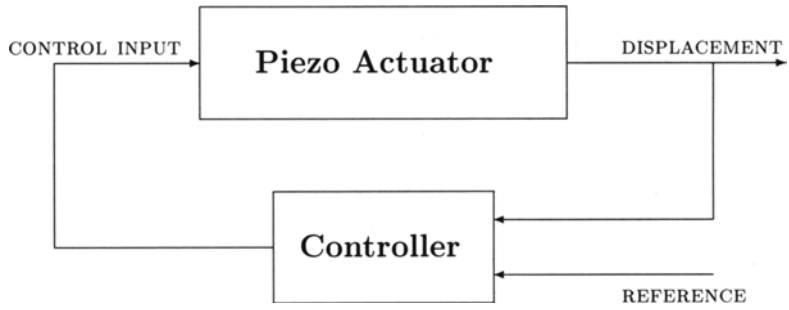


Figure 8.2. Piezoelectric bimorph actuator plant with controller.

Our approach is as follows: we will first use the stochastic equivalent linearisation method proposed by Chang [115] to obtain a linearised model for the nonlinear hysteretic dynamics. Then we reformulate our design into an H_∞

almost disturbance decoupling problem in which the disturbance inputs are the reference input and the error between the hysteretic dynamics and that of its linearised model, and where the controlled output is simply the double integration of the tracking error. Thus, our task becomes to design a controller such that, when it is applied to the piezoelectric actuator, the overall system is asymptotically stable, and the controlled output, which corresponds to the tacking error, is as small as possible and decays as fast as possible.

The outline of this chapter is as follows. In Section 8.2 a first-order linearised model is obtained for the nonlinear hysteresis using the stochastic equivalent linearisation method. A simulation result is also given to show the match between the nonlinear and linearised models. In Section 8.3 we formulate our controller design into a standard almost disturbance decoupling problem by properly defining the disturbance input and the controlled output. Two integrators are augmented into the original plant to enhance the performance of the overall system. Then a robust controller that is explicitly parameterised by a certain tuning parameter, and that solves the proposed almost disturbance decoupling problem, is carried out using a so-called asymptotic time scale and eigenstructure assignment technique. In Section 8.4 we present the final controller and simulation results of our overall control system using MATLAB SIMULINK. We also obtain an explicit relationship between the peak values of the control signal and the tuning parameter of the controller, as well as an explicit linear relationship of the maximum trackable frequency, *i.e.* the corresponding tracking error can be settled to 1%, versus the tuning parameter of the controller. The simulation results of this section clearly show that all the design specifications are met and that the overall performance is very satisfactory.

8.2 Linearisation of Nonlinear Hysteretic Dynamics

We will proceed to linearise the nonlinear hysteretic dynamics of Equation (8.2) in this section. As pointed out by Chang [115], there are basically three methods available in the literature to linearise the hysteretic types of nonlinear system. These are (i) the Fokker–Planck equation approach (see *e.g.*, Caughey [116]), (ii) the perturbation techniques (see *e.g.*, Crandall [117] and Lyon [118]) and (iii) the stochastic linearisation approach. All of them have certain advantages and limitations. However, the stochastic linearisation technique has the widest range of applications compared with the other methods. This method is based on the concept of replacing the nonlinear system with an "equivalent" linear system in such a way that the "difference" between these two systems is minimised in a certain sense. The technique was initiated by Booton [119]. In this chapter, we will just follow the stochastic linearisa-

tion method given by Chang [115] to obtain a linear model of the following form:

$$\dot{z} = k_1\dot{u} + k_2 z, \tag{8.4}$$

for the hysteretic dynamics of Equation (8.2), where k_1 and k_2 are the linearisation coefficients and are to be determined. The procedure is quite straightforward and proceeds as follows. First we introduce a so-called "difference" function e between $\dot{z}$ of Equation (8.2) and $\dot{z}$ of Equation (8.4):

$$e(k_1, k_2) = \alpha d\dot{u} - \beta|\dot{u}|z - \gamma\dot{u}|z| - (k_1\dot{u} + k_2 z). \tag{8.5}$$

Then minimising $\boldsymbol{E}[e^2]$, where $\boldsymbol{E}$ is the expectation operator, with respect to k_1 and k_2, we obtain

$$\frac{\partial \boldsymbol{E}[e^2]}{\partial k_1} = \frac{\partial \boldsymbol{E}[e^2]}{\partial k_2} = 0, \tag{8.6}$$

from which the stochastic linearisation coefficients k_1 and k_2 are determined. It turns out that if h and $\dot{u}$ are of zero means and jointly Gaussian, then k_1 and k_2 can be easily obtained. Let us assume that h and $\dot{u}$ have a joint probability density function

$$f_{\dot{u}z}(\dot{u}, z) = \frac{1}{2\pi\sigma_{\dot{u}}\sigma_z\sqrt{1-\rho_{\dot{u}z}^2}} \exp\left\{-\frac{\sigma_{\dot{u}}^2 z^2 - 2\sigma_{\dot{u}}\sigma_z\rho_{\dot{u}z}\dot{u}z + \sigma_z^2\dot{u}^2}{2\sigma_{\dot{u}}^2\sigma_z^2(1-\rho_{\dot{u}z}^2)}\right\},$$

where $\rho_{\dot{u}z}$ is the normalised covariance of $\dot{u}$ and z, and $\sigma_{\dot{u}}$ and σ_z are the standard deviations of $\dot{u}$ and z respectively. Then the linearisation coefficients k_1 and k_2 can be expressed as follows:

$$k_1 = \alpha d - \beta c_1 - \gamma c_2, \tag{8.7}$$

and

$$k_2 = -\beta c_3 - \gamma c_4, \tag{8.8}$$

where c_1, c_2, c_3 and c_4 are given by

$$c_1 = 0.79788456\sigma_z \cos\left[\tan^{-1}\left(\frac{\sqrt{1-\rho_{\dot{u}z}^2}}{\rho_{\dot{u}z}}\right)\right], \tag{8.9}$$

$$c_2 = 0.79788456\sigma_z, \qquad c_4 = 0.79788456\rho_{\dot{u}z}\sigma_{\dot{u}}, \tag{8.10}$$

and

$$c_3 = 0.79788456\sigma_{\dot{u}} \left\{ 1 - \rho_{\dot{u}z}^2 + \rho_{\dot{u}z} \cos \left[\tan^{-1} \left(\frac{\sqrt{1 - \rho_{\dot{u}z}^2}}{\rho_{\dot{u}z}} \right) \right] \right\}. \quad (8.11)$$

After a few iterations, we found that a sinusoidal excitation $\dot{u}$ with frequencies ranging from 0 to 100 Hz (the expected working frequency range) and peak magnitude of 50 V, which has a standard deviation of $\sigma_{\dot{u}} = 35$, would yield a suitable linearised model for Equation (8.2). For this excitation, we obtain $\sigma_z = 5 \times 10^{-7}$, $\rho_{\dot{u}z} = 5 \times 10^{-3}$

$$c_1 = 1.9947 \times 10^{-9}, \qquad c_2 = 3.9894 \times 10^{-7}, \quad (8.12)$$

$$c_3 = 27.9260, \qquad c_4 = 0.1396, \quad (8.13)$$

and

$$k_1 = 3.5382 \times 10^{-7}, \qquad k_2 = -0.9597. \quad (8.14)$$

The stochastic linearisation model of the given nonlinear hysteretic dynamics of Equation (8.2) is then given by

$$\dot{\hat{z}} = k_1 \dot{u} + k_2 \hat{z} = 3.5382 \times 10^{-7} \dot{u} - 0.9597 \hat{z}. \quad (8.15)$$

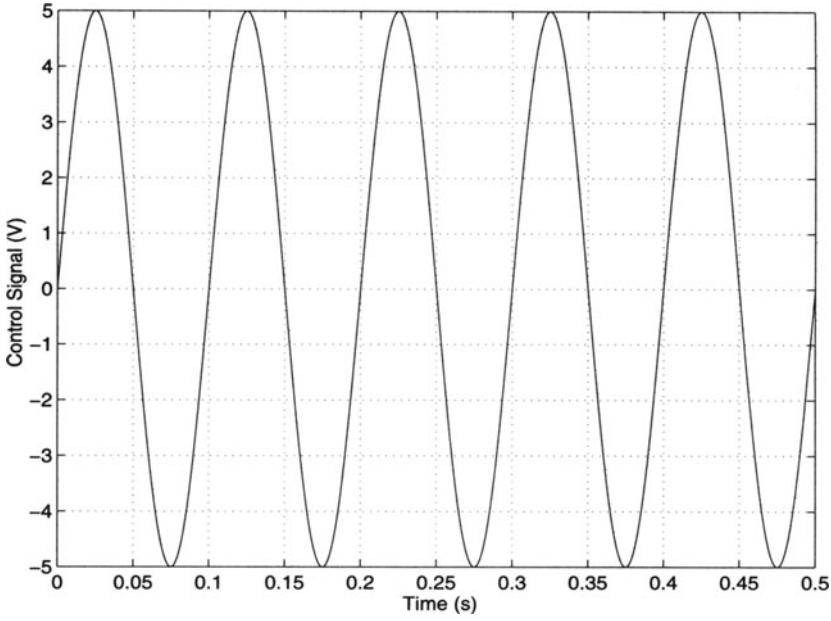
For future use, let us define the linearisation error as

$$e_z = z - \hat{z}. \quad (8.16)$$

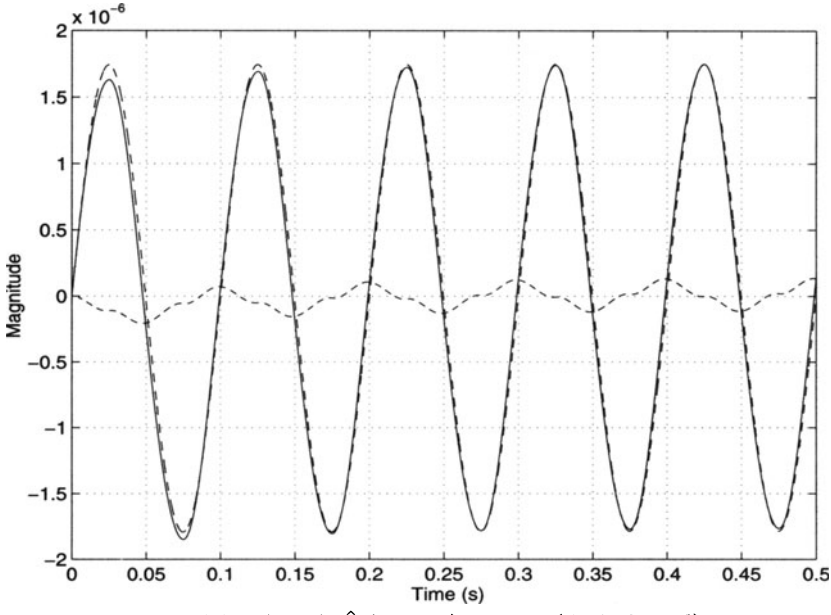
Figure 8.3 shows the open-loop simulation results of the nonlinear hysteresis and its linearised model, as well as the error for a typical sine wave input signal u. The results are quite satisfactory. Here we should note that, because of the nature of our approach in controller design later in the next section, the variation of the linearised model within a certain range, which might result in larger linearisation error e_z, will not much affect the overall performance of the closed-loop system. We will formulate e_z as a disturbance input and our controller will automatically reject it from the output response.

8.3 Almost Disturbance Decoupling Controller Design

This section is the heart of this chapter. We will first formulate our control system design for the piezoelectric bimorph actuator into a standard H_∞



(a) control input signal, u



(b) z (solid), $\hat{z}$ (dashed), and e_z (dash-dotted)

Figure 8.3. Responses of hysteresis and its linearised model to a sine input.

almost disturbance decoupling problem, and then apply the results of Chapter 4 to check the solvability of the proposed problem. Finally, we will utilise the results in Chapter 4 to find an internally stabilising controller that solves the proposed almost disturbance decoupling problem. Of course, most importantly, the resulting closed-loop system and its responses should meet all the design specifications as stated at the beginning of this chapter. To do this, we will have to convert the dynamic model of Equation (8.1) with the linearised model of the hysteresis into a state-space form. Let us first define a new state variable

$$v = \hat{z} - k_1 u. \tag{8.17}$$

Then from Equation (8.15), we have

$$\dot{v} = \dot{\hat{z}} - k_1 \dot{u} = k_2 \hat{z} = k_2 v + k_1 k_2 u. \tag{8.18}$$

Substituting Equations (8.16) and (8.17) into Equation (8.1), we obtain

$$\ddot{x}_1 + \frac{b}{m}\dot{x}_1 + \frac{k}{m}x_1 + \frac{k}{m}v = \frac{k(d-k_1)}{m}u - \frac{k}{m}e_z. \tag{8.19}$$

The overall controller structure of our approach is then depicted in Figure 8.4. Note that in Figure 8.4 we have augmented two integrators after e, the tracking error between the displacement x_1 and the reference input signal r. We have observed a very interesting property of this problem, *i.e.*, the more integrators that we augment after the tracking error e, the smaller the tracking error we can achieve for the same level of control input u. Because our control input u is limited to the range from -112.5 to 112.5 V, it turns out that two integrators are needed in order to meet all the design specifications. It is clear to see that the augmented system has an order of five. Next, let us define the state of the augmented system as

$$x = (\, x_1 \quad \dot{x}_1 \quad v \quad x_4 \quad x_5 \,)', \tag{8.20}$$

and the measurement output

$$y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_4 \\ x_5 \end{pmatrix}, \tag{8.21}$$

i.e., the original measurement of displacement x_1 plus two augmented states. The auxiliary disturbance input is

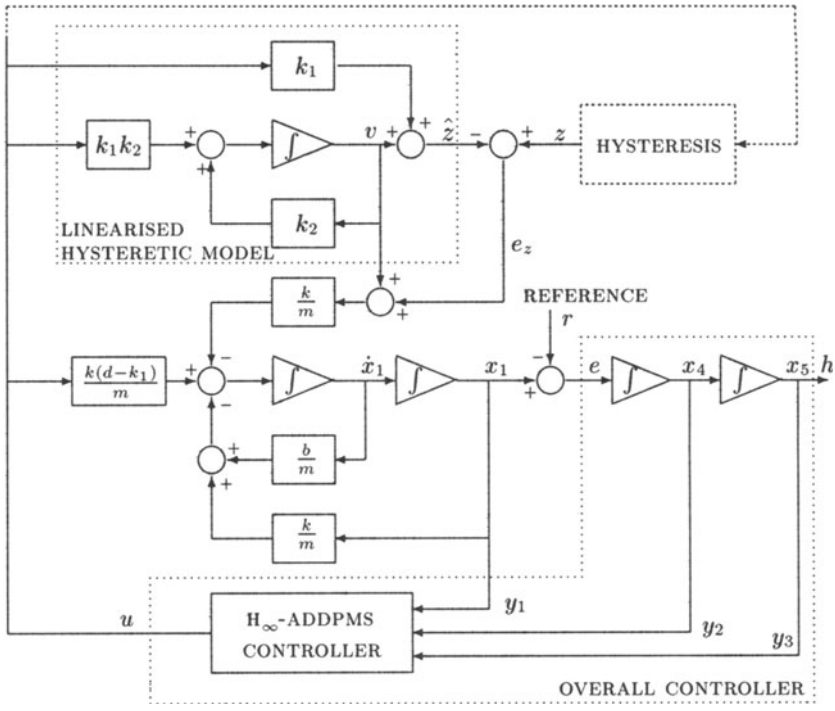


Figure 8.4. Augmented linearised model with controller.

$$w = \begin{pmatrix} e_z \\ r \end{pmatrix}, \tag{8.22}$$

and the output to be controlled, h, is simply the double integration of the tracking error. The state-space model of the overall augmented system is then given by

$$\Sigma: \quad \begin{cases} \dot{x} = A\ x + B\ u + E\ w, \\ y = C_1\ x \qquad\quad + D_1\ w, \\ h = C_2\ x + D_2\ u\ , \end{cases} \tag{8.23}$$

with

$$C_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \tag{8.24}$$

$$C_2 = [0 \quad 0 \quad 0 \quad 0 \quad 1], \quad D_2 = 0, \tag{8.25}$$

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ -k/m & -b/m & -k/m & 0 & 0 \\ 0 & 0 & k_2 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ -274921.63 & -73.2915 & -274921.63 & 0 & 0 \\ 0 & 0 & -0.9597 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \tag{8.26}$$

$$B = \begin{bmatrix} 0 \\ k(d-k_1)/m \\ k_1 k_2 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.12841 \\ -3.39561 \times 10^{-7} \\ 0 \\ 0 \end{bmatrix}, \tag{8.27}$$

$$E = \begin{bmatrix} 0 & 0 \\ -k/m & 0 \\ 0 & 0 \\ 0 & -1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ -274921.63 & 0 \\ 0 & 0 \\ 0 & -1 \\ 0 & 0 \end{bmatrix}. \tag{8.28}$$

For the problem that we are considering here, it is simple to verify that the system Σ of Equation (8.23) has the following properties.

1. The subsystem (A, B, C_2, D_2) is invertible and of minimum phase with one invariant zero at -1.6867. It also has one infinite zero of order four.

2. The subsystem (A, E, C_1, D_1) is left invertible and of minimum phase with one invariant zero at -0.9597 and two infinite zeros of orders one and two.

Then, it follows from the results of Section 4.4 that the H_∞-ADDPMS for the system in Equation (8.23) is solvable. In fact, one can design either a full order observer-based controller or a reduced order observer-based controller to solve this problem. For the full order observer-based controller, the order of the disturbance decoupling controller (see Figure 8.4) will be five and the order of the final overall controller (again see Figure 8.4) will be seven (the disturbance decoupling controller plus two integrators). On the other hand, if we use a reduced order observer in the disturbance decoupling controller, the total order of the resulting final overall controller will be reduced to four. From the practical point of view, the latter is much more desirable than the former. Thus, in what follows we will only focus on the controller design based on a reduced order observer. We can separate our controller design into two steps:

1. In the first step, we assume that all five states of Σ in Equation (8.23) are available and then design a static and parameterised state feedback control law,

$$u = F(\varepsilon)x, \tag{8.29}$$

such that it solves the almost disturbance decoupling problem for the state feedback case, *i.e.*, $y = x$, by adjusting the tuning parameter ε to an appropriate value;

2. in the second step, we design a reduced order observer-based controller. It has a parameterised reduced order observer gain matrix $K_2(\varepsilon)$ that can be tuned to recover the performance achieved by the state feedback control law in the first step.

We will use the structural decomposition approach of Chapter 4 to construct both the state feedback law and the reduced order observer gain. We would like to note that, in principle, one can also apply the ARE-based H_∞ optimisation technique (see *e.g.*, Zhou and Khargonekar [72]) to solve this problem. However, because the numerical conditions of our system Σ are very bad, we are unable to obtain any satisfactory solution from the ARE approach. We cannot get any meaningful solution for the associated H_∞ continuous-time ARE in MATLAB. The following is a closed form solution of the static state feedback parameterised gain matrix obtained using the method given in Chapter 4.

$$F(\varepsilon) = \Big[(2.1410\times10^6 - 62.3004/\varepsilon^2) \quad (570.7619 - 31.1502/\varepsilon) \\ 2.1410\times10^6 \quad -62.3004/\varepsilon^3 \quad -31.1502/\varepsilon^4\Big], \tag{8.30}$$

where ε is the tuning parameter that can be adjusted to achieve almost disturbance decoupling. It can be verified that the closed-loop system matrix, $A + BF(\varepsilon)$ is asymptotically stable for all $0 < \varepsilon < \infty$ and the closed-loop transfer function from the disturbance w to the controlled output h, $T_{hw}(\varepsilon, s)$, satisfying

$$\|T_{hw}(\varepsilon, s)\|_\infty = \|[C_2 + D_2F(\varepsilon)][sI - A - BF(\varepsilon)]^{-1}E\|_\infty \to 0, \tag{8.31}$$

as $\varepsilon \to 0$.

The next step is to design a reduced order observer-based controller that will recover the performance of the above state feedback control law. First, let us perform the following nonsingular (permutation) state transformation to the system Σ of Equation (8.23),

$$x = T\tilde{x}, \tag{8.32}$$

where

$$T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \tag{8.33}$$

such that the transformed measurement matrix has the form of

$$C_1 T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} = [I_3 \quad 0]. \tag{8.34}$$

Clearly, the first three states of the transformed system, or x_1, x_4 and x_5 of the original system Σ in Equation (8.23), need not be estimated as they are already available from the measurement output. Let us now partition the transformed system as follows:

$$T^{-1}AT = \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right] = \left[\begin{array}{ccc|cc} 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \hline -274921.63 & 0 & 0 & -73.2915 & -274921.63 \\ 0 & 0 & 0 & 0 & -0.9597 \end{array}\right], \tag{8.35}$$

$$T^{-1}B = \left[\begin{array}{c} B_1 \\ \hline B_2 \end{array}\right] = \left[\begin{array}{c} 0 \\ 0 \\ 0 \\ \hline 0.12841 \\ -3.39561 \times 10^{-7} \end{array}\right], \tag{8.36}$$

$$T^{-1}E = \left[\begin{array}{c} E_1 \\ \hline E_2 \end{array}\right] = \left[\begin{array}{cc} 0 & 0 \\ 0 & -1 \\ 0 & 0 \\ \hline -274921.63 & 0 \\ 0 & 0 \end{array}\right]. \tag{8.37}$$

Also, we partition

$$F(\varepsilon)T = [F_1(\varepsilon)|F_2(\varepsilon)] \tag{8.38}$$

$$= \Big[(2.14\times 10^6 - 62.3/\varepsilon^2) \quad -62.3/\varepsilon^3 \quad -31.15/\varepsilon^4 \;\Big|\; (570.76 - 31.15/\varepsilon) \quad 2.14\times 10^6 \Big]. \tag{8.39}$$

Then the reduced order observer-based controller (see Chapter 4) is given in the form of

$$\Sigma_{\text{cmp}} : \begin{cases} \dot{v} = A_{\text{cmp}}(\varepsilon)\, v + B_{\text{cmp}}(\varepsilon)\, y, \\ u = C_{\text{cmp}}(\varepsilon)\, v + D_{\text{cmp}}(\varepsilon)\, y, \end{cases} \tag{8.40}$$

with

$$A_{\text{cmp}}(\varepsilon) = A_{22} + K_2(\varepsilon)A_{12} + B_2F_2(\varepsilon) + K_2(\varepsilon)B_1F_2(\varepsilon), \tag{8.41}$$

$$B_{\text{cmp}}(\varepsilon) = A_{21} + K_2(\varepsilon)A_{11} - \big[A_{22} + K_2(\varepsilon)A_{12}\big]K_2(\varepsilon) + \big[B_2 + K_2(\varepsilon)B_1\big]\big[F_1(\varepsilon) - F_2(\varepsilon)K_2(\varepsilon)\big], \tag{8.42}$$

$$C_{\text{cmp}}(\varepsilon) = F_2(\varepsilon), \tag{8.43}$$

$$D_{\text{cmp}}(\varepsilon) = F_1(\varepsilon) - F_2(\varepsilon)K_2(\varepsilon), \tag{8.44}$$

where $K_2(\varepsilon)$ is the parameterised reduced order observer gain matrix and is to be designed such that $A_{22} + K_2(\varepsilon)A_{12}$ is asymptotically stable for sufficiently small ε and also

$$\|[sI - A_{22} - K_2(\varepsilon)A_{12}]^{-1}[E_2 + K_2(\varepsilon)E_1]\|_\infty \to 0, \tag{8.45}$$

as $\varepsilon \to 0$. Again, using the software package of [34], we obtained the following parameterised reduced order observer gain matrix

$$K_2(\varepsilon) = \begin{bmatrix} 73.2915 - 1/\varepsilon & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{8.46}$$

Then the explicitly parameterised matrices of the state-space model of the reduced order observer-based controller are given by

$$A_{\text{cmp}}(\varepsilon) = \begin{bmatrix} 73.2915 - 4/\varepsilon - 1/\varepsilon & 0 \\ -1.9381\times 10^{-4} + 1.0577\times 10^{-5}/\varepsilon & -1.6867 \end{bmatrix},$$

$$C_{\text{cmp}}(\varepsilon) = [570.7619 - 31.1502/\varepsilon \quad 2140967],$$

$$D_{\text{cmp}}(\varepsilon) = [2099135.4 + 2853.81/\varepsilon - 93.45/\varepsilon^2 \quad -62.3/\varepsilon^3 \quad -31.15/\varepsilon^4],$$

$$B_{\text{cmp}}(\varepsilon) = \begin{bmatrix} \psi_1 & -8/\varepsilon^3 & -4/\varepsilon^4 \\ \psi_2 & 2.1155\times 10^{-5}/\varepsilon^3 & 1.0577\times 10^{-5}/\varepsilon^4 \end{bmatrix},$$

where

$$\psi_1 = -5731.6533 - 13/\varepsilon^2 + 439.7492/\varepsilon, \tag{8.47}$$

and

$$\psi_2 = -0.7128 + 3.1732 \times 10^{-5}/\varepsilon^2 - 9.6904 \times 10^{-4}/\varepsilon. \tag{8.48}$$

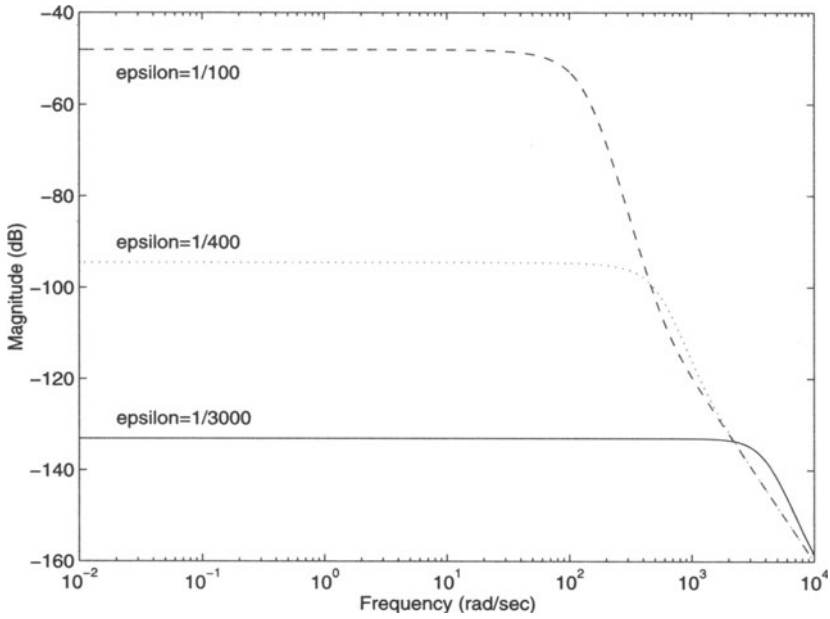


Figure 8.5. Maximum singular values of closed-loop transfer function $T_{hw}(\varepsilon, s)$.

The overall closed-loop system comprising the system Σ of Equation (8.23) and the above controller would be asymptotically stable as long as $\varepsilon \in (0, \infty)$. In fact, the closed-loop poles are exactly located at -1.6867, two pairs at $-1/\varepsilon \pm j1/\varepsilon$, -0.9597 and $-1/\varepsilon$. The plots of the maximum singular values of the closed-loop transfer function matrix from the disturbance w to the controlled output h, namely $T_{hw}(\varepsilon, s)$, for several values of ε, *i.e.*, $\varepsilon = 1/100$, $\varepsilon = 1/400$ and $\varepsilon = 1/3000$, in Figure 8.5 show that as ε tends to smaller values, the H_∞-norm of $T_{hw}(\varepsilon, s)$ becomes also smaller. Hence, almost disturbance decoupling is indeed achieved. These are the properties of our control system in the frequency domain. In the next section we will address its time domain properties, which are, of course, much more important, as all the design specifications are in the time domain.

8.4 Final Controller and Simulation Results

In this section we will put our design of the previous section into a final controller as depicted in Figure 8.2. It is simple to derive the state-space model of the final overall controller by observing its interconnection with the disturbance decoupling controller Σ_{cmp} of Equation (8.40) (see Figure 8.4). We will also present simulation results of the responses of the overall design to several different types of reference input signal. They clearly show that all the design specifications are successfully achieved. Furthermore, because our controller is explicitly parameterised by a tuning parameter, it is very easy to adjust it to meet other design specifications without going through it all over again from the beginning. This will also be discussed next.

As mentioned earlier, the final overall controller of our design will be of the order of four, of which two are from the disturbance decoupling controller and two from the augmented integrators. It has two inputs: one is the displacement x_1 and the other is the reference signal r. It is straightforward to verify that the state-space model of the final overall controller is given by

$$\Sigma_{\text{oc}}(\varepsilon): \quad \begin{cases} \dot{v} = A_{\text{oc}}(\varepsilon)\ v + B_{\text{oc}}(\varepsilon)\ x_1 + G_{\text{oc}}\ r, \\ u = C_{\text{oc}}(\varepsilon)\ v + D_{\text{oc}}(\varepsilon)\ x_1\ , \end{cases} \tag{8.49}$$

where $A_{\text{oc}}(\varepsilon)$ is given by

$$\begin{bmatrix} 73.29 - 5/\varepsilon & 0 & -8/\varepsilon^3 & -4/\varepsilon^4 \\ 1.06\times 10^{-5}/\varepsilon & -1.69 & 2.12\times 10^{-5}/\varepsilon^3 & 1.08\times 10^{-5}/\varepsilon^4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \tag{8.50}$$

$$G_{\text{oc}} = \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \qquad B_{\text{oc}}(\varepsilon) = \begin{bmatrix} \psi_1 \\ \psi_2 \\ 1 \\ 0 \end{bmatrix}, \tag{8.51}$$

with ψ_1 and ψ_2 given by Equations (8.47) and (8.48) respectively,

$$C_{\text{oc}}(\varepsilon) = [570.76 - 31.15/\varepsilon \quad 2140967 \quad -62.3/\varepsilon^3 \quad -31.15/\varepsilon^4], \tag{8.52}$$

and

$$D_{\text{oc}}(\varepsilon) = 2099135.4 - 93.45/\varepsilon^2 + 2853.81/\varepsilon. \tag{8.53}$$

There are some very interesting and useful properties of this parameterised controller. After repeatedly simulating the overall design, we found that the

maximum peak values of the control signal u are independent of the frequencies of the reference signals. They are only dependent on the initial error between displacement x_1 and the reference r. The larger the initial error is, the bigger is the peak that occurs in u. Because the working range of our actuator is within ± 1 μm, we will assume that the largest magnitude of the initial error in any situation should not be larger that 1 μm. This assumption is reasonable, as we can always reset our displacement x_1 to zero before the system is to track any reference and hence the magnitude of initial tracking error can never be larger than 1 μm. Let us consider the worst case, *i.e.*, that the magnitude of the initial error is 1 μm. Then, interestingly, we are able to obtain a clear relationship between the tuning parameter $1/\varepsilon$ and the maximum peak of u. The result is plotted in Figure 8.6. We also found that the tracking error is independent of initial errors. It only depends on the frequencies of the references, *i.e.*, the larger is the frequency that the reference signal r has, the larger the tracking error that occurs. Again, we can obtain a simple and linear relationship between the tuning parameter ε and the maximum frequency that a reference signal can have such that the corresponding tracking error is no larger than 1%, which is one of our main design specifications. The result is plotted in Figure 8.7.

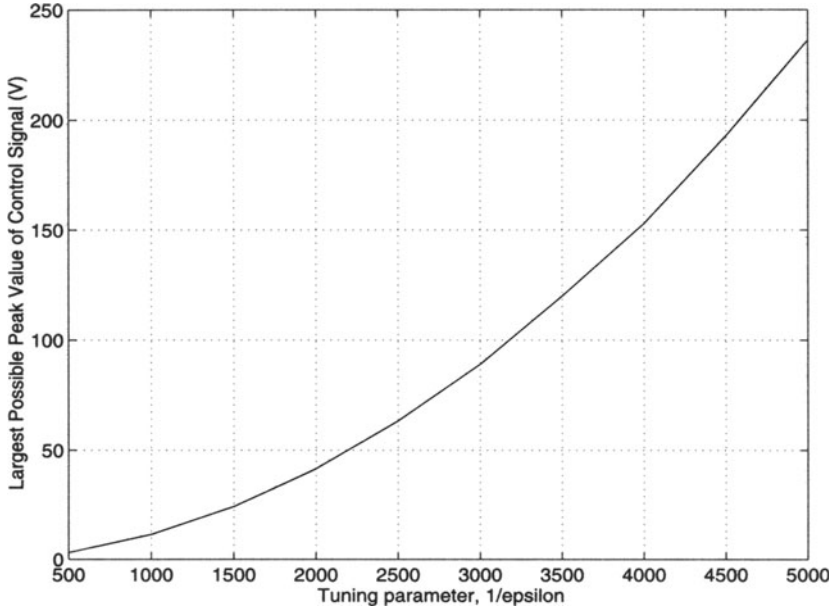


Figure 8.6. Parameter $1/\varepsilon$ versus maximum peaks of u in worst initial errors.

Clearly, from Figure 8.6, we know that owing to the constraints on the control input, *i.e.*, it must be kept within ± 112.5 V, we have to select our controller

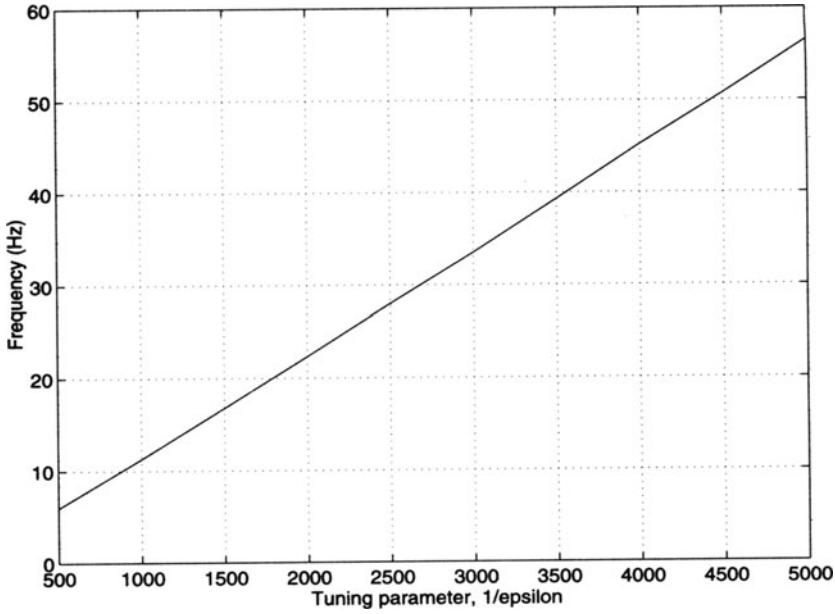


Figure 8.7. Parameter $1/\varepsilon$ versus maximum frequency of r that has 1% tracking error.

with $\varepsilon > 1/3370$. From Figure 8.7, we know that in order to meet the first design specification, *i.e.*, the steady-state tracking errors should be less than 1% for reference inputs that have frequencies up to 30 Hz, we have to choose our controller with $\varepsilon < 1/2680$. Hence, the final controller as given in Equation (8.49) to Equation (8.53) will meet all the design goals for our piezoelectric actuator system of Equations (8.1) and (8.2), for all $\varepsilon \in (1/3370, 1/2680)$. Let us choose $\varepsilon = 1/3000$. We obtain the overall controller as in the form of Equation (8.49) with

$$A_{\text{oc}} = \begin{bmatrix} -14926.71 & 0 & -2.16 \times 10^{11} & 3.24 \times 10^{14} \\ 0.0315 & -1.69 & 5.71 \times 10^{5} & 8.57 \times 10^{8} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \tag{8.54}$$

$$B_{\text{oc}} = \begin{bmatrix} -1.16 \times 10^{8} \\ 281.97 \\ 1 \\ 0 \end{bmatrix}, \quad G_{\text{oc}} = \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \quad D_{\text{oc}} = -8.3 \times 10^{8}, \tag{8.55}$$

$$C_{\text{oc}} = \begin{bmatrix} -92879.9 & 2140967 & -1.68 \times 10^{12} & -2.52 \times 10^{15} \end{bmatrix}. \tag{8.56}$$

The simulation results presented in the following are done using the MATLAB SIMULINK package, which is widely available everywhere these days. Two different reference inputs are simulated using the Runge–Kutta 5 method in SIMULINK with a minimum step size of 10 μs and a maximum step size of 100 μs as well as a tolerance of 10^{-5}. These references are: (1) a cosine signal with a frequency of 30 Hz and peak magnitude of 1 μm, and (2) a sine signal with a frequency of 34 Hz and peak magnitude of 1 μm. The results for the cosine signal are given in Figures 8.8 to 8.10. In Figure 8.8, the solid-line curve is x_1 and the dash-dotted curve is the reference. The tracking error and the control signal corresponding to this reference are given in Figures 8.9 and 8.10 respectively. Similarly, Figures 8.11 to 8.13 are the results corresponding to the sine signal. All these results show that our design goals are fully achieved. To be more specific, the tracking error for a 30 Hz cosine wave reference is about 0.8%, which is better than the specification, and the worst peak magnitude of the control signal is less than 90 V, which is of course less than the saturated level, *i.e.*, 112.5 V. Furthermore, the 1% tracking error settling times for both cases are less than 0.003 s. It is interesting to note that the performance of the actual closed-loop system is even better than that of its linear counterpart.

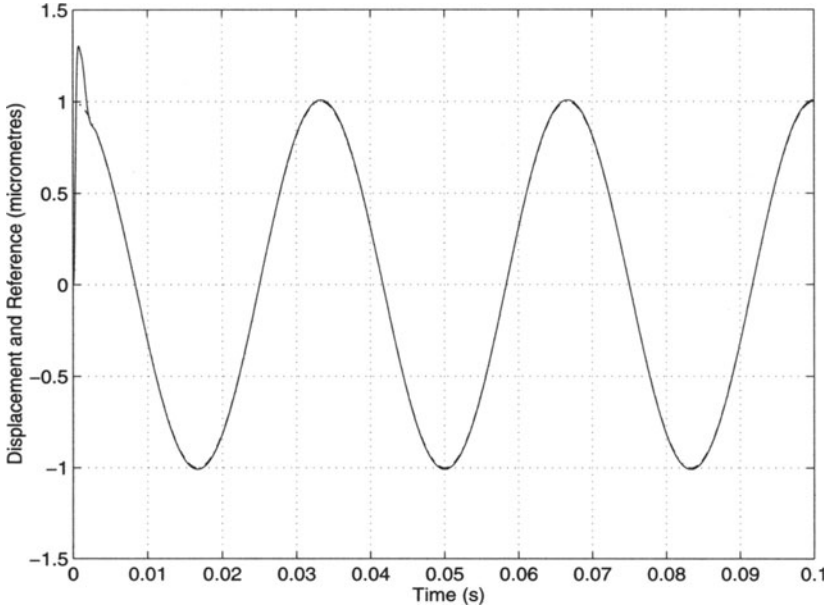
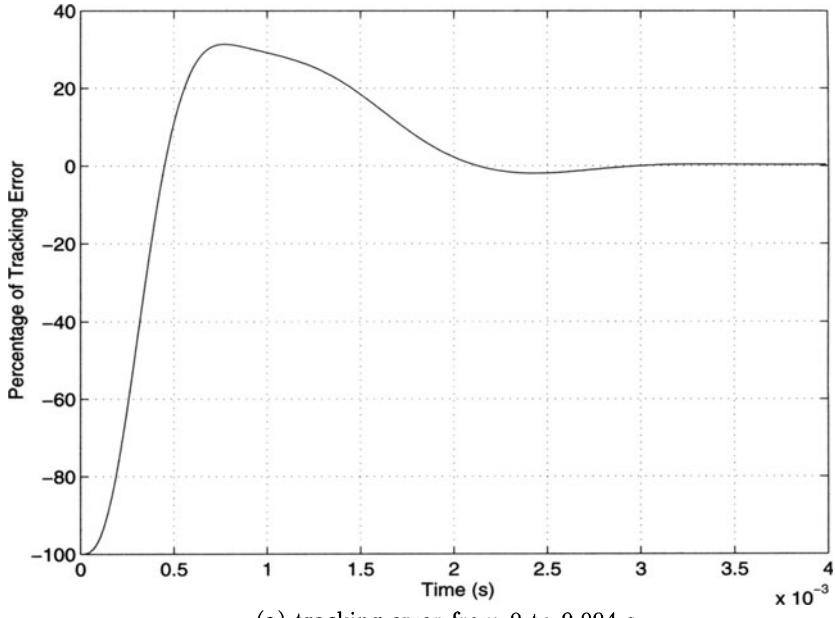
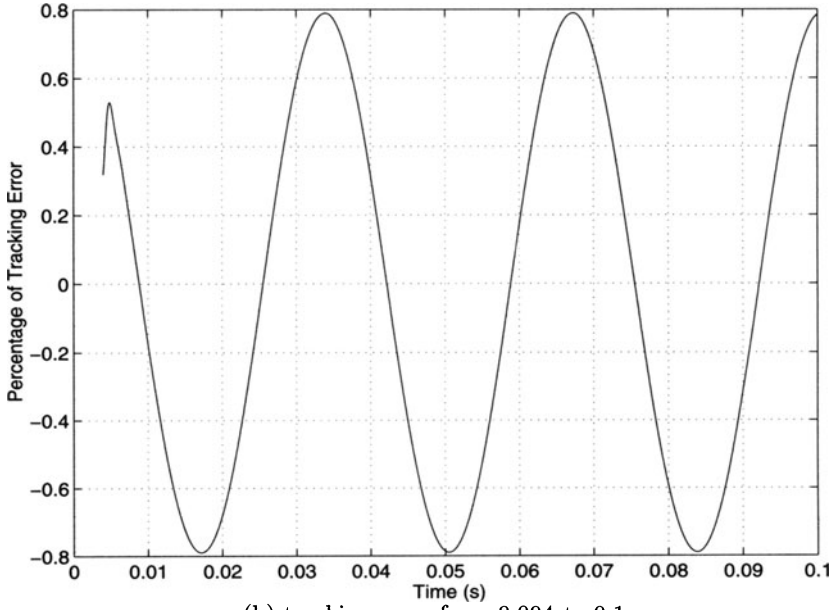


Figure 8.8. Responses of the displacement and the 30 Hz cosine reference.

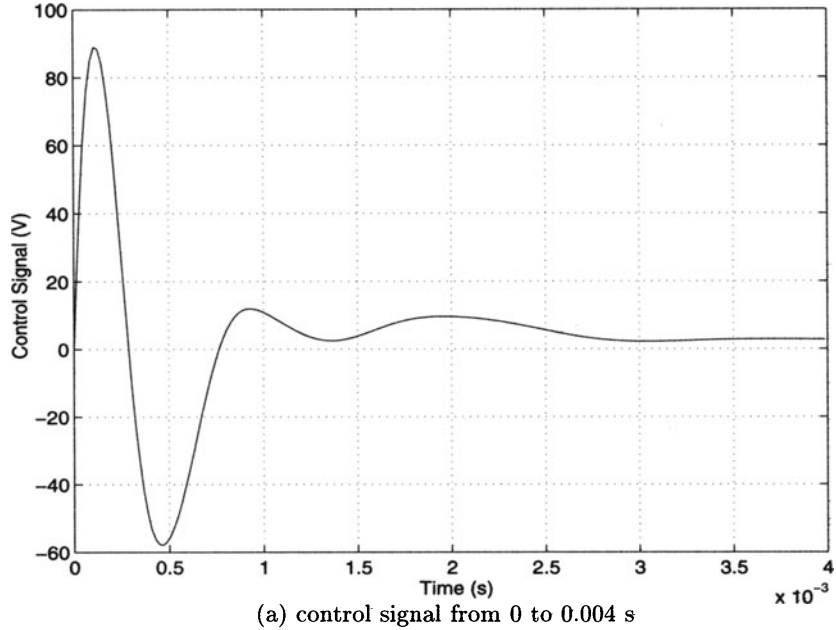


(a) tracking error from 0 to 0.004 s

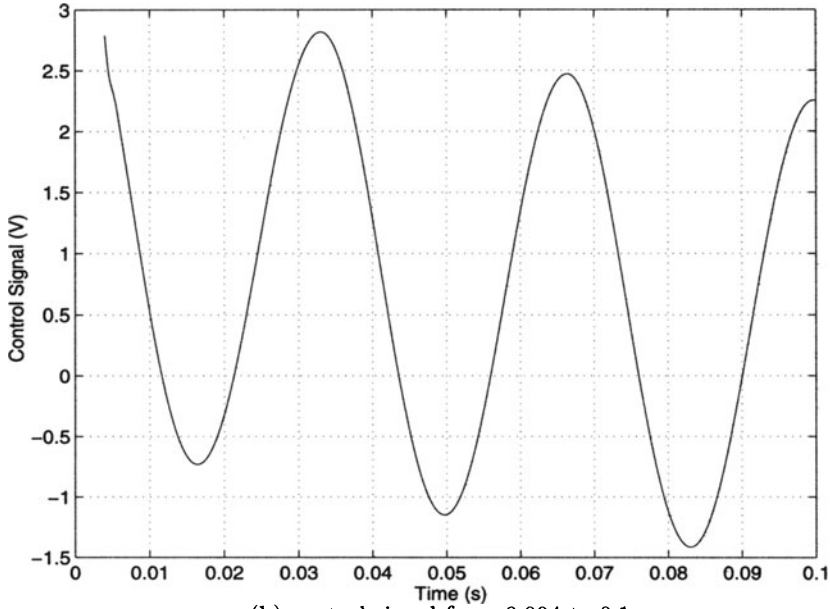


(b) tracking error from 0.004 to 0.1 s

Figure 8.9. Tracking error for the 30 Hz cosine reference.



(a) control signal from 0 to 0.004 s



(b) control signal from 0.004 to 0.1 s

Figure 8.10. Control signal for the 30 Hz cosine reference.

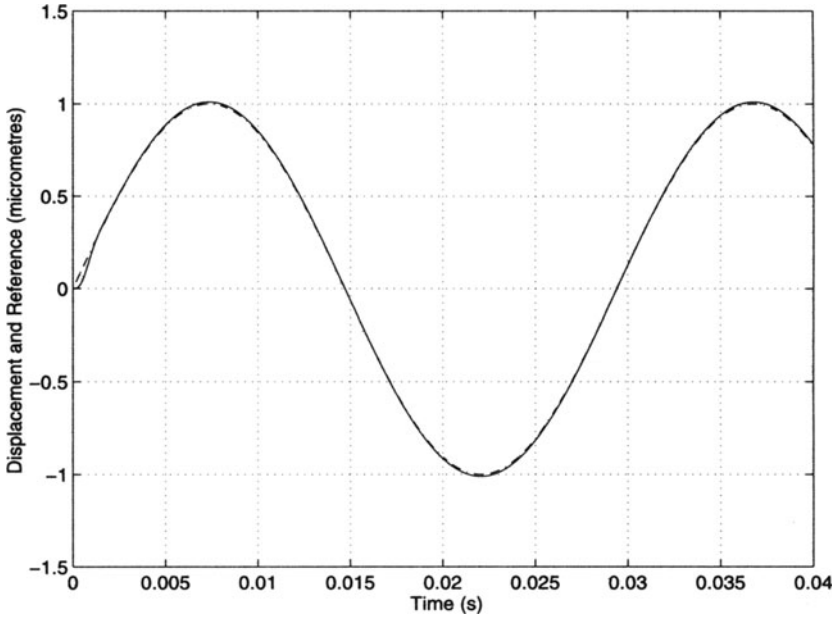


Figure 8.11. Responses of the displacement and the 34 Hz sine reference.

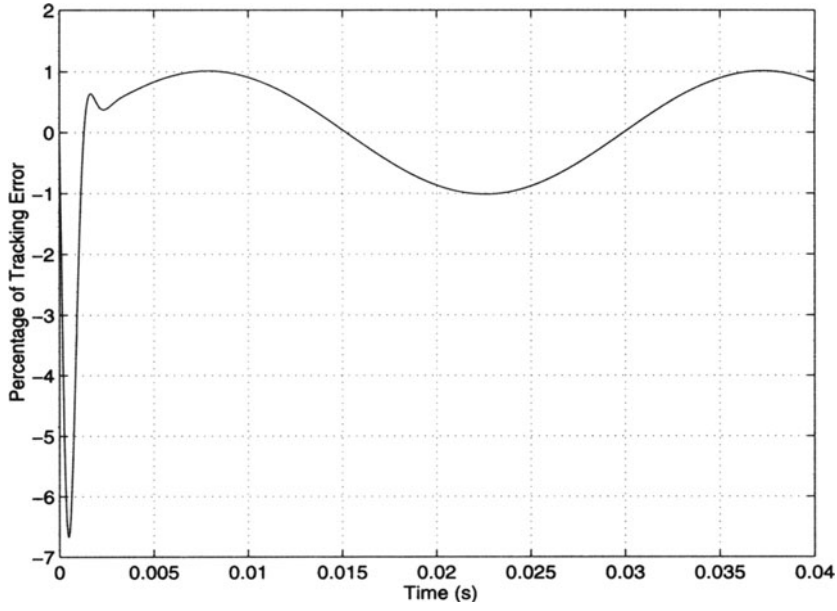


Figure 8.12. Tracking error for the 34 Hz sine reference.

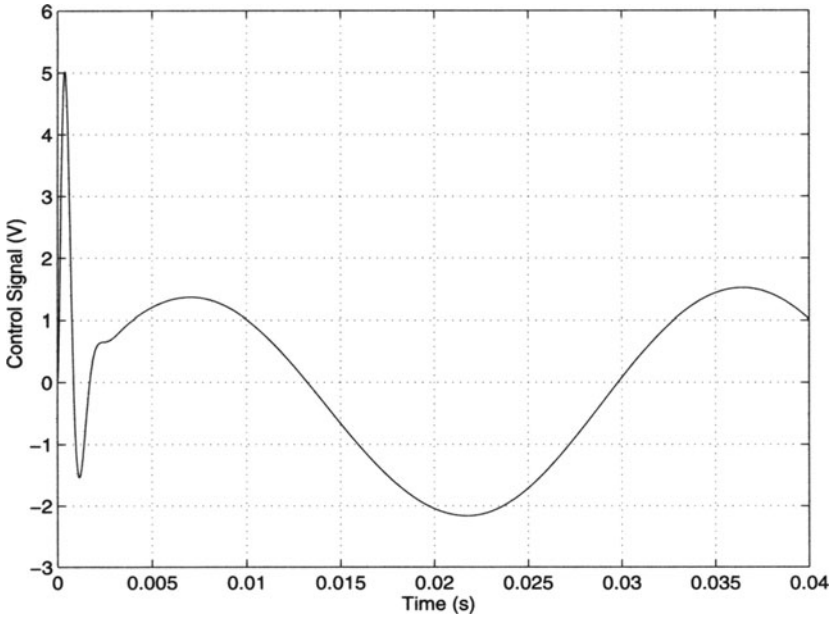


Figure 8.13. Control signal for the 34 Hz sine reference.

CHAPTER 9
DUAL-STAGE ACTUATED SERVO SYSTEMS

9.1 Introduction

The present demand for large-capacity disk drives is leading to an increase in areal density at a rate of 100% per year. This requires a positioning accuracy of the order of few nanometres. The servo bandwidth of the current disk drive actuators makes it very hard to achieve this. The VCM actuator used in conventional disk drives has hundreds of flexible resonances in high frequencies (see *e.g.*, [1, 29]), which limits the increase of bandwidth and hence the positioning accuracy. In order to develop high bandwidth (track following) servo systems, dual-stage actuation has been proposed as a possible solution. Dual-stage actuator refers to the fact that there is a microactuator mounted on a large conventional VCM actuator. The VCM actuator is used for coarse and relatively slow positioning and the microactuator is used for fine and fast positioning. The two most fundamental choices in a dual-stage system are the actuator configuration and the control algorithm. There have been proposals for electromagnetic, electrostatic, piezoelectric, shape memory and rubber microactuators, *etc.*, each with their own advantages and disadvantages. Many research studies have been done and reported in the literature (see *e.g.*, [20, 114, 120–137] just to name a few). In this chapter, we focus on the design of complete HDD servo systems with a dual-stage actuator with a piezoelectric actuator in its second stage (see Figure 9.1).

Diverse control strategies and methods have been reported in the design of HDD servo systems with a dual-stage actuator (see *e.g.*, [121, 125, 127, 128, 130, 132–134]). Guo *et al.* [125] have proposed four control strategies to design the dual-stage actuator control system: the so-called parallel loop, master–slave loop, dual feedback loop and master-slave with decoupling methods. Hu *et al.* [130] and Guo *et al.* [127] have also utilised the well-known LQG/LTR method to design the dual-stage actuator control system. These studies have accelerated the progress to improve HDD servo system performances, but more studies need to be done before such dual-stage actuated HDDs can be considered for commercialisation.

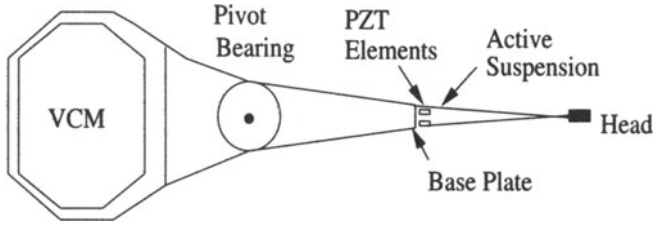


Figure 9.1. A dual-stage HDD actuator.

We present in this chapter the design of complete dual-stage actuated HDD servo systems with two different approaches, *i.e.*, the MSC method, and the CNF control technique. In each design, the VCM actuator will be controlled by a control law obtained using one of these two approaches and the microactuator will be controlled by an appropriate proportional gain together with a low-pass filter. Here we note that the pure PTOS approach generally cannot achieve a satisfactory performance. As such, we will drop the results of such an approach in this chapter.

9.2 Modelling of a Dual-stage Actuator

In this section we will develop a model for the dual-stage actuator. Since our control design will be such that the VCM actuator and the microactuator are decoupled, we need to identify two separate models for the VCM actuator and for the microactuator. A Maxtor dual-stage HDD is used in our design and implementation. It has a fine positioner based on a piezoelectric suspension mounted at the end of a primary VCM arm (see Figure 9.1), and the microactuators produce the relative motion of the R/W head along the radial direction. Here we note that only the displacement of the R/W head is available as the measurement output. Also, the VCM arm in this HDD is quite similar to that in the HDD studied in Chapters 6 and 7. Figures 9.2 and 9.3 respectively show the frequency response characteristics of the VCM actuator and the microactuator.

Using the data measured from the actual system, and the algorithm given in Chapter 2 (see also [13, 37]), we obtain a fourth-order model for the VCM actuator,

$$G_{\rm v}(s) = \frac{6.4013 \times 10^7}{s^2} \cdot \frac{2.467 \times 10^8}{s^2 + 2.513 \times 10^3 s + 2.467 \times 10^8}, \tag{9.1}$$

and a fourth-order model for the microactuator,

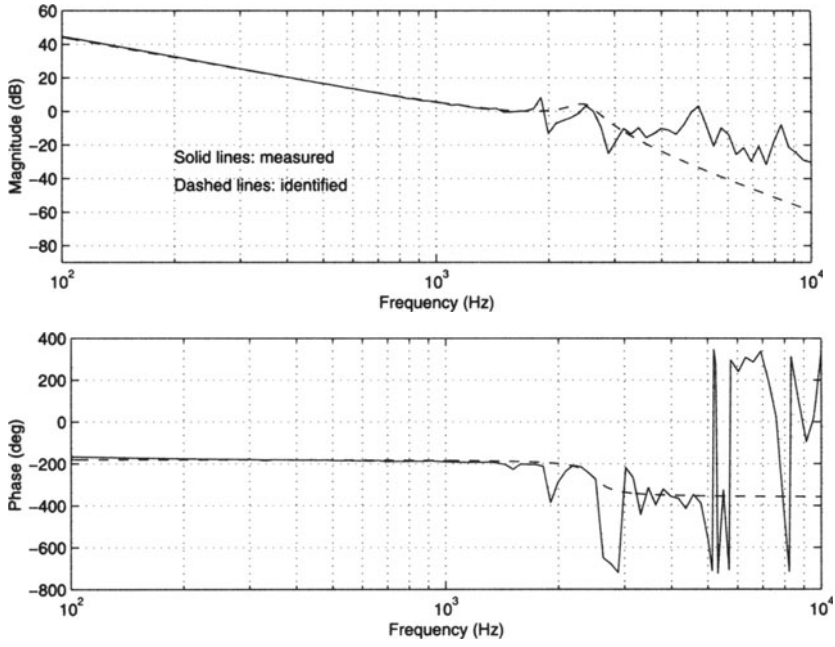


Figure 9.2. Frequency responses of the actual and identified VCM actuator models.

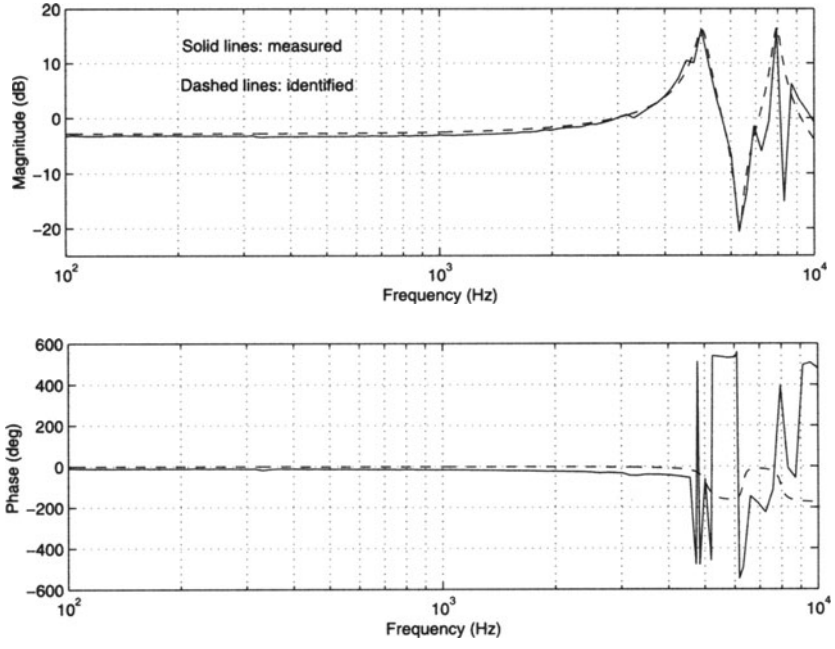


Figure 9.3. Frequency responses of the actual and identified microactuator models.

$$G_{\mathrm{m}}(s) = \frac{0.722(1.593\times10^{9}s^{2}+1.708\times10^{12}s+2.512\times10^{18})}{s^{4}+4256s^{3}+3.506\times10^{9}s^{2}+7.496\times10^{12}s+2.512\times10^{18}}. \quad (9.2)$$

We note here that the inputs to both actuators are voltages (in volts) and their outputs are displacements (in micrometres). In order to simplify our design procedure, we approximate the models of the VCM actuator and the microactuator, respectively, as follows:

$$\Sigma_{\mathrm{v}}: \begin{cases} \dot{x}_{\mathrm{v}} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x_{\mathrm{v}} + \begin{bmatrix} 0 \\ 6.4013\times10^{7} \end{bmatrix} u_{\mathrm{v}} \\ y_{\mathrm{v}} = \begin{bmatrix} 1 & 0 \end{bmatrix} x_{\mathrm{v}} \end{cases} \quad (9.3)$$

and Σ_{m}:

$$y_{\mathrm{m}} = 0.722\, u_{\mathrm{m}}. \quad (9.4)$$

Note that the control input to the VCM actuator should be constrained within ± 3 V, which is the same as the one used in Chapters 6 and 7, and the control input to the microactuator should be kept within ± 2 V. Moreover, the maximum displacement that can be generated by the microactuator should not exceed 1 μm, which is equivalent to one track pitch of the HDD servo system. Clearly, we are having a plant with both sensor and actuator nonlinearities.

9.3 Dual-stage Servo System Design

We now carry out the design of servo systems for the HDD with a dual-stage actuator. Similarly, we would like to design our servo systems to the following requirements.

1. The control input to the VCM actuator should not exceed ± 3 V, whereas the control input to the microactuator should be within ± 2 V.
2. The displacement of the microactuator should not exceed 1 μm. Moreover, it should settle down to zero in the steady state so that the microactuator can be further used for the next move.
3. The overshoot and undershoot of the step response should be kept less than 0.05 μm, *i.e.*, 5% of one track pitch. As pointed out earlier, the R/W head of the HDD servo system can start writing data on to the disk when it is within 5% of one track pitch of the target.

Unfortunately, the only available measurement in the dual-stage actuated HDD is the displacement of the R/W head, which is a combination of the displacement of the VCM actuator and that of the microactuator. Practically, we have to control both actuators using the same measurement, which will make the servo system design very difficult. Observing that the frequency response of the microactuator is as in Figure 9.3, we find that it is nothing more than a constant gain at low frequencies with a gain of 0.722. This property will be valid so long as we do not push the speed of the microactuator too fast. As such, we propose in Figure 9.4 a control configuration for the dual-stage actuated servo system.

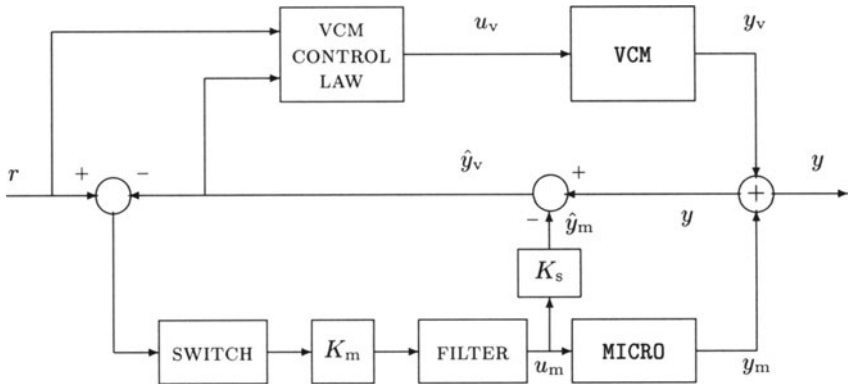


Figure 9.4. The schematic representation of a dual-stage actuator control.

To be more specific, we estimate the displacement of the microactuator $\hat{y}_m$ directly from its input u_m and then the estimation of the displacement of the VCM actuator can be obtain as $y - \hat{y}_m$. In the control configuration of Figure 9.4, we add a low-pass filter on the path to the control input of the microactuator. The main purpose of adding such a filter is to reduce the effects of the resonant modes of the microactuator and other noises. The switch on the same path is to make sure that the microactuator is only to be activated when the tracking error is within the reach of the microactuator, which has a maximum displacement of 1 μm. Our design philosophy is actually rather simple.

1. Since the maximum displacement of the microactuator is 1 μm, it will not provide much help in the track seeking stage. Thus, when the tracking error is large, *i.e.*, the displacement y is far away from the target, the switch in the microactuator path will be turned off and only the VCM actuator will be solo in action.

2. When the R/W head is entering the target region, *e.g.*, within 1 μm of the target reference, the switch will be turned on and the microactuator will be in action to speed up the response of the overall system.

3. It can be observed from the configuration in Figure 9.4 that y_m, the displacement of the microactuator, will settle down to zero as the tracking error approaches zero. As pointed out earlier, such a feature would enable the microactuator to be used for the next move.

The detailed control parameters are as follows.

1. The estimation gain of the microactuator displacement $K_s = 0.722$.

2. The switch is chosen such that the path is activated when $|r - \hat{y}_v| < 1.2$ μm. Otherwise, it will be open.

3. A second-order Chebyshev low-pass filter is used on the control input path of the microactuator to attenuate the microactuator resonant modes. Its transfer function is given by

$$H_f(s) = \frac{(3000\pi)^2}{s^2 + 4500\pi s + (3000\pi)^2}, \tag{9.5}$$

which has a cut-off frequency at 1500 Hz. This filter is to be discretised using the ZOH method in implementation.

4. The control laws for the VCM actuator and the proportional gain of the microactuator are given as follows:

   a) `MSC`. The first part of the MSC law, *i.e.*, the PTOS control law, is the same as that given in Equations (7.7) and (7.8). The second part of the MSC law, *i.e.*, the RPT controller, is given by

$$\begin{cases} \dot{x}_v = & -8410.8 & x_v - 4.1367\times10^7\,\hat{y}_v + 7.9952\times10^6\,r, \\ u_R = -6.8905\times10^{-5}\,x_v - & 0.4005 & \hat{y}_v + \quad 0.1249 \quad r. \end{cases}$$

   The proportional gain K_m is chosen as:

   i. $K_m = 0.4416$ for SL = 1 and 10 μm;

   ii. $K_m = 0.3450$ for SL = 20 μm.

   The switching conditions remain the same as those in Chapter 7.

   b) `CNF`. The CNF controller is the same as that in Equations (7.20) to (7.22), except y is replaced by $\hat{y}_v$. The proportional gain K_m is chosen as:

i. $K_m = 0.3462$ for SL = 1 and 10 μm;

ii. $K_m = 0.2778$ for SL = 20 μm.

The simulation and implementation results of the dual-stage actuated systems obtained are presented in Section 9.4. The velocity estimator used for the PTOS control law is the same as that in Chapter 7. We note that a complete analysis on the stability of the overall closed-loop system as depicted in Figure 9.4 with the CNF approach can be found in [138].

9.4 Simulation and Implementation Results

We now present the simulation and implementation results of the servo systems obtained in the previous section. The simulation results are done in a continuous-time setting. The implementation results are carried out at a sampling frequency of 10 kHz. The results of the dual-stage actuated HDD servo systems will then be compared with those of the servo systems with a single-stage actuator. The latter are done on the same drive by keeping the microactuator inactive throughout the whole implementation process. The controller parameters for the single-stage actuated systems are identical to those given in Chapter 7. As expected, and as to be seen soon, the dual-stage actuated servo systems do yield a better performance compared with their single-actuated counterparts.

9.4.1 Track Seeking and Following Test

We first test the performance of track seeking and track following of the dual-stage actuated servo system. The simulation and implementation results with various seek lengths, *i.e.*, for SL = 1, 10 and 20 μm, are to be presented. Note that, for larger seek lengths, the performance of the dual-stage actuated servo system will be about the same as that of the single-stage actuated counterpart.

In order to see how the dual-stage actuated system works, we present detailed simulation and implementation results for SL = 1 μm, which show the responses of the R/W head, the VCM actuator and the microactuator, as well as the control signals of two actuators. The contributions of the microactuator to the overall response can also be clearly observed from these results. For larger seek lengths, *i.e.*, SL = 10 and 20 μm, the contributions of the microactuator are almost invisible, and thus it is meaningless to include them in the plots. Nonetheless, results for SL = 10 and 20 μm will then be

given together with those for SL = 1 μm for easy reference. Results of the corresponding single-stage actuated system will also be given for comparison.

MSC. The results for the servo systems with the MSC technique are presented in Figures 9.5 to 9.10. The settling times and the percentages of improvement are summarised in Table 9.1.

Table 9.1. Performances of single- and dual-stage servo systems with MSC.

(a) simulation results

Seek length (μm)	Settling Time (ms)		Overall improvement
	Single	Dual	
1	1.15	1.05	9%
10	3.83	3.65	5%
20	4.43	4.35	2%

(b) experimental results

Seek length (μm)	Settling Time (ms)		Overall improvement
	Single	Dual	
1	1.25	1.1	12%
10	4.15	4.0	4%
20	5.10	4.9	4%

CNF. Similarly, the results of the HDD servo systems with the CNF method are shown in Figures 9.11 to 9.16. The settling times and the percentages of improvement are summarised in Table 9.2.

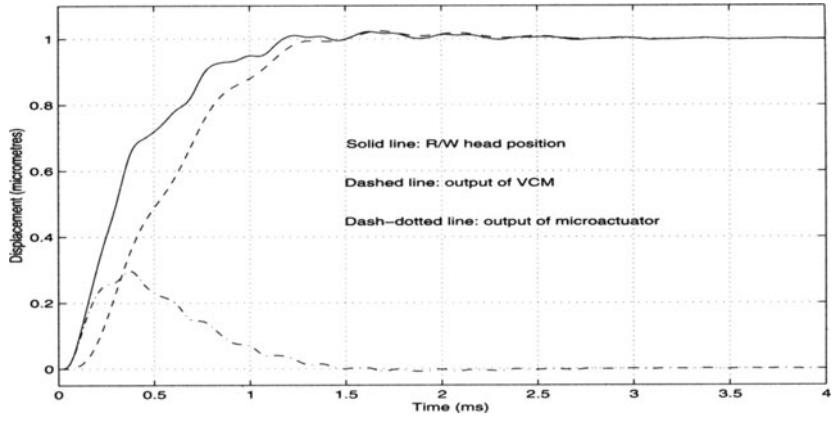
Table 9.2. Performances of single- and dual-stage servo systems with CNF.

(a) simulation results

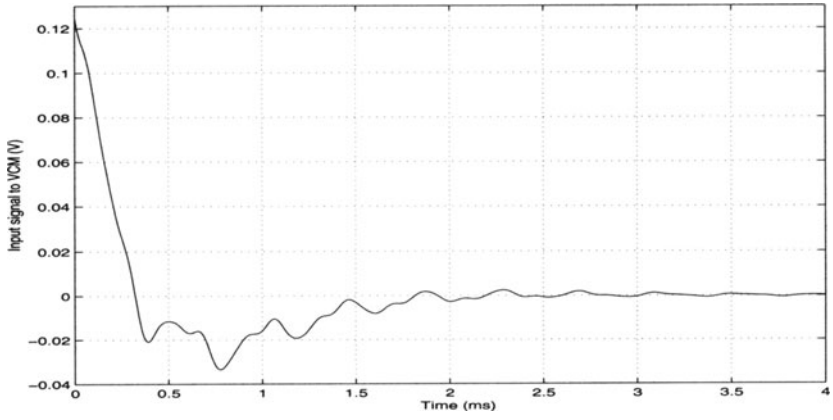
Seek length (μm)	Settling Time (ms)		Overall improvement
	Single	Dual	
1	0.96	0.83	14%
10	2.59	2.43	6%
20	3.25	3.17	2%

(b) experimental results

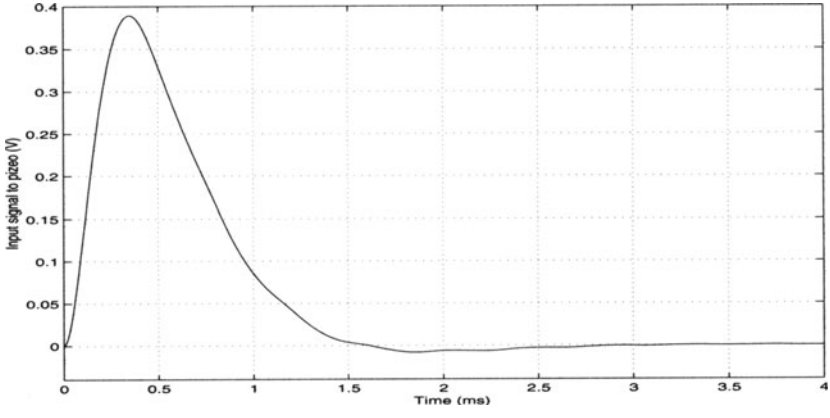
Seek length (μm)	Settling Time (ms)		Overall improvement
	Single	Dual	
1	1.20	0.85	29%
10	3.25	2.80	14%
20	4.13	3.90	6%



(a) output response

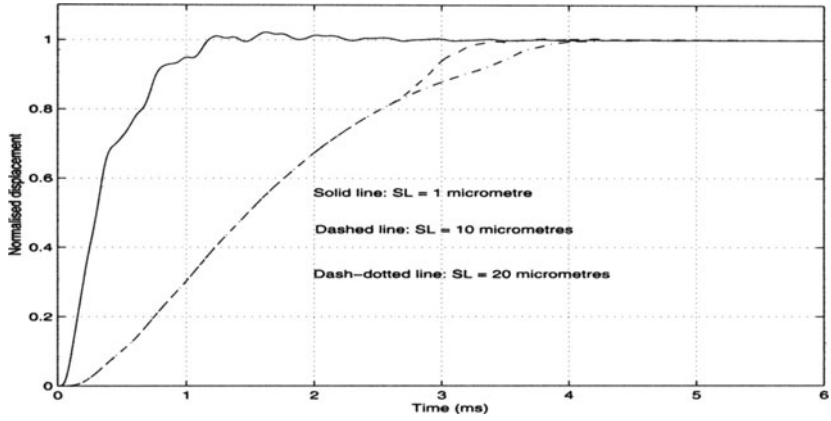


(b) control signal to VCM actuator

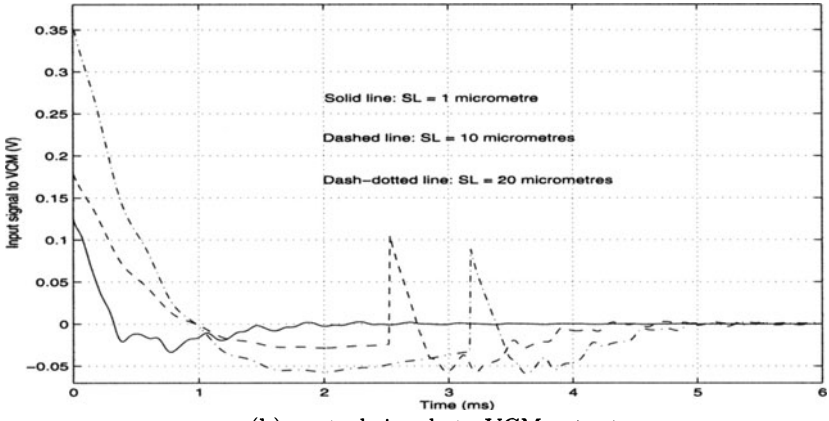


(c) control signal to microactuator

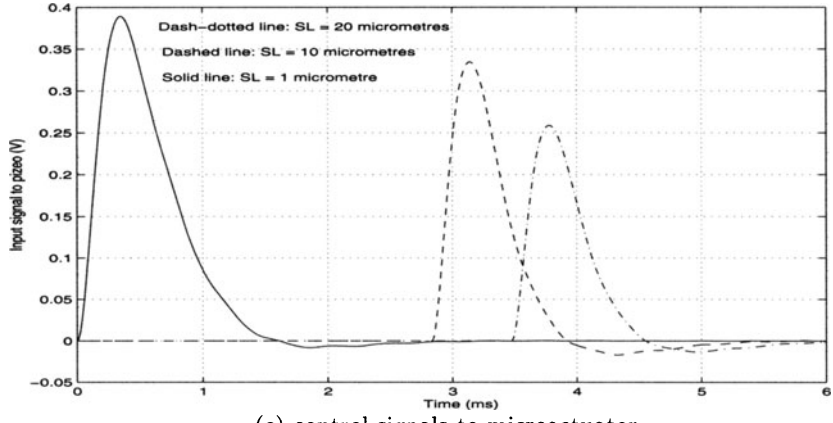
Figure 9.5. Simulation results of dual-stage system (MSC) for SL = 1.



(a) normalised output responses

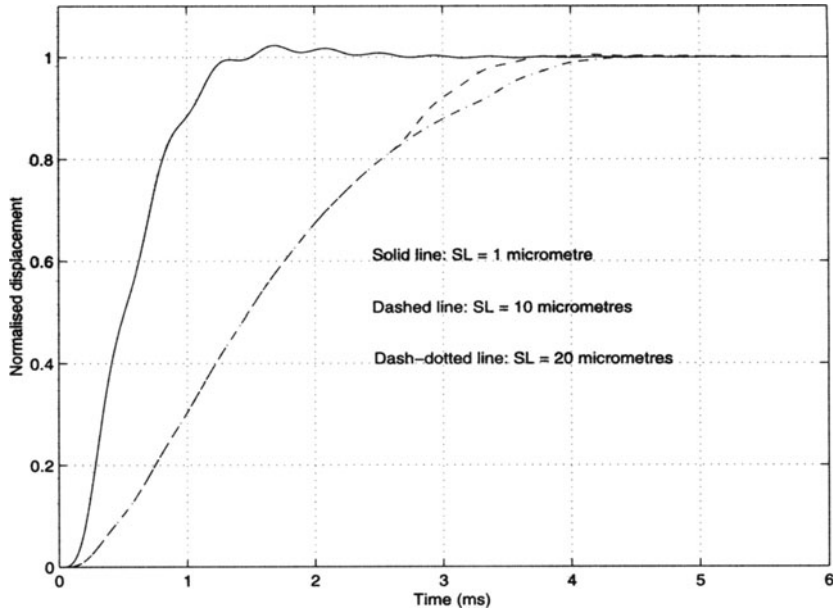


(b) control signals to VCM actuator

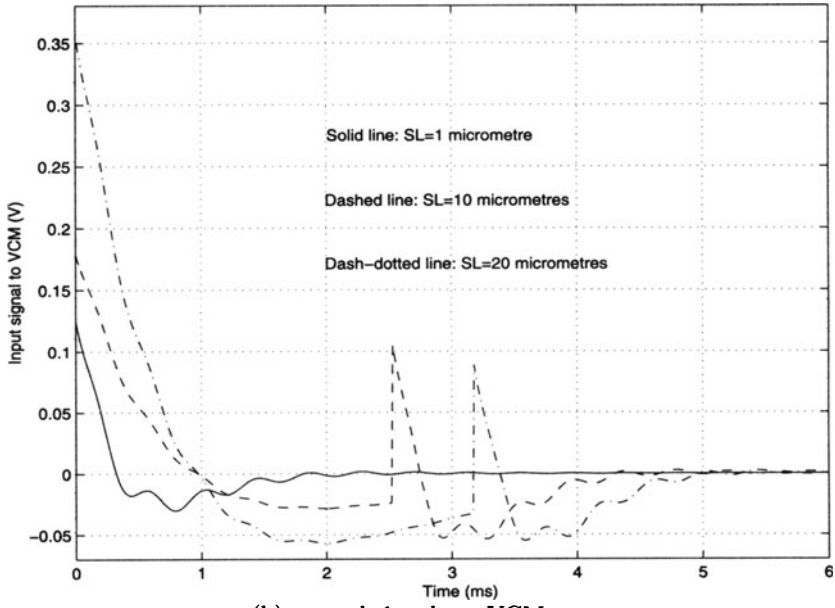


(c) control signals to microactuator

Figure 9.6. Simulation results of dual-stage system (MSC) for SL = 1, 10, 20.

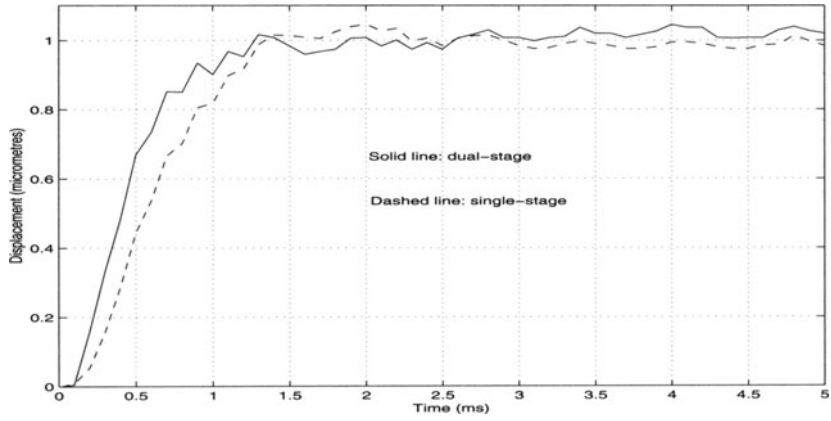


(a) normalised output responses

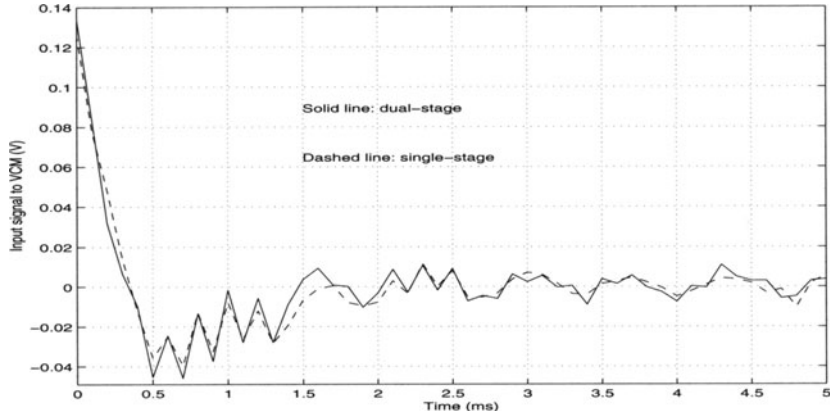


(b) control signals to VCM actuator

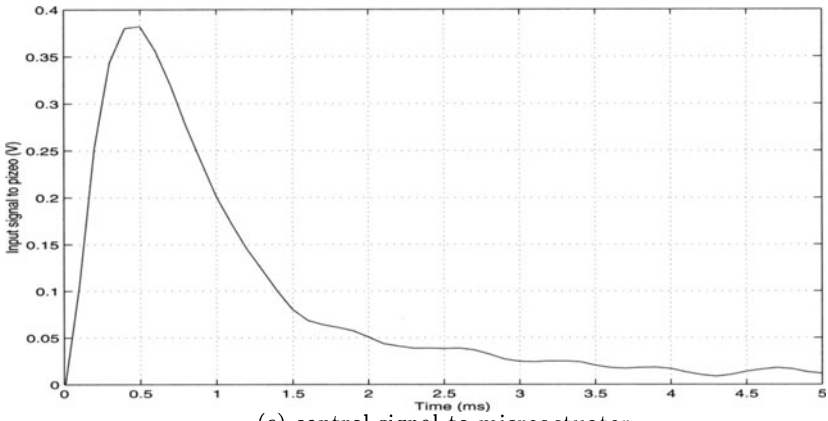
Figure 9.7. Simulation results of single-stage system (MSC) for SL = 1, 10, 20.



(a) output responses

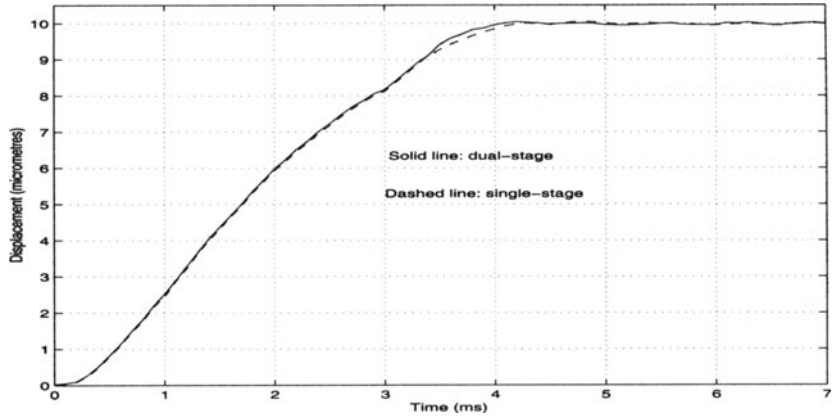


(b) control signals to VCM actuator

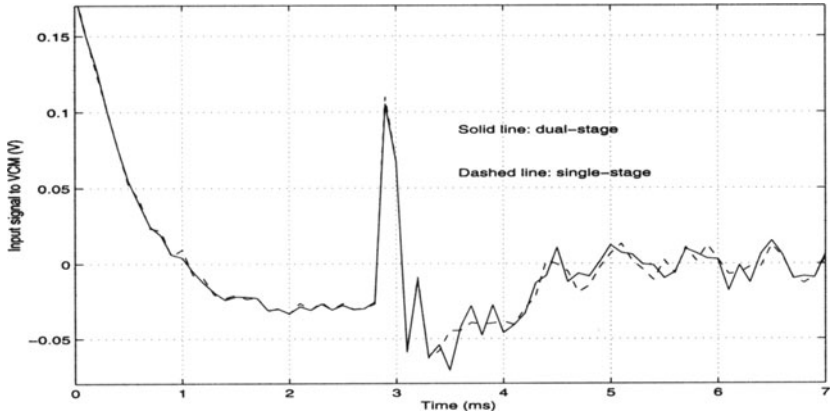


(c) control signal to microactuator

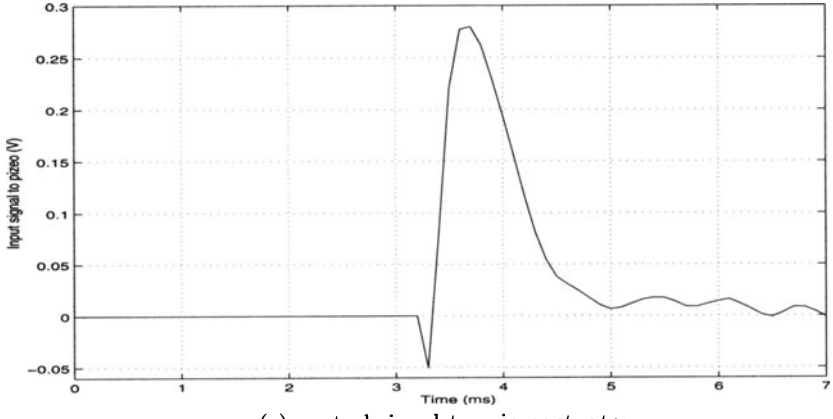
Figure 9.8. Experimental results of single- & dual-stage systems (MSC) (SL=1).



(a) output responses

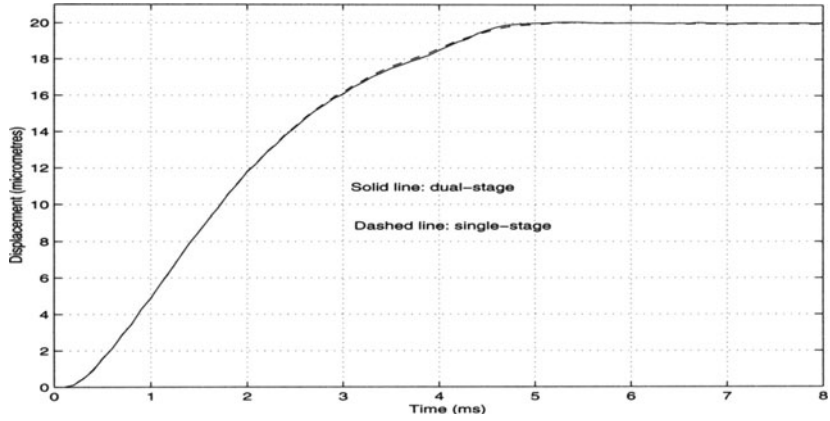


(b) control signals to VCM actuator

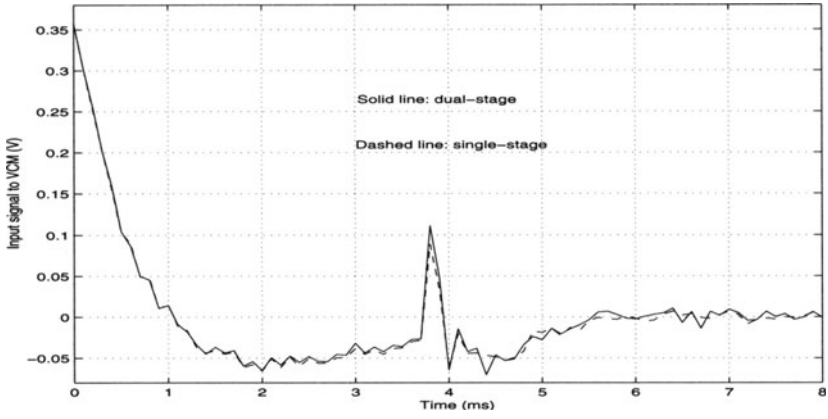


(c) control signal to microactuator

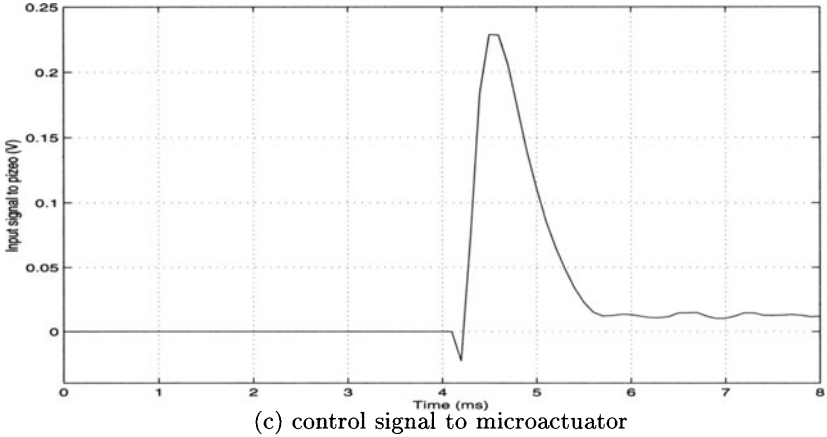
Figure 9.9. Experimental results of single- and dual-stage systems (MSC SL=10).



(a) output responses

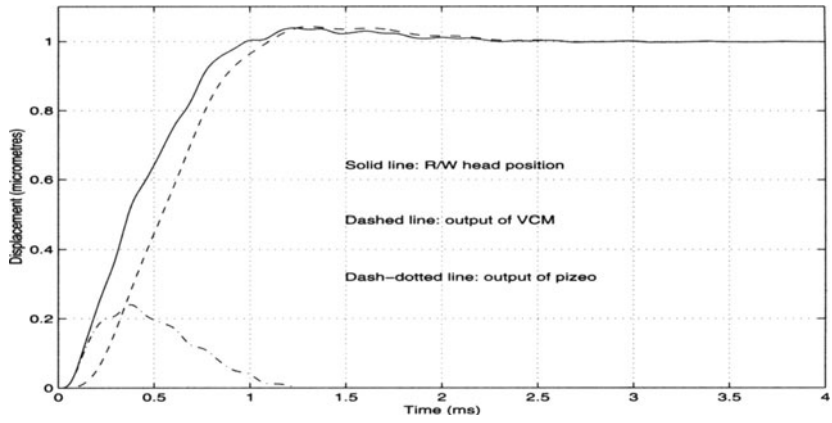


(b) control signals to VCM actuator

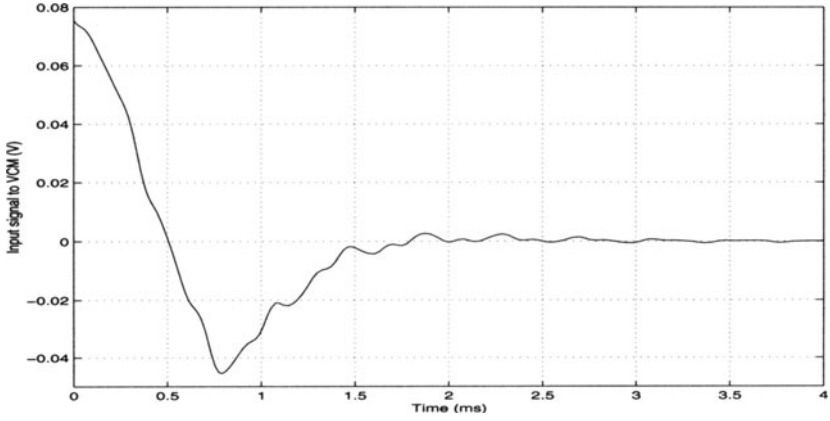


(c) control signal to microactuator

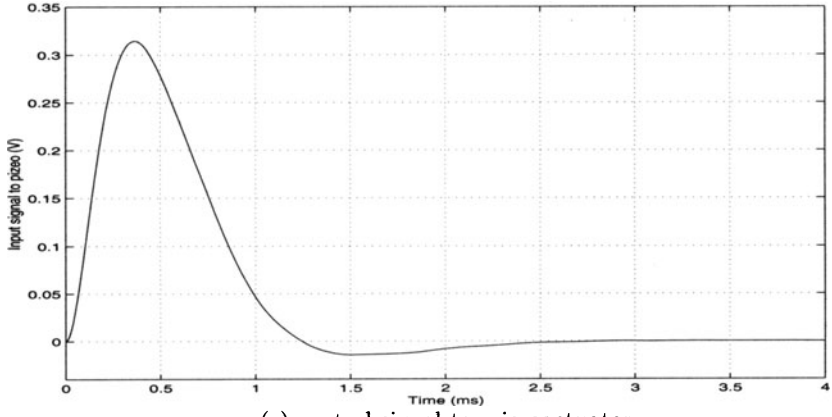
Figure 9.10. Experimental results of single- and dual-stage systems (MSC SL=20).



(a) output response

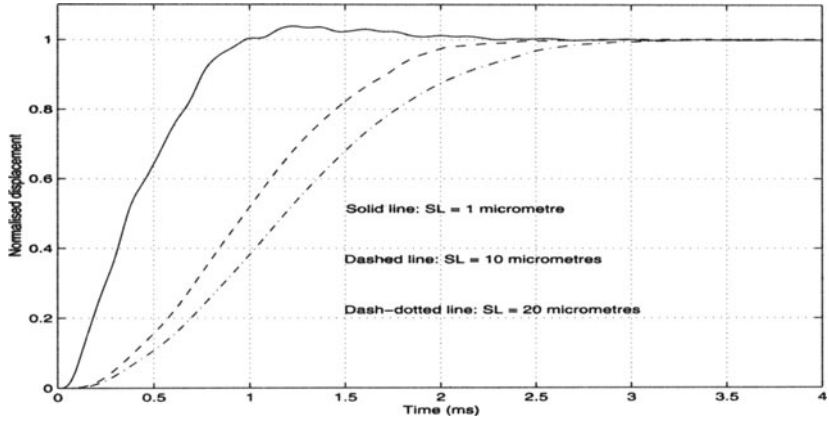


(b) control signal to VCM actuator

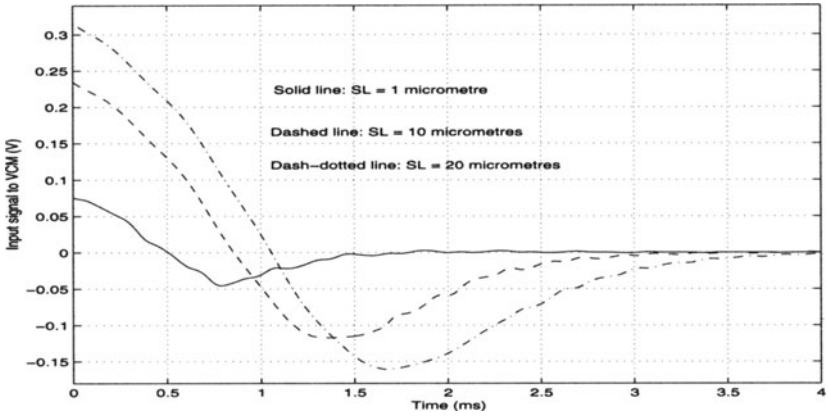


(c) control signal to microactuator

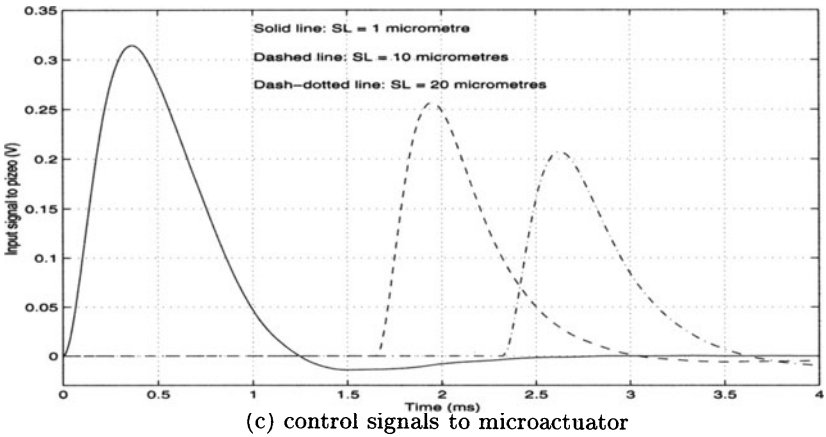
Figure 9.11. Simulation results of dual-stage system (CNF) for SL = 1.



(a) normalised output responses

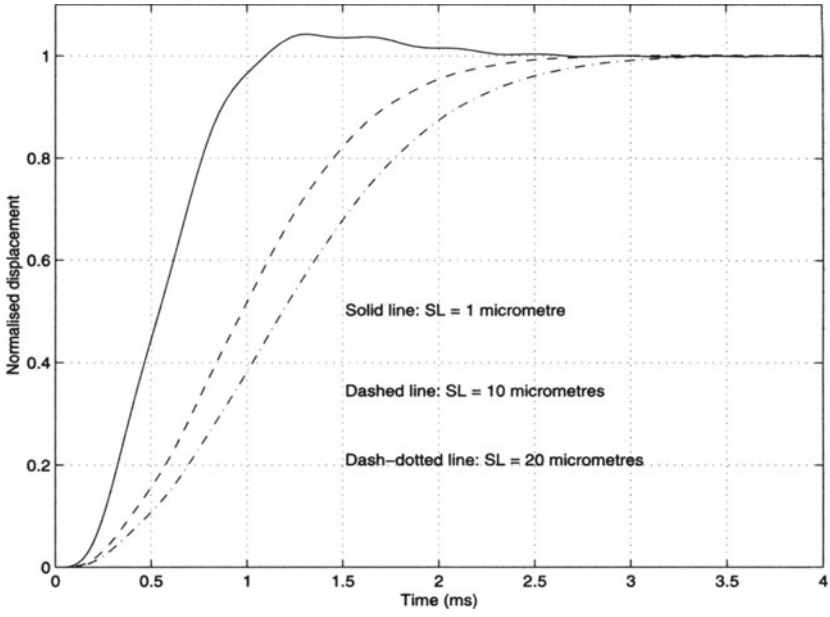


(b) control signals to VCM actuator

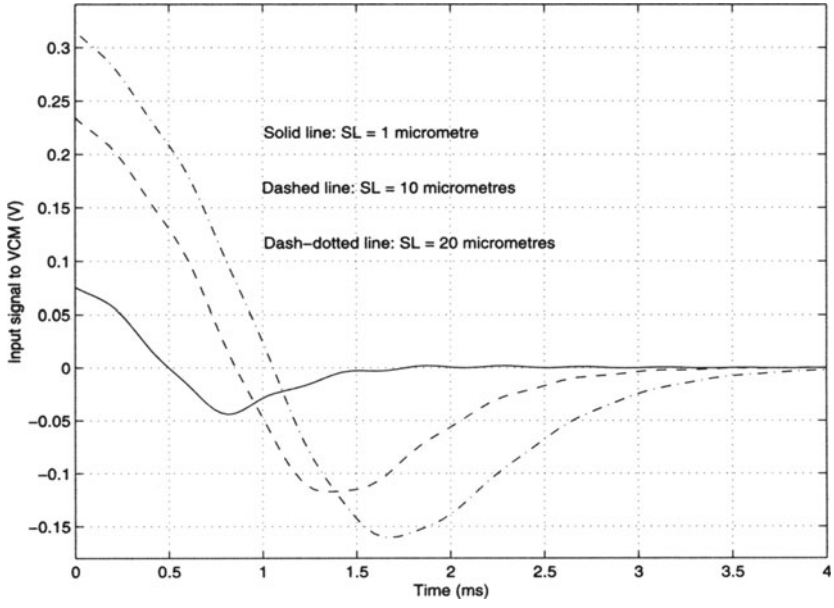


(c) control signals to microactuator

Figure 9.12. Simulation results of dual-stage system (CNF) for SL = 1, 10, 20.

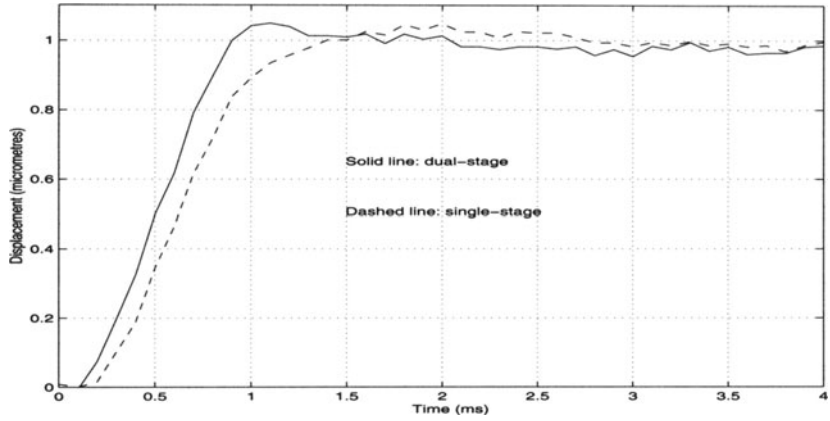


(a) normalised output responses

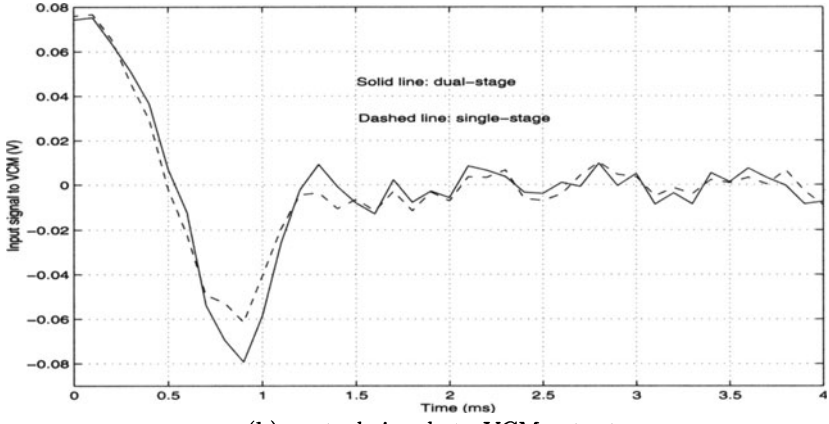


(b) control signals to VCM actuator

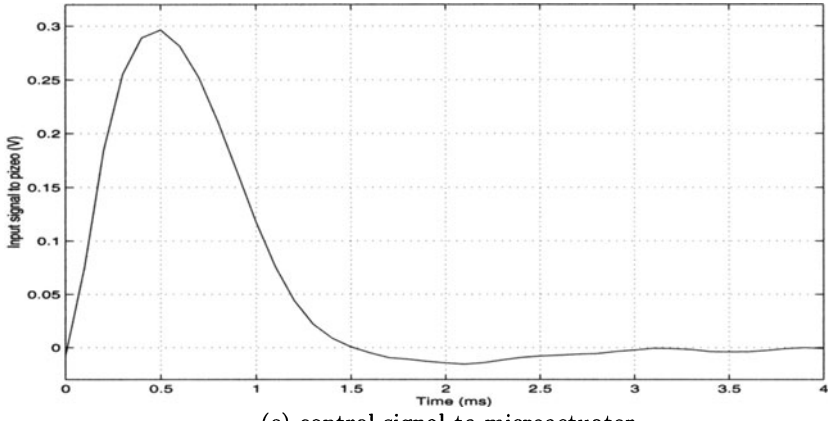
Figure 9.13. Simulation results of single-stage system (CNF) for SL = 1, 10, 20.



(a) output responses

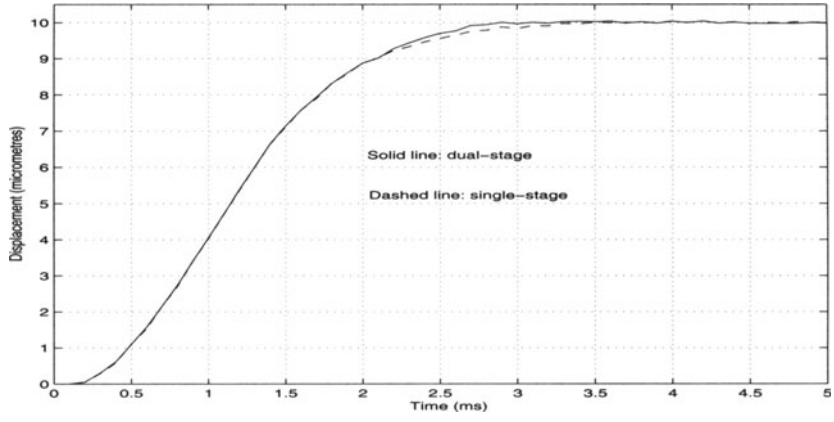


(b) control signals to VCM actuator

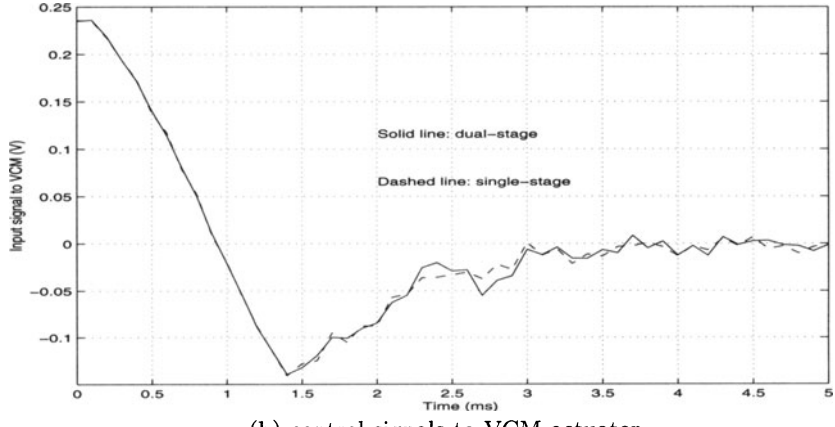


(c) control signal to microactuator

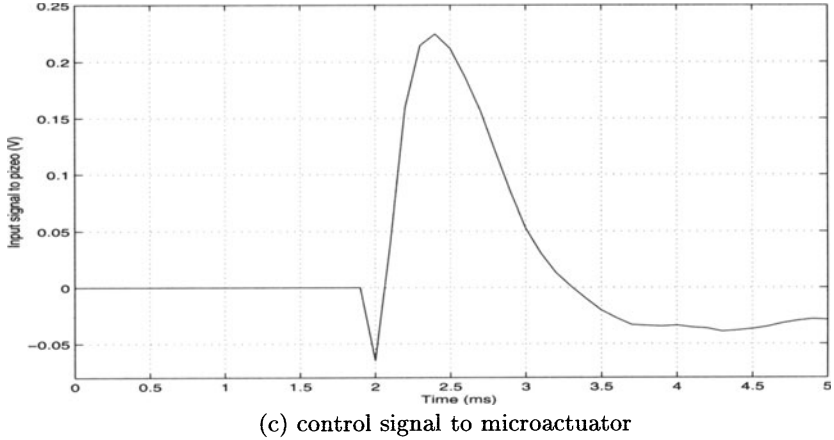
Figure 9.14. Experimental results of single- and dual-stage systems (CNF SL=1).



(a) output responses

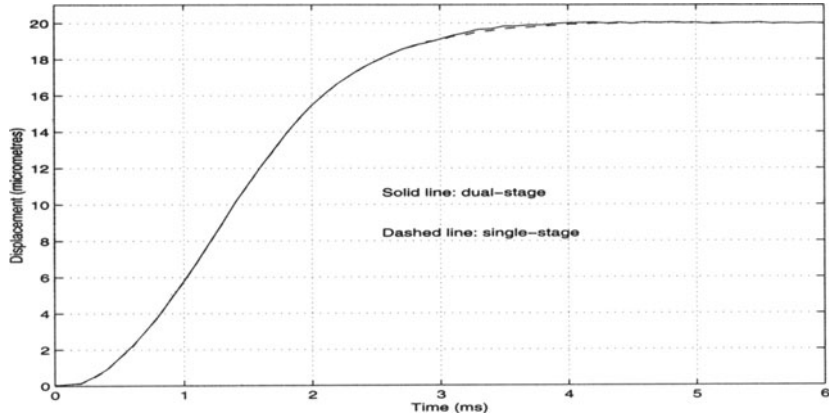


(b) control signals to VCM actuator

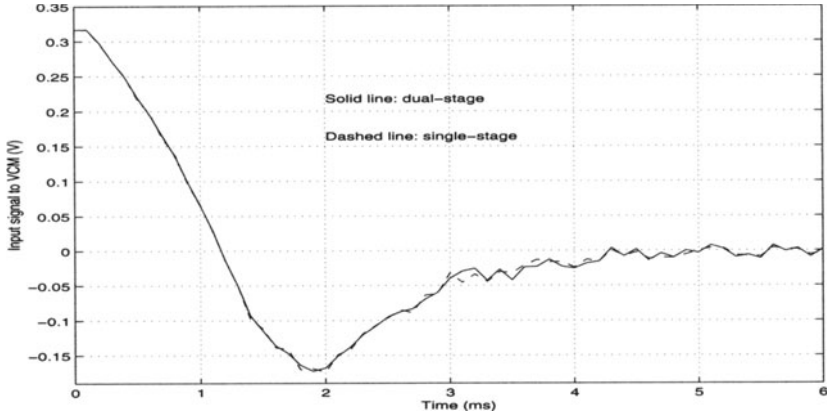


(c) control signal to microactuator

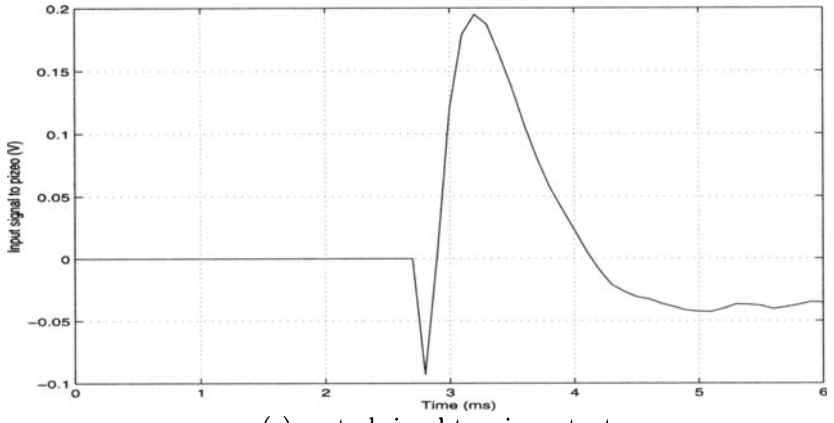
Figure 9.15. Experimental results of single- and dual-stage systems (CNF SL=10).



(a) output responses



(b) control signals to VCM actuator



(c) control signal to microactuator

Figure 9.16. Experimental results of single- and dual-stage systems (CNF SL=20).

9.4.2 Runout Disturbance Test

As in the previous chapters, we artificially add a runout disturbance

$$w(t) = 0.5 + 0.1\cos(100\pi t) + 0.05\sin(220\pi t), \tag{9.6}$$

which is the same as those used in Chapters 6 and 7, to our servo systems and the implementation results of the corresponding responses are respectively shown in Figures 9.17 and 9.18. The standard deviations of the error signals are summarised in Table 9.3.

Table 9.3. Standard deviations of error signals in runout disturbance tests.

MSC		CNF	
Single	Dual	Single	Dual
0.0085	0.0075	0.0083	0.0073

9.4.3 Position Error Signal Test

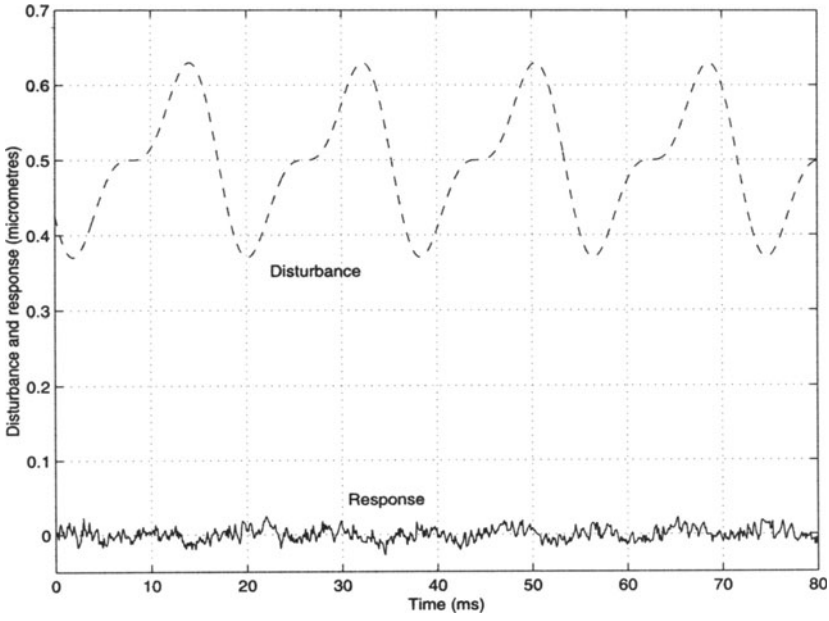
Lastly, as were done in Chapters 6 and 7, we conduct the PES tests for the complete single- and dual-stage actuated servo systems. The results, *i.e.*, the histograms of the PES tests, are given in Figures 9.19 and 9.20. The $3\sigma_{\text{pes}}$ values of the PES tests, which are a measure of track mis-registration (TMR) in HDDs and which are closely related to the maximum achievable track density, are summarised in Table 9.4.

Table 9.4. The $3\sigma_{\text{pes}}$ values of the PES tests.

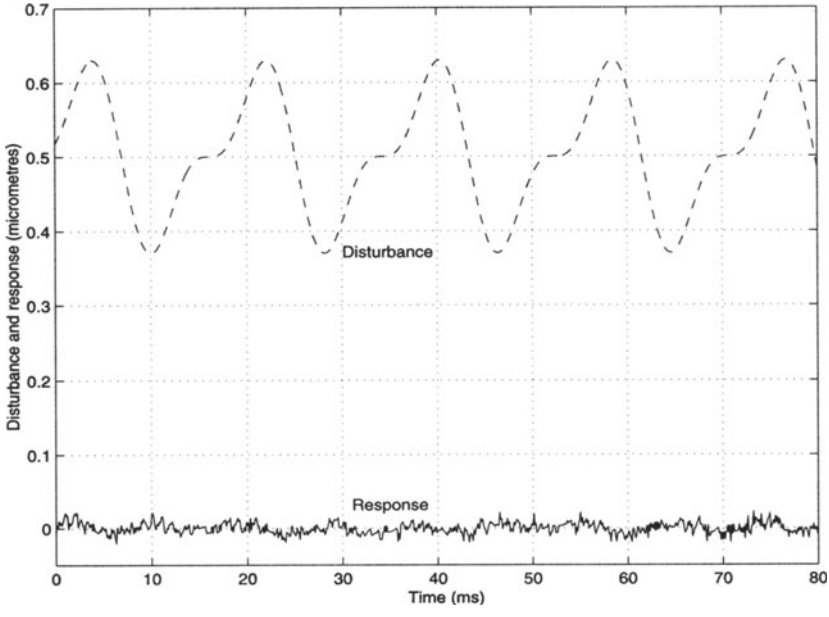
MSC		CNF	
Single	Dual	Single	Dual
0.0255	0.0225	0.0249	0.0219

9.5 Discussion

It can be easily observed from the results in Section 9.4 that the dual-stage actuated servo systems do provide a faster settling time compared with that

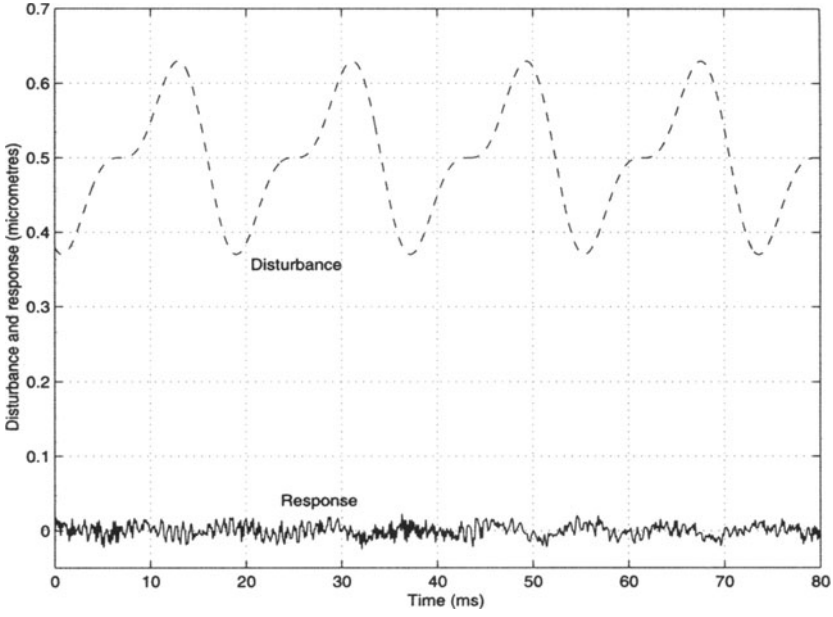


(a) single-stage system

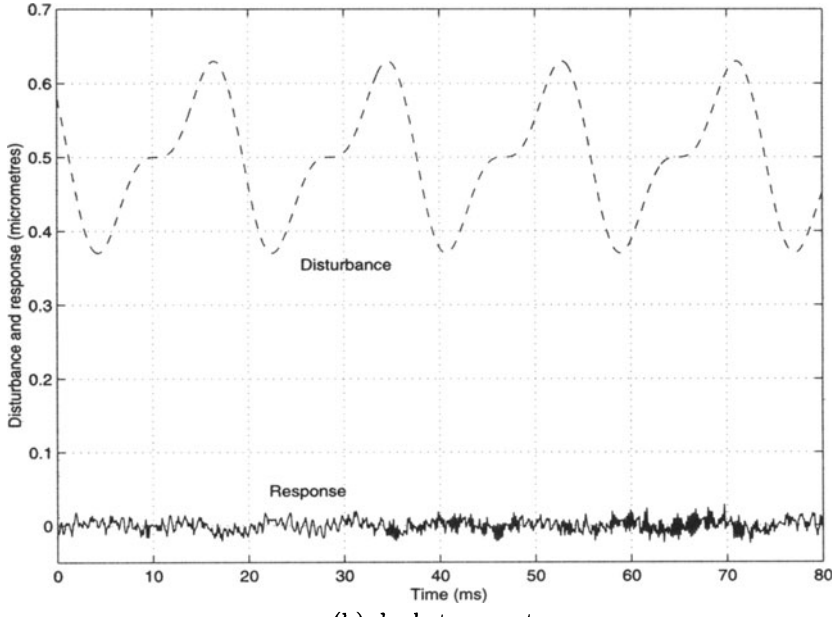


(b) dual-stage system

Figure 9.17. Implementation results. Responses to a runout disturbance (MSC).

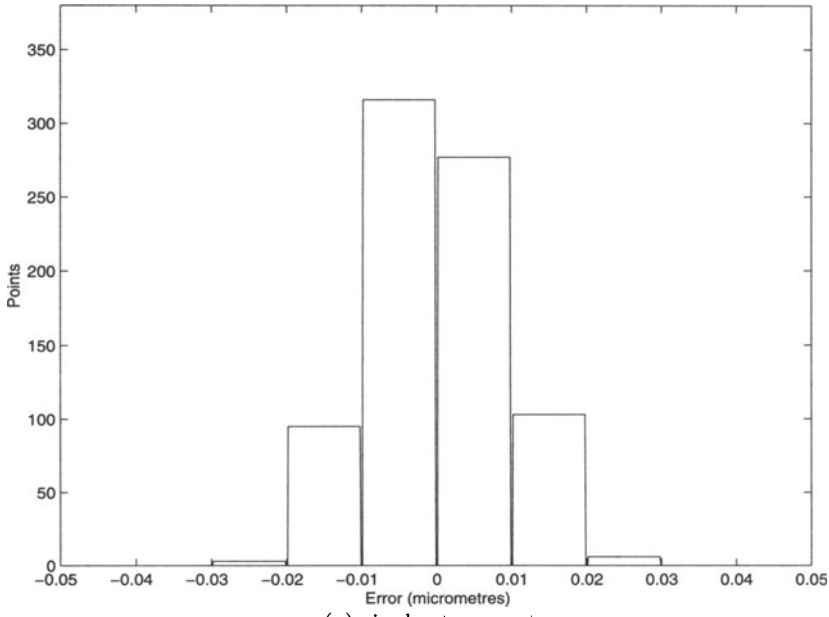


(a) single-stage system

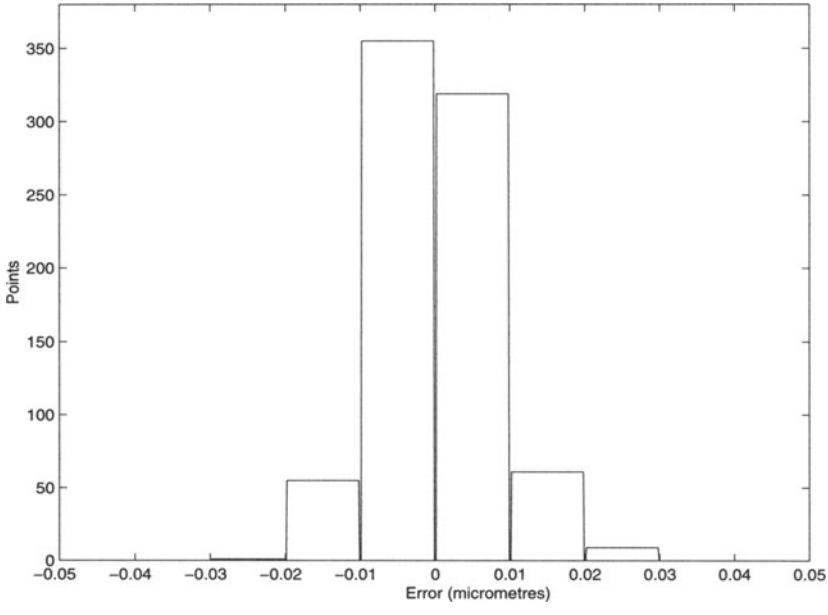


(b) dual-stage system

Figure 9.18. Implementation results. Responses to a runout disturbance (CNF).

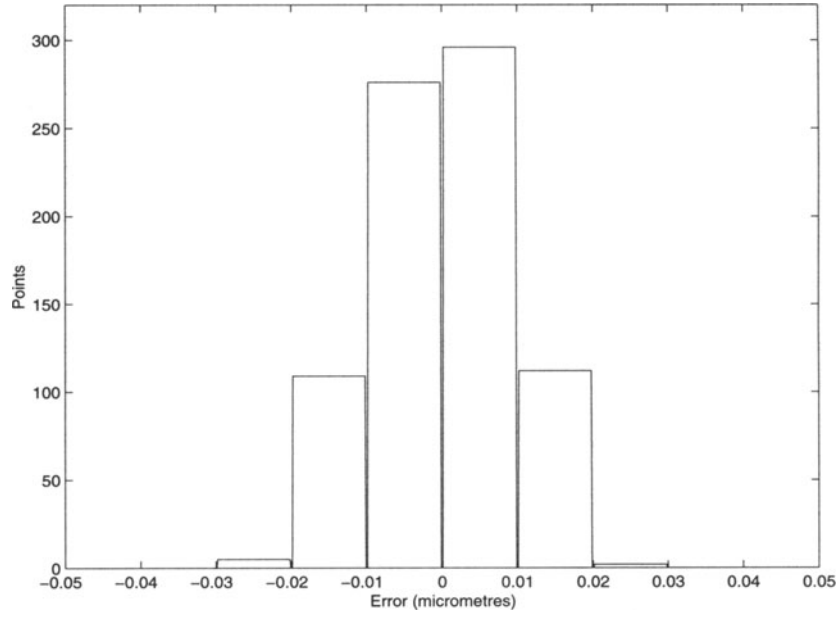


(a) single-stage system

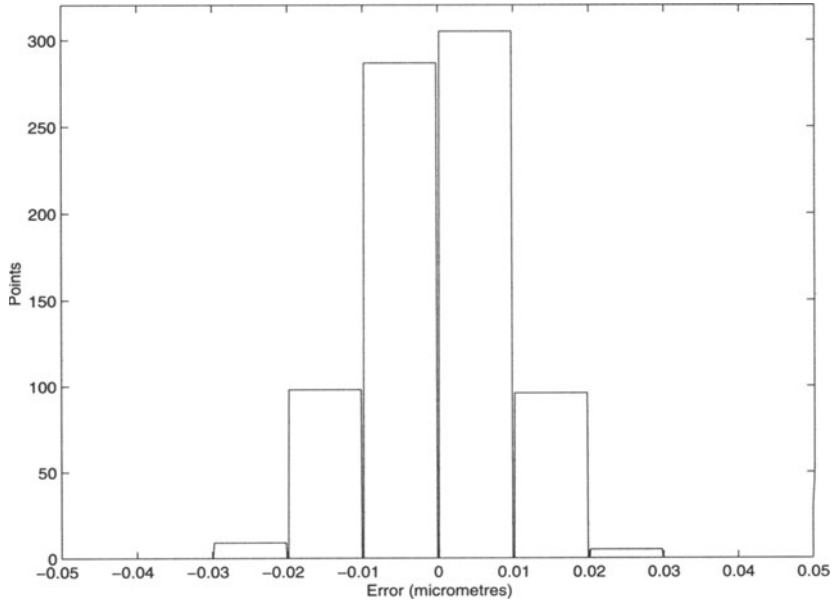


(b) dual-stage system

Figure 9.19. Implementation results. PES test histograms (MSC).



(a) single-stage system



(b) dual-stage system

Figure 9.20. Implementation results. PES test histograms (CNF).

of the single-stage actuated counterparts. The improvement in the track following stage turns out to be very noticeable. This was actually the original purpose of introducing the microactuator to HDD servo systems. However, we personally feel that the price we have paid (*i.e.*, by adding an expensive and delicate piezoelectric actuator to the system) for such an improvement is too high. Besides this, we believe that the following improvements on the physical properties of the microactuator are needed before it can be considered for use in commercial HDDs.

1. There should be a significant improvement on the bandwidth of the piezoelectric actuator. The resonant modes of the piezoelectric actuator used here are simply not high enough. Much more can be done if these resonant modes can be made further away at higher frequencies.

2. The maximum displacement of the piezoelectric actuator should be made as large as possible. This would allow us to design a more efficient servo system in which the microactuator would play a more active role.

Nonetheless, these issues are beyond our interest and control. Somehow, we feel that there is some way to go before the dual-stage actuated hard drives can be pushed to the market. It would perhaps be more efficient to just focus on improving the properties of the VCM actuators instead.

CHAPTER 10
RESONANCE AND DISTURBANCE REJECTION

10.1 Introduction

The prevalent trend in hard disk design is towards smaller hard disks with increasingly larger capacities. This implies that the track width has to be smaller, leading to lower error tolerance in the positioning of the head. The controller for track following has to achieve tighter regulation in the control of the servomechanism. To read (or write) the data reliably from (or to) the disk, the absolute track following error with respect to the target track centre, which is commonly called track mis-registration (TMR), must be less than 10% of the track pitch. For example, for a 3.5″ HDD with 25 kTPI, the track pitch is about 1 μm and its TMR must be less than 0.1 μm. Thus, for 70 kTPI, TMR must be less than 0.036 μm. This requires rigorous analysis of the sources of TMR and development of advanced techniques to overcome or eliminate these sources to meet the increasing demand for higher TPI. Figure 10.1 shows a typical disk drive servo channel indicating the various sources of disturbances and errors. Some of the larger components of TMR are due to the following error sources, listed roughly in the order of impact (see *e.g.*, [139] and references cited therein):

1. external shock and vibrations present in portable devices;
2. TMR caused by bearing hysteresis and poor velocity estimates during track settling mode;
3. servo pattern nonlinearities and inaccuracies caused by head, media, and servo writing effects;
4. mechanical resonances in suspension, actuator, disk, and housing;
5. electronic noise in recording channel entering the servo demodulator;
6. nonrepeatable spindle runout caused by bearings;
7. variations in RRO caused by thermal and other drifts.

In our opinion, it would be ideal to design an HDD servo system that has the desired track seeking and following performance and at the same time is robust enough to overcome all the above listed disturbances and uncertainties. Nonetheless, this would be a subject for our future research. The main purpose of this chapter is to give a brief survey on how these issues are tackled in the literature.

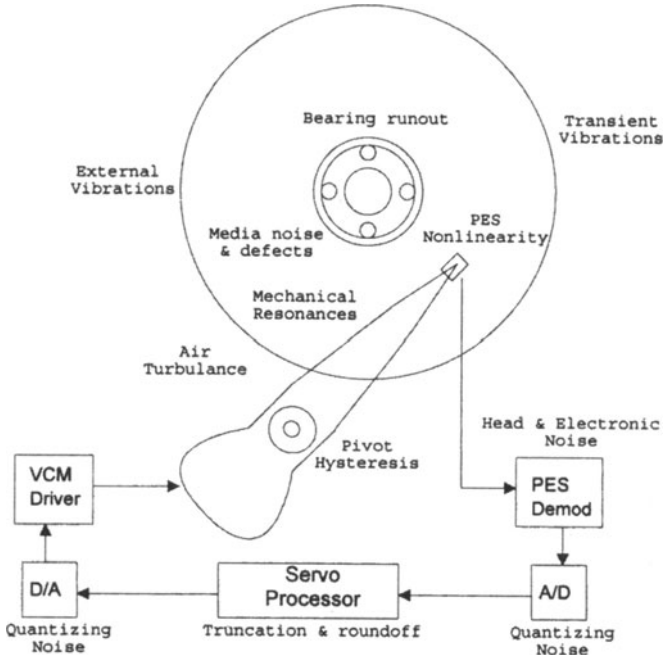


Figure 10.1. Sources of error in an HDD servo system.

10.2 Disturbance Rejection

As discussed earlier, higher TPI requires a tighter TMR, which is formally defined as three times the position error variance of the true PES, *i.e.*, $3\sigma_{\text{pes}}$, plus some terms due to the settling process. The sources of disturbances, which are the error sources contributing to $3\sigma_{\text{pes}}$ can be classified into three categories: input disturbance, output disturbance and measurement noise. The input disturbance is typically a colour noise due to flexure, an electronic bias superimposed with selective energy aring from the natural frequencies of the various mechanical perturbations such as resonances, vibrations and

friction. The output disturbance is also a colour noise due to spindle rotation and its effects such as runout, windage and media noise. The measurement noise is a typical white noise due to the position measurement techniques and/or sensors. These disturbances and noise can be modelled as in Figure 10.2. The objective would be to reject the effect of the disturbances and the measurement noise and achieve minimum position error variance.

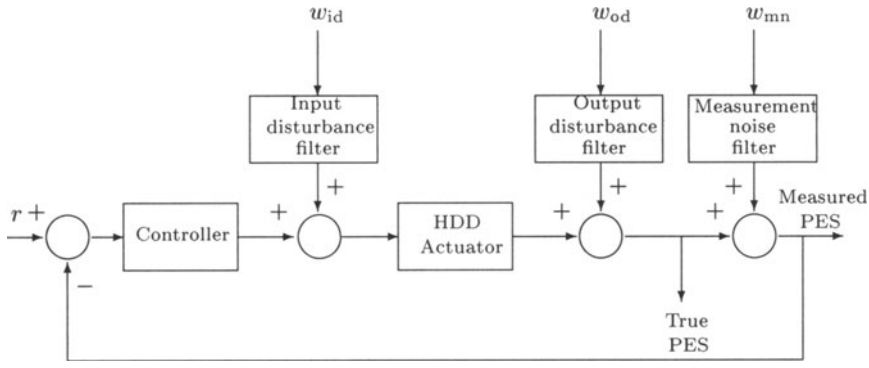


Figure 10.2. Modeling of disturbances in an HDD servo system.

It has been shown [140] that, under some simplifications, σ_{pes} can be approximated as

$$\sigma_{\text{pes}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} y_{\text{pes}}^2(i)}. \tag{10.1}$$

Recently, Li *et al.* [141] gave a solution that minimises σ_{pes} by converting the minimisation of σ_{pes} into an equivalent H_2 optimal control problem for an auxiliary system, which contains the dynamics of the actuator, the input and output disturbances and the measurement noise, and which has a properly defined disturbance input and an output to be controlled. The problem to minimise σ_{pes} can then be solved using any appropriate H_2 optimal control methods (see *e.g.*, Chapter 4).

Next, to understand runout disturbances, we recall the two main functions of HDD servo systems, *i.e.*, track seeking and track following, which are usually achieved by two different controllers. The track seeking controller moves the R/W head to the target track in minimum time; after this, when the control is switched to the track following controller, it must make the R/W head follow the target track and keep the errors as small as possible. Thus, all HDDs must have a position measurement mechanism. The position feedback signals in most HDD servomechanisms are derived through prerecorded

position information recorded on one side of a disk surface at the time of manufacture using a servo writer. Ideally, servo tracks are perfect concentric circles. However, in the process of servo writing, the head that writes the signals cannot be kept perfectly still, due to *e.g.*, the presence of vibration and NRRO effects, which result in tracks that are not perfect circles. This apparent track motion will cause the R/W head to move in an attempt to minimise the position error, which results in positioning of the R/W head away from the real data track. Such an imperfection is termed a *runout*. This runout can depending upon its nature be classified as repeatable and nonrepeatable. In what follows we will discuss these two types of runout and the methods available to compensate these disturbances collected from a literature survey.

10.2.1 Repeatable Runout

When the sampling frequency is equal to the spindle rotation frequency, or one of its order, the runout motion produced by the apparent track is repeated. This repeated runout, which is locked to the spindle rotation in both frequency and phase, is what we call a RRO. Thus, the major source of RROs is the eccentricity of the track. Other sources include the offset of the track centre with respect to the spindle centre, bearing geometry and wear, and motor geometry [142]. RROs caused by factors other than the eccentricity would cause a large amount of RROs at the rotational frequency of the spindle or its order, which is common to all tracks.

An RRO is a repetitive event in that both its amplitude and phase are locked to the rotation of the spindle. Therefore, this prior knowledge of an RRO can be used as a feed-forward signal to compensate the tracking error. Roughly, there exist three approaches to reject RRO: (1) repetitive control, (2) feedback control based on the internal model principle, and (3) identification and feed-forward control. Many versions of above compensation techniques for RROs have been reported, and some include adaptive feed-forward cancellation and repetitive control (see *e.g.*, [142–145]), PID with repetitive control (see *e.g.*, [146]) and recurrent neural networks (see *e.g.*, [147]). To be a bit more specific, assuming that the RRO, $d(t)$, is a time-varying unknown disturbance consisting of a sum of n sinusoids of known frequencies, *i.e.*,

$$d(t) = \sum_{i=1}^{n} \Big[a_i(t) \cos(\omega_i t) + b_i(t) \sin(\omega_i t) \Big], \tag{10.2}$$

the adaptive feed-forward compensation approach is to design a control

$$u(t) = \hat{d}(t) = \sum_{i=1}^{n} \Big[\hat{a}_i(t) \cos(\omega_i t) + \hat{b}_i(t) \sin(\omega_i t) + g_i k_i y(t) \Big]. \tag{10.3}$$

Next, an appropriate adaptive algorithm is used to adjust the estimates $\hat{a}_i(t)$ and $\hat{b}_i(t)$ so that these estimates are made equal to the nominal values, *i.e.*,

$$\hat{a}_i(t) = a_i(t), \quad \hat{b}_i(t) = b_i(t). \tag{10.4}$$

The RRO disturbance approximated as in Equation (10.2) can then be cancelled by the reproduced signal. This technique is also demonstrated by using neural networks in [147].

10.2.2 Nonrepeatable Runout

NRROs are a product of disk drive vibration and electrical noise in the measurement channel (see *e.g.*, [1]). More specifically, the causes of NRRO are spindle bearing defects, windage-induced disk flutter, electronics noise in the measurement channel, *etc.*, present during servo track writing. NRROs can be minimised via improved servo writing, use of better bearings, and improved design of electronics. Since RROs are the harmonics of the motor rotational frequency in the frequency domain, an NRRO is the subtraction of the harmonics from the total indicated runout (TIR), which can be defined as the distance difference between the R/W head and the previously written track in an HDD, and hence an NRRO in the time domain can be easily constructed by the inverse Fourier transform of an NRRO in the frequency domain [148]. Alternatively, for a better understanding of the time trends of an NRRO, specifically a motor NRRO, Ohmi [149] proposed subtracting the averaged RRO from the experimentally obtained radial vibration of a rotor, the TIR. That is, an NRRO can be derived as follows [149]:

$$\text{RRO}\,(j) = \frac{1}{N}\sum_{i=1}^{N} \text{TIR}\,(i,j), \quad j = 1,2,\cdots,M, \tag{10.5}$$

$$\text{NRRO}\,(i,j) = \text{TIR}\,(i,j) - \text{RRO}\,(j), \quad j = 1,2,\cdots,M, \tag{10.6}$$

where N is the number of revolutions of the rotor, M is the number of samples per revolution, i represents the number of disk revolutions and j represents the number of phases from a fixed point of the slit. An NRRO can be taken care of by the servo controller through improved loop bandwidth. However, the increase in servo bandwidth required to reject an NRRO is mainly determined by three factors, *i.e.*, the servo sampling rate, the spectrum of the measurement noise, and the existence of plant resonance modes. The effect of resonance modes and their compensation will be discussed in the following section. Recent research suggests the use of improved mechanical design, with a damped disk substrate and fluid bearing spindles [150], which provides less stress on the servo loop, to reject NRROs.

10.3 Resonance Compensation

The actuator and HDD structures, of course, are not perfectly rigid and have hundreds of flexible modes. This flexibility gives rise to vibrations, which results in a longer time to settle at the target track and amounts to a significant component of the TMR. The servo bandwidth of modern disk drives is approaching 2 kHz, and it has been proved in the literature that resonance modes that exist within a decade away from the servo crossover frequency degrade the system performance. In short, in modern disk drives, resonance or vibration modes are the major sources of NRROs. Each resonance mode can be modelled as a second-order transfer function. The VCM actuator transfer function displaying multiple resonance modes can be modelled as [139],

$$G(s) = \frac{k_t}{Js^2} \sum_{i=1}^{n} \frac{\omega_i^2}{s^2 + 2\zeta_i \omega_i s + \omega_i^2}, \tag{10.7}$$

where k_t is torque constant, J is inertia, ζ_i and ω_i are respectively the damping ratio and the natural frequency of the ith resonance mode. For simplicity, the frequency response characteristics with a single resonance mode in actuator dynamics are shown in Figure 10.3, in which the characteristics without resonance mode is shown by dashed lines. There is a significant difference in phase angle of the transfer function with the resonance mode. This tends to cause a loss of gain margin in the compensated loop and hence reduces its stability. Although there are hundreds of such resonances in an actual disk drive, many of the characteristics can be defined by considering only three or four modes, as other modes have an insignificant amplitude or are of too high a frequency to be of interest [1]. Some of the important resonance modes that must be considered in the design of high-density disk drives are the quasi-rigid body mode, the pivot bearing, the lateral elastic bending mode, and the vibrations of the individual disk platters.

Currently, the resonances of an HDD head actuator assembly caused by the pivot bearing have become a critical issue, since these resonances have been found to be the major design factor limiting the higher servo control bandwidth [151]. These resonances are excited during the track seeking mode, and when the control is transfered to the track following mode these vibrations result in an increased settling time. Recently, Mah *et al.* [152] have developed a novel moving-coil head actuator, which is designed deliberately to make sure that the force acting on the VCM is an orthogonal force so that there is no resulting force acting on the pivot bearing, thereby minimising residual vibrations. The rotational speeds of modern disk drives are progressively increasing, and hence the effect of the vibration of individual disk platters at their natural frequencies is a significant contributing factor to TMR in high-density disk drives. These resonances are driven primarily by internal

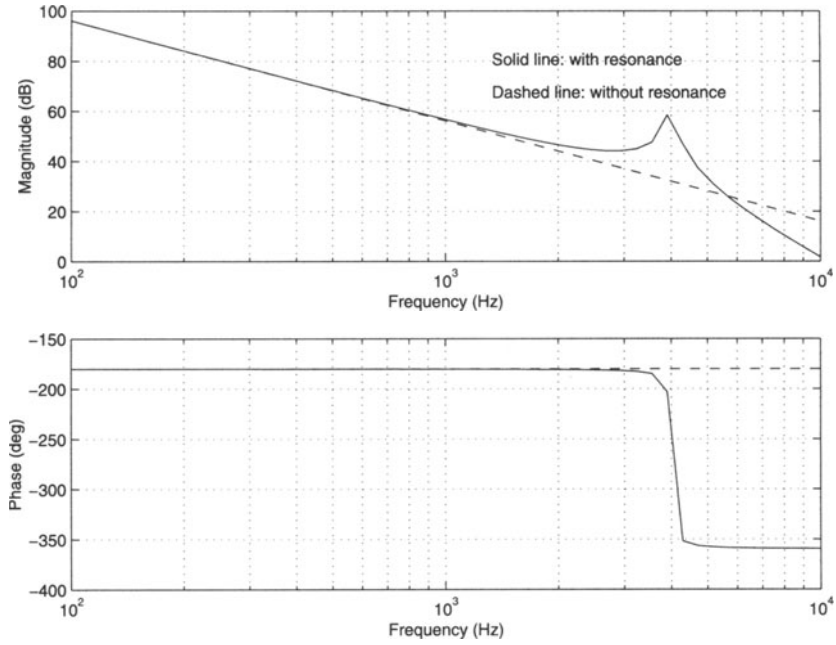


Figure 10.3. The ideal and "actual" frequency responses of HDD actuators.

windage excitation, and its behaviour is dominated by the disk material properties and geometry, and not by the spindle, encloser, or structural design (see *e.g.*, [153]). Use of alternate disk substrate materials can control these effects. Structural resonance modes can be compensated by using a notch filter as a precompensator. Since almost all structural resonant modes have lightly damped poles, the idea is to cancel lightly damped poles and place a pair of well-damped poles instead by using a notch filter. Hanselmann and Mortix [154] proposed the use of three notch filters to suppress the plant model resonance modes. These filters are preferred instead of low-pass filters because the sharper the cut-off in the magnitude frequency response, the lower the phase introduced in the loop. The transfer function of an analog notch filter is commonly chosen as (see *e.g.*, [155])

$$G_{\mathrm{n}}(s) = \frac{s^2 + (2\pi f_0)^2}{s^2 + \frac{2\pi f_0}{Q} s + (2\pi f_0)^2}, \tag{10.8}$$

where f_0 is the centre frequency and Q is the Q-factor. These notch filters can be realised by using switched capacitance filters. To use with digital control, digital notch filers can be realised using microprocessors or high-speed DSPs. Weaver and Ehrlich [156] proposed the use of multirate filters to eliminate the resonance modes beyond the Nyquist frequency.

Since VCM actuators have three or four resonance modes between 1 and 10 kHz, alternate solutions currently being investigated include the use of a secondary actuator, such as a piezoelectric actuator, *i.e.*, the use of dual-stage actuated systems, which have already been discussed in previous chapters, *e.g.*, Chapters 8 and 9.

Finally, we would like to note that it is possible to design a simple servo system that is capable of rejecting disturbances, such as the RROs and NRROs, and compensating the resonance modes, while also giving a good track seeking and following performance. In fact, the results presented in the previous chapters have confirmed this, although it is not addressed in a systematic fashion. Our future research, as mentioned earlier, is to concentrate on deriving a complete treatment of all the issues discussed in this very last chapter.

REFERENCES

1. Franklin GF, Powell JD, Workman ML. Digital control of dynamic systems. 3rd ed. Reading (MA): Addison-Wesley; 1998.

2. Fujimoto H, Hori Y, Yarnaguchi T, Nakagawa S. Proposal of seeking control of hard disk drives based on perfect tracking control using multirate feedforward control. Proceedings of the 6th International Workshop on Advanced Motion Control; Nagoya, Japan; 2000. p. 74-9.

3. Goh TB, Li Z, Chen BM, Lee TH, Huang T. Design and implementation of a hard disk drive servo system using robust and perfect tracking approach. IEEE Trans Contr Syst Technol 2001; 9:221-33.

4. Gu Y, Tomizuka M. Digital redesign and multi-rate control for motion control – a general approach and application to hard disk drive servo system. Proceedings of the 6th International Workshop on Advanced Motion Control; Nagoya, Japan; 2000. p. 246-51.

5. Hara S, Hara T, Yi L, Tomizuka M. Two degree-of-freedom controllers for hard disk drives with novel reference signal generation. Proceedings of American Control Conference; San Diego, California; 1999. p. 4132-6.

6. Huang Y, Messner WC, Steele J. Feed-forward algorithms for time-optimal settling of hard disk drive servo systems. Proceedings of the 23rd International Conference on Industrial Electronics Control and Instrumentation; New Orleans, Louisiana; 1997. p. 52-7.

7. Ho HT. Fast bang-bang servo control. IEEE Trans Magn 1997; 33:4522-7.

8. Ishikawa J, Tomizuka M. A novel add-on compensator for cancellation of pivot nonlinearities in hard disk drives. IEEE Trans on Magn 1998; 34:1895-7.

9. Iwashiro M, Yatsu M, Suzuki H. Time optimal track-to-track seek control by model following deadbeat control. IEEE Trans Magn 1999; 35:904-9.

10. Pao LY, Franklin GF. Proximate time-optimal control of third-order servomechanisms. IEEE Trans Automat Contr 1993; 38:560-80.

11. Patten WN, Wu HC, White L. A minimum time seek controller for a disk drive. IEEE Trans Magn 1995; 31:2380-7.

12. Takakura S. Design of a tracking system using n-delay two-degree-of-freedom control and its application to hard disk drives. Proceedings of the 1999 IEEE International Conference on Control Applications; Kohala Coast, Hawaii; 1999. p. 170-5.

13. Wang L, Yuan L, Chen BM, Lee TH. Modeling and control of a dual actuator servo system for hard disk drives. Proceedings of the 1998 International Conference on Mechatronics Technology; Hsinchu, Taiwan; 1998. p. 533-8.

14. Weerasooriya S, Low TS, Mamun AA. Design of a time optimal variable structure controller for a disk drive actuator. Proceedings of the International Conference on Industrial Electronics, Control, and Instrumentation; Hawaii; 1993. p. 2161-5.

15. Yamaguchi T, Soyama Y, Hosokawa H, Tsuneta K, Hirai H. Improvement of settling response of disk drive head positioning servo using mode switching control with initial value compensation. IEEE Trans Magn 1996; 32:1767-72.

16. Zhang DQ, Guo GX. Discrete-time sliding mode proximate time optimal seek control of hard disk drives. IEE Proc Contr Theory Appl 2000; 147:440-6.

17. Chang JK, Ho HT. LQG/LTR frequency loop shaping to improve TMR budget. IEEE Trans Magn 1999; 35:2280-82.

18. Hanselmann H, Engelke A. LQG-control of a highly resonant disk drive head positioning actuator. IEEE Trans Ind Electron 1988; 35:100-4.

19. Weerasooriya S, Phan DT. Discrete-time LQG/LTR design and modeling of a disk drive actuator tracking servo system. IEEE Trans Ind Electron 1995; 42:240-7.

20. Chen BM, Lee TH, Hang CC, Guo Y, Weerasooriya S. An H_∞ almost disturbance decoupling robust controller design for a piezoelectric bimorph actuator with hysteresis. IEEE Trans Contr Syst Technol 1997; 7:160-74.

21. Hirata M, Atsumi T, Murase A, Nonami K. Following control of a hard disk drive by using sampled-data H_∞ control. Proceedings of the 1999 IEEE International Conference on Control Applications; Kohala Coast, Hawaii; 1999. p. 182-6.

22. Li Y, Tomizuka M. Two degree-of-freedom control with adaptive robust control for hard disk servo systems. IEEE/ASME Trans Mechatron 1999; 4:17-24.

23. Hirata M, Liu KZ, Mita T, Yamaguchi T. Head positioning control of a hard disk drive using H_∞ theory. Proceedings of the 31st IEEE Conference on Decision and Control; Tucson, Arizona; 1992. p. 2460-1.

24. Kim BK, Chung WK, Lee HS, Choi HT, Suh IH, Chang YH. Robust time optimal controller design for hard disk drives. IEEE Trans Magn 1999; 35:3598-607.

25. Teo YT, Tay TT. Application of the l_1 optimal regulation strategy to a hard disk servo system. IEEE Trans Contr Syst Technol 1996; 4:467-72.

26. Chen R, Guo G, Huang T, Low TS. Adaptive multirate control for embedded HDD servo systems. Proceedings of the 24rd International Conference on Industrial Electronics Control and Instrumentation; Aachen, Germany; 1998. p. 1716-20.

27. McCormick J, Horowitz R. A direct adaptive control scheme for disk file servos. Proceedings of the 1993 American Control Conference; San Francisco, California; 1993. p. 346-51.

28. Weerasooriya S, Low TS. Adaptive sliding mode control of a disk drive actuator. Proceedings of Asia-Pacific Workshop on Advances in Motion Control; Singapore, 1993. p. 177-82.

29. Workman ML. Adaptive proximate time optimal servomechanisms [PhD dissertation]. Stanford University; 1987.

30. Mamun AA. Servo engineering. Lecture Notes; Department of Electrical & Computer Engineering, National University of Singapore; 2000.

31. Zhang JL. Electronic computer magnetic storing devices. Beijing: Military Industry Publishing House; 1981 (in Chinese).

32. Internet URL: www.mkdata.dk; www.storagereview.com; www.pcguide.com; www.storage.ibm.com; 2001.

33. Porter J. Disk drives' evolution. Presented at 100th Anniversary Conference: Magnetic Recording and Information Storage; Santa Clara, California; 1998.

34. Lin Z, Chen BM. Linear systems and control toolbox. Technical Report, Dept of Electrical and Computer Engineering, University of Virginia, Charlottesville (VA); 2000.

35. Ljung L. System identification: theory for the user. Englewood Cliffs (NJ): Prentice Hall; 1987.

36. Sinha NK, Kuszta B. Modeling and identification of dynamic systems. New York: Van Nostrand Reinhold Company; 1983.

37. Eykhoff P. System identification – parameter and state estimation. New York: John Wiley; 1981.

38. Hsia TC. System identification. Lexington (MA): Lexington Books; 1977.

39. Sage AP, Melsa JL. System identification. New York: Academic Press; 1971.

40. Rake H. Step response and frequency response methods. Automatica 1980; 16:519-26.

41. Eykhoff P. Trends and progress in system identification. New York: Pergamon Press; 1981.

42. Chen J, Gu G. Control oriented system identification: an H_∞ approach. New York: John Wiley & Sons; 2000.

43. Liu X, Chen BM, Lin Z. On the problem of general structural assignments of linear systems through sensor/actuator selection. Proceedings of the 39th IEEE Conference on Decision and Control; Sydney, Australia; 2000. p. 1979-84.

44. Chen BM. Robust and H_∞ Control. London: Springer; 2000.

45. Kailath T. Linear systems. Englewood Cliffs (NJ): Prentice Hall; 1980.

46. Luenberger DG. Canonical forms for linear multivariable systems. IEEE Trans Automat Contr 1967; 12:290-3.

47. Brunovsky P. A classification of linear controllable systems. Kybernetika (Praha) 1970; 3:173-187.

48. Sannuti P, Saberi A. A special coordinate basis of multivariable linear systems – Finite and infinite zero structure, squaring down and decoupling. Int J Contr 1987; 45:1655-704.

49. Saberi A, Sannuti P. Squaring down of non-strictly proper systems. Int J Contr 1990; 51:621-9.

50. Chen BM. A simple algorithm for the stable/unstable decomposition of a linear discrete-time system. Int J Contr 1995; 61:255-60.

51. Chen BM. On properties of the special coordinate basis of linear systems. Int J Contr 1998; 71:981-1003.

52. Rosenbrock HH. State-space and Multivariable Theory. New York: John Wiley; 1970.

53. MacFarlane AGJ, Karcanias N. Poles and zeros of linear multivariable systems: a survey of the algebraic, geometric and complex variable theory. Int J Contr 1976; 24:33-74.

54. Morse AS. Structural invariants of linear multivariable systems. SIAM J Contr 1973; 11:446-65.

55. Commault C, Dion JM. Structure at infinity of linear multivariable systems: a geometric approach. IEEE Trans Automat Contr 1982; 27:693-6.

56. Pugh AC, Ratcliffe PA. On the zeros and poles of a rational matrix. Int J Contr 1979; 30:213-27.

57. Verghese G. Infinite frequency behavior in generalized dynamical systems [PhD dissertation]. Stanford University; 1978.

58. Owens DH. Invariant zeros of multivariable systems: a geometric analysis. Int J Contr 1978; 28:187-98.

59. Moylan P. Stable inversion of linear systems. IEEE Trans Automat Contr 1977; 22:74-8.

60. Scherer C. H_∞-optimization without assumptions on finite or infinite zeros. SIAM J Contr Optimiz 1992; 30:143-66.

61. Chen BM, Weller SR. Mappings of the finite and infinite zero structures and invertibility structures of general linear multivariable systems under the bilinear transformation. Automatica 1998; 34:111-24.

62. Åström KJ, Wittenmark B. Computer controlled systems: theory and design. Englewood Cliffs (NJ): Prentice Hall; 1984.

63. Ziegler JG, Nichols NB. Optimum settings for automatic controllers. Trans ASME 1942; 64:759-68.

64. Ziegler JG, Nichols NB. Process lags in automatic control circuits. Trans ASME 1943; 65:433-44.

65. Franklin GF, Powell JD, Emami-Naeini A. Feedback control of dynamic systems. 3rd ed. Reading (MA): Addison-Wesley; 1994.

66. Anderson BDO, Moore JB. Optimal control: linear quadratic methods. Englewood Cliffs (NJ): Prentice Hall; 1989.

67. Fleming WH, Rishel RW. Deterministic and stochastic optimal control. New York: Springer-Verlag; 1975.

68. Kwakernaak H, Sivan R. Linear optimal control systems. New York: John Wiley; 1972.

69. Saberi A, Sannuti P, Chen BM. H_2 optimal control. London: Prentice Hall; 1995.

70. Doyle J, Glover K, Khargonekar PP, Francis BA. State space solutions to standard H_2 and H_∞ control problems. IEEE Trans Automat Contr 1989; 34:831-47.

71. Chen BM, Saberi A, Sannuti P, Shamash Y. Construction and parameterization of all static and dynamic H_2-optimal state feedback solutions, optimal fixed modes and fixed decoupling zeros. IEEE Trans Automat Contr 1993; 38:248-61.

72. Zhou K, Khargonekar P. An algebraic Riccati equation approach to H_∞-optimization. Syst Contr Lett 1988; 11:85-91.

73. Chen BM, Saberi A, Bingulac S, Sannuti P. Loop transfer recovery for nonstrictly proper plants. Contr Theory Adv Technol 1990; 6:573-94.

74. Zames G. Feedback and optimal sensitivity: Model reference transformations, multiplicative seminorms, and approximate inverses. IEEE Trans Automat Contr 1981; 26:301-20.

75. Limebeer DJN, Anderson BDO. An interpolation theory approach to H_∞ controller degree bounds. Linear Algebra Appl 1988; 98:347-86.

76. Doyle JC. Lecture notes in advances in multivariable control. ONR-Honeywell Workshop; 1984.

77. Francis BA. A course in H_∞ control theory. Lecture Notes in Control and Information Sciences; Berlin: Springer; 1987.

78. Glover K. All optimal Hankel-norm approximations of linear multivariable systems and their $\mathcal{L}_\infty$ error bounds. Int J Contr 1984; 39:1115-93.

79. Kwakernaak H. A polynomial approach to minimax frequency domain optimization of multivariable feedback systems. Int J Contr 1986; 41:117-56.

80. Kimura H. Chain-scattering approach to H_∞-control. Boston: Birkhäuser; 1997.

81. Zhou K, Doyle J, Glover K. Robust and optimal control. Englewood Cliffs (NJ): Prentice Hall; 1996.

82. Willems JC. Almost invariant subspaces: an approach to high gain feedback design – part I: almost controlled invariant subspaces. IEEE Trans Automat Contr 1981; 26:235-52.

83. Willems JC. Almost invariant subspaces: an approach to high gain feedback design – part II: almost conditionally invariant subspaces. IEEE Trans Automat Contr 1982; 27:1071-85.

84. Liu K, Chen BM, Lin Z. On the problem of robust and perfect tracking for linear systems with external disturbances. Int J Contr 2001; 74:158-74.

85. Chen BM, Lin Z, Liu K. Robust and perfect tracking of discrete-time systems. Automatica 2002; in press.

86. Athans M. A tutorial on LQG/LTR methods. Proceedings of American Control Conference; Seattle, Washington; 1986. p. 1289-96.

87. Chen BM. Theory of loop transfer recovery for multivarible linear systems [PhD dissertation]. Pullman (WA): Washington State University; 1991.

88. Chen BM, Saberi A, Sannuti P. A new stable compensator design for exact and approximate loop transfer recovery. Automatica 1991; 27:257-80.

89. Doyle JC, Stein G. Multivariable feedback design: concepts for a classical/modern synthesis. IEEE Trans Automat Contr 1981; 26:4-16.

90. Goodman GC. The LQG/LTR method and discrete-time control systems. Technical report. MIT (MA): Report No.: LIDS-TH-1392; 1984.

91. Kwakernaak H. Optimal low sensitivity linear feedback systems. Automatica 1969; 5:279-85.

92. Matson CL, Maybeck PS. On an assumed convergence result in the LQG/LTR technique. Proceedings of the 26th IEEE Conference on Decision and Control; Los Angeles, California; 1987. p. 951-2.

93. Niemann HH, Sogaard-Andersen P, Stoustrup J. Loop Transfer Recovery: Analysis and Design for General Observer Architecture. Int J Contr 1991; 53:1177-203.

94. Saberi A, Chen BM, Sannuti P. Loop transfer recovery: analysis and design. London: Springer; 1993.

95. Stein G, Athans M. The LQG/LTR procedure for multivariable feedback control design. IEEE Tran Automat Contr 1987; 32:105-14.

96. Zhang Z, Freudenberg JS. Loop transfer recovery for nonminimum phase plants. IEEE Trans Automat Contr 1990; 35:547-53.

97. Hu T, Lin Z. Control systems with actuator saturation: analysis and design. Boston: Birkhäuser; 2001.

98. Lin Z, Pachter M, Banda S. Toward improvement of tracking performance – nonlinear feedback for linear systems. Int J Contr 1998; 70:1-11.

99. Kirk DE. Optimal control theory. Englewood Cliffs (NJ): Prentice Hall; 1970.

100. Venkataramanan V, Chen BM, Lee TH, Guo G. A new approach to the design of mode switching control in hard disk drive servo systems. Control Engineering Practice; Submitted for publication.

101. Itkis U. Control systems of variable structure. New York: Wiley; 1976.

102. Yamaguchi T, Numasato H, Hirai H. A mode-switching control for motion control and its application to disk drives: Design of optimal mode-switching conditions. IEEE/ASME Trans Mechatron 1998; 3:202-9.

103. Salle JL, Lefschetz S. Stability by Liapunov's direct method. New York: Academic Press; 1961.

104. LaSalle J. Stability by Liapunov's direct method with applications. New York: Academic Press; 1961.

105. Chen BM, Lee TH, Peng K, Venkataramanan V. Composite nonlinear feedback control: theory and an application. 15th IFAC World Congress; Submitted for publication; 2001.

106. Weerasooriya S, Low TS, Huang YH. Adaptive time optimal control of a disk drive actuator. IEEE Trans Magn 1994; 30:4224-6.

107. Xiong Y, Weerasooriya S, Low TS. Improved discrete proximate time optimal controller of a disk drive actuator. IEEE Trans Magn 1996; 32:4010-2.

108. Mizoshita Y, Hasegawa S, Takaishi K. Vibration minimized access control for disk drives. IEEE Trans Magn 1996; 32:1793-8.

109. Yamaguchi T, Nakagawa S. Recent control technologies for fast and precise servo system of hard disk drives. Proceedings of the 6th International Workshop on Advanced Motion Control; Nagoya, Japan; 2000. p. 69-73.

110. Venkataramanan V, Peng K, Chen BM, Lee TH. Discrete-time composite nonlinear feedback control with an application in design of a hard disk drive servo system. Proceedings of the 40th IEEE Conference on Decision and Control; Orlando, Florida; 2001. in press.

111. Low TS, Guo W. Modeling of a three-layer piezoelectric bimorph beam with hysteresis. J Microelectromech Syst 1995; 4:230-7.

112. Tsuchiura KM, Tsukuba HH, Toride HO, Takahashi T. Disk system with sub-actuators for fine head displacement. US Patent No: 5189578; 1993.

113. Miu DK, Tai YC. Silicon micromachined SCALED technology. IEEE Trans Ind Electron 1995; 42:234-9.

114. Fan LS, Ottesen HH, Reiley TC, Wood RW. Magnetic recording head positioning at very high track densities using a microactuator based, two stage servo system. IEEE Trans Ind Electron 1995; 42:222-33.

115. Chang TP. Seismic response analysis of nonlinear structures using the stochastic equivalent linearization technique [PhD dissertation]. New York (NY): Columbia University; 1985.

116. Caughey TK. Derivation and application of the Fokker–Planck equation to discrete nonlinear dynamic systems subjected to white random excitation. J Acoust Soc Am 1963; 35:1683-92.

117. Crandall ST. Perturbation techniques for random vibration of nonlinear systems. J Acoust Soc Am 1963; 35:1700-05.

118. Lyon RH. Response of a nonlinear string to random excitation. J Acousti Soc Am 1960; 32:953-60.

119. Booton, Jr., RC. Nonlinear control systems with random inputs. IRE Trans Circuit Theory 1954; CT-1:9-18.

120. Aggarwal SK, Horsley DA, Horowitz R, Pisano AP. Microactuators for high density disk drives. Proceedings of American Control Conference; Albuquerque, New Mexico; 1997. p. 3979-84.

121. Ding, J, Tomizuka M, Numasato H. Design and robustness analysis of dual stage servo system. Proceedings of American Control Conference; Chicago; 2000. p. 2605-09.

122. Evans RB, Griesbach JS, Messner WC. Piezoelectric microactuator for dual stage control. IEEE Trans Magn 1999; 35:977-82.

123. Fan LS, Hirano T, Hong J, Webb PR, Juan WH, Lee WY, Chan S, Semba T, Imaino W, Pan TS, Pattanaik S, Lee FC, McFadyen I, Arya S, Wood R. Electrostatic microactuator and design considerations for HDD application. IEEE Trans Magn 1999; 35:1000-5.

124. Guo L, Chang JK, Hu X. Track-following and seek/settle control schemes for high density disk drives with dual-stage actuators. Proceedings of the 2001 IEEE/ASME International Conference on Advances Intelligent Mechatronics; Como, Italy; 2001. p. 1136-41.

125. Guo L, Martin D, Brunnett D. Dual-stage actuator servo control for high density disk drives. Proceedings of the 1999 IEEE/ASME International Conference on Advanced Intelligent Mechatronics; Atlanta, USA; 1999. p. 132-7.

126. Guo W, Weerasooriya S, Goh TB, Li QH, Bi C, Chang KT, Low TS. Dual stage actuators for high density rotating memory devices. IEEE Trans Magn 1998; 34:450-5.

127. Guo W, Yuan L, Wang L, Guo G, Huang T, Chen BM, Lee TH. Linear quadratic optimal dual-stage servo control systems for hard disk drives. Proceedings of the 24th Annual Conference of the IEEE Industrial Electronics Society; Aachen, Germany; 1998. p. 1405-10.

128. Hernandez D, Park SS, Horowitz R, Packard A. Dual-stage track-following design for hard disk drives. Proceedings of American Control Conference; San Diego, California; 1999. p. 4116-21.

129. Horsley DA, Hernandez D, Horowitz R, Packard AK, Pisano AP. Closed-loop control of a microfabricated actuator for dual-stage hard disk drive servo systems. Proceedings of American Control Conference; Philadelphia, USA; 1998. p. 3028-32.

130. Hu X, Guo W, Huang T, Chen BM, Discrete time LQG/LTR dual-stage controller design and implementation for high track density HDDs. Proceedings of American Control Conference; San Diego, California; 1999. p. 4111-5.

131. Kobayashi M, Horowitz R. Track seek control for hard disk dual-stage servo systems. IEEE Trans Magn 2001; 37:949-54.

132. Li Y, Horowitz R. Track-following controller design of MEMS based dual-stage servos in magnetic hard disk drives. Proceedings of the 2000 IEEE International Conference on Robotics and Automation; San Francisco; 2000. p. 953-8.

133. Mori K, Munemoto T, Otsuki H, Yamaguchi Y, Akagi K. A dual-stage magnetic disk drive actuator using a piezoelectric device for a high track density. IEEE Trans Magn 1991; 27:5298-300.

134. Schroeck SJ, Messner WC. On controller design for linear time-invariant dual-input single-output systems. Proceedings of American Control Conference; San Diego, Calfornia; 1999. p. 4122-6.

135. Semba T, Hirano T, Hong J, Fan LS. Dual-stage servo controller for HDD using MEMS microactuator. IEEE Trans Magn 1999; 35:2271-3.

136. Suthasun T, Mareels I, Mamun AA. System identification and control design for dual actuated disk drive. Proceedings of the 3rd International Conference on Control Theory & Applications; Pretoria, South Africa; 2001. in press.

137. Takaishi K, Imamura T, Mizasgita Y, Hasegawa S, Ueno T, Yamada T. Microactuator control for disk drive. IEEE Trans Magn 1996; 32:1863-6.

138. Peng K, Venkataramanan V, Chen BM, Lee TH. Design and implementation of a dual-stage actuator HDD servo system via composite nonlinear feedback approach. 15th IFAC World Congress; Submitted for publication.

139. Li Q, Hong OE, Hui Z, Mannan MA, Weerasooriya S, Ann MY, Shixin C, Wood R. Analysis of the dynamics of 3.5″ hard disk drive actuators. Technical Report. Data Storage Institute (Singapore); 1997.

140. Chang JK, Weerasooriya S, Ho HT. Improved TMR through a frequency shaped servo design. Proceedings of the 23rd International Conference on Industrial Electronics Control and Instrumentation; New Orleans, Louisiana; 1997. p. 47-51.

141. Li Z, Guo G, Chen BM, Lee TH. Optimal control design to achieve highest track-per-inch in hard disk drives. J Inform Stor Process Syst 2001; 3:27-41.

142. Sacks AH, Bodson M, Messner W. Advanced methods for repeatable runout compensation. IEEE Trans on Magn 1995; 31:1031-6.

143. Bodson M, Sacks A, Khosla P. Harmonic generation in adaptive feedforward cancellation schemes. IEEE Trans Automat Contr 1994; 39:1939-44.

144. Guo L, Tomizuka M. High speed and high precision motion control with an optimal hybrid feedforward controller. IEEE/ASME Trans Mechatron 1997; 2:110-22.

145. Kempf C, Messner W, Tomizuka M, Horowitz R. A comparison of four discrete-time repetitive control algorithms. IEEE Contr Syst Mag 1993; 13:48-54.

146. Guo L. A new disturbance rejection scheme for hard disk drive control. Proceedings of American Control Conference; Philadelphia; 1998. p. 1553-7.

147. Weerasooriya S. Learning and compensation for repeatable runout of a disk drive servo using a recurrent neural network. Technical Report; Magnetics Technology Center, National University of Singapore; 1995.

148. Jamg G, Kim D, Oh JE. New frequency domain method of nonrepeatable runout measurement in a hard disk drive spindle motor. IEEE Trans Magn 1999; 35:833-8.

149. Ohmi T. Non-repeatable runout of ball-bearing spindle-motor for 2.5″ HDD. IEEE Trans Magn 1996; 32:1715-20.

150. Abramovitch D, Hurst T, Henze D. An overview of the PES Pareto method for decomposing baseline noise sources in hard disk position error signals. IEEE Trans Magn 1998; 34:17-23.

151. Zeng S, Lin RM, Xu LM. Novel method for minimizing track seeking residual vibrations of hard disk drives. IEEE Trans Magn 2001; 37:1146-56.

152. Mah YA, Lin H, Li QH. Design of a high bandwidth moving-coil actuator with force couple actuation. IEEE Trans Magn 1999; 35:874-8.

153. McAllister JS. The effect of disk platter resonances on track misregistration in 3.5″ disk drives. IEEE Trans Magn 1996; 32:1762-6.

154. Hanselmann H, Mortix W. High-bandwidth control of the head positioning mechanism in a winchester disk drive. IEEE Contr Syst Mag 1987; 7:15-9.

155. Guo G. Lecture notes in servo engineering. Department of Electrical and Computer Engineering, National University of Singapore; 1998.

156. Weaver PA, Ehrlich RM. The use of multirate notch filters in embedded servo disk drives. Proceedings of American Control Conference; San Francisco, California; 1993. p. 4156-60.

INDEX